



An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails

M. de CARVALHO , S. PEREIRA, P. PEREIRA, and P. de ZEA BERMUDEZ

We introduce a novel regression model for the conditional left and right tail of a possibly heavy-tailed response. The proposed model can be used to learn the effect of covariates on an extreme value setting via a Lasso-type specification based on a Lagrangian restriction. Our model can be used to track if some covariates are significant for the lower values, but not for the (right) tail—and vice versa; in addition to this, the proposed model bypasses the need for conditional threshold selection in an extreme value theory framework. We assess the finite-sample performance of the proposed methods through a simulation study that reveals that our method recovers the true conditional distribution over a variety of simulation scenarios, along with being accurate on variable selection. Rainfall data are used to showcase how the proposed method can learn to distinguish between key drivers of moderate rainfall, against those of extreme rainfall. Supplementary materials accompanying this paper appear online.

Key Words: Conditional tail; Extended generalized Pareto distribution; Heavy-tailed response; Lasso; L_1 -Penalization; Nonstationary extremes; Statistics of extremes; Variable selection.

1. INTRODUCTION

Learning about the drivers of risk is key in a variety of fields, including climatology, environmental sciences, finance, forestry, and hydrology. Mainstream approaches for learning about such drivers or predictors of risk include, for instance, [Davison and Smith \(1990\)](#), [Chavez-Demoulin and Davison \(2005\)](#), [Eastoe and Tawn \(2009\)](#), [Wang and Tsai \(2009\)](#), [Chavez-Demoulin et al. \(2016\)](#), and [Huser and Genton \(2016\)](#).

The main contribution of this paper rests on a Bayesian regression model for the conditional lower values (i.e., left tail) and conditional (right) tail of a possibly heavy-tailed response. Our model pioneers the development of regression methods in an extreme value framework that allow for some covariates to be significant for the lower values but not for the tail (and vice versa). Some further comments on the proposed model are in order. First, the proposed model bypasses the need for (conditional) threshold selection; such selection

M. de Carvalho (✉) School of Mathematics, University of Edinburgh, Edinburgh, UK
(E-mail: miguel.decarvalho@ed.ac.uk).

S. Pereira · P. de Zea Bermudez Faculdade de Ciências and CEAUL, Universidade de Lisboa, Lisboa, Portugal
(E-mail: sapereira@fc.ul.pt).

P. Pereira EST Setúbal/IPS and CEAUL, Lisboa, Portugal (E-mail: paula.pereira@estsetubal.ips.pt).

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 2, Pages 222–239
<https://doi.org/10.1007/s13253-021-00469-9>

is particularly challenging in a regression framework as it entails selecting a function of the covariate (say, u_x) rather than a single scalar. Second, our method models both the conditional lower values and the conditional tail, offering a full portrait of the conditional distribution, while still being able to extrapolate beyond observed data into the conditional tail. Finally, our method is directly tailored for variable selection in an extreme value framework; in particular, it can be used to track what covariates are significant for the lower values and tails.

In an extreme value framework, interest focuses on modeling the most extreme observations, disregarding the central part of the distribution. Usually, efforts center on modeling extreme values using the generalized extreme value distribution or the generalized Pareto distribution (Embrechts et al. 1997; Coles 2001; Beirlant et al. 2004). Many observations are disregarded using the latter approaches, and the choice of the block size or threshold is far from straightforward. Moreover, in many situations of applied interest, it would be desirable to model both the lower values of the data along with the extreme values in a regression framework. Our model builds over Papastathopoulos and Tawn (2013) and Naveau et al. (2016) who proposed an extended generalized Pareto distribution (EGPD) to jointly model low, moderate, and extreme observations—without the need of threshold selection; other interesting options for modeling both the bulk and the tail of a distribution, include extreme value mixture models (e.g., Frigessi et al. 2002, Behrens et al. 2004, Carreau and Bengio 2009, Cabras and Castellanos 2011, MacDonald et al. 2011, do Nascimento et al. 2012) as well as composition-based approaches (Stein 2020, 2021).

The proposed model can be regarded as a Bayesian Lasso-type model for the lower values and tail of a possibly heavy-tailed response supported on the positive real line. The Bayesian Lasso was introduced by Park and Casella (2008) as a Bayesian version of Tibshirani's Lasso (Tibshirani 1996). Roughly speaking, the frequentist Lasso is a regularization method, that shrinks some regression coefficients, and sets others to zero; as such it is naturally tailored for variable selection. The frequentist Lasso solves a least squares type of optimization problem, where an L_1 -penalization is incorporated via a constraint. The starting point for the Bayesian Lasso is the fact that the posterior mode generated from a Laplace prior coincides with the solution to the Lasso least squares optimization problem with L_1 -penalization (Reich and Ghosh 2019, Sect. 4.2.3). It should be noted however that the Bayesian Lasso does not set coefficients to zero.

Shrinkage in a Bayesian context has been traditionally achieved via priors with peaks at zero such as the Laplace. Other examples of continuous shrinkage priors include the horseshoe (Carvalho et al. 2010), Dirichlet–Laplace (Bhattacharya et al. 2015), and R2D2 (Zhang et al. 2020). There are also discrete shrinkage priors (George and McCulloch 1993) which are arguably more interpretable. Despite all these considerable advances in Bayesian variable selection, the Bayesian Lasso is simply one way to go, one that is associated L_1 -penalization, but other priors could be used to meet a similar target, leading to other geometries for the corresponding penalization. For a recent review of Bayesian regularization methods, see Polson and Sokolov (2019).

As we clarify below (Sect. 2.2), our model has also links with quantile regression. Indeed, by modeling both the conditional lower values and the conditional tail, our model bridges quantile regression (Koenker and Bassett 1978) with extremal quantile regression (Cher-

nozhlukov 2005). Finally, as a by-product, this paper contributes to the literature on Bayesian inference for Pareto distributions (Arnold and Press 1989; de Zea Bermudez and Turkman 2003; Castellanos and Cabras 2007; de Zea Bermudez and Kotz 2010; Villa 2017).

The rest of this paper unfolds as follows. In Sect. 2, we introduce the proposed methods. We assess the performance of the proposed methods on simulated examples in Sect. 3, and report the main findings of our numerical studies. A data illustration is included in Sect. 4.

2. EXTREME VALUE BAYESIAN LASSO

To streamline the presentation, the proposed methods are introduced in a step-by-step fashion, with the most flexible version of our model being introduced in Sect. 2.4.

2.1. THE EXTENDED GENERALIZED PARETO FAMILY

Our starting point for modeling is the so-called extended generalized Pareto distribution (EGPD), as proposed in Papastathopoulos and Tawn (2013) and Naveau et al. (2016); the EGPD is a distribution over the positive real line, whose cumulative distribution function is $F(y) = G\{H(y)\}$, where $G : [0, 1] \rightarrow [0, 1]$ is a carrier function, which tilts the generalized Pareto distribution (GPD) function

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi},$$

defined on $\{y \in (0, \infty) : 1 + \xi y/\sigma > 0\}$. Here, $\sigma > 0$ is a scale parameter, $\xi \in \mathbb{R}$ is the shape parameter; the case $\xi = 0$ should be understood by taking the limit $\xi \rightarrow 0$. The shape parameter ξ is termed extreme value index and it controls the rate of decay of the tail. Following Naveau et al. (2016), we assume that the carrier function G obeys the following conditions:

- A. $\lim_{v \rightarrow 0^+} \{1 - G(1 - v)\}/v = a$, with $a > 0$.
- B. $\lim_{v \rightarrow 0^+} G\{v w(v)\}/G(v) = b$, with $b > 0$ and $w(v) > 0$ such that $w(v) = 1 + o(v)$ as $v \rightarrow 0^+$.
- C. $\lim_{v \rightarrow 0^+} G(v)/v^c = c$, with $c > 0$.

Assumption A ensures a Pareto-type tail, whereas Assumptions B–C ensure that the lower values are driven by the carrier G . In more detail, Assumption A implies that

$$\lim_{y \rightarrow y^*} \frac{1 - F(y)}{1 - H(y)} = a,$$

and thus it can be understood as a tail-equivalence condition (Embrechts et al. 1997, Sect. 3.3), where $y^* = \inf\{y : F(y) < 1\}$ is the so-called right endpoint, here assumed to be positive. Since tail-equivalence implies that both F and H are in the same domain of attraction (Resnick 1971), it follows from Assumption A that ξ can be literally interpreted

as the extreme value index of $F(y) = G\{H(y)\}$. We note further that Assumption C implies that small values follow a Weibull-type GPD, and that the role and need of Assumption B are in fact questionable (Tencaliec et al. 2019); actually, the latter paper makes no reference to it.

For parsimony reasons, we focus on modeling G using a parametric class, so that $G(v) \equiv G_{\kappa}(v)$, with $\kappa \in \mathbb{R}^q$. The canonical example of a parametric carrier is $G_{\kappa}(v) = v^{\kappa}$, with $\kappa > 0$ controlling the shape of the lower tail, with a larger value of κ leading to less mass close to zero; we refer to the EGPD distribution with the latter carrier as the canonical EGPD.

Below, we use the notation $Y \sim \text{EGPD}_G(\kappa, \sigma, \xi)$ to denote that Y follows an EGPD with parameters (κ, σ, ξ) and carrier G , and \mathcal{G} to represent the space of all carrier functions $G : (0, \infty) \rightarrow [0, 1]$ obeying Assumptions A–C.

2.2. A FIRST CONDITIONAL MODEL FOR THE LOWER VALUES AND TAIL OF A POSSIBLY HEAVY-TAILED RESPONSE

The first version of our model specifies the conditional distribution function

$$F(y \mid \mathbf{x}) = G_{\kappa(\mathbf{x})}[H(y)], \quad (1)$$

where $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ is a vector of covariates and $\kappa(\mathbf{x})$ is a vector-valued function with components given by inverse-link functions

$$\kappa(\mathbf{x}) = [\kappa_1(\mathbf{x}^T \boldsymbol{\beta}_1), \dots, \kappa_q(\mathbf{x}^T \boldsymbol{\beta}_q)], \quad (2)$$

and $G_{\kappa(\mathbf{x})} \in \mathcal{G}$, for every $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p$; here and below, $\boldsymbol{\beta}_j = (\beta_{1,j}, \dots, \beta_{p,j}) \in \mathbb{R}^p$, for $j = 1, \dots, q$. For reasons that will become evident later, we do not allow for now (σ, ξ) to depend on the covariates \mathbf{x} , but we will extend the specification in (1) in Sect. 2.4. The next examples illustrate that different types of carrier functions exist, and thus that there exist different forms of tilting the GPD distribution function via a carrier function; the examples also highlight that, given the freedom to choose a G obeying Assumptions A–C, the model is quite flexible and that thus it would be challenging to numerically illustrate all its instances.

Example 1. (Power carrier, exponential link) The canonical embodiment of the first version of our model in (1) is obtained by specifying

$$G_{\kappa(\mathbf{x})}(v) = v^{\kappa(\mathbf{x})}, \quad \kappa(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}). \quad (3)$$

Example 2. (Beta carrier, exponential link) Another variant of (1) is obtained by specifying

$$G_{\kappa(\mathbf{x})}(v) = 1 - Q_{\kappa(\mathbf{x})}(1 - v^{\kappa(\mathbf{x})}), \quad \kappa(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad (4)$$

where

$$Q_{\kappa(\mathbf{x})}(v) = \frac{1 + \kappa(\mathbf{x})}{\kappa(\mathbf{x})} v^{1/\kappa(\mathbf{x})} \left(1 - \frac{v}{1 + \kappa(\mathbf{x})} \right)$$

is the distribution function of a Beta distribution with parameters $(1/\kappa(\mathbf{x}), 2)$.

Example 3. (Power mixture carrier, exponential links) Still another variant of (1) is obtained by specifying

$$G_{\kappa(\mathbf{x})}(v) = \pi v^{\kappa_1(\mathbf{x})} + (1 - \pi) v^{\kappa_2(\mathbf{x})},$$

with $0 < \pi < 1$ and

$$\kappa_1(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}_1), \quad \kappa_2(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}_2).$$

The conditions on the intercepts $\beta_{1,1} > \beta_{1,2}$ and $\beta_{j,1} = \beta_{j,2}$, for $j = 2, \dots, q$, leads to $\kappa_1(\mathbf{x}) > \kappa_2(\mathbf{x})$ and is used for identifiability purposes.

A consequence of (1) is that

$$F^{-1}(p | \mathbf{x}) = \begin{cases} \frac{\sigma}{\xi} [\{1 - G_{\kappa(\mathbf{x})}^{-1}(p)\}^{-\xi} - 1], & \xi \neq 0, \\ -\frac{\sigma}{\xi} \log\{1 - G_{\kappa(\mathbf{x})}^{-1}(p)\}, & \xi = 0, \end{cases} \quad (5)$$

where $F^{-1}(p | \mathbf{x}) = \inf\{y : F(y | \mathbf{x}) \geq p\}$, for $0 < p < 1$. Equation (5) warrants some remarks on links with quantile regression. The first version of our model in (1) can be regarded as a model that bridges quantile regression (Koenker and Bassett 1978) with extremal quantile regression (Chernozhukov 2005), in the sense that it offers a way to model both moderate and high quantiles. Quantile regression (Koenker 2005) allows for each τ th conditional quantile to have its own slope $\boldsymbol{\beta}_\tau$, according to the following linear specification

$$F^{-1}(\tau | \mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}_\tau, \quad 0 < \tau < 1. \quad (6)$$

In the same way that high empirical quantiles fail to extrapolate into the tail of a distribution, the standard version of quantile regression in (6) is unable to extrapolate into the tail of the conditional response.

We describe next Bayesian Lasso modeling and tackle inference for the first version of our model. For parsimony, below we will focus on the version of the model that sets $q = 1$ in (2), so that $\kappa(\mathbf{x}) = \kappa_1(\mathbf{x}^T \boldsymbol{\beta})$. We underscore that what is studied in the next section applies with some minor modifications to the full model presented above; the focus on $q = 1$ (to which Examples 1–2 apply) in the next section eases however notation, and it obviates the need for discussing again the identifiability issues that we have alluded to already in Example 3.

2.3. REGULARIZATION AND BAYESIAN INFERENCE

Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ be a random sample from $F(\mathbf{x}, y)$. We propose a Bayesian Lasso-type specification for the model in (1) so as to regularize the log likelihood. Specifically, let $G_{\kappa(\mathbf{x})} \in \mathcal{G}$, with $\kappa(\mathbf{x})$ being a function of \mathbf{x} , so that the likelihood becomes

$$L(\boldsymbol{\beta}, \sigma, \xi) = \prod_{i=1}^n h(y_i) g_{\kappa(\mathbf{x}_i)}\{H(y_i)\},$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q)$, $h(y) = \sigma^{-1}(1 + \xi y/\sigma)_+^{-1/\xi-1}$ is the density of the generalized Pareto distribution, $g_{\kappa(\mathbf{x})} = dG_{\kappa(\mathbf{x})}/dy$, and $(a)_+ = \max(0, a)$ is the positive part function. In a Bayesian context, the constrained optimization problem underlying the Lasso can be equivalently rewritten as the posterior mode of regression parameters, provided that one assumes that those are (a priori) independent with identical Laplace priors (i.e., double exponential), that is, $\pi(\boldsymbol{\beta}) \propto \prod_{j=1}^q e^{-\lambda/2 |\beta_j|}$; see Tibshirani (1996) and (Reich and Ghosh (2019), Section 4.2.3). Following Park and Casella (2008), we assume a Gamma prior on λ^2 . The hierarchical representation of the first version of our Bayesian Lasso conditional EGPD for a heavy-tailed response is thus:

Extreme Value Bayesian Lasso for the Conditional Lower Values and Tail (1st version)

1) Likelihood

$$y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \sigma, \xi, \lambda^2 \sim \text{EGPD}_G[\kappa(\mathbf{x}_i), \sigma, \xi], \quad i = 1, \dots, n,$$

$$\kappa(\mathbf{x}) = k(\mathbf{x}^T \boldsymbol{\beta}). \quad (7)$$

2) Priors

$$\begin{cases} \boldsymbol{\beta}_j \mid \lambda \stackrel{\text{iid}}{\sim} \text{Laplace}(\lambda), \\ \lambda^2 \sim \text{Gamma}(a_\lambda, b_\lambda), \\ \sigma \sim \text{Gamma}(a_\sigma, b_\sigma), \\ \xi \sim \pi_\xi. \end{cases}$$

When the goal is to set a prior on the space of heavy-tailed distributions, then it may be sensible to set $\pi_\xi = \text{Gamma}(a_\xi, b_\xi)$, whereas $\pi_\xi = \text{N}(\mu_\xi, \sigma_\xi)$ may be sensible if all domains attraction are equally likely a priori. Since the posterior has no closed-form expression, we resort to Markov Chain Monte Carlo (MCMC) methods to sample from the posterior.

A shortcoming of the first version of the model is that it only allows for the effect of covariates on the lower values, but not on the tail; this is a consequence of the fact that only the part of the model that drives the lower values, $\kappa(\mathbf{x})$, is indexed by a covariate. This issue is addressed in the next section.

2.4. EXTENSIONS FOR COVARIATE-ADJUSTED TAIL

In practice some covariates can be significant for the lower values but not for the tail—or the other way around. Thus, we extend the specification from Sects. 2.2–2.3 by also allowing parameters underlying the tail to depend on covariates. Specifically, we consider the following specification:

$$F(y | \mathbf{x}) = G_{\kappa(\mathbf{x})}\{H(y | \mathbf{x})\}, \quad (8)$$

where $H(y | \mathbf{x})$ is a reparametrized conditional generalized Pareto distribution, with parameters $\nu(\mathbf{x}) = \sigma(\mathbf{x})\{1 + \xi(\mathbf{x})\}$ and $\xi(\mathbf{x})$, that is

$$H(y | \mathbf{x}) = 1 - \left[1 + \frac{\xi(\mathbf{x})\{1 + \xi(\mathbf{x})\}y}{\nu(\mathbf{x})} \right]^{-1/\xi}.$$

Here, $\kappa(\mathbf{x})$ is a function as in (7) whereas $\nu(\mathbf{x}) = \phi(\mathbf{x}^T \boldsymbol{\alpha})$ and $\xi(\mathbf{x}) = \mu(\mathbf{x}^T \boldsymbol{\gamma})$, with ϕ and μ being inverse-link functions. The canonical embodiment of the full version of our model is obtained by specifying $G_{\kappa(\mathbf{x})}(v) = v^{\kappa(\mathbf{x})}$ along with

$$\kappa(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad \nu(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\alpha}), \quad \xi(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\gamma}),$$

assuming $\xi(\mathbf{x}) > 0$. The schematic representation below summarizes our model:

Extreme Value Bayesian Lasso for the Conditional Lower Values and Tail (2nd version)

1) Likelihood

$$y_i | \mathbf{x}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_\alpha \sim \text{EGPD}_G[\kappa(\mathbf{x}_i), \sigma(\mathbf{x}_i), \xi(\mathbf{x}_i)], \quad i = 1, \dots, n,$$

$$\kappa(\mathbf{x}) = k(\mathbf{x}^T \boldsymbol{\beta}), \quad \nu(\mathbf{x}) = \ell(\mathbf{x}^T \boldsymbol{\alpha}), \quad \xi(\mathbf{x}) = \mu(\mathbf{x}^T \boldsymbol{\gamma}). \quad (9)$$

2) Priors

$$\begin{cases} \beta_i & | \lambda_\beta \stackrel{\text{iid}}{\sim} \text{Laplace}(\lambda_\beta), \quad \lambda_\beta \sim \text{Gamma}(a_\beta, b_\beta), \\ \alpha_i & | \lambda_\alpha \stackrel{\text{iid}}{\sim} \text{Laplace}(\lambda_\alpha), \quad \lambda_\alpha \sim \text{Gamma}(a_\alpha, b_\alpha), \\ \gamma_i & | \lambda_\gamma \stackrel{\text{iid}}{\sim} \text{Laplace}(\lambda_\gamma), \quad \lambda_\gamma \sim \text{Gamma}(a_\gamma, b_\gamma). \end{cases}$$

The specification in (8) warrants some remarks:

1. *The need for a reparametrization:* While it would seem natural to simply index (σ, ξ) with a covariate and to proceed as in Sects. 2.2–2.3, similarly to Chavez-Demoulin and Davison (2005) and Cabras and Castellanos (2011), who work in a GPD setting we found that approach to be computationally unstable; in particular, in our case the latter parametrization leads to poor mixing and to small effective sample sizes.
2. *Any potential reparametrization should keep parameters for the lower values and tails separated:* Since we aim to learn which covariates are important for the lower

values and tail, any reparametrization to be made should not mix the parameters for the conditional lower values and tail, i.e., we look for reparametrizations of the type

$$(\kappa, \sigma, \xi) \mapsto \{\kappa, \nu(\sigma, \xi), \xi\} \quad \text{or} \quad (\kappa, \sigma, \xi) \mapsto \{\kappa, \sigma, \zeta(\sigma, \xi)\}.$$

3. *Fisher information*: Parameter orthogonality (Young and Smith 2005, Sect. 9.2) requires the computation of the Fisher information matrix. The canonical EGPD is a particular case of the so-called generalized Feller–Pareto distribution (Kleiber and Kotz 2003), with $(a, b, p, q, r) = (1, \sigma/\xi, \kappa, 1, 1/\xi)$, as

$$h_{\text{GFP}}(y) = \frac{a r y^{a-1}}{b^a B(p, q)} \left\{ 1 + \left(\frac{y}{b} \right)^a \right\}^{-r q - 1} \left[1 - \left\{ 1 + \left(\frac{y}{b} \right)^a \right\}^{-r} \right]^{p-1}, \quad y > 0,$$

where $B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt$, the entries of the Fisher information matrix follow from (Mahmoud and Abd El-Ghafour (2015), Sect. 3); yet the matrix is rather intricate and thus we decided to look for a more sensible way to proceed than attempting to achieve parameter orthogonality.

The reparametrization $(\sigma, \xi) \mapsto (\sigma(1 + \xi), \xi)$ is orthogonal for the case of the GPD for $\xi > -1/2$ (Chavez-Demoulin and Davison 2005), but it is only approximately orthogonal for the EGPD in a neighborhood of $\kappa = 1$. Heuristically, this can be seen by contrasting the likelihood of the canonical EGPD (l) with that of the GPD (l^*); to streamline the argument, we concentrate on the single observation case (y) but the derivations below holds more generally. The starting point is that

$$l(\boldsymbol{\psi}) = l(\kappa, \boldsymbol{\theta}) = l^*(\boldsymbol{\theta}) + \log \kappa + (\kappa - 1) \log\{H(y)\}, \quad (10)$$

where $\boldsymbol{\psi} = (\sigma, \xi, \kappa)$ and $\boldsymbol{\theta} = (\sigma, \xi)$, and thus it follows from (10) that $l(\boldsymbol{\psi}) \approx l^*(\boldsymbol{\theta})$, in a neighborhood of $\kappa = 1$; a similar relation holds for the corresponding Fisher information,

$$I_{\boldsymbol{\psi}} = -\mathbb{E}_{\boldsymbol{\psi}} \left(\frac{\partial^2 l}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^T} \right) = \begin{pmatrix} I_{\boldsymbol{\theta}} & I_{\boldsymbol{\theta}, \kappa} \\ I_{\kappa, \boldsymbol{\theta}} & I_{\kappa} \end{pmatrix}, \quad I_{\boldsymbol{\theta}}^* = -\mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial^2 l^*}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right),$$

where $\mathbb{E}_{\boldsymbol{\theta}}(\cdot) = \int \cdot h_{\boldsymbol{\theta}}(y) dy$ and $\mathbb{E}_{\boldsymbol{\psi}}(\cdot) = \int \cdot h_{\boldsymbol{\psi}}(y) g\{H(y)\} dy$. Indeed, matrix calculus (e.g., Schott 2016, Chapter 9) can be used to show that that

$$I_{\boldsymbol{\theta}} = -\mathbb{E}_{\boldsymbol{\psi}} \left(\frac{\partial^2 l^*}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) - (\kappa - 1) \mathbb{E}_{\boldsymbol{\psi}}(M_{\boldsymbol{\theta}}),$$

where

$$M_{\boldsymbol{\theta}} = \left(H \frac{\partial^2 H}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \frac{\partial H}{\partial \boldsymbol{\theta}} \frac{\partial H}{\partial \boldsymbol{\theta}^T} \right) H^{-2}, \quad H \equiv H(y),$$

thus suggesting that $I_{\boldsymbol{\theta}} \approx I_{\boldsymbol{\theta}}^*$ in a neighborhood of $\kappa = 1$. Despite the fact the reparametrization is only quasi-orthogonal for (σ, ξ) in an EGPD framework, numerical studies reported

in Sect. 3 indicate very good performances in terms of accuracy of fitted regression estimates and QQ-plots of randomized quantile residuals (Dunn and Smyth 1996).

3. SIMULATION STUDY

3.1. SIMULATION SCENARIOS AND ONE-SHOT EXPERIMENTS

In this section, we describe the simulation scenarios and present one-shot experiments so as to illustrate the proposed method; a Monte Carlo study is presented in Sect. 3.2. For now, we focus on describing the setting over which we simulate the data and on discussing the numerical experiments. Code for implementing our methods is available from the Supplementary Material; we used `jags` (Plummer 2019) as we have more experience with it, but `stan` (Gelman et al. 2015) would look like another natural option that could potentially yield faster mixing and more effective sample sizes at little extra programming cost. Throughout we use uninformative Gamma priors, $\text{Gamma}(0.1, 0.1)$, for the hyperparameters λ_k , λ_ν , and λ_ξ .

Data are simulated according to (8), with power carrier $G_{\kappa(\mathbf{x})}(v) = v^{\kappa(\mathbf{x})}$ and with link functions

$$\kappa_{\mathbf{x}} = \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad \nu_{\mathbf{x}} = \exp(\mathbf{x}^T \boldsymbol{\alpha}), \quad \xi_{\mathbf{x}} = \exp(\mathbf{x}^T \boldsymbol{\gamma}).$$

We consider four scenarios, described below, based on the following vectors for the regression coefficients:

- **Scenario 1** Light effects for lower values, light effects for tail: $\boldsymbol{\beta} = \mathbf{a}$; $\boldsymbol{\alpha} = \mathbf{b}$; $\boldsymbol{\gamma} = \mathbf{c}$.
- **Scenario 2** Light effects for lower values, large effects for tail: $\boldsymbol{\beta} = \mathbf{a}$; $\boldsymbol{\alpha} = \mathbf{b}$; $\boldsymbol{\gamma} = 2\mathbf{c}$.
- **Scenario 3** Large effects for lower values, light effects for tail: $\boldsymbol{\beta} = 2\mathbf{a}$; $\boldsymbol{\alpha} = \mathbf{b}$; $\boldsymbol{\gamma} = \mathbf{c}$.
- **Scenario 4** Large effects for lower values, large effects for tail: $\boldsymbol{\beta} = 2\mathbf{a}$; $\boldsymbol{\alpha} = 2\mathbf{b}$; $\boldsymbol{\gamma} = 2\mathbf{c}$.

where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^{10}$ have components

$$\begin{aligned} a_1 = a_3 = 0.3; a_6 = a_{10} = -0.3; \quad b_2 = -0.3; b_5 = b_8 = 0.3; \\ c_1 = c_4 = c_9 = 0.3; c_{10} = -0.3; \end{aligned}$$

and zero otherwise.

For each of the scenarios above, we simulated $n = 500$ observations, $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, from a conditional EGPD with a canonical carrier function; this is about the same number of observations available in the data application of Sect. 4. The covariates are simulated from independent standard uniform distributions. Figure 1 shows cross sections of the true and conditional densities for Scenarios 1–4 estimated using our methods.

As can be seen from these figures, the method recovers satisfactorily well the true conditional density—especially keeping in mind the sample size.

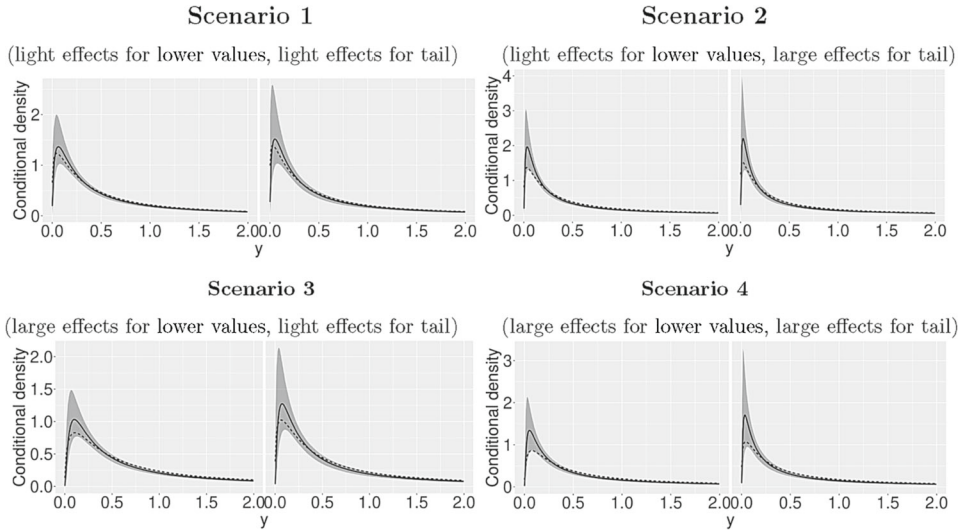


Figure 1. Cross sections of posterior mean conditional density (solid) along with pointwise credible bands against true (dashed) for a one-shot experiment with $n = 500$; the cross sections result from conditioning on $\mathbf{x} = (0.25, \dots, 0.25)$ (left) and $\mathbf{x} = (0.50, \dots, 0.50)$ (right).

3.2. MONTE CARLO SIMULATION STUDY

To assess the finite-sample performance of the proposed methods, we now present the results of a Monte Carlo simulation study, where we repeat 250 times the one-shot experiments from Sect. 3.1. We consider the different sample sizes of $n = 100$, $n = 250$, and $n = 500$.

Figure 2 presents side-by-side boxplots of the coefficient estimates for each scenario for the case $n = 250$, and we can see that the estimates tend to be close to the true values thus suggesting that the proposed methods are able to learn what covariates are significant for the lower values, but not for the tail (and vice versa); further simulation experiments (results not shown) indicate that performance may deteriorate if the number of common effects between left and right tails is large. The coefficient estimates for the cases $n = 100$ and $n = 500$ are reported in the Supplementary Material; as expected, the larger the sample size, the more accurate the estimates. The frequency of variable selection table presented in the Supplementary Material suggests a satisfactory performance of the method in terms of variable selection.

We now move to the conditional density. To compare the fitted conditional density against the true as the sample size increases, we resort to the mean integrated squared error (MISE):

$$\text{MISE} = \mathbb{E} \left[\int \int \{\hat{f}(y | \mathbf{x}) - f(y | \mathbf{x})\}^2 d\mathbf{x} dy \right],$$

where the expectation is taken so to summarize the average behavior of the randomness on the double integral that stems from the posterior mean estimate \hat{f} (Wand and Jones 1995, Sect. 2.3). Figure 3 shows a boxplot of the MISE of each run of the simulation experiment for Scenario 1. As can be seen from Fig. 3, MISE tends to decrease as the sample size

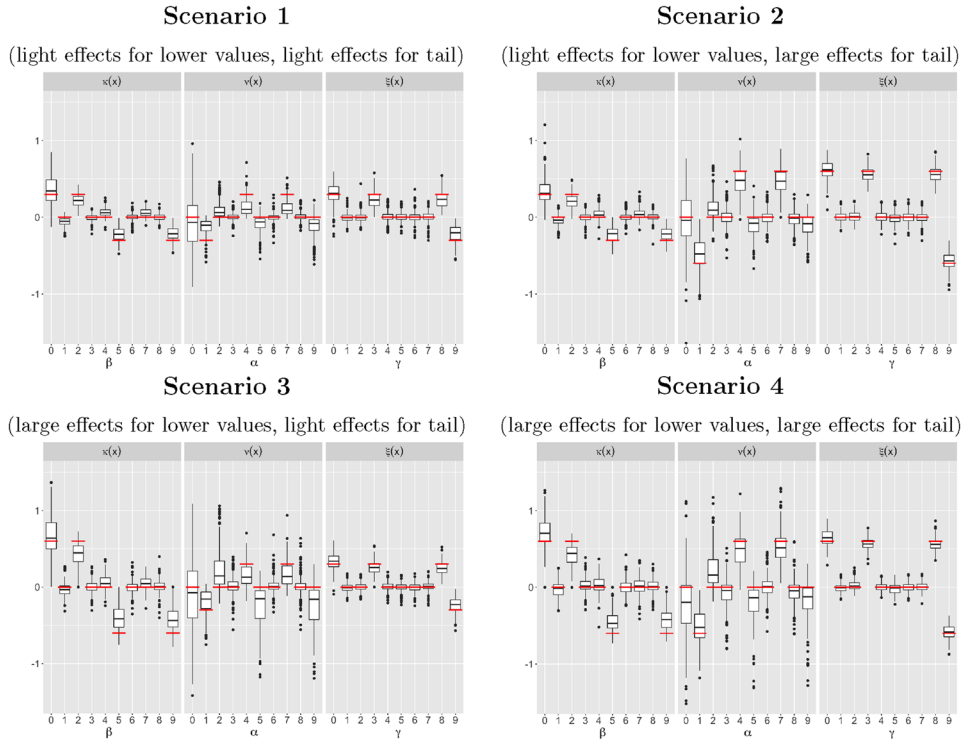


Figure 2. Side-by-side boxplots with regression coefficient estimates for Monte Carlo simulation study ($n = 250$) plotted against the true values (—). The values 1–9 represent the coefficient indices.

increases, thus indicating that a better performance of the proposed methods is to be expected on larger samples; the same holds for the remainder scenarios as can be seen from the boxplots of MISE available in the Supplementary Material. To examine the quality of each fit corresponding to a simulated dataset, $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, we use randomized quantile residuals (Dunn and Smyth 1996) adapted to our model, that is, $\{\varepsilon_i\}_{i=1}^n = [\Phi^{-1}\{\hat{F}(y_i|\mathbf{x}_i)\}]_{i=1}^n$, where Φ^{-1} is the quantile function of the standard Normal distribution and \hat{F} is the posterior mean estimate of the conditional distribution function. Figure 4 depicts the corresponding posterior mean QQ-plot of randomized quantile residuals against the theoretical standard normal quantiles, and it evidences an acceptably good fit of the model for Scenarios 1–4. Finally, to supplement the analysis, we also report in the Supplementary Material numerical experiments that tend to support the claim that the performance of the proposed methods is relatively satisfactory at separating covariate effects for the lower values and tail (Sect. 2.1 of Supplementary Material), and that this also holds in a misspecified setting. Despite the overall good performance, in all fairness not everything is perfect; specifically, our supporting numerical experiments reveal that: i) the correspondence between “non-constant” effects on lower quantiles and coefficients of $\kappa(\mathbf{x})$ is weaker for Scenario 3; ii) bias appears on the lower conditional quantile function estimates, under misspecification.

Figure 3. Side-by-side boxplots of MISE on the log scale for Monte Carlo simulation study for Scenario 1.

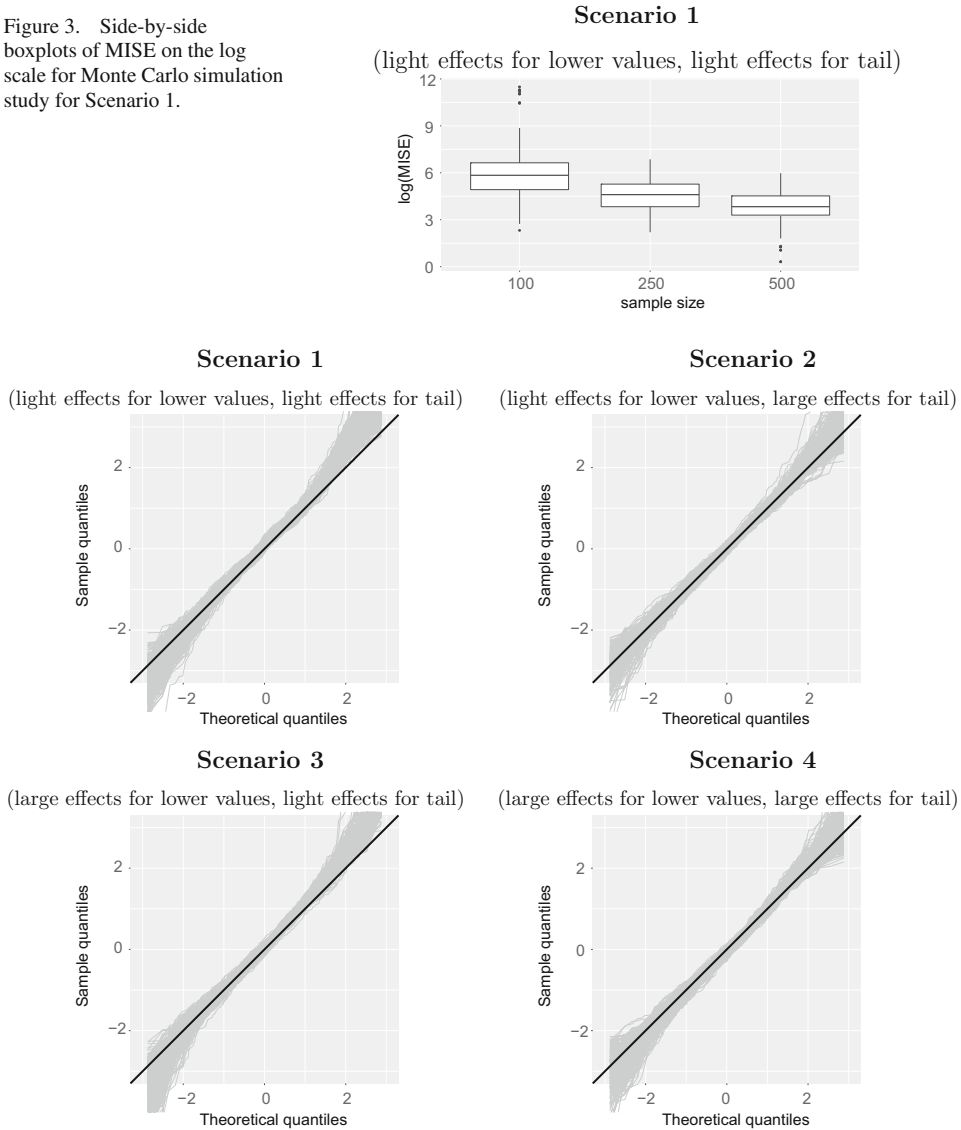


Figure 4. QQ-plots of randomized quantile residuals for Monte Carlo simulation study; each trajectory corresponds to a posterior mean QQ-plot from each simulated dataset.

4. DRIVERS OF MODERATE AND EXTREME RAINFALL IN MADEIRA

4.1. MOTIVATION, APPLIED CONTEXT, AND DATA DESCRIPTION

We now showcase the proposed methodology with a climatological illustration with data from Funchal, Madeira (Portugal); the island of Madeira is an archipelago of volcanic origin located in the Atlantic Ocean about 900km southwest of mainland Portugal. Prior to fitting the proposed model, we start by providing some background on the scientific problem of

interest and by describing the data. Madeira has suffered a variety of extreme rainfall events over the last two centuries, including the flash floods of October 1803 (800–1000 casualties) and those of February 2010—the latter with a death toll of 45 people (Fragoso et al. 2012; Santos et al. 2017) and with an estimated damage of 1.4 billion Euros (Baioni et al. 2011). Such violent rainfall events are often followed by damaging landslides events, including debris, earth, and mud flows.

To analyze such rainfall events in Funchal, Madeira, we have gathered data from the National Oceanic and Atmospheric Administration (www.noaa.gov). Specifically, we have gathered total monthly precipitation (.01 inches), as well as the following potential drivers for extreme rainfall: Atlantic multi-decadal Oscillation (AMO), El Niño-Southern Oscillation (expressed by NINO34 index) (ENSO), North Pacific Index (NP), Pacific Decadal Oscillation (PDO), Southern Oscillation Index (SOI), and North Atlantic Oscillation (NAO). The sample period covers January 1973 to June 2018, thus including episodes such as the violent flash floods of 1979, 1993, 2001, and 2010 (Baioni et al. 2011). After eliminating the dry events (i.e., zero precipitation) and the missing precipitation data (two observations), we are left with a total of 532 observations.

The potential drivers for extreme rainfall above have been widely examined in the climatological literature, mainly on large landmasses. In particular, it has been suggested that in North America ENSO, PDO, and NAO play a key role governing the occurrence of extreme rainfall events (Kenyon and Hegerl 2010; Zhang et al. 2010; Whan and Zwiers 2017); yet for the UK, while NAO is believed to impact the occurrence of extreme rainfall events, no influence of PDO nor AMO has been detected (Brown 2018). The many peculiarities surrounding Madeira climate (e.g., Azores Anticyclone, Canary Current, Gulf Stream, *etc.*), along with the negative impact that flash floods and landslide events produce on the island, motivate us to ask: (i) whether such drivers are relevant for explaining extreme rainfall episodes in Madeira; (ii) whether such drivers are also relevant for moderate rainfall events.

4.2. TRACKING DRIVERS OF MODERATE AND EXTREME RAINFALL

One of the goals of the analysis is to use the lenses of our model so to learn what are the drivers of moderate and extreme rainfall in Funchal. To conduct such inquiry, we use the full model from Sect. 2.4 [see Eq. (9)], with power carrier $G_{\kappa(\mathbf{x})}(v) = v^{\kappa(\mathbf{x})}$ and with link functions

$$\kappa(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}), \quad v(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\alpha}), \quad \xi(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\gamma}.$$

Covariates have been standardized and we used a flat Normal prior, $N(0, 100^2)$, for the intercepts, and uninformative Gamma priors, $\text{Gamma}(0.1, 0.1)$, for the hyperparameters $\lambda_k, \lambda_v, \lambda_\xi$; here an identity link was used for $\xi(\mathbf{x})$ as fitting a GPD to the response via likelihood methods suggested no compelling evidence in favor of an heavy-tailed response. After a burn-in period of 10 000 iterations, we collected a total of 20 000 posterior samples. The results of MCMC convergence diagnostics were satisfactory; in particular, the effective sample sizes were acceptably high, ranging from about 1000 to 5000, and all values of Geweke's diagnostic were satisfactory, ranging from about -3.2 to 3.0 . Figure 5 depicts

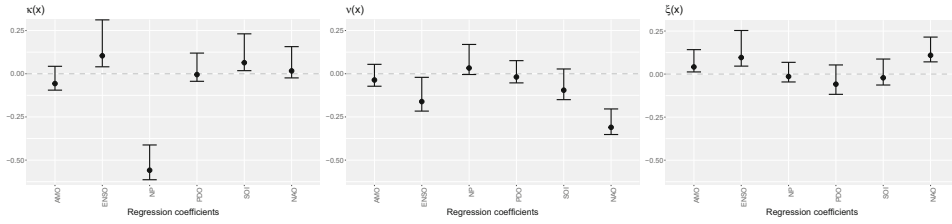
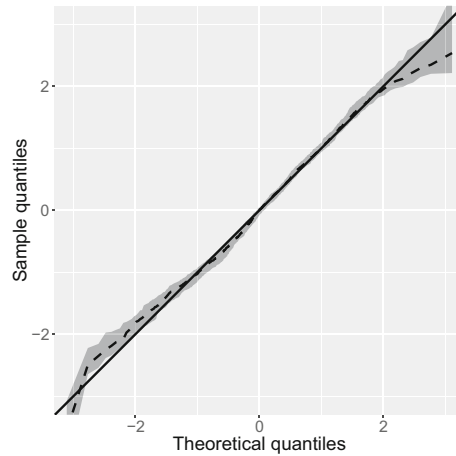


Figure 5. Credible intervals for regression coefficients for inverse-link functions; the dots (●) represent the posterior means and the dashed line represents the reference line $\beta = 0$.

Figure 6. QQ-plot of randomized quantile residuals; the dashed line represents the posterior mean plotted along with pointwise credible bands.



the credible intervals for each regression coefficient. To assess the fit of the proposed model we resort once more to randomized quantile residuals (Dunn and Smyth 1996). Figure 6 evidences an acceptably good fit of the model; while the pointwise bands in Fig. 6 are narrow, they result from acceptably high effective sample sizes as mentioned above.

Figure 5 (middle and right panels) suggests that the key drivers for extreme rainfall in Funchal are NAO and ENSO, along with a possible modest influence of AMO on the magnitude of extreme rainfall events. To put it differently, such results hint that a higher NAO, ENSO, and AMO tend to be associated with extreme rainfall episodes in Funchal. Such findings are thus reasonably in line with those of Kenyon and Hegerl (2010), Zhang et al. (2010), and Whan and Zwiers (2017) for North America. Yet, similarly to Brown (2018), we find no relevant impact of PDO, given the rest, on extreme rainfall spells. Interestingly, of all these potential drivers for extreme rainfall, only ENSO seems to be statistically associated with moderate rainfall; see Fig. 5 (left). Also, NP is the most relevant driver of moderate rainfall, and yet the analysis suggests it plays no role on extreme rainfall. Additionally, Fig. 5 (left) suggests that an increase in NP leads to a heavier left tail (i.e., dryer months), whereas an increase in ENSO leads to less mass close to zero (i.e., rainy months). Finally, to supplement the analysis, we also present conditional quantiles in the Supplementary Material so to directly assess how rainfall itself is impacted by all these potential drivers.

5. CLOSING REMARKS

In this paper, we have introduced a Lasso-type model for the lower values and tail of a possibly heavy-tailed response. The proposed model: i) bypasses the need for threshold selection ($u_{\mathbf{x}}$); ii) models both the conditional lower values and conditional tails; iii) it is naturally tailored for variable selection. In addition, contrarily to GPD-based approach of [Davison and Smith \(1990\)](#) the proposed model does not suffer from lack of threshold stability, as indeed our model does not depend on a threshold; the fact that the Davison–Smith approach does not retain the threshold stability property is known since ([Eastoe and Tawn 2009](#), Sect. 2.2), and it implies that selection of covariates in the parameters of the GPD is not invariant to threshold choice. Yet, other approaches that circumvent the threshold stability issue are available beyond our model, including an inhomogeneous point process formulation for extremes ([Coles 2001](#), Chapter 7). Interestingly, the proposed model can be regarded as a Lasso-type model for quantile regression that is tailored for both the lower values and for extrapolating into the tails. As a by-product, our paper has links with the Bayesian literature on Bayesian distributional regression ([Umlauf and Kneib 2018](#)), and with the recent paper of [Groll et al. \(2019\)](#).

Despite the considerations above on threshold selection, there is no “free lunch”; indeed, here the parameters of the model and the choice of the carrier dictate the behavior of the quantiles in the whole range of the distribution, while the impact of some parameters is smaller in the lower or upper tail. Yet, competing EGPD models that stem from different choices for the carrier function can be formally compared and ranked via standard model selection criteria, such as log pseudo-marginal likelihood (e.g., [Geisser and Eddy 1979](#); [Gelfand and Dey 1994](#)).

Some final comments on future research are in order. A version of our model that includes additionally a regression model for point masses at zero would be natural for a variety of contexts, such as for modeling long periods without rain or droughts. Semicontinuous responses that consist of a mixture of zeros and a continuous positive distribution are indeed common in practice (e.g., [Olsen and Schafer 2001](#)). Another natural avenue for future research would endow the model with further flexibility by resorting to a generalized additive model, where the smooth function of each covariate is modeled using B-spline basis functions. The latter extension would however require a group lasso ([Yuan and Lin 2006](#)), shrinking groups of regression coefficients (per smooth function) toward zero. While we focus here on modeling positive random variables, another interesting extension of the proposed model would consider the left endpoint itself as a parameter rather than fixing it at zero. Finally, from an inference perspective it would also seem natural defining a semiparametric Bayesian version of our model that would set a prior directly on the covariate-specific carrier function $G_{\mathbf{x}}$ rather than specifying a power carrier function in advance, and to model the conditional density as $F(y | \mathbf{x}) = G_{\mathbf{x}}\{H(y | \mathbf{x})\}$. This approach would however require setting a prior on the space \mathcal{G} of all carrier functions so as to ensure that $G_{\mathbf{x}}$ obeys Assumptions A–C, for all \mathbf{x} . While priors on spaces of functions are an active field of research ([Ghosal and Van der Vaart 2015](#)), definition of priors over \mathcal{G} is to our knowledge an open problem; a natural line of attack to construct such a prior would be via random Bernstein polynomials ([Petrone and Wasserman 2002](#)), with weight constraints derived from Lemma 1 of [Tencaliec et al.](#)

(2019); indeed, by modeling the carrier function with a Bernstein polynomial basis a higher level of flexibility could be achieved in comparison to that offered by Examples 1–3.

ACKNOWLEDGEMENTS

We are grateful to the Editor, the Associate Editor, and two Referees for their insightful and constructive remarks on an earlier version of the paper. We thank participants of IWSM (International Workshop on Statistical Modeling) 2019 and of EVA 2019 for constructive comments and discussions. We also thank, without implicating, Vanda Inácio de Carvalho, Ioannis Papastathopoulos, Philippe Naveau, Antónia Turkman, and Feridun Turkman for helpful comments and fruitful discussions. This work was partially supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal), through the projects PTDC/MAT-STA/28649/2017 and UID/MAT/00006/2020.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received October 2020. Revised August 2021. Accepted August 2021. Published Online November 2021.]

REFERENCES

- Arnold BC, Press SJ (1989) Bayesian estimation and prediction for Pareto data. *J Am Stat Assoc* 84:1079–1084
- Baioni D, Castaldini D, Cencetti C (2011) Human activity and damaging landslides and floods on Madeira Island. *Nat Hazards Earth Syst Sci* 11:3035–3046
- Behrens CN, Lopes HF, Gamerman D (2004) Bayesian analysis of extreme events with threshold estimation. *Stat Modell* 4:227–244
- Beirlant J, Goegebeur Y, Segers J, Teugels J (2004) *Statistics of extremes: Theory and applications*. Wiley, Hoboken
- Bhattacharya A, Pati D, Pillai NS, Dunson DB (2015) Dirichlet–Laplace priors for optimal shrinkage. *J Am Stat Assoc* 110:1479–1490
- Brown SJ (2018) The drivers of variability in UK extreme rainfall. *Int J Climatol* 38:e119–e130
- Cabras S, Castellanos ME (2011) A Bayesian approach for estimating extreme quantiles under a semiparametric mixture model. *ASTIN Bull* 41:87–106
- Carreau J, Bengio Y (2009) A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes* 12:53–76
- Carvalho CM, Polson NG, Scott JG (2010) The horseshoe estimator for sparse signals. *Biometrika* 97:465–480
- Castellanos ME, Cabras S (2007) A default Bayesian procedure for the generalized Pareto distribution. *J Stat Plann Inference* 137:473–483
- Chavez-Demoulin V, Davison AC (2005) Generalized additive modelling of sample extremes. *J R Stat Soc Ser C (Appl Stat)* 54:207–222
- Chavez-Demoulin V, Embrechts P, Hofert M (2016) An extreme value approach for modeling operational risk losses depending on covariates. *J Risk Insur* 83:735–776
- Chernozhukov V (2005) Extremal quantile regression. *Ann Stat* 33:806–839
- Coles S (2001) *An introduction to statistical modeling of extreme values*. Springer, London

- Davison, AC, Smith R L (1990), Models for exceedances over high thresholds (with Discussion). *J R Stat Soc Ser B (Stat Methodol)*, 393–442
- de Zea Bermudez P, Kotz S (2010) Parameter estimation of the generalized Pareto distribution-Part II. *J Stat Plann Inference* 140:1374–1388
- de Zea Bermudez P, Turkman MA (2003) Bayesian approach to parameter estimation of the generalized Pareto distribution. *Test* 12:259–277
- do Nascimento F. F., Gamerman D., Lopes H. F. (2012) A semiparametric Bayesian approach to extreme value estimation. *Stat Comput* 22:661–675
- Dunn PK, Smyth GK (1996) Randomized quantile residuals. *J Comput Graph Stat* 5:236–244
- Eastoe EF, Tawn JA (2009) Modelling non-stationary extremes with application to surface level ozone. *J R Stat Soc Ser C (Appl Stat)* 58:25–45
- Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, New York
- Fragoso M, Trigo R, Pinto J, Lopes S, Lopes A, Ulbrich S, Magro C (2012) The 20 February 2010 Madeira flash-floods: Synoptic analysis and extreme rainfall assessment. *Nat Hazards Earth Syst Sci* 12:715–730
- Frigessi A, Haug O, Rue H (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* 5:219–235
- Geisser S, Eddy WF (1979) A predictive approach to model selection. *J Am Stat Assoc* 74:153–160
- Gelfand AE, Dey DK (1994) Bayesian model choice: asymptotics and exact calculations. *J R Stat Soc Ser B (Stat Methodol)* 56:501–514
- Gelman A, Lee D, Guo J (2015) *Stan*: A probabilistic programming language for Bayesian inference and optimization. *J Edu Behav Stati* 40:530–543
- George EI, McCulloch RE (1993) Variable selection via Gibbs sampling. *J Am Stat Assoc* 88:881–889
- Ghosal S, Van der Vaart AW (2015) Fundamentals of nonparametric Bayesian inference. Cambridge University Press, Cambridge
- Groll A, Hambuckers J, Kneib T, Umlauf N (2019) LASSO-type penalization in the framework of generalized additive models for location, scale and shape. *Comput Stat Data Anal* 140:59–73
- Huser R, Genton MG (2016) Non-stationary dependence structures for spatial extremes. *J Agric Biol Environ Stat* 21:470–491
- Kenyon J, Hegerl GC (2010) Influence of modes of climate variability on global precipitation extremes. *J Climate* 23:6248–6262
- Kleiber C, Kotz S (2003) Statistical size distributions in economics and actuarial sciences. Wiley, New York
- Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
- Koenker R, Bassett G. (1978), Regression quantiles, *Econometrica*, 33–50
- MacDonald A, Scarrott C, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. *Comput Stat Data Anal* 55:2137–2157
- Mahmoud MR, Abd El-Ghafour AS (2015) Fisher information matrix for the generalized Feller–Pareto distribution. *Commun Stat Theory Methods* 44:4396–4407
- Naveau P, Huser R, Ribereau P, Hannart A (2016) Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resour Res* 52:2753–2769
- Olsen MK, Schafer JL (2001) A two-part random-effects model for semicontinuous longitudinal data. *J Am Stat Assoc* 96:730–745
- Papastathopoulos I, Tawn JA (2013) Extended generalised Pareto models for tail estimation. *J Stat Plann Inference* 143:131–143
- Park T, Casella G (2008) The Bayesian Lasso. *J Am Stat Assoc* 103:681–686
- Petrone S, Wasserman L (2002) Consistency of Bernstein polynomial posteriors. *J R Stat Soc Ser B (Stat Methodol)* 64:79–100

- Plummer M (2019), `rjags`: Bayesian graphical models using MCMC. R package version 4-10, <http://CRAN.R-project.org/package=rjags>
- Polson NG, Sokolov V (2019) Bayesian regularization: From Tikhonov to horseshoe. *Computational Statistics, Wiley Interdisciplinary Reviews*, p e1463
- Reich BJ, Ghosh SK (2019) *Bayesian statistical methods*. Chapman & Hall/CRC, Boca Raton
- Resnick SI (1971) Tail equivalence and its applications. *J Appl Probab* 8:136–156
- Santos M, Santos JA, Fragoso M (2017) Atmospheric driving mechanisms of flash floods in Portugal. *Int J Climatol* 37:671–680
- Schott JR (2016) *Matrix analysis for statistics*. Wiley, New York
- Stein ML (2020) Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, 1–31
- Stein ML (2021) A parametric model for distributions with flexible behavior in both tails. *Environmetrics* 32:e2658
- Tencaliec P, Favre A-C, Naveau P, Prieur C, Nicolet G (2019) Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, e2582
- Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B (Stat Methodol)* 58:267–288
- Umlauf N, Kneib T (2018) A primer on Bayesian distributional regression. *Stat Modell* 18:219–247
- Villa C (2017) Bayesian estimation of the threshold of a generalised Pareto distribution for heavy-tailed observations. *Test* 26:95–118
- Wand MP, Jones MC (1995) *Kernel smoothing*, 1st edn. Chapman & Hall, London
- Wang H, Tsai C-L (2009) Tail index regression. *J Am Stat Assoc* 104:1233–1240
- Whan K, Zwiers F (2017) The impact of ENSO and the NAO on extreme winter precipitation in North America in observations and regional climate models. *Climate Dyn* 48:1401–1411
- Young GA, Smith RL (2005) *Essentials of statistical inference*. Cambridge University Press, Cambridge
- Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. *J R Stat Soc Ser B (Stat Methodol)* 68:49–67
- Zhang X, Wang J, Zwiers FW, Groisman PY (2010) The influence of large-scale climate variability on winter maximum daily precipitation over North America. *J Climate* 23:2902–2915
- Zhang Y. D, Naughton B. P, Bondell H. D, Reich B. J. (2020), Bayesian regression using a prior on the model fit: The R2–D2 shrinkage prior. *J Am Stat Assoc* 1–13