A Bayesian Spatial Model for Exceedances Over a Threshold

Fernando Ferraz do Nascimento and Bruno Sansó *

June 2, 2017

Abstract

Extreme value theory focuses on the study of rare events and uses asymptotic properties to estimate their associated probabilities. Easy availability of georeferenced data has prompted a growing interest in the analysis of spatial extremes. Most of the work so far has focused on models that can handle block maxima, with few examples of spatial models for exceedances over a threshold. Using a hierarchical representation, we propose a spatial process, that is obtained by perturbing a Pareto process. Our approach uses conditional independence at each location, within a hierarchical model for the spatial field of exceedances. The model has the ability to capture both, asymptotic dependence and independence. We use a Bayesian approach for inference of the process parameters that can be efficiently applied to a large number of spatial locations. We assess the flexibility of the model and the accuracy of the inference by considering some simulated examples. We illustrate the model with an analysis of data for temperature and rainfall in California.
Key Words: Generalized Pareto distribution, Max-Stable process, Bayesian hierarchical model, Spatial extremes, MCMC, Asymptotic Dependence.

1 Introduction

The statistical analysis of extreme values focuses on inference for rare events that correspond to the tails of probability distributions. As such, it is a key ingredient in the assessment of the risk of phenomena that can have strong societal impacts like floods, heat waves, high concentration of pollutants, crashes in the financial markets, among others. The fundamental challenge of extreme value theory (EVT) is to use information, collected over limited periods of time, to extrapolate to long time horizons. This is possible thanks to theoretical results that give asymptotic descriptions of the probability distributions of extreme values. The

*Fernando Ferraz do Nascimento (`fernandofn@soe.ucsc.edu`) is Postoctoral Researcher, and Bruno Sansó (`bruno@ams.ucsc.edu`) is Professor, Department of Applied Mathematics and Statistics, University of California, 1156 High Street, Mail Stop: SOE 2, Santa Cruz, CA-95064, USA .

development of EVT dates back at least to Fisher and Tippett (1928). Inferential methods for the extreme values of univariate observations are well established and software is widely available (see, for example, Coles, 2001). The most traditional approach to analyze extreme data for one variable, is to consider its maxima over a given period of time. As an example, we can consider the annual maxima of daily temperature at a given location. The results in von Mises (1954) and Jenkinson (1955) show that the distribution of the block maxima, as the number of observations go to infinity, belongs to the family of generalized extreme value (GEV) distributions. As the density of such family is readily available, likelihood-based methods can be used to estimate the three parameters that characterize the members of the family. This method relies on a drastic reduction of the original data to a small set of block maxima. An alternative, that uses additional information from the data, is to fix a threshold, say, u, and obtain the exceedances over that threshold. Pickands (1975) shows that, when $u \to \infty$, the exceedance distribution converges to the generalized Pareto distribution (GPD). This asymptotic result is used for inference, after setting a high threshold and filtering the original sample with respect to it.

It is typical of environmental data to be collected over networks of geographically scattered locations. In these cases, an extension of the geostatistical methods used for inference on spatial fields (see, for example, Banerjee et al., 2004) is needed to infer the joint distribution of extremes at different locations. This requires an extension of the EVT to multidimensional variables and, more generally, to stochastic processes indexed in space. Inference for multivariate block maxima relies on multivariate extreme value distributions that are based on the notion of max-stability. The work of Pickands (1981), Coles and Tawn (1991) and Heffernand and Tawn (2004) are some examples of the use of these methods. For exceedance over a threshold in multivariate settings, the work of Rootzen and Tajvidi (2006) defines the multivariate generalized Pareto distribution. Further analysis of these classes of distributions is presented in Falk and Guillou (2008). In this context, Michel (2008) provides a detailed discussion of different inferential approaches.

To tackle the associations that arise when considering observations that are collected at different spatial locations it is natural to consider hierarchical models. These are based on assuming that the block maxima at a given site follow a GEV whose location, scale and shape parameters depend on that site. The second level of the hierarchy consists of assuming that such parameters correspond to spatial random fields. Geostatistical models are then used to describe their variability. Examples of this approach are found in Huerta and Sansó (2007) and Sang and Gelfand (2009). For the exceedances approach, Cooley et al. (2007) develop a model where the scale parameter of the GPD distribution varies spatially. The hierarchical approach is very appealing computationally. It has been criticized, though, for not adequately capturing the spatial dependence structure of complex fields, like the ones that correspond to rainfall. Comprehensive summaries of the methods used for spatial extremes can be found in Cooley et al. (2012) and Davison and Gholamreazee (2012), which include a thorough list of relevant references.

1.1 Max-stable processes

A fundamental concept in EVT is that of max-stability. A distribution G has the max-stable property if $G^n(y) = G(A_n y + B_n)$ for some constants A_n and B_n. In other words, a collection of random variables $Y_1, \ldots, Y_n$, i.i.d. from G is such that the distribution of its maximum is also G. The central role of GEV distributions in EVT is due to the fact that they correspond to the only family that satisfies the max-stable property. Max-Stable processes represents a infinite dimensional extension of multivariate extreme value theory. According to Smith (1990), they form a natural class when block maxima are observed at each site of a spatial process. Following Huser and Davison (2013) we say that a spatial process $Y(x)$ defined for $x \in \mathcal{X}$ is *max-stable* if for any finite set $\mathcal{D} = \{x_1, ..., x_D\} \subset \mathcal{X}$, and any function defined on $\mathcal{D}$ we have that

$$Pr(Y(x)/n \leq y(x),\ x \in \mathcal{D})^n = Pr(Y(x) \leq y(x), x \in \mathcal{D})$$

for all integers n. As mentioned above in, the univariate case, the family of GEV distributions is the only max-stable class. Thus, the marginals of a max-stable process must be GEV. These can be transformed to the unit Fréchet distribution, implying that, without loss of generality, $Pr(Y(x) \leq y) = \exp(-1/y), y > 0$. For D different sites we have that $P(Y(x_1) \leq y_1, ..., Y(x_D) \leq y_D) = \exp\{-V(y_1, ..., y_D)\}$, where the function V measures dependence among the different sites. For $y_i = y$ for all i, we have that $P(Y(x_1) \leq y, ..., Y(x_D) \leq y) = \exp\{-1/y\}^{V(1,...,1)}$. Letting $\theta = V(1, ..., 1)$, we have that $\theta = D$ implies complete independence and $\theta = 1$ implies complete dependence. A drawback of the max-stable processes that have been proposed in the literature is that, typically, only the bivariate marginals have closed form expressions. As a consequence, standard likelihood-based approaches to inference are not practical. The most common alternative is the use of pairwise log-likelihoods (Padoan et al., 2010; Cooley et al., 2010). Spatial analyses of exceedences over a threshold are presented in Huser and Davison (2013) and Jeon and Smith (2012), using the relationship between the densities of the GPD and GEV to build pairwise likelihoods for the exceedances. A restriction of the pairwise likelihood is that the variability of pairwise estimator is usually underestimated (Pauli et al., 2011). More recent developed, based on leveraging the link between max-stable processes and Poisson processes, are found in Wadsworth and Tawn (2014) and Thibaud and Opitz (2015). In those papers, a full likelihood approach is developed for some specific classes of models. The use of censoring, though, requires dealing with non-trivial normalizing constants. Reich and Shaby (2012) extend the approach in Stephenson (2009) to a spatial process. The method is based on building a hierarchical model representation of max stable process that generalizes the multivariate extreme value distribution with asymmetric logistic dependence function.

In this paper, we propose a hierarchical Bayesian model for excedances in a spatial domain using a process that results from perturbing a generalized Pareto process (Ferreira and de Haan, 2014). The hierarchical approach is based on conditional independence, which allows for efficient inference. Moreover, the model allows for asymptotic dependence as well as independence between any two points. Section 2 presents the definition of the generalized Pareto process, and discusses the difficulties to obtain a viable estimation method. We then introduce our proposed model and develop the estimation procedure. Section 3 present some aspects and properties to obtain conditional dependence, and what's the main approaches to model the GPP under this feature. Section 4 presents how to perform the inference considering a perturbed GPP, and the toward to using that by a bayesian paradigm and the algorithm. Section 5 presents some numerical simulations of the proposed model using different parameter configurations. Section 6 presents an illustration on data for winter temperature and rainfall in California during the last five years. Finally, Section 7 discusses the results obtained with the proposed model, as well as possible future extensions.

2 The Pareto process

To perform inference on the distribution of the exceedances over a threshold, in a univariate setting, consider a random variable Y and a threshold u, and consider the $F_u(y) = Pr(Y \leq y - u | Y > u)$. Then, for large enough u, and for the number of observations on Y tending to infinity, $F_u(y)$ can be approximated by a properly scaled GPD whose cumulative distribution is given as

$$H(y|\mu, \sigma, \gamma) = \begin{cases} 1 - \left(1 + \gamma \frac{(y-\mu)}{\sigma}\right)_+^{-1/\gamma}, & \text{if} \quad \gamma \neq 0 \\ 1 - \exp\{-(y-\mu)/\sigma\}, & \text{if} \quad \gamma = 0 \end{cases}, \tag{1}$$

In the univariate case, the relationship between GEVs and GPDs is made explicit by observing that, if G is a GEV distribution, then $H(y) = 1 + \log(G(y))$, for all y such that $\log(G(y)) \in [-1, 0]$. In Michel (2008) this relationship is used to generalize the definition of a GPD distribution to the multivariate setting.

An infinite dimensional generalization of the GPD is given by the Pareto process proposed in Ferreira and de Haan (2014). A constructive definition of a simple Pareto process is given as follows:

Definition 1. *Let $C_+(S)$ the space of continuous non-negative functions in S, a compact subset of $\mathbb{R}^d$. Let $W(s)$ be a stochastic process in $C_+(S)$ and w_0 a positive constant. Then $W(s) = Y\theta(s)$, $\forall s \in S$ is a simple Pareto Process (SPP) if*

1. *θ is a stochastic process in $C_+(S)$ with $\sup_s \theta(s) = w_0$ and $E(\theta(s)) > 0$, $\forall s \in S$,*
2. *Y is a standard Pareto random variable,*
3. *Y and θ are independent.*

This definition provides a very simple constructive approach to generalize the univariate GPD to an infinite-dimensional setting. In fact, a Pareto process can be obtained from a bounded process and a standard Pareto random variable. We will refer to Y as the radius of W and to θ as its spectral process. The relevance of the Pareto process for a peaks over threshold approach to spatial extremes is due to the following results, stated in Theorem 3.2 of Ferreira and de Haan (2014):

$$\lim_{t \to \infty} Pr(T_t X(s) \in A | \sup_{s \in S} T_t X(s) > 1) = Pr(W \in A), \tag{2}$$

where W is a SPP. Here $X(s)$ is a process in the domain of attraction of a max-stable process, and

$$T_tX(s) = \left(1 + \gamma(s)\frac{X(s) - b_t(s)}{a_t(s)}\right)_+^{1/\gamma(s)}$$

is the standardized version of $X(s)$, with a_t, b_t and γ continuous functions in S, and $a_t(s) > 0, \forall s$. The condition that $\sup_{s\in S} T_tX(s) > 1$ is equivalent to $\sup_{s\in S} X(s) > b_t(s)$. Thus, $b_t(s)$ can be seen as a sequence of thresholds.

Can these results be used in order to perform likelihood based inference about the parameters of a Pareto Process? We start by assuming that for a large enough t the limit in (2) is well approximated and that the threshold and scale processes are fixed at $\mu(s)$ and $\sigma(s)$ respectively. We then condition on the event $T_tX(s) \in B$, where $B = \{f \in C_+(S) : \sup_{s\in S} f(s) > 1\}$. That is, we focus on the observed trajectories of the process $X(s)$ whose supremum, after transformation, is larger than 1. These correspond to observations, say, $x_j(s_i)$, for $j = 1, \ldots, m$ trajectories, observed at $i = 1, \ldots, n$ locations. Then $Tx_j(s_i) = Y_j\theta_j(s_i)$, according to the constructive definition of a SPP. Conditioning on Y_j, and postulating a flexible model for $\theta(s)$ for which the finite dimensional distribution is tractable, it is possible to build a hierarchical model for inference on the process of exceedances.

While the approach described in the previous paragraph fits naturally within a hierarchical model formulation, it is subject to two substantial problems. The first one is that, when data are collected on a finite number of locations it is difficult to assess the occurrence of the event $T_tX(s) \in B$. More commonly, instead of B, we consider $B_1 = \{f \in C_+(S) : \bigvee_{i=1}^n f(s_i) > 1\}$, or $B_2 = \{f \in C_+(S) : f(s_i) > 1, i = 1, \ldots, m\}$, for some $m \leq n$ or $B_3 = \{f \in C_+(S) : \bigwedge_{i=1}^n f(s_i) > 1\}$, where $\vee$ denotes the maximum, and $\wedge$ the minimum. We notice that, in all three cases, $B_i \subseteq B$, and that $B_3 = B_2$ when $m = n$. Following Equation (4.1) in Ferreira and de Haan (2014), we obtain

$$\lim_{t\to\infty} Pr(T_tX(s) \in A | T_tX(s) \in B_i) = \frac{Pr(W \in A \cap B_i)}{Pr(W \in B_i)}.$$

This implies that the distribution of W needs to be restricted at the locations where the thresholds are exceeded. Thus, inference in this case requires accounting explicitly for a normalizing constant. Thus, as in the case of Wadsworth and Tawn (2014) and Thibaud and Opitz (2015), this method requires dealing with an intractable likelihood. The second problem is that the transformation $Tx_j(s_i)$ induces a set constraints on the parameters that are used to describe the processes $\sigma(s)$ and $\gamma(s)$, within a hierarchical model. An effective exploration of the resulting parameter space, for realistically large problems, is a very challenging task. In fact our attempts to implement this approach have not been successful.

3 Conditional independence revisited

Hierarchical models for spatial fields are often built on the assumption that the marginal distributions corresponding to different locations are conditional independent. Cooley et al.

(2007), for example, consider that exceedances over a field threshold $\mu(s)$ follow, at each location, the distribution in Equation (1), with spatially varying parameters $\gamma(s)$ and $\sigma(s)$. Then impose spatial dependence by assuming that such parameters correspond to spatial processes. Using this approach for spatial fields of extreme values usually leads to processes that have asymptotic independence. This is undesirable as it implies that the conditional probability that the process at one location exceeds a high value, given that it has exceeded that value at another location, tends to zero. Thus, the model has a built in underestimation property. Our aim is to consider an approach that uses conditional independence, but provides a range of extremal dependence properties. For clearness, in what follows we limit ourselves to a bivariate vector, say, $W = (W_1, W_2)$. We assume that both components have the same marginal distribution and thus consider the coefficient

$$\chi_{ij} = \lim_{w \to \infty} Pr(W_i > w | W_j > w).$$

We note in passing that, while our interest focuses on spatial processes, the models considered in the following sections have an interest for multivariate extreme value settings.

3.1 Bivariate Pareto distribution

Suppose that W follows a bivariate Pareto distribution with identical univariate marginals. Then $W = Y \times (\theta_1, \theta_2)'$, where θ_1 and θ_2 are bounded positive random variables with the same univariate marginal distributions. For simplicity we assume that the bound is 1. Then

Proposition 1.

$$\chi_{12} = \frac{E(\theta_1 \wedge \theta_2)}{E(\theta_2)}. \tag{3}$$

This result is established following Ferreira and de Haan (2014). It implies that a bivariate Pareto distribution has asymptotic dependence, asymptotic independence can not be achieved except for degenerate cases, and the asymptotic behavior is determined by the distributions of the vector θ.

3.2 Independent bivariate Pareto

Consider a modification of the above representation to $W_k = Y_k \theta_k$, where Y_k is a standard Pareto random variable.

Proposition 2. *If Y_1 and Y_2 are independent,*

$$\chi_{12} = \lim_{w \to \infty} \frac{Pr(Y_1 > w/\theta_1, Y_2 > w/\theta_2)}{Pr(Y_2 > w/\theta_2)} = \lim_{w \to \infty} \frac{1}{w} \frac{E(\theta_1 \theta_2)}{E(\theta_2)} = 0.$$

The result is an immediate consequence of independence of Y_1 and Y_2, and the fact that $\theta_k \leq 1$. Thus, regardless of the dependence between θ_1 and θ_2, the distribution of W has asymptotic independence.

3.3 Conditional independent bivariate Pareto

Consider a modification of the previous model where the random variables (Y_1, Y_2) have a joint distribution given as

$$p(y_1, y_2) \propto \left(\prod_{i=1}^{2} y_i \right)^{-1} \left(b + \sum_{i=1}^{2} \log y_i \right)^{-2+a}.$$

This distribution is motivated by the idea of considering radial variables having a dependent joint distribution. In fact, the distribution is obtained from a continuous mixture of conditionally independent generalized Pareto, as

$$p(y_1, y_2) = \int_0^\infty \prod_{i=1}^{2} p(y_i|\beta)p(\beta)d\beta$$

where $Pr(Y_i > y|\beta) = 1/y^\beta$, and $\beta \sim Ga(a, b)$. That is, the bivariate marginal of (Y_1, Y_2) corresponds to assuming conditional independence for two generalized Pareto random variables, and then postulating a common distribution for their shape parameter. In this case we have:

Proposition 3.

$$\chi_{12} = \lim_{w \to \infty} \frac{(b + \log w)^a}{(b + 2 \log w)^a} = \left(\frac{1}{2} \right)^a.$$

This result is established by expressing the numerator and the denominator as integrals with respect to β, and then using the dominated convergence theorem. It implies that the distribution of W has asymptotic dependence, with $0 < \chi_{12} < 1$ depending on the value of a. Thus, the introduction of a dependence structure in the distribution of the radial variables induces the desired strong dependence property. Nevertheless, asymptotic dependence is regulated by a only, and not by the dependence in the spectral distribution. This is somewhat unappealing.

3.4 Perturbed bivariate Pareto

An additional way to develop a model based on conditional independence that produces asymptotic dependence is to consider a bivariate Pareto vector that is perturbed by a multiplicative generalized Pareto noise. Thus, $W_k = V_k^\xi Y \theta_k$, where V_k are independent standard Pareto random variables. We can think of this as a nugget, or observational error. The bivariate Pareto distribution is recovered when $\xi \to 0$. For the asymptotic dependence we have the following results:

Proposition 4. *If* $\xi < 0$ *then*

$$\chi_{12} = \frac{E(V_1^\xi \theta_1 \wedge V_2^\xi \theta_2)}{E(V_2^\xi \theta_2)},$$

where the expectation is taken with respect to V_2, V_2, θ_1 *and* θ_2.

If $\xi > 0$ *then*

$$\chi_{12} = \begin{cases} \frac{(\xi-1)(E\theta_1^\xi\theta_2^\xi+2\xi-1)}{\xi(2\xi-1)E\theta_2^\xi} & \xi < 1 \\ 0 & \xi \geq 1 \end{cases},$$

where the expectation is taken with respect to θ_1 *and* θ_2.

The result for $\xi < 0$ is established following calculations similar to the ones required for Proposition 1, recalling the fact that $1/(V_k^\xi \theta_k) > 1, k = 1, 2$. The result for $\xi > 0$ uses the distribution of the product $V_k^\xi Y$. We notice that, for $\xi < 0$, χ_{12} corresponds to a perturbed version of the expression in Equation (3). For $\xi > 0$ we observe that $\xi = 1$ corresponds to a break point. Generally, depending on the value of ξ this model has the flexibility of capturing asymptotic dependence or independence. Moreover, the value of χ_{12} depends on both, the exponent of the perturbation and the spectral distribution of the bivariate Pareto. From an inferential point of view, it is important to notice that, conditional on Y and θ_k, $W_k/Y\theta_k$ follows a generalized Pareto distribution. Thus, this model provides the inferential advantages of conditional independence, as does the model in Section 3.3, but, additionally, it generates a wide set of possible extreme dependence behaviors. Moreover, it corresponds to a perturbation of the theoretical limit of distributions in a max-stable domain of attraction. A natural assumption when considering noisy data.

4 Inference for perturbed Pareto processes

The discussion in the previous section indicates that it is possible to model a spatial process that has an asymptotic dependence behavior similar to that of a Pareto process, using a conditional independence assumption. This can be achieved by perturbing a Pareto process $W(s)$ with a multiplicative random noise, say $V(s)^\xi$, that is obtained as a power of independent Pareto random variables. Thus we consider the process $H(s) = V(s)^\xi W(s) = V(s)^\xi Y\theta(s)$, for Y and $\theta(s)$ defined as in Section 2, with $\omega_0 = 1$. The perturbation, that can be interpreted as observational error or nugget, allows to consider each location separately, simplifying the likelihood based inference.

Consider observations $x_j(s_i)$ as in Section 2. After the standardization $Tx_j(s_i)$, we assume that such observations are realization of $H(s_i)$, that correspond to perturbations of a simple Pareto process. To obtain the likelihood we notice that, if V and Y are both distributed as standard Pareto, then, for $u > 0$, and conditional on θ,

$$Pr(V^\xi Y > u/\theta) = \begin{cases} 1 & \xi > 0, 0 < u \leq \theta \\ \frac{\xi}{\xi-1}(\theta/u)^{1/\xi} - \frac{1}{\xi-1}(\theta/u) & \xi > 0, \xi \neq 1, u > \theta \\ (\log u - \log\theta + 1)(\theta/u) & \xi = 1, u > \theta \\ \frac{1}{1-\xi}(\theta/u) & \xi < 0, u > \theta \\ 1 + \frac{\xi}{1-\xi}(\theta/u)^{1/\xi} & \xi < 0, 0 < u \leq \theta \end{cases}. \quad (4)$$

From (4) we obtain the density of the random variable $H = V^{\xi} Y \theta$, conditional on θ and on the event $H > 1$, as

$$f(u) = \begin{cases} (\xi\theta^{1/\xi-1} - 1)^{-1}((\theta/u)^{1/\xi-1} - 1)u^{-2} & \xi > 0, \xi \neq 1 \\ (\log u - 1)(1 - \log\theta)^{-1}u^{-2} & \xi = 1 \\ u^{-2} & \xi < 0 \end{cases}.$$

Changing variables we have that the contribution to the likelihood for an observation $x_j(s_i)$ that is above the threshold $\mu(s_i)$ is given as

$$f_T(x_j(s_i)) = \frac{Tx_j(s_i)^{-1}}{\sigma(s_i)(1 + \frac{\gamma}{\sigma(s_i)}(x_j(s_i) - \mu(s_i))} \begin{cases} \frac{((\theta(s_i)/Tx_j(s_i))^{1/\xi-1}-1)}{(\xi\theta(s_i)^{1/\xi-1}-1)} & \xi > 0, \xi \neq 1 \\ \frac{\log Tx_j(s_i)-1}{(1-\log\theta(s_i))} & \xi = 1 \\ 1 & \xi < 0 \end{cases}.$$

To complete the model we need to specify the processes $\theta(s), \sigma(s)$ and $\gamma(s)$. A natural choice for $\theta(s)$, that is required to be a bounded process, is a logistic transformations of a Gaussian process. A possibility is to let $u(s) = \sum_{l=1}^{L} A_l k_l(s)$, where $A_l \sim N(0, \tau^2)$, and $k_l(s), l = 1, \ldots, L$ is a collection of kernels. Thus $(u(s_1), \ldots, u(s_n)) \sim N_n(0, KK')$, where $K_{ij} = k_j(s_i)$. We then let $\theta(s) = \exp(-u(s))/(1 + \exp(-u(s))$.

4.1 Posterior distribution

A hierarchical specification of the model is obtained by denoting as $B_j(s_i)$ a latent variable that flags the crossing of the thresholds. Thus, $B_j(s_i)$ equals 1 if $TX_j(s_i) > 1$ and 0 otherwise. We have that $B_j(s_i) \sim Bernoulli(p(s_i))$, where $p(s_i)$ can be obtained from Equation (4), letting $u = 1$. The likelihood can then be written as

$$L(X|\Theta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} f_T(x_j(s_i))^{I(Tx_j(s_i)>1)} p(s_i)^{B_j(s_i)} (1 - p(s_i))^{1-B_j(s_i)}$$

where I is the indicator function, and Θ denotes the collection of ξ and all the parameters that define the processes $\theta(s), \sigma(s), \gamma(s)$. A posterior distribution is obtained by considering priors on the different components of θ. This is explored using a Monte Carlo method based on an adaptive Metropolis approach.

4.2 Estimation of return levels

An important product of the analysis of extreme data is the estimation of the quantiles of the distribution that correspond to rare events. These are traditionally given as return levels. The return level t, denoted by r_t, is the value of the quantile $1 - 1/t$ of the distribution, i.e., every t years, it is expected that at least once, the value of the variable of interest would be equal or higher than q_t. For the GPD distribution the return level is given by

$$r_t = \mu + \frac{\sigma}{\xi}\left((1/t)^{-\xi} - 1\right).$$

The above formula for the return level assumes that all observations are above a threshold. Considering the results obtained from the hierarchical model conditioned in Y, our proposed model incorporates information about the probability of crossing the threshold. Thus, we need to weigh the return level according to such probability to obtain the correct quantile. According to our model $P(W_j(s) > u) = p_{u(s)}$. Thus, the return level for t is given by

$$r_t(s) = \mu(s) + \frac{\sigma'(s)}{\gamma'(s)} \left(\left(\frac{1/t}{p_{\mu(s)}} \right)^{-\gamma'(s)} - 1 \right). \tag{5}$$

5 Simulations

To explore the characteristics of the proposed model, as well as the ability of our estimation approach to recover the true parameter values, we conduct a series of simulations. The simulation were performed on a 30x30 grid. For each site, we generated 500 replicates. Thus, $n = 900$ and $m = 500$. We used $L = 30$ regularly spaced kernel knots. We set the threshold to $\mu = 50$, and $\sigma = 10$, both constant in space. We considered three different values of the perturbation parameter $\xi = \{-0.3, 0, 0.3\}$, and considered two different space-constant values of the shape parameter $\gamma = \{-0.2, 0.5\}$. Notice that $\xi = 0$ correspond to the simulation of an actual Pareto process. We simulated using Gaussian kernels $k_l(s) = \exp(-||s - s_l||^2/b_w)/\sqrt{2\pi b_w^2}$, with $b + w = 5$. The steps to generate the simulations are as follows: (1) Recalling that $u(s) = \sum_{l=1}^{L} A_l k_l(s)$ generate $A_l \sim N(0, \tau), l = 1, \ldots, L = 30$ independently, calculate $u(s)$ for all 900 gridded locations, and set $\theta(s) = \exp(-u(s))/(1 + \exp(-u(s))$; (2) Generate Y_j iid $GPD(1, 1, 1)$ for $j = 1, \ldots, 500$; (3) Generate $V_j(s)$ iid $GPD(1, 1, 1)$, for $j = 1, \ldots, 500$ and each location s; (5) For each s compute $H_j(s) = Y_j \theta(s) V_j(s)^{\xi}, j = 1, \ldots, 500$; (7) For each j and each s, compute the transformation $x_j(s) = \mu + (\sigma/\xi) \left(H_j(s)^{\xi} - 1\right)$.

We fit our model using the following priors: for τ and b_w we used a Gamma(0.001,0.001); The joint prior to shape and scale parameters σ and γ was the Jeffreys prior proposed in Castellanos and Cabras (2007); For ξ we used a Normal with mean 0 and low precision. Table 1 shows the posterior mean for the different configurations of the parameter values, together with 95% credibility intervals. The results in the table indicate that the estimation of the model parameters is performed with high level of accuracy for all configurations. We notice that results are better when the value of ξ than when it is positive. Interestingly, we observe that the model parameters are properly estimated even in the case of $\xi = 0$, which is a singularity for the proposed model. This illustrates the ability of the perturbed Pareto model to infer the structure of an actual Pareto process.

Figures 1 and 2 show the fields of posterior expectations for the $t = 20$ year return levels corresponding to two different combinations of parameter values. We calculate the true 20-year return level values using Equation (4.2), and compare the true ones in the left panels to the estimated ones in the right panels. We can see that the estimation recovers the spatial distribution of the true returns, with some underestimation of the largest values. Figure

Table 1: Posterior mean and 95% credibility intervals. T- True, M - Posterior mean, CI - Credibility interval.

	$\gamma = -0.2$								
	$\xi = -0.3$			$\xi = 0$			$\xi = 0.3$		
	T	M	CI	T	M	CI	T	M	CI
γ	-0.2	-0.209	$(-0.215; -0.204)$	-0.2	-0.211	$(-0.216; -0.204)$	-0.2	-0.224	$(-0.228; -0.220)$
ξ	-0.3	-0.370	$(-0.402; -0.328)$	0	-0.067	$(-0.075; -0.060)$	0.3	0.208	$(0.204; 0.210)$
σ	10	9.46	$(9.38; 9.55)$	10	9.55	$(9.46; 9.62)$	10	9.92	$(9.84; 9.98)$
b_w	5	4.86	$(4.73; 5.03)$	5	4.89	$(4.77; 4.99)$	5	4.86	$(4.80; 4.92)$
τ	1	1.36	$(0.89; 2.00)$	1	1.49	$(1.06; 2.09)$	1	1.35	$(0.95; 1.89)$
	$\gamma = 0.5$								
	$\xi = -0.3$			$\xi = 0$			$\xi = 0.3$		
	T	M	CI	T	M	CI	T	M	CI
γ	0.5	0.427	$(0.417; 0.439)$	0.5	0.430	$(0.419; 0.439)$	0.5	0.427	$(0.418; 0.438)$
ξ	-0.3	-0.309	$(-0.405; -0.369)$	0	-0.062	$(-0.068; -0.057)$	0.3	0.216	$(0.211; 0.220)$
σ	10	9.54	$(9.42; 9.66)$	10	9.64	$(9.53; 9.75)$	10	10.10	$(9.98; 10.23)$
b_w	5	4.76	$(4.61; 4.89)$	5	4.88	$(4.80; 4.97)$	5	5.00	$(4.92; 5.08)$
τ	1	1.41	$(0.92; 2.03)$	1	1.45	$(1.01; 1.98)$	1	1.42	$(0.99; 2.06)$

2 shows that the perturbed model is able to estimate the true returns corresponding to a Pareto process as it present the case where the simulations correspond to $\xi = 0$.

In Figures 3 and 4 we explore the effect of different parameter values on the estimation of the return levels. We notice from Figure 3 that when γ increase, the spatial dependence is unchanged, but the magnitude of the returns is increased. This is to be expected, as the underlying spatial process is unchanged, and γ controls the tails of the distributions. Figure 4 explores the effect of the perturbation parameter. We observe that when the true field correspond to a Pareto process perturbed with a positive ξ, the estimation of the spatial dependece is fuzzier. Overall, our simulation study shows that estimation in this case is harder than in the $\xi < 0$ case.

6 California temperature and rainfall

As illustrative examples we analyze two datasets. The first one consists of data for minimum daily temperature at 665 locations in the State of California, from 2012 to 2014. The second dataset consists of the daily accumulated volume of precipitation, in the same period in California, at 992 locations. We limited our analysis to the winter period, and included only observations for the months of December, January and February. For the analysis of the minimum daily temperature we considered the transformation $x = max(y) - y$, where y is the variable representing temperatureu, and the maximum is taken over the whole available record. This results in a positive variable whose exceedances over a threshold correspond to

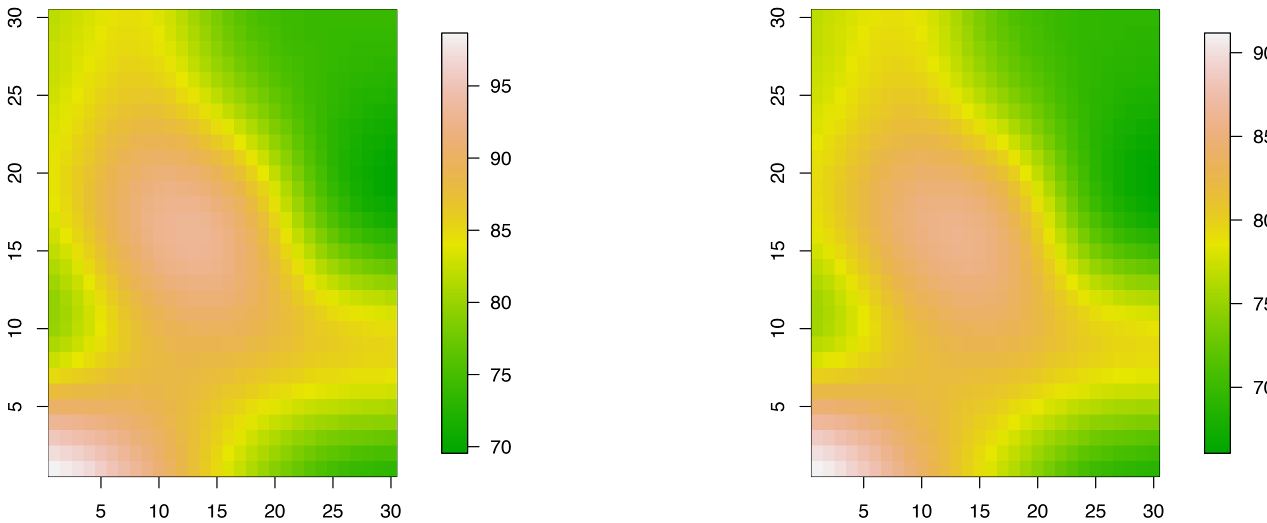

Figure 1: 20-years return levels in space for simulated data corresponding to $\gamma = 0.5$, $\xi = -0.3$ and $\tau = 10$. Left: True returns; Right: estimated returns.

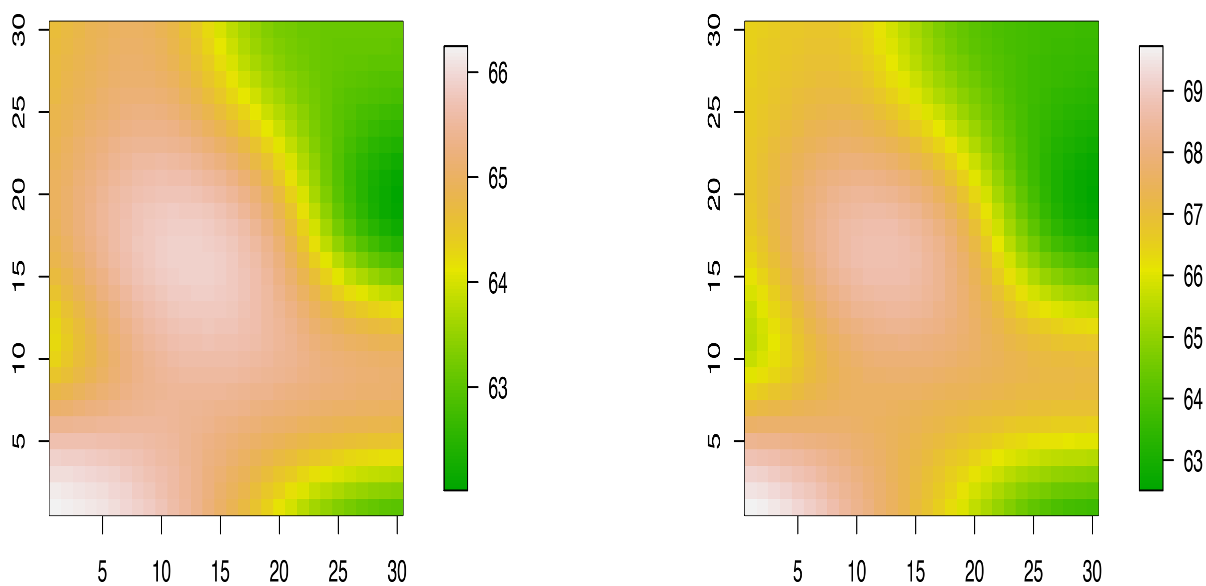

Figure 2: 20-years return levels in space for simulated data corresponding to $\gamma = -0.2$, $\xi = 0$ and $\tau = 10$. Left: True returns; Right: estimated returns.

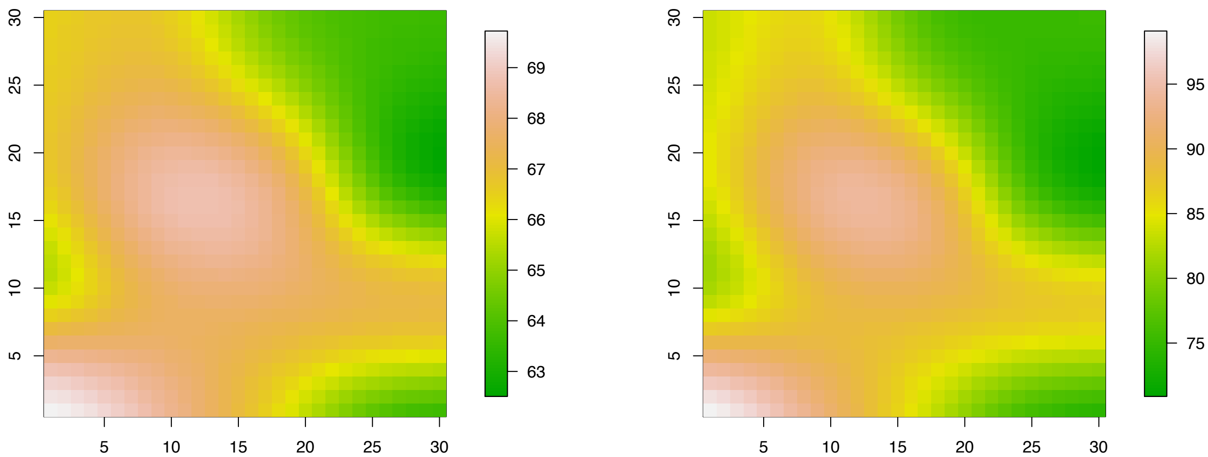

Figure 3: 20-year return levels in space for the simulated data with $\xi = 0$ and $\tau = 10$. Left: $\gamma = -0.2$; Right: $\gamma = 0.5$.

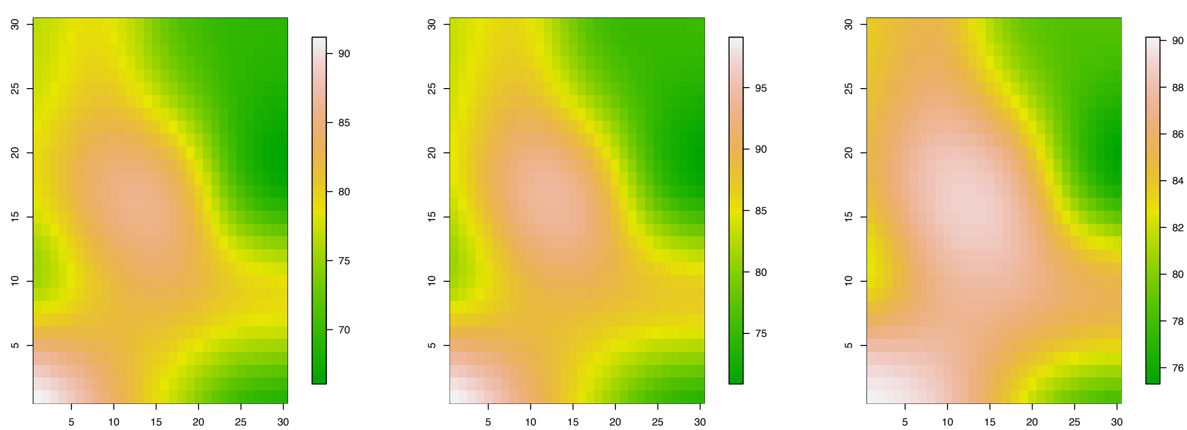

Figure 4: 20-year return levels in space for the simulated data with $\gamma = 0.5$ and $\tau = 10$. Left: $\xi = -0.3$, Center: $\xi = 0$. Right: $\xi = 0.3$.

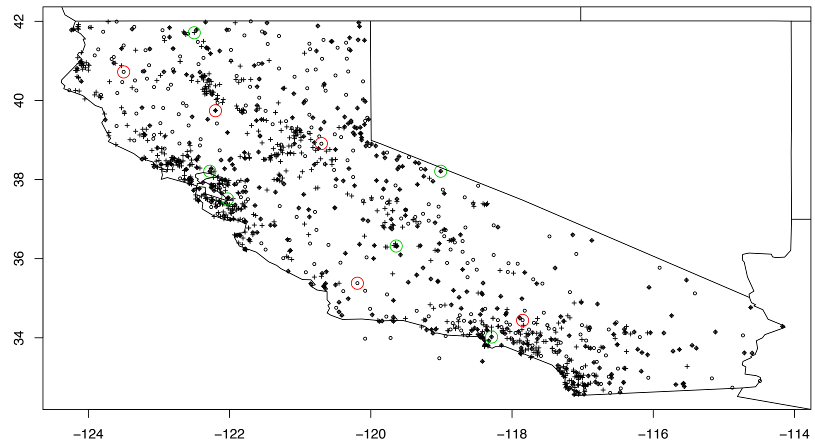

Figure 5: Locations in the State of California where the data were collected. Circle points: Temperature data; Crossed points: Rainfall data.

the left hand tail of the temperature distribution. The data were obtained from the National Climatic Data Center, and are available on the web at `http://www.ncdc.noaa.gov/`. The total number number of observations, for each station, during the two years period, is 180 daily data. Figure 5 shows the locations where the data were collected. The peaks over a threshold for these data are clearly not independent in time, as threshold excesses often occur in clusters. Following the ideas in Coles (2001), we tackle this problem by considering cluster maxima. A more structured approach could be considered by modeling the serial correlation as in Reich et al. (2014). In this example we calculate the maxima for blocks of four days, obtaining a total of 45 maxima per station. The threshold for the temperature application was chosen as $10 \cdot 0$ °C Celsius (or equivalently to 50 degrees Fahrenheit). For the rainfall example we chose 100 mm. In both cases the thresholds are close to the corresponding 80% quantile of the declustered data.

We fit our proposed model using the prior distributions suggested in Section 4.1, with constant the tail and scale parameters. We use a Gaussian kernel given by $k(s - s^*) = \exp(-0 \cdot 5||s - s^*||^2/b_w^2)/\sqrt{2\pi b_w^2}$ for $L = 100$ knots distributed on regular grid over the domain. This implies that the distance between knots is equal to 110km. Table 2 shows the posterior means and credibility interval for the parameters.

Figure 6 shows two maps that illustrate the results for the daily minimum winter temperature over California. As expected the probability that the exceedance will be larger than the 10 °threshold is pretty low, with the exception of the southern coastal region and the south-east corner of the state. The map of the return levels shows a north-south gradient that is affected by a coastal and a sierra effect. Analogous maps are presented in Figure 7 for winter precipitations. Here the maps indicate a clear north-south split, with a strong effect of the desertical areas in the south-east of the state, and very high return levels in the northern coast of the state.

Table 2: Means and 95% credibility intervals for the different parameters in the model. M - Posterior mean, CI - Credibility interval

	γ		ξ		σ		b_w		τ	
	M	CI	M	CI	M	CI	M	CI	M	CI
Minimum Temperature	0.045	(0.023; 0.068)	-0.36	$(-0.44; -0.30)$	2.30	(2.23; 2.39)	0.54	(0.51; 0.58)	12.06	(10.21; 13.12)
Rainfall	0.102	(0.086; 0.116)	0.022	$(-0.082; 0.120)$	16.41	(16.05; 16.74)	1.45	(1.40; 1.53)	10.48	(9.08; 11.83)

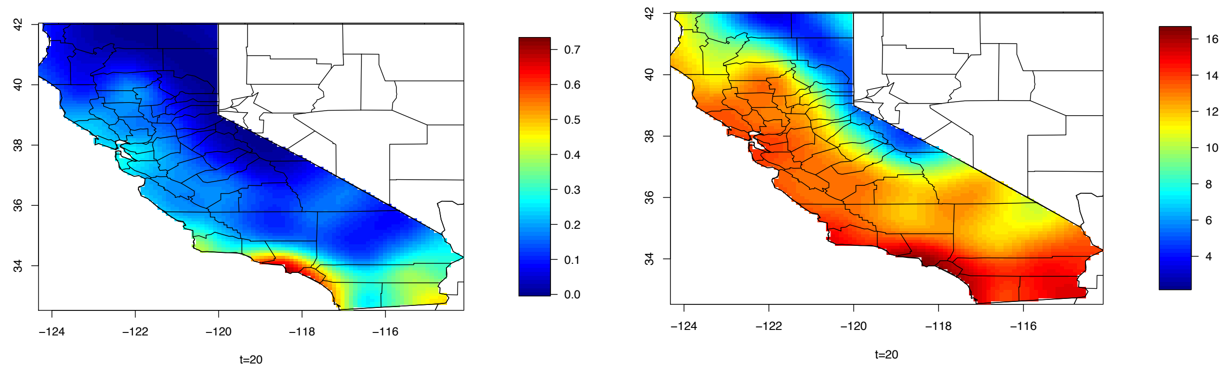

Figure 6: Minimum winter temperature in the State of California. Left: Probability of an exceedance larger than 10 °C; Right: 20-year return levels

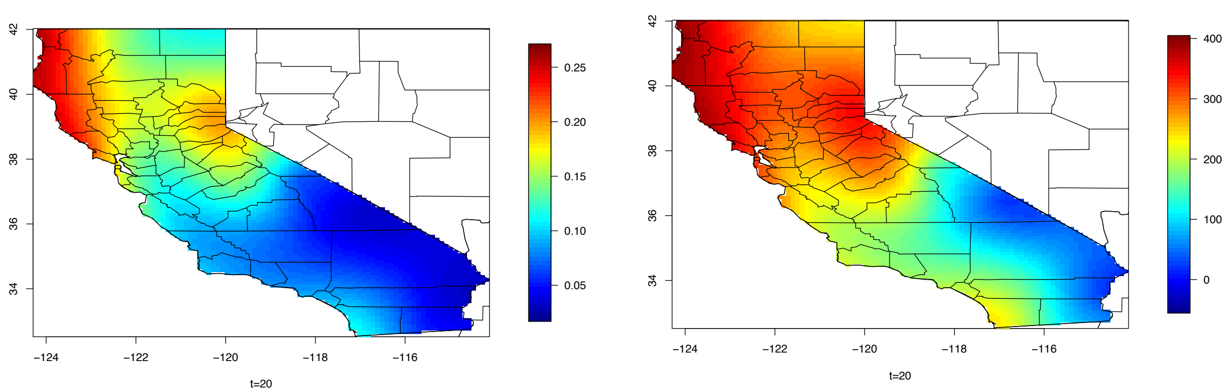

Figure 7: Winter precipitation over the State of California. Left: Probability of exceeding the 100 mm threshold; Right: 20-year return levels

7 Conclusion and discussion

We have presented a new model for the excesses above a threshold of spatially referenced observations. The model leverages the constructive multiplicative definition of a Pareto processes. The Pareto process is an infinite-dimensional extension of the Pareto distribution that is commonly used for the analysis of exceedances over a threshold at a single location. Inference for a Pareto process is, in principle, possible using a hierarchical approach, but it is complicated by the need to impose complex restrictions in the parameter space and estimate a non-trivial normalizing constant. Our model is based on adding a multiplicative perturbation the Pareto process, in order to facilitate an inferential approach based on conditional independence. The perturbation can be interpreted as a nugget effect or as an observational error. The model can flexibly capture wide ranges of spatial dependence. It can also handle wide ranges of dependence. Moreover, simulations show that the proposed model is able to capture spatial structures that are typical of fields of extreme values, including the particular case $\xi = 0$ which correspond to the Pareto process. The hierarchical structure of the model, coupled with the kernel representation of the spatial field, allows for computations to be performed on spatial domains with large numbers of locations. Our Bayesian inferential approach, uses the full likelihood and avoids splitting the inference for model parameters in a sequence of steps. It is, thus, able to coherently propagate the uncertainty in the estimation of all components of the model, using probability distributions.

Acknowledgement

This research was partically funded by the National Science Foundation grant DMS-1513076. The research of the first author was supported CAPES, Brazil, grant number $3169 - 13 - 8$.

References

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*. Chapman and Hall, New York.

Castellanos, M. A. and Cabras, S. (2007). A default bayesian procedure for the generalized pareto distribution. *Journal of Statistical Planning and Inference*, 137:473–483.

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B*, 53:377–392.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., and Sun, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *REVSTAT-Statistical Journal*, 10:135–165.

Cooley, D., Davis, R. A., and Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101:2103–2117.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of American Statistical Association*, 102:824–840.

Davison, A. C. and Gholamreazee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 468:581–608.

Falk, M. and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized pareto distributions. *Journal of Multivariate Analysis*, 99:715–734.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20:1717–1737.

Fisher, R. A. and Tippett, L. H. C. (1928). On the estimation of the frequency distributions of the largest and smallest sumber of a sample. *Proceedings of the Cambridge Philosophycal Society*, 24:180–190.

Heffernand, J. and Tawn, J. (2004). A conditional approach for multivariate extreme. *Journal or the Royal Statistical Society, Series B*, 66:497–546.

Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics*, 14:285–299.

Huser, R. and Davison, A. C. (2013). Spacetime modelling of extreme events. *Journal of the Royal Statistical Society, Series B*, to appear.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorology Society*, 81:158–171.

Jeon, S. and Smith, R. L. (2012). Dependence structure of spatial extremes using threshold approach. *Submitted, arXiv:1209.6344*.

Michel, R. (2008). Some notes on multivariate generalized pareto distributions. *Journal of Multivariate Analysis*, 99:1288–1301.

Padoan, S., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *To appear in Journal of the American Statistical Association*.

Pauli, L., Racungo, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 21:149–164.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131.

Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute: Proceeding of the 43rd Session (Buenos Aires)*, Book 2:857–878.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics*, 6:1430–1451.

Reich, B. J., Shaby, B. A., and Cooley, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western us. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):119–135.

Rootzen, H. and Tajvidi, N. (2006). Multivariate generalized pareto distributions. *Bernoulli*, 12:917–930.

Sang, H. and Gelfand, A. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16:407–426.

Smith, R. L. (1990). Max-stable processes and spatial extreme. *Unpublished Manuscript*.

Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Aust. N. Z. J. Stat.*, 51:77–88.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102:855–870.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. *American Mathematical Society*, 2:271–294.

Wadsworth, J. and Tawn, J. (2014). Efficient inference for spatial extreme values processes associated to log-Gaussian random functions. *Biometrika*, 101:1–15.