A Semiparametric Bayesian Extreme Value Model Using a Dirichlet Process Mixture of Gamma Densities

Jairo A. Fuquene P. *

May 8, 2014

Abstract

In this paper we propose a model with a Dirichlet process mixture of gamma densities in the bulk part below threshold and a generalized Pareto density in the tail for extreme value estimation. The proposed model is simple and flexible for posterior density estimation and posterior inference for high quantiles. The model works well even for small sample sizes and in the absence of prior information. We evaluate the performance of the proposed model through a simulation study. Finally, the proposed model is applied to a real environmental data.

keywords Generalized Pareto Distribution, Threshold Estimation, Dirichlet Process Mixture.

1 Introduction

In recent years, extreme value mixture models have been proposed as a combination of a distribution with a "bulk part" below threshold and a generalized Pareto distribution (GPD) in the tail. Different distributions have been proposed for modelling the "bulk part" where the threshold is a parameter to be estimated. The first approach which induces a transition between the bulk and tail parts is provided by Frigessi, Haug & Rue (2003). Frigessi et al. (2003) uses maximum likelihood estimation with a Weibull distribution in the bulk part, a GPD for the tail and a location-scale Cauchy cdf in the transition function. However, in the Frigessi et al. (2003) approach, to use maximum likelihood estimation in the bulk part could produce multiple modes and hence some identifiability problems. Behrens, Lopes & Gamerman (2004) and Carreau & Bengio (2009) consider Gamma and Normal distributions in the bulk part respectively. But, to

*Department of Statistics. University of Warwick. UK. J.A.Fuquene-Patino@warwick.ac.uk

consider an unimodal distribution is not realistic in practice where the density has different unknown shapes in many applications. do Nascimento, Gamerman & Lopes (2011) use Bayesian inference in the bulk part following the proposal of Wiper, Insua & Ruggeri (2001) who propose to assign prior probabilities on the number of components of the mixture of gammas and to use the reversible jump algorithm for posterior inference purposes. Wiper et al. (2001) use BIC and DIC criterion for model comparison on a fixed number of gamma components. This approach has a flexible model with multimodality in the bulk part distribution. A complete review on likelihood-based methods and heuristic arguments for model complexity in mixture models such as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) is presented in Frühwirth-Schnatter (2006). On the other hand, do Nascimento et al. (2011) show that by using posterior predictive inference the discontinuity problem at the threshold is eliminated. MacDonald, Scarrott, Lee, Darlow, Reale & Rusell (2011) propose a non-parametric approach in the bulk part with kernel bandwidth estimators and a GPD in the tail where Bayesian inference is applied. For a more exhaustive discussion of extreme value threshold estimation see for example Scarrott & MacDonald (2012). On the other hand, there is an extensive literature on Dirichlet mixture process for density estimation particulary using gaussian distributions, the main paper is given by Escobar & West (1995). The Dirichlet process is very flexible, theoretically coherent and simple and in recent years it has been an important tool of many applications for Bayesian density estimation (Ferguson (1973) and Antoniak (1974)). Hanson (2006) proposes the Dirichlet process mixture of gamma densities (DPMG) for density estimation of univariate densities on the positive real line.

In this paper we propose a model with a DPMG below threshold and a GPD in the tail. We have important reasons for using the proposed model: First, because DPMG could be a powerful tool for density estimation in the bulk part (in order to accommodate a very wide variety of shapes and spreads in the bulk part), the tail fit is expected to be adequate. Second, the proposed model can be used in the absence of prior information. Third, Dirichlet Process Mixture controls the expected number of components (Antoniak (1974)); therefore, the extensive task for model comparison purposes using BIC and AIC on a fixed number of gamma components in the bulk part is not necessary. In addition, because DPMG is random we can build credible intervals of the posterior density in the bulk part. This paper is organized as follows. Section 2 is devoted to present the proposed model. In Section 3 we present a simulation study of the proposed model. In Section 4 the proposed model is applied to a real environmental data. Finally Section 5 present the conclusions.

2 Model

Extreme value theory is an area of far researching with many environmental and economics applications. We find among the most notable applications Coles & Pericchi (2003) where extreme value theory is used to anticipate catastrophes and Coles (2001) for finance applications. We consider the class of models from extreme value theory presented in Coles (2001) and the fundamental result in Pickands (1975). Pickands (1975) showed that the limiting distribution of exceeding large thresholds converge to the GPD. Extreme value theory is used to describe atypical situations and the most important classical result is presented in the Fisher & Tippett (1928) theorem. The three possible distributions for maxima block of observations are presented in Coles (2001) in Theorem 3.1 which introduces the Generalized Extreme Value (GEV) Family. Pickands (1975) showed 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 $u \to x_F$, the conditional distribution function $F(x|u) = P(X \leq u + x|X > u)$ is the distribution function of a GPD. In words, this result states that for a large u the conditional distribution $F(x|u)$ can be approximated by a GPD as u tends to the end point of F. Moreover, we use this result to define the proposed model. The density of the GPD used in this work with scale parameter σ and shape parameter γ is given by:

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

where the vector of parameters $\boldsymbol{\phi} = (\xi, \sigma, u)$, $x - u > 0$ for $\xi \geq 0$ and $0 \leq x - u < -\sigma/\xi$ for $\xi < 0$. We have that GDP is bounded from below by u, bounded from above by $u - \sigma/\xi$ if $\xi < 0$ and unbounded from above if $\xi \geq 0$. The density of the proposed model is the following:

$$f(x|\boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{cases} k(x|\boldsymbol{\theta}) & x \leq u \\ [1 - K(u|\boldsymbol{\theta})]g(x|\boldsymbol{\phi}) & x > u \end{cases} \tag{2}$$

where $\boldsymbol{\phi} = (u, \xi, \lambda)$ and $K(u|\boldsymbol{\theta})$ denotes the cumulative distribution function (cdf) of $k(x|\boldsymbol{\theta})$ at u. The cumulative distribution function of (2) is as follows:

$$F(x|\boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{cases} K(x|\boldsymbol{\theta}) & x \leq u \\ K(u|\boldsymbol{\theta}) + [1 - K(u|\boldsymbol{\theta})]G(x|\boldsymbol{\phi}) & x > u \end{cases} \tag{3}$$

where $G(x|\boldsymbol{\phi})$ is the cdf of GPD. Note that

$$\lim_{x \to u^-} F(x|\boldsymbol{\phi}, \boldsymbol{\theta}) = K(u|\boldsymbol{\theta}); \qquad \lim_{x \to u^+} F(x|\boldsymbol{\phi}, \boldsymbol{\theta}) = K(u|\boldsymbol{\theta}), \tag{4}$$

therefore (3) is continuous at u.

2.1 The Dirichlet Process Mixture of Gamma densities

The novel proposal is to use a DPMG in the bulk part of (2), we present a short introduction to the DP here. A distribution G on Θ follows a dirichlet process $DP(\alpha, G_0)$ if, given an arbitrary measurable partition, $B_1, B_2, ..., B_k$ of Θ the joint distribution of $(G(B_1), G(B_2), ..., G(B_k))$ is Dirichlet $(\alpha G_0(B_1), \alpha G_0(B_2), ..., \alpha G_0(B_k))$ where $G(B_i)$ and $G_0(B_i)$ denote the probability of set (B_i) under G and G_0 respectively, G_0 is a specific distribution on Θ and α is a precision parameter (Ferguson (1973)). Here $\boldsymbol{\theta} = \{\lambda, \gamma\}$ and we use the approach of Hanson (2006) for the density g_0 therefore two independent exponential distributions are considered as follows

$$g_0(\lambda, \gamma | a_\lambda, a_\gamma) = a_\lambda \exp(-a_\lambda \lambda) a_\gamma \exp(-a_\gamma \gamma), \tag{5}$$

where g_0 is the density corresponding to cdf G_0 with hyperparameters $\eta = \{a_\lambda, a_\gamma\}$. The hyperparameters of (4) follow gamma priors $a_\lambda \sim \Gamma(b_\lambda, c_\lambda)$ and $a_\gamma \sim \Gamma(b_\gamma, c_\gamma)$, where $\Gamma(a, b)$ denotes the gamma density with parameters a and b. Let be now $K(;, \boldsymbol{\theta})$ be a parameter family of distributions functions (CDF's) indexed by $\boldsymbol{\theta} \in \Theta$, with associated densities $k(; \boldsymbol{\theta})$. Let be $x_1, x_2, ..., x_n$ the data and $\boldsymbol{\theta}_i = (\lambda_i, \gamma_i)$ such that $k(x_i, \theta_i)$ denotes the gamma density with the scale parameter λ_i and the shape parameter γ_i:

$$k(x_i | \lambda_i, \gamma_i) = \frac{\gamma_i^{\lambda_i}}{\Gamma(\gamma_i)} x_i^{\lambda_i - 1} \exp\{-\gamma_i x_i\} \quad x_i > 0. \tag{6}$$

Because G is proper we can define the mixture distribution

$$F(.; G) = \int K(;, \boldsymbol{\theta}) G(d\boldsymbol{\theta}) \tag{7}$$

where $G(d\boldsymbol{\theta})$ can be interpreted as the conditional distribution of $\boldsymbol{\theta}$ given G. We can express (6) as $f(.; G) = \int k(.; G)$ differentiating with respect to (.). Due to G being random, $F(.; G)$ is random. $F(.; G)$ is the model for the stochastic mechanism corresponding to $x_1, x_2, ..., x_n$ assuming x_i given G are i.i.d. from $F(.; G)$ with the DP structure. In this paper we implement the Dirichlet Process Mixture model by using the Pólya urn scheme (see Escobar & West (1995) and MacEachern (1994)). In DPMG we have mixing parameters $\boldsymbol{\theta}_i = (\lambda_i, \gamma_i)$ associated with each x_i. The model can be expressed in hierarchical form as follows:

$$\begin{aligned} x_i | \lambda_i, \gamma_i &\sim k(x_i, \boldsymbol{\theta_i}), \quad i = 1, .., n \\ \boldsymbol{\theta_i} | G &\sim G, \quad i = 1, .., n \\ G | \alpha, \eta &\sim DP(\alpha, G_0), G_0 = G_0(.|\eta) \\ \alpha, \eta &\sim p(\alpha) p(\eta) \end{aligned} \tag{8}$$

2.2 Priors for the parameters in the generalized Pareto distribution

Now we present the priors for the threshold u, the scale parameter σ and shape parameter ξ of the GPD. The prior distribution for u is a normal density $N(m_u, \sigma_u^2)$ as suggested in Behrens et al. (2004). Castellanos & Cabras (2007) obtain the Jeffrey's non-informative prior for (σ, ξ) and they show this prior produces proper posterior results. The prior is the following:

$$p(\sigma, \xi) \propto \sigma^{-1}(1+\xi)^{-1}(1+2\xi)^{-1/2} \tag{9}$$

where $\xi > -0.5$ and $\sigma > 0$. According to Castellanos & Cabras (1996) situations were $\xi < -0.5$ are very unusual in practice. The posterior distribution on the log-scale using the density (2) is then:

$$\begin{aligned}\log(p(\boldsymbol{\theta}, \boldsymbol{\phi}|x)) \propto \sum_A \log(k(x|\boldsymbol{\theta})) + \sum_B \log\left((1-K(u|\boldsymbol{\theta}))\frac{1}{\sigma}\left(1+\xi\frac{(x-u)}{\sigma}\right)^{-(1+\xi)/\xi}\right) \\ + \log(p(u)p(\xi)p(\sigma))\end{aligned} \tag{10}$$

for $\xi \neq 0$ and

$$\begin{aligned}\log(p(\boldsymbol{\theta}, \boldsymbol{\phi}|x)) \propto \sum_A \log(k(x|\boldsymbol{\theta})) + \sum_B \log\left((1-K(u|\boldsymbol{\theta}))\frac{1}{\sigma}\exp(-(x-u)/\sigma)\right) \\ + \log(p(u)p(\xi)p(\sigma))\end{aligned} \tag{11}$$

for $\xi = 0$. With $A = \{x_i : x_i \leq u\}$ and $B = \{x_i : x_i > u\}$. Using the proposed model we can compute high quantiles below threshold. In order to find values beyond the threshold we have that

$$F(x|\boldsymbol{\phi}, \lambda, \gamma) = K(u|\lambda, \gamma) + [1 - K(u|\lambda, \gamma)]G(x|\boldsymbol{\phi}) \tag{12}$$

where $G(x|\phi)$ is the CDF of the GPD. For example to find the p-quantile, q, we use

$$p^* = \frac{p - K(u|\lambda, \gamma)}{1 - K(u|\lambda, \gamma)} \tag{13}$$

and solve $G(q|\boldsymbol{\phi}) = p^*$.

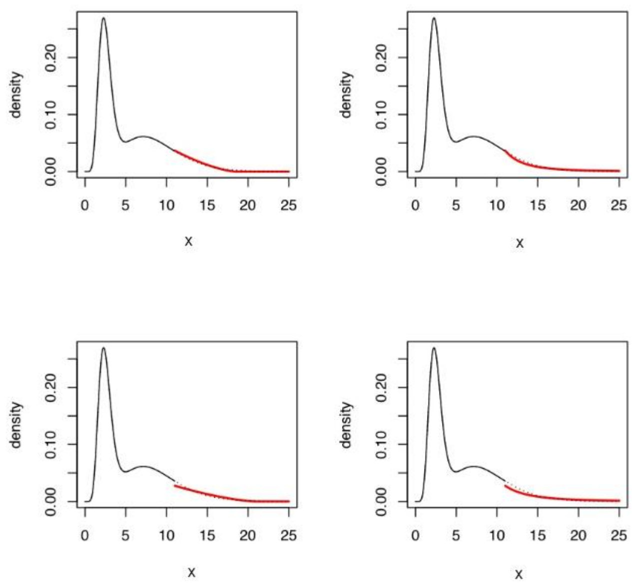

Figure 1: Density function of the proposed model (3). (a) $\xi = -0.4$ and $\sigma = 3$, (b) $\xi = 0.4$ and $\sigma = 3$, (c) $\xi = -0.4$ and $\sigma = 4$ and (d) $\xi = 0.4$ and $\sigma = 4$. Threshold at $u = 11$ and the center is a two-component mixture of gamma densities.

Figure 1. displays the density of the proposed model considering different parameter values. This model has a discontinuity of the density at the threshold. However, both the appropriate choice of the proposed model (in particular the cdf is continuous) and the appropriate Bayesian estimation (see do Nascimento et al. (2011)) solve this problem.

3 Simulation study

In this section we evaluate the performance of the proposed model through a simulation study. The precision α of g_0 in the DP affects the expected number of components in the mixture. Hanson (2006) considers values of α fixed to 0.1 and 1 and also random values using different assignments of Gamma priors for α such as $\Gamma(2, 2)$ and $\Gamma(2, 0.5)$. Here we consider the DP precision using $\alpha = 0.1$. The parameters of g_0 can be expressed in terms of the mean $\mu = \lambda/\gamma$ and variance $V = \lambda/\gamma^2$ of $h(x|\theta)$ (see Hanson (2006)) as two diffuse densities $f(\mu|a_\lambda, a_\gamma) = a_\lambda a_\gamma/(a_\lambda \mu + a_\gamma)^2$ and $f(V^{-1}|a_\lambda, a_\gamma) = \Gamma(2, a_\lambda \mu^2 + a_\gamma \mu)$ respectively. Suppose now that $a_\lambda = a_\gamma = 1$, so $f(\mu|1, 1) = 1/(1 + \mu)^2$ which is the Beta Prime distribution with scale and shape parameters equals to 1.

The Beta prime distribution was proposed by Perez & Pericchi (2012) as a default prior for modelling the scales in Bayesian parametric settings. Also, the Beta prime distribution for modelling the square of the scales in Bayesian dynamic models is extensively studied in Fuquene, Perez & Pericchi (2014). Therefore, we can think that we are modelling the mean of the bulk part in a non informative (but robust) manner. We consider a small sample size $n = 200$ and we verify the convergence using techniques such as correlation plots, traces plots and the usual Gelman & Rubin (1992) diagnostic. Hanson (2006) obtains an accurate smooth in an univariate density using DPMG with different specifications for α and large sample sizes 1000 and 10000. Here, we have that $\alpha = 0.1$, $\xi = 0.4$, $\sigma = 3$ and the threshold $u = 11$ at the 90% quantile in the simulated data. The simulated mixture density for the central part is:

$$h(x) = 0.5\Gamma(x|10, 4) + 0.5\Gamma(x|6, 0.7). \tag{14}$$

Following Hanson (2006) the hyperparameters for a_λ and a_γ are $b_\lambda = b_\gamma = c_\lambda = c_\gamma = 0.001$ in order to have a non informative g_0. The prior of the threshold u has mean equal to 90% quantile in the simulated data and the variance σ_u^2 gives 99% of probability in the range between 50% and 99% of the simulated data. As usual in the Metropolis algorithm, we adjust the variance of the sampling proposal densities considering the hessian of the maximum likelihood estimates using some MCMC simulations. We obtained convergence of all parameters using 10000 iterations after a burn-in period of 5000 iterations. Figure 2 illustrates the quality of the approach even with a small sample size of $n = 200$. The posterior density in the proposed model reproduces the underline density with precision according to the credible interval in the bulk part and posterior predictive mean in the tail. The density estimation in bulk part of the proposed model could be even better when large sample sizes are considered (see Hanson (2006)). Figure 3 displays the posterior densities of threshold u, scale σ, and shape ξ.

Table 1: Measures of fitting using BIC and AIC for the simulated data.

Number of components	AIC	BIC
1	1077.611	1087.426
2	**1035.914**	**1050.637**
3	1043.192	1067.730

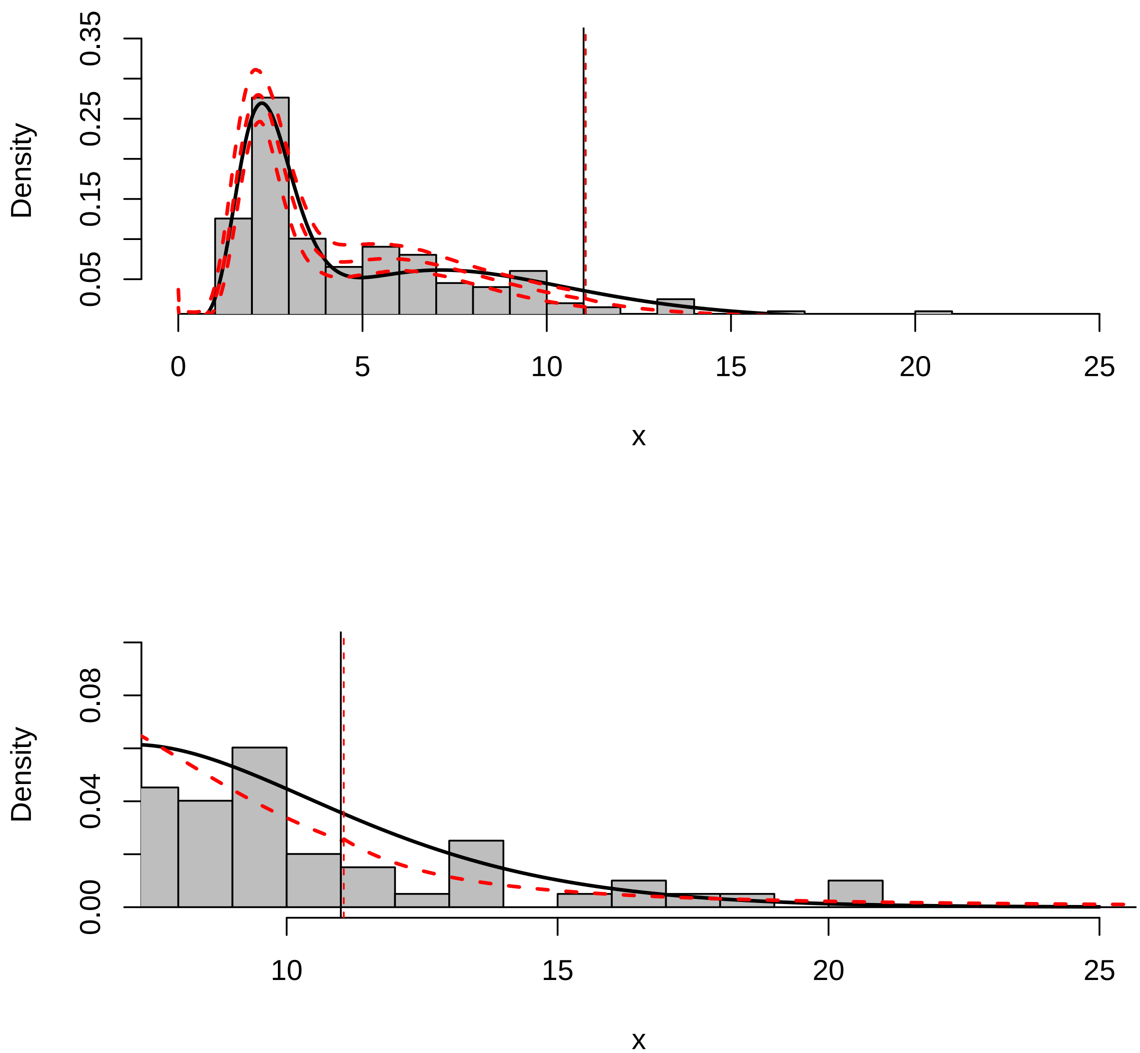

Figure 2: Parameter values $\alpha = 0.1$, $\xi = 0.4$, $\sigma = 3$ and the threshold $u = 11$ at the 90% quantile. Full black line is the true density. The vertical full black line is the true threshold location and the vertical dashed red lines are the posterior threshold location. Top: dashed red lines are the posterior predictive mean and 95% posterior predictive credible intervals using the Dirichlet process mixture of gamma densities in the bulk part and a GPD in the tail. Bottom: dashed red lines are the posterior predictive mean using the Dirichlet process mixture of gamma densities in the bulk part and a GPD in the tail.

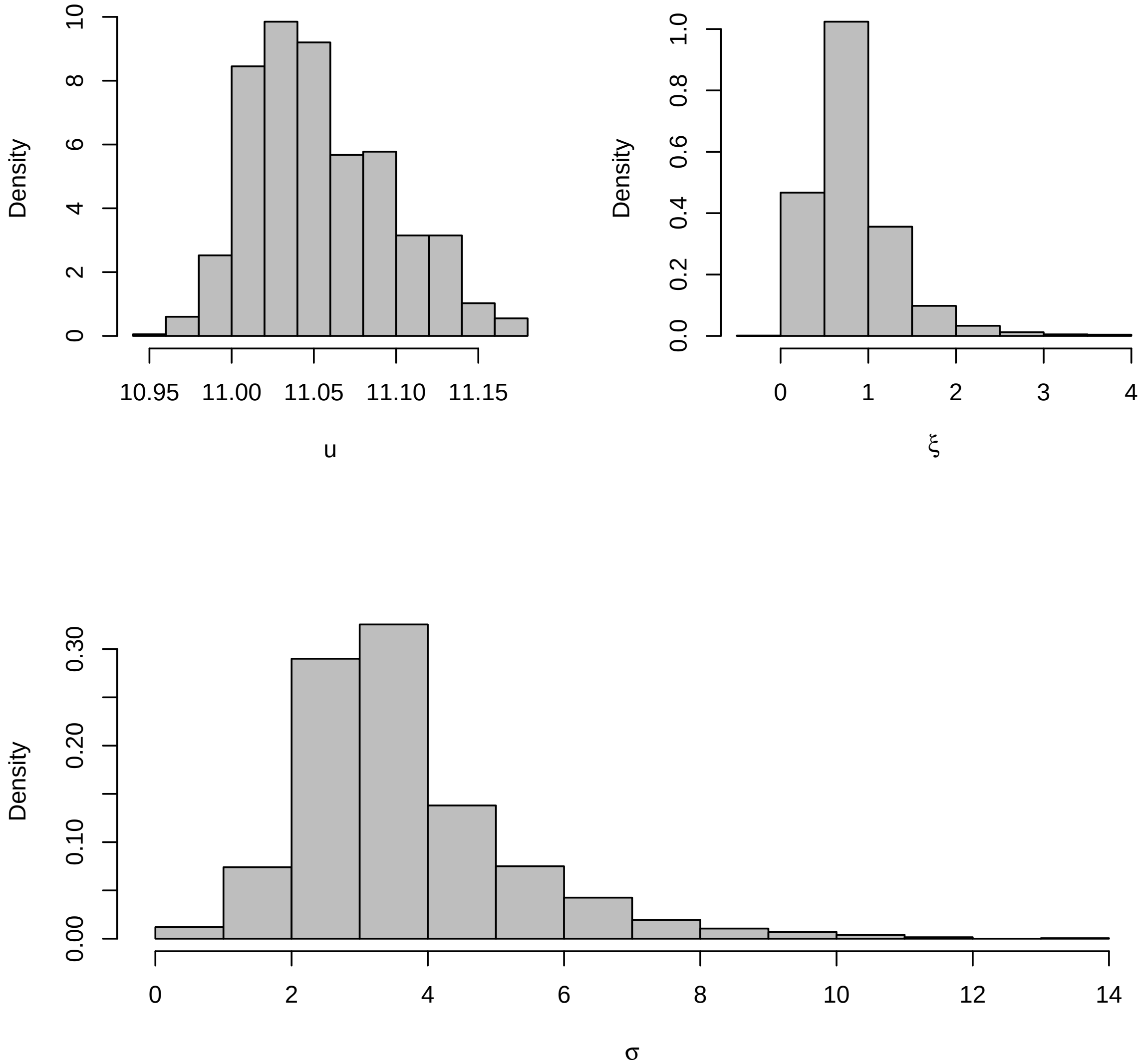

Figure 3: Posterior distribution of u, ξ and σ. With $\alpha = 0.1$, $\xi = 0.4$, $\sigma = 3$ and the threshold $u = 11$ at the 90% quantile.

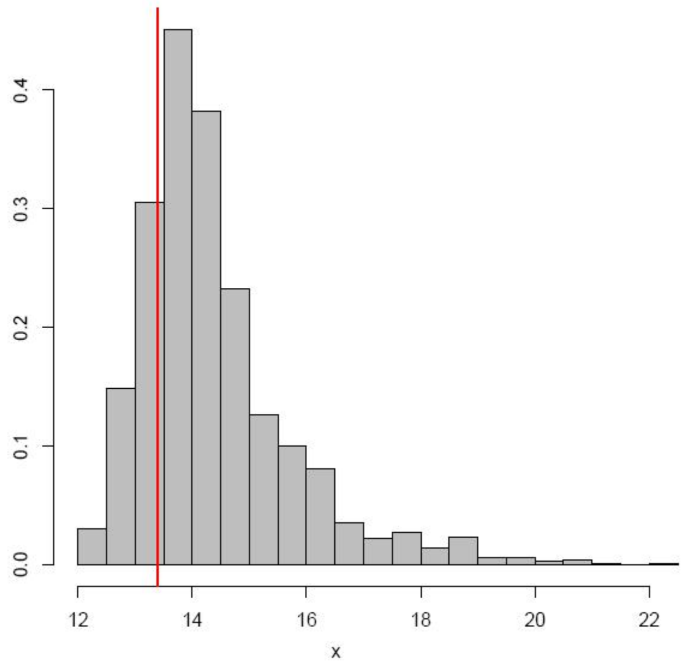

Figure 4: Posterior histogram of the 95% quantile for the simulation. Red line the true quantile. With $\alpha = 0.1$, $\xi = 0.4$, $\sigma = 3$ and the threshold $u = 11$ at the 90% quantile.

We can see the posterior distribution represents nicely the true parameters. In particular the threshold is centered around the true value 11. Figure 4 shows that the posterior distributions of the predictive quantiles at 95% is accurately estimated. On the other hand, Table 1 shows that BIC and AIC criterion suggest two use a model with two components in the bulk part.

4 Application to the flow levels in the Gurabo river

River flow levels are important measures to prevent disasters in populations when flow rate exceeds the capacity of the river channel. We applied the proposed model in river flow levels measured at cubic feet per second (ft^3/s) in Gurabo river at Gurabo Puerto Rico. The data is available at waterdata.usgs.gov. The flows are monitored between December 2 2012, 12:00 am to December 4 2012, 8:45 pm. The measures are made each 15 minutes for a total sample size of n=254. We obtained convergence of all parameters using 5000 iterations after a burn-in period of 2000 iterations. We spent approximately 30 minutes to obtain the results using R Core Team (2014) package and a PC with Intel(R) Xeon(R) 2.80 GHZ and 4 GB RAM.

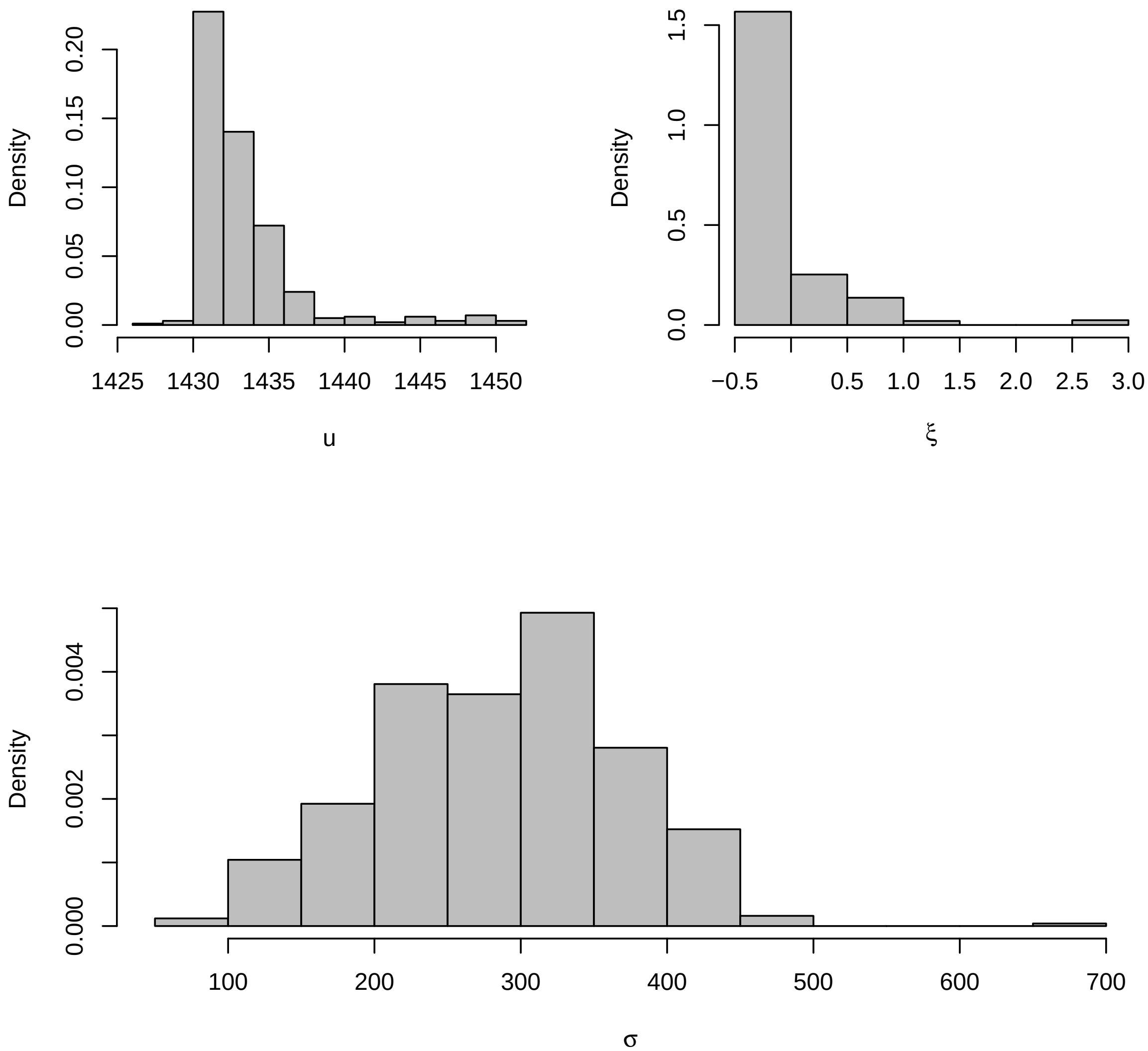

Figure 5: Posterior histogram of the GPD parameters in the tail of the proposed model for the application.

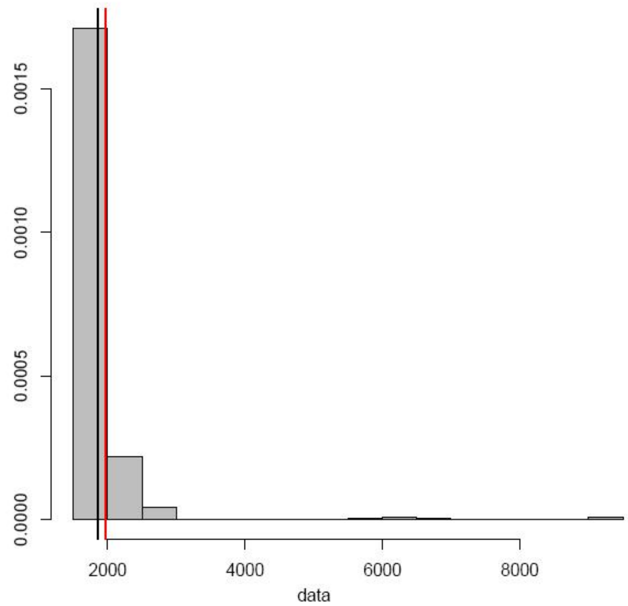

Figure 6: Posterior distribution of the 99.9% quantile for the application. Black line is the maximum observed data and red line is the posterior mean for the 99.9% simulated quantile.

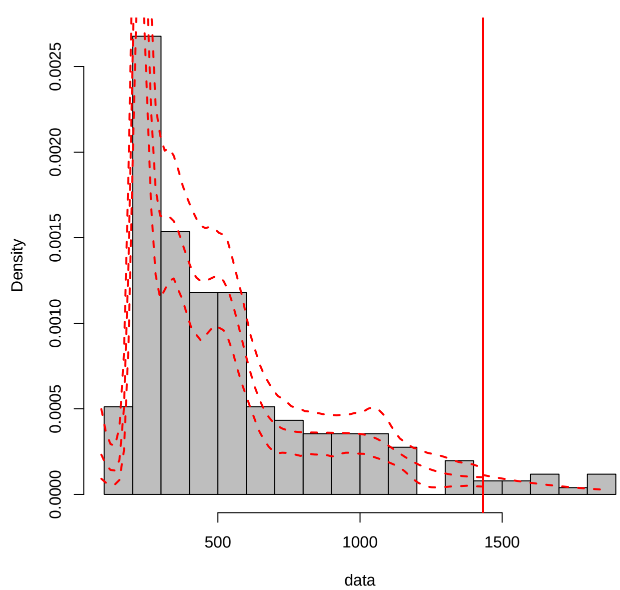

Figure 7: Dashed red lines are the posterior predictive mean and 95% posterior predictive credible intervals using the Dirichlet process mixture of gamma densities in the bulk part and a GPD in the tail. The vertical red line is the posterior threshold location.

Table 2: Measures of fitting using BIC and AIC for the application.

Number of components	AIC	BIC
1	3597.096	3606.911
2	3533.582	3548.305
3	**3501.836**	**3526.374**
4	3505.836	3540.19

On the other hand, Table 2 shows that BIC and AIC criterion suggest to use a model with three components in the bulk part. Figure 5 displays the posterior distributions of the parameters in the tail of the proposed model. The threshold, scale and shape are around the values 1430 (quantile at 96% according to the simulation), 300 and -0.25 respectively. Figure 6 shows the posterior distribution for the 99.9% high quantile, we can see the maximum value is less than the posterior mean for the quantile at 99.9% and the posterior distribution is asymmetric which is expected. The prediction ability in this example with even a small sample size is illustrated in Figure 6. where predictions of high quantiles can be considered. Figure 7 displays the posterior density using DPMG in the bulk part and a GPD in the tail. We can see our proposed model reproduces the data in the bulk and tail parts. As a conclusion according to the posterior analysis and based on the last two days of observations, we can see flow levels over 1998 ft^3/s in the Gurabo River with 0.1% probability.

5 Conclusion

We proposed a model with a Dirichlet process mixture of gamma densities in the bulk part of the distribution and a heavy tailed generalized Pareto distribution in the tail for extreme value estimation. The proposal is very flexible and simple for density estimation in the bulk part and posterior inference in the tail. According to the simulations and application to real data the model works well even for small sample sizes and in the absence of prior information. The Dirichlet Process mixture controls the expected number of components and so the extensive task for model comparison purposes using BIC and AIC on a fixed number of gamma components in the bulk part is not necessary. The proposed model was applied to a real environmental data set but interesting applications can be found in different areas such as clinical trials or finance.

References

Antoniak, C. E. (1974), 'Mixture of Dirichlet process with applications to Bayesian Nonparametric problems', *Annals of Statistics* **2**, 1152–1174.

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

Carreau, J. & Bengio, Y. (2009), 'A hibrid Pareto mixture for conditional asymmetric fat-tailed data', *Extremes* **12**, 53–76.

Castellanos, E. & Cabras, S. (1996), 'A Bayesian analysis of extreme rainfall data', *Applied Statistics* **45**, 463–478.

Castellanos, E. & Cabras, S. (2007), 'A default Bayesian procedure for the generalized Pareto distribution', *Journal of Statistical Planing and Inference* **137**, 473–483.

Coles, S. (2001), *An introduction to Statistical Modelling of Extreme Values*, London, Springer.

Coles, S. & Pericchi, L. (2003), 'Anticipating Catastrophes through Extreme Value Modelling', *Journal of the Royal Statistical Society* **Series C (Applied Statistics)**, 405–416.

do Nascimento, F., Gamerman, D. & Lopes, H. F. (2011), 'A semiparametric Bayesian approach to extreme value estimation', *Statist. Comput.* .

Escobar, M. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. (1973), 'A Bayesian analysis of some nonparametric problems', *Annals of Statistics* **1**, 209–230.

Fisher, R. A. & Tippett, L. C. H. (1928), 'On the estimation on the frequency distribution of the largest and smaller number of a sample', *Proceedings Cambridge Philosophy* **24**, 180–190.

Frigessi, A., Haug, O. & Rue, H. (2003), 'A dynamic mixture model for unsupervised tail estimation without threshold estimation', *Extremes* **5**, 219–235.

Frühwirth-Schnatter, S. (2006), *Finite mixtures and markov switching models*, Springer, Ney York, USA.

Fuquene, J., Perez, M. E. & Pericchi, L. R. (2014), 'An alternative to the inverted Gamma for the variance to modelling outliers and structural breaks in dynamic models', *Brazilian Journal of Probability and Statistics* **In press**, 1–14.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulation using multiples sequences', *Statistical Science* **7**, 457–511.

Hanson, T. E. (2006), 'Modelling censoring lifetime data using a mixture of Gamma baselines', *Bayesian Analysis* **3**, 575–594.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. & Rusell, G. (2011), 'A flexible extreme value mixture model', *Computational statistics and data analysis* **55**, 2137–2157.

MacEachern, S. (1994), 'Estimating normal means with a conjugate style Dirichlet process prior', *Communications in statistics* **23(460)**, 727–741.

Perez, M. E. & Pericchi, L. R. (2012), 'The scaled Beta2 distribution as a robust prior for scales, and a explicit horseshoe prior for locations', *Center for Biostatistics and Bioinformatics* **Technical report**, 1–10.

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

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Scarrott, C. & MacDonald, A. (2012), 'A review of extreme value threshold estimation and uncertainty quantification', *Statistical Journal* **10**, 33–60.

Wiper, M., Insua, D. R. & Ruggeri, F. (2001), 'Mixture of Gamma distributions with applications', *Journal of the American statistical association* **10**, 440–454.

A MCMC algorithm

1. For the bulk part we need to compute $k(x|\theta)$ and also $K(u|\theta)$, we consider the pólya urn expression in the DPMG to compute posterior realizations for the density $h(x|\theta)$. Let $\{\theta_1^*, ..., \theta_{n^*}^*\}$ the unique values of θ_i, $\omega_i = j$ if and only if $\theta_i = \theta_j^*$ $i = 1, 2, , ..., n$ and $n_j = |\{i : \omega_i = j\}|$ and $j = 1, 2, ..., n^*$ with n^* number of distinct values. We use the following transition probabilities:

   (a) Pólya urn: marginalized G (using $^-$ to indicate summaries without ω_i) and defining a specific configuration $\{\omega_1, .., \omega_n\}$ with transition probabilities:

   $$p(\omega_i = \ell|\omega_{-i}) \propto \begin{cases} n_j^- & j = 1, ..., n^{*-}, \\ \alpha & j = n^{*-} + 1 \end{cases} \tag{15}$$

   (b) Resampling cluster membership indicators ω_i:

   $$p(\omega_i = j|, ..., x_i) \propto \begin{cases} n_j^- k(x_i; \theta_j^*) & j = 1, ..., n^{*-}, \\ \alpha \int k(x_i; \theta_i) dG_0(\theta_i|\eta) & j = n^{*-} + 1 \end{cases} \tag{16}$$

   where we use the close results in Hanson (2006):

   $$k(x_i; \theta_j^*) = h(x_i|\theta_j^*) \tag{17}$$

   $$\int k(x_i; \theta_i) dG_0(\theta_i|\mu, \tau^2) = \tag{18}$$

   $$\frac{a_\lambda a_\gamma}{x_i(x_i + a_\lambda)(a_\lambda - \log(x_i/(x_i + a_\gamma)))^2} \tag{19}$$

   with probability proportional to $n_j^- k(x_i; \theta_j^*)$ we make $\theta_i = \theta_j^{*-}$. On the other hand with probability proportional to $\alpha \int k(y_i; \theta_i, \phi) dG_0(\theta_i|\eta)$ we open a new component and we sample $\theta_i = (\lambda_i, \gamma_i)$. First we sample $\lambda_i|\eta \sim \Gamma(2, a_\lambda - \log(x_i/(x_i + a_\gamma))^2)$ then we sample $\gamma_i|\lambda_i, \eta \sim \Gamma(\lambda_i + 1, x_i + a_\gamma)$.

2. Now we are interested in to show the sampling for the parameters in the GPD defined in the tails of (2). Following do Nascimento et al. (2011) we compute the posterior distribution of u, ξ and σ using three steeps of the Metropolis Hasting algorithm. The algorithm is as follow:

(a) Sampling ξ: proposal transition kernel is given by a truncated normal

$$\xi^*|\xi^b \sim N(\xi^s, V_\xi)I(-\sigma^b/(M-u^b), \infty) \tag{20}$$

where V_ξ is a variance in order to improve the mixing. M is the maximum value in the sample the acceptance probability is

$$\alpha_\xi = \min\left\{1, \frac{p(\theta^*, \phi^*|x)\Phi((\xi^b + \sigma^b/(M-u^b))/\sqrt{V_\xi})}{p(\theta^b, \phi^b|x)\Phi((\xi^* + \sigma^*/(M-u^*))/\sqrt{V_\xi})}\right\}$$

where is the density function of the standard normal distribution.

(b) Sampling σ: If $\xi^{(b+1)} > 0$ then σ^* is sampled from the Gamma distribution $\Gamma(\sigma^{2(b)}/V_\sigma, \sigma^b/V_\sigma)$ where V_σ is a variance in order to improve the mixing. On the other hand if $\xi^{(b+1)} < 0$ then σ^* is sampled from a truncated normal

$$\sigma^*|\sigma^b \sim N(\sigma^s, V_\sigma)I(-\xi^{(b+1)}(M-u^b), \infty) \tag{21}$$

the acceptance probabilities are respectively:

$$\alpha_\sigma = \min\left\{1, \frac{p(\theta^*, \phi^*|x)\Phi((\sigma^b + \xi^{(b+1)}(M-u^b)/\sqrt{V_\sigma})}{p(\theta^b, \phi^b|x)\Phi((\sigma^* + \xi^{(b+1)}(M-u^b)/\sqrt{V_\sigma})}\right\}$$

and

$$\alpha_\sigma = \min\left\{1, \frac{p(\theta^*, \phi^*|x)\Gamma(\sigma^b|\sigma^{2(*)}/V_\sigma, \sigma^*/V_\sigma)}{p(\theta^b, \phi^b|x)\Gamma(\sigma^*|\sigma^{2(b)}/V_\sigma, \sigma^{(b)}/V_\sigma)}\right\}$$

(c) The threshold u^* is sampled following the requirement of the lower truncation for the GPD. Therefore u^* is sampled using a truncated normal density

$$\sigma^*|\sigma^b \sim N(u^s, V_u)I(a^{(b+1)}, \infty) \tag{22}$$

If $\xi^{(b+1)} \geq 0$ then $a^{(b+1)}$ is the minimum value at the sample in the iteration $b+1$ otherwise if $\xi^{(b+1)} < 0$ $a^{(b+1)} = M + \sigma^{(b+1)}/\xi^{(b+1)}$. The acceptance probability is then

$$\alpha_\xi = \min\left\{1, \frac{p(\theta^*, \phi^*|x)\Phi((u^b - a^{b+1})/\sqrt{V_u})}{p(\theta^b, \phi^b|x)\Phi((u^b - a^{b+1})/\sqrt{V_u})}\right\}$$