A Bayesian semi-parametric approach to extreme regime identification

Fernando Ferraz do Nascimento[a], Dani Gamerman[b] and Richard Davis[c]

[a]*Universidade Federal do Piauí*
[b]*Universidade Federal do Rio de Janeiro*
[c]*Columbia University*

Abstract. The limiting tail behaviour of distributions is known to follow one of three possible limiting distributions, depending on the domain of attraction of the observational model under suitable regularity conditions. This work proposes a new approach for identification and analysis of the shape parameter of the GPD as a mixture distribution over the three possible regimes. This estimation is based on evaluation of posterior probabilities for each regime. The model-based approach uses a mixture at the observational level where a Generalized Pareto distribution (GPD) is assumed above the threshold, and mixture of Gammas distributions is used under a threshold. The threshold is also estimated. Simulation exercises were conducted to evaluate the accuracy of the model for various parameter settings and sample sizes, specifically in the estimation of high quantiles. They show an improved performance over existing approaches. The paper also compares inferences based on Bayesian regime choice against Bayesian averaging over the regimes. Results of environmental applications show the correctly identifying the GPD regime plays a vital role in these studies.

1 Introduction

Natural disasters have become an issue of increasing concern worldwide. According to Parmesan, Root and Willing (2000), changes in extremes of temperature are more responsible for impacts in nature than change in mean temperature. Sang and Gelfand (2009) says that trends in climate extremes are more important than trends in the mean climate.

Given the importance of extreme values in the aforementioned situations, Extreme Value Theory (EVT) has proved vital for the prediction of these rare but high destructive events. Through EVT, one can formulate a specific model to estimate rare events and their odds by generalized extreme value distribution (GEV), when analyzing maxima of data blocks, or by generalized Pareto distribution (GPD), when analyzing excesses above a large threshold.

Pickands (1975) proved that if X is a random variable whose distribution function F, with endpoint x_F, is in the domain of attraction of a GEV distribution, then

as $u \to x_F$, the conditional distribution function $F(x|u) = P(X \leq u + x | X > u)$ is the distribution function of a generalized Pareto distribution (GPD), whose density is provided below:

$$g(x|\Psi) = \begin{cases} \dfrac{1}{\sigma}\left(1 + \xi \dfrac{(x-u)}{\sigma}\right)^{-(1+\xi)/\xi}, & \text{if } \xi \neq 0, \\ \dfrac{1}{\sigma}\exp\{-(x-u)/\sigma\}, & \text{if } \xi = 0, \end{cases} \quad (1.1)$$

where $x - u > 0$ for $\xi \geq 0$ and $0 \leq x - u < -\sigma/\xi$ 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 $\xi \geq 0$. These different possibilities for the tail shape are named Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$).

Smith (1984) proposed parameter estimation in this family of models via maximum likelihood. He showed that the maximum likelihood estimators do not obey the regularity conditions if $\xi \in (-1, -0.5)$, and do not exist if $\xi < -1$. Cooley et al. (2012) revises some estimation methods, such as using a Poisson process for the occurrence of events.

Extreme value theory is particularly useful for determination of larger quantiles, that is, q-values satisfying $P(X > x_p) = 1 - p$ for large values of p. The quantile x_q can be found by inversion of the distribution function of the GPD distribution function $p = G(q|\xi, \sigma, u): \in [0, 1]$. This gives

$$x_p = \begin{cases} u + \dfrac{((1-p)^{-\xi} - 1)\sigma}{\xi}, & \text{if } \xi \neq 0, \\ u - \sigma \log(1-p), & \text{if } \xi = 0. \end{cases} \quad (1.2)$$

Most other exceedance data analyses make no assumptions at all about the distribution of values below the threshold, or make estimation empirically of the proportion of data higher a threshold, considering the threshold fixed. Various models have been proposed for the distribution of values below the threshold. Scarrott and Macdonald (2012) makes an encompassing review of possible approaches, including Frigessi, Haug and Rue (2002), Behrens, Gamerman and Lopes (2004), Tancredi, Anderson and O'Hagan (2006), Macdonald et al. (2011). In this work, we used the approach of Nascimento, Gamerman and Lopes (2012), that used a mixture of Gammas below the threshold. They showed the importance of this model, having more precise prediction of data in nonexceedance part, and good results in estimation of the threshold.

The sign of the GPD shape is also important in identifying the support of the underlying process. Usual approaches for studying the shape parameter either assume knowledge of the limiting regime or assume the shape to vary continuously over its possible values. One of the main difficulties there is to testing the hypothesis whether ξ is significantly different from zero. The most works related with GPD estimation does not consider the part of Gumbel, even so the estimation of ξ is very close to 0.

In this study, identification of the Gumbel tail behavior is explicitly considered, since in many of the applications cited above this seems to take place. Our approach allows for evaluation of the probabilities of the three limiting regimes, afforded by a mixed distribution for the shape parameter ξ. Stephenson and Tawn (2004) provided a similar approach in the study threshold exceedances via the GEV distribution. They only considered the possibilities $\xi = 0$ and $\xi \neq 0$, the priors for (ξ, σ) follow Coles and Tawn (1996) and the prior probabilities of the regimes are fixed. Our approach allows for different specification for each of the possible 3 regimes.

Section 2 introduces the observational model, which is based on the Nascimento, Gamerman and Lopes (2012) model that combines a finite mixture of Gammas below the threshold, and a GPD above. A Bayesian approach will be presented for parameter estimation of all model parameters. A distinctive feature is the mixed prior distribution for ξ, with a positive probability of this parameter being equal to 0. Section 3 shows simulation results from the proposed model, which illustrate the efficiency of the estimation method and the accuracy in estimating extreme quantiles. The method is compared with the method proposed in Nascimento, Gamerman and Lopes (2012), that consider an absolutely continuous distribution for ξ. Section 4 illustrates our procedure applied to two applications: rainfall data from Portugal and river flow in Puerto Rico. Results show that in some applications, there is a high probability of the tail shape ξ parameter equalling zero. Section 5 summarizes the main conclusions of this work.

2 The model

In this section, we propose a new approach to analyse the observational model of Nascimento, Gamerman and Lopes (2012), that takes into account the complete range of the data, bearing in mind the need to estimate extreme quantiles. Nascimento, Gamerman and Lopes (2012) showed that a mixture of Gamma distributions provides an efficient procedure for non-parametric density estimation below a threshold for positive data. Their mixture model with Gamma´s and GPD (MGPD, in short) has density given by

$$f(x|\theta, \mathbf{p}, \Psi) = \begin{cases} \sum_{j=1}^{k} p_j f_G(x|\mu_j, v_j), & \text{if } x \leq u, \\ \left[1 - \sum_{j=1}^{k} p_j F_G(u|\mu_j, v_j)\right] g(x|\Psi), & \text{if } x > u, \end{cases} \quad (2.1)$$

where f_G and F_G are respectively, the density and cumulative functions of a Gamma distributions, and g is the GPD density with parameters $\Psi = (\xi, \sigma, u)$. An alternative for data over the entire line would be a Gaussian mixture as in Diebolt

and Robert (1994), Richardson and Green (1997), Frühwirth-Schnatter (2001) and Macdonald et al. (2011).

One could argue that the threshold u has no physical meaning, being a mere artifact of an approach based on an asymptotic result. While we essentially agree with this view, it is also important for users to learn about where the GPD regime can be assumed to hold. In fact, this is a key point in applying the GPD and users will benefit from the estimation results about the threshold. Of course, if one is uncomfortable with this parametrisation, one can always integrate the threshold out and proceed from there. The reader is referred to the comprehensive review of Scarrott and Macdonald (2012) for further discussion.

Parameters are estimated under the Bayesian paradigm, assigning a probability distribution for the parametric vector that combines information from the dataset and the prior distribution via Bayes theorem. The novelty is in the characterization of the shape parameter ξ, and this will be made explicit through its prior distribution. We are specifically concerned here in the determination of the limiting regime to bring the data belong. So, the prior distribution for ξ will explicitly consider this distinction among the regimes through a mixed distribution.

A difficulty arises from the discreteness of the Gumbel regime, associated with $\xi = 0$. If a continuous distribution is assumed for ξ, testing for Gumbel could be performed by either checking if the posterior credibility interval of high level (say 0.95) includes 0 or by evaluating the probability of $\xi > 0$, where π denotes the posterior distribution. A large value for this probability would indicate that the Gumbel regime is not supported by the data. The latter procedure is successfully used for testing by West and Harrison (1997) in the context of state-space models.

2.1 Prior distribution

The parametric vector for density (2.1) is composed of three groups of parameters: (a) the parameter $\theta = (\mu, \eta)$ of the k-mixture of Gamma densities $f_G(x|\mu_j, \eta_j)$ with means μ_j and shape parameters η_j, $j = 1, \ldots, k$, where $\mu = (\mu_1, \ldots, \mu_k)$ and $\eta = (\eta_1, \ldots, \eta_k)$; (b) the mixture weights $\mathbf{p} = (p_1, \ldots, p_k)$; (c) the GPD parameter $\Psi = (\xi, \sigma, u)$.

The prior distribution for θ is the same as the one used in Nascimento, Gamerman and Lopes (2012), based on work of Wiper, Rios Insua and Ruggeri (2001). An important question about these parameters is their identification. Diebolt and Robert (1994) and Frühwirth-Schnatter (2001) showed that the identifiability of the parameters of the finite mixture of densities is only possible if the parametric space is restricted to $C(\mu) = \{\mu | 0 < \mu_1 < \mu_2 < \cdots < \mu_k\}$. Therefore, the prior for μ is taken in the form

$$p(\mu_1, \ldots, \mu_k) = K \prod_{i=1}^{k} f_{IG}(\mu_i | a_i/b_i, b_i) I(\mu_1 < \cdots < \mu_k),$$

where $K^{-1} = \int_{C(\mu)} \prod_{i=1}^{k} p(\mu_i) \, d(\mu_1, \ldots, \mu_k)$ and f_{IG} is the inverse Gamma density with parameters defined as in the corresponding Gamma.

The prior for shape parameters η is taken as a product of Gamma distributions with $\eta_j \sim G(c_j/d_j, c_j)$, for some positive constants c_j and d_j, for $j = 1, \ldots, k$. The prior for the weights is taken as $\mathbf{p} \sim D_k(\gamma_1, \ldots, \gamma_k)$, that represents the Dirichlet distribution with density proportional to $\prod_{i=1}^{k} p_i^{\gamma_i}$.

The prior distribution for the threshold u is assumed to be $N(\mu_u, \sigma_u^2)$, as suggested by Behrens, Gamerman and Lopes (2004). This distribution is truncated at the lower data limit but usual choices of hyperparameters make the effect of this truncation negligible. Nascimento, Gamerman and Lopes (2012) describes how to specify this value even with scarce prior information. The prior distribution for the scale parameter σ is given in the usual non-informative format for the scale parameter $p(\sigma) \propto \sigma^{-1}, \sigma > 0$. This can be shown to be the marginal specification for the Jeffreys prior proposed for GPD model in Castellanos and Cabras (2007) for (σ, ξ).

A mixed distribution is proposed for ξ, considering separately the probabilities associated with the three possible limiting tail behaviors: Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) and Weibull ($\xi < 0$).

When ξ is positive or negative, continuous densities are assumed for these regions. This mixed setting can be rephrased with the insertion of latent quantities (Z_ξ, Q_ξ). The three-dimensional quantity $Z_\xi = (Z_\xi^+, Z_\xi^0, Z_\xi^-)$, where $Z_\xi^+ + Z_\xi^0 + Z_\xi^- = 1$, indicates the signal of ξ, where $Z_\xi^+ = 1$ indicates that $\xi > 0$, $Z_\xi^0 = 1$ indicates that $\xi = 0$ and $Z_\xi^- = 1$ indicates that $\xi < 0$. The importance of Z_ξ is that the probability of the data tail behavior can be obtained through its distribution. The quantity Q_ξ shows the probability of ξ be positive, negative or null, given by $Q_\xi = (q_\xi^+, q_\xi^0, q_\xi^-)$, where $q_\xi^+ + q_\xi^0 + q_\xi^- = 1$. The joint prior of these parameters is given by

$$p(\xi, Z_\xi, Q_\xi) = p(\xi | Z_\xi, Q_\xi) p(Z_\xi | Q_\xi) p(Q_\xi)$$
$$= p(\xi | Z_\xi) p(Z_\xi | Q_\xi) p(Q_\xi),$$

where conditional independence between ξ and Q_ξ given Z_ξ is assumed. The distribution of $\xi | Z_\xi$ is assigned in the following specification

$$\xi | Z_\xi^+ = 1 \sim \text{Gamma}(a_\xi, b_\xi), \qquad \xi | Z_\xi^0 = 1 \sim \delta_{\xi=0} \quad \text{and}$$
$$\xi | Z_\xi^- = 1 \sim U(-0.5, 0),$$
(2.2)

where the parameters a_ξ, b_ξ may be chosen so that the prior distribution of the positive part is vague, δ is the Dirac function.

Thus, the range of negative values of ξ is limited to $[-0.5, 0]$, as situations where $\xi < -0.5$ are extremely rare in environmental data (Coles and Tawn, 1996). A number of other options were considered for the prior of $\xi | Z_\xi^- = 1$. We also

used a negative Gamma prior with same parameters as those used for $\xi|Z_\xi^+ = 1$, and the results were the same. Inference does not also seem to be affected by the specific choices of hyperparameters made above. Simulations with other values of a_ξ and b_ξ returned virtually unchanged results. Similarly, a shifted Beta prior over $[-0.5, 0]$ were considered for the negative range of ξ and results remained unchanged.

The conditional distribution $Z_\xi|Q_\xi$ is assigned a multinomial prior with parameters $(q_\xi^+, q_\xi^0, q_\xi^-)$

$$p(Z_\xi|Q_\xi) \propto (q_\xi^+)^{z_\xi^+}(q_\xi^0)^{z_\xi^0}(q_\xi^-)^{z_\xi^-}$$

$$\text{for } (z_\xi^+, z_\xi^0, z_\xi^-) = (0, 0, 1), (0, 1, 0), (1, 0, 0),$$

and 0, otherwise. Finally, Q_ξ was given a Dirichlet distribution with parameter $\alpha_\xi = (\alpha_+, \alpha_0, \alpha_-)$. Then,

$$p(Q_\xi) \propto (q_\xi^+)^{\alpha_+}(q_\xi^0)^{\alpha_0}(q_\xi^-)^{\alpha_-}$$

$$\text{for } (q_\xi^+, q_\xi^0, q_\xi^-) \in \{(x, y, z)|x, y, z \geq 0, x + y + z = 1\},$$

and 0, otherwise. In the lack of prior information, the vector α_ξ may be chosen to provide the desired amount of information on Q_ξ. We opted to assign a distribution to Q_ξ (unlike Stephenson and Tawn, 2004, who assign prior values directly to Q_ξ) in order to introduce as little prior information as possible and to let the data govern inference. Note that and the choice of this value affects the posterior probabilities, as shown in their Table 1.

The above prior distribution can be marginalized with respect to Z_ξ and Q_ξ by analytic integration, leading to a mixture distribution for ξ containing continuous and discrete components. The above formulation is retained because it simplifies computations and also highlights the different regimes of tail behavior by readily providing their posterior probabilities. These features are explored in the next sections.

Table 1 *Posterior probability of Z_ξ in different values α_0, with $\alpha_+ = \alpha_- = 0.1$*

	$\alpha_0 = 0.01$			$\alpha_0 = 0.1$			$\alpha_0 = 1$		
ξ	Z^+	Z^0	Z^-	Z^+	Z^0	Z^-	Z^+	Z^0	Z^-
0	0.176	0.487	0.337	0.139	0.492	0.369	0.135	0.514	0.350
0.1	0.976	0.019	0.005	0.982	0.015	0.003	0.975	0.020	0.005

Z^v indicates the value of $P(z_\xi^v = 1)$, for $v = +, 0, -$.

2.2 Posterior distribution

The posterior distribution function to the parameters is obtained by combining the observations of a sample $\mathbf{x} = (x_1, \ldots, x_n)$ of size n from the model (2.1), with the prior distribution of the parametric vector described in Section 2.1. The posterior density is given by

$$
\pi(\theta, \mathbf{p}, \Psi, Z_\xi, Q_\xi | \mathbf{x})
$$

$$
\propto \prod_{i:x_i \leq u} \left(\sum_{j=1}^{k} p_j f_G(x_i | \mu_j, \eta_j) \right) \prod_{i:x_i \geq u} \left[1 - \sum_{j=1}^{k} p_j F_G(u | \mu_j, \eta_j) \right] g(x_i | \Psi)
$$

$$
\times \prod_{j=1}^{k} \left[p_j^{\gamma_j} \eta_j^{c_j - 1} \exp\left(-\frac{d_j}{\eta_j} \right) \mu_j^{a_j - 1} \exp\left(-\frac{b_j}{\mu_j} \right) \right] \frac{1}{2} \exp\left(-\frac{(u - \mu_u)}{2\sigma_u} \right)^2
$$

$$
\times \frac{1}{\sigma} p(\xi | Z_\xi) \left(q_\xi^+ \right)^{z_\xi^+ + \alpha_+} \left(q_\xi^0 \right)^{z_\xi^0 + \alpha_0} \left(q_\xi^- \right)^{z_\xi^- + \alpha_-},
$$

(2.3)

where the first line above comes from the likelihood, the second line refers to the prior density of parameters, where $p(\xi | Z_\xi)$ is the density of (2.2), with a mixed distribution.

Inference cannot be performed analytically and approximating MCMC methods are used (Gamerman and Lopes, 2006). Convergence was assessed by running four parallel chains with different starting values. Chain convergence was visually assessed through the trace plots of the chains. Parameters were separated into blocks and each block was updated according to a Metropolis rule, since most do not have full conditional densities in recognizable form. Among the exception, one can cite Q_ξ, whose full conditional distribution is known.

Stephenson and Tawn (2004) use reversible jump MCMC to moves from space where $\xi = 0$ to a space where $\xi \neq 0$. Here, we make the movements of each regime using the posterior probabilities of Z_ξ, and when $z_\xi^0 = 1$, we consider a degenerate posteriori of ξ ($P(\xi = 0 | z_\xi^0 = 1) = 1$). Basically the value of $z_\xi^{(t)}$ determines the regime at time t. Each of the 3 possible regimes is proposed with equal probabilities (1/3). Posterior distribution of ξ has discrete full conditional component when $Z_\xi = 0$, but has continuous full conditional when Z_ξ is $+1$ or -1. The Associate Editor noted that the "terms in Metropolis acceptance rate are densities with respect to an unusual dominating measure—not the usual Lebesgue, but Lebesgue modified by adding a point mass on $\xi = 0$. Then the rate is well-defined (as a Radon–Nikodym derivative) because these point masses cancel out". This is a similar point to that made on page 317 of Hastie and Green (2012). The MCMC sampler deals correctly with this issue. Metropolis–Hastings acceptance probabilities determines the regime at time $t + 1$. Details of the sampling algorithm are provided in the Appendix and are similar to the application of the algorithm seen in Nascimento, Gamerman and Lopes (2012). Proposal variances were tuned with

a variation of the method proposed by Roberts and Rosenthal (2009), with variances smaller than their target value since our blocks are multidimensional.

2.3 Bayesian inference for higher quantiles

The model proposed here can be seen as mixture of different models M_1, M_2 and M_3 characterized by the 3 possible limiting tail regimes. This distinction is entirely based on a single parameter ξ with all other parameters remaining the same under the possible sub-models. There are a few possibilities for making inference about quantities of interest in such cases. Of particular interest in EVT is estimation of higher quantiles, extrapolated beyond the data points, that are merely functions of model parameters.

Consider estimating a quantity of interest δ, for example, a large quantile. The posterior probability of each model is given by

$$\pi(M_l|\mathbf{y}) = \frac{\pi(\mathbf{y}|M_l)\pi(M_l)}{\sum_{j=1}^{3} \pi(\mathbf{y}|M_j)\pi(M_j)} \qquad \text{for } l = 1, 2, 3, \qquad (2.4)$$

and the posterior distribution of δ is given by

$$p(\delta|\mathbf{y}) = \sum_{j=1}^{3} p(\delta|\mathbf{y}, M_j)\pi(M_j|\mathbf{y}).$$

Following Draper (1995), inference should be based on this distribution, obtained as a weighted average of all possible models, hence the name Bayesian model averaging (BMA). An alternative route is provided by Bayesian model choice (BMC) where one of the models (say s) is chosen according to some criteria and inference about δ is based on the conditional posterior $p(\delta|\mathbf{y}, M_s)$. The most obvious criteria is to choose the most probable model. An interesting discussion on the relative merits of BMA and BMC at a less formal level can be seen in Wasserman (2000) and references therein. The next sections present and compare both approaches to inference about higher quantiles.

3 Simulation

Simulation studies were performed in different settings of parameter values to better understand features of parameter estimation, and to verify that the proposed methodology provides accurate and credible results. Particular attention is given to the variation of the shape parameter ξ, which is the main target of this study.

The simulation exercise was performed with samples of sizes 1000, 2500 and 10,000. Data below the tail were simulated by a mixture of two Gamma distributions with $\mu = (2, 8)$, $\eta = (4, 8)$, and weight vector $\mathbf{p} = (2/3, 1/3)$. The threshold was fixed at the 80% quantile of the generating mixture of Gamma density and the

scale parameters was $\sigma = 2$. The shape parameter ξ was fixed with the following values $\xi = (-0.4, -0.1, 0, 0.1, 0.4, 1)$.

The prior distributions for mixture parameters were $\mu_j \sim \text{IG}(2.1, 5.5)$ and $\eta_j \sim G(6, 0.5)$, for $j = 1, \ldots, k$, and $\pi(\mathbf{p}) \sim D_k(1, \ldots, 1)$. These distributions have mean around the actual parameter value but large variance to represent lack of information: the prior variance for the μ_j's is 250 and the prior variance for η_j's is 24. The prior distributions for the GPD parameters were specified as follows: Jeffreys prior for σ, with $p(\sigma) \propto 1/\sigma$; the prior distribution described in (2.2) was assigned to ξ, with $a_\xi = b_\xi = 0.1$; and Q_ξ was assigned a vague Dirichlet prior with $\alpha_+ = \alpha_0 = \alpha_- = 0.1$, in a case where initially all regimes has the same probability. Table 1 suggests that the posterior is not sensitive to the value of prior parameters α_0 chosen to reflect vague prior information. The threshold was assigned as a normal prior distribution (Nascimento, Gamerman and Lopes, 2012).

Inference was performed via the MCMC algorithm. 50,000 iterations were used as burn-in and the next 30,000 iterations were kept for inference for simulations with a sample size $n = 1000$. 25,000 iterations to burn-in and 15,000 iterations were retained for inference were run for the simulations with $n = 2500$, while for simulations with sample size $n = 10,000$, 15,000 iterations to burn-in and 10,000 iterations were retained for inference. Inference was made after thinning at every 20 iterations in all simulations. The code was developed in OxMetrics5 (Doornik, 1996) in a HP Compaq 6005 Pro MT PC and 2Gb RAM. The processing time allowed for 35 (or 4) iterations per second when $n = 1000$ (or $n = 10,000$).

Figure 1 shows the posterior histogram distribution of tail parameters in simulations with $\xi = 0.4$. The distributions of the parameters are centered around the true values of the parameters used in the simulation. Precision was as high as in the model by Nascimento, Gamerman and Lopes (2012), with threshold prior variance of $\sigma_u^2 = 10$. Note that the entire posterior distribution of ξ is contained in the positive semi-line, because the sampled values of Z_ξ in all interactions of the chain were $z_\xi^+ = 1$. Estimation is more precise for larger data sizes, as expected.

Figure 2 shows the estimation of ξ, when the true value is equal to 0.1, in three different configurations of sample size. The parameter estimates are less accurate now as this value is close to 0, and there is a substantial probability that $\xi = 0$, specially for smaller sample sizes. The estimation becomes more accurate and the posterior distribution is located around the true value as sample size increases, with a few (when $n = 2500$) or rare (when $n = 10,000$) samples associated with the estimated value $\xi = 0$. Figure 2 also shows the estimation of ξ, when the true value is equal to 0, in three different configurations of sample size. Virtually the same comments are valid here with replacement of $\xi = 0.1$ by $\xi = 0$. It is interesting to notice the similarities between the two patterns despite the different expected behavior between $\xi = 0.1$ and $\xi = 0$.

Table 2 shows the posterior probability of parameter Z_ξ in all simulations. The tail behavior is correctly identified with posterior probability 1 for situations where
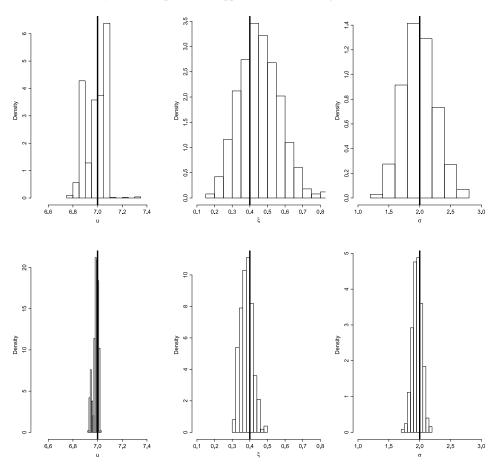
Figure 1 *Posterior histogram for the GPD parameters with $\xi = 0.4$: top row—$n = 1000$; bottom row—$n = 10,000$. Vertical lines: true value of parameter.*

ξ is away from 0 ($\xi = (-0.4, 0.4, 1)$), even for smaller sample sizes. The posterior $P(Z_\xi^0 = 1|\mathbf{x})$ in the case $\xi = 0$ is high for the three sample sizes, and seem to converge to 1 when the sample size increases. Care is required to study the situations where ξ is close to 0 ($\xi = (-0.1, 0.1)$). In those cases when the sample size is smaller, the estimation is less accurate. Thus, there is greater uncertainty about the sign of the parameter ξ, and in the simulations with $n = 1000$, the probability $P(Z_\xi^0 = 1|\mathbf{x})$ of incorrect detection was high. This does not occur for samples with $n = 10,000$, which provide more accuracy in detection of the sign of ξ, giving a probability of correct sign very close to 1.

3.1 Simulations from mixture of Gammas

Another simulation exercise is to consider the case where there is no threshold in the true model. In this case, the same simulations were performed as in the
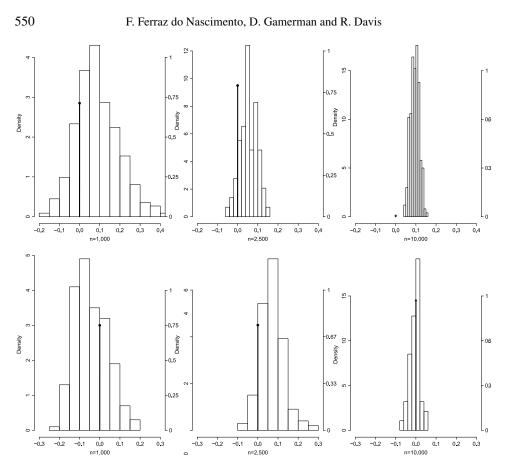
Figure 2 *Posterior histogram for ξ in simulations*: *top row—ξ = 0.1*; *bottom row—ξ = 0*. *Vertical lines*: *probability lumps at ξ = 0. Their respective posterior probabilities have the scales shown on the right y-axis and are given in Table* 2.

Table 2 *Posterior probability of Z_ξ in all simulations*

	$n = 1000$			$n = 2500$			$n = 10,000$		
ξ	Z^+	Z^0	Z^-	Z^+	Z^0	Z^-	Z^+	Z^0	Z^-
−0.4	0	0	1	0	0	1	0	0	1
−0.1	0.02	0.45	0.52	0.03	0.66	0.31	0	0	1
0	0.07	0.75	0.18	0.22	0.75	0.03	0.02	0.95	0.03
0.1	0.22	0.72	0.06	0.16	0.82	0.02	1	0	0
0.4	1	0	0	1	0	0	1	0	0
1	1	0	0	1	0	0	1	0	0

Z^v indicates the value of $P(z_\xi^v = 1)$, for $v = +, 0, -$.

Table 3 *Fit measures and ξ estimation for simulations of MG model*

Model	$n = 1000$		$n = 2500$		$n = 10,000$	
	MG	Prop.	MG	Prop.	MG	Prop.
DIC	4399	4445	11,074	11,070	44,432	44,422
BIC	4450	4464	11,131	11,153	44,497	44,519
$\hat{\xi}$		$-0.136\ (0.079)$		$-0.097\ (0.057)$		$-0.163\ (0.018)$
$P(Z_\xi = z_\xi^-)$		0.88		0.88		1

MG is the mixture of Gammas model. Prop. is the model proposed in this work. $\hat{\xi}$ is the posterior mean of ξ, with posterior standard deviations in brackets.

previous section, but allowing the original dataset to be drawn from mixture of Gamma distributions. The purpose of this section is to check what happens when we try to estimate the threshold in this situation.

In the first exercise, the same threshold prior was retained, with $\sigma_u^2 = 10$. The posterior mean of the threshold approached the maximum value of the observations, giving all the weight of the estimation for the Mixture of Gammas.

In a second exercise, the variance for the threshold was reduced to $\sigma_u^2 = 1$, reflecting a situation where almost all information about this threshold is already provided by the prior distribution. Table 3 shows the fit measures BIC (Schwarz, 1978) and DIC (Spiegelhalter et al., 2002) of for the estimation exercise. Results show that for n small, the MG model outperformed our proposal model as expected. However, as n gets larger, our MGPD model provides comparable results. Table 3 shows also the results of the posterior mean of ξ, and the probability of regime with $\xi < 0$. We can observe that when n increases, the tail distribution of Mixture of Gammas converges to the Weibull regime ($\xi < 0$). These results indicate that our MGPD model is a robust alternative for estimation, particularly if tail behavior is relevant. So, they can be safely used even the data is known to be drawn from an alternative model, specially for large datasets.

3.2 Extreme quantiles

A important evaluation criterion in the estimation of extreme values is the analysis of high quantiles of the distribution. These will also be compared here. Figure 3 illustrates the estimation of the posterior distribution of extreme quantiles x_p, given in (1.2) in two different simulations. High quantiles are well estimated in the proposed model even in the cases where the model mistakenly indicates with high probability the incorrect tail behavior. So imprecise regime identification for smaller sample sizes does not seem to affect estimation of high quantiles.

Table 4 shows the estimation of extreme quantiles x_p in two different configurations ($\xi = 0$ and $\xi = 0.1$). They constitute the cases where the proposed mixed model with a lump probability at $\xi = 0$ may contrast more from the MGPD model
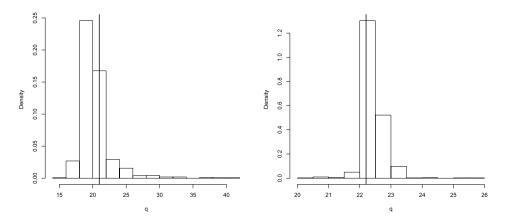
Figure 3 *p-quantiles of simulations*: *left*—$\xi = 0.1$, $n = 1,000$ *and* $p = 0.999$; *right*—$\xi = 0$, $n = 10,000$ *and* $p = 0.9999$. *Vertical lines*: *true values of quantiles*.

Table 4 *Extreme quantile x_p of simulations*

	$\xi = 0$				$\xi = 0.1$			
p	T	E	M	P	T	E	M	P
				$n = 1000$				
0.95	9.85	10.01	9.90	9.82	10.06	10.41	10.25	10.36
0.99	13.06	13.04	12.85	12.92	14.09	14.03	14.44	14.26
0.999	17.67	16.93	16.65	17.30	21.10	20.37	21.91	20.27
0.9999	22.28	N/A	20.14	21.65	29.93	N/A	32.04	27.13
				$n = 10,000$				
0.95	9.85	9.80	9.91	9.87	10.06	9.93	9.96	9.96
0.99	13.06	12.77	13.11	13.12	14.09	13.86	13.83	13.84
0.999	17.67	17.00	17.54	17.78	21.10	20.86	20.47	20.56
0.9999	22.28	22.42	21.79	22.43	29.93	27.22	28.70	28.99

T—True, E—Empirical, M—MGPD, P—Proposed.

with continuous prior for ξ. The quantiles of the mixed model were compared with empirical quantiles, and with the quantiles obtained from the continuous MGPD model of Nascimento, Gamerman and Lopes (2012). The comparison favors the empirical quantiles for $n = 1000$ with the continuous MGPD providing better estimates for $\xi = 0.1$ and the mixed MGPD providing better estimates when $\xi = 0$. The difference between the 2 models becomes smaller for larger sample sizes with a slight superiority for the quantiles proposed here.

Figure 4 shows point estimates of returns with our approach and MGPD. The results are very similar but our approach performs uniformly better when $\xi = 0.1$ and better for early periods when $\xi = 0$.
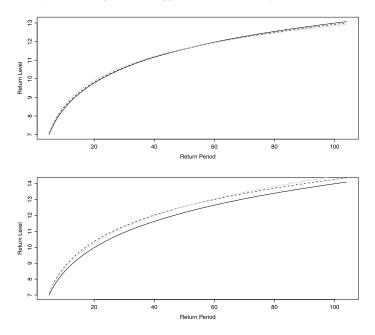
Figure 4 *Estimated returns (in log scale) for simulations with $n = 1000$. Left: $\xi = 0$. Right: $\xi = 0.1$. Solid line: true; grey line: MGPD; Dashed line: our approach.*

3.3 BMA × BMC: empirical results

The estimation approaches described in Section 2.3 are evaluated empirically in this section. Table 2 shows that the posterior probability for the correct regime is basically 1 for all situations considered, when ξ is away from 0. This means that inference via BMA or BMC will return basically the same values, since $\pi(M_l|\mathbf{y}) \approx 1$, for some l and BMA and BMC will coincide. Thus, a more detailed study is only required when the two approaches may differ, that is, when ξ is close or equal to 0.

Table 5 presents a comparison of the 2 approaches for the estimation of high quantiles. The closer the proportion of runs that includes the true value to the nominal credibility level, the better is the estimation procedure. The table shows a clear superiority of the BMA in all scenarios considered. In some situations, the BMC has a poor result, with true coverage well below the nominal level. The largest advantage is observed when $\xi = 0$. Note that differences between the two approaches get smaller as sample sizes increase, as expected. Based on these findings, it seems more reasonable to use BMA to report the results of the applications of the next section.

Table 5 *Summary of comparison BMA × BMC: proportion of simulations (based on 100 independent runs) where the true value of the quantile x_p was included in the 95% credibility interval*

	$\xi = -0.1$		$\xi = 0$		$\xi = 0.1$	
p	BMA	BMC	BMA	BMC	BMA	BMC
			$n = 1000$			
0.95	0.95	0.95	0.99	0.98	0.99	0.98
0.99	0.98	0.98	0.94	0.93	0.94	0.94
0.999	0.92	0.88	0.95	0.81	0.89	0.83
0.9999	0.92	0.86	0.95	0.71	0.89	0.82
Total	0.94	0.92	0.96	0.86	0.93	0.89
			$n = 2500$			
0.95	0.96	0.96	0.98	0.98	0.95	0.95
0.99	0.95	0.94	0.97	0.96	0.91	0.91
0.999	0.97	0.97	0.94	0.80	0.82	0.80
0.9999	0.97	0.99	0.96	0.67	0.67	0.66
Total	0.96	0.96	0.96	0.85	0.86	0.85

BMA—Bayesian Model Averaging, BMC—Bayesian Model Choice.

4 Applications

This section shows results of real data analyses of extreme data from Environmental Sciences. The same datasets used in the applications of Nascimento, Gamerman and Lopes (2012) were considered for comparative purposes. It consists of datasets where the shape parameter ξ was around 0 in continuous analysis using the MGPD model, and thus there was a reasonable degree of uncertainty about the tail behavior. One would expect that the proposed model would assign substantial probability to the Gumbel regime in these cases.

The first analysis is based on a dataset consisting of the measurement of the levels of flow of two rivers located in Northeast Puerto Rico: Fajardo and Espirito Santo. The data were recorded daily from April 1967 to September 2002, and are freely available from waterdata.usgs.gov. A total of 864 fortnightly maxima was analyzed. The second analysis is based on datasets consisting of the measurement of the amount of rain in two monitoring stations in Portugal: Barcelos in the North, and Grandola in the South. The data were recorded daily from 1931 to 2008, and are freely available from www.snirh.pt. The complete dataset has a large number of missing values, leading to a total of 918 monthly maxima data points for Barcelos station and 925 monthly maxima data points for Grandola station.

Figure 5 illustrates the estimation of ξ in the four applications considered. The posterior distribution of ξ is significantly positive (around 0.5) only in the Fajardo river application. In all other applications, more than half of the distribution is concentrated in the Gumbel regime ($\xi = 0$). More specifically, for the Espirito
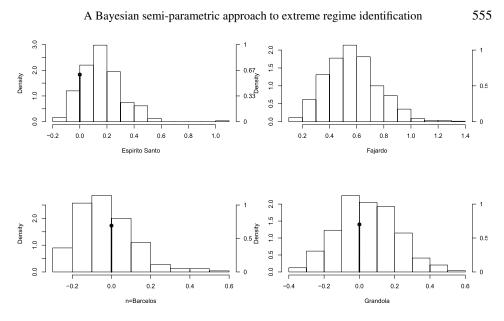
Figure 5 *Posterior distribution for ξ in the applications. The respective posterior mass probabilities at* 0 *have the scales shown on the right y-axis.*
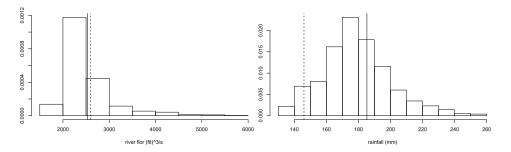

Figure 6 *Posterior histogram of high quantiles of applications. Left: Espirito Santo river, quantile 99.9%. Right: Barcelos station, quantile 99.99%. Vertical lines: full—posterior mean; dashed—maximum observed data.*

Santo river, $P(z_\xi^0 = 1|\mathbf{x}) = 0.61$, while for the Barcelos station $P(z_\xi^0 = 1|\mathbf{x}) = 0.69$ and for the Grandola station $P(z_\xi^0 = 1|\mathbf{x}) = 0.70$. One can conjecture that similar results, with large posterior probabilities associated with values of ξ around 0, would be obtained in some of the other environmental applications of Tancredi, Anderson and O'Hagan (2006), Castellanos and Cabras (2007), Macdonald et al. (2011) and Mahmoudi (2011), for example.

Figure 6 shows estimation of extreme quantiles for the two applications. For the Espirito Santo river, the 99.9% quantile is estimated close to the data maximum. Taking the posterior mean as an estimator of the quantile, it is expected that a river flow greater than or equal to 2531 ft^3/s occurs once every 40 years. This is close

to the estimated value of 2726 ft^3/s obtained for the MGPD model. The 99.99% quantile for Barcelos station goes beyond the observed maximum, well above the observations. This means that one day with rain levels greater than or equal to 180 mm will occur on average once every 820 years.

Table 6 shows the fit measures for the four data sets. In two applications, the proposed mixed model provides a lower BIC, compared with the continuous MGPD model, whereas lower DIC values are obtained for the continuous MGPD model. In the other two applications, the BIC and DIC values are lower for the MGPD model. The difference is relatively small in both cases, indicating that the estimation by both methods give similar results. This is particularly true for the central part of the data, modelled in the same way by the MGPD and by our proposed model, which contain large percentage of the observations, resulting in a much greater weight in the calculation of the fit measures than the tail.

Table 7 shows the posterior mean of extreme quantiles x_p for the applications, comparing the MGPD and the proposed model. The 0.95 and 0.99 quantiles of both models are close. There is greater difficulty in estimation of more extreme quantiles, because these quantiles are already extrapolating beyond the observed data. In these situations, the difference between the models increases. Based on simulation results, which showed that very high quantiles for the proposed model are more efficiently estimated by the proposed model than by the MGPD when $\xi = 0$, while the two models are equivalent around $\xi = 0$, we can safely make in-

Table 6 *Fit measures for the applications*

	Espirito Santo		Fajardo		Barcelos		Grandola	
Model	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC
Prop.	11,268	11,335	11,265	11,344	7640	7716	4318	4395
MGPD	11,255	11,329	11,264	11,380	7612	7709	4304	4397

Prop.—Proposed mixed model, MGPD—continuous MGPD model.

Table 7 *Extreme quantile x_p of applications*

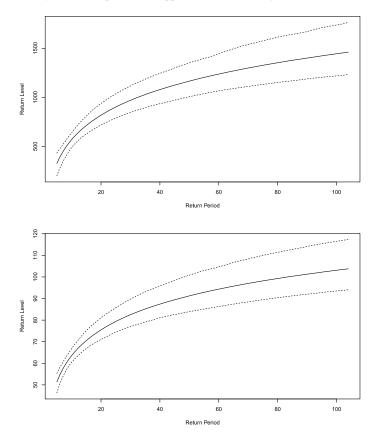|	Espirito Santo (ft^3/s)		Fajardo (ft^3/s)		Barcelos (mm)		Grandola (mm)	
p	MGPD	Prop.	MGPD	Prop.	MGPD	Prop.	MGPD	Prop.
0.95	813	818	858	857	77.5	75.5	53.9	51.1
0.99	1450	1444	1836	2002	105.4	103.1	75.2	71.7
0.999	2726	2531	4706	8181	139.9	142.9	94.8	103.2
0.9999	4734	4292	11,631	23,387	176.1	185.3	110.4	139.7

Prop.—Proposed model, MGPD—MGPD model.

Figure 7 *Expected return levels for the applications. Top*: *Espirito Santo river flow. Bottom*: *rainfall at Barcelos station. Solid line*: *posterior mean*; *dashed lines*: *95% credibility limits.*

ferences about rare events with the quantiles proposed in this paper. The quantiles in the table respectively represent the probability of an event greater than or equal to the estimated every 10 months, 4 years, 40 years and 400 years for applications in rivers and twice of these values respectively for the applications of the rainfalls. For example, a river flow greater than or equal to 857 ft^3/s in Fajardo river will occur on average once every 10 months, and the level of rainfall at Grandola station will be greater than or equal to 51.1 mm on average once every 20 months, based on the estimation of the quantile 0.95 using the proposed model.

Figure 7 presents estimated return levels for Espirito Santo river flow and rainfall at Barcelos station. Useful summaries of return levels can be obtained from this figure. These plots indicate that the expected Espirito Santo river flow is expected to surpass 1000 ft^3/s and 1500 ft^3/s every 37 and 100 fortnightly periods on average, respectively. The respective expected return levels of a 80 mm or 100 mm rainfall at Barcelos station are 23 or 80 months on average.

5 Conclusions

This paper presented a new approach to estimate extreme events, using the GPD distribution for the exceedances, where the distribution of the shape parameter ξ has a mixed nature, assigning probabilities to the 3 different extreme regimes. These probabilities are estimated using a Bayesian framework.

The only differences were observed in situations where the simulated value of ξ is close to 0 and the sample size is small, where there was greater uncertainty on the parameter, and where a high probability of $\xi = 0$ was obtained. These problems however occur for many of the other approaches and we conjecture they will occur for all approaches unless very informative prior distributions are used. Extreme quantile and fit measures showed the advantage of the proposed model, specifically with respect to the BIC, and when the quantiles are very high.

These findings emphasize the importance of the models proposed here where special attention is given to the $\xi = 0$ regime, usually discarded or overlooked in other formulations. This work can serve as a basis for all other models derived from the use of the GPD distribution and exceedances for the study of extreme value. These include regression models (Nascimento, Gamerman and Lopes, 2011) and time series models (Huerta and Sansó, 2007).

Appendix: MCMC algorithm

Sampling was made in blocks with Metropolis–Hastings proposals for each block due to unrecognizable form of the respective full conditionals. Each GPD parameter was sampled separately, the pair (μ, η) for each mixture component was sampled in a block and the weights **p** were sampled in a single block.

Details of the MCMC sampling scheme are given below. At iteration s, parameters are updated as follows:

Sampling Q_ξ: it can be seen from the posterior distribution in (2.3) that the full conditional distribution of Q_ξ is a Dirichlet distribution and it will be sampled in step $s + 1$ from

$$Q_\xi^{(s+1)} | \Theta \sim D_3\big(z_\xi^{+(s)} + \alpha_+, z_\xi^{0(s)} + \alpha_0, z_\xi^{-(s)} + \alpha_-\big).$$

Sampling ξ, Z_ξ: Although it is possible, obtain Z_ξ marginally using the prior distribution, given by

$$\big(q_\xi^+ f_G(\xi | a_\xi, b_\xi)\big)^{z_\xi^+} \big(q_\xi^0 I_{(\xi=0)}\big)^{z_\xi^0} \big(q_\xi^- f_U(\xi | -0.5, 0)\big)^{z_\xi^-},$$

the parameters ξ and Z_ξ must be sampled jointly because these parameters are highly correlated and it is also simpler to sample them jointly. The proposed kernel is $q(\xi, Z_\xi) = q(\xi | Z_\xi) q(Z_\xi)$. The proposal for $q(Z_\xi)$ is a multinomial $M_3(1, 1/3, 1/3, 1/3)$ and provides a value $Z_\xi^* = (z_\xi^{+*}, z_\xi^{0*}, z_\xi^{-*})$. If $z_\xi^{+*} = 1$, ξ^*

is obtained from a Gamma distribution $G(a_\xi, b_\xi)$. If $z_\xi^{0*} = 1$, then $\xi^* = 0$ with probability 1. If $z_\xi^{-*} = 1$, then ξ^* is generated from the Uniform $U(-\delta, 0)$, where $\delta = \sigma^{(s)}/(M - u^{(s)})$ and $M = \max(x_1, \ldots, x_n)$. So, $\xi^{(s+1)} = \xi^*$ and $Z_\xi^{(s+1)} = Z_\xi^*$ are jointly accepted with probability α_ξ, where

$$\alpha_\xi = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})q(\xi^{(s)}, Z_\xi^{(s)})}{\pi(\tilde{\Theta}|\mathbf{x})q(\xi^*, Z_\xi^*)}\right\},$$

where $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^*, Z_\xi^*, Q_\xi^{(s+1)})$ and $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^{(s)}, Z_\xi^{(s)}, Q_\xi^{(s+1)})$.

Sampling $\sigma, u, \mu, \eta, \mathbf{p}$: The algorithm to sample these parameters is similar to the algorithm for the MGPD model of Nascimento, Gamerman and Lopes (2012).

Acknowledgments

References

Behrens, C., Gamerman, D. and Lopes, H. F. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* **4**, 227–244. MR2062102

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. MR2298951

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics* **45**, 463–478.

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. MR2912374

Diebolt, J. and Robert, C. (1994). Estimation of finite mixture distributions by Bayesian sampling. *Journal of the Royal Statistical Society, Series B* **56**, 363–375. MR1281940

Doornik, J. A. (1996). *Ox: Object Oriented Matrix Programming, 4.1 Console Version*. London: Nuffield College, Oxford University.

Draper, S. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B* **57**, 45–97. MR1325378

Frigessi, A., Haug, O. and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**, 219–235. MR1995776

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* **96**, 194–209. MR1952732

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Baton Rouge: Chapman and Hall/CRC. MR2260716

Hastie, D. I. and Green, P. J. (2012). Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica* **66**, 309–338. MR2955422

Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* **14**, 285–299. MR2405331

Macdonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis* **55**, 2137–2157. MR2785120

Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation* **81**, 2414–2430. MR2811794

Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2011). Regression models for exceedance data via the full likelihood. *Environmental and Ecological Statistics* **18**, 495–512. MR2832907

Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Statistics and Computing* **22**, 661–675. MR2865043

Parmesan, C., Root, T. L. and Willing, M. R. (2000). Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society* **81**, 443–450.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131. MR0423667

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B* **59**, 731–792. MR1483213

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computation and Graphical Statistics* **18**, 349–367. MR2749836

Sang, H. and Gelfand, A. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**, 407–426. MR2749848

Scarrott, C. J. and Macdonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal* **10**, 33–60. MR2912370

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464. MR0468014

Smith, R. L. (1984). Threshold models for sample extremes. *Statistical Extremes and Applications* 621–638. MR0784846

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**, 583–639. MR1979380

Stephenson, A. and Tawn, J. (2004). Bayesian inference for extremes: Accounting for the three extremal types. *Extremes* **7**, 291–307. MR2212389

Tancredi, A., Anderson, C. and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**, 87–106. MR2329799

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology* **44**, 92–107. MR1770003

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer. MR1482232

Wiper, M., Rios Insua, D. and Ruggeri, F. (2001). Mixtures of Gamma distributions with applications. *Journal of Computational and Graphical Statistics* **10**, 440–454. MR1939034

F. Ferraz do Nascimento
Curso de Estatística
Universidade Federal do Piaui
Campus Ministro Petrônio Portela, CCN2
64049-550, Teresina
Brazil
E-mail: fernandofn@ufpi.edu.br

D. Gamerman
Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro
Caixa Postal 68530
21945-970, Rio de Janeiro
Brazil
E-mail: dani@im.ufrj.br

R. Davis
Department of Statistics
Columbia University
1255 Amsterdam Avenue, MC 4690
Room 1004 SSW
New York, New York 10027
USA