A Bayesian semi-parametric approach to extreme regime identification

Fernando Ferraz do Nascimento[1] Dani Gamerman[2]

1- Universidade Federal do Piaui 2 - Universidade Federal do Rio de Janeiro
E-mails:fernandofn@ufpi.edu.br dani@im.ufrj.br

Abstract. Limiting tail behavior of distributions are 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 limiting regimes that these data exceedances belong to. The model-based approach uses a mixture at the observational level where a Generalized Pareto distribution (GPD) is assumed above the threshold. The novelty of this work is the study of the behavior of the GPD through another mixture distribution over the three possible regimes. Adequate solution to this evaluation is shown to require a mixed prior distribution with a point mass, unlike previous work in the area. In some applications, estimation results showed one regime is clearly indicated by the data whereas in other applications there is no clear indication of the regime. This estimation is based on evaluation of posterior probabilities for each regime. Simulation exercises were conducted to evaluate the accuracy of the model in various parameter settings and sample sizes, specifically in the estimation of high quantiles. They show an improved performance over existing approaches. Results of environmental applications show that the point mass approach plays a vital role in this study.

Key-Words: Extreme value theory, GPD distribution, environmental data, MCMC, Bayesian Inference.

1 Introduction

Natural disasters have become an issue of increasing concern worldwide. In Brazil, rainfall during the summer months are very common, and in a few days, heavy rainfall frequently causes disorder in the population. More specifically, cities in the Rio de Janeiro state have suffered heavy losses with high rainfall over the last three years, leading to large-scale property destruction and hundreds of lifes lost. According to Parmesan et al. (2000), changes in extremes of temperature are more responsible for changes in the nature than change in mean temperature.

Given the importance of extreme values in the aforementioned situations, Extreme Value Theory (EVT) has proved vital for the prediction of these values. Through EVT, one can formulate a specific model to estimate outliers 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 certain threshold. The classical result in this area is the Fisher & Tippett (1928) theorem. It establishes the three possible distributions for maximum of blocks of observations. von Mises (1954) and Jenkinson (1955) unified these distributions in a single class called generalized extreme value (GEV).

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 d.f. $F(x|u) = P(X \leq u + x|X > u)$ is the d.f. of a generalized Pareto distribution (GPD), whose density is provided below. Loosely speaking, this result states that if u is large enough, the conditional distribution $F(x \mid u)$ can in general be approximated by a properly scaled GPD, as u tends to the endpoint of F. In addition to u, the GPD depends on a scale parameter σ and a shape parameter ξ. Let the parameter vector be denoted $\Psi = (\xi, \sigma, u)$. The density of the GPD can be written as

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

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 $\xi \geq 0$.

Smith (1984) proposed parameter estimation 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 $\xi < -1$. According to Coles & Tawn (1996), situations where $\xi < -0.5$ are extremely rare in environmental data. This finding is empirically supported by many studies, including our data analyses.

Extreme value theory is particularly useful for determination of higher quantiles, ie., q-values satisfying $P(X > q) = 1-p$ for large values of p. This task is particularly hard to achieve under other contexts due to scarcity of the data but EVT provide the theoretical support needed for this task. Thus, q can be found by inversion of the d.f. of the GPD equation $p = G(q \mid \xi, \sigma, u)$ for any given probability $p \in [0, 1]$. This gives

$$q = \begin{cases} u + \frac{((1-p)^{-\xi}-1)\sigma}{\xi}, & \text{if} \quad \xi \neq 0 \\ u - \sigma\log(1-p), & \text{if} \quad \xi = 0 \end{cases}. \tag{2}$$

These quantiles are important design parameters specially in extreme cases with p approaching 1 and we shall concentrate on their estimation. Nascimento et al. (2011a) used a GPD distribution to calculate lower quantiles $P(X < q) = 1 - p$ for the minimum temperatures of cities in the Rio de Janeiro state, in Brazil, reversing the curve of minimum density, and then returning the data in the original scale to calculate quantiles and threshold.

1.1 Nonparametric density estimation

Analysis of extreme values through the GPD distribution typically starts by setting a threshold and estimating only the tail of the observations. Recently, Cabras et al. (2011) proposed a model considering a regression structure for the GPD parameters. Mahmoudi (2011) handles a generalization of the GPD distribution, performing a classical estimation approach, proposing a Beta GPD (BGPD). Although in some situations, estimating the tail is the only interest to the researcher, one may be interested in answering other questions such as what is the distribution below the threshold. More importantly, the key question is the determination of the threshold itself.

Regarding the distribution below the threshold, various models have been proposed. Frigessi et al. (2002) used a Weibull distribution, Behrens et al. (2004) used a Gamma distribution. A more flexible approach is to consider a non-parametric approach to this density. Tancredi et al. (2006) uses a mixture of uniforms distributions. Diebolt et al. (2005) uses a continuous mixture of Gamma distributions, but considered a fixed threshold. Mcdonald et al. (2011) used a mixture of kernels below the threshold whereas Nascimento et al. (2012) proposed a finite mixture of Gamma distributions, based on results from Wiper et al. (2001). Nascimento et al. (2012) showed that finite mixture of Gammas belongs to the domain of attraction of the GEV distribution, indicating that the limit of the tail converges to the GPD density. Regarding threshold estimation, Bermudez et al. (2001) performs a Bayesian approach to the Peaks of Threshold (POT) method. Tancredi et al. (2006) considers the number of observations beyond the threshold is a parameter to be estimated. Behrens et al. (2004) and Nascimento et al. (2012) considers the threshold as a parameter to be estimated.

1.2 Study of the shape parameter of the GPD

Embrechts et al. (1997) show that if the shape parameter ξ is non-negative, the distribution has a heavy tailed behavior and the distributions in its domain of attraction have infinite support. If ξ is negative, the distribution is light tailed and the distributions in its domain of attraction have infinite support. When $\xi = 0$, the GPD has some special properties, and the distributions in its domain of attraction may have finite or infinite support. Situations where $\xi = 0$ are common in many applications and should be studied in detail.

Usual approaches for study of the shape 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 the evaluation of whether $\xi = 0$. Coles (2001), on pages 80 and 81, report numerical instabilities in the likelihood function when $\xi \approx 0$.

In this study, the possibility that the shape parameter of the GPD may assume the value 0 is explicitly considered, since in many of the applications cited above this seems to take place. Our approach allows for evaluation of the probabilities that $\xi = 0$, be positive or be negative, afforded by a mixed distribution for this parameter.

Figure 1 shows the GPD density in some situations, varying ξ. The density is higher where the parameter is negative in all configurations, because in this region the density is limited and their points are concentrated in a smaller range. When ξ is increasing, the density is more dispersed for larger values, exhibiting a heavier tail behavior. Moreover, it is clear that there is no discontinuity at 0, even using the exact density of the GPD for the case $\xi = 0$.

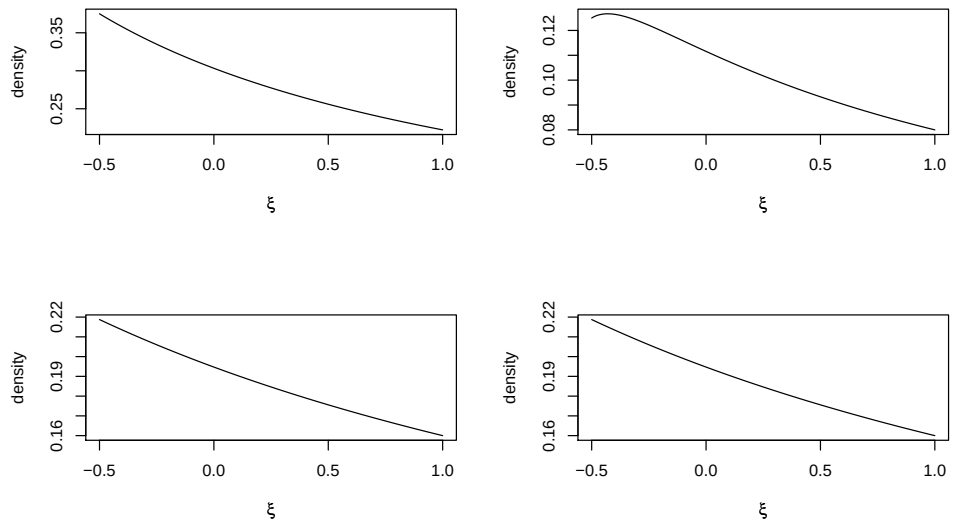

Figure 1: Probability density function of the GPD with $u = 6$, ξ varying in the following configurations: From the left to the right. First line: $y = 7$ and $\sigma = 2$ and $y = 9$ and $\sigma = 2$. Second line: $y = 7$ and $\sigma = 4$ and $y = 9$ and $\sigma = 4$

1.3 Outline of the paper

Section 2 shows the observational model, based on Nascimento et al. (2012), combining a finite mixture of Gammas below the threshold, and 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 simulations exercises from the proposed model, showing the efficiency of the estimation method and the accuracy in detecting extreme quantiles. The method is compared with standard approaches that consider absolutely continuous distribution of ξ. Section 4 shows two applications to extreme events: 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 equal 0. The section 5 summarizes the main conclusions of this work.

2 Model

An observational model that takes into account the complete data set is proposed, bearing in mind the need to estimate extreme events. The MGPD model of Nascimento et al. (2012) showed to be efficient in the non-parametric estimation before the tail, in the estimation of the threshold and in the estimation of the GPD parameters and is given by

$$f(x|\theta, \mathbf{p}, \Psi) = \begin{cases} h(x \mid \mu, \eta, \mathbf{p}), & \text{if} \quad x \leq u \\ [1 - H(u \mid \mu, \eta, \mathbf{p})]\, g(x|\Psi), & \text{if} \quad x > u \end{cases} \tag{3}$$

where h and H are respectively the density and cumulative functions of a finite mixture of Gamma distributions, and g is the GPD density with parameters $\Psi = (u, \sigma, \xi)$.

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 difference will be 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 under which the data is subject to. 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 HPD credibility interval of high level (say 0.95) includes 0 or by evaluating the probability of the set $\{\xi : \pi(\xi|\mathbf{x}) > 0\}$. 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 & Harrison (1997) in the context of state-space models.

The advantage of using a mixed distribution for ξ is that it provides the exact probabilities of Gumbel ($\xi = 0$), Frchet ($\xi > 0$) and Weibull ($\xi < 0$) tail behaviours. When ξ is positive for example, it is known that the data exhibit heavy tail behavior, which is a common feature in financial data. When ξ is negative, its distribution is bounded above. In the situations cited in the Section 1, the data provides evidence pointing at $\xi = 0$ but quantification of this event is difficult with a likelihood approach and even with a Bayesian approach based on a continuous prior for ξ. The next section introduces an approach to handle this situation.

2.1 Prior distribution

The parametric vector for density (3) is composed of three groups of parameters: a) the parameter $\theta = (\mu, \eta)$ of the k-mixture of Gamma densities $f_G(x \mid \mu_j, \eta_j)$ with means μ_j and shape parameters η_j, $j = 1, ..., k$, where $\mu = (\mu_1, ..., \mu_k)$ and $\eta = (\eta_1, ..., \eta_k)$; b) the mixture weights $\mathbf{p} = (p_1, ..., p_k)$; c) the GPD parameter $\Psi = (u, \sigma, \xi)$.

The prior distribution for θ is the same as Nascimento et al. (2012), based on work of Wiper et al. (2001). A important question about these parameters is their identification.

Diebolt & 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 $C(\mu) = \{\mu | 0 < \mu_1 < \mu_2 < ... < \mu_k\}$. Therefore, the prior for μ is taken in the form

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

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 $\eta \sim G(c_j/d_j, c_j)$, for some positive constants c_j and d_j, for $j = 1, \ldots, k$. The prior for weights is taken as $\mathbf{p} \sim D_k(\gamma_1, \ldots, \gamma_k)$, where $D_k(w_1, ..., w_k)$ represents the Dirichlet distribution with density proportional to $\prod_{i=1}^{k} p_i^{w_i}$.

The prior distribution for the threshold u is assumed to be $u \sim N(\mu_u, \sigma_u^2)$, as suggested by Behrens et al. (2004). This distribution is truncated at the lower data limit but usual choices of hyperparameters make the effect of this truncation negligible. Nascimento et al. (2012) shows that caution is necessary for setting hyperparameter values in some situations with small samples. The prior distribution can not be too vague because it will lead to an incorrect estimation of the threshold. This is controlled by appropriately setting the value of σ_u^2. Nascimento et al. (2012) describes how to specify this value even with scarce prior information. The prior distribution to scale parameter σ is given in the usual non-informative format for 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 & Cabras (2007) for (σ, ξ).

The main novelty of this work is the explicit evaluation of the probabilities associated with the limiting regimes. This is afforded by appropriately setting the prior distribution for the shape parameter ξ. Instead of considering a continuous prior distribution, the possibility of ξ taking the value 0 is assigned a positive probability. Thus, a mixed distribution is proposed for ξ, considering separately the probabilities associated with the three posible limiting tail behaviors: Gumbel ($\xi = 0$), Frchet ($\xi > 0$) 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

$$\begin{aligned} p(\xi, Z_\xi, Q_\xi) &= p(\xi \mid Z_\xi, Q_\xi) p(Z_\xi \mid Q_\xi) p(Q_\xi) \\ &= p(\xi \mid Z_\xi) p(Z_\xi \mid Q_\xi) p(Q_\xi), \end{aligned}$$

where conditional independence between ξ and Q_ξ given Z_ξ is assumed. The distribution

of $\xi \mid Z_\xi$ is assigned in the following specification

$$\xi \mid z_\xi^+ = 1 \sim Gamma(a_\xi, b_\xi),\ \xi \mid z_\xi^0 = 1 \sim \delta_{\xi=0} \text{ and } \xi \mid z_\xi^- = 1 \sim U(-0.5, 0), \tag{4}$$

where the parameters a_ξ, b_ξ may chosen so that the prior distribution of the positive part is vague, δ is the Dirac function. According to Coles & Tawn (1996), situations where $\xi < -0.5$ are extremely rare in environmental data and this is consistent with our experience. Thus, the range of negative values of ξ is limited to $[-0.5, 0]$, based on Smith (1984) but other limits may be chosen if required.

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

$$p(Z_\xi \mid Q_\xi) \propto (q_\xi^+)^{z_\xi^+}(q_\xi^0)^{z_\xi^0}(q_\xi^-)^{z_\xi^-},\ \text{for } (z_\xi^+, z_\xi^0, z_\xi^-) = (0,0,1),(0,1,0),(1,0,0),$$

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 little information on Q_ξ. A example could be $\alpha_+ = \alpha_0 = \alpha_- = 0.01$.

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. These features are explored in the next sections.

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 (3), with the prior distribution of the parametric vector described in section 2.1. The posterior density is given by

$$\pi(\theta, \mathbf{p}, \Psi, Z_\xi, Q_\xi | \mathbf{x}) \propto \prod_{i:x_i \leq u} \left(\sum_{j=1}^k p_j f_G(x_i | \mu_j, \eta_j) \right) \prod_{i:x_i \geq u} \left[1 - \sum_{j=1}^k p_j F_G(u | \mu_j, \eta_j) \right] g(x_i \mid u, \sigma, \xi)$$

$$\prod_{j=1}^k \left[p_j^{\gamma_j} \eta_j^{c_j - 1} \exp(-\frac{d_j}{\eta_j}) \mu_j^{a_j - 1} \exp(-\frac{b_j}{\mu_j}) \right] \frac{1}{2} \exp\left(-\frac{u - \mu_u}{\sigma_u} \right)^2 \frac{1}{\sigma} p(\xi \mid Z_\xi)(q_\xi^+)^{z_\xi^+ + \alpha_+}(q_\xi^0)^{z_\xi^0 + \alpha_0}(q_\xi^-)^{z_\xi^- + \alpha_-}, \tag{5}$$

where the first line above comes from the likelihood, the second line refers to the prior density of parameters, where $p(\xi \mid Z_\xi)$ is the density of (4), with a mixed distribution.

Inference cannot be performed analytically and approximating MCMC methods are used (Gamerman & Lopes, 2006). Convergence was assessed by running two parallel

chains with different starting values. 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. Details of the sampling algorithm are provided in the Appendix and are similar to the application of the algorithm seen in Nascimento et al. (2012). The modifications occur with respect to sampling of ξ, with mixed distribution, and with the additional parameters Z_ξ and Q_ξ. Proposal variances were tuned with a variation of the method proposed by Roberts & 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 3 different models characterized by the 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. Specially of interest in EVT is estimation of higher quantiles, extrapolated beyond the data points, that are merely functions of model parameters.

Consider a quantity of interest δ, eg a higher quantile, and assume the existence of the 3 models described above. The posterior probability of each model is given by

$$\pi(M_l \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid M_l)\pi(M_l)}{\sum_{j=1}^{3} \pi(\mathbf{y} \mid M_j)\pi(M_j)}, \text{ for } l = 1, 2, 3, \tag{6}$$

and the posterior distribution of δ is given by

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

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 $p(\delta \mid \mathbf{y}, M_s)$. 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 if 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 1,000, 2,500 and 10,000. Data before the tail was simulated a mixture of two Gamma distributions with $\mu = (2, 8)$, $\eta = (4, 8)$, and weight vector $\mathbf{p} = (1/3, 2/3)$. The threshold chosen was the 80% data quantile and the scale parameters is $\sigma = 2$. The shape parameter ξ was simulated with the following values $\xi = (-1, -0.4, -0.1, 0, 0.1, 0.4, 1)$. The aim of these configurations is to verify if the model is accurate in a variety of situations.

Prior distribution for mixture parameters were $\mu_j \sim IG(2.1, 5.5)$ and $\eta_j \sim G(6, 0.5)$, for $j = 1, \ldots, k$, and $\pi(\mathbf{p}) \sim D_k(1, \ldots, 1)$. These distribution have mean around the actual parameter value but with large variance to represent lack of information: the prior variance for μ_j's is 250 and the prior variance for η_j's is 24. Prior distribution for the GPD parameters was given by Jeffreys prior for σ, with $p(\sigma) \propto 1/\sigma$. The prior distribution described in (4) was assigned to ξ, with $a_\xi = b_\xi = 0.1$, and Q_ξ was assigned a Dirichlet prior with $\alpha_+ = \alpha_0 = \alpha_- = 0.1$. The threshold was assigned a normal prior distribution with mean given by the true value. The prior variance for the threshold was chosen in a way that the 95% credibility interval for the threshold ranges over the top half of the data values. This is only mildly informative, using information typically available and giving enough flexibility to avoid posterior degeneracy. Very large variances for the threshold could also be considered and cause no problem to the inference for large data sets. Specifically, a normal prior variance $\sigma_u^2 = 10$ was adopted when $n = 1,000$. A prior variance $\sigma_u^2 = 100$ was used when $n = 2,500$ and $n = 10,000$.

Inference was performed via the MCMC algorithm. 50,000 iterations to burn-in and the next 30,000 iterations were kept for inference for simulations with a sample size $n = 1,000$. 25,000 iteration to burn-in and 15,000 iterations were retained for inference were run for the simulations with $n = 2,500$, 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 $n = 1,000$ (or $n = 10,000$).

Figure 2 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 effective as in the model by Nascimento et al. (2012). 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 3 shows the estimation of ξ, when the true value is equal to 0.1, in three different configurations of sample size. The posterior distribution is 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 $n = 2,500$)

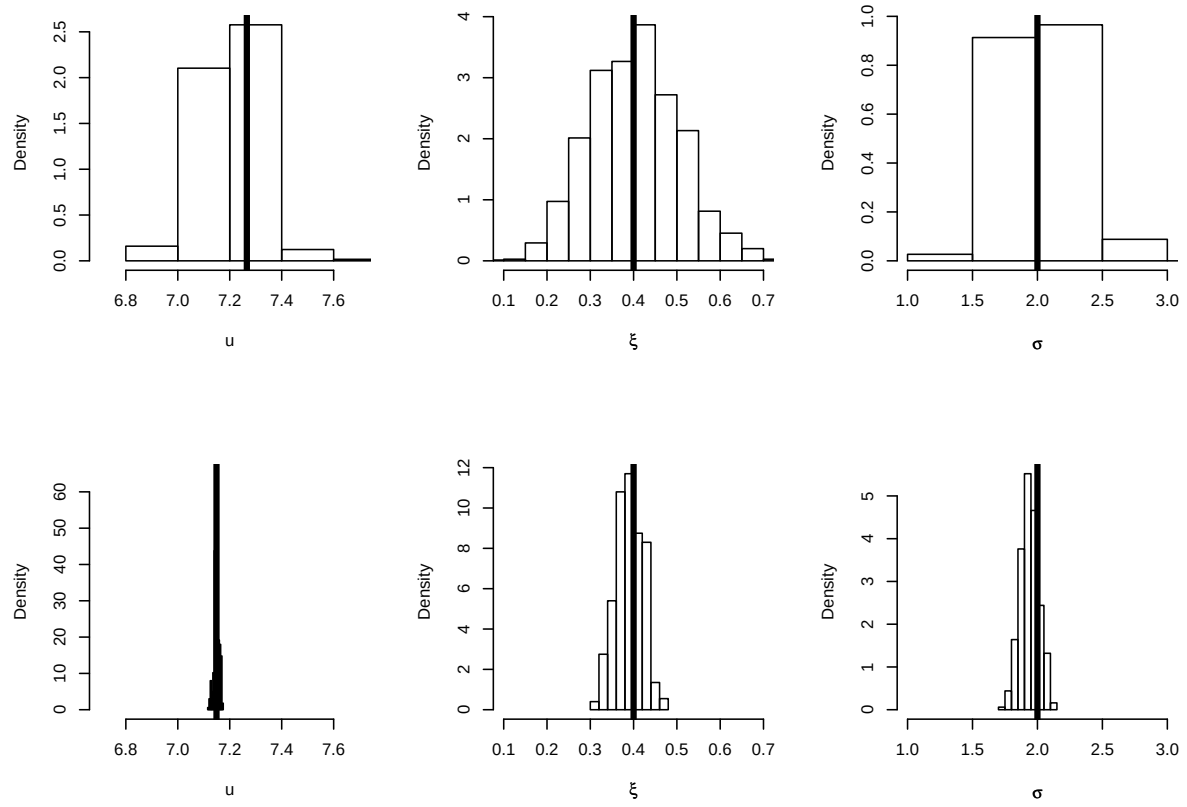

Figure 2: Posterior histogram for the GPD parameters with $\xi = 0.4$: top row - $n = 1,000$; bottom row - $n = 10,000$. Vertical lines: True value of parameter.

or rare (when $n = 10,000$) samples associated with the estimated value $\xi = 0$. Figure 3 also shows the estimation of ξ, when the true value is equal to 0, in three different configurations of sample size. Virtually 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 1 shows the posterior probability of parameter Z_ξ in all simulations. The tail behavior is correctly identified with posterior probability 1 for situations where ξ is away from 0 ($\xi = (-0.4, 0.4, 1)$), even for smaller sample sizes. The probability probability $P(z_\xi^0 = 1)$ 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 $n = 1,000$, the probability $P(z_\xi^0 = 1)$ of incorrect detection was high. This does not occur for samples with $n = 10,000$, which provide more accuracy in

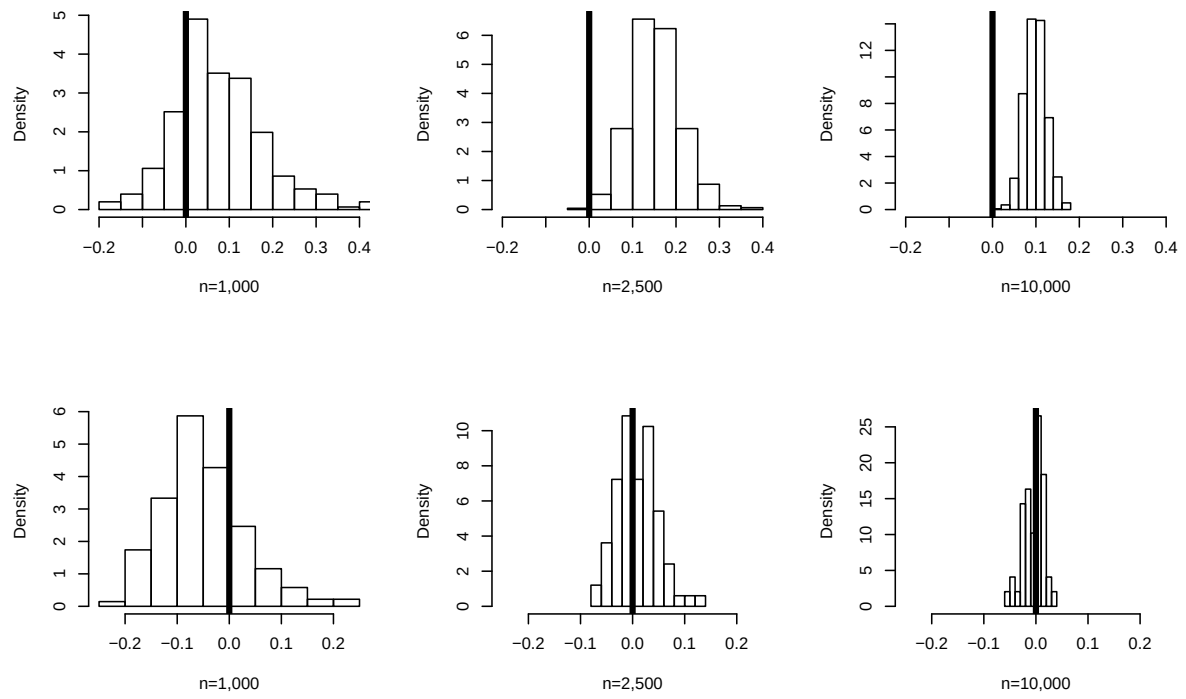

Figure 3: Posterior histogram for ξ in simulations: top row - $\xi = 0.1$; bottom row - $\xi = 0$. Vertical lines: probability lumps at $\xi = 0$.

detection of the sign of ξ, giving a probability of correct sign very close to 1.

Table 1: Posterior probability of Z_ξ in all simulations.

	n=1,000			n=2,500			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	0.47	0.51	0.02	0.17	0.81	0	0.003	0.997
0	0.05	0.78	0.17	0.04	0.92	0.04	0.025	0.951	0.024
0.1	0.16	0.80	0.04	0.92	0.08	0	0.996	0.004	0
0.4	1	0	0	1	0	0	1	0	0
1	1	0	0	1	0	0	1	0	0

$$Z^+ - P(z_\xi^+ = 1), Z^0 - P(z_\xi^0 = 1), Z^- - P(z_\xi^- = 1)$$

3.1 Fit measures and extreme quantiles

The proposed model is compared with the model of Nascimento et al. (2012), which showed good efficiency and gain estimation, when compared with other models, such as models of Wiper et al. (2001) and Behrens et al. (2004). The comparison was made through fit measures BIC (Schwarz, 1978) and DIC (Spiegelhalter et al., 2002). Another 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 4 illustrates the estimation of posterior distribution of extreme quantiles 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 these imprecise regime identification for smaller sample sizes do not seem to affect estimation of high quantiles.

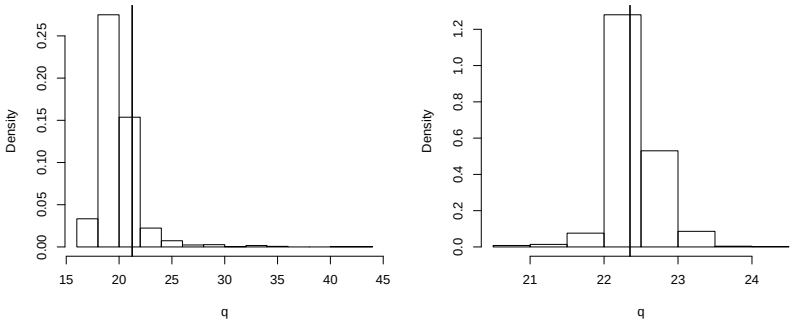

Figure 4: p-quantiles of simulations: left - $\xi = 0.1$, $n = 1,000$ and $p = 0.999$; right - $\xi = 0$, $n = 10,000$ e $p = 0.9999$. Vertical lines: true values of quantiles.

Table 2 shows the estimation of extreme quantiles in two different configurations ($\xi = 0$ and $\xi = 0.1$), that are frequently encountered in practice. They constitute the cases where the proposed model with a lump probability at $\xi = 0$ may contrast more from the MGPD model with continuous prior for ξ. The quantiles of the model proposed in this study were compared with empirical quantiles, and with the quantiles obtained from the MGPD model of Nascimento et al. (2012). The comparison favors the empirical quantiles for $n = 1,000$ with the MGPD providing better estimates for $\xi = 0.1$ and the proposed model 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.

Table 3 provides the fit measures BIC and DIC in the different configurations used. There seems to be a balance when comparing the two models. Fit results were very similar for larger sample sizes.

The results seem to indicate that little is lost in estimation efficiency when using the

Table 2: Extreme quantile of simulations

$n = 1,000$								
	$\xi = 0$				$\xi = 0.1$			
Prob	T	E	M	P	T	E	M	P
0.95	10.04	10.18	10.02	9.96	10.24	10.41	10.34	10.45
0.99	13.26	13.29	12.96	13.06	14.25	14.12	14.33	14.27
0.999	17.86	17.20	16.81	17.42	21.23	20.64	21.11	19.98
0.9999	22.47	N/A	20.37	21.76	30.03	N/A	29.72	26.18
$n = 10,000$								
	$\xi = 0$				$\xi = 0.1$			
0.95	9.92	9.95	9.98	9.93	10.12	10.01	10.04	10.02
0.99	13.14	13.19	13.19	13.16	14.13	13.89	13.92	13.9
0.999	17.74	17.29	17.61	17.78	21.12	21.01	20.68	20.66
0.9999	22.35	23.86	21.87	22.41	29.91	27.37	29.21	29.21

T-True, E-Empirical, M-MGPD, P-Proposed

Table 3: Measures of fit for the simulations

		n=1,000		n=2,500		n=10,000	
	Model	DIC	BIC	DIC	BIC	DIC	BIC
$\xi = -0.4$	Prop.	4163	4210	10421	10498	41663	41759
	MGPD	4184	4219	10422	10507	41661	41768
$\xi = -0.1$	Prop.	4300	4333	10738	10820	42825	42924
	MGPD	4278	4342	10740	10828	42824	42933
$\xi = 0$	Prop.	4319	4352	10736	10821	43403	43502
	MGPD	4298	4361	10734	10829	43402	43509
$\xi = 0.1$	Prop.	4376	4414	10948	11042	43601	43700
	MGPD	4362	4424	10948	11041	43602	43709
$\xi = 0.4$	Prop.	4517	4535	11269	11352	44764	44864
	MGPD	4487	4546	11273	11362	44768	44873

Prop. - Proposed Model, *MGPD* - MGPD Model

models proposed here. In addition, these models are able to identify and quantify well the possible regimes for tail behavior. So, they seem to provide a useful alternative for handling and identifying extreme values.

3.2 BMA $\times$ BMC: empirical results

The estimation approaches described in Section 2.3 are evaluated empirically in this

section. Estimation with BMC is performed with the regime with larger posterior probability amongst the possible 3 regimes.

Table 1 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 the same values. Thus, a more detailed study is only required when the 2 approaches may differ, ie when ξ is close or equal to 0.

Table 4 presents a comparison of the 2 approaches for the estimation of high quantiles. The table shows a clear superiority of the BMA in all scenarios considered. The largest advantage is observed when $\xi = 0.1$. Note that differences between the 2 approaches get smaller as sample size increases, 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 4: Summary of comparison BMA $\times$ BMC:
proportion of simulations (based on 30 independent runs) where the true value of the quantile was included in the 95% credibility interval

$n = 1,000$						
	$\xi = -0.1$		$\xi = 0$		$\xi = 0.1$	
Prob	BMA	BMC	BMA	BMC	BMA	BMC
0.95	0.97	0.93	1	1	1	0.96
0.99	0.90	0.90	0.90	0.87	1	1
0.999	0.90	0.80	0.93	0.77	0.93	0.90
0.9999	0.90	0.73	0.93	0.67	0.90	0.83
Total	0.92	0.84	0.94	0.83	0.96	0.93
$n = 2,500$						
	$\xi = -0.1$		$\xi = 0$		$\xi = 0.1$	
Prob	BMA	BMC	BMA	BMC	BMA	BMC
0.95	0.97	0.97	0.96	0.93	0.97	0.90
0.99	0.97	0.97	1	1	0.93	0.93
0.999	1	1	1	0.90	0.83	0.83
0.9999	1	1	1	0.70	0.70	0.70
Total	0.98	0.98	0.99	0.88	0.86	0.84

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 et al. (2012) were considered for comparative purposes. They consist on datasets where the shape parameter

ξ was around 0 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 dataset consisting on the measurement of the levels of flow of two rivers located in Northeast Puerto Rico: Fajardo and Espiritu Santo. The data was recorded daily from April 1967 to September 2002, and is freely available from *waterdata.usgs.gov*. A total of 864 fortnightly maxima was analyzed. The second analysis is based on datasets consisting on the measurement of the amount of rain in two monitoring stations in Portugal: Barcelos, in the North, and Grandola, in the South. The data was recorded daily from 1931 to 2008, and is 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. Only in the Fajardo river application the distribution of ξ is significantly positive, around 0.5. In all other applications, more than half of the distribution is concentrated in the Gumbel regime ($\xi = 0$). More specifically, for the Espiritu 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 would be obtained in some of the other environmental applications of Tancredi et al. (2006), Castellanos & Cabras (2007), Mcdonald et al. (2011) and Mahmoudi (2011), for example.

Figure 6 shows the estimation of extreme quantiles for two applications. For the Espiritu 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 $2,531\ ft^3/s$ occurs once every 40 years. This is close to the estimated value of $2,726\ ft^3/s$ obtained for the MGPD model. The 99.99% quantile for Barcelos station goes beyond the observed maximum, well above the observations. This means that rain levels greater than or equal to 180 mm will occur on average once every 820 years.

Table 5 shows the fit measures for the applications. In most applications, the proposed model provides a lower BIC, compared with the MGPD model, whereas the lower DIC are obtained for the MGPD model. The difference is very small in both cases, indicating that the estimation by both methods give similar results, particularly for the central part of the data, which has a very large percentage of the observations, resulting in a much greater weight in the calculation of the fit measures than the tail.

Table 6 shows the posterior mean of extreme quantiles 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 efficient than the MGPD, we can safely make inferences about rare events with the quantiles proposed in this paper. The

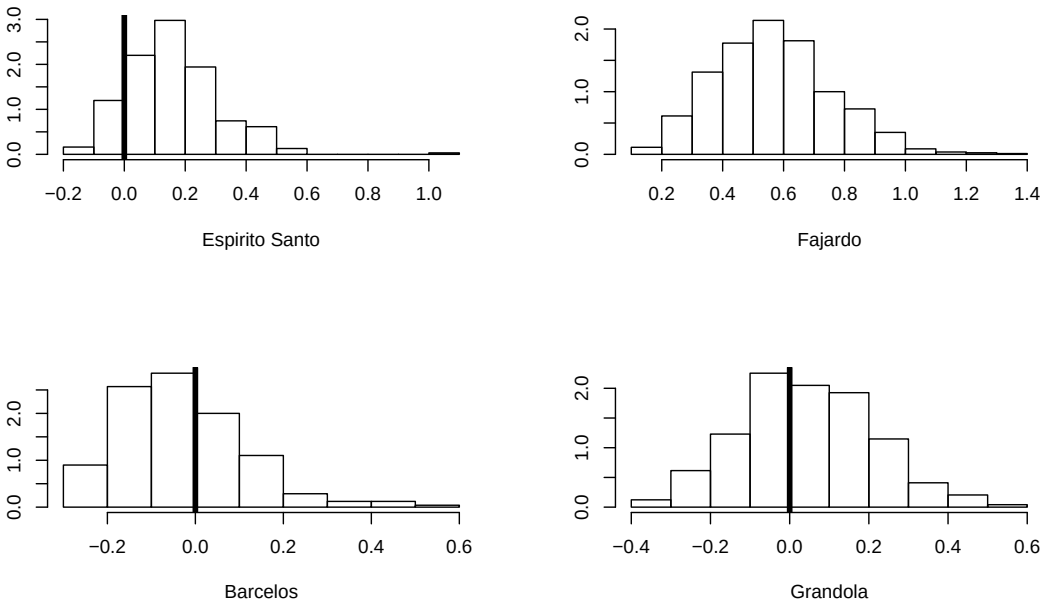

Figure 5: Posterior distribution for ξ in the applications

Table 5: Fit measures for the applications

	Espirito Santo		Fajardo		Barcelos		Grandola	
Model	DIC	BIC	DIC	BIC	DIC	BIC	DIC	BIC
Prop.	11264	11335	11264	11339	7641	7709	4319	4389
MGPD	11255	11329	11264	11380	7612	7709	4304	4397

Prop. - Proposed Model, $MGPD$ - MGPD Model

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.

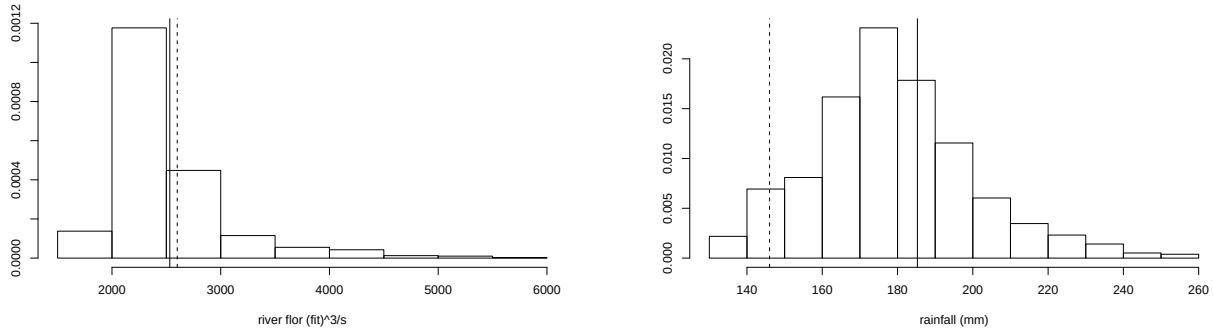

Figure 6: Posterior histogram of high quantiles of applications, under $MGPD_3$. Left: Espirito Santo river, quantile 99.9%. Right: Barcelos station, quantile 99.99%. Vertical lines: full - posterior mean; dashed - maximum observed data.

Table 6: Extreme quantile of aplications

	Espirito Santo		Fajardo		Barcelos		Grandola	
	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	11631	23387	176.1	185.3	110.4	139.7

Prop. - Proposed Model, $MGPD$ - MGPD Model

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 based on the data.

It was observed through the simulations that the model is effective in estimating the central part of the distribution, the threshold and the GPD parameters. It was also possible to obtain good estimates for the tail behaviors. The only restrictions 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 large probability of $\xi = 0$ were 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 proposed model,

specifically with respect to the BIC, and when the quantiles are very high.

Applications results showed datasets that exhibit high probability of the Gumbel regime ($\xi = 0$). These findings emphasize the importance of the models proposed here where special attention is given to this 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 et al., 2011a), time series models (Nascimento et al., 2011b) and spatial models.

Acknowledgments

The research that led to this paper was motivated by discussions with Prof. Richard Davis, whose contribution is gratefully acknowledged. The research of the second author was supported by grants from CNPq and FAPERJ, from Brazil.

References

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

Bermudez, P., Turkman, M. A. & Turkman, K. F. (2001). A predictive approach to tail probability estimation. *Extremes* 4 295–314.

Cabras, S., Castellanos, M. A. & Gamerman, D. (2011). A default bayesian approach for regression on extremes. *Statistical Modelling* 11 557–580.

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

Coles, S. G. (2001). *Extreme Value Theory an Applications*. Kluver Academic Publishers.

Coles, S. G. & Tawn, J. A. (1996). A bayesian analysis of extreme rainfall data. *Applied Statistics* 45 463–78.

Diebolt, J., El-Aroui, M., Garrido, M. & Girard, S. (2005). Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes* 1 57–78.

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

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.

Embrechts, P., Küppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* New York: Springer.

Fisher, R. A. & 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.

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

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.

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

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.

Mahmoudi, E. (2011). The beta generalized pareto distribution with application to lifetime data. *Mathematics and computers in simulation* 81 2414–2430.

Mcdonald, A., Scarott, C. J., Lee, D., Darlow, B., Reale, M. & Russell, G. (2011). A exible extreme value mixture model. *Computational Statistics and Data Analysis* 55 2137–2157.

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

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

Nascimento, F. F., Lopes, H. F. & Gamerman, D. (2011b). Generalized pareto models with time-varying tail behavior. *Technical Report* LES UFRJ, In preparation.

Parmesan, C., Root, T. L. & 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.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6 461–4.

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

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

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

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

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

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer - Second Edition.

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

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 $\mathbf{p}$ 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 (5) that the full conditional distribution of Q_ξ is a Dirichlet distribution and it will be sampled in step $s+1$ from

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

Sampling ξ, Z_ξ: 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 $q(\xi, Z_\xi) = q(\xi \mid Z_\xi)q(Z_\xi)$. 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 $\delta = \sigma^{(s)}/(M - u^{(s)})$ e $M = \max(x_1, \ldots, x_n)$. 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 et al. (2012).