A Bayesian Hierarchical Generalized Pareto Spatial Model for Exceedances Over a Threshold

Fernando Ferraz do Nascimento

Universidade Federal do Piauí, Campus Petronio Portella, CCN II, 64000 Teresina-PI, Brazil.

Bruno Sansó

University of California, Santa Cruz, 1156 High Street, SOE 2, Santa Cruz, CA-95064, USA.

Summary. 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 has generalized Pareto distributions as its marginals. The process is in the domain of attraction of the max-stable family. It 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.

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 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 distribution. This method relies on a drastic reduction of the original data to a small set of block maxima. An alternative, that uses additional information, 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, given by G^n, is a rescaled and shifted version of 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 of processes to model block maxima observed at spatially referenced sites. 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 estimators is usually underestimated (Pauli et al., 2011).

Of particular relevance for the model developed in this paper is the hierarchical max-stable model in Reich and Shaby (2012). Let $Y(s)$ be the block maximum at location $s \in \mathcal{S}$. Assuming that the process is max-stable, we have that marginally, $Y(s) \sim GEV(\mu(s), \sigma(s), \xi(s))$, or $Y(s) = \mu(s) + \sigma(s)/\xi(s)[X(s)^{\xi(s)} - 1]$, where $X(s) \sim GEV(1, 1, 1)$. The process $X(s)$ is written as a product of two terms, $X(s) = U(s)\theta(s)$. $U(s) \sim GEV(1, \alpha, \alpha)$, $\alpha \in (0, 1)$. $\theta(s)$ is a nugget effect that captures the spatial dependence, and it is written as a power function of a linear combination of positive stable distributions, where the weights are defined by spatial kernels.

In this paper, we propose a hierarchical Bayesian model for excedances in a spatial domain using a process with GPD univariate marginals. The hierarchy that defines the process is similar to the one proposed in Reich and Shaby (2012), and it is inspired in the representation discussed in Ferreira and de Haan (2014). We show that our proposed model can capture a wide range of spatial dependencies. Moreover, the model allows for asymptotic dependence as well as independence between any two points. In addition, the resulting process belongs to the domain of attraction of a max-stable process, for some parameter values. 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 presents some numerical simulations of the proposed model using different parameter configurations. Section 4 presents an illustration on data for winter temperature and rainfall in California during the last five years. Finally, Section 5 discusses the results obtained with the proposed model, as well as possible future extensions.

2. The Model

To perform inference on the distribution of the exceedances over a threshold, 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

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

where $y - u > 0$ for $\xi \geq 0$ and $0 \leq y - u < -\sigma/\xi$ for $\xi < 0$. Thus, the GPD is always bounded from below by u, is bounded from above by $u - \sigma/\xi$ if $\xi < 0$ and unbounded from above if $\xi \geq 0$. 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.

2.1. The generalized Pareto process

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

Definition 1. *Let $W(s)$ be a stochastic process indexed in $s \in \mathcal{S}$ and w_0 a positive constant. Then $W(s) = Z\theta(s)$, $\forall s \in \mathcal{S}$ is a simple Pareto Process if*

(a) *θ is a stochastic process with $\sup_s \theta(s) = w_0$ and $E(\theta(s)) > 0$, $\forall s \in \mathcal{S}$,*

(b) *Z is a standard Pareto random variable,*

(c) *Z 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. To obtain a generalized Pareto process we rescale a simple Pareto process W as

$$W^*(s) = u + \sigma \frac{W(s)^\gamma - 1}{\gamma} \ .$$

Following Ferreira and de Haan (2014) we have that the dependence between realizations of the process at two sites can be quantified by the identity

$$Pr(W(s_1) > x, W(s_2) > x) = \frac{E(\theta(s_1) \wedge \theta(s_2))}{x}$$

where $a \wedge b = \min(a, b)$. As a consequence, there can not be independence between $W(s_i)$ and $W(s_j)$ for any two locations s_i and s_j. Moreover, the process is completely dependent when $\theta \equiv w_0$. In addition, for for n sites $s_1, \ldots, s_n$, we have that

$$Pr(W^*(s_i) > x_i \mid W^*(s_i) > w_0, \theta(s_i); i = 1, ..., n) = w_0 \frac{\min_i \left\{ \theta(s_i) \left[1 + \frac{\xi}{\sigma}(x_i - u) \right]^{-1/\xi} \right\}}{\min_i \{\theta(s_i)\}} \ .$$

From the above expression it is clear that likelihood-based inference is not possible as, in general, the cumulative joint distribution function is not absolutely continuous. Thus,

it does not have a derivative with respect to $(x_1, ..., x_n)$. This is also the case, even for pairwise likelihood approach. Due to this limitation, we propose an alternative hierarchical representation of the generalized Pareto process.

2.2.   A hierarchical generalized Pareto process

We consider an unscaled process $X(s)$, and, in the spirit of Reich and Shaby (2012), we factorize it as in $X(s) = U(s)\theta(s)$, where $U(s)$ are spatially uncorrelated random variables, and $\theta(s)$ is a process that controls the spatial dependence. More specifically, we set $U(s) \sim GPD(1, 1/\beta, 1/\beta)$, i.i.d., so that $Pr(U(s) > u) = 1/u^\beta$, $u > 1$. Then, we select a set of L points $s_1^*, \ldots, s_L^*$ in $\mathcal{S}$ and propose the model $\theta(s) = \exp\{-\sum_{l=1}^{L} A_l w_l(s)^{1/\alpha}\}^{1/\beta}$. Here the random variables $A_l, l = 1, \ldots, L$, follow a positive stable distribution, and $w_l(s) = k(s - s_l^*) =$ for some kernel k such that, $w_l(s) \geq 0$ and $\sum_{l=1}^{L} w_l(s) = 1$, $\forall s \in \mathcal{S}$. Notice that $\theta(s)$ is a bounded process, thus, it satisfies the first condition in Definition 1. But, as opposed to the random variable Z in Definition 1, $U(s)$ depends on location. Because of this dependence, our model does not conform to the constructive definition of a Pareto process. Nevertheless, as the following results will show, we have defined a process with a number of properties that make it suitable for the analysis of spatial exceedances over a threshold.

The following proposition gives the joint distribution of a vector corresponding to the process at an arbitrary collection of sites.

Proposition 1. *For $s_1, \ldots s_n$ in $\mathcal{S}$*

$$Pr(X(s_i) > x_i; i = 1, ..., n) = \left(\prod_{i=1}^{n}(1/x_i^\beta)\right) \exp\left\{-\sum_{l=1}^{L}\left(\sum_{i=1}^{n} w_l(s_i)^{1/\alpha}\right)^\alpha\right\}, \qquad (2)$$

providing an explicit formula for the finite dimensional distributions of $X(s)$.

PROOF. Denote $\boldsymbol{A} = (A_1, \ldots, A_L)$, thus

$$Pr(X(s_i) > x_i, i = 1, ..., n) = \int_{\mathcal{A}} Pr(X(s_1) > x_1, ..., X(s_n) > x_n|\boldsymbol{A})p(\boldsymbol{A}|\alpha)d\boldsymbol{A}.$$

Use the independence of the components of $\boldsymbol{A}$ to write the integral as a product of integrals. From the definition of $X(s)$ and the fact that $Pr(U(s_i) > x_i/\theta(s_i)|\boldsymbol{A}) = (\theta(s_i)/x_i)^\beta$, we have that the above integral is equal to $\prod_{i=1}^{n}(1/x_i^\beta)\int_{\mathcal{A}} \theta(s_1)^\beta \times \ldots \times \theta(s_n)^\beta P(\boldsymbol{A}|\alpha)d\boldsymbol{A}$. Recall that, if A follows a positive stable distribution with parameter α, then $\int_0^\infty \exp\{-At\}p(A|\alpha)dA = \exp\{-t^\alpha\}$. The result follows from the use of this identity in each of the n terms of the product.

Corollary 1. *The univariate marginal distribution of $X(s)$ for a given point $s \in \mathcal{S}$ is obtained from $Pr(X(s) > x) = (ex^\beta)^{-1}$.*

Corollary 2. *$\alpha = 1$ corresponds to complete independence.*

Corollary 3. *$\alpha = 0$ corresponds to complete dependence yielding*

$$Pr(X(s_i) > x_i, i = 1, ..., n) = \left(\prod_{i=1}^{n} x_i^{-\beta}\right) \exp\left\{-\sum_{l=1}^{L} \max_{i=1,...,n} w_l(s_i)\right\}$$

PROOF. The proof of the corollary is obtained by recalling that $\lim_{\alpha \to 0} \left(\sum_{i=1}^{n} w_l(s_i)^{1/\alpha}\right)^\alpha = \max_{i=1,...,n} w_l(s_i)$.

From Equation (2) it is clear that the spatial dependence is controlled by the α-norm of the weigths. The next proposition provides information about the asymptotic max-stability of the process.

Proposition 2. *The process $X(s)$ belongs to the domain of attraction of a max-stable process when $\beta = 1$.*

PROOF. The proof follows along the lines of a similar proof in Reich and Shaby (2012). Recalling that $X(s) = U(s)\theta(s)$, we have

$$Pr(X(s_1) \leq tc_1, ..., X(s_n) \leq tc_n)^t = \left(\int_{\mathcal{A}} \prod_{i=1}^{n} \left(1 - \frac{\theta(s_i)}{tc_i}\right) p(\boldsymbol{A}|\alpha) d\boldsymbol{A}\right)^t. \qquad (3)$$

This expression has the form $(1 + a_1/t + a_2/t^2 + \ldots + a_n/t^n)^t$ for appropriate coefficients $a_i, i = 1, \ldots, n$. The limit, when $t \to \infty$, is $\exp\{a_1\}$, where $a_1 = (1/e)\sum_{i=1}^{n} 1/c_i$. This shows that the limit of the expression in Equation (3) is the product of n independent $GEV(e^{-1}, e^{-1}, 1)$ distributions.

We notice that, when the above result is compared to the function V described in Section 1.1, we have that $V(z_1, ..., z_D) = \frac{1}{e}(1/z_1 + ... + 1/z_D)$. As a consequence, when $\beta = 1$ the process is in the domain of attraction of a max-stable process that has asymptotic independence. Asymptotic dependence can be obtained by considering a probability distribution on β, instead of fixing its value. We will postpone the discussion of this feature of the model to Section 2.3.

The two propositions and the three corollaries above have explored the properties of the proposed hierarchical process. We have established that the model has the appropriate marginal distributions to generalize the GPD analysis, for $\beta = 1$ it belongs to the domain of attraction of a max-stable process, and has a dependence structure, controlled by the

parameters α that allows for a wide range of possibilities. For inferential purposes, we now extend the model to include shape, scale and location parameters, letting

$$Y(s) = \left(u + \frac{\sigma(X(s)^\xi - 1)}{\xi}\right),$$

for $\sigma > 0$,ξ and $u \in \mathbb{R}$. The univariate marginal is given in the next proposition whose proof follows along the lines of the proof to Proposition 1. Its corollary provides the basis of a model that is conditional on exceeding a threshold u.

Proposition 3.

$$Pr(Y(s) > y) = \left(1 + \frac{\xi(y - u)}{\sigma}\right)^{-\beta/\xi} e^{-1} \ .$$

Corollary 4. *Given a threshold u, the conditional probability of $Y(s)$, given that $Y(s) > u$ is a $GPD(u, \sigma/\beta, \xi/\beta)$.*

PROOF.

$$Pr(Y(s) > y \mid Y(s) > u) = \int_{\mathcal{A}} \frac{Pr(Y(s) > y \mid \theta(s))}{Pr(Y(s) > u \mid \theta(s))} p(\boldsymbol{A}|\alpha) d\boldsymbol{A} = \left(1 + \frac{\xi(y - u)}{\sigma}\right)^{-\beta/\xi} \ .$$

The final step to determine the distribution of the different levels of a hierarchical model consists of obtaining $Pr(Y(s) > y|\theta(s))$. The following result is a consequence of Corollary 1.

Corollary 5. $Pr(Y(s) > y \mid \theta(s)) = \theta(s)^\beta \left(1 + \frac{\xi(y-u)}{\sigma}\right)^{-\beta/\xi}.$

As a special case of the above, when $y = u$ we have that $Pr(Y(s) > u|\theta(s)) = \theta(s)^\beta$.

To obtain a hierarchical specification of a model for the exceedences of $Y(s)$ above the threshold $u(s)$ we assume that we have a number of replicates of the process Y. These are denoted as $Y_j(s)$. The probability that $Y_j(s)$ is larger than $u(s)$ is $\theta(s)^\beta = \exp\{-\sum_{l=1}^{L} A_l w_l(s)\}$. Denote as $V_j(s)$ a latent variable that is equal to 1 if $Y_j(s) > u(s)$ and 0 otherwise. We have that $V_j(s) \sim Bernoulli(\theta(s)^\beta)$. With these properties, we propose the following hierarchical model:

$$V_j(s_i) \mid \theta(s_i)^\beta \overset{indep}{\sim} Bernoulli(\theta(s_i)^\beta)$$

$$Y_j(s_i) \mid Y_j(s_i) > u(s_i) \overset{indep}{\sim} GPD(u(s_i), \sigma/\beta, \xi/\beta)$$

$$A_l \overset{iid}{\sim} PS(\alpha), \tag{4}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$.

For a dataset corresponding to $s_1, ..., s_n$ points in space, assume that we have m_i replicates at each location s_i. Let $Y_j(s_i)$ be the quantity of interest at location s_i, replicate
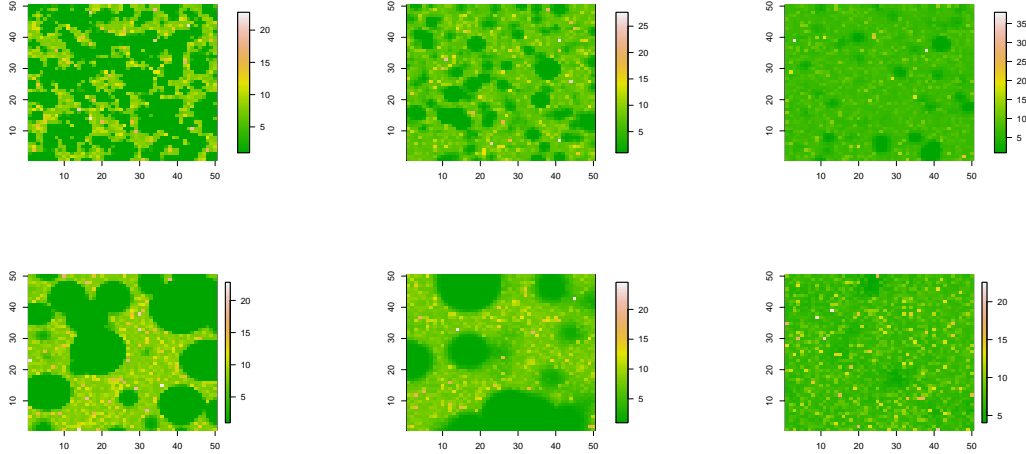
Fig. 1. Random draws from the HGP model with different parameter configurations. Columns from left to right: $\alpha = (0.2, 0.5, 0.8)$. Top row, $\tau = 1$, second row, $\tau = 5$. $\sigma = 1$, $\xi = .2$,

j. If this quantity is higher than a threshold $u(s_i)$, its resulting distribution is a GPD. The auxiliary variable $V_j(s_i)$ indicates if a point is higher than $u(s_i)$, where the resulting probability follows from a Bernoulli distribution with parameter $\theta(s_j)^\beta$. In practice the condition that values above a threshold are independent replicates is seldom satisfied. Thus, it is common to decluster the data in a preprocessing stage (Coles, 2001).

To explore the characteristics of this model, we show, in Figure 1, random draws of the process on a 50×50 grid, for different values of α and the bandwidth of a Gaussian kernel, τ. From the figure, we observe that the process shows clustering patterns that are typical of spatial processes of extremes. We also observe that lower values of α and τ increase the smoothness of the realizations.

2.3. Shape parameters and asymptotic dependence

From the specification of the model in Equation (4), we note that the resulting marginal GPD depends on four parameters. The shape of the GPD is given by ξ/β, while the scale is given by σ/β. Thus, such parameterization can cause indetifiability problems for estimation. To tackle this problem we complete the proposed hierarchical model by assuming that the prior distribution of β has support limited to the positive real line. The prior distribution for ξ is a discrete distribution concentrated on the values $\{-1, 0, 1\}$. Thus ξ indicates the sign of the chape, and the β quantifies its value. This is analogous

to the approach in Nascimento et al. (2015). We note in passing that distributions with negative shape parameters have bounded support. Such tail behavior is radically different to the very heavy tails, that correspond to positive shape values. It is important for the model to have the flexibility to distinguish this feature.

According to Proposition 2 taking $\beta = 1$ provides asymptotic independence. To gain a better understanding of the this issues we recall that, for two different locations, s_i and s_j, a measure of the asymptotic dependence between $X(s_i)$ and $X(s_j)$ is given by (Coles et al., 1999)

$$\chi_{i,j} = \lim_{z \to z^*} P(X(s_i) > z \mid X(s_j) > z),$$

where z^* is the upper limit of the distribution. The value $\chi = 0$ indicates asymptotic independence. Positive values of χ indicate some level of asymptotic dependence. For the model proposed in this paper, it can be seen that, any fixed value of β, not only $\beta = 1$, will produce $\chi_{i,j} = 0$. This suggests that a model based on independent Pareto shocks at each location is not able to capture possible extremal dependence between different locations. To induce the sharing of information between the different sites, we assume that β is a random variable following a gamma distribution with parameters a and b. In this fashion the component of the vector $(U(s_1), \ldots, U(s_n))$, for any collection of n sites, will not be independent. They will be exchangeable with a joint density proportional to $\prod_{i=1}^{n} u_i^{-1} (b + \sum_{i=1}^{n} \log u_i)^{-n-a}, u_i > 1, \forall i$.

For the proposed model we have that

$$\chi_{i,j} = \lim_{z \to \infty} \frac{\int_0^\infty \beta^{a-1} e^{-\beta[b+2\log(z)] - \beta^\alpha \sum_{l=1}^{L} (w_l(s_i)^{1/\alpha} + w_l(s_j)^{1/\alpha})^\alpha} d\beta}{\int_0^\infty \beta^{a-1} e^{-\beta[b+\log(z)] - \beta^\alpha} d\beta} .$$

For $\alpha = 0$, $\chi_{i,j} = (1/2)^a \exp\{1 - \sum_{l=1}^{L} \max(w_l(s_i), w_l(s_j))\}$. For $\alpha = 1$, $\chi_{i,j} = (1/2)^a$. This indicates that asymptotic dependence can indeed be captured by the model. For $0 < \alpha < 1$, the exact value of χ depends of integrals that cannot be computed analytically. Numerical calculations indicate that $(1/2)^a \exp\{1 - \sum_{l=1}^{L} \max(w_l(s_i), w_l(s_j))\} < \chi_{ij} < (1/2)^a$, implying that, for any value of α, the model produces asymptotic dependence.

2.4. 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

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

the above formula for the return level assumes that all observations are above a threshold. Our propose 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 $P(Y_j(s) > u(s) \mid \theta(s)) = \theta(s)^\beta$. Thus, the return level for t is given by

$$r_t(s) = u(s) + \frac{\sigma}{\xi}\left(\left(\frac{1/t}{\theta(s)^\beta}\right)^{-\xi} - 1\right). \tag{5}$$

2.5. *Posterior distribution and estimation algorithm*

Let Θ be the vector of parameters $\{\xi, \sigma, \alpha, \tau, \beta\}$ and let $\mathbf{A}$ be the positive stable random effects in $\theta(s)$. Then, Θ and $\boldsymbol{A}$ are the unknown quantities in our model that need to be estimated from the observations. Using a Bayesian approach, we obtain a likelihood for Θ and $\boldsymbol{A}$ and consider prior distributions that allow us to obtain a posterior distribution that includes all estimation uncertainties. From the model proposed in Equation (4), the likelihood, conditional on the indicator variables $V_j(s_i)$, that is obtained from m replicates of the process is proportional to

$$\prod_{i=1,\ldots,n} \prod_{j=1,\ldots,m} \left(f_G\left(Y_j(s_i) \mid u(s_i), \frac{\sigma(s_i)}{\beta}, \frac{\xi(s_i)}{\beta}\right)\right)^{I(Y_j(s_i)>u(s_i))} \times$$
$$\times\theta(s_i)^{\beta V_j(s_i)}(1 - \theta(s_i)^\beta)^{(1-V_j(s_i))}, \tag{6}$$

where f_G is the density of the GPD distribution.

Regarding our default choices for the prior distribution for the GPD parameters, we assume independence a priori. We use the prior for σ, $p(\sigma) \propto 1/\sigma$, that corresponds to the Jeffreys prior, as defined in Castellanos and Cabras (2007). For α we use a $U(0,1)$ prior, following the choice in Reich and Shaby (2012). For the bandwidth τ, we propose to use a weakly informative gamma prior. The proposed prior for β is a $Gamma(1,1)$, whose expected valued $E(\beta) = 1$. Thus, in the mean, this corresponds to the case of max-stability and asymptotic independence, but allows significant probability for the other cases. The prior for ξ assigns probability 1/3 to each of the values -1,0 and 1. In order to explore the posterior distribution of the model parameters we develop an adaptive Monte Carlo. The details are presented in Appendix A.

3. 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 10x10 grid. For each site, we generated 200 replicates. We used as many kernel knots as grid cells in the grid. We chose $u = 6$, $\sigma = 10$, $\xi = \{-1, 0, 1\}$ and $\beta = \{1, 3\}$. We simulated using a $\tau = 10$, and three different values of α, $\alpha = \{0.2, 0.5, 0.8\}$. The steps to generate the simulations are as follows: 1) Generate $A_l \sim PS(\alpha), l = 1, \ldots, L$ independently; 2) Generate $U_j(s)$ iid $GPD(1, 1/\beta, 1/\beta)$; 3) Compute the kernel weights $w_l(s)$, for $l = 1, ..., L$. 4) Compute the $\theta(s) = \exp\left(-\sum_{l=1}^{L} A_l w_l(s)^{1/\alpha}\right)^{1/\beta}$; 5) Compute $X_j(s) = U_j(s)\theta(s), j = 1, \ldots, 200$; 6) For each j, compute the transformation $Y_j(s) = u + (\sigma/\xi)\left(X_j(s)^\xi - 1\right)$.

We fit our model using the priors proposed as default. When fitting the model we let number of knots $L = 25$. Thus, the fitted model has one knot for every four data locations, compared to the true model that generated the data. Table 1 shows the posterior mean for the different configurations of the parameter values, together with 95% credibility intervals. From the table, we notice that we are able to recover the true parameter values for most parameters in most of the configurations. We observe that the most challenging case is the one where $\xi = -1$ and $\alpha = 0.8$, where the method underestimates the value of β.

Figure 2 shows the fields of posterior expectations for the return levels corresponding to two different values of t and three different estimation methods. The true values of the returns can be calculated using the true values of parameters in (2.4). The first approach uses an empirical estimate of $\theta(s)^\beta$, the probability of crossing the threshold. The second uses the full model. The third method uses the empirical CDF of the data. We observe that estimating $\theta(s)^\beta$ produces results that are very similar to the ones obtained from the model. We warn, though, that these examples use 200 replicates. A very large number compared to what is typically available in real applications. We repeated the analysis for several values of t. We observed that, empirically estimating the return levels for a large t is inviable, even with a large number of replicates.

In Figures 3 and 4 we explore the effect of different parameter values on the estimation of the return levels. From Figure 3 we observe that β does not seem to have a strong effect on the spatial patterns of the resulting fields. Quite the opposite is true for ξ. This is not surprising, as ξ controls the tails of the GPD. Thus, when $\xi < 0$, which corresponds to a

Table 1. Posterior mean and 95% credibility intervals. T- True, M - Posterior mean, CI - Credibility interval. The below of $P(\xi)$ indicates the probability of the signal of ξ would be 0, 1("+") or -1("-").

|	$\beta = 1,\ \xi = 1$								
	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
	T	M	CI	T	M	CI	T	M	CI
τ	1	1.014	(1.003;1.026)	1	0.921	(0.885;0.944)	1	0.838	(0.785;0.891)
α	0.2	0.213	(0.204;0.222)	0.5	0.560	(0.507;0.593)	0.8	0.812	(0.776;0.834)
β	1	0.974	(0.941;1.008)	1	0.997	(0.955;1.039)	1	0.989	(0.948;1.027)
σ	10	9.68	(9.30;10.06)	10	9.998	(9.581;10.393)	10	9.96	(9.49;10.40)
$P(\xi)$	+	0	-	+	0	-	+	0	-
	1	0	0	1	0	0	1	0	0
	$\beta = 1,\ \xi = 0$								
	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
	T	M	CI	T	M	CI	T	M	CI
τ	1	0.998	(0.984;1.009)	1	0.863	(0.841;0.883)	1	1.050	(0.997;1.102)
α	0.2	0.217	(0.203;0.226)	0.5	0.577	(0.527;0.613)	0.8	0.802	(0.770;0.827)
β	1	0.986	(0.081;3.614)	1	0.345	(0.015;2.55)	1	1.187	(0.179;3.934)
σ	10	9.96	(9.76;10.14)	10	10.02	(9.80;10.20)	10	10.04	(9.83;10.26)
$P(\xi)$	+	0	-	+	0	-	+	0	-
	0	1	0	0	1	0	0	1	0
	$\beta = 3,\ \xi = 1$								
	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
	T	M	CI	T	M	CI	T	M	CI
τ	1	0.998	(0.987;1.007)	1	9.905	(0.880;0.931)	1	0.984	(0.933;1.033)
α	0.2	0.217	(0.204;0.226)	0.5	0.569	(0.518;0.607)	0.8	0.840	(0.805;0.962)
β	3	2.852	(2.638;3.074)	3	2.984	(2.740;3.249)	3	2.922	(2.695;3.174)
σ	10	9.31	(8.43;10.24)	10	9.96	(8.90;11.10)	10	9.71	(8.71;10.82)
$P(\xi)$	+	0	-	+	0	-	+	0	-
	1	0	0	1	0	0	1	0	0
	$\beta = 3,\ \xi = -1$								
	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
	T	M	CI	T	M	CI	T	M	CI
τ	1	0.997	(0.985;1.009)	1	0.952	(0.928;0.973)	1	0.882	(0.833;0.930)
α	0.2	0.217	(0.205;0.226)	0.5	0.551	(0.510;0.587)	0.8	0.846	(0.813;0.867)
β	3	3.048	(2.945;3.194)	3	2.998	(2.874;3.132)	3	1.709	(1.708;1.711)
σ	10	10.08	(9.87;10.38)	10	10.00	(9.78;10.28)	10	9.67	(966;9.68)
$P(\xi)$	+	0	-	+	0	-	+	0	-
	0	0	1	0	0	1	0	0	1
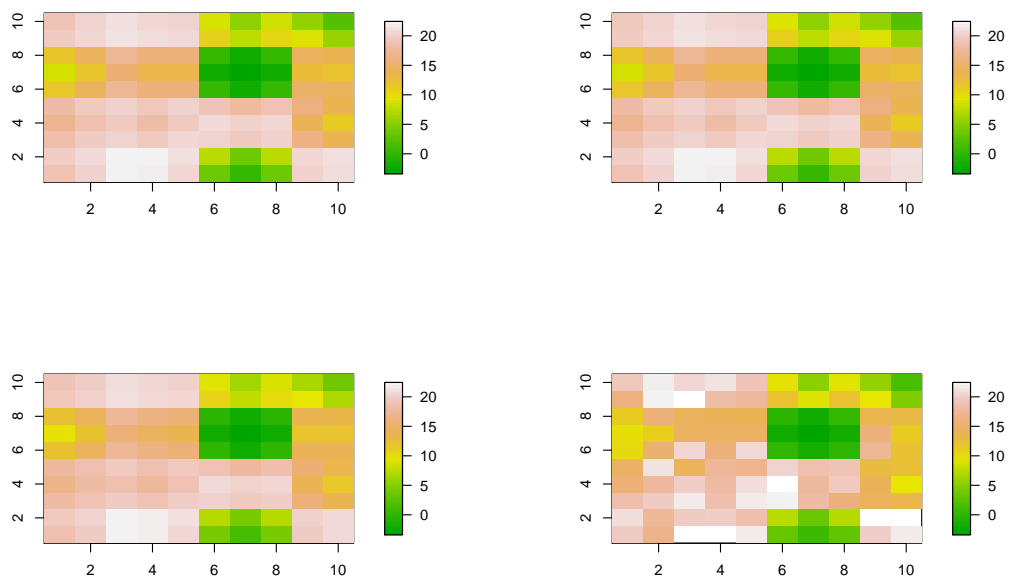
Fig. 2. 20-years return levels in space for simulated data corresponding to $\beta = 3$, $\xi = 1$, $\alpha = 0.5$ and $\tau = 10$. Top left: True returns; Top right: $P(Y(s) > u) = \theta(s)^\beta$ is estimated empirically; Bottom left: $\theta(s)^\beta$ estimated from the model; Bottom Right: Full empirical estimation.
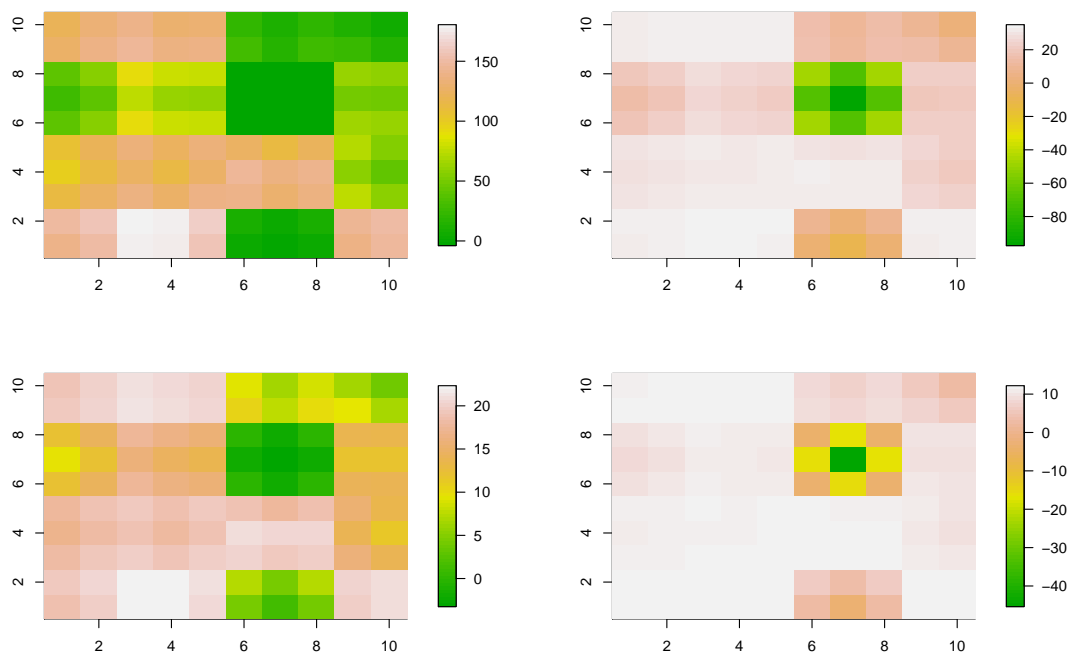
Fig. 3. 20-year return levels in space for the simulated data with $\alpha = 0.5$ and $\tau = 1$. Top Left: $\xi = 1, \beta = 1$; Top Right: $\xi = 0, \beta = 1$: Bottom Left: $\xi = 1, \beta = 3$; Bottom Right: $\xi = -1, \beta = 3$.

bounded distribution, we obtain very low return levels. From Figure 4 we observe that the lower the value of α the higher the clustering of the field. This is expected, as α control the spatial dependence of the process.

4. California temperature and rainfall

As illustrative examples we analyze two datasets. The first one consists of data for maximum 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. These 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
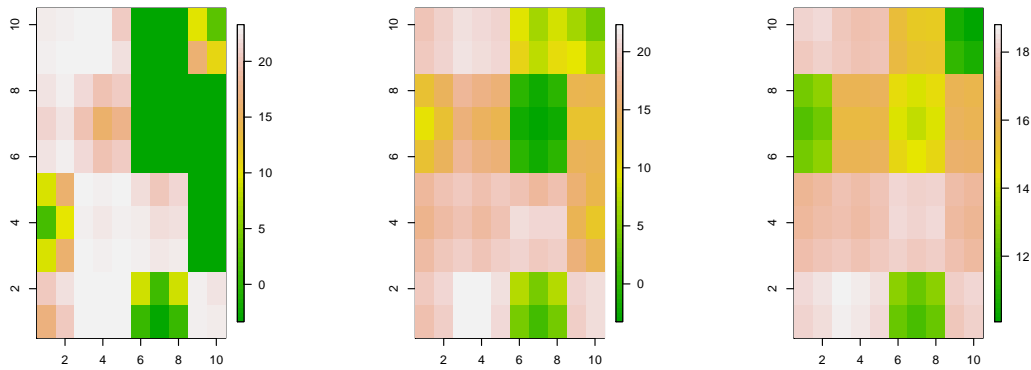
Fig. 4. 20-year return levels in space for the simulated data with $\xi = 0.4$ and $\tau = 1$. Left: $\alpha = 0.2$, Center: $\alpha = 0.5$. Right: $\alpha = 0.8$.

(Coles, 2001, Section 5.3.2), we tackle this problem by considering cluster maxima. 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 10.0 °C Celsius (or equivalently to 50 degrees Fahrenheit). The threshold chosen for this application was 100 mm. In both cases these thresholds are close to the corresponding 80% quantile of the declustered data.

We perform a preliminary analysis to quantify the asymptotic behavior of the dependence between observations at different locations. For this purpose we chose as illustration some different sites, transformed the observations so that they have standard Frechet distributions, then explored the resulting scatter plots. Points along the diagonal give an indication of asymptotic dependence. Figures 6 and 7 show the resulting plots for four different pair of locations. Such locations are circled in Figure 7. For the temperature data, Figure 6 shows that correlations for the stations For the rainfall data, Figure 7 shows that relationship between the sites We observe that there are clear indications that there is asymptotic dependence between some of the locations. The above results indicate that, to appropriately capture the tail behavior of the joint distribution of the peaks over a threshold for the data under consideration, we need to consider a model that can capture asymptotic dependence. This is achieved by our model by assuming that β is a random variable.
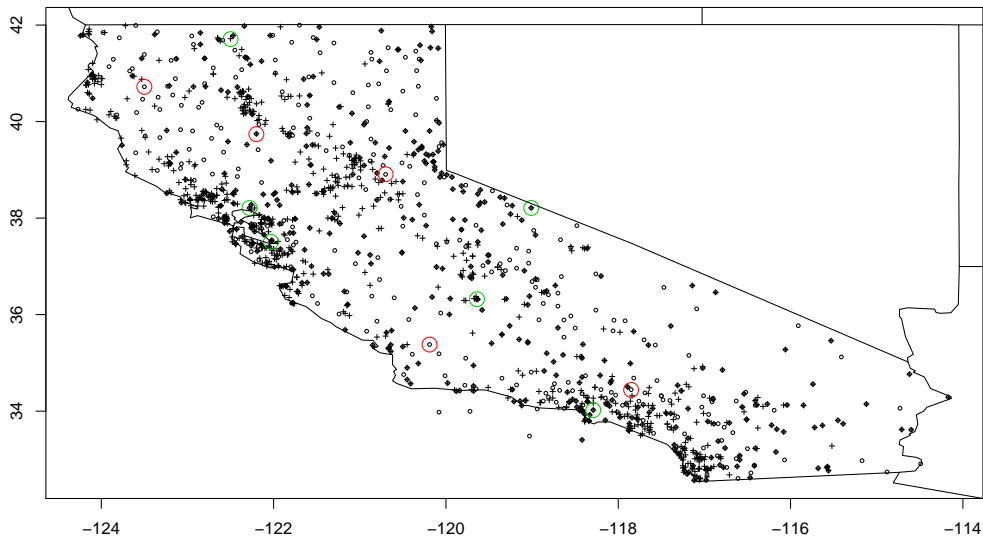
Fig. 5. Locations in the State of California where the data were collected. Circle points: Temperature data; Crossed points: Rainfall data.
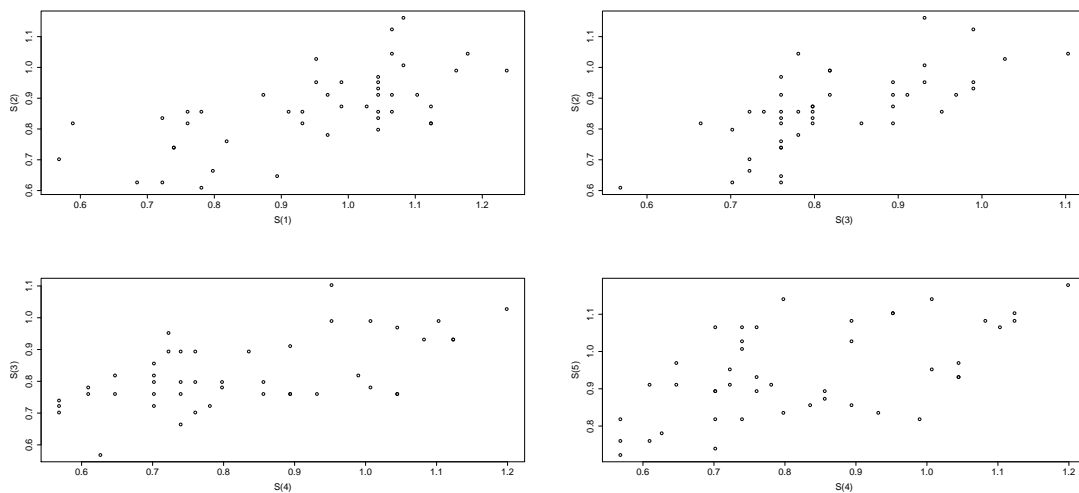

Fig. 6. Examples of scatter plots for temperature observations, after transforming to Fréchet standard distributions for locations: $s_1 = (38.91, -120.70)$, $s_2 = (34.44, -117.85)$, $s_3 = (40.72, -123.50)$, $s_4 = (35.38, -120.19)$ and $s_5 = (39.74, -122.20)$. This stations are circled in red on Figure 5.
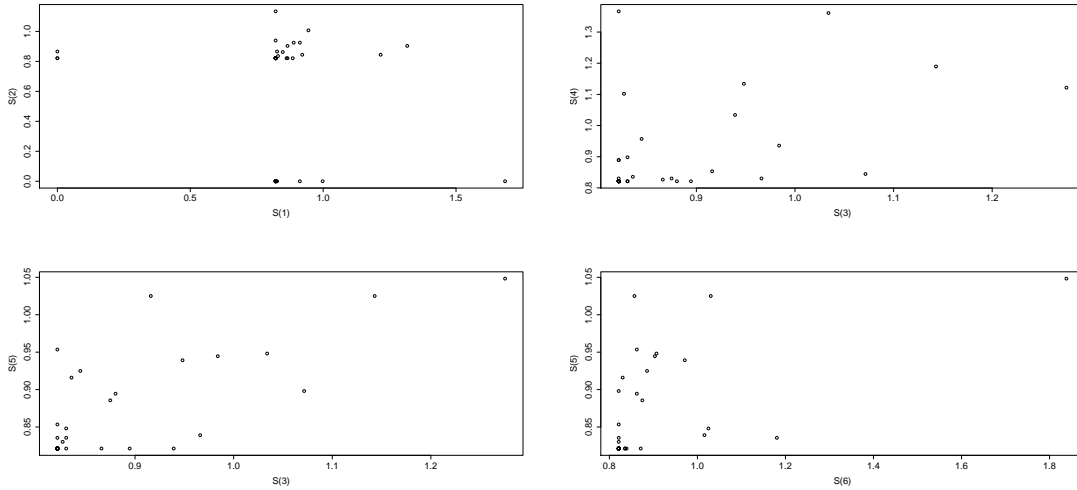
Fig. 7. Examples of scatter plots for rainfall observations, after transforming to Fréchet standard distributions for locations: $s_1 = (37.51, -122.03)$, $s_2 = (41.71, -122.50)$, $s_3 = (38.21, -119.01)$, $s_4 = (38.21, -122.28)$, $s_5 = (36.32, -119.64)$ and $s_6 = (34.02, -118.29)$. This stations are circled in green on Figure 5.

We fit our proposed model using the prior distributions suggested in Section 2.5. We use a Gaussian kernel given by $k(s - s^*) = \exp(-0.5||s - s^*||^2/\tau^2)/\sqrt{2\pi\tau^2}$ 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. The estimated values for α, in both cases, indicate the presence of significant spatial dependence. The posterior distribution of τ is concentrated on much higher values for the temperature examples than for the rainfall. Thus, in the former case, we obtain a spatial field that is much smoother than in the latter. More specifically, for the rainfall data, the posterior mean value of τ implies that the kernels is smaller than 0.01 for distances that are larger than abut 800 km, while for the rainfall data the equivalent distance is about 570 km. Regarding the estimation of the shape of the tails, we observe that the estimation of ξ is very sharp in both examples. The Monet Carlo method selects one specific value in all the iterations. For the maximum temperature observations ξ is estimated as 0, indicating that the distribution has exponential tails. For the rainfall observations, ξ is estimated as 1, providing evidence of a long tail behavior. Further evidence of differences in the tails of the temperature and rainfall distributions is given by the posteriors of β in each case.

To quantify the effect of asymptotic dependence we show in Table 3 the ratio between $Pr(Y(s_i) > u(s_i), Y(s_j) > u(s_j))$ and $Pr(Y(s_i) > u(s_i) \times Pr(Y(s_j) > u(s_j))$ calculated

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

Parameter	Maximum Temperature				
	τ	α	β	σ	$P(\xi)\{+,0,-\}$
M (CI)	6.82 (6.53;7.16)	0.597 (0.595;0.600)	1.10 (0.20;4.06)	24.22 (23.56;24.88)	{0,1,0 }
Parameter	Rainfall				
	τ	α	β	σ	$P(\xi)\{+,0,-\}$
M (CI)	2.61 (2.52;2.71)	0.605 (0.603;0.607)	3.65 (3.25;4.09)	144.46 (138.37;150.64)	{1,0,0 }

Table 3. Ratios of the probabilities of exceeding the threshold calculated using the joint distribution over the product of the marginals. The values in parentheses correspond to the distances between sites in 100 kms

	Temperature				Rainfall				
	s_2	s_3	s_4	s_5	s_2	s_3	s_4	s_5	s_6
s_1	1.22 (5.30)	1.50 (3.33)	1.43 (3.57)	1.60 (1.71)	1.48 (4.23)	1.56 (3.10)	1.61 (0.74)	1.50 (2.67)	1.25 (5.12)
s_2	-	1.09 (8.45)	1.51 (2.52)	1.14 (6.86)	-	1.40 (4.94)	1.52 (3.51)	1.32 (6.10)	1.13 (8.77)
s_3	-	-	1.24 (6.28)	1.61 (1.63)	-	-	1.53 (3.27)	1.58 (1.99)	1.35 (4.25)
s_4	-	-	-	1.32 (4.80)	-	-	-	1.46 (3.25)	1.21 (5.79)
s_5	-	-	-	-	-	-	-	-	1.46 (2.67)

from our fitted model, for the points in of Figures 6 and 7. The values in the table show a strong dependence between sites, especially at short distances. This confirms the results of the exploratory analysis. Figures 8 and 9 show maps of the probabilities of threshold exceedance and those of some high quantiles for rainfall and maximum temperature, respectively.

To explore the behavior of the return levels we consider return level plots that include credibility intervals for each site. shows shows an exampe for one station.In Figure 10 we compare the posterior mean of the return levels for six different stations, which are respectively the points showed in Figures 8 and 9. For the temperature data, we can see that the highest return is in the southern coast, while the lowest return level is in extreme north. For the rainfall data, the highest return is in the north, and the lowest is in extreme south of state

5.   Conclusion and discussion

We have presented a new model for the excesses above a threshold of spatially referenced observations. The model is based on a multiplicative structure that is suggested by a constructive definition of generalized Pareto processes. To achieve spatial coherence, the model uses a transformation of a linear combination of positive stable random variable,
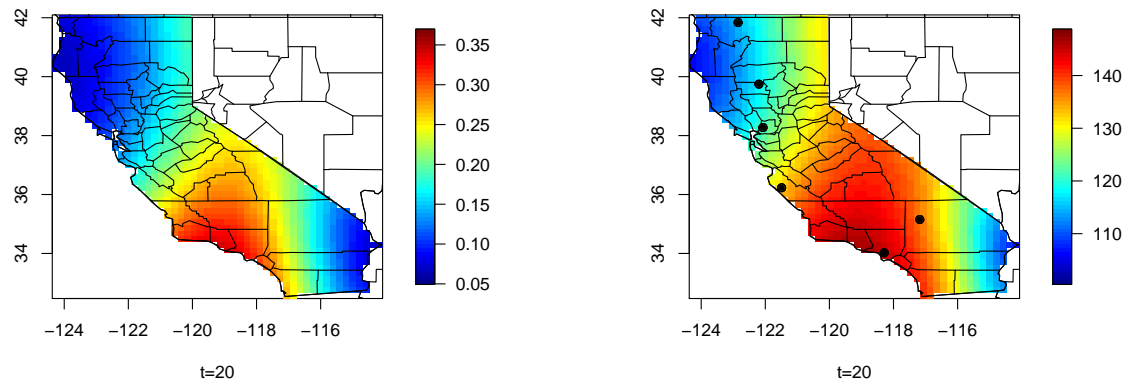
Fig. 8. $P(X(s_i) > 10\,^\circ\text{C})$ (left) and 20-year return level (right) for the maximum winter temperature in the State of California.
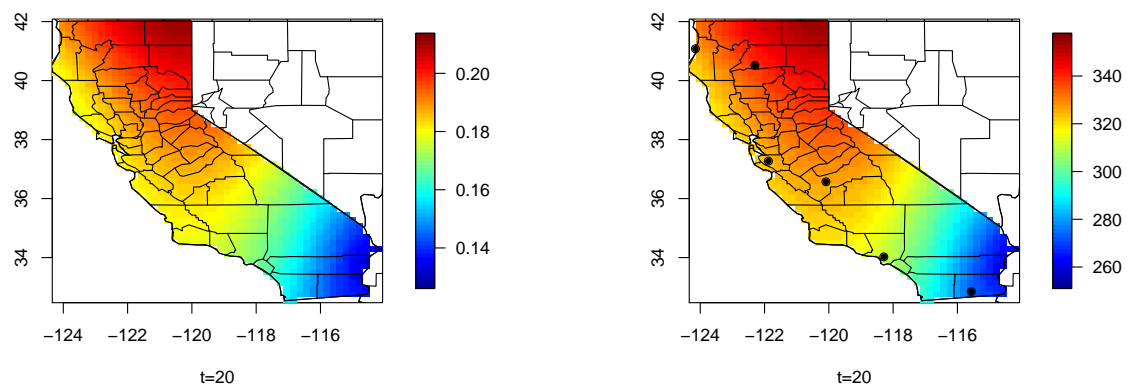

Fig. 9. $P(X(s_i) > 100$ mmu) (left) and 20-year return (right) for the winter precipitation over the State of California.
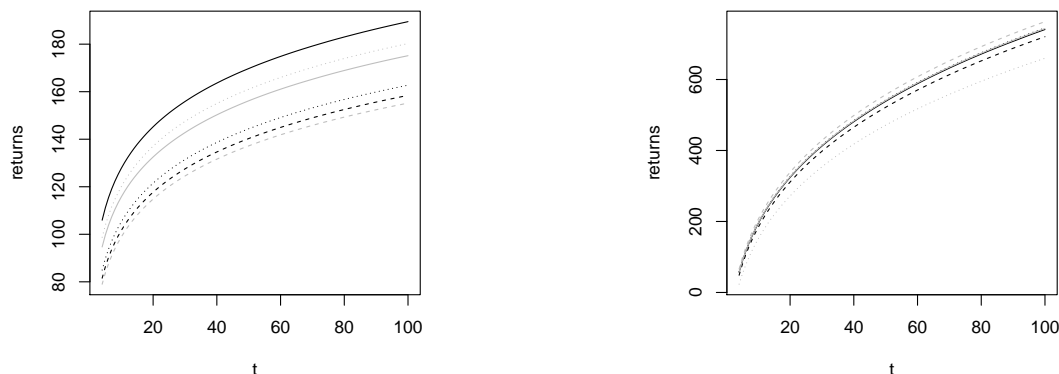
Fig. 10. Posterior mean of returns for six different locations. Left: Temperature (Locations marked as dots in Figure 8). Right: Rainfall (Locations marked as dots in Figure 9).

whose coefficients are defined by means of a kernel function acting on the spatial domain. The model is able to capture the whole possible range of spatial dependence, as well as tail dependence. For some parameter values, it belongs to the domain of attraction of a max-stable process. Moreover, simulations show that the proposed model is able to capture clustered structures in space that are typical of fields of extreme values. 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.

There are a number of natural extensions to he model presented here. It is important to incorporate information from spatially-varying covariates, such as elevation. Unfortunately it is not obvious how to include an additive or multiplicative effect of a covariate to the structure of the proposed model. The most natural way to proceed is to follow the approach in Cooley et al. (2007), and assume that σ and β are transformations of Gaussian proceses, whose means depend linearly on the covariates of interest. Yet another extension is to model the threshold as an unknown, possibly space-varying, parameters, in the spirit of Nascimento et al. (2012).

A.   Adaptive Monte Carlo Algorithm

- Sampling B_l

  Positive Stable random variables do not have a closed form density. Reich and Shaby (2012) use the auxiliary variable technique of Stephenson (2009), that introduces variables $B_l \in (0, 1)$ so that

$$p(A, B \mid \alpha) = \frac{\alpha A^{-1/(1-\alpha)}}{1 - \alpha} c(B) \exp \left[-c(B) A^{-\alpha/(1-\alpha)} \right]$$

  where $c(B) = \left\{ \left[\frac{\sin(\alpha \pi B)}{\sin(\pi B)} \right]^{1/(1-\alpha)} \frac{\sin((1-\alpha)\pi B)}{\sin(\alpha \pi B)} \right\}$. Then, marginalizing B_l, $A_l \sim PS(\alpha)$. Within the Monte Carlo algorithm, B_l is an auxiliary variable, that does not appear in the likelihood described in (6), its full conditional distribution is proportional to $p(A_l, B_l \mid \alpha)$. As $B_l \in (0, 1)$, a Beta distribution is used as proposal. We take the mean as the sample of B_l from the previous step of the chain and the variance is controlled using the technique of Roberts and Rosenthal (2009).

- Sampling A

  For each location $l = 1, ..., n$ we sample A_l from a log-Normal distribution, whose mean is the sample of A_l in the previous step. The variance of the proposal distribution is controlled using the technique of Roberts and Rosenthal (2009). The full conditional distribution of the vector $(A_1, ..., A_n)$ is given by

$$p(A_1, ..., A_n \mid \alpha, \xi, \sigma, \beta, B, \tau) \propto \prod_{i=1,...,n} \left[\left(\theta(s_i)^{\beta V_j(s_i)} (1 - \theta(s_i)^\beta)^{(1-V_j(s_i))} \right) \times p(A_i, B_i \mid \alpha) \right].$$

  where $\theta(s_i) = \exp\{-\sum_{l=1}^{L} A_l w_l(s_i)^{1/\alpha}\}^{1/\beta}$. Sampling of $(A_1, ..., A_n)$ is performed using the a block Metropolis step.

- Sampling σ

  The parameter σ is sampled from a proposal distribution, that corresponds to a gamma, whose mean is the value of σ from the previous step for the chain, and the proposed variance is controlled using the technique of Roberts and Rosenthal (2009). The proportional of posterior of these parameters is given by

$$p(\sigma \mid \xi, \beta, \mathbf{A}, \mathbf{B}, \alpha, \tau) \propto \prod_{i=1,...,n} \prod_{j=1,...,m} \left(f_G(Y_j(s_i) \mid u, \sigma/\beta, \xi/\beta)^{I(Y_j(s_i) > u(s_i))} \right) \times p(\sigma).$$

Where f_G is the GPD distribution given by (1). Samples are obtained from a Metropolis step.

- Sampling ξ

  Proposed samples of ξ are sampled from a discrete uniform distribution on $\{-1, 0, 1\}$, with probabilities $\{1/3, 1/3, 1/3\}$. The posterior is given as

$$p(\xi \mid \sigma, \beta, \mathbf{A}, \mathbf{B}, \alpha, \tau) \propto \prod_{i=1,\dots,n} \prod_{j=1,\dots,m} \left(f_G(Y_j(s_i) \mid u, \sigma/\beta, \xi/\beta)^{I(Y_j(s_i)>u(s_i))} \right) \times p(\xi).$$

  which is sampled directly.

- Sampling β

  The parameter β is sampled from a Gamma distribution whose mean is obtained from the previous step of the chain, and the proposed variance is controlled using the technique of Roberts and Rosenthal (2009). The full conditional is given as

$$p(\beta \mid \sigma, \xi, \mathbf{A}, \mathbf{B}, \alpha, \tau) \propto \prod_{i=1,\dots,n} \prod_{j=1,\dots,m} \left(f_G(Y_j(s_i) \mid u, \sigma/\beta, \xi/\beta)^{I(Y_j(s_i)>u(s_i))} \right) \times$$
$$\prod_{i=1,\dots,n} \prod_{j=1,\dots,m} \left(\theta(s_i)^{\beta V_j(s_i)} (1 - \theta(s_i)^{\beta})^{(1-V_j(s_i))} \right) \times p(\beta).$$

  Given that $\theta(s_i) = \exp\{-\sum_{l=1}^{L} A_l w_l(s_i)^{1/\alpha}\}^{1/\beta}$, the full conditional can be simplified to

$$p(\beta \mid \sigma, \xi, \mathbf{A}, \mathbf{B}, \alpha, \tau) \propto \prod_{i=1,\dots,n} \prod_{j=1,\dots,m} \left(f_G(Y_j(s_i) \mid u, \sigma/\beta, \xi/\beta)^{I(Y_j(s_i)>u(s_i))} \right) p(\beta).$$

  This expression is used to sample β using a Metropolis step.

- Sampling α

  The parameter α, that controls the spatial dependence, are sampled from a Beta distribution, whose mean is obtained from the previous step of the chain, and the proposed variance is controlled using the technique of Roberts and Rosenthal (2009). The full conditional is given as

$$p(\alpha \mid \xi, \sigma, \beta, \mathbf{A}, \mathbf{B}, \tau) \propto \prod_{i=1,\dots,n} \prod_{j=1,\dots,m} \left(\theta(s_i)^{\beta V_j(s_i)} (1 - \theta(s_i)^{\beta})^{(1-V_j(s_i))} \right) \times p(\mathbf{A}, \mathbf{B} \mid \alpha) p(\alpha).$$

  This distribution is sampled using a Metropolis step.

- Sampling τ

  The parameter τ, the Kernel Bandwidth, is sampled from a Gamma distribution, whose mean is obtained from the previous step of the chain, and the proposed variance is controlled using the technique of Roberts and Rosenthal (2009). The full conditional is given as

  $$p(\tau \mid \xi, \sigma, \beta, \mathbf{A}, \mathbf{B}, \alpha) \propto \prod_{i=1,\ldots,n} \prod_{j=1,\ldots,m} \left(\theta(s_i)^{\beta V_j(s_i)} (1 - \theta(s_i)^{\beta})^{(1 - V_j(s_i))} \right) \times p(\tau).$$

  This is sampled using a Metropolis step.

References

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

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., Heffernan, J. and Tawn, J. (1999) Dependence measures for extreme value analysis. *Extremes*, **2:4**, 339–365.

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*, **4**, 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.

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

Nascimento, F. F., Gamerman, D. and Davis, R. (2015) A bayesian semi-parametric approach to extreme regime identification. *Brazilian Journal of Probability and Statistics*.

Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012) A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, **22**, 661–675.

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.

— (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.

Roberts, G. O. and Rosenthal, J. S. (2009) Examples of adaptive mcmc. *Journal of Computation and Graphical Statistics*, **18**, 349–367.

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.