Regression models for exceedance data via the full likelihood

Fernando Nascimento, Dani Gamerman
Instituto de Matemática, Federal University of Rio de Janeiro
Caixa Postal 68530, 21945-970, Rio de Janeiro, RJ
Emails: fernandofn@dme.ufrj.br and dani@im.ufrj.br

Hedibert F. Lopes
The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL, 60637
Email: hlopes@ChicagoBooth.edu

Abstract

Many situations in practice require appropriate specification of operating characteristics under extreme conditions. Typical examples include environmental sciences where studies include extreme temperature, rainfall and river flow to name a few. In these cases, the effect of geographic and climatological inputs are likely to play a relevant role. This paper is concerned with the study of extreme data in the presence of relevant auxiliary information. The underlying model involves a mixture distribution: a generalized Pareto distribution is assumed for the exceedances beyond a high threshold and a non-parametric approach is assumed for the data below the threshold. Thus, the full likelihood including data below and above the threshold is considered in the estimation. The main novelty is the introduction of a regression structure to explain the variation of the exceedances through all tail parameters. Estimation is performed under the Bayesian paradigm and includes model choice. This allows for determination of higher quantiles under each covariate configuration and upper bounds for the data, where appropriate. Simulation results show that the models are appropriate and identifiable. The models are applied to the study of two temperature datasets: maxima in the U.S.A. and minima in Brazil, and compared to other related models.

Key words: Bayesian inference, generalized Pareto distribution, higher quantiles, MCMC, mixture of distributions, regression models.

1 Introduction

Extreme data analysis has become an important tool in a variety of areas of Science over the last decades, helping the prediction of gains and losses. The areas where this analysis is gaining proeminence are Environmental Sciences and Economics (Embrechts et al., 1997, Reiss and

Thomas, 2007). Global warming is an issue of foremost importance due to changes experienced by the planet over the last years. These changes are inducing alterations in the occurrence of extreme temperature events, both maxima and minima. This is of great relevance since changes in extreme temperature are more responsible for changes in Nature than changes in mean temperature (Parmesan et al., 2000). Extremely warm summers or extremely cold winters may have strong influence in agriculture, energy consumption and health problems for the population. Also, these events may be influenced by a number of covariates such as the specific location in space and time where they take place. Understanding the pattern of occurrence of these effects is important to mitigate the impact that these events may bring to the society.

1.1 Extreme value theory

Extreme value theory (EVT) was designed to describe atypical situations that induce substantial impact when they occur, despite being rare. The classical result in this area is the theorem by Fisher and Tippet (1928), characterizing three possible types of limiting distributions for block maxima. These distributions were unified into the generalized extreme value (GEV) distribution by von Mises (1954) and Jenkinson (1955). Pickands (1975) proved that if X is a random variable whose distribution function F belong to the domain of attraction of a GEV distribution, then the conditional distribution $F(x|u) = P(X \leq u + x|X > u)$, as $u \to \infty$, comes from a generalized Pareto distribution (GPD). The GPD depends on a scale parameter σ, a shape parameter ξ and the threshold u. Denoting $\Psi = (u, \sigma, \xi)$ and letting I be the indicator that $\xi < 0$, ie, $I = 1$, if $\xi < 0$ and $I = 0$, if $\xi \geq 0$, then the GPD density is

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

and $g(x|\Psi) = 0$ otherwise, such that the support of the GPD is $(u, u - \sigma/\xi)$ when $\xi < 0$ and (u, ∞) when $\xi \geq 0$. Smith (1985) showed that maximum likelihood estimators of the GPD parameters do not obey regularity conditions for ξ in $(-1, -0.5)$ and do not exist when $\xi < -1$. However, it is quite uncommon that $\xi < -0.5$ in environmental data, as suggested by Coles and Tawn (1996).

The GPD provides a precise specification for data beyond the limiting threshold and is thus used in practice for inference beyond a suitably chosen large value for u. The results however say nothing about what happens below the threshold u. Different approaches were proposed in the literature to fill this gap. Bermudez et al. (2001), for example, consider only a frequency based approximation for data below the threshold. Frigessi et al. (2002) consider a mixture of a GPD and Weibull distributions, with data dependent weights. Tancredi et al. (2006) use

mixtures of uniform distributions for data below the threshold, while Behrens et al. (2004) use a Gamma distribution.

Given the lack of information below the threshold, non-parametric approximations seem a natural choice. Wiper et al. (2001) showed that mixture of Gamma distributions, denoted here by MG_k, provide good approximations for distributions with positive support. Their density is

$$h(x|\theta, p) = \sum_{j=1}^{k} p_j f_G(x|\mu_j, \alpha_j), \tag{2}$$

where $\theta = (\mu', \alpha', p')'$, $\mu = (\mu_1, \ldots, \mu_k)'$, $\alpha = (\alpha_1, \ldots, \alpha_k)'$, $p = (p_1, \ldots, p_k)'$ and f_G is the Gamma density

$$f_G(x|\mu, \alpha) = \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-(\alpha/\mu)x\} \text{ for } x > 0. \tag{3}$$

The parameters μ and α are the mean and shape of the Gamma distribution, respectively, and may take any positive value, while the weights p_js are positive and sum to 1. The number of components k may be fixed in advance, chosen according to some optimality criterion, or assumed to be an additional model parameter and estimated.

It is not difficult to show that finite mixtures of Gamma distributions belong to the domain of attraction of a GEV distribution (Embrechts et al., 1997), allowing values beyond a certain threshold to be suitably modeled by a GPD distribution. This idea was used in Nascimento et al. (2009) to generalize Behrens et al. (2004) single component gamma distribution by a mixture of Gamma distributions below the threshold and a GPD above the threshold. This class of models, denoted here by $MGPD_k$, has density

$$f(x|\theta, \Psi) = \begin{cases} h(x|\theta) & \text{for } x \leq u \\ (1 - H(u|\theta))g(x|\Psi) & \text{for } x > u \end{cases}, \tag{4}$$

where g is given in (1), h is given in (2) and H its cumulative distribution function. Nascimento et al. (2009) showed advantages of the $MGPD_k$ class over the MG_k class, as expected from the theory.

1.2 Regression models for exceedances

The extremal behavior of a variable of interest may be related to other variables. For example, Castellanos and Cabras (2007) showed that the GPD parameters for rainfall data vary according to the season. In finance, extremes in stock values may be related a stock exchange index or

interest rates or a macroeconomic indicator. In meteorological studies, an important element of a data point is its geographical location. Latitude, longitude, altitude and distance from the ocean may be valuable indicators of extremal behavior of climatological variables. This can be incorporated into the model through a regression structure. Given the results associated with the tail of the distribution and the nonparametric nature of the central part, it seems natural to consider the possibility of building regressions over the GPD parameters.

Cabras et al. (2009) extended the work of Castellanos and Cabras (2007) by allowing for a regression of the GPD parameters but considered only data beyond a pre-specified threshold. They suggested the use of Jeffreys prior distribution for the regression coefficients, in the lack of further information, and used a reparametrization suitable for this purpose. This led to a uniform prior distribution over ξ and $\nu = \sigma(1 + \xi)$, which are shown to be orthogonal (see Chaves-Demoulin and Davison, 2005). Cabras et al. (2009) suggested the link functions $\xi^* = \log(\xi + 1)$ and $\nu^* = \log \nu$ because ξ^* and ν^* have uniform Jeffreys prior distributions. Thus the coefficients of linear regression of covariates over them will also have uniform Jeffreys prior distributions. Other transformations could also be considered. One alternative is to regress directly over σ and ξ possibly after a logarithm transformation.

These transformations are used and compared in the remainder of the paper which extends Cabras et al. (2009) by incorporating uncertainty over the threshold. The paper also introduces a regression structure for the threshold. Moreover, it discusses identifiability issues and provides empirical evidence through simulation exercises that the model can be identified.

1.3 Outline of the paper

Section 2 presents the proposed model, using a Bayesian approach, by merging ideas from Nascimento et al. (2009) with those from Cabras et al. (2009). Thus the complete dataset is used in the analysis: the data below the threshold helps the identification of the threshold, while the data above the threshold is modeled according to regressions over the GPD parameters. This includes a regression model for the threshold. Choice of prior distributions and estimation procedures are also described in this Section. Section 3 shows the results of simulation exercises, highlighting the model ability to recover generated values and discussing some of the model restrictions. Section 4 presents two applications to environmental data: maximum temperature data over a large range of cities in the United States and minimum temperature data over cities in the State of Rio de Janeiro, Brazil. Section 5 summarizes the main results and suggests some potential directions for further studies.

2 Modeling structure

2.1 Likelihood

Consider the availability of n response variables $x_1, \ldots, x_n$ and associated r-dimensional vectors of covariates $z_1, \ldots, z_n$. The first components these vector of covariates is 1 to allow for the presence of an intercept. The class of model proposed in this paper, denoted by $MGPDR_k$, assumes a generalized regression form where responses are conditionally independents and, for $i = 1, \ldots, n$, x_i has density $f(x|\theta, \Psi_i)$ given by (4), a trivariate link function $t(\Psi_i) = \eta_i$ and a trivariate linear preditor $\eta_i = z_i'\beta$.

The $r \times 3$ matrix of regression coefficients β has columns β_u, β_ν and β_ξ. The covariates are assumed to be the same for notational simplicity but need not be so. Appropriate 0 entries into components of β suffice to accommodate for covariates with no explanatory power for some but not all components of Ψ. The triviate link function relates each of the 3 components of Ψ to its linear predictor. Many possibilities are available including the identity link $t(u, \sigma, \xi) = (u, \sigma, \xi)'$.

Following the discussion of the previous section, we have opted for the transformation $t(u, \sigma, \xi) = \eta = (u, \nu^*, \xi^*)'$, where $\nu^* = \xi^* + \log\sigma$ and $\xi^* = \log(\xi + 1)$. These specifications lead to the expression for g in (1) as

$$g(x_i|\Psi_i) = \begin{cases} \frac{\zeta_{\xi i}}{\zeta_{\nu i}}\left(1 + (x_i - z_i'\beta_u))(\zeta_{\xi i} - 1)\frac{\zeta_{\xi i}}{\zeta_{\nu i}}\right)^{-\frac{\zeta_{\xi_i}}{(\zeta_{\xi i}-1)}} & \text{if } \zeta_{\xi i} \neq 1 \\ \nu_i^{-1}\exp(-x_i - z_i'\beta_u/\zeta_{\nu i}), & \text{if } \zeta_{\xi i} = 1, \end{cases} \tag{5}$$

where $\zeta_{\xi i} = \exp(z_i'\beta_\xi)$ and $\zeta_{\nu i} = \exp(z_i'\beta_\nu)$. Also, $x_i - z_i'\beta_u \geq 0$, for $\zeta_{\xi i} \geq 1$ and $(1 + (x_i - z_i'\beta_u)(\zeta_{\xi i} - 1)\zeta_{\xi i}/\zeta_{\nu i} > 0$, if $\zeta_{\xi i} < 1$.

Note that the above conditions imply restrictions over the parameter space for β that depend on the values of the covariates. This nuisance must be considered during the inference process. Other options of link include replacement of $\eta_1 = u$ by $\eta_1 = \log u$ and replacement of $\eta_2 = \nu^*$ by $\eta_2 = \log\sigma$, but they do not remove all the restrictions over the parameter space. The collection of all model parameters is (θ, β). The likelihood function for (θ, β) is given by the product of terms in (4) with g given by (5).

2.2 Prior distribution

The different model components θ and β are assumed to be independent a priori. The prior for θ follows from Wiper et al. (2001) and Nascimento et al. (2009) and is given by $\pi(\theta) = \pi(\mu)\pi(\alpha)\pi(p)$, for $\pi(\mu) \propto \prod_{i=1}^k f_{IG}(\mu_i|a_i/b_i, b_i)I_A(\mu)$ and $\pi(\alpha) = \prod_{i=1}^k f_{IG}(\alpha_i|c_i/d_i, d_i)$,

where $I_A(\mu) = 1$ when μ is in A and zero otherwise, $A = \{\mu \in R^k : \mu_1 < \mu_2 < \cdots < \mu_k\}$, and f_{IG} is the inverse Gamma density with parameters as defined in (3). The prior distribution for the weights p is assumed to be a Dirichlet distribution $D_k(\gamma)$, with $\gamma = (\gamma_1, \ldots, \gamma_k)'$ and density $\pi(p) \propto \prod_{i=1}^k p_i^{\gamma_i}$. The hyperparameters a_i, b_i, c_i, d_i and γ_i, for $i = 1, \ldots, k$, are assumed to be known.

The only remaining prior component is the prior distribution for the regression coefficient β. Cabras et al. (2009) assumed knowledge of the threshold and used a uniform prior for (β_ν, β_ξ). This amounts to an assumption of independence between the β_u, β_ν and β_ξ. This prior specification is used here and completed with a Normal prior for β_u, i.e. $N(b_u, V_u)$. Nascimento et al. (2009) showed that informative prior distributions are required for the estimation of the threshold when the sample size is small due to the lack of identification in the GPD part of the likelihood. Similar comments are valid here for the regression setting. The first component of b_u is the mean of the intercept of the regression on the threshold u. It is typically set around a higher data quantile. The remaining components of b_u are set at 0 to represent lack of knowledge about the relevance of the covariates. Appropriate choice are also needed for V_u, generally assumed to be in diagonal form, with large but finite entries V_{u_i}.

2.3 Posterior distribution

Given the prior and likelihood specifications of the previous subsection, the posterior distribution can be obtained via Bayes theorem. The log posterior density can be written up to an additive constant as

$$\begin{aligned}
\log \pi(\theta, \beta | x, z) &= \sum_{\{i: x_i < z_i'\beta_u\}} \log\left(\sum_{j=1}^k p_j f_G(x_i|\mu_j, \alpha_j)\right) \\
&+ \sum_{\{i: x_i \geq z_i'\beta_u\}} \log\left(1 - \sum_{j=1}^k p_j F_G(z_i'\beta_u|\mu_j, \alpha_j)\right) + \sum_{\{i: x_i \geq z_i'\beta_u\}} \log g(x_i|\Psi_i) \\
&+ \sum_{j=1}^k \left[(a_j - 1)\log(\alpha_j) - b_j\alpha_j - (c_j + 1)\log(\mu_j) - \frac{d_j}{\mu_j}\right] \\
&- \frac{(\beta_{u_0} - a_0)^2}{2V_{\beta_{u_0}}} - \sum_{i=1}^{p-1}\left(\frac{\beta_{u_i}^2}{2V_{\beta_{u_i}}}\right), \qquad (6)
\end{aligned}$$

where the support of the parameter space depends on the identifiability restrictions of parameters and on the values of the covariates. They are given by all values satisfying

$$\exp(2z_i'\beta_\xi) - \exp(z_i'\beta_\xi) > -x_i^{-1}\exp(z_i'\beta_\nu) \text{ and } z_i'\beta_\xi > -\log 2, \qquad (7)$$

for all $i = 1, \ldots, n$. These restrictions can only be verified numerically (Cabras et al., 2009), and are imposed to ensure the existence of the likelihood function.

Inference is performed via approximating Markov chain Monte Carlo (MCMC) techniques, whose details are provided in the Appendix. Sampling algorithms for the components of θ are given by Nascimento et al. (2009). The three vectors of regression coefficients β_u, β_ν and β_ξ are sampled in block. Once the posterior distribution for model parameters is obtained, inference can be made about other unknown quantities of interest. Special interest involve the calculation of higher quantiles, ie, values q such that $P(X \leq q|z)$ is small, such as 1%, 0.1% or 0.01%. These quantities are complicated functions of the model parameters and the values of the covariates. They can not be obtained analytically but can be evaluated numerically. This is particularly suitable for the sampling-based approach used here. In addition, approximations for the posterior distribution of higher quantiles can be easily obtained.

When ξ is negative, the data distribution has an upper bound given by $u - \sigma/\xi$. Depending on the application, this also may be a quantity of interest. For temperature data, it may be important to determine the highest or lowest possible temperature at a given geographical location under some given climatic conditions. Some configurations of values of the covariates may lead to a posterior distribution for ξ entirely concentrated over negative values. This is a case where the upper bound will exist and its distribution can be calculated. When a portion of the distribution is above 0, a mixed posterior distribution is obtained with a lump probability associated with an infinite bound and the remaining probability spread over a finite range.

3 Simulations

Samples generated from the model were drawn under a number of parameter configurations. Estimation can be performed for each configuration. The results associated with parameters and higher quantiles provide empirical evidence about the model capability to recover generated values. This exercise provides useful information for inference in the real data sets of the next Section.

The exercise was performed with samples of size $n = 1,000$ and $n = 10,000$. The central part of the data was generated from a mixture of two Gamma distributions with parameters $\mu = (2, 8)$, $\alpha = (4, 8)$ and $p = (1/3, 2/3)$. Extreme data was generated with two covariates, z_1 and z_2, drawn from $U(0, 2)$ and $U(0, 4)$ distributions respectively. The exceedance data is generated as follows: if the i-th datapoint is larger than $u_i = \beta_{0,u} + \beta_{1,u} z_{1,i}$, then its value is replaced by a value generated from a GPD with parameters $\xi_i = \exp(\beta_{0,\xi} + \beta_{1,\xi} z_{1,i}) - 1$ and $\nu_i = \exp(\beta_{0,\nu} + \beta_{2,\nu} z_{2,i})$, for $i = 1, ..., n$. The values assumed for the regression coefficient β_u were $(6, 0.5, 0)$, $(9, 0.5, 0)$, $(6, -0.5, 0)$ and $(9, -0.5, 0)$. The values chosen for $\beta_{0,u}$ correspond

to higher data quantiles. The value assumed for the regression coefficient β_ν was $(3, 0, 0.5)$. The values assumed for the regression coefficient β_ξ were $(0.2, 0.3, 0)$ and $(0.2, 0.3, 0)$ so that both positive and negative values for ξ are considered in the simulation.

The prior distributions used for the Gamma parameters were $\mu_j \sim IG(2.1, 5.5)$ with order constraints and $\alpha_j \sim IG(6, 0.5)$, for $j = 1, ..., k$. This specification is centered around the true values but with large variance, thus representing lack of prior information. The number of Gamma mixture components was fixed at $k = 2$. Nascimento et al. (2009) provided empirical evidence that this number is well estimated by our procedures. The prior for β_u can not be entirely flat, specially for small to moderate data sizes. Nevertheless, centering $\beta_{0,u}$ around higher data quantiles and the coefficients of the regressors around 0 with reasonably large variances suffices. In the exercises of this section, the prior distributions for $\beta_{0,u}$ and $\beta_{1,u}$ are $N(7, 10)$ and $N(0, 5)$ when $n = 1,000$ and $N(7, 100)$ and $N(0, 100)$ when $n = 10,000$. The MCMC algorithm was set with 2 parallel chains starting from different starting points. After a burn-in period of 10,000 iterations, further 10,000 iterations were performed with thinning of 20 iterations. Convergence was ascertained by visual inspection and standard convergence tests. Trace plots are not presented for conciseness.

Table 1 presents a summary of the estimation of the tail regression parameters for $n = 1,000$. The parameter value lies inside the intervals in almost all cases. The table also shows large variability for some estimates. The estimation is not very precise apart from the intercept $\beta_{0,u}$ and $\beta_{0,\nu}$ and does not point clearly to the significance of the effect of the covariates. It seems to indicate lack of information for samples of size 1,000. The same table presents a summary of the estimation of the tail regression parameters, for $n = 10,000$. The parameter value lies inside the intervals in all cases. The intervals are also smaller than those corresponding to the case $n = 1,000$. Overall, the models seem capable of accurately retrieving information from the data. The data presents a finite upper limit in the cases where ξ is negative. This limit is just a function of the model parameters and the covariates and its posterior distribution can be calculated. Figure 1 shows the distribution of the data maxima for two covariate configurations. The posterior distributions is well concentrated around the generated value with their means very close to the true values. Results get more precise for a smaller threshold as more extreme data is available and estimation of the extreme is more reliable.

4 Aplications

This section illustrates the methodology for real data in two applications with extreme temperature data: maxima over United States cities and minima over cities in the State of Rio de Janeiro, Brazil. Models are compared against a few alternatives using the BIC (Schwarz, 1978) and DIC

$n = 1,000$

	$\beta_{1,\xi} = 0.3$				$\beta_{1,\xi} = -0.3$			
	$\beta_{0,u} = 6$		$\beta_{0,u} = 9$		$\beta_{0,u} = 6$		$\beta_{0,u} = 9$	
	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$
$\beta_{0,u}$	(5.91,6.73)	(5.73,6.31)	(8.42,9.32)	(8.50,9.20)	(5.95,6.24)	(5.94,6.23)	(8.46,9.15)	(8.89,9.15)
$\beta_{1,u}$	(-0.16,0.57)	(-0.73,-0.11)	(0.09,1.46)	(-0.53,-0.04)	(0.31,0.57)	(-0.69,-0.45)	(0.38,0.97)	(-0.64,-0.46)
$\beta_{0,\nu}$	(2.74,3.54)	(3.06,3.75)	(1.96,3.54)	(2.89,3.87)	(2.86,3.39)	(2.80,3.18)	(2.75,3.52)	(2.72,3.32)
$\beta_{2,\nu}$	(0.21,0.58)	(0.17,0.48)	(0.30,0.93)	(0.13,0.55)	(0.34,0.59)	(0.44,0.61)	(0.31,0.62)	(0.35,0.62)
$\beta_{0,\xi}$	(0.01,0.55)	(-0.40,0.17)	(-0.21,0.86)	(0.12,0.81)	(-0.01,0.50)	(-0.08,0.51)	(-0.04,0.64)	(-0.18,0.44)
$\beta_{1,\xi}$	(0.02,0.51)	(0.32,0.80)	(-0.18,0.84)	(-0.37,0.22)	(-0.57,-0.10)	(-0.54,-0.08)	(-0.61,-0.06)	(-0.46,0.12)

$n = 10,000$

	$\beta_{1,\xi} = 0.3$				$\beta_{1,\xi} = -0.3$			
	$\beta_{0,u} = 6$		$\beta_{0,u} = 9$		$\beta_{0,u} = 6$		$\beta_{0,u} = 9$	
	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$	$\beta_{1,u} = 0.5$	$\beta_{1,u} = -0.5$
$\beta_{0,u}$	(5.98,6.02)	(5.96,6.02)	(8.97,9.01)	(8.97,9.03)	(5.99,6.06)	(5.98,6.01)	(8.98,9.02)	(8.99,9.04)
$\beta_{1,u}$	(0.48,0.53)	(-0.51,-0.45)	(0.49,0.53)	(-0.51,-0.47)	(0.46,0.52)	(-0.51,-0.48)	(0.49,0.53)	(-0.53,-0.48)
$\beta_{0,\nu}$	(2.93,3.16)	(2.93,3.16)	(2.93,3.32)	(2.93,3.27)	(2.92,3.07)	(2.90,3.03)	(2.89,3.16)	(2.89,3.06)
$\beta_{2,\nu}$	(0.41,0.52)	(0.44,0.54)	(0.35,0.51)	(0.41,0.56)	(0.46,0.53)	(0.47,0.53)	(0.42,0.54)	(0.46,0.54)
$\beta_{0,\xi}$	(0.14,0.31)	(0.14,0.31)	(0.05,0.26)	(0.12,0.34)	(0.07,0.23)	(0.11,0.27)	(0.05,0.28)	(0.05,0.31)
$\beta_{1,\xi}$	(0.18,0.33)	(0.22,0.36)	(0.25,0.47)	(0.15,0.35)	(-0.33,-0.20)	(-0.35,-0.22)	(-0.33,-0.10)	(-0.41,-0.19)

Table 1: 95% credibility intervals for regression parameters when $n = 1000$ (top) or $n = 10,000$ (bottom), with parameters $\beta_{0,\nu} = 3$, $\beta_{2,\nu} = 0.5$ and $\beta_{0,\xi} = 0.2$ fixed through all configurations.

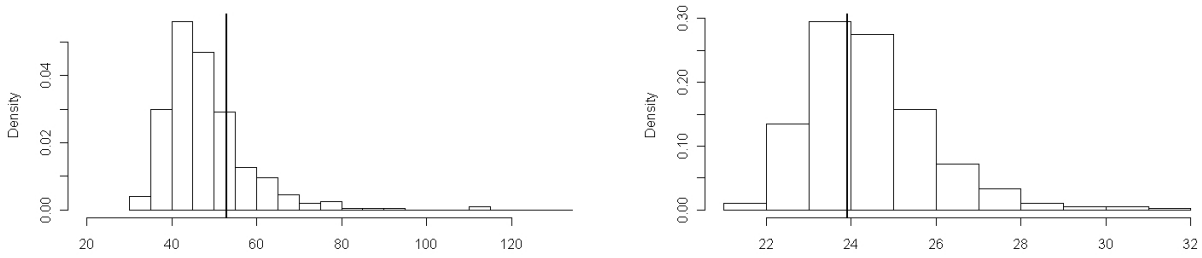


Figure 1: Posterior histograms of maxima in simulations with $n = 10000$, $\beta_{1,u} = -0.5$, $\beta_{0,\nu} = 3$, $\beta_{1,\nu} = 5$, $\beta_{0,\xi} = 0.2$ and $\beta_{1,\xi} = -0.3$: left panel: $\beta_{0,u} = 9$ with $z_{1,i} = 1$ and $z_{2,i} = 2$; right panel: $\beta_{0,u} = 6$ with $z_{1,i} = 1.5$ and $z_{2,i} = 1$. Vertical lines indicate corresponding true values.

(Spiegelhalter et al., 2002) criteria.

4.1 Maximum temperature in the United States

The dataset contains daily average temperatures, in Farenheit degrees, from 1995 to 2008 in 84 cities scattered over the continental United States (`www.engr.udayton.edu/weather`). Data used in the analysis was monthly maxima, totaling 14,356 observations. Many factors may affect temperature, specially season and location. The extensive region covered by the area under study makes it relevant to include latitude. Data comes from cities as close to the Equator as Miami, with latitude $25^o47'$ and as close to the Arctic Circle as Seattle, with latitude $47^o37'$.

The seasonality obviously present in temperature data is incorporated into the model through trigonometric functions. The effect of the season is in most cases well captured by a single sine wave. Latitude is also taken into account. This leads to a model where $z = (z_0, z_1, z_2, z_3, z_4)$, where $z_0 = 1$ is the intercept; $z_1 = \cos(2\pi m/12)$ and $z_2 = \sin(2\pi m/12)$, where m is the month of the observation; $z_3 = (l - m_l)/10$, where l is the latitude of the observation and m_l is the average latitude over all cities considered; and $z_4 = z_2 z_3$ represents an interaction term between season and latitude.

Estimation was performed with model $MGPDR_k$ for different values of k and with tail parameters described as functions of z. The prior distribution for components of β_u were independent normals with $\beta_{0,u} \sim N(40, 10000)$ and $\beta_{i,u} \sim N(0, 1000)$, for $i = 1, \ldots, 4$. The results of this model were compared against a mixture of Gammas (Wiper et al., 2001) and against mixture of Gammas with GPD tail parameters fixed, without covariates as in Nascimento et al. (2009). Table 2 shows a summary of the model comparison. It shows a significant improvement in terms of fit after inclusion of the covariates *and* after inclusion of a specific component to handle exceedance data. Table 3 presents a summary of the tail parameters estimation for model $MGPDR_3$, indicated as the best model in Table 2.

Model	p_D	$10^5 DIC$	$10^5 BIC$
$MGPDR_1$	19.30	0.9352	0.9371
$MGPDR_2$	11.54	0.9169	0.9192
$\mathbf{MGPDR_3}$	**5.16**	**0.9141**	**0.9163**
$MGPDR_4$	8.55	0.9151	0.9177
$MGPD_2$	5.99	1.1325	1.1335
MG_3	6.86	1.1321	1.1331

Table 2: Fit criteria for maximum temperature data in the U.S. MG_3 and $MGPD_2$ are the best models in their class according to the DIC criteria.

Other parametrizations were also considered for model $MGPDR_k$. The first one is to take

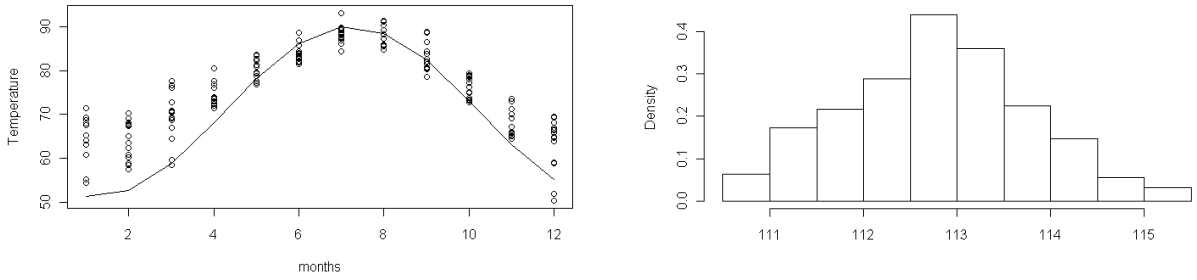
the logarithm transformation for the threshold, ie, $u = \exp(z'\beta_u)$. Once again, the best BIC and DIC values were also obtained with $k = 3$ with respective values 0.9333 and 0.9309 $\times 10^5$. Another alternative considered was to build a regression over the original GPD parameter $\sigma = \exp(z'\beta_\sigma)$ instead of ν. A uniform prior was considered for β_σ to represent lack of information but not in Jeffreys' sense. Once again, the best model in this class obtained BIC and DIC values given respectively by 0.9243 and 0.9265 $\times 10^5$. Therefore, the parametrization proposed in this paper for the models seems to have practical support in addition to their theoretical justification.

$\beta_{0,\xi}$	$\beta_{1,\xi}$	$\beta_{2,\xi}$	$\beta_{3,\xi}$	$\beta_{4,\xi}$
-0.54	-0.15	-0.10	0.10	-0.01
(-0.55,-0.53)	(-0.16,-0.13)	(-0.14,-0.07)	(0.07,0.14)	(-0.02,0.01)
$\beta_{0,\nu}$	$\beta_{1,\nu}$	$\beta_{2,\nu}$	$\beta_{3,\nu}$	$\beta_{4,\nu}$
1.77	0.15	0.18	0.15	0.21
(1.76,1.78)	(0.14,0.17)	(0.16,0.20)	(0.12,0.18)	(0.20,0.23)
$\beta_{0,u}$	$\beta_{1,u}$	$\beta_{2,u}$	$\beta_{3,u}$	$\beta_{4,u}$
63.01	-17.37	-16.03	-4.10	-11.71
(62.99,63.02)	(-17.39,-17.35)	(-16.06,-15.99)	(-4.14, -4.06)	(-11.74,-11.69)

Table 3: Summary of the estimation results for model $MGPDR_3$: posterior means, with 95% credibility intervals in parentheses.

Virtually all covariates seem to have a significant effect in explaining the tail parameters. The only exception in the interaction term in ξ. The thresholds associated with all cities along the year can be estimated. Figure 2(a) shows an example of the maxima for each month/year and they seems to compare well against the estimated thresholds. The maximum temperatures are higher in July. The posterior distribution for ξ seem to be concentrated over negative values with mean -0.27 for the same city of Little Rock. Therefore, the distribution has an upper limit. Figure 2(b) shows the distribution of this upper limit. According to our estimation results, about 83% of the observations were above the threshold. It is expected from theory that only a small proportion of the data should lie above the threshold. Since the prior distribution for the threshold was fairly vague, this result can only be attributed to the data due to the information they provide to the likelihood.

The data distribution is not bounded from above when $\xi > 0$. Interesting quantities for study in these cases are higher quantiles. They inform about the probability of rare, extreme events. Figure 3 illustrates this point for the city of Los Angeles. It shows for example that on average the probability of a winter temperature larger than 74 degrees is 1%, ie, it occur on average once in a century. Likewise, a summer temperature above 101 degrees is expected once in a century.



(a) (b)

Figure 2: (a) Monthly maxima temperature data for Little Rock, Arkansas. Full line: posterior mean thresholds. (b) Posterior histogram for the temperature upper limit in Little Rock, Arkansas, in July.

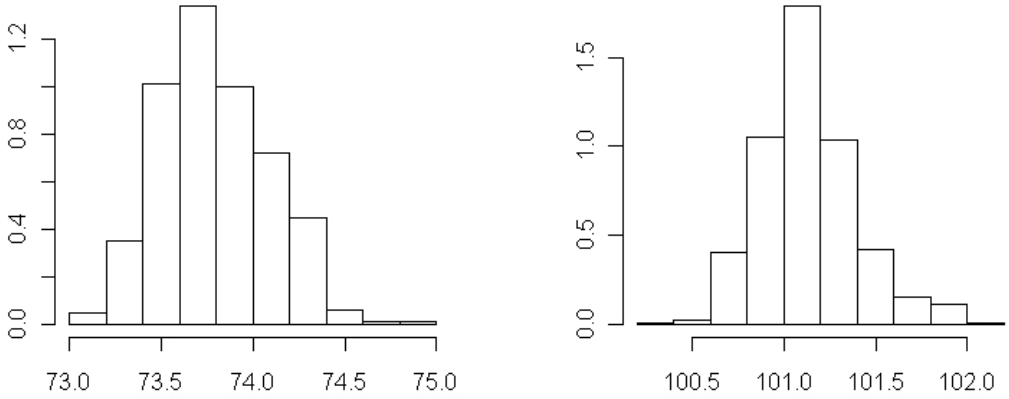


Figure 3: Posterior histogram for the 99% quantile temperature in Los Angeles, California. Left panel: January. Right panel: June.

One could still analyse each combination of city/month separately but then very few observations would be available. The regression structure allows combination of different cities and months into a unified framework. This leads to a more precise and reliable procedure for estimating extremes.

4.2 Minimum temperature in Rio de Janeiro

This second dataset consists on minimum daily temperatures, in Celsius degrees, in a few monitoring stations spread over the State of Rio de Janeiro, Brazil, from 1961 to 2000. Data was analysed using minimum monthly temperatures, resulting in 11,336 observations. The tail of the observation now lies to the left. The transformation $x = 30 - y$, where y is the original data, was performed to bring the tail to the left, as considered in our models. There is substantial assurance that only positive observations will be obtained because the largest minimum temperature observed in our extensive dataset was 25.6 degrees, well below the limit of our transformation. Figure 4 shows original and transformed data.

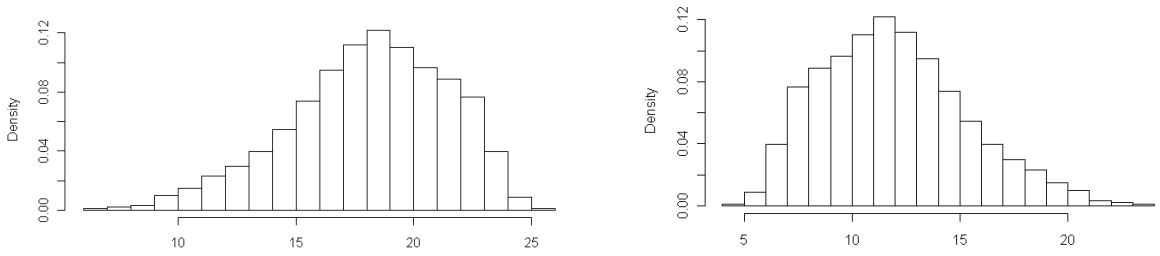


Figure 4: Histogram of the observations of the minimum temperature (in Celsius degrees) data for the State of Rio de Janeiro: left panel: original data; right panel: transformed data.

The latitudes vary here only from $20^o76'$ to $23^o36'$, much less than in the previous example. The geographical characteristic of relevance here is altitude, given the diversified topography of the region with plains, hills and chains of mountains. Once again, the season is relevant and enters the model as before, through a single sine wave. Thus, the covariates are the same of the previous application but for the replacement of latitude by altitude. Thus, the vector of covariates is $z = (z_0, z_1, z_2, z_3, z_4)'$, where $z_0 = 1$ is the intercept; $z_1 = \cos(2\pi m/12)$ and $z_2 = \sin(2\pi m/12)$, where m is the month of the observation; $z_3 = (a - m_a)/100$, where a is the

altitude of the observation and m_a is the mean value of the altitudes; and $z_4 = z_2 z_3$ represents an interaction term between season and altitude.

Prior distributions used for the regression parameters β_ν and β_ξ were assumed to be proportional to a constant. Prior distributions used for the components of the regression parameter β_u were independent $\beta_{0,u} \sim N(10, 50)$ and $\beta_{i,u} \sim N(0, 30)$, $i = 1, \ldots, 4$. Models $MGPDR_k$ were compared against models MG_k and $MGPD_k$. Table 4 shows a summary of the model comparison. It shows a significant improvement in terms of fit after inclusion of the covariates *and* after inclusion of a specific component to handle exceedance data, as in the previous application. This can be observed more closely at Table 5 that shows relevance of virtually all covariates for all three GPD parameters.

Model	p_D	$10^5 DIC$	$10^5 BIC$
$MGPDR_1$	6.54	0.4253	0.4249
$MGPDR_2$	9.78	0.4230	0.4250
$MGPDR_3$	18.26	0.4222	0.4248
$MGPDR_4$	11.03	0.4227	0.4253
$MGPD_2$	1.70	0.5866	0.5874
MG_4	8.13	0.5863	0.5876

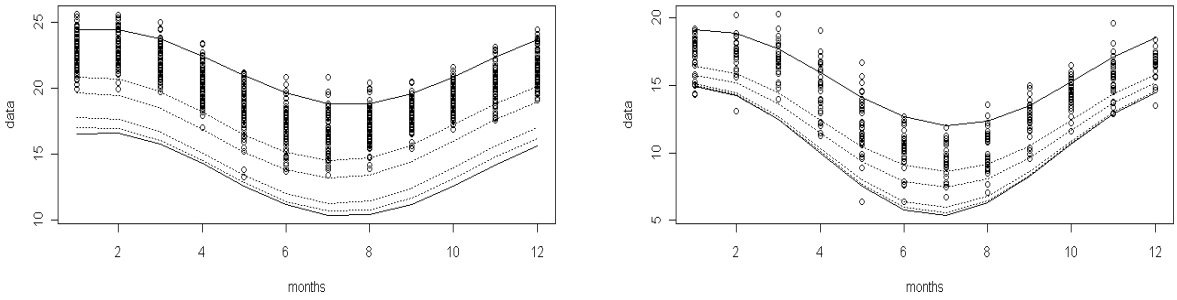
Table 4: Fit criteria for minimum temperature data in the State of Rio de Janeiro. MG_4 and $MGPD_2$ are the best models in their class according to the DIC criteria.

$\beta_{0,\xi}$	$\beta_{1,\xi}$	$\beta_{2,\xi}$	$\beta_{3,\xi}$	$\beta_{4,\xi}$
-0.293	0.011	-0.017	-0.012	-0.016
(-0.301,-0.284)	(-0.001,0.023)	(-0.022,-0.010)	(-0.020,-0.005)	(-0.039,0.007)
$\beta_{0,\nu}$	$\beta_{1,\nu}$	$\beta_{2,\nu}$	$\beta_{3,\nu}$	$\beta_{4,\nu}$
0.577	-0.147	-0.022	-0.008	0.018
(0.562,0.589)	(-0.165,-0.131)	(-0.028,-0.018)	(-0.014,-0.002)	(-0.005,0.040)
$\beta_{0,u}$	$\beta_{1,u}$	$\beta_{2,u}$	$\beta_{3,u}$	$\beta_{4,u}$
5.752	-2.260	0.705	-0.104	-2.083
(5.750,5.753)	(-2.262,-2.256)	(0.704,0.706)	(-0.105, -0.103)	(-2.085,-2.082)

Table 5: Summary of the estimation results for model $MGPDR_3$: posterior means, with 95% credibility intervals in parentheses.

Figure 5 shows the results for two specific stations: Ilha Guaíba and Nova Friburgo. The

first one is located at sea level, on an island, while the second is in the mountainous region, at 857 meters high. The model seems to capture their relatively different patterns well. For the station Ilha Guaíba, for example, the posterior means of the minimum temperature and of the 5% quantile are 10.4 and 14.6 degrees. This means a temperature as low as 14.6 degrees is expected on average only once in every 20 years. The figure also informs that two May observations were very close to the posterior mean of the 0.1% quantile, being therefore extremely rare, and expected to occur once every milenium For the Nova Friburgo station, the posterior mean for the lower limit is 5.4 degrees while the posterior mean for the 5% quantile is 8.7 degrees. There were also four observations below the posterior mean of the lower limits.



(a) Ilha Guaíba station

(b) Nova Friburgo station

Figure 5: Monthly minima temperature data for Ilha Guaíba and Nova Friburgo stations. Full lines: posterior mean thresholds (above) and posterior mean of the data lower limit. Dotted lines: posterior means for the 0.05, 0.01, 0.0001 and 0.000001 quantiles.

5 Final remarks

This paper is concerned with the study of variation of extremal behaviour in the presence of relevant external information. Simulation studies show that true values can be recovered with great precision for large datasets. Special care must be exercised when specifying prior vagueness for moderate or small datasets. Important parametric functions such as the data upper limits can also be recovered.

Simulation and results from real data analysis shows the improvement obtained with the incorporation of the different model components and the need for all of them. Vague prior

information was dominated by the likelihood information and led to reliable estimates of the regression coefficients.

Models can be extended into a number of directions. The most proeminent ones are those related to the incorporation of spatial and/or temporal heterogeneity. This can be achieved through some form of stochastic dependence. This dependence that can be added to the regression structure, in cases were the covariates were not capable of handling all the sources of data variation.

Appendix: MCMC algorithm

The MCMC algorithm consist on iterations over the parameter space, performed over blocks of parameters. The blocks were formed by μ, α, p, β_u, β_ν and β_ξ. At iteration s, parameters are updated to iteration $s+1$ as follows:

Sampling θ. The components of θ are sampled separately for each mixture component. The α_js and μ_js must be positive. Therefore, α_j^* is proposed from $\alpha_j^*|\alpha_j^{(s)} \sim G(\alpha_j^{(s)}, \alpha_j^{(s)^2}/V_{\alpha_j})$. Note that, $E(\alpha_j^* \mid \alpha_j^{(s)}) = \alpha_j^{(s)}$, and $Var(\alpha_j^* \mid \alpha_j^{(s)}) = V_{\alpha_j}$, for $j = 1, \ldots, k$. Same procedure is adopted for the proposal for μ_j, given by $\mu_j^*|\mu_j^{(s)} \sim G(\mu_j^{(s)}, \mu_j^{(s)^2}/V_{\mu_j})I_A$, where $I_A = I(\mu_1^{(s+1)} < \ldots < \mu_{j-1}^{(s+1)} < \mu_j^{(s)} < \mu_{j+1}^{(s)} < \ldots < \mu_k^{(s)})$ and the difference that they must also obey the order constraint. The values $\alpha_j^{(s+1)} = \alpha_j^*$ and $\mu_j^{(s+1)} = \mu_j^*$ are accepted with probability

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

where $\Theta^* = (\alpha_{<j}^{(s+1)}, \alpha_j^*, \alpha_{>j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_j^*, \mu_{>j}^{(s)}, p^{(s)}, \beta_u^{(s+1)}, \beta_\nu^{(s+1)}, \beta_\xi^{(s+1)})$ and $\tilde{\Theta} = (\alpha_{<j}^{(s+1)}, \alpha_{\geq j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_{\geq j}^{(s)}, p^{(s)}, \beta_u^{(s+1)}, \beta_\nu^{(s+1)}, \beta_\xi^{(s+1)})$, with $y_{<l} = (y_1, ..., y_{l-1})$ and $y_{\geq l} = (y_l, ..., y_k)$, for any vector $y = (y_1, ..., y_k)$.

Sampling p. p^* is sampled from a Dirichlet proposal with parameters $(V_p p_1^{(s)}, \ldots, V_p p_k^{(s)})$, where V_p is a tuning constant that determines the variance of the proposal distribution. Then, set $p^{(s+1)} = p^*$ with probability $\min\left\{1, \frac{\pi(\theta^{(s+1)}, p^*, \beta^{(s)}|\mathbf{x}) f_D(p^{(s)}|p^*)}{\pi(\theta^{(s+1)}, p^{(s)}, \beta^{(s)}|\mathbf{x}) f_D(p^*|p^{(s)})}\right\}$.

Sampling β_u. Sample β_u^* from a $N(\beta_u^{(s)}, V_u)$ distribution. The sampled value must satisfy (7). Otherwise, a new value must be sampled until the conditions are satisfied. Then, set $\beta_\nu^{(s+1)} = \beta_\nu^*$ with probability $\min\left\{1, \frac{\pi(\theta^{(s)}, p^{(s)}, \beta_u^*, \beta_\nu^{(s)}, \beta_\xi^{(s)}|\mathbf{x})}{\pi(\theta^{(s)}, p^{(s)}, \beta_u^{(s)}, \beta_\nu^{(s)}, \beta_\xi^{(s)}|\mathbf{x})}\right\}$.

Sampling β_ν. Sample β_ν^* from a $N(\beta_\nu^{(s)}, V_\nu I_p)$ distribution. The sampled value must satisfy (7). Otherwise, a new value must be sampled until the conditions are satisfied. Then, set $\beta_\nu^{(s+1)} = \beta_\nu^*$ with probability $\min\left\{1, \frac{\pi(\theta^{(s)},p^{(s)},\beta_u^{(s)},\beta_\nu^*,\beta_\xi^{(s)}|\mathbf{x})}{\pi(\theta^{(s)},p^{(s)},\beta_u^{(s)},\beta_\nu^{(s)},\beta_\xi^{(s)}|\mathbf{x})}\right\}$.

Sampling β_ξ. Sample β_ξ^* from a $N(\beta_\xi^{(s)}, V_\xi I_p)$ distribution. The sampled value must satisfy (7). Otherwise, a new value must be sampled until the conditions are satisfied. Then, set $\beta_\xi^{(s+1)} = \beta_\xi^*$ with probability $\min\left\{1, \frac{\pi(\theta^{(s)},p^{(s)},\beta_u^{(s)},\beta_\nu^{(s)},\beta_\xi^*|\mathbf{x})}{\pi(\theta^{(s)},p^{(s)},\beta_u^{(s)},\beta_\nu^{(s)},\beta_\xi^{(s)}|\mathbf{x})}\right\}$.

The method of Roberts and Rosenthal (2006) prescribes optimality of MCMC algorithms for univariate components when acceptance rates of 0.44 are obtained. Smaller acceptance rates around 0.15 are used since the proposals here are not univariate. These settings were used to select the tuned variance parameters. This choice was efficient for all components with reasonably fast convergence to stationarity. The only parameter where variances were not tuned automatically according to this rule was p. A few choices of values were imposed until convergence was also reached.

References

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

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

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

Cabras, S. , Castellanos, M. A. and Gamerman, D. (2009). A default approach for regression on parameters of generalized Pareto distribution. *Technical report*, LES, UFRJ.

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

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics*, **45**, 463-78.

Coles, S. G. (2001). *Extreme Value Theory an Applications*. In Galambos, J., Lechner, J. and Mimiu, E. (editors). Kluver Academic Publishers.

Dalal, S. and Hall, W. (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society, Series B*, **45**, 278-286.

de Vore, R. and Lorentz, G. (1993). *Constructive Approximation*. New York: Springer Verlag.

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

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

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

Fisher, R. A. and Tippett, L. H. C. (1928). On the estimation of the frequency distributions of the largest and smallest sumber of a sample, *Proceedings of the Cambridge Philosophycal Society*. **24**, 180-190.

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

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

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorology Society*, **81**, 158-171.

Nascimento, F. F. , Gamerman, D. and Lopes, H. F. (2009). Semiparametric Bayesian approach to extreme estimation. *Technical report*, LES, UFRJ.

Parmesan, C., Root, T. L., and Willing, M. R. (2000). Impacts of Extreme Weather and Climate on Terrestrial Biota. *Bulletin of the American Meteorological Society*, **81**, 443-450.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, **3**, 131-199.

Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields (3rd edition)*, Basel: Birkhäuser Verlag.

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

Roberts, G. O. and Rosenthal, J. S. (2006). Examples of adaptive MCMC. *Technical Report*, Lancaster University.

Roeder, K. e Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, **92**, 894-902.

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

Smith, R. L., (1984). Threshold models for sample extremes. In J. Tiago de Oliveira, Reidel, Dordrecht (editors), *Statistical extremes and applications*, 621-638.

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

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

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

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

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