# A semiparametric Bayesian approach to extreme value estimation

Fernando F. Nascimento[a,*], Dani Gamerman[a], Hedibert Freitas Lopes[b,*]

[a]*Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
[b]*The University of Chicago Booth School of Business, Chicago, USA*

## Abstract

This paper is concerned with extreme value density estimation. The generalized Pareto distribution (GPD) beyond a given threshold is combined with a nonparametric estimation approach above the threshold. This semiparametric setup is shown to generalize a few existing approaches and enables density estimation over the complete sample space. Estimation is performed via the Bayesian paradigm, which helps identify model components. Estimation of all model parameters, including the threshold and higher quantiles, and prediction for future observations are provided. Simulation studies suggests a few useful guidelines to evaluate the relevance of the proposed procedures. Models are then applied to environmental data sets. The paper is concluded with a few directions for future work.

*Key words:*

Bayesian, GPD, Higher Quantiles, MCMC, Mixture of Distributions, Nonparametric Estimation.

*Corresponding author: Hedibert Freitas Lopes. The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL, 60637, USA. Phone: +1 (773) 834-5458. Fax: +1 (773) 702-0458.

*Email addresses:* `fernandofn@dme.ufrj.br` (Fernando F. Nascimento), `dani@im.ufrj.br` (Dani Gamerman), `hlopes@chicagogsb.edu` (Hedibert Freitas Lopes)

# 1. Introduction

Extreme value theory was shown to provide a very useful tool in many areas of application where precise knowledge of the tail behavior of a distribution is of central interest. The areas where most impact was achieved are environmental science and finance. Problems associated with large amounts of rain have always plagued the society due to the social and economic loss they potentially inflict. Understanding their pattern of occurrence specially for higher values allows for prevention and/or mitigation of their potentially harmful effects. Similar reasoning applies to other environmental variables such as wind speed, sea tides and river flows. Applications to finance are even more obvious since they are directly related to money. Risk management involves dealing with potential loss both in actuarial applications and in stock market trading (Embrechts et al, 1997).

A fundamental result to this end was proved by Pickands (1975). He showed that the limiting distribution of exceedances over suitably large thresholds behaves in a very stable fashion, converging to a the generalized Pareto distribution (GPD). The result does not provide any information below the threshold.

There are many possibilities for handling both parts (below and above the threshold) and for combining them. Mixture of distributions and extreme value theory will be used as building blocks in our modeling strategy. Therefore, the main ideas behind these approaches are briefly introduced next.

## 1.1. Mixture of distributions

Mixture models provide an interesting illustration of the development of more complex models that was helped by the advance in computationally complex methods. Finite mixture of normal distributions are used in nonparametric density estimation by Titterington et al. (1985), Diebolt and Robert (1994), Roeder and Wasserman (1997) and Richardson and Green (1997), to name a few.

In many applications, data is restricted to positive values. In these cases, a more appropriate basis for building a mixture model is the Gamma family of distributions.

Based on theoretical results by De Vore e Lorenz (1993) and Asmussen (1987), Wiper et al. (2001) used mixtures of Gamma densities to approximate any density defined over $[0, \infty)$. See also Dalal and Hall (1983) and Dey et al. (1995) for related work.

The mixture model used in this paper is denoted by $MG_k$ with distribution function (d.f.) $H$ and density $h$ defined as

$$h(x \mid \theta, p) = \sum_{j=1}^{k} p_j f_G(x \mid \mu_j, \eta_j) \tag{1}$$

where $\theta = (\mu, \eta)$ denote the Gamma parameters $\mu = (\mu_1, \ldots, \mu_k)$ and $\eta = (\eta_1, \ldots, \eta_k)$, $p = (p_1, \ldots, p_k)$ denotes the mixture weights and $f_G$ is the density of the Gamma distribution given by

$$f_G(x|\mu, \eta) = \frac{(\eta/\mu)^\eta}{\Gamma(\eta)} x^{\eta-1} \exp(-(\eta/\mu)x), \text{ for } x > 0. \tag{2}$$

The $\mu_j$´s and $\eta_j$´s can take any positive value and the $p_j$s must be non-negative and add up to 1. Note that the parametrization is in terms of the mean $\mu$ and the shape parameter $\eta$. This choice will simplify model specifications in the sequel.

The papers above provide theoretical and empirical evidence that mixture of Gamma densities can be used for density estimation. They will cover adequately the data span but are not designed for handling extrapolation towards the tail of the distribution where little or no data is available. This shortcoming will be illustrated later on in this paper. Extreme value theory provide a precise description of the tail, designed to overcome these difficulties.

*1.2. Extreme Value Theory*

Extreme value theory is designed to describe atypical situations that may have a substantial impact in the phenomenon under study. The classical result in this area is the Fisher-Tippet (1928) theorem. It establishes the three possible distributions for maxima of blocks of observations. von Mises (1954) and Jenkinson (1955) unified these distributions is a single class called generalized extreme value (GEV).

3

Pickands (1975) proved that if $X$ is a random variable whose distribution function $F$ is in the domain of attraction of a GEV distribution, then as $u \to \infty$, the conditional d.f. $F(x|u) = P(X \leq u + x | X > u)$ is that of a GPD. In addition to $u$, the GPD depends on a scale parameter $\sigma$ and a shape parameter $\xi$. Let the parameter vector be denoted $\Psi = (\xi, \sigma, u)$ and $I$ denote the indicator of $\xi < 0$, ie, it is 1 if $\xi < 0$ is true and 0, if $\xi \geq 0$. The density of the GPD can be written as

$$g(x|\Psi) = \begin{cases} \frac{1}{\sigma}\left(1 + \xi\frac{(x-u)}{\sigma}\right)^{-(1+\xi)/\xi}, & \text{for } 0 \leq x - u \leq (-1)^I \left(\frac{\sigma}{\xi}\right)^{1/I} & (\text{when } \xi \neq 0) \\ \frac{1}{\sigma}\exp\{-(x-u)/\sigma\}, & \text{for } 0 \leq x - u & (\text{when } \xi = 0) \end{cases}, \quad (3)$$

and $g(x|\Psi) = 0$, otherwise. 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 $\xi \geq 0$. Hereafter, description of GPD will concentrated over its range $C_\xi$ of possible values to simplify notation.

If, for example, $X$ is maxima over a given set of observations, then it belongs to the domain of attraction of the GEV distribution. Hence, based on this theorem, its larger values (beyond a suitably large threshold $u$) can be approximated by a GPD. Thus, extrapolation around and beyond the largest observed order statistics can be made. This is in sharp contrast with non-parametric approaches such as mixtures of distribution where no sound basis for inference in these tail regions exist as they rely on the data information which is scarce or absent.

Smith (1985) proposed parameter estimation via maximum likelihood. He showed that they do not obey the regularity conditions if $\xi \in (-1, -0.5)$, and do not exist if $\xi < -1$. According to Coles and Tawn (1996), situations where $\xi < -0.5$ are extremely rare in environmental data. It is also worth noting that the scale parameter and the threshold are related. If the threshold is changed to $u' > u$, then the new exceedances are also described by a GPD with same shape parameter $\xi$ and scale parameter $\sigma' = \sigma + \xi(u' - u)$. The next section describes how this lack of identification may be partially resolved through appropriate model specifications.

Extreme value theory is also concerned with determination of higher quantiles, ie., $q$-values satisfying $P(X > q) = p$ for small values of $p$. The theory above allows also

4

estimation of these higher quantiles beyond the threshold as they are simply functions of the GPD parameters. The d.f. of the GPD is given by

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

for $x \in C_\xi$. Thus, $q$ can be found by inversion of the equation $p = G(q \mid \xi, \sigma, u)$ for any given probability $p \in [0, 1]$. This gives

$$q = \frac{((1-p)^{-\xi} - 1)\sigma}{\xi}.$$

These quantiles are important design parameters specially in extreme cases with $p$ approaching 0 and we shall concentrate on their estimation. They will also illustrate the advantage of incorporation of the GPD into the model.

### 1.3. Related work

Different approaches have been proposed in the literature recently under the Bayesian paradigm. Bermudez et al. (2001) suggest an approach that only considers the probability cumulated up to the threshold and estimates it based on the data frequency. Frigessi et al. (2002) considers mixture of two distributions: GPD and Weibull, with data dependent weights. Tancredi et al. (2003) uses a mixture of uniform densities for the central part of the data and the number of observations beyond the threshold is a parameter to be estimated. Behrens et al. (2004) uses a Gamma distribution below the threshold and a GPD above it. The threshold is a parameter to be estimated. This paper builds on their work by applying a mixture of Gamma densities to the central part. It will be shown that this generalization provides gains in flexibility and adequacy.

### 1.4. Outline of the paper

Section 2 will present our model for extreme data analysis, based on a combination of mixture of Gamma below the threshold and GPD for the tail. Inference procedure is carried out via MCMC and some computational details are also provided there. Section

3 illustrates the method with a few simulated examples. This exercise will also provide a useful benchmark for the model comparisons carried out. Section 4 presents two applications to extreme data analysis: river flow levels in Puerto Rico and rainfall in Portugal. The results are compared with those obtained with other approaches. Section 5 draws some concluding remarks and points at possible extensions.

## 2. Model

In view of the introductory discussion, it seems reasonable to contemplate a model that incorporates the flexibility of the mixture of Gamma for when there is data to estimate it and uses the GPD for the tail of the distribution. The result of Pickands (1975) suggests one would expect to get better results, specially in the tail, by using (3) instead of a simple mixture of Gamma when data on maxima is involved.

Let $h$ be the density of a $MG_k$ as in (1) and $g$ be the density of the GPD, as in (3). The density of our proposed model, denoted by $MGPD_k$, is given by

$$
f(x|\theta, p, \Psi) = \begin{cases} h(x \mid \mu, \eta, p), & \text{if } x \leq u \\ [1 - H(u \mid \mu, \eta, p)] \, g(x|\Psi), & \text{if } x > u \end{cases}
\tag{4}
$$

where $H$ is the d.f. of the mixture of Gammas, already presented in Section 1. Typical forms for this density are shown in Figure 1. Notice that model specification allows for a discontinuity of the density at the threshold. Continuity constraints could be imposed but this is an unnecessary condition. The nonparametric nature of the central part of the density allows for appropriate adaptation. As a result, the density shows no noticeable break for typical real data applications and all relevant calculations can be made without any theoretical or applied difficulties.

The advantage of this model formulation is flexibility. A non-parametric approach with only mild continuity assumptions is considered for the center of the distribution without imposing any specific parametric form or constraints such as unimodality. A parametric approach can be safely assumed for the tail due to its theoretical backing. The combination of these two blocks gives rise to the semi-parametric nature of our approach.

The flexibility is also present in the choice of the threshold, performed through parametric estimation. This allows for the division of the sample space into two data regimes: the central part and the tail. This task is performed automatically, incorporating uncertainty about all model components, and is governed by the data.

Figure 1

Recall from Section 1.2 that the GPD parameters are not uniquely identified. The model (4) introduces an important element to help identification. Provided enough data is observed, The Gamma mixture for the central part of the distribution can be separated from the GPD and this change can be picked up from the likelihood. As a result, clear identification of the threshold is obtained thus leading to correct identification of the other GPD parameters. This is not an easy task and, as will be shown, requires a substantial amount of data information. When there is no such information, mild probabilistic constraints in the form of a prior distribution provide a suitable complement.

Once again, it is important to obtain higher quantiles of this distribution. This is another advantage of this class of distributions over the Gamma mixture $MG_k$. The $p$-quantile $q$, satisfying $P(X < q) = p$, is obtained for this latter class of models after solving

$$p = H(q \mid \mu, \eta, p) = \sum_{j=1}^{k} p_j \int_0^q f_G(x \mid \mu_j, \eta_j) dx. \tag{5}$$

There is no explicit solution for (5) analytically. These quantiles must be computed numerically. In this work, this was achieved by an exhaustive search over a range of values.

Another advantage of the $MGPD_k$ class is the case with which higher quantiles can be obtained. For values beyond the threshold, the d.f. of the $MGPD_k$ model is given by

$$F(x \mid \theta, p, \Psi) = H(u \mid \mu, \eta, p) + [1 - H(u \mid \mu, \eta, p)] G(x \mid \Psi).$$

Therefore, it is straightforward to obtain the $p$-quantile as

$$q = \frac{((1 - p^*)^{-\xi} - 1)\sigma}{\xi}, \tag{6}$$

7

where

$$p^* = \frac{p - H(u \mid \mu, \eta, p)}{1 - H(u \mid \mu, \eta, p)}.$$

Note that this quantile is a highly nonlinear funcion of the model parameters. Therefore, its posterior distribution can only be obtained via approximating techniques. Once this distribution is (approximately) obtained for any given probability $p$, it can provide useful information about the extreme behavior of the data both in terms of point estimates through posterior means or medians and in terms of uncertainty through their credibility intervals.

### 2.1. Prior distribution

A relevant aspect of mixture models is the inherent lack of identifiability. Therefore, some restriction must be imposed to allow identification of the model parameters. Diebolt and Robert (1994) and Frühwirth-Schnatter (2001) among others impose order restrictions over the means in Gaussian mixtures. This procedure is also applied here as in Wiper et al. (2001) and the parameter space is hereafter restricted to $C(\mu) = \{\mu | 0 < \mu_1 < \mu_2 < \cdots < \mu_k\}$. Therefore, the prior for $\mu$ is taken in the form

$$p(\mu_1, \ldots, \mu_k) = K \prod_{i=1}^{k} f_{IG}(\mu_i \mid a_i/b_i, b_i) I(\mu_1 < \mu_2 < \ldots < \mu_k),$$

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 $\eta$ 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 $p \sim D_k(\gamma_1, \ldots, \gamma_k)$, where $D_k(w_1, \ldots, w_k)$ represents the Dirichlet distribution with density proportional to $\prod_{i=1}^{k} p_i^{w_i}$.

There are many possibilities available for the GPD parameters $\sigma$ and $\xi$. Coles and Tawn (1996) define Gamma prior distributions for quantiles based on expert opinion. They then used the relation between the quantiles and the parameters to induce their

priors. Castellanos and Cabras (2007) obtained the non-informative prior for $(\sigma, \xi)$ as

$$\pi(\sigma, \xi) \propto \sigma^{-1}(1 + \xi)^{-1}(1 + 2\xi)^{-1/2}, \quad \xi > -0.5, \sigma > 0. \tag{7}$$

They also showed that this prior leads to proper posterior distributions.

The prior distribution for the threshold was taken as a normal distribution $N(\mu_u, \sigma_u^2)$, as suggested by Behrens et al. (2004). Care must be exercised when specifying these models hyperparameters. The mean $\mu_u$ may have a strong influence over the resulting inference. Thus, the recommendation is that it should be placed around suitably large order statistics of the sample. Also, this prior should not be too concentrated unless there is substantial prior knowledge about this parameter. This is rarely the case since this is an artificial parameter, governing only when approximation of the tail by the GPD can be safely assumed. It should not be made entirely vague either because of the identification issue referred to in Section 1.2. Once again, it seems reasonable to have the threshold concentrated around the upper end of the sample. In doing that, it also rules out the possibility of a negative value for the threshold for all practical purposes. An alternative is to truncate the prior from below to avoid negative values but this was not needed in our applications.

### 2.2. Posterior and predictive distributions

Assuming the presence of a sample $x = (x_1, \ldots, x_n)$ from (4), the posterior density is obtained in the log scale as

$$
\begin{aligned}
\log \pi(\theta, p, \Psi | x) \;=\; & Z + \sum_{i:x_i \geq u} \log \left( \sum_{j=1}^{k} p_j f_G(x_i | \mu_j, \eta_j) \right) + \sum_{i:x_i \geq u} \log \left[ 1 - \sum_{j=1}^{k} p_j F_G(u | \mu_j, \eta_j) \right] \\
& - \sum_{i:x_i \geq u} \left[ \log(\sigma) - \frac{1 + \xi}{\xi} \log \left( 1 + \frac{\xi(x_i - u)}{\sigma} \right) \right] \\
& + \sum_{j=1}^{k} \left[ (c_j - 1) \log(\eta) - d_j \eta_j - (a_j + 1) \log(\mu) - b_j / \mu_j \right] \\
& - \frac{1}{2} \left( \frac{u - \mu_u}{\sigma_u} \right)^2 - \log(\sigma) - \log(1 + \xi) - (1/2) \log(1 + 2\xi), \tag{8}
\end{aligned}
$$

9

where Z is a normalizing constant, the first two lines above come from the likelihood and the remaining ones come from the prior.

Inference cannot be performed analytically and approximating MCMC methods are used. Parameters were separated into blocks and each block was updated according to a Metropolis rule, since none has a full conditional density in recognizable form. Unlike previous work on mixtures (eg. Diebolt and Robert (1994) and Wiper et al. (2001)), the introduction of latent variables indicating the mixture components does not lead to full conditional distributions that can be easily sampled from. This difficulty is mainly due to the combination of the mixture with another distribution for the tail. Details of the sampling algorithm are provided in the Appendix. Proposal variances were tuned with a variation of the method proposed by Roberts and Rosenthal (2006), with variances smaller than their target value since our blocks are multidimensional.

Prediction for a new observation is as important as parameter and quantile estimation. Different combinations of parameters may lead to the same, undistinguishable predictions. These evaluations are better performed through the predictive density. For any given data set $x$, the density for a new observation $x_{n+1}$ is given by

$$
\begin{aligned}
f(x_{n+1}|x) &= \int f(x_{n+1}, \theta, p, \Psi|x) d\theta dp d\Psi \\
&= \int f(x_{n+1}|\theta, p, \Psi) p(\theta, p, \Psi|x) d\theta dp d\Psi \\
&= E_{(\theta, p, \Psi)|x} \left( f(x_{n+1}|\theta, p, \Psi) \right).
\end{aligned}
$$

This integration can not be performed analytically and Monte Carlo approximation may be used instead. This gives $\hat{f}(x_{n+1}|x) = \frac{1}{I} \sum_{i=1}^{I} f(x_{n+1}|\theta^{(i)}, p^{(i)}, \Psi^{(i)})$, where $(\theta^{(i)}, p^{(i)}, \Psi^{(i)})$ is a value sampled from $\pi(\theta, p, \Psi|x)$, for $i = 1, \cdots, I$. For model $MGPD_k$, $f(x_{n+1} | \theta^{(i)}, p^{(i)}, \Psi^{(i)})$ is given by (4). The expression of the sampling density simplifies to (1) for model $MG_k$.

## 3. Simulations

Simulations based on samples from the proposed model were made with a number of purposes. Initially they can provide empirical evidence of model identifiability. But it also allows for appropriate validation of model selection criteria used. This latter exercise will enable the use of these criteria for the analysis of real data in the next Section. The rationale behind this idea is that if the criteria appropriately selects the correct model when we know which one it is, then it should behave well in practical situations, when we do not know which, if any, is the best model.

It could be argued that mixture models $MG_k$ are capable of handling any positive data by selecting an appropriately large number of components in the mixture. If that were true, models $MPGD_k$ would be an unnecessary complication. The exercises in this section will show that in practice clear gains are obtained with models $MPGD_k$ when extreme data is present.

The simulation exercise was performed with samples of sizes 1,000 and 10,000. The exercises are based on mixture of two Gamma distributions with $\mu = (2,8)$ and $\eta = (4,8)$. Different values were used for the GPD tail with $\sigma = 2, 3$ and 5, $\xi = -0.4$ and 0.4 and threshold at the 90% and 99% data quantiles. When the threshold is set at a very high value (99% quantile), reliable estimation can only be performed for larger data sets ($n = 10,000$). For this threshold, only the results with larger data sets are reported.

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 $\pi(p) \sim D_k(1, \ldots, 1)$. These distribution have mean around the actual parameter value but with large variance to represent lack of information. Prior distribution for the GPD parameters was given by Jeffreys prior for $(\xi, \sigma)$ and for the threshold was used a normal distribution with mean given by actual value and a suitably large variance. This is reasonably vague but does provide some information. The prior variance for the threshold was chosen in a way that the 95% credibility intervals for the threshold range a priori from around the 50% to the 99% data quantiles. Thus, they are only mildly informative, giving enough flexibility for influence of the likelihood. Very

large variances for the threshold could also be considered and cause no problem to the inference for large data sets. Problems associated with such vague prior distributions for small to moderate sample sizes are illustrated in the next Section.

Figures 2, 3 and 4 show the predictive densities in three of the simulation exercises. They clearly show the inadequacy of model $MGPD_1$ with predictions far from the true density. Results from models $MGPD_2$ and $MGPD_3$ are undistinguishable and very close to the true density, specially when $n = 10,000$ (see Figures 3 and 4). The models $MG_k$ involving only mixtures of Gamma provide a good fit in the central part of the distribution but do not perform so well in the tail, as expected. This is illustrated in Figures 2 and specially 3 and 4 for the best fit model in this class according to deviance information criterion (DIC) (Spiegelhalter et al., 2002).

Figure 2

Figure 3

Figure 4

Figures 5, 6 and 7 show posterior histograms of the tail parameters for the above examples with $k = 2$. Parameters are well estimated with actual values close to their respective point estimates and comfortably inside their 95% credibility intervals for the simulations with threshold at 90% data quantile. Notice also that the credibility intervals concentrate more around true values as sample sizes increase, as expected. Tail estimation is virtually unchanged when 3 components are used in the Gamma mixture. When only one Gamma distribution is used, the threshold is not well estimated and this affects also the estimation of the other tail parameters. Estimation is not as precise when the threshold is located at the 99% data quantile, as shown in Figure 7. The predictive density however is well estimated, as shown in Figure 4 .

Figure 5

Figure 6

Figure 7

Figures 8 and 9 illustrates the effect of large prior variances for the threshold for

12

sample size 1,000 and 10,000. Figure 8 shows that suitable values are required for the prior variance for the threshold to ensure appropriate inference with good recovery of the true value, when $n = 1,000$. Large prior variances may lead to erroneous estimation, indicating possible divergence. For larger data sets ($n = 10,000$), the specific value of the prior variance for the threshold does not seem to affect the posterior inference. Figure 9 shows that the threshold is well estimated and point and interval estimates are virtually the same for a large range of values for the prior variance for the threshold. This is a clear indication that the likelihood here is strong enough to correctly identify the true values.

The prior variance for the threshold affects the posterior correlation. When $\sigma_u^2 = 10$, it changed from 0.08 when $n = 1,000$ to 0.69, when $n = 10,000$. Despite the correlation, all parameters are well within their respective 95% posterior credibility interval limits. Similar results were obtained in other situations for a variety of values for the GPD parameters.

<center>Figure 8</center>

<center>Figure 9</center>

Table 1 and 2 shows the fit results of different models to a number of data generating conditions. It shows that in most cases, the best fitted models are the true ones, indicating the model ability to identify itself correctly. Table 1 shows that 75% of the results (9 out of 12) with the DIC identified the model correctly. This figure is slightly increased to 10 out of 12 with the BIC (Schwarz, 1978) . Theses figures also indicate that BIC and DIC seem to provide reliable sources for comparison of model fit even in these mixture settings. It can also be noticed that the DIC identifies the correct number of effective parameters through its $P_d$ component, as shown in Table 1.

A good example is the exercise with $n = 1,000$, $\sigma = 3$ and $\xi = 0.4$. The effective number of components is basically the same for $k = 2$ and $k = 3$. Figure 10 shows that when the true model has $k = 2$ components, and a model $MGPD_3$ is estimated, the weight of the 3rd component is about 0. Also, the weights of the first two components mimic the weights of model $MGPD_2$. In other words, the correct model is recovered.

<center>13</center>

Figure 10

Table 1

Table 2 shows the performance of the different models for data generated with threshold at the 99% data quantile. Model $MGPD_2$ obtained the best performance and no model in the $MG_k$ class outperforms it.

Table 2

One of the main interests in extreme data analysis is the correct identification of extreme points or higher quantiles. These were also estimated for all models considered and also compared against the true and the empirical quantiles. Table 3 shows some of these results. None of the methods seem to outstand when $n = 1,000$. The empirical estimate is better in 5 of them, while the $MGPD$ is better in 4 and the $MG$ in 3. The situation changes when $n = 10,000$. The MGPD estimates outperforms even the empirical estimates with better performance in 8 of the 12 simulations. A direct comparison between the $MGPD$ and $MG$ classes shows indifference between them when $n = 1,000$ and vast superiority of the $MGPD$ when $n = 10,000$ in 11 out of the 12 simulations. These results seem to indicate that the $MGPD$ class is appropriate for the study extreme values data and that its superiority over mixtures of Gamma distributions is more evident for larger data sets.

Table 3

This section serves a few purposes. The first and most obvious one is to ensure that the model is capable to identify data generated from itself. One could have anticipated that using only a mixture of Gamma densities would provide a fit just as adequate by enlarging the number of components on the mixture. The exercises performed in this section seem to indicate that this is not true. Also, the specification of mildly informative prior distribution for the threshold was discussed and shown to lead to sensible inference without the need for strong prior inputs. Basic assumptions about the threshold, available in any problem, are enough to ensure appropriate inference. In passing, the results indicate that DIC and BIC are reliable sources of comparison, even in these mixture settings.

## 4. Applications

This section shows results of real data analyses of extreme data from environmental sciences: river flow levels in Puerto Rico and pluviometric levels in Portugal. Comparisons of the models proposed here against verions of those in Wiper et al. (2001) and Behrens et al. (2004) are also carried out.

### 4.1. River flow in Puerto Rico

This analysis is based on datasets consisting in the measurement of the levels of flow of two rivers located in Northeast Puerto Rico in $ft^3/s$: Fajardo and Espírito Santo. The data was recorded daily from April 1967 to September 2002 and is freely available from *www.waterdata.usgs.gov*. We analysed fortnightly maxima data.

Figure 11 shows the predictive density of the three classes of models where differences are apparent. Model $MGPD_3$ presents higher values up to around 200. Model $MG_3$ follows model $MGPD_3$ closer than model $MGPD_1$ but gets away around the most likely threshold locations. All models place the mean posterior threshold at around the 80% data quantile.

Figure 11

Fit measures are provided in Table 4. According to the BIC, model $MG_2$ was the best fit for both rivers while according to the DIC, model $MGPD_3$ was the best fit for both rivers. In any case, there is clear indication of relevance of using mixtures for the central part of the data.

Table 4

### 4.2. Pluviometric levels in Portugal

This analysis is based on datasets consisting in the measurement of the amount of rain in two monitoring stations in Portugal: Barcelos, in the North, and Grândola, in the South. These two station were chosen to characterize contrasting climatological patterns between a rainier region (North) and a drier region (South). The data was recorded daily from 1931 to 2008 and is freely available from *www.snirh.pt*. We analyzed fortnightly

maxima data. Figure 12 presents the mean amount of rain in the different regions of Portugal between 1940 and 1997.

Figure 12

Figure 13 shows the predictive density of the three classes of models where differences are apparent. Model $MGPD_1$ estimates a very low threshold, resulting in a prediction that is not compatible with the data histogram. The predictive densities for models $MGPD_3$ and $MG_3$ are close in the central part of the distribution.

Figure 13

Table 5 presents the measures of fit for data from the two stations. The models proposed here presented superior performance for both rivers and for both assessment criteria: BIC and DIC. These results provide substantial evidence of their potential relevance in practical data analyses.

Table 5

*4.3. Estimation of higher quantiles*

Precise determination of higher quantiles is one of the main interests in extreme data analysis. These quantiles were evaluated for both data sets. Illustration of this task is provided below for river Espírito Santo and for station Barcelos in Table 6.

According to our preferred model, Espírito Santo river flows above $2,718$ $ft^3/s$ occur on average with 0.1% probability or around once every 10 years. Figure 14 shows the posterior histogram for the 99.9% quantile, denoted $q_{x,0.999}$. The distribution is skew as expected and concentrated around the corresponding data quantile. It may be compared against the corresponding quantile estimated by maximum likelihood (ML) methods, as in Coles et al. (1994). Setting the threshold value at the 3 posterior quartiles, the respective ML estimates of $q_{x,0.999}$ are given by $2,607$, $2,509$ and $2,562$ $ft^3/s$. In this case, the posterior point estimate of this quantile is closer to the classical estimate obtained with the a choice of the threshold at its lower tail.

Figure 14

16

Similarly, rainfall levels around Barcelos station that are above 185 $mm$ occur on average with 0.01% probability or around once each century. Figure 15 shows the posterior histogram for the 99.99% quantile, denoted $q_{x,0.9999}$. The distribution is also skew as expected and concentrated around the corresponding data quantile. It may also be compared against the corresponding quantile estimated by maximum likelihood (ML) methods. Setting the threshold value at the 3 posterior quartiles, the respective ML estimates of $q_{x,0.9999}$ are given by 148, 150 and 152 $mm$. In this case, the posterior point estimate of this higher quantile is closer to the classical estimate obtained with a choice of the threshold at its upper tail. No such choice is required to obtain the posterior distribution or its mean since these are based on integration over the other parameters, thus automatically incorporating their uncertainty.

Figure 14

Figure 15

Table 6

In the absence of knowledge of the true quantiles, comparisons can be made using the empirical quantile as benchmark to compare with. Results are inconclusive for the data of river Espírito Santo with alternation between models $MG$ and $MGPD$. Results are very clear for the Barcelos station with all higher quantiles from the $MGPD$ class closer to the empirical estimates than the estimates from the $MG$ class. These results provide further reassurance that the models proposed here provide sensible results and can be used for extreme value data analysis.

## 5. Conclusion

This paper presents a methodology for extreme value estimation based on a complete model for the entire sample space. Presence of a model component that takes into account theoretical results about the limiting behavior of extremes seem to improve the performance of the models. Additionally, the region where extreme behavior takes place is explicitly characterized through threshold estimation. This may prove useful for prac-

titioners wishing to establish the extreme region.

Simulation results suggest that this class of models can be identified from the data even in the presence of vague prior information. The only exception is the threshold that requires some form of prior information. This should not be too much of a problem because the threshold is usually located in the region of higher empirical quantiles and this information suffices for correct identification. Simulation also showed that BIC and DIC provide sound indication of model performance in these mixture settings.

These models may be improved and extended in a number of directions. Cabras et al. (2008) considered the inclusion of covariates in a GPD model formulation. One natural extension would be the consideration of a regression structure for the GPD component on our models. Another extension is the consideration of temporal data dependence. Initial efforts in this direction were proposed by Lopes (2007). Further developments in this direction to include trend and seasonality can provide useful additions to the understanding of these processes.

## Acknowledgements

## References

[1] Asmussen, S., 1987. Applied Probability and Queues, Wiley, New York.

[2] Behrens, C., Gamerman, D., Lopes, H.F., 2004. Bayesian analysis of extreme events with threshold estimation. Statistical Modelling. 4, 227-244.

[3] Bermudez, P., Turkman, M.A., Turkman, K.F., 2001. A predictive approach to tail probability estimation. Extremes. 4, 295-314.

[4] Cabras, S., Castellanos, M.A., Gamerman, D., 2008. A default approach for regression on parameters of generalized Pareto distribution. Technical Report. Laboratory of Statistics, Federal University of Rio de Janeiro, Brazil.

[5] Castellanos, M.A., Cabras, S., 2007. A default Bayesian procedure for the generalized Pareto distribution. Journal of Statistical Planning and Inference. 137, 473-483.

[6] Coles, S.G., Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data. Applied Statistics. 45, 463-78.

[7] Coles, S.G., 1994. Extreme Value Theory an Applications. In Galambos, J., Lechner, J., Mimiu, E. (Eds.),??????, Kluver Academic Publishers, ???, ???-???.

[8] Dalal, S., Hall, W., 1983. Approximating priors by mixtures of natural conjugate priors. Journal of the Royal Statistical Society, Series B. 45, 278-286.

[9] de Vore, R., Lorentz, G., 1993. Constructive Approximation. Springer Verlag, New York.

[10] Dey, D., Kuo, L., Sahu, S., 1995. A Bayesian predictive approach to determining the number of components in a mixture distribution. Statistics and Computing. 5, 297-305.

[11] Diebolt, J., Robert, C., 1994. Estimation of finite mixture distributions by Bayesian sampling. Journal of the Royal Statistical Society, Series B. 56, 363-375.

[12] Embrechts, P., Küppelberg, C., Mikosch, T., 1997. Extremal Events for Insurance and Finance. Springer, New York.

[13] 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.

[14] Frigessi, A., Haug, O., Rue, H., 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. Extremes. 5, 219-235.

[15] 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.

[16] 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.

[17] Lopes, H.F., 2007. Generalized Pareto models with time-varying tail behavior. Technical Report. The University of Chicago Booth School of Business, US.

[18] Pickands, J., 1975. Statistical Inference using extreme order statistics. Annals of Statistics. 3, 131-199.

[19] Richardson, S., 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.

[20] Roberts, G.O. and Rosenthal, J.S., 2006. Examples of adaptive MCMC. Technical Report. Lancaster University.

[21] Roeder, K., Wasserman, L., 1997. Practical Bayesian density estimation using mixtures of normals. Journal of the American Statistical Association. 92, 894-902.

[22] Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics. 6, 461-4.

[23] Smith, R.A., 1985. Maximum Likelihood estimation in a class of non-regular cases. Biometrika. 72, 67-90.

[24] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion and rejoinder). Journal of the Royal Statistical Society, Series B. 64, 583-639.

[25] Tancredi, A., Anderson, C., O'Hagan, A., 2006. Accounting for threshold uncertainty in extreme value estimation. Extremes. 9, 87-106.

[26] Titterington, D., Smith, A.F.M. and Makov, U., 1985. Statistical Analysis of Finite Mixture Distributions, Wiley, New Nork.

[27] von Mises, R., 1954. La distribution de la plus grande de n valeurs. American Mathematical Society: Selected Papers. 2, 271-294.

[28] 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 $(\mu, \eta)$ 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:

- $\xi$: Proposal transition kernel for $\xi$ is given by a truncated normal

$$\xi^* \mid \xi^{(s)} \sim N(\xi^{(s)}, V_\xi) I(-\sigma^{(s)}(M - u^{(s)}), \infty),$$

where $V_\xi$ is a variance appropriately chosen to ensure chain mixing and $M = \max(x_1, \ldots, x_n)$. So, $\xi^{(j+1)} = \xi^*$ with probability $\alpha_\xi$, where

$$\alpha_\xi = \min\left\{1, \frac{\pi(\Theta^*|x)\Phi((\xi^{(s)} + \sigma^{(s)}/(M - u^{(s)})/\sqrt{V_\xi}))}{\pi(\tilde{\Theta}|x)\Phi((\xi^* + \sigma^{(s)}/(M - u^{(s)})/\sqrt{V_\xi}))}\right\},$$

where $\Phi$ is the d.f. of the standard normal distribution, $\Theta^* = (\mu^{(s)}, \eta^{(s)}, p^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^*)$ and $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, p^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^{(s)})$.

- $\sigma$

  If $\xi^{(s+1)} > 0$, then $\sigma^*$ is sampled from the Gamma distribution $G(\sigma^{(s)}, \sigma^{(s)^2}/V_\sigma)$, where $V_\sigma$ is the variance of the proposal distribution appropriately chosen to ensure chain mixing.

  If $\xi^{(s+1)} < 0$, then $\sigma^*$ is sampled from a $N(\sigma^{(s)}, V_\sigma)I(-\xi^{(s+1)}(M - u^{(s)}), \infty)$

  So, $\sigma^{(s+1)} = \sigma^*$ with probability $\alpha_\sigma$ where, if $\xi^{(s+1)} < 0$,

  $$\alpha_\sigma = \min\left\{1, \frac{\pi(\Theta^*|x)\Phi((\sigma^{(s)} + \xi^{(s+1)}(M - u^{(s)}))/\sqrt{V_\sigma}))}{\pi(\tilde{\Theta}|x)\Phi((\sigma^* + \xi^{(s+1)}(M - u^{(s)}))/\sqrt{V_\sigma}))}\right\},$$

  and if $\xi^{(s+1)} > 0$,

  $$\alpha_\sigma = \min\left\{1, \frac{\pi(\Theta^*|x)f_G(\sigma^{(s)}|\sigma^*, \sigma^{*2}/V_\sigma)}{\pi(\tilde{\Theta}|x)f_G(\sigma^*|\sigma^{(s)}, \sigma^{(s)2}/V_\sigma)}\right\},$$

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

- $u$

  The threshold $u^*$ is sampled from a $N(u^{(s)}, V_u)I(a^{(s+1)}, \infty)$ distribution where $a^{(s+1)} = \min(x_1, \ldots, x_n)$ if $\xi^{(s+1)} \geq 0$ and $a^{(s+1)} = M + \sigma^{(s+1)}/\xi^{(s+1)}$, if $\xi^{(s+1)} < 0$. The lower limit of the truncation is chosen to satisfy the sample space of the GPD in (3). $V_u$ is the variance chosen to ensure appropriate chain mixing. Then, accept $u^{(s+1)} = u^*$ with probability $\alpha_u$, where

  $$\alpha_u = \min\left\{1, \frac{\pi(\Theta^*|x)\Phi((u^{(s)} - a^{(s+1)})/\sqrt{V_u})}{\pi(\tilde{\Theta}|x)\Phi((u^* - a^{(s+1)})/\sqrt{V_u})}\right\},$$

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

- $(\mu_j, \eta_j)$ For $j = 1, ..., k$

  Since $\eta_j$ is positive, the proposal kernel is taken as the Gamma distribution

  $$\eta_j^*|\eta_j^{(s)} \sim G(\eta_j^{(s)}, \eta_j^{(s)^2}/V_{\eta_j}),$$

22

where $\eta_j^{(s)}$ is the value $\eta_j$ at iteration $s$ and $V_{\eta_j}$ is the variance chosen to ensure appropriate chain mixing. Note that $E(\eta_j^* \mid \eta_j^{(s)}) = \eta_j^{(s)}$, and $Var(\eta_j^* \mid \eta_j^{(s)}) = V_{\eta_j}$, $j = 1, \ldots, k$.

Since $\mu_j$ is also positive, the proposal kernel is taken as the Gamma distribution

$$\mu_j^* | \mu_j^{(s)} \sim G(\mu_j^{(s)}, \mu_j^{(s)^2}/V_{\mu_j}) I(\mu_1^{(s+1)} < \ldots < \mu_{j-1}^{(s+1)} < \mu_j^{(s)} \ldots < \mu_k^{(s)}),$$

where $\mu_j^{(s)}$ is the value of $\mu_j$ at iteration $s$ and $V_{\mu_j}$ is the variance chosen to ensure appropriate chain mixing.

The values $\eta_j^{(s+1)} = \eta_j^*$ and $\mu_j^{(s+1)} = \mu_j^*$ are accepted with probability $\alpha_{\mu_j, \eta_j}$, where

$$\alpha_{\mu_j, \eta_j} = \min\left\{1, \frac{\pi(\Theta^*|x) f_G(\mu_j^{(s)}|\mu_j^*, \mu_j^{*2}/V_\mu) f_G(\eta_j^{(s)}|\eta_j^*, \eta_j^{*2}/V_\eta) I(\mu_1^{(s+1)} < \ldots < \mu_j^* \ldots < \mu_k^{(s)})}{\pi(\tilde{\Theta}|x) f_G(\mu_j^*|\mu_j^{(s)}, \mu_j^{(s)2}/V_\mu) f_G(\eta_j^*|\eta_j^{(s)}, \eta_j^{(s)2}/V_\eta) I(\mu_1^{(s+1)} < \ldots < \mu_j^{(s)} \ldots < \mu_k^{(s)})}\right\},$$

where $\Theta^* = (\eta_{<j}^{(s+1)}, \eta_j^*, \eta_{>j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_j^*, \mu_{>j}^{(s)}, p^{(s)}, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$ and

$\tilde{\Theta} = (\eta_{<j}^{(s+1)}, \eta_{\geq j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_{\geq j}^{(s)}, p^{(s)}, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$.

- $p$

The vector of weights is proposed from a Dirichlet distribution $p^* \sim D_k(V_p p_1^{(s)}, \ldots, V_p p_k^{(s)})$. So, $p^{(s+1)} = p^*$ with probability

$$\alpha_p = \min\left\{1, \frac{\pi(\Theta^*|x) f_D(p^{(s)}|p^*)}{\pi(\tilde{\Theta}|x) f_D(p^*|p^{(s)})}\right\},$$

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

| | $\sigma = 2$ | | | | $\sigma = 3$ | | | | $\sigma = 5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $P_d$ | DIC | BIC | $k$ | $P_d$ | DIC | BIC | $k$ | $P_d$ | DIC | BIC |
| | | | | | | n=1000 | | | | | |
| | | | | | | $\xi = 0.4$ | | | | | |
| 1 | 0.77 | 4653.7 | 4691.5 | 1 | 3.61 | 4603.9 | 4648.1 | 1 | 4.58 | 4699.8 | 4747.6 |
| 2 | 6.79 | 4468.2 | 4542.5 | 2 | 7.81 | 4547.1 | 4624.6 | 2 | 4.70 | 4660.6 | 4728.1 |
| 3 | 6.50 | 4469.1 | 4563.1 | 3 | 7.83 | 4547.4 | 4645.4 | 3 | 4.17 | 4662.5 | 4749.3 |
| 3* | 7.70 | 4466.6 | 4543.9 | 3* | 7.41 | 4552.2 | 4628.4 | 3* | 2.68 | 4684.8 | 4742.3 |
| | | | | | | $\xi = -0.4$ | | | | | |
| 1 | 2.85 | 4388.9 | 4435.3 | 1 | 3.56 | 4462.7 | 4510.1 | 1 | 5.08 | 4754.9 | 4807.6 |
| 2 | 6.77 | 4345.5 | 4421.7 | 2 | 6.53 | 4426.3 | 4501.9 | 2 | 8.81 | 4522.6 | 4603.8 |
| 3 | 6.61 | 4345.8 | 4442.3 | 3 | 6.76 | 4425.7 | 4522.6 | 3 | 9.03 | 4522.1 | 4624.6 |
| 2* | 5.18 | 4366.1 | 4418.1 | 2* | 5.27 | 4440.9 | 4493.1 | 2* | 5.12 | 4556.0 | 4607.9 |
| | | | | | | n=10000 | | | | | |
| | | | | | | $\xi = 0.4$ | | | | | |
| 1 | -0.95 | 46817 | 46857 | 1 | 2.04 | 47610 | 47665 | 1 | 6.55 | 48621 | 48691 |
| 2 | 7.77 | 44718 | 44814 | 2 | 5.68 | 45530 | 45621 | 2 | 7.20 | 46548 | 46644 |
| 3 | 4.61 | 44727 | 44842 | 3 | 4.67 | 45534 | 45649 | 3 | 7.07 | 46548 | 46671 |
| 4* | 7.68 | 44734 | 44859 | 3* | 7.17 | 45560 | 45656 | 5* | 9.98 | 46594 | 46753 |
| | | | | | | $\xi = -0.4$ | | | | | |
| 1 | 5.71 | 45372 | 45440 | 1 | 1.79 | 44416 | 44467 | 1 | 3.06 | 47215 | 47275 |
| 2 | 8.65 | 43171 | 43272 | 2 | 7.51 | 43986 | 44083 | 2 | 7.80 | 45008 | 45106 |
| 3 | 8.57 | 43172 | 43300 | 3 | 4.32 | 43997 | 44112 | 3 | 7.57 | 45009 | 45133 |
| 5* | 8.94 | 43211 | 43368 | 2* | 4.79 | 44064 | 44129 | 4* | 6.81 | 45057 | 45181 |

Table 1: Measures of fit for the simulations with threshold set at the 90% data quantile. ∗ indicates the best $MG_k$ model, ie, with smallest DIC value.

| | $\sigma = 2$ | | | | $\sigma = 3$ | | | | $\sigma = 5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | $P_d$ | DIC | BIC | $k$ | $P_d$ | DIC | BIC | $k$ | $P_d$ | DIC | BIC |
| | | | | | n=10000 | | | | | | |
| | | | | | $\xi = 0.4$ | | | | | | |
| 1 | 5.41 | 46656 | 46718 | 1 | -0.56 | 46771 | 46813 | 1 | 3.54 | 46906 | 46967 |
| 2 | 4.08 | 44623 | 44708 | 2 | 6.76 | 44696 | 44791 | 2 | 6.28 | 44799 | 44893 |
| 3 | 3.93 | 44623 | 44736 | 3 | 7.18 | 44695 | 44819 | 3 | 7.01 | 44798 | 44921 |
| 4* | 9.39 | 47116 | 47244 | 3* | 5.65 | 44731 | 44824 | 4* | 6.67 | 44855 | 44978 |
| | | | | | $\xi = -0.4$ | | | | | | |
| 1 | 3.04 | 44535 | 44592 | 1 | 2.04 | 47610 | 47665 | 1 | 1.01 | 46415 | 46469 |
| 2 | 6.13 | 44104 | 44198 | 2 | 5.23 | 44187 | 44279 | 2 | 5.74 | 44288 | 44381 |
| 3 | 5.78 | 44105 | 44226 | 3 | 3.28 | 44193 | 44307 | 3 | 5.52 | 44289 | 44409 |
| 5* | 5.40 | 44122 | 44272 | 3* | 5.07 | 44204 | 44298 | 5* | 8.13 | 44307 | 44462 |

Table 2: Measures of fit for the simulations with threshold set at the 99% data quantile. $*$ indicates the best $MG_k$ model, ie, with smallest DIC value.

| $Prob$ | T | E | 1 | 2 | 3 | $MG_k$ | T | E | 1 | 2 | 3 | $MG_k$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | n=1000 | | | | | | n=10000 | | | |
| | | | | | | $\sigma = 2$ | | | | | | |
| 0.95 | 10.4 | 10.5 | 10.0 | 10.4 | 10.4 | 10.5 | 10.7 | 10.7 | 10.0 | 10.7 | 10.7 | 10.8 |
| 0.99 | 16.4 | 15.1 | 15.0 | 16.0 | 16.0 | 15.2 | 16.6 | 16.4 | 15.2 | 16.4 | 16.4 | 16.5 |
| 0.999 | 35.4 | 33.8 | 42.8 | 44.4 | 46.2 | 37.3 | 35.6 | 32.0 | 30.4 | 33.6 | 33.8 | 32.5 |
| 0.9999 | 83.1 | 54.0 | 5950.0 | 710.3 | 516.8 | 65.2 | 83.3 | 60.0 | 65.5 | 74.6 | 75.5 | 47.7 |
| | | | | | | $\sigma = 3$ | | | | | | |
| 0.95 | 11.2 | 11.3 | 12.5 | 11.0 | 11.0 | 11.1 | 11.5 | 11.5 | 10.4 | 11.4 | 11.4 | 11.4 |
| 0.99 | 20.2 | 18.3 | 20.0 | 19.8 | 19.9 | 19.0 | 20.4 | 20.1 | 18.0 | 20.0 | 19.9 | 21.4 |
| 0.999 | 48.7 | 46.3 | 35.9 | 54.3 | 56.8 | 48.8 | 48.9 | 43.5 | 42.3 | 47.2 | 46.7 | 41.2 |
| 0.9999 | 120.2 | 76.8 | 54.9 | 183.3 | 209.8 | 85.0 | 120.5 | 85.4 | 103.7 | 116.2 | 113.7 | 60.4 |
| | | | | | | $\sigma = 5$ | | | | | | |
| 0.95 | 12.8 | 12.9 | 13.8 | 12.5 | 12.4 | 12.4 | 13.1 | 13.1 | 11.4 | 13.0 | 13.0 | 12.8 |
| 0.99 | 27.8 | 24.6 | 26.0 | 27.4 | 27.6 | 29.0 | 28.0 | 27.4 | 24.2 | 27.3 | 27.3 | 29.7 |
| 0.999 | 75.2 | 71.3 | 53.5 | 92.1 | 96.1 | 64.4 | 75.5 | 66.4 | 63.2 | 71.0 | 70.8 | 53.7 |
| 0.9999 | 194.5 | 121.7 | 100.8 | 363.0 | 402.9 | 109.6 | 194.7 | 136.3 | 157.0 | 177.1 | 176.1 | 77.5 |

Table 3: Quantiles of simulations with $\xi = 0.4$ and threshold set at the 90% data quantile. $Prob = P(X \leq q)$, T = True, E= Empirical, $1 = MGPD_1$, $2 = MGPD_2$, $3 = MGPD_3$. $MG_k$ refers to the best model in this class.

|   | Espírito Santo | | | Fajardo | | |
|---|---|---|---|---|---|---|
| $k$ | Pd | DIC | BIC | Pd | DIC | BIC |
| | $MG_k$ | | | | | |
| 1 | 2.12 | 11330 | 11353 | 2.01 | 11441 | 11463 |
| 2 | 3.87 | 11275 | 11322 | 4.65 | 11310 | 11355 |
| 3 | 3.87 | 11275 | 11342 | 4.53 | 11307 | 11371 |
| 4 | 4.68 | 11273 | 11362 | 9.31 | 11293 | 11391 |
| 5 | 11.04 | 11288 | 11410 | 7.22 | 11265 | 11382 |
| | $MGPD_k$ | | | | | |
| 1 | 0.83 | 11299 | 11336 | 1.05 | 11327 | 11361 |
| 2 | 4.11 | 11264 | 11329 | 0.53 | 11327 | 11380 |
| 3 | 6.71 | 11255 | 11349 | 6.28 | 11264 | 11357 |
| 4 | 5.32 | 11265 | 11377 | 6.89 | 11268 | 11384 |

Table 4: Fit measures for the rivers Fajardo and Espírito Santo.

| $k$ | Barcelos | | | Grândola | | |
|---|---|---|---|---|---|---|
| | Pd | DIC | BIC | Pd | DIC | BIC |
| | $MG_k$ | | | | | |
| 1 | 2.02 | 8149 | 8172 | 1.98 | 6676 | 6701 |
| 2 | 3.94 | 7951 | 8001 | 4.21 | 6391 | 6442 |
| 3 | 4.50 | 7931 | 8003 | 4.18 | 6391 | 6462 |
| 4 | 3.88 | 7933 | 8022 | 4.26 | 6391 | 6483 |
| 5 | 4.55 | 7868 | 7980 | 4.05 | 6391 | 6503 |
| | $MGPD_k$ | | | | | |
| 1 | 3.41 | 8143 | 8184.1 | 0.20 | 4491 | 4522 |
| 2 | 5.76 | 7639 | 7712 | 5.31 | 4325 | 4397 |
| 3 | 7.08 | 7612 | 7709 | 7.01 | 4304 | 4400 |
| 4 | 6.63 | 7614 | 7729 | 6.42 | 4303 | 4417 |

Table 5: Fit measures for the stations Barcelos and Grândola

| | Espírito Santo (in $ft^3/s$) | | | | | Barcelos (in $mm$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Prob* | E | 1 | 2 | 3 | $MG_k$ | E | 1 | 2 | 3 | $MG_k$ |
| 0.95 | 798 | 793.29 | 813.52 | 807.54 | 842.7 | 73.1 | 74.54 | 77.54 | 73.29 | 74.71 |
| 0.99 | 1360 | 1426.04 | 1450.79 | 1443.85 | 1398.8 | 99.4 | 101.73 | 105.38 | 102.24 | 104.09 |
| 0.999 | 2600 | 2677.56 | 2726.55 | 2718.29 | 2197.0 | 117.5 | 137.84 | 139.91 | 137.77 | 151.50 |
| 0.9999 | N/A | 4612.30 | 4734.16 | 4710.35 | 3014.0 | 143.5 | 171.41 | 176.12 | 184.54 | 233.00 |

Table 6: Higher quantiles for river Espírito Santo and Barcelos station. Prob= $P(X \leq q)$, E = Empirical, $1 = MGPD_1$, $2 = MGPD_2$, 3=$MGPD_3$. $MG_k$ refers to the best model in this class.
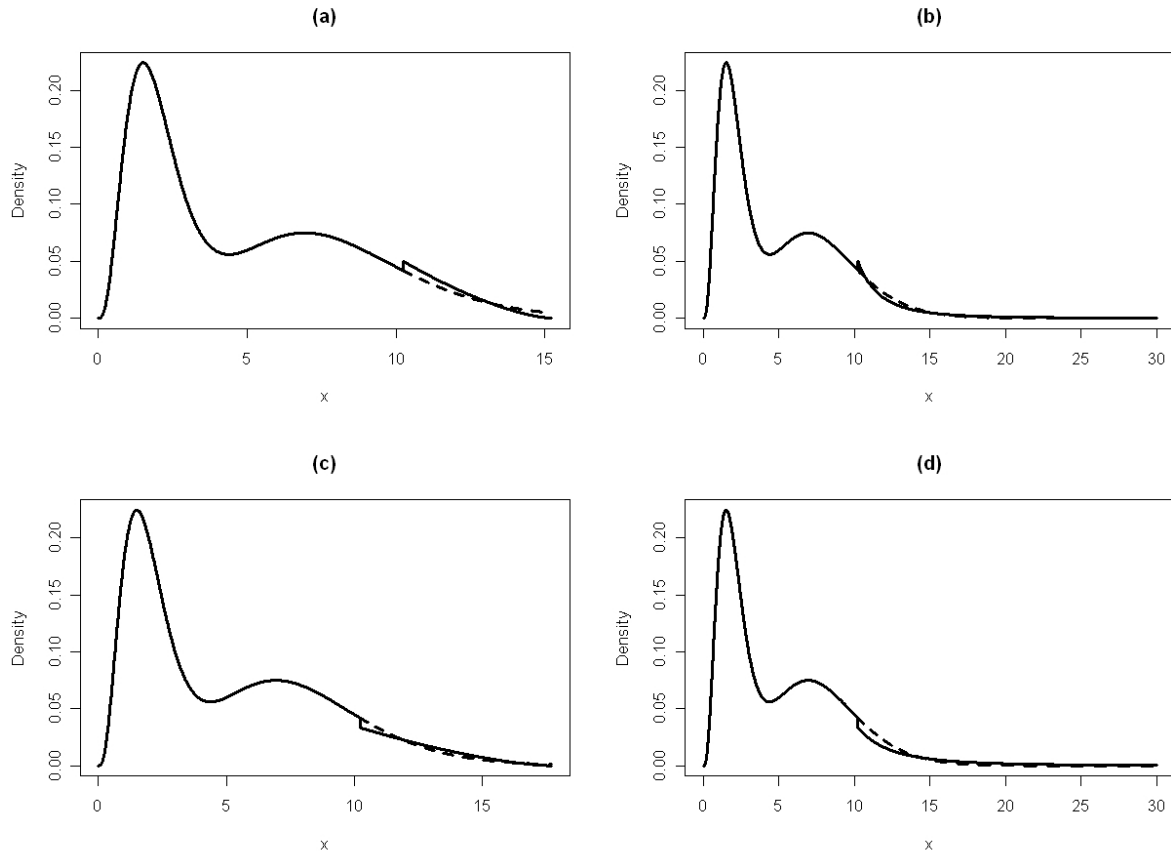
Figure 1: Probability density function of the proposed model for a number of parameter values: (a) - $\sigma = 2$ and $\xi = -0.4$; (b) - $\sigma = 2$ and $\xi = 0.4$; (c) - $\sigma = 3$ and $\xi = -0.4$; (d) - $\sigma = 3$ and $\xi = 0.4$. The central part is based on a mixture of 2 Gamma distributions. The dashed lines represent the continuation of the mixture of Gamma densities beyond the threshold.
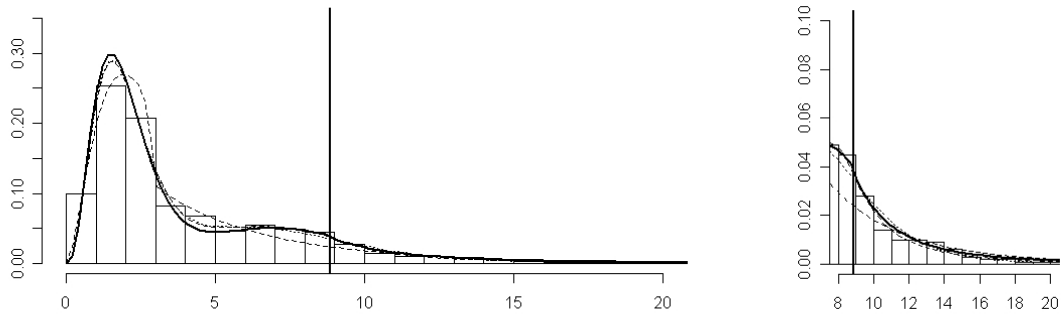
Figure 2: Predictive density for data simulated with $\xi = 0.4$, $\sigma = 3$, $n = 1,000$ and threshold set at 8.85, the 90% data quantile. Left panel: full density; right panel: detail of density in the tail. Full line is true density, dashed lines are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and dotted line is the predictive density from model $MG_3$, the best in the $MG_k$ class. Vertical line indicates location of the threshold.
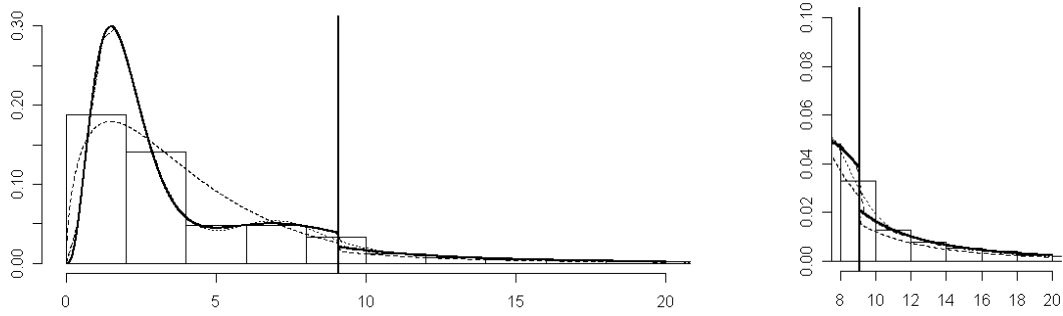
Figure 3: Predictive density for data simulated with $\xi = 0.4$, $\sigma = 5$, $n = 10,000$ and threshold set at 9.08 (the 90% data quantile). Left panel: full density; right panel: detail of density in the tail. Full line is true density, dashed lines are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and dotted line is the predictive density from model $MG_5$, the best in the $MG_k$ class. Vertical line indicates location of the threshold.
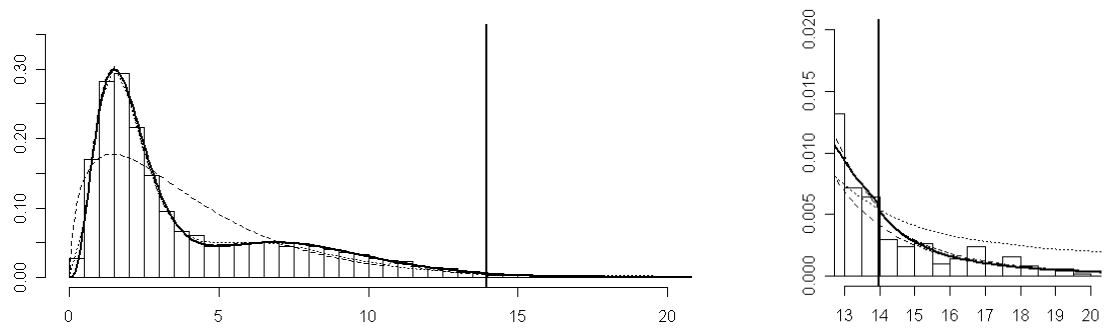
Figure 4: Predictive density for data simulated with $\xi = 0.4$, $\sigma = 2$, $n = 10,000$ and threshold set at 13.97 (the 99% data quantile). Left panel: full density; right panel: detail of density in the tail. Full line is true density, dashed lines are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and dotted line is the predictive density from model $MG_3$, the best in the $MG_k$ class. Vertical line indicates location of the threshold.
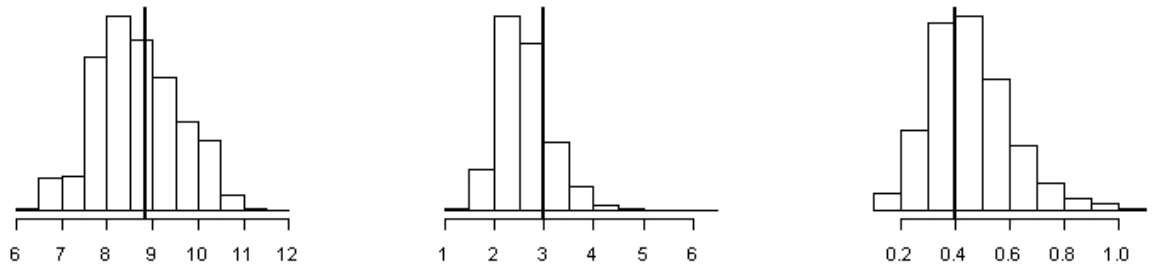
Figure 5: Posterior histogram of GPD parameters when $\xi = 0.4$, $\sigma = 3$, $n = 1,000$ and threshold set at 8.85 (the 90% data quantile). From left to right: $u$, $\sigma$ and $\xi$. Vertical lines indicates location of true values.
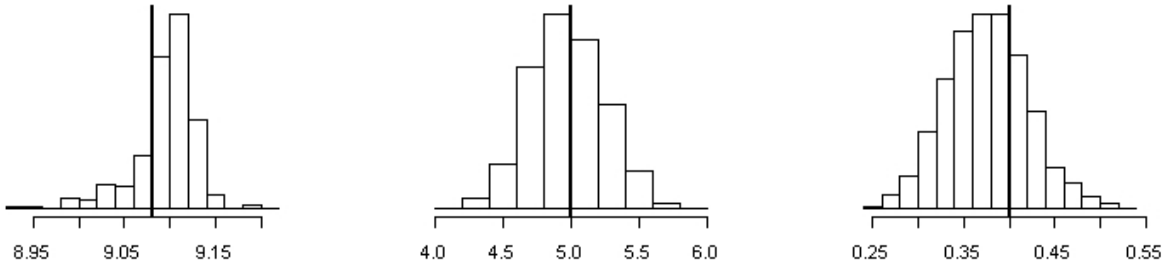
Figure 6: Posterior histogram of GPD parameters when $\xi = 0.4$, $\sigma = 5$ and $n = 10,000$ and threshold set at 9.08 (the 90% data quantile). From left to right: $u$, $\sigma$ and $\xi$. Vertical lines indicates location of true values.
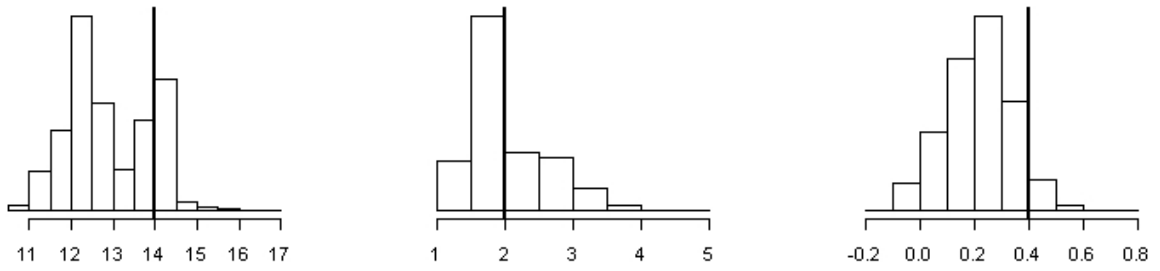
Figure 7: Posterior histogram of GPD parameters when $\xi = 0.4$, $\sigma = 2$, $n = 10,000$ and threshold set at 13.97 (the 99% data quantile). From left to right: $u$, $\sigma$ and $\xi$. Vertical lines indicates location of true values.
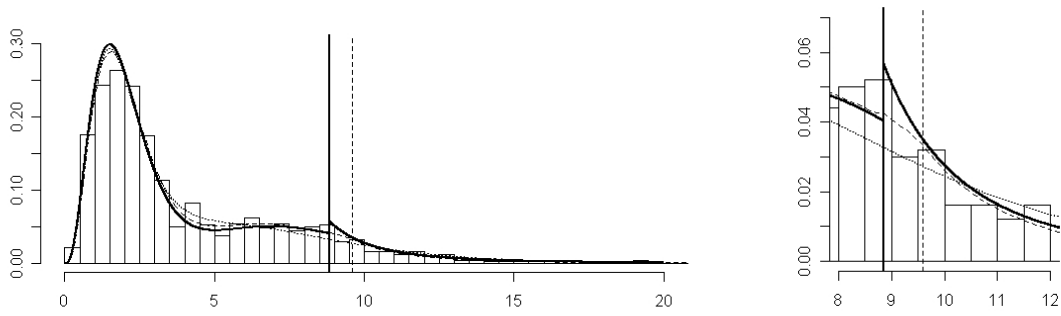
Figure 8: Predictive density for data (represented in histogram form) simulated with $k = 2$, $\sigma = 2$, $\xi = 0.4$, $u = 8.85$ and $n = 1,000$. Left panel: full density; right panel: density around the true thresohld value. The full line indicates the true density, the dashed line is the predictive density for model with prior variance $\sigma_u^2 = 10$ for the threshold, the dotted line is the predictive density for model with prior variance $\sigma_u^2 = 10,000$ for the threshold. Vertical lines indicates location of threshold: full - true; dashed - model with $\sigma_u^2 = 10$. The 95% posterior credibility interval for the threshold is (8.06,11.97), when $\sigma_u^2 = 10$.
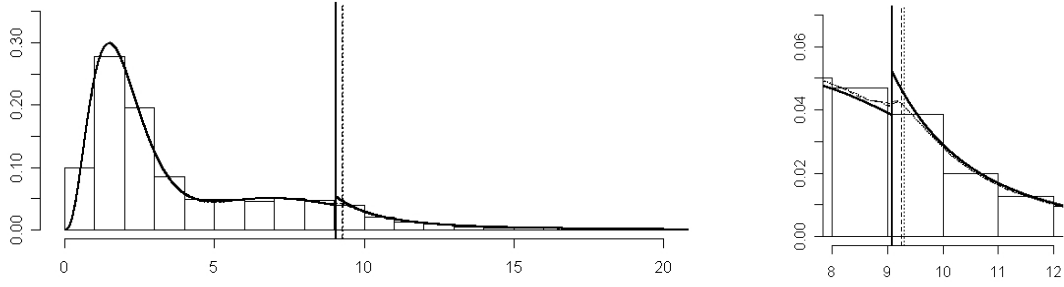
Figure 9: Predictive density for data (represented in histogram form) simulated with $k = 2$, $\sigma = 2$, $\xi = 0.4$, $u = 9.08$ and $n = 10,000$. Left panel: full density; right panel: densidade around the true thresohld value. The full line indicates the true density and, the dashed line is the predictive density for $\sigma_u^2 = 10$ and the dotted line is the predictive density for $\sigma_u^2 = 10,000$. Vertical lines indicates location of threshold: full - true; dashed - model with $\sigma_u^2 = 10$, dotted - model with $\sigma_u^2 = 10,000$. The 95% posterior credibility interval for the threshold are (8.73,10.14), when $\sigma_u^2 = 10$, and (8.76,10.18), when $\sigma_u^2 = 10,000$.
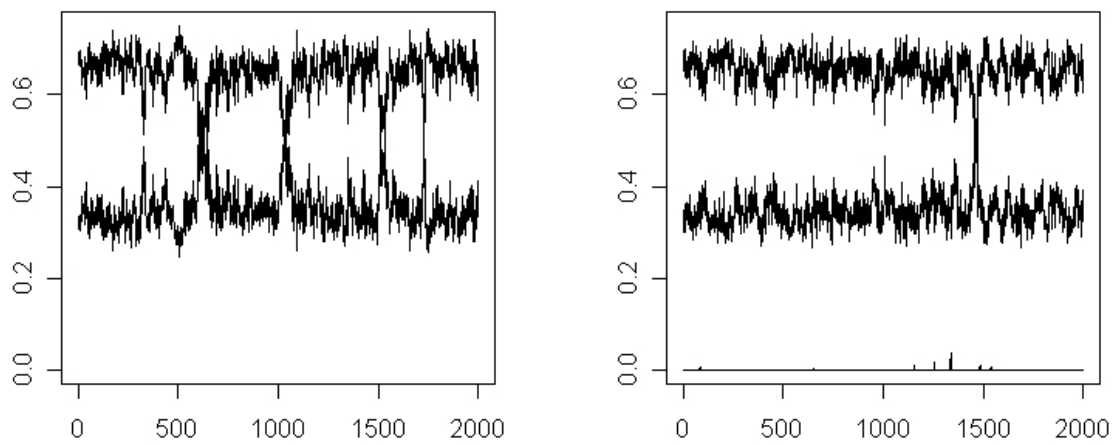
Figure 10: Trace plots of weights p when $n = 1,000$, $\sigma = 3$ e $\xi = 0.4$: left - weights from model $MGPD_2$; right - weights from model $MGPD_3$.
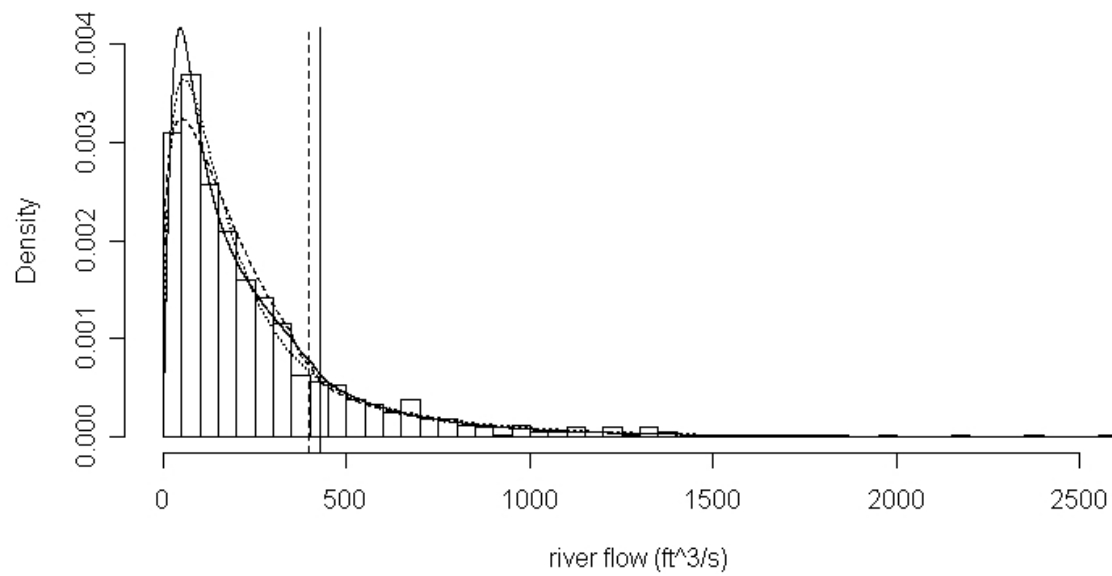
Figure 11: Predictive densities for data from Espírito Santo river: full line - $MGPD_3$; dashed line - $MGPD_1$, dotted line - $MG_3$, vertical lines: respective posterior means of the threshold.
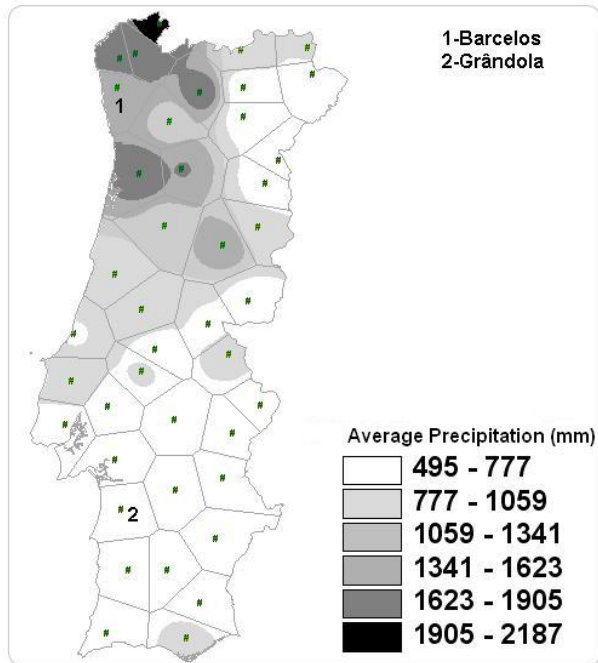
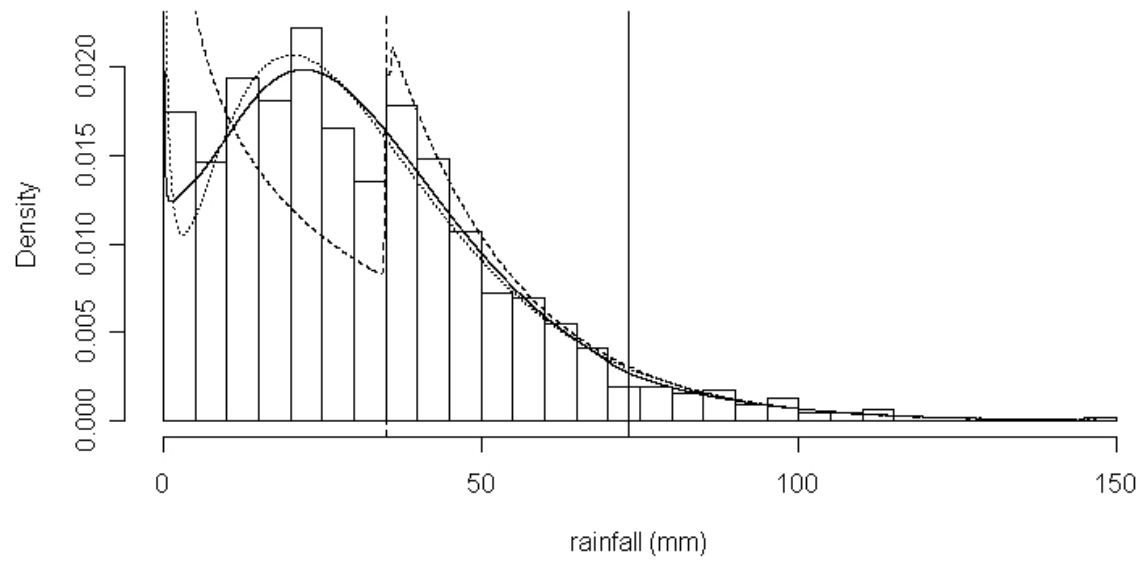Figure 12: mean pluviometric levels in Portugal, with location of two stations used in the analysis.

Figure 13: Predictive densities for data from Barcelos station: full line - $MGPD_3$; dashed line - $MGPD_1$, dotted line - $MG_3$, vertical lines: respective posterior means of the threshold.
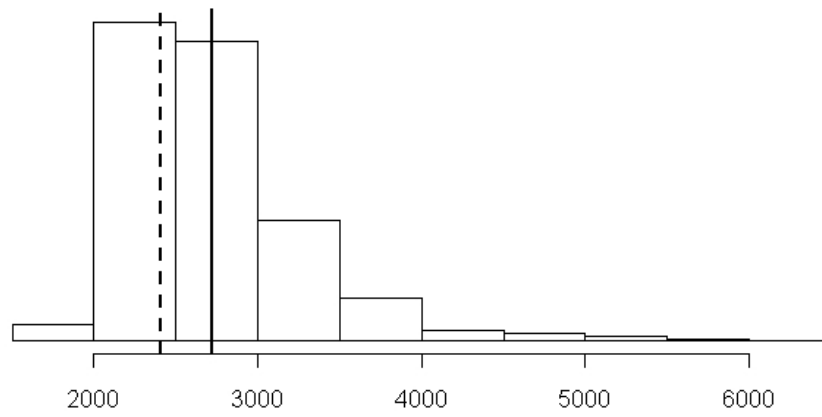
Figure 14: Posterior histogram of the 99.9% quantile of river Espírito Santo, under $MGPD_3$. Vertical lines: full - posterior mean; dashed - empirical. The posterior standard deviation of the 99.9% quantile is 560.157 $ft^3/s$.

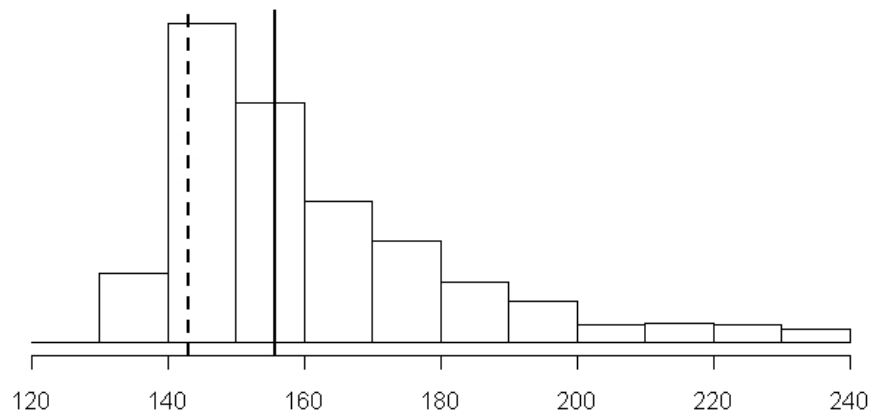Figure 15: Posterior histogram of the 99.99% quantile of Barcelos station, under $MGPD_3$. Vertical lines: full - posterior mean; dashed - empirical. The posterior standard deviation of the 99.9% quantile is 22.51 $mm$.