Stat Comput (2012) 22:661–675
DOI 10.1007/s11222-011-9270-z

A semiparametric Bayesian approach to extreme value estimation

Fernando Ferraz do Nascimento · Dani Gamerman · Hedibert Freitas Lopes

Received: 7 February 2011 / Accepted: 5 July 2011 / Published online: 3 August 2011

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 below 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 is provided. Simulation studies suggest a few useful guidelines to evaluate the relevance of the proposed procedures. They also provide empirical evidence about the improvement of the proposed methodology over existing approaches. Models are then applied to environmental data sets. The paper is concluded with a few directions for future work.

Keywords Bayesian · GPD · Higher quantiles · MCMC · Threshold estimation · Nonparametric estimation of curves

F.F. do Nascimento (✉)
Departamento de Informática e Estatística, Universidade Federal do Piauí, Campus Ministro Petrônio Portela, SG2, 64049-550 Teresina, PI, Brazil
e-mail: fernandofn@ufpi.edu.br

D. Gamerman
Instituto de Matemática, Federal University of Rio de Janeiro, Caixa Postal 68530, 21945-970, Rio de Janeiro, RJ, Brazil
e-mail: dani@im.ufrj.br

H.F. Lopes
Booth School of Business, The University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL, 60637, USA
e-mail: hlopes@ChicagoBooth.edu

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. Nonparametric estimation and extreme value theory will be used as building blocks in our modelling strategy. Therefore, the main ideas behind these approaches are briefly introduced next.

1.1 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 and Tippett (1928) theorem. It establishes the three possible distributions for maxima 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. $F(x|u) = P(X \le u + x|X > u)$ 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

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

where $x - u > 0$ for $\xi \ge 0$ and $0 \le x - u < -\sigma/\xi$ 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 $\xi \ge 0$. Hereafter, description of GPD will concentrated over its range C_ξ of possible values to simplify notation.

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 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 ξ 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, i.e., q-values satisfying $P(X > q) = 1 - p$ for large values of p. The theory above allows also estimation of these higher quantiles beyond the threshold as they are simply functions of the GPD parameters. 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

$$q = u + \frac{((1-p)^{-\xi} - 1)\sigma}{\xi}. \tag{2}$$

These quantiles are important design parameters specially in extreme cases with p approaching 1 and we shall concentrate on their estimation. They will also illustrate the advantage of incorporation of the GPD into the model.

1.2 Nonparametric estimation of curves

Mixture modeling can be included among the many nonparametric techniques for density estimation. It provides 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

$$h(x \mid \theta, \mathbf{p}) = \sum_{j=1}^{k} p_j f_G(x \mid \mu_j, \eta_j) \tag{3}$$

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

$$f_G(x|\mu, \eta) = \frac{(\eta/\mu)^\eta}{\Gamma(\eta)} x^{\eta-1} \exp\big(-(\eta/\mu)x\big), \quad \text{for } x > 0. \tag{4}$$

The μ_j's and η_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 μ and the shape parameter η. This choice will simplify model specifications in the sequel.

The above papers provide theoretical and empirical evidence that mixture of Gammas 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 provides a precise description of the tail, designed to overcome these difficulties.

1.3 Related work

Different approaches have been proposed in the literature recently under the Bayesian paradigm to analyse extreme values and to determine the threshold. Bermudez et al. (2001) suggest a Bayesian approach to the peaks over threshold

(POT) method. They only consider the probability cumulated up to the threshold and estimate it based on the data frequency. Frigessi et al. (2002) considers mixture of two distributions: GPD and Weibull, with data dependent weights. Threshold choice is performed indirectly. Tancredi et al. (2006) 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, where the threshold is a parameter to be estimated. Diebolt et al. (2005) uses a continuous mixture of Gamma distributions for extreme modeling but assumes the threshold to be fixed and the model is only valid for positive shape GPD parameter. This paper builds on their work by applying a mixture of Gammas 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 under the Bayesian paradigm, with prior information playing an important role in the estimation procedures. Estimation was implemented 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. It also shows that the models proposed here outperform existing approaches. 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 out at possible extensions.

2 Model

In view of the introductory discussion, it seems reasonable to contemplate a model that incorporates a nonparametric specification, using mixture of Gamma's 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 (1) instead of a simple mixture of Gamma when data on excesses is involved.

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

$$f(x|\theta,\mathbf{p},\Psi)=\begin{cases}h(x\mid\mu,\eta,\mathbf{p}), & \text{if } x\le u\\ [1-H(u\mid\mu,\eta,\mathbf{p})]g(x|\Psi), & \text{if } x>u\end{cases} \tag{5}$$

where H is the d.f. of the mixture of Gammas, already presented in Sect. 1. Pickands' (1975) theorem is only applicable when H belongs to the domain of attraction of a GEV distribution. It is well known that the Gamma distribution belongs to the maximum domain of attraction of a Gumbel distribution. It can be easily shown that the result can be extended for mixture of Gamma distributions based on the results of Embrechts et al. (1997), p. 156. Thus, the limiting conditional distribution of excesses from a mixture of Gamma distributions follows a GPD and the GPD assumption for the tail behavior of mixtures of Gammas is justified.

Forms for this density are shown in Fig. 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. Appropriate application of Bayesian estimation procedures along with appropriate choice of models will basically remove this problem as explained in Sect. 2.2. Simulation studies and real data applications will show that the model is well estimated and does not exhibit any discontinuity.

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.

Recall from Sect. 1.1 that the GPD parameters are not uniquely identified. The model (5) 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

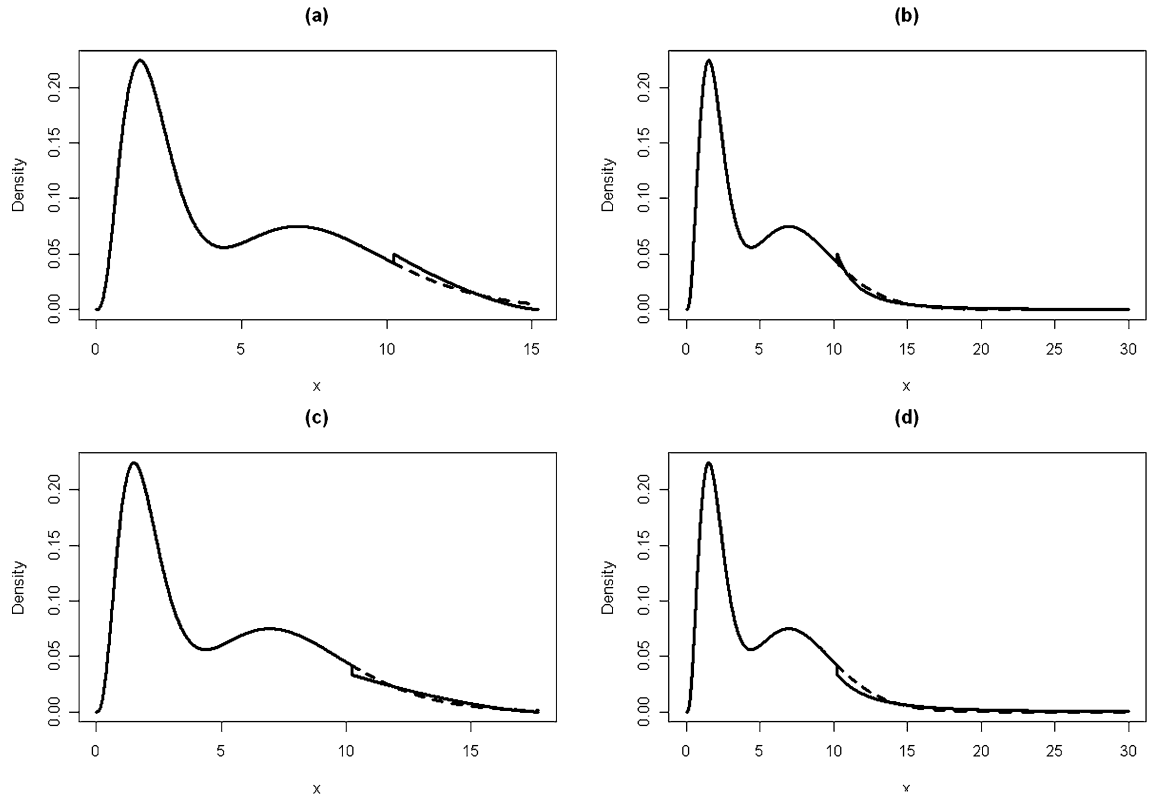

Fig. 1 Probability density function of the proposed model for a number of parameter values: (**a**) $\sigma = 2$ and $\xi = -0.4$. (**b**) $\sigma = 2$ and $\xi = 0.4$. (**c**) $\sigma = 3$ and $\xi = -0.4$. (**d**) $\sigma = 3$ and $\xi = 0.4$. The central part is based on a mixture of 2 Gamma distributions. The *dashed lines* represent the continuation of the mixture of Gamma densities beyond the threshold

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

$$p = H(q \mid \mu, \eta, \mathbf{p}) = \sum_{j=1}^{k} p_j \int_0^q f_G(x \mid \mu_j, \eta_j) dx. \tag{6}$$

There is no explicit solution for (6) 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

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

Therefore, it is straightforward to obtain the p-quantile as (2), where the quantity p in equation is replaced by

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

for quantiles beyond the threshold. Typically, one is interested in high quantiles well above the threshold but similar calculations can be performed for lower quantiles, even below the threshold.

Note that this quantile is a highly nonlinear function 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 behaviour 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 μ is taken in the form

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

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

The prior for shape parameters η is taken as a product of Gamma distributions with $\eta \sim G(c_j/d_j, c_j)$, 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 $D_k(w_1, \ldots, w_k)$ 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 σ and ξ. 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 (σ, ξ) 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 μ_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 Sect. 1.2. 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 $\mathbf{x} = (x_1, \ldots, x_n)$ from (5), the posterior density is obtained in the log scale as

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

where Z is a log 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 (Gamerman and 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 none has a full conditional density in recognizable form. Unlike previous work on mixtures (e.g. 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.

The code was developed in OxMetrics4 (Doornik 1996) in a PC with processor Intel Atom N270 1.66 Hz and 1 Gb RAM. 15,000 iterations were used for burn-in and the last 10,000 iterations were used for inference after thinning at every 20 iterations. The processing time allowed for 20 (or 2) iterations per second when $n = 1{,}000$ (or $n = 10{,}000$). 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 (2009), 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 $\mathbf{x}$, the density for a new observation x_{n+1} is given by

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

This predictive density is the appropriate quantity to be used for making inference about the resulting density for the

data. Another way to justify it is by considering the density as a parameter function. Under squared error loss function, the posterior mean is the best estimator for any given value of x_{n+1}. The above equation clearly states that the predictive density is the posterior mean of the sampling density.

This result has important practical consequences. The predictive density gets naturally smoothed out by taking into account all possible threshold values and averaging them with respect to their posterior density, irrespective of the possible jumps associated with each particular value. Smooth densities are obtained as a result and these will be presented in the following sections. Similar situations are encountered in other models that are based on partitions. An example is Gramacy and Lee (2008) where they show that the resulting predictive density virtually smooths out model discontinuities for all practical purposes.

This integration above can not be performed analytically and Monte Carlo approximation may be used instead. This gives $\hat{f}(x_{n+1}|\mathbf{x}) = \frac{1}{J}\sum_{i=1}^{J} f(x_{n+1}|\theta^{(i)}, \mathbf{p}^{(i)}, \Psi^{(i)})$, where $(\theta^{(i)}, \mathbf{p}^{(i)}, \Psi^{(i)})$ is a value sampled from $\pi(\theta, \mathbf{p}, \Psi|\mathbf{x})$, for $i = 1, \ldots, J$. For model $MGPD_k$, $f(x_{n+1} \mid \theta^{(i)}, \mathbf{p}^{(i)}, \Psi^{(i)})$ is given by (5). The expression of the sampling density simplifies to (3) 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 they also allow 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. Simulations based on samples from mixture of Gamma's were also made and showed that mixtures with GPD are not preferred in this case. These findings only reinforce the adequacy of the criteria used.

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. Note that the number of observations in the tail can be quite small despite the large sample sizes since the threshold is always set at a high data quantile. 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 $\pi(\mathbf{p}) \sim D_k(1, \ldots, 1)$. These distribution have mean around the actual parameter value but with large variance to represent lack of information: μ_j's have variance 250 and η_j's have variance 24. Prior distribution for the GPD parameters was given by Jeffreys prior for (ξ, σ) 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 below.

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 Figs. 3 and 4), showing the efficiency of the Gamma mixture model to reproduce the true underlying density. 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 Figs. 2 and specially 3 and 4 for the best fit model in this class according to deviance information criterion (DIC) Spiegelhalter et al. (2002).

Figures 5 and 6 illustrates the effect of large prior variances for the threshold for sample size 1,000 and 10,000. Figure 5 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 6 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.

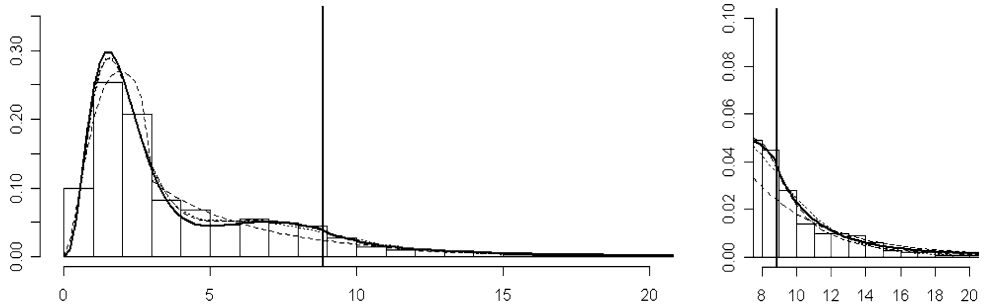

Fig. 2 Predictive density for data simulated with $\xi = 0.4$, $\sigma = 3$, $n = 1{,}000$ and threshold set at 8.85, the 90% data quantile. *Left panel*: full density; *right panel*: detail of density in the tail. *Full line* is true density, *dashed lines* are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and *dotted line* is the predictive density from model MG_3, the best in the MG_k class. The results from $MGPD_k$, $k = 2, 3$ are undistinguishable visually and the $MGPD_1$ provided the worst result. *Vertical line* indicates location of the threshold

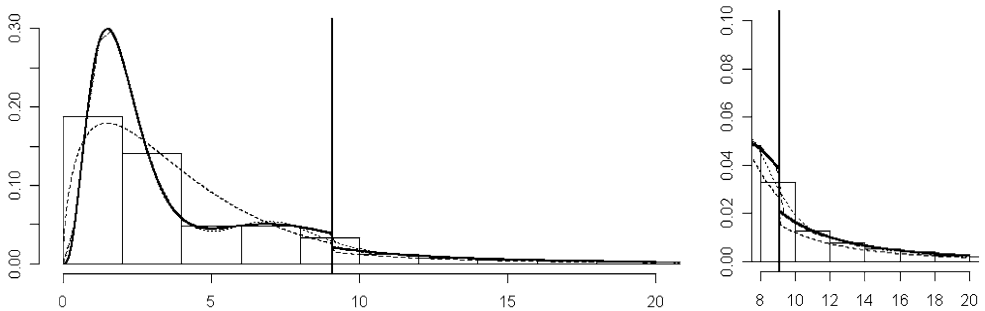

Fig. 3 Predictive density for data simulated with $\xi = 0.4$, $\sigma = 5$, $n = 10{,}000$ and threshold set at 9.08 (the 90% data quantile). *Left panel*: full density; *right panel*: detail of density in the tail. *Full line* is true density, *dashed lines* are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and *dotted line* is the predictive density from model MG_5, the best in the MG_k class. The results from $MGPD_k$, $k = 2, 3$ are undistinguishable visually and the $MGPD_1$ provided the worst result. *Vertical line* indicates location of the threshold

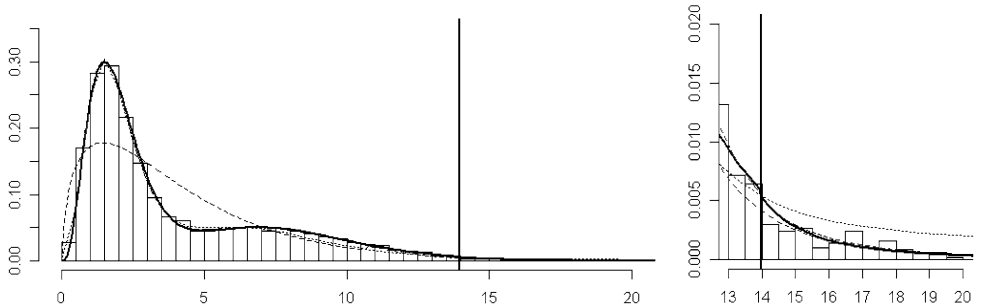

Fig. 4 Predictive density for data simulated with $\xi = 0.4$, $\sigma = 2$, $n = 10{,}000$ and threshold set at 13.97 (the 99% data quantile). *Left panel*: full density; *right panel*: detail of density in the tail. *Full line* is true density, *dashed lines* are predictive densities from $MGPD_k$, $k = 1, 2, 3$ and *dotted line* is the predictive density from model MG_3, the best in the MG_k class. The results from $MGPD_k$, $k = 2, 3$ are undistinguishable visually and the $MGPD_1$ provided the worst result. *Vertical line* indicates location of the threshold

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.

Tables 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.

A good example is the exercise with $n = 1{,}000$, $\sigma = 3$ and $\xi = 0.4$. The effective number of components is basi-

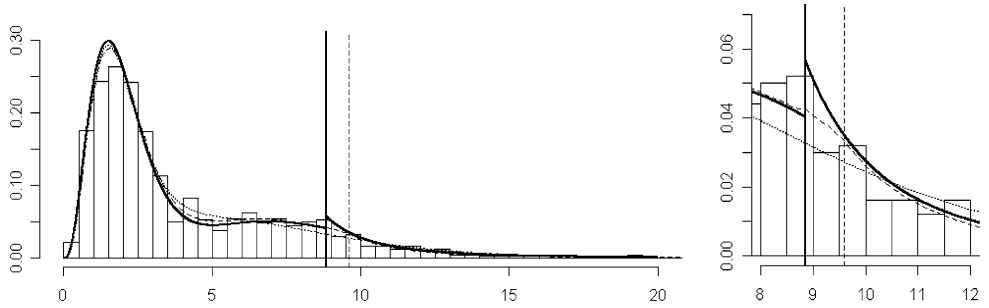

Fig. 5 Predictive density for data (represented in histogram form) simulated with $k = 2$, $\sigma = 2$, $\xi = 0.4$, $u = 8.85$ and $n = 1{,}000$. *Left panel*: full density; *right panel*: density around the true threshold value. The *full line* indicates the true density, the *dashed line* is the 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 $\sigma_u^2 = 10$, and $(14.60, 27.25)$, when $\sigma_u^2 = 10{,}000$

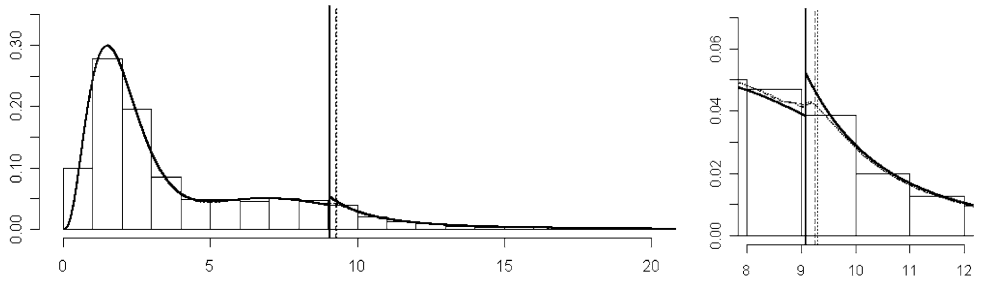

Fig. 6 Predictive density for data (represented in histogram form) simulated with $k = 2$, $\sigma = 2$, $\xi = 0.4$, $u = 9.08$ and $n = 10{,}000$. *Left panel*: full density; *right panel*: density around the true threshold value. The *full line* indicates the true density and, the *dashed line* is 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 $\sigma_u^2 = 10{,}000$

Table 1 Measures of fit for the simulations with threshold set at the 90% data quantile

$\sigma = 2$			$\sigma = 3$			$\sigma = 5$		
k	DIC	BIC	k	DIC	BIC	k	DIC	BIC
			$n = 1{,}000$					
			$\xi = 0.4$					
1	4653.7	4691.5	1	4603.9	4648.1	1	4699.8	4747.6
2	4468.2	**4542.5**	2	**4547.1**	**4624.6**	2	**4660.6**	**4728.1**
3	4469.1	4563.1	3	4547.4	4645.4	3	4662.5	4749.3
3[a]	**4466.6**	4543.9	3[a]	4552.2	4628.4	3[a]	4684.8	4742.3
			$\xi = -0.4$					
1	4388.9	4435.3	1	4462.7	4510.1	1	4754.9	4807.6
2	**4345.5**	4421.7	2	4426.3	4501.9	2	4522.6	*4603.8*
3	4345.8	4442.3	3	**4425.7**	4522.6	3	**4522.1**	4624.6
2[a]	4366.1	**4418.1**	2[a]	4440.9	**4493.1**	2[a]	4556.0	4607.9
			$n = 10{,}000$					
			$\xi = 0.4$					
1	46817	46857	1	47610	47665	1	48621	48691
2	**44718**	**44814**	2	**45530**	**45621**	2	**46548**	**46644**
3	44727	44842	3	45534	45649	3	46548	46671
4[a]	44734	44859	3[a]	45560	45656	5[a]	46594	46753
			$\xi = -0.4$					
1	45372	45440	1	44416	44467	1	47215	47275
2	**43171**	**43272**	2	**43986**	**44083**	2	**45008**	**45106**
3	43172	43300	3	43997	44112	3	45009	45133
5[a]	43211	43368	2[a]	44064	44129	4[a]	45057	45181

[a]Indicates the best MG_k model, i.e., with smallest DIC value

Table 2 Measures of fit for the simulations with threshold set at the 99% data quantile

$\sigma = 2$			$\sigma = 3$			$\sigma = 5$		
k	DIC	BIC	k	DIC	BIC	k	DIC	BIC
			$n = 10{,}000$					
			$\xi = 0.4$					
1	46656	46718	1	46771	46813	1	46906	46967
2	**44623**	**44708**	2	44696	**44791**	2	44799	**44893**
3	44623	44736	3	**44695**	44819	3	**44798**	44921
4[a]	47116	47244	3[a]	44731	44824	4[a]	44855	44978
			$\xi = -0.4$					
1	44535	44592	1	47610	47665	1	46415	46469
2	**44104**	**44198**	2	**44187**	**44279**	2	**44288**	**44381**
3	44105	44226	3	44193	44307	3	44289	44409
5[a]	44122	44272	3[a]	44204	44298	5[a]	44307	44462

[a]Indicates the best MG_k model, i.e., with smallest DIC value

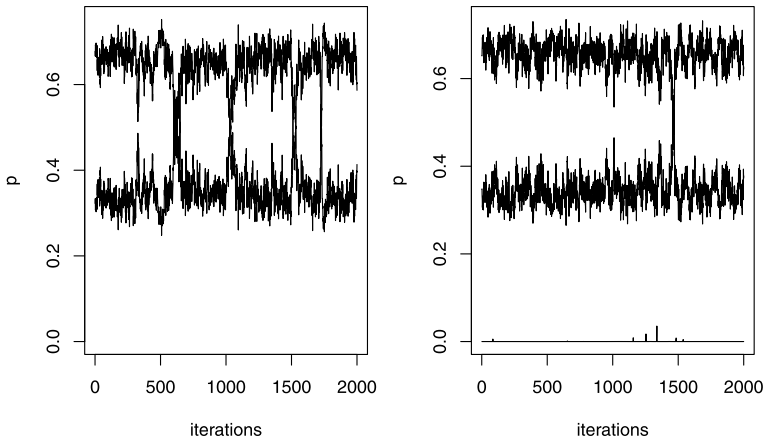

Fig. 7 Trace plots of weights **p** when $n = 1{,}000$, $\sigma = 3$ and $\xi = 0.4$: *left*—weights from model $MGPD_2$; *right*—weights from model $MGPD_3$

cally the same for $k = 2$ and $k = 3$. Figure 7 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.

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.

The validity of the BIC and DIC criteria for detecting the correct model in these settings was tested further with data generated from a mixture of Gamma's. The DIC correctly pointed at the Gamma mixture as the best model when $n = 1{,}000$. The BIC provided very similar values for the correct MG_2 model and the incorrect $MGPD_2$. The latter estimates the threshold beyond the largest observed value indicating the in practice it also pointed at the MG_2 as the best model description. Both criteria agree to indicate the correct Gamma mixture model when $n = 10{,}000$. Thus, irrespective of the sample size, the BIC and DIC criteria are capable of indicating the correct model in our simulation exercise.

3.1 Extreme quantile estimation

One of the main interests in extreme data analysis is the correct identification of extreme points or higher quantiles. These can also estimated for all models considered and also compared against the true quantiles and those estimated by other procedures. They are generally well estimated with MGPD, as illustrated in Fig. 8, indicating that the model can also be used to accurately estimate high quantiles.

There are a few existing techniques that were developed for quantile estimation with the help of graphical tools for threshold determination. They are usually referred to as peaks over threshold (POT) methods and are very popular in practice. Some of the most used techniques are the mean excess plot (MEP) (Davison and Smith 1990) and the dispersion index plots (DIP) (Cunnane 1979). They are easy to use but the accurate determination of the threshold is not an easy task.

Table 3 illustrate this issue with some numerical results. POT estimates rely on appropriate determination of the threshold. When this task is not performed adequately, they fail to provide sound estimates. The approach of this

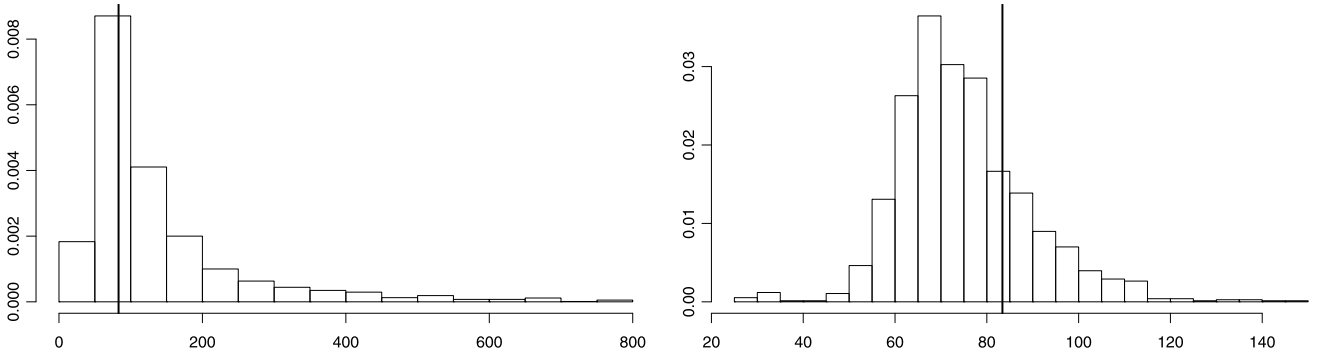

Fig. 8 Posterior histograms of the 99.99% quantile for simulations with $\xi = 0.4$, $\sigma = 2$ and threshold at the 90% quantile ($u = 8.85$). *Left panel*: $n = 1{,}000$; *right panel*: $n = 10{,}000$. The *vertical line* indicates the true quantile value

Table 3 Higher quantile estimation for simulated data with $n = 1{,}000$, $\xi = 0.4$, $\sigma = 3$ and varying thresholds

MGPD—posterior mean quantile based on *MGPD*, POT—POT estimated quantile

	$u = 6$			$u = 9$			$u = 12$		
Quantile	True	*MGPD*	POT	True	*MGPD*	POT	True	*MGPD*	POT
0.99	20.06	23.13	22.07	21.56	20.48	20.21	17.55	17.77	17.11
0.999	65.21	53.19	42.68	51.49	41.44	38.06	37.30	31.59	28.54
0.99999	419.44	314.54	130.58	319.43	191.20	116.41	211.45	319.09	72.86

paper is model-based and therefore more involved but provides more reliable estimates. Our simulations seem to indicate that the *MGPD* estimates outperform the POT estimates in situations where graphical determination of the threshold is difficult but also provide comparable estimates when graphical determination is accurate.

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 Gammas 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. The results also show that use of the information below the threshold is beneficial for correct threshold specification. 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 or for ad-hoc procedures for its determination. 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 versions of those in Wiper et al. (2001) and Behrens et al. (2004) are also carried out.

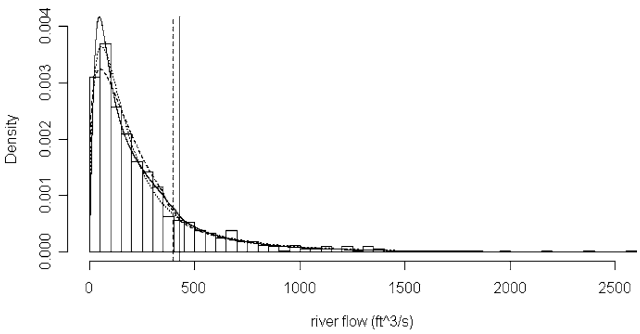

Fig. 9 Predictive densities for data from Espiritu Santo river: *full line—$MGPD_3$*; *dashed line—$MGPD_1$*, *dotted line—MG_3*, *vertical lines*: respective posterior means of the threshold

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 Espiritu Santo. The data was recorded daily from April 1967 to September 2002 and is freely available from waterdata.usgs.gov. We analysed a total of 864 fortnightly maxima data.

Figure 9 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.

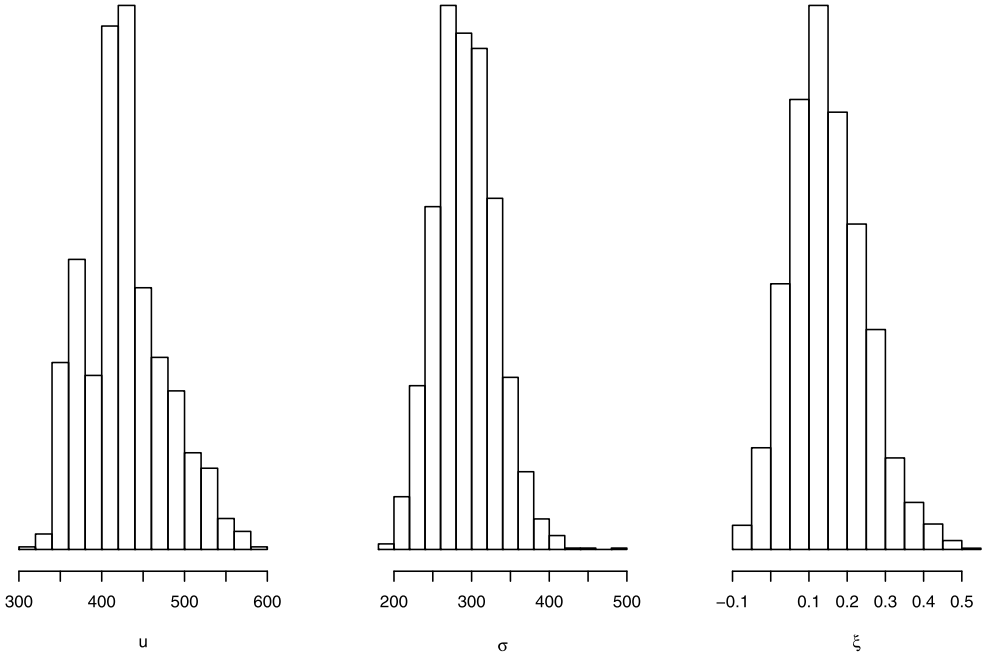

Fig. 10 Posterior histogram for the GPD parameters of the analysis of the Espiritu Santo river flow. From *left* to *right*: u, σ and ξ

Table 4 Fit measures for the applications

k	Espiritu Santo			Fajardo			Barcelos			Grandola		
	Pd	DIC	BIC	Pd	DIC	BIC	Pd	DIC	BIC	Pd	DIC	BIC
						MG_k						
1	2.12	11330	11353	2.01	11441	11463	2.02	8149	8172	1.98	6676	6701
2	3.87	11275	**11322**	4.65	11310	**11355**	3.94	7951	8001	4.21	6391	6442
3	3.87	11275	11342	4.53	11307	11371	4.50	7931	8003	4.18	6391	6462
4	4.68	11273	11362	9.31	11293	11391	3.88	7933	8022	4.26	6391	6483
5	11.04	11288	11410	7.22	11265	11382	4.55	7868	7980	4.05	6391	6503
						$MGPD_k$						
1	0.83	11299	11336	1.05	11327	11361	3.41	8143	8184.1	0.20	4491	4522
2	4.11	11264	11329	0.53	11327	11380	5.76	7639	7712	5.31	4325	**4397**
3	6.71	**11255**	11349	6.28	**11264**	11357	7.08	**7612**	**7709**	7.01	4304	4400
4	5.32	11265	11377	6.89	11268	11384	6.63	7614	7729	6.42	**4303**	4417

Figure 10 provides a summary of parameter estimation for the analysis of the Espiritu Santo river flow. There is strong evidence of an infinite tail based on the large probability associated with positive values of ξ. The posterior mean threshold is 426.82.

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.

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 Grandola, in the South. These two stations 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. It is freely available from www.snirh.pt. We analysed a total of 918 fortnightly maxima data points for Barcelos station and 925 fortnightly maxima data points for Grandola station due to a large number of missing values in the complete dataset.

Figure 11 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.

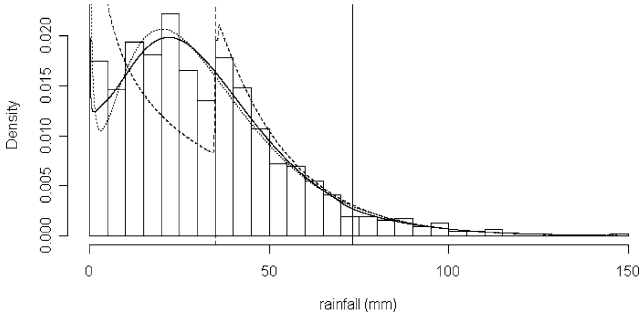

Fig. 11 Predictive densities for data from Barcelos station: *full line—MGPD*$_3$; *dashed line—MGPD*$_1$, *dotted line—MG*$_3$, *vertical lines*: respective posterior means of the threshold

Table 5 Higher quantiles for river Espiritu Santo and Barcelos station

Prob	Espiritu Santo (in t^3/s)				
	E	1	2	3	MG_k
0.95	798	793.29	813.52	807.54	842.7
0.99	1360	1426.04	1450.79	1443.85	1398.8
0.999	N/A	2677.56	2726.55	2718.29	2197.0
0.9999	N/A	4612.30	4734.16	4710.35	3014.0
0.99999	N/A	7844.89	8151.99	8032.95	3804.0
	Barcelos (in mm)				
0.95	73.5	74.54	77.54	73.29	74.71
0.99	99.4	101.73	105.38	102.24	104.09
0.999	N/A	137.84	139.91	137.77	151.50
0.9999	N/A	171.41	176.12	184.54	233.00
0.99999	N/A	203.13	236.24	454.26	333.00

$Prob = P(X \leq q)$, E = Empirical, $1 = MGPD_1$, $2 = MGPD_2$, $3 = MGPD_3$, MG_k refers to the best model in this class

Table 4 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.

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 for river Espiritu Santo and for station Barcelos in Table 5.

According to our preferred model, Espiritu Santo river flows above 2,718 ft^3/s occur on average with 0.1% probability or around once every 10 years. Figure 12 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 (2001). 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.

Similarly, rainfall levels around Barcelos station that are above 159 mm occur on average with 0.01% probability or around once each four centuries. Figure 12 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.

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 Espiritu 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 practitioners 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.

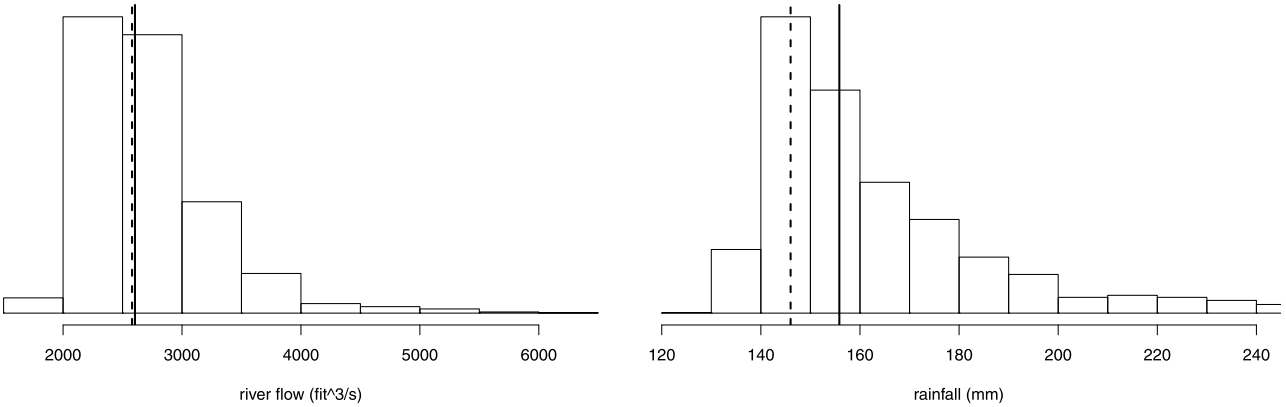

Fig. 12 Posterior histogram of high quantiles of applications, under $MGPD_3$. *Left*: Espiritu Santo river, quantile 99.9%. *Right*: Barcelos station, quantile 99.99%. *Vertical lines*: *full*—posterior mean; *dashed*—maximum observed data

These models may be improved and extended in a number of directions. Cabras et al. (2011) 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 (Nascimento et al. 2011). Another extension is the consideration of temporal data dependence. Initial efforts in this direction were proposed by Lopes et al. (2011). Further developments in this direction to include trend and seasonality can provide useful additions to the understanding of these processes.

Acknowledgements The authors have benefited from discussions with Professors Anthony Davison, Fabrizio Ruggeri and Abel Rodriguez and from many comments made by the referees. The work of F.F. Nascimento and D. Gamerman was supported by Brazilian research supporting agencies Capes, CNPq and Faperj. The work of H.F. Lopes was supported by the University of Chicago Booth School of Business.

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 ξ: Proposal transition kernel for ξ is given by a truncated normal

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

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

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

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

Sampling σ:

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

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

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

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

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

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

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

Sampling u:

The threshold u^* is sampled from a $N(u^{(s)}, V_u) \times I(a^{(s+1)}, \infty)$ 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 (1). V_u is the variance chosen to ensure appropriate chain mixing. Then, accept $u^{(s+1)} = u^*$ with probability α_u, where

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

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

Sampling (μ_j, η_j): For $j = 1, \ldots, k$.

Since η_j is positive, the proposal kernel is taken as the Gamma distribution

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

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

Since μ_j is also positive, the proposal kernel is taken as the Gamma distribution

$$\mu_j^*|\mu_j^{(s)} \sim G\left(\mu_j^{(s)}, \frac{\mu_j^{(s)^2}}{V_{\mu_j}}\right) I\big(\mu_1^{(s+1)} < \cdots < \mu_{j-1}^{(s+1)} < \mu_j^{(s)} < \cdots < \mu_k^{(s)}\big),$$

where $\mu_j^{(s)}$ is the value of μ_j at iteration s and V_{μ_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 α_{μ_j,η_j}, where

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

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

Sampling $\mathbf{p}$:

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

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

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

References

Asmussen, S.: Applied Probability and Queues. Wiley, New York (1987)

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

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

Cabras, S., Castellanos, M.A., Gamerman, D.: A default Bayesian approach for regression on extremes. Stat. Model. (2011, accepted)

Castellanos, M.A., Cabras, S.: A default Bayesian procedure for the generalized Pareto distribution. J. Stat. Plan. Inference **137**, 473–483 (2007)

Coles, S.G.: Extreme Value Theory an Applications. Kluver Academic, Dordrecht (2001)

Coles, S.G., Tawn, J.A.: A Bayesian analysis of extreme rainfall data. Appl. Stat. **45**, 463–478 (1996)

Cunnane, C.: Note on the Poisson assumption in partial duration series model. Water Resour. Res. **15**, 489–494 (1979)

Dalal, S., Hall, W.: Approximating priors by mixtures of natural conjugate priors. J. R. Stat. Soc., Ser. B **45**, 278–286 (1983)

Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds (with discussion). J. R. Stat. Soc., Ser. B **52**, 393–342 (1990)

Dey, D., Kuo, L., Sahu, S.: A Bayesian predictive approach to determining the number of components in a mixture distribution. Stat. Comput. **5**, 297–305 (1995)

Diebolt, J., Robert, C.: Estimation of finite mixture distributions by Bayesian sampling. J. R. Stat. Soc., Ser. B **56**, 363–375 (1994)

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

Doornik, JA: Ox: Object Oriented Matrix Programming, 4.1 console version. Nuffield College, Oxford University, London (1996)

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

Fisher, R.A., Tippett, L.H.C.: On the estimation of the frequency distributions of the largest and smallest number of a sample. Proc. Camb. Philos. Soc. **24**, 180–190 (1928)

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

Frühwirth-Schnatter, S.: Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. J. Am. Stat. Assoc. **96**, 194–209 (2001)

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

Gramacy, R., Lee, K.: Bayesian treed Gaussian process models with an application to computer modeling. J. Am. Stat. Assoc. **103**, 1119–1130 (2008)

Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological events. Q. J. R. Meteorol. Soc. **81**, 158–171 (1955)

Lopes, H.F., Nascimento, F.F., Gamerman, D.: Generalized Pareto models with time-varying tail behavior. Technical Report LES:UFRJ, in preparation (2011)

von Mises, R.: La distribution de la plus grande de nvaleurs. Am. Math. Soc. **2**, 271–294 (1954)

Nascimento, F.F., Gamerman, D., Lopes, H.F.: Regression models for exceedance data via the full likelihood. Environ. Ecol. Stat. (2011, to appear)

Pickands, J.: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)

Richardson, S., Green, P.: On Bayesian analysis of mixtures with an unknown number of components. J. R. Stat. Soc., Ser. B **59**, 731–792 (1997)

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

Roeder, K., Wasserman, L.: Practical Bayesian density estimation using mixtures of normals. J. Am. Stat. Assoc. **92**, 894–902 (1997)

Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)

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

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.: Bayesian measures of model complexity and fit. J. R. Stat. Soc. B **64**, 583–639 (2002)

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

Titterington, D., Smith, A.F.M., Makov, U.: Statistical Analysis of Finite Mixture Distributions. Wiley, New York (1985)

Wiper, M., Rios Insua, D., Ruggeri, F.: Mixtures of gamma distributions with applications. J. Comput. Graph. Stat. **10**, 440–454 (2001)