

# Practical strategies for generalized extreme value-based regression models for extremes

Daniela Castro-Camilo<sup>1</sup>  | Raphaël Huser<sup>2</sup>  | Håvard Rue<sup>2</sup> 

<sup>1</sup>School of Mathematics and Statistics,  
University of Glasgow, Glasgow, UK

<sup>2</sup>Computer, Electrical and Mathematical  
Sciences and Engineering (CEMSE)  
Division, King Abdullah University of  
Science and Technology (KAUST),  
Thuwal, Saudi Arabia

## Correspondence

Daniela Castro-Camilo, School of  
Mathematics and Statistics, University of  
Glasgow, Glasgow, UK.

Email:

[daniela.castrocamilo@glasgow.ac.uk](mailto:daniela.castrocamilo@glasgow.ac.uk)

## Funding information

King Abdullah University of Science and  
Technology (KAUST) Office of Sponsored  
Research (OSR) under Award No.  
OSR-CRG2017-3434

## Abstract

The generalized extreme value (GEV) distribution is the only possible limiting distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. As such, it has been widely applied to approximate the distribution of maxima over blocks. In these applications, GEV properties such as finite lower endpoint when the shape parameter  $\xi$  is positive or the loss of moments due to the magnitude of  $\xi$  are inherited by the finite-sample maxima distribution. The extent to which these properties are realistic for the data at hand has been widely ignored. Motivated by these overlooked consequences in a regression setting, we here make three contributions. First, we propose a blended GEV (bGEV) distribution, which smoothly combines the left tail of a Gumbel distribution (GEV with  $\xi = 0$ ) with the right tail of a Fréchet distribution (GEV with  $\xi > 0$ ). Our resulting distribution has, therefore, unbounded support. Second, we proposed a principled method called property-preserving penalized complexity ( $P^3C$ ) prior to decide on the existence of the GEV distribution first and second moments a priori. Third, we propose a reparameterization of the GEV distribution that provides a more natural interpretation of the (possibly covariate-dependent) model parameters, which in turn helps define meaningful priors. We implement the bGEV distribution with the new parameterization and the  $P^3C$  prior approach in the R-INLA package to make it readily available to users. We illustrate our methods with a simulation study that reveals that the GEV and bGEV distributions are comparable when estimating the right tail under large-sample settings. Moreover, some small-sample settings show that the bGEV fit slightly outperforms the GEV fit. Finally, we conclude with an application to  $\text{NO}_2$  pollution levels in California that illustrates the suitability of the new parameterization and the  $P^3C$  prior approach in the Bayesian framework.

## KEYWORDS

blended generalized extreme value distribution, block maxima, extreme value theory, generalized extreme value distribution, INLA, property-preserving penalized complexity prior

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Environmetrics* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The generalized extreme value distribution (GEVD) is a three parameters family that describes the asymptotic behavior of properly renormalized maxima of a sequence of independent and identically distributed random variables. It is parameterized in terms of a location  $\mu \in \mathbb{R}$ , a scale  $\sigma > 0$  and a shape parameter  $\xi \in \mathbb{R}$  that controls the type of tail of the GEVD. Indeed, if  $\xi$  is zero, the GEVD has unbounded (parameter-free) support, whereas if  $\xi$  is positive, the limiting distribution is heavy-tailed with infinite upper endpoint but finite lower endpoint that depends on the parameters. If  $\xi < 0$ , then the short-tailed GEVD has a finite upper endpoint that also depends on the parameter. We can see then that when  $\xi \neq 0$ , the GEVD does not obey the general regularity conditions for likelihood-based inference (Stuart et al., 2004); see Smith (1985) for an in-depth study of maximum likelihood inference for a class of irregular models that includes the GEVD. Moreover, by using the asymptotic GEVD as an approximation for the distribution of maxima over finite blocks, GEV properties, such as finite lower bound in the case  $\xi > 0$  and finite upper bound in the case  $\xi < 0$ , are inherited by the original maxima distribution, which might not have bounded support.

Despite the above features, many authors have shown the practical usefulness of the GEV family as an approximation to the distribution of block maxima. For instance, Broussard and Booth (1998) use the GEV distribution to analyse the behavior of extreme losses in a German stock index, whereas Bruun and Tawn (1998) assume a GEVD to estimate the probability of coastal flooding. Extensions using regression models to include covariate effects have been used in El Adlouni et al. (2007), El Adlouni and Ouarda (2009), and Cannon (2010), among others. Flexible parameter smoothing using the class of vector generalized linear and additive models was proposed in Yee and Stephenson (2007), and this framework was recently exploited by Zhong et al. (2022) for an application to heatwave hazard assessment in Europe. Further extensions to combine spatial information across locations have been analyzed in Casson and Coles (1999), while Westra et al. (2013) studied the effect of climate change on global precipitation annual maxima, and Jóhannesson et al. (2021) embedded space-time covariates and spatially correlated random effects within GEV parameters to predict extreme river flows over the whole U.K. territory. Additionally, Huang et al. (2016) conducted a detailed study on the usefulness and potential limitations of the GEVD to model seasonal maximum and minimum temperatures from a millennial-scale climate simulation. From a theoretical perspective, Stein (2017) introduces two new results on the limiting distribution of block maxima, motivated by the study of annual temperature extremes. The need for such new results comes from the upper bound imposed by the GEV when it is used as an approximation to the distribution of properly renormalized annual maximum temperatures. Indeed, Stein (2017) argues that the existence of an upper bound means that for the GEVD to be a suitable approximation of the annual maximum temperature distribution, we would need a theoretical result that applies to the maximum of a large number of random variables with varying upper bounds. To solve this artefact created by the GEVD, Stein (2017) develops a framework based on triangular arrays, where the GEVD appears as a particular case.

As in Stein (2017), here we focus on an artefact created by the GEV distribution when it is used as an approximation to the distribution of finite block maxima. Our focus, though, is on the case where the shape parameter is positive and therefore, the limiting GEV distribution has a lower bound. To illustrate this artefact, consider  $X_1, \dots, X_n$  independent and identically distributed (i.i.d.) according to a standard Cauchy distribution  $F$ , which has support in  $\mathbb{R}$ . Then, if  $M_n = \max\{X_1, \dots, X_n\}$  and  $M_n^* = (M_n - b_n)/a_n$  with sequences  $a_n > 0$  and  $b_n \in \mathbb{R}$ , one has that  $\Pr(M_n \leq z) = F^n(z)$  and  $\Pr(M_n^* \leq z) = F^n(a_n z + b_n)$  (where  $a_n$  and  $b_n$  can be chosen as  $a_n = n\pi^{-1}$  and  $b_n = 0$ ). This implies that both  $M_n$  and its standardized version  $M_n^*$  have also support in  $\mathbb{R}$ . It can be shown (see, e.g., Schmiedt, 2016) that in this case,  $\Pr(M_n^* \leq z) \rightarrow G(z) = \exp(-1/z)$ ,  $z > 0$ , as  $n \rightarrow \infty$ , that is, properly renormalized maxima of a sequence of i.i.d. standard Cauchy variables with support in  $\mathbb{R}$  converge to a unit Fréchet random variable, which has support in  $(0, \infty)$ . Therefore, the GEV limit imposes an artificial bounded support to the finite-sample maxima of Cauchy random variables, whose actual finite- $n$  distribution may have non-negligible mass over  $(-\infty, 0)$ . Note that many other similar theoretical examples can be found to illustrate the same mismatch between the finite- $n$  and asymptotic supports of block maxima (e.g., maxima of exponential, beta, or Weibull random variables, just to name a few examples).

The artificial bound imposed by the GEV approximation is particularly troublesome when the GEV parameters are covariate-dependent, that is,  $\mathbf{y}(\mathbf{x}) \sim \text{GEV}(\mu(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x}))$ , for  $\mathbf{x} \in \mathbb{R}^d$ ,  $d \geq 1$ . In practical applications, the components of  $\mathbf{y}(\mathbf{x})$  are usually constructed as the block-maxima of a certain variable, say  $\mathbf{z}(\mathbf{x})$ , for a given block size. In this case, the GEV support,  $\{z : 1 + \xi(\mathbf{x})\{z - \mu(\mathbf{x})\}/\sigma(\mathbf{x}) > 0\}$ , is also covariate-dependent. In particular, when  $\xi(\mathbf{x}) > 0$ , the lower bound is  $\mu(\mathbf{x}) - \sigma(\mathbf{x})/\xi(\mathbf{x})$ . A GEVD fitted to data  $\mathbf{y}$  implies that the maximum possible value of  $\mathbf{z}$  should also have a lower bound for any configuration of  $\mathbf{x}$ , even those that have not been observed. Moreover, this lower bound is changing with the covariates that have been observed. Currently, there is no theoretical guarantee that a new configuration of covariates

outside the sample support would not predict a value  $y^*$  below the fitted lower bound. In practice, if we wish to predict with covariates outside the sample support, then the whole model is refitted to account for  $y^*$ . Furthermore, note that if  $y^*$  is observed, this single observation will control the estimates and consequently the lower bound, reducing the importance of the remaining observations. Therefore, following the reasoning of Stein (2017), we could argue that for the GEV to be a reasonable model for  $\mathbf{y}$ , we would need, in principle, a result that applied to maxima of a large number of random variables with varying lower bound.

The examples presented here illustrate that when the GEVD with positive shape is scaled back to be used as an approximation for finite-sample maxima, it introduces a purely theoretical, unrealistic and inconvenient artefact. Furthermore, while this artefact may be considered negligible for the actual model being fitted, it can have major implications for prediction and computing in regression settings. We address this issue by modifying the lower tail of the GEVD with positive shape. Specifically, we construct a blended GEV (bGEV) distribution with support in the whole real line, based on a combination of the Gumbel ( $\xi = 0$ ) and the Fréchet ( $\xi > 0$ ) distributions. Our approach shares similarities with spliced models (see, e.g., Reynkens et al., 2017; Castro-Camilo et al., 2019; Opitz et al., 2018) and mixture models for extremes (Behrens et al., 2004; Carreau & Bengio, 2009; de Melo Mendes & Lopes, 2004; Do Nascimento et al., 2011; Frigessi et al., 2002; Leonelli & Gaman, 2020) in the sense that all of these methods join two distributions. But besides this, our approach is fundamentally different from that of spliced or mixture models. First of all, mixture models blend two or more distributions at the density level. This means that when the mixture weights are not constant in the argument, the definition of the associated distribution function requires the computation of a normalizing constant, which may not have a closed-form expression. Our bGEV model blends the Fréchet and Gumbel distributions at the distribution level so that the resulting density function can straightforwardly be computed and is automatically normalized. Second, the mixture and spliced models cited above are proposed for threshold exceedances. For instance, Castro-Camilo et al. (2019) and Opitz et al. (2018) propose a blend of a Gamma distribution for the bulk and a generalized Pareto distribution (GPD) for the tail, and their approach incorporate the modeling of the exceedance probability. Frigessi et al. (2002) provides a continuous transition between the bulk (described using a Weibull distribution) and GPD-tail models, avoiding the threshold selection associated with fitting a generalized Pareto distribution. As noted by Scarrott and MacDonald (2012), their model is appropriate when there is a lower bound on the support. On the contrary, the goal of our bGEV model is to modify the left tail of the GEV distribution with positive shape parameter while keeping its upper tail behavior.

Note that since the GEVD is the only possible nondegenerate limit to properly renormalized maxima, we do not propose a different limiting model. Instead, we create a distribution that behaves very closely to the GEVD, bypassing the lower bound problem. This means that we expect to get similar results for both distributions in terms of parameter estimates and return level estimation, which is confirmed by our first simulation study in Section 5. The benefit of our proposed bGEV model is that it is numerically more stable and reliable when we want to make predictions. It is also more statistically meaningful in the sense that the support does not depend on parameters.

The bGEVD is expressed in terms of a new parametrization of the GEVD, that provides a more meaningful interpretation of the model's parameters. Indeed, the usual location-scale GEV parametrization loses interpretability when the mean and standard deviation do not exist, which is the case when  $\xi \geq 1$  and  $\xi \geq 0.5$ , respectively. Our proposed bGEV parametrization is based on a GEV quantile  $q_\alpha$  (location) and a quantile range (spread)  $s_\beta = q_{1-\beta/2} - q_{\beta/2}$  ( $0 < \alpha, \beta < 1$ ) whose existence, unlike the mean and standard deviation, are not influenced by the shape parameter. A consequence of this is that we have bGEV location and spread parameters over which we can meaningfully assign priors in the Bayesian setting. In particular, we have that independent priors over the bGEV location and spread parameters induce a joint prior distribution for the original location and scale parameters.

The selection of a prior distribution for the bGEV shape parameter motivates an additional contribution of this paper. As mentioned above, the shape parameter has a major role in determining the type of tail of the GEVD. It is also key in determining important features of the distribution, such as the existence of moments. Indeed, the  $k$ th moment exists if and only if  $\xi < 1/k$  (Muraleedharan et al., 2011). In a regression context where  $\xi$  might depend on covariates, we might have that, for example, the first and second moments are not continuous as a function of the covariates, that is, mean and variance exist only for some configuration of the covariates. The consequences of this issue have not received much attention in the literature. By focusing on these low-order moments, we proposed a principle method to decide a prior on the existence of the first and second moments. Our method is called property-preserving penalized complexity (P<sup>3</sup>C) prior approach, and exploits the framework introduced by Simpson et al. (2017) to ensure desired features of the bGEVD.

To summarize, the goal of this article is threefold: (1) to offer an alternative to the classical GEVD that corrects the left tail, providing a parameter-free support; (2) to propose a reparametrization of the GEVD in terms of a quantile and a

quantile range that eases the interpretation of the model's parameters; and (3) to introduce a principled method to decide the existence of low-order moments a prior.

Our proposed bGEV model is implemented adopting a hierarchical Bayesian framework in the context of latent Gaussian models, using the new parametrization and the P<sup>3</sup>C prior approach. Marginal posterior distributions of interest are computed using the integrated nested Laplace approximation (INLA; Rue et al., 2009). The bGEV model is freely available and efficiently implemented in the R-INLA package (Bivand et al., 2015; Lindgren & Rue, 2015).

The remainder of the article is organized as follows. In Section 2 we describe the new reparametrization of the GEVD, while in Section 3 we describe our modeling approach based on the bGEVD. The concept of property-preserving penalized complexity priors is introduced in Section 4. In Section 5 we present three simulation studies that compare the performance of the GEVD and the bGEVD at estimating return levels and provide insights on the influence of the bGEV model parameters. Section 6 presents the improvements implied by our model using NO<sub>2</sub> concentrations in California. Conclusions and a discussion of our methods are given in Section 7.

## 2 | A REPARAMETRIZATION OF THE GEV DISTRIBUTION

The usual GEV parametrization in terms of location, scale, and shape parameters is somewhat convenient as it resembles well-known location-scale families. Within these families, we usually associate the location and scale parameters with the mean and standard deviation. Nonetheless, in skewed distributions such as the GEVD, the mean and standard deviation are no longer reasonable proxies for the location and scale of the distribution. Moreover, for large enough values of the shape parameter, the GEV mean and variance grow to infinity, which prevents us from interpreting the location-scale GEV parametrization as we usually do in other models. This is particularly troublesome in a regression context where the natural choice is to use a regression model in the location parameter, the scale parameter, or both. In such cases, covariate coefficients may not have a clear meaning if the GEV mean and variance are not defined. This is even more critical in a Bayesian context, as it might not be clear how to assign reasonable priors to GEV parameters. In this section, we propose a reparametrization of the GEV that has a broader interpretation for all values of  $\xi \in \mathbb{R}$ , where prior information can be assigned to parameters that can still be widely interpreted even in the cases where the first and second moments are not finite. Specifically, following the spirit of Coles and Tawn (1996), we reparametrize the GEVD in terms of new quantile-based location and spread parameters. The new location parameter is the  $\alpha$ -quantile ( $0 < \alpha < 1$ ), denoted  $q_\alpha \in \mathbb{R}$ , while the new spread is the quantile range  $s_\beta = q_{1-\beta/2} - q_{\beta/2} \in \mathbb{R}^+$  ( $0 < \beta < 1$ ). So, for instance, if  $\alpha = \beta = 0.5$ , then the new location  $q_{0.5}$  is the median, and the spread  $s_{0.5}$  is the interquartile range. The GEVD reparametrized in terms of  $q_\alpha$ ,  $s_\beta$  and  $\xi$  can be written as

$$F(y|q_\alpha, s_\beta, \xi) = \exp \left[ - \left\{ \frac{y - q_\alpha}{s_\beta (\ell_{1-\beta/2, \xi} - \ell_{\beta/2, \xi})^{-1}} + \ell_{\alpha, \xi} \right\}_+^{-1/\xi} \right],$$

where  $a_+ = \max(a, 0)$  and for any  $a > 0$ ,  $\ell_{\alpha, \xi} = (-\log a)^{-\xi}$ . Note that the case  $\xi = 0$  simplifies to

$$F(y|q_\alpha, s_\beta) = \exp \left[ - \exp \left\{ - \left( \frac{y - q_\alpha}{s_\beta (\ell_{\beta/2} - \ell_{1-\beta/2})^{-1}} - \ell_\alpha \right) \right\} \right]$$

with  $\ell_\alpha = \log(-\log a)$ , for any  $a > 0$ . There is a one-to-one mapping between  $(q_\alpha, s_\beta, \xi)^\top$  and the parametrization in terms of  $(\mu, \sigma, \xi)^\top$ . For the case  $\xi \neq 0$ , the mapping is given by

$$\mu = q_\alpha - \frac{s_\beta (\ell_{\alpha, \xi} - 1)}{\ell_{1-\beta/2, \xi} - \ell_{\beta/2, \xi}}, \quad \sigma = \frac{\xi s_\beta}{\ell_{1-\beta/2, \xi} - \ell_{\beta/2, \xi}}.$$

The case  $\xi = 0$  is interpreted as the limit when  $\xi \rightarrow 0$ , that is,

$$\mu = q_\alpha + \frac{s_\beta \ell_\alpha}{\ell_{\beta/2} - \ell_{1-\beta/2}}, \quad \sigma = \frac{s_\beta}{\ell_{\beta/2} - \ell_{1-\beta/2}}.$$

Note that the location and spread parameters are still interpretable in this new parametrization even when the mean and variance are undefined. The probabilities  $\alpha$  and  $\beta$  that define the quantiles  $q_\alpha, q_{\beta/2}, q_{1-\beta/2}$  are fixed hyperparameters and should be defined by the user.

### 3 | THE BLENDED GEV MODEL

As mentioned in Section 1, we want to modify the left tail of the GEVD with positive shape parameter to avoid inheriting a left-bounded support for the finite-sample maxima. We want to do this by blending two distributions,  $F(x)$  and  $G(x)$ , at the distribution level, so that the resulting distribution  $H(x)$  resembles the GEVD's right tail, has infinite support, and the resulting density function can straightforwardly be computed and is automatically normalized. Our blended GEV (bGEV) distribution function is defined as

$$H(x|q_\alpha, s_\beta, \xi) = F(x|q_\alpha, s_\beta, \xi)^{p(x)} G(x|\tilde{q}_\alpha, \tilde{s}_\beta)^{1-p(x)}, \tag{1}$$

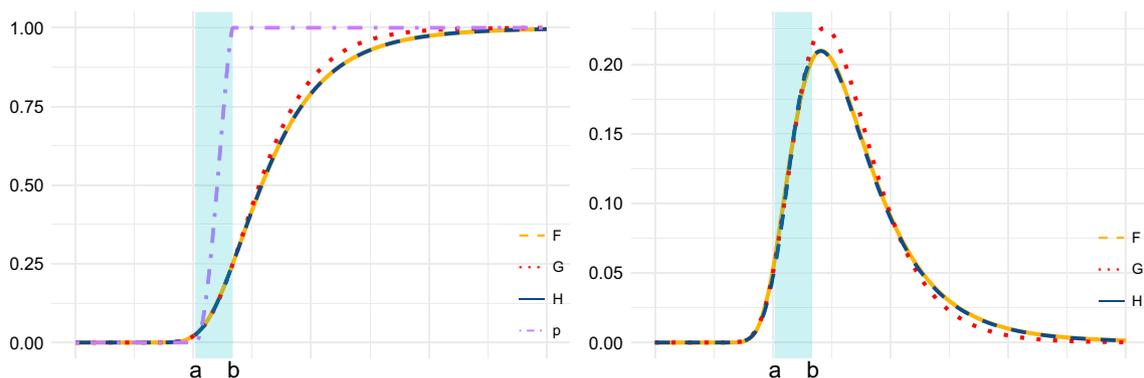
where  $F$  is the Fréchet distribution with location  $q_\alpha$ , spread  $s_\beta$ , and shape  $\xi$ , and  $G$  is the Gumbel distribution with location  $\tilde{q}_\alpha \equiv \tilde{q}_\alpha(q_\alpha, s_\beta)$  and spread  $\tilde{s}_\beta \equiv \tilde{s}_\beta(q_\alpha, s_\beta)$ . The function  $p$  is a weight function that controls the way the distributions  $F$  and  $G$  are blended together.

As per the notation in (1),  $H$  is parameterized solely in terms of  $q_\alpha, s_\beta$ , and  $\xi$ , and not in terms of  $\tilde{q}_\alpha$  and  $\tilde{s}_\beta$ . Indeed, in the following we will see that  $\tilde{q}_\alpha$  and  $\tilde{s}_\beta$  can be completely characterized via  $q_\alpha, s_\beta, \xi$  and a set of hyperparameters, of which the first two are the probabilities  $\alpha$  and  $\beta$ . In principle,  $\alpha$  and  $\beta$  can take any value, but as noted in Section 2, some choices might be easier to interpret than others (e.g., setting  $\alpha = \beta = 0.5$ ). The remaining hyperparameters are related to the weight function  $p$ . The function  $p$  is chosen as the cumulative distribution function of a beta distribution with shape parameters  $c_1 > 0$  and  $c_2 > 0$  ( $F_{\text{Beta}}(\cdot|c_1, c_2)$ ) defined over the *left-tail mixing area*  $[a, b]$ . As the name implies, the left-tail mixing area is where  $F$  and  $G$  are merged; see the blue shaded rectangle in Figure 1. The weight function  $p$  is then defined as  $F_{\text{Beta}}(\cdot|c_1, c_2)$  evaluated at the point  $(x - a)/(b - a)$ , that is,

$$p(x) = p(x; a, b, c_1, c_2) = F_{\text{Beta}}\left(\frac{x - a}{b - a} | c_1, c_2\right). \tag{2}$$

The left-tail mixing area  $[a, b]$  is defined by setting  $a = F^{-1}(p_a)$  and  $b = F^{-1}(p_b)$  for relatively small probabilities  $p_a$  and  $p_b$ . The shape parameters  $c_1$  and  $c_2$  of the weight function  $p$  can take any positive value, but we restrict them to greater than 3. In that case, the second derivative of the log-density of  $H$  is always continuous (see Section A.2 in the Appendix).

Some comments related to our above choices are in order. First, note that the weight function  $p$  in (2) is 0 for  $x \leq a$  and one for  $x \geq b$ , which means that the left tail of  $H$  is equal to the left tail of  $G$  (which is unbounded) and the right tail of  $H$  is equal to the right tail of  $F$ . In other words, the distribution  $H$  has the correct limiting right tail, yet without the inconvenient, left-bounded, parameter-dependent support imposed by the limit. Note also that  $\lim_{\xi \rightarrow 0} H(x) = G(x)$  for all weight functions  $p$ , which means that our proposed distribution reduces to the classical Gumbel case when  $\xi \rightarrow 0$ ,



**FIGURE 1** Illustration of the distribution (left) and density (right) for  $H$  (bGEV),  $F$  (Fréchet), and  $G$  (Gumbel). On the left, the purple dash-dotted line is the weight function  $p$ .

for all  $p$ . This also ensures that the resulting distribution  $H$  is not too far from the classical GEVD when  $\xi \approx 0$ . Second, note that by setting  $a = F^{-1}(p_a)$  and  $b = F^{-1}(p_b)$  we get that  $F(a) = G(a)$ ,  $F(b) = G(b)$  and the resulting distribution  $H$  is continuous. Moreover, we can express the Gumbel parameters  $\tilde{q}_\alpha$  and  $\tilde{s}_\beta$  as

$$\tilde{q}_\alpha = a - \frac{(b-a)(\ell_\alpha - \ell_{p_a})}{\ell_{p_a} - \ell_{p_b}}, \quad \tilde{s}_\beta = \frac{(b-a)(\ell_{\beta/2} - \ell_{1-\beta/2})}{\ell_{p_a} - \ell_{p_b}} \quad (3)$$

which are injective functions of  $q_\alpha$  and  $s_\beta$ . Finally, note that the reparametrization proposed in Section 2 is based on the quantiles  $q_\alpha$ ,  $q_{\beta/2}$ , and  $q_{1-\beta/2}$  of the GEVD. Since the bGEVD is equal to the GEVD only for values above the left-tail mixing area (i.e., only for  $x \geq b$ ), the definition of  $q_\alpha$  and  $s_\beta$  should be constrained to the value of  $b$ . Since  $b$  is also defined as a quantile of the GEVD, the above implies that we should have  $\alpha \geq p_b$  and  $\beta/2 \geq p_b$ . Otherwise,  $q_\alpha$  and  $s_\beta$  would be defined as quantiles of the bGEVD, which are different from those of the GEVD below  $b$ .

Summarizing,  $H$  is completely defined through the parameters  $(q_\alpha, s_\beta, \xi)$  and the hyperparameters  $(\alpha, \beta, p_a, p_b, c_1, c_2)$ . For our data application in Section 6, we set  $\alpha = \beta = 0.5$ , in which case  $H$  is parameterized in terms of the median and the interquartile range of  $F$ . Additionally, we set  $p_a = 0.05$ ,  $p_b = 0.2$ , and  $c_1 = c_2 = 5$ . The choice for  $c_1$  and  $c_2$  leads to a symmetric and computationally convenient weight function.

## 4 | PROPERTY-PRESERVING PENALIZED COMPLEXITY PRIORS

Penalized complexity (PC) priors (Simpson et al., 2017) provide a principled and widely applicable method to specify prior information that is difficult to obtain from expert knowledge. It uses the natural nested structure inherent in many model components to define these model components as flexible extensions of some base model. The method penalizes deviations from this base model by placing an exponential prior on the Kullback–Leibler divergence from the base model. Base models should then be defined for every model component of interest. This task is simplified by noticing that, in practice, model components are completely defined through parameters. Take, for example, a Gaussian random walk of order 2 (a definition can be found in (7)), which is entirely defined by its precision (inverse of its standard deviation). This fact can be used to define a PC prior using as base model a random walk with an infinite precision (see Section 6 for an application of this particular PC prior).

PC priors can also be defined over data parameters, such as the positive shape parameter of the GEV/bGEV distributions. The natural base model, in this case, is a GEV/bGEV with a shape equal to 0 (i.e., a Gumbel distribution). Although it is possible to derive a formula for this PC prior, we can take advantage of the relationship between the GEV and generalized Pareto (GP) distributions and approximate the PC prior of the GEV shape by that of the GP shape. Opitz et al. (2018) derived the PC prior for the GP positive shape parameter using a GP with shape equal to 0 as base model. This PC prior is defined for  $0 \leq \xi < 1$ , thus preventing infinite-mean models, and depends on a penalty parameter  $\lambda > 0$ . It can be expressed as

$$\pi(\xi) = \frac{\lambda}{\sqrt{2}} \exp\left(-\frac{\lambda}{\sqrt{2}} \frac{\xi}{(1-\xi)^{1/2}}\right) \left(\frac{1-\xi/2}{(1-\xi)^{3/2}}\right), \quad \xi \in [0, 1). \quad (4)$$

We can use the PC prior framework to define priors that help preserve specific properties of the model components or the data distribution, such as the existence of moments. This is what we call the property-preserving penalized complexity (P<sup>3</sup>C) prior approach. In the case of the GEVD, the existence of the  $k$ th moment is guaranteed if  $\xi < 1/k$ . Although, in practice, the existence of high-order moments for the GEVD with  $\xi$  strictly positive cannot be guaranteed, we can preserve low-order moments by shrinking the interval of possible values that  $\xi$  can take *a priori*. Specifically, we can preserve the first two moments of the GEVD by conditioning on  $\xi$  being less than 0.5. In practice, this means that we normalize (4) by the corresponding cumulative distribution evaluated at 0.5, that is, using the alternative prior

$$\tilde{\pi}(\xi) = \frac{\pi(\xi)}{\int_0^{0.5} \pi(s) ds}, \quad \xi \in [0, 0.5). \quad (5)$$

A method such as the one presented here is particularly important in prediction problems. Indeed, if the prior for  $\xi$  has positive support in  $[0, u_\xi]$ , then for finite sample sizes, the prediction of a new data point will inherit the moment properties

of  $\xi \in [0, u_\xi]$ . Therefore, if, for instance,  $u_\xi = 1$ , the posterior distribution will assign a positive density to  $\xi \in [0, 1]$ , which means that all predictions for a new data point will have infinite mean and variance once the uncertainty is taken into account, even if the data suggest the opposite.

The P<sup>3</sup>C prior concept that we show here for the shape parameter can be extended to other models' parameters when important model properties are not *continuous* as a function of such parameters. We argue that distributional properties such as the existence of moments are too crucial to be determined by the randomness of the data-generating process. Therefore, the importance of the P<sup>3</sup>C prior approach lies in providing a framework where we can make rational decisions about important model properties through prior knowledge, avoiding any influence of the data or the model error inherent to any statistical model fit.

Note that, in principle, standard bounded priors over  $\xi$  (such as a Beta(0, 1) distribution conditioned to be less than 1/2) can also be used to preserve the first two moments. Here we choose the PC prior approach because it is a principled method flexible enough to represent different levels of prior knowledge, but alternative approaches could also be taken.

## 5 | SIMULATION

We conduct simulation studies to compare the performance of the GEVD and the bGEVD at estimating return levels and to provide some insights on the effects of the hyperparameters  $\alpha, \beta, p_a, p_b$ , and  $c_1 = c_2$  on the bGEV model. Since the GEVD is the only possible non-degenerate limit to properly renormalized maxima, we expect the GEVD to outperform the bGEV for large sample sizes. However, it is very rare to find a sufficiently large number of block-maxima in practical applications. In those cases, we expect both models to perform similarly. For our first simulation study, we draw  $n$  samples from a Fréchet distribution with location 0, scale 1 and shape  $\alpha = 1/0.1$  and retain the largest value. We repeat this  $N$  times to get a sample of  $N$  independent block-maxima and fit the GEV and bGEV distributions using maximum likelihood. We then use the fitted parameters to compute return levels associated with a return period  $T$  and compare them with the return levels from the true maxima distribution (Fréchet(0,  $n^{1/\alpha}$ ,  $\alpha$ )). We repeat the above  $M$  times to obtain Monte Carlo performance measures. Table 1 show our results in terms of the difference ( $\text{RMSE}_{\text{GEV}} - \text{RMSE}_{\text{bGEV}}$ ), where  $\text{RMSE}_{\text{GEV}}$  and  $\text{RMSE}_{\text{bGEV}}$  are the root mean square error associated to the GEV and bGEV fits, respectively. These measures were computed based on  $M = 500$  Monte Carlo replicates with  $n = 30, 50, 100, 500, N = 30, 50, 100, 500, 1000$ , and  $T = 30, 50, 100$ . We can see that both models perform similarly for all block sizes ( $n$ ), number of block-maxima ( $N$ ) and return periods ( $T$ ). As expected, the GEVD performs slightly better for large number of block-maxima (see the last column, where  $N = 1000$ ). We can also see that for  $N = 30$ , the bGEVD performs slightly better than the GEVD for most block sizes. Note that both models were fitted using the same optimization method (Nelder & Mead, 1965) and the same starting values.

Our second simulation study aims to assess the effect of the hyperparameters  $p_a, p_b$ , and  $c_1 = c_2$  over the bGEV model. To this end, we generate  $N = 100$  samples from the GEV distribution with  $\mu = 0, \sigma = 1$ , and  $\xi = 0.1$ . We then fit the bGEV using maximum likelihood and different values of the hyperparameters  $p_a, p_b$ , and  $c_1 = c_2$ . Specifically, the values are  $p_a = 0.05, 0.1, 0.15, p_b = 0.2, 0.25, 0.3$ , and  $c_1 = c_2 = 3, 5$ . We compare the different models based on their ability to estimate a theoretical 50-year return level and summarize our results by computing the RMSE based on  $M = 500$  Monte Carlo replicates. Our results in Table 2 show no major differences across all 18 configurations of  $p_a, p_b$ , and  $c_1 = c_2$ , although for fixed values of  $p_b$ , there seems to be a slight increase in the RMSE as  $p_a$  increases, that is, as the left-tail mixing area become smaller. Note that the left-tail mixing region, defined by the  $p_a$ - and  $p_b$ -quantiles of the Fréchet distribution  $F$ ,

**TABLE 1** Difference ( $\text{RMSE}_{\text{GEV}} - \text{RMSE}_{\text{bGEV}}$ ) for return levels associated with different return periods  $T$  based on  $M = 500$  Monte Carlo replicates

$N$	30	50	100	500	1000
$n$	$T = 30/50/100$				
30	0/0/0.01	0/-0.01/-0.01	0/0/0	0/-0.01/0	0/0/-0.01
50	0/0/0.01	0/0/0	0/0/0	0/0/0	0/0/-0.01
100	-0.01/0/0	0/-0.01/0	0/0/-0.01	0/0/0	0/0/-0.01
500	0/0/0.01	0/0/0	0/0/0	0/0/-0.01	-0.01/0/-0.01

Note:  $N$  is the number of block-maxima and  $n$  is the block size.

**TABLE 2** RMSE for 50-year return level based on  $M = 500$  Monte Carlo replicates to assess the effect of the hyperparameters  $p_a$ ,  $p_b$ , and  $c_1 = c_2$  when fitting the bGEV model

$p_a$	0.05		0.1		0.15	
	$c_1 = c_2 = 3$	$c_1 = c_2 = 5$	$c_1 = c_2 = 3$	$c_1 = c_2 = 5$	$c_1 = c_2 = 3$	$c_1 = c_2 = 5$
0.2	1.11	1.11	1.13	1.13	1.14	1.14
0.25	1.11	1.11	1.13	1.13	1.14	1.14
0.3	1.11	1.11	1.13	1.13	1.14	1.14

Note: For reference, the RMSE associated with a GEV fit is 2.53.

**TABLE 3** RMSE for 50-year return level based on  $M = 500$  Monte Carlo replicates to assess the effect of the hyperparameters  $\alpha$  and  $\beta$  when fitting the bGEV model

$\alpha/\beta$	0.5	0.7	0.9
0.3	1.47	1.11	0.57
0.5	1.88	1.63	0.91
0.7	1.92	2.17	2.16
0.9	0.75	0.65	0.59

Note: For reference, the RMSE associated with a GEV fit is 2.53.

conditions the choice for  $\alpha$  and  $\beta$ . Indeed, as mentioned in Section 3, we must have  $p_a < p_b \leq \min\{\alpha, \beta/2\}$ . Therefore, it is advisable to choose  $p_a$  and  $p_b$  to be relatively small probabilities. In light of the results from Table 2, it is also advisable to choose them not too close. Finally, note that all the bGEV RMSEs are better than the RMSE obtain with GEV fit, which is equal to 2.53.

Our third simulation study analyses the effect of the hyperparameters  $\alpha$  and  $\beta$  on the bGEV model. To this end, we follow the same setup used in our second simulation study with  $p_a = 0.05$ ,  $p_b = 0.2$ ,  $c_1 = c_2 = 5$ , and  $\alpha = (0.3, 0.5, 0.7, 0.9)$  and  $\beta = (0.5, 0.7, 0.9)$ . The spreads  $s_\beta$  associated with the different values of  $\beta$  are, respectively, the length of the central 50%, 30%, and 10% interval. As before, we compare the different models based on their ability to estimate a theoretical 50-year return level and summarize our results by computing the RMSE based on  $M = 500$  Monte Carlo replicates. From Table 3, we can see that combinations of  $\alpha$  and  $\beta$  yield different levels of accuracy with no clear monotonic pattern. Overall, the bGEV fit with any combination of  $\alpha$  and  $\beta$  leads to a better RMSE than the GEV fit, whose RMSE is 2.53.

To summarize, our first simulation study shows that the GEVD and bGEVD are comparable when estimating the right tail. Note that in this simulation study, maxima were taken over Fréchet random variables with location parameter equal to 0, which means that the generated values are bounded by 0. This represents a setting where the original observations have a natural (physical) lower bound, which is the case for, for example, precipitation or pollutant concentrations. From our results in Table 1, we can see that the bGEV model misspecification (support in the whole real line) does not affect the resulting upper tail estimation. Finally, the results presented in Tables 2 and 3 show the effect of the hyperparameters associated with the new parametrization, the left-tail mixing region and the weight function and can be used as guidelines when fitting the bGEVD.

## 6 | APPLICATION

In this section, we illustrate the bGEV model using the new parametrization (Section 2) and the P<sup>3</sup>C prior approach (Section 4). To this end, we consider monthly maximum concentrations of nitrogen dioxide (NO<sub>2</sub>), measured in micrograms per cubic meter or  $\mu\text{g}/\text{m}^3$ , in Bakersfield, a city in Kern County, California. NO<sub>2</sub> is a chemical compound that primarily gets in the air from fuel-burning (U.S. Environmental Protection Agency, <https://www.epa.gov>); see Vettori et al. (2019) and Vettori et al. (2020) for related studies of extreme air pollutant concentrations in California. Monthly maximum NO<sub>2</sub> measurements are computed from January 2000 to March 2016, giving rise to 136 complete measurements.

Available covariates include monthly mean and monthly maximum measurements of wind speed, temperature, pressure, and relative humidity (eight covariates in total). To model (possible non-linear) relationships between the covariates and the NO<sub>2</sub> concentrations, we assume that monthly maximum concentrations of NO<sub>2</sub> follow a bGEVD where the location  $q_\alpha$  and spread  $s_\beta$  vary with covariates. For simplicity, we assume the shape  $\xi$  to be covariate-free. Although likelihood-based inference assuming independence is straightforward (see the density associated with the model in (1) in the Section A.1 of the Appendix), we here consider Bayesian inference in the context of latent Gaussian models using the integrated nested Laplace approximation (INLA; Rue et al., 2009). Within the INLA framework, we assume that observations are conditionally independent given a latent Gaussian field and a set of hyperparameters. Loosely speaking, the latent field describes the underlying trends and dependence structure of the observations and can be represented in terms of an additive linear predictor. For the bGEV model, we link the linear predictor to  $q_\alpha$ , which means that the linear predictor describes the effect of the covariates over the  $\alpha$ -quantile. Additionally, our bGEV-INLA implementation allows the spread and the shape parameters to vary linearly with covariates.

The goal of the Bayesian inference is to get posterior marginal distributions for the components of the linear predictor, the components of the linear models for the spread and shape parameters, and the set of hyperparameters. This translates into the computation of high-dimensional integrals, and INLA tackles these computations conveniently using the Laplace approximation in a nested way, leveraging modern numerical techniques for sparse matrices. It is important to notice that latent Gaussian models represent a very vast class of statistical models; therefore, INLA can be used in a variety of applications (see, e.g., Wang et al., 2018 and Krainski et al., 2018). For a gentle review of the INLA methodology, see Rue et al. (2017).

Model selection was carried out using the Watanabe–Akaike Information Criterion (WAIC), the Deviance Information Criterion (DIC), and graphical tools based on the probability integral transform over all the possible configurations of the eight covariates when we include them linearly and non-linearly in the location  $q_\alpha$  and linearly in the spread  $s_\beta$ . The best model obtained using a forward-selection procedure has a covariate-dependent location and covariate-free spread and shape and is given by

$$\begin{aligned} q_\alpha(t) &= \gamma_0 + \gamma_1 \mathbb{1}_{\text{yot}_t=i} + \gamma_2 \times \text{monthly max wind speed}(t) + f_1(\text{month}(t)) \\ &\quad + f_2(\text{monthly mean temperature}(t)), \quad i = 1, \dots, 17, \\ \log(s_\beta(t)) &= s_{\beta,0}, \quad \xi(t) = \xi_0, \end{aligned} \quad (6)$$

for  $t = 1, \dots, 136$ , where  $\mathbb{1}_{\text{yot}_t=i}$  denotes the year of time  $t$ ,  $\gamma_0$  is an intercept,  $\gamma_1$  and  $\gamma_2$  are linear regression coefficient and  $f_1(\cdot)$  and  $f_2(\cdot)$  are non-linear effects defined as cyclic Gaussian random walks of order two with a sum-to-zero constraint for identifiability purposes. A cyclic Gaussian random walk of second order  $f$  can be defined as follows: let  $\mathbf{w} = (w_1, \dots, w_K)^T$  be a discretization of the covariate  $\mathbf{x}$  into  $K$  bins. Then,

$$\begin{aligned} f(w_2) - 2f(w_1) + f(w_K) &\sim \mathcal{N}(0, \tau^{-1}), \\ f(w_{j+1}) - 2f(w_j) + f(w_{j-1}) &\sim \mathcal{N}(0, \tau^{-1}), \quad j = 2, \dots, K-1, \end{aligned} \quad (7)$$

is a Gaussian random walk of order 2 with precision parameter  $\tau$  that controls the level of smoothness of the random walk. Note that  $f_1$  is defined over the original covariate (month) as it is naturally binned, while  $f_2$  is defined over a binned version of monthly mean temperature.

Prior distributions need to be specified for the three linear regression coefficient in (6), the precision parameters of the two random walks ( $\tau_1$  and  $\tau_2$  for  $f_1$  and  $f_2$ , respectively),  $s_{\beta,0}$  and  $\xi_0$ . We choose a vague Gaussian prior with zero mean and precision  $3 \times 10^{-3}$  for the linear regression coefficients ( $\gamma_0, \gamma_1, \gamma_2$ ). Preliminary exploratory analysis (not shown) suggests that the values of NO<sub>2</sub> vary smoothly with months, exhibiting a concave relationship. Using a grid search,  $\tau_1$  was kept fixed to a value that reflects this relationship. For the precision parameter  $\tau_2$ , we set a weak prior distribution where the probability that the standard deviation  $1/\tau_2$  is greater than the empirical standard deviation of the response is 0.01. For the spread, we assume that  $s_{\beta,0} \sim \text{Gamma}(3, 3)$  a prior. Finally, for  $\xi$ , we use the P<sup>3</sup>C prior approach in (5) to preserve the existence of the first two moments of the bGEVD, with  $\lambda = 7$ .

For comparison, we also implement the GEVD with the linear model in (6) linked to the GEV location  $\mu(t)$ . Table 4 shows posterior means and 95% credible intervals for the bGEV and GEV linear regression coefficients and hyperparameters. Figure 2 shows the posterior mean and 95% credible bands of the nonlinear effects associated with both fits as well as probability integral transform values (PITs). PITs are predictive measures of fit, commonly used

**TABLE 4** Posterior means and 95% credible intervals for the linear regression coefficients and hyperparameters of the bGEV and GEV INLA fits

	<b>Pmean</b>	<b>2.5% quant</b>	<b>97.5% quant</b>
bGEV intercept	55.64	51.51	59.75
GEV intercept	46.62	44.08	49.17
bGEV year	-1.14	-1.37	-0.92
GEV year	-1.12	-1.36	-0.87
bGEV max ws	-0.86	-1.23	-0.48
GEV max ws	-2.33	-3.50	-1.16
bGEV $\tau_2$	0.50	0.07	2.10
GEV $\tau_2$	0.18	0.03	0.
bGEV $s_\beta$	14.33	12.65	15.94
GEV $\sigma$	3.30	0.15	6.59
bGEV $\xi$	0.02	0.00	0.20
GEV $\xi$	0.05	-0.04	0.13

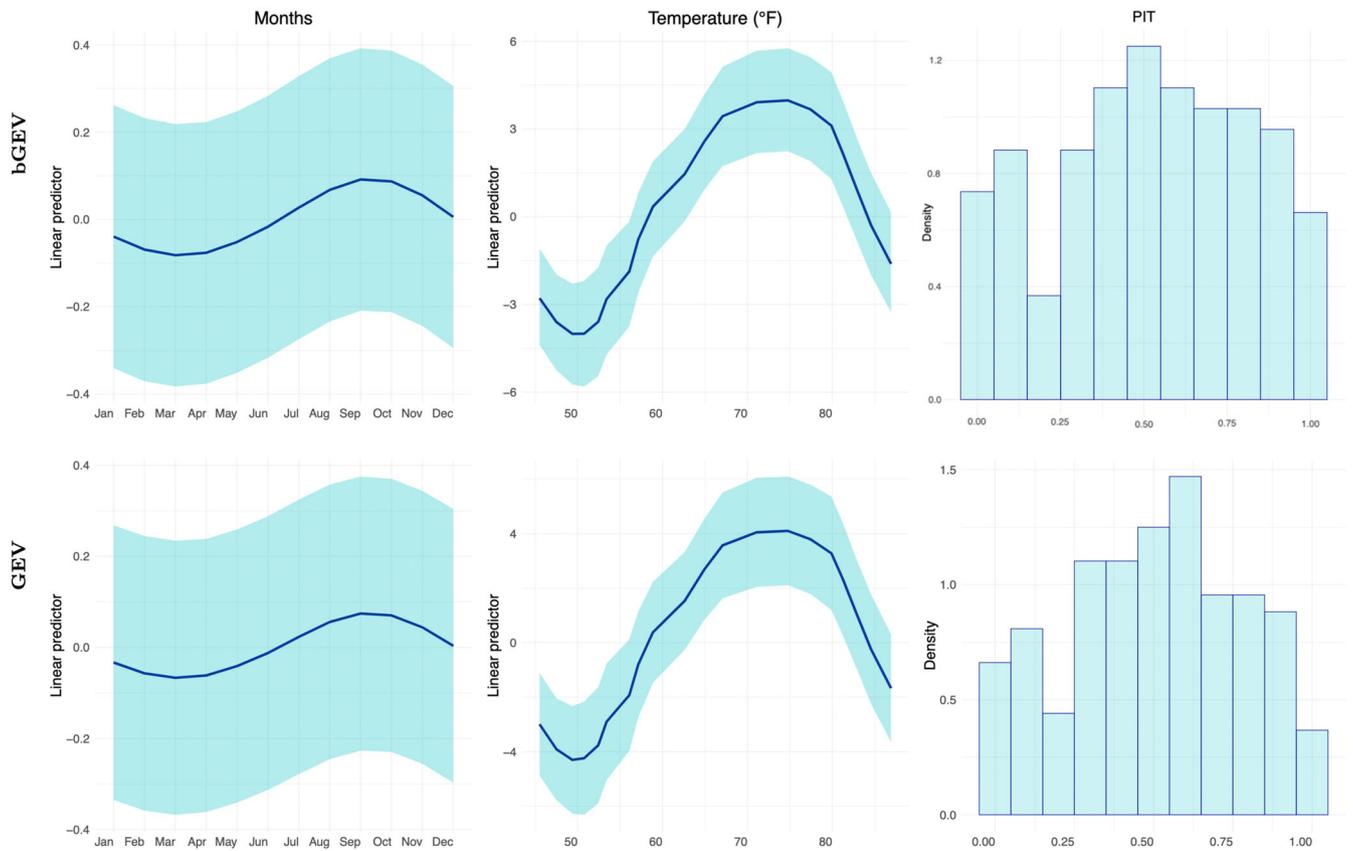
to assess model calibration (Gneiting et al., 2007). If a model is well-calibrated, the PITs histogram serves as a tool to empirically check for uniformity. As expected, we can see that both fits produce very similar results in terms of model estimates, although the GEV fit had less stable results and required additional steps to guarantee convergence. Both models are also comparable in terms of WAIC (934 bGEVD versus 927.8 GEVD) and DIC (938.7 bGEVD versus 926.8 GEVD).

## 7 | DISCUSSION

Inspired by the usually overlooked restrictions inherited by finite-sample maxima distributions when they are approximated by the GEVD, we here make three main contributions. Firstly, we propose the blended generalized extreme value (bGEV) distribution as an alternative to the classical GEVD to fix the lower bound constraint imposed by the GEV distribution when the shape parameter  $\xi$  is positive. Specifically, by using the GEV distribution as an approximation for the distribution of maxima over blocks, properties of the GEV distribution, such as a lower endpoint, are inherited by the finite-sample maxima distribution, which might not be lower-bounded. This issue is particularly troublesome in the presence of linear or nonlinear covariates and becomes crucial with high-dimensional covariate settings. The bGEVD smoothly combines the left tail of the Gumbel distribution ( $\xi = 0$ ) and the right tail of the Fréchet distribution ( $\xi > 0$ ) and depends on a weight function that controls the influence of both distributions. It is important to notice that since the GEVD is the only possible nondegenerate limit to properly renormalized maxima, we do not propose a different limiting model. Instead, we derive a distribution that resembles the GEVD but with infinite support, thus avoiding any lower bound restrictions while matching the right tail of the GEVD. The bGEV model is numerically more stable and is guaranteed to work correctly in all applications. Our simulations and application show that, as expected, the GEV and bGEV models produce similar results.

Secondly, we propose a new parametrization of the GEVD in terms of a quantile and a spread parameter to allow for a more straightforward and meaningful interpretation of the model parameters. The spread is here defined as the difference between two quantiles. This parametrization has significant consequences in a regression setting where parameters are assumed to vary with covariates, as it ensures interpretability even when the first two moments of the distribution are not defined. The parametrization has natural connections with empirical quantiles, and it is particularly advantageous in the Bayesian framework as interpretable prior distributions can be easily assigned.

Thirdly, we introduce the concept of  $P^3C$  priors to retain important model properties when these properties are not *continuous* as functions of model parameters. We use this concept to avoid inconsistencies related to the first two moments



**FIGURE 2** Nonlinear effects over months and mean temperature and PIT values for the bGEV (top) and GEV (bottom) models fitted to the  $\text{NO}_2$  pollution data

of the GEVD. Specifically, since the first two GEV moments are defined for  $\xi < 0.5$ , we restrict the values of  $\xi$  to the interval  $(0, 0.5)$  by normalizing the PC prior using the corresponding cumulative distribution. We argue that the data at hand should not define critical distributional features such as the existence of moments. The  $\text{P}^3\text{C}$  prior approach provides a framework to define those features through prior knowledge.

The GEVD is closely connected to the generalized Pareto distribution (GPD) for threshold exceedances. Indeed, if the conditions for the extremal types theorem hold, then the GPD is the only possible limiting distribution for exceedances over a threshold  $u$  when  $u$  grows to infinity. But unlike the GEVD, the GPD does not have a parameter-dependent support when  $\xi > 0$ , so the methodology presented in Section 3 is not suitable nor needed.

Note that the new parametrization introduced in Section 2 is needed when using the bGEV-INLA implementation, but it is not crucial for defining the bGEV model. Indeed, we can define the bGEV distribution in terms of the GEV parameters  $\mu, \sigma, \xi > 0$ . The purpose of the new parametrization is to provide a way to interpret model parameters even in the absence of the first and second moments, which in turn helps define meaningful priors.

The bGEV model with the new parametrization and the  $\text{P}^3\text{C}$  prior approach is freely available and efficiently implemented in the R-INLA package. In general, INLA requires the model likelihood to be log-concave (Rue et al., 2009), which is not the case for the GEV and the bGEV distributions. Although the lack of log-concavity does not necessarily mean that INLA will not converge, it is highly advisable to try to mitigate potential numerical instabilities by, for example, choosing informative priors for the likelihood parameters. Additional standardizations of the response variable can also help to reduce convergence issues.

## ACKNOWLEDGMENTS

We thank Sabrina Vettori for providing a simplified version of the pollution data. We acknowledge Lars Holden for the notion of blending at the distribution level, an idea that came to light during a hallway conversation 20 years ago. This

publication is partially based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2017-3434.

## DATA AVAILABILITY STATEMENT

The INLA implementation and documentation for the bGEV model can be found in the R-INLA package via `inla.doc` ("bgev"). A tutorial for fitting the bGEV model with R-INLA can be found in <https://www.r-inla.org/documentation>. Additionally, codes to reproduce the simulation studies and the data application are freely available from <https://www.github.com/dcastrocamilo/bGEV>.

## ORCID

Daniela Castro-Camilo  <https://orcid.org/0000-0002-7536-4613>

Raphaël Huser  <https://orcid.org/0000-0002-1228-2071>

Håvard Rue  <https://orcid.org/0000-0002-0222-1881>

## REFERENCES

- Behrens, C. N., Lopes, H. F., & Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3), 227–244.
- Bivand, R., Gómez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software*, 63(1), 1–31.
- Broussard, J. P., & Booth, G. G. (1998). The behavior of extreme values in Germany's stock index futures: An application to intradaily margin setting. *European Journal of Operational Research*, 104(3), 393–402.
- Bruun, J. T., & Tawn, J. A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3), 405–423.
- Cannon, A. J. (2010). A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes: An International Journal*, 24(6), 673–685.
- Carreau, J., & Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes*, 12(1), 53–76.
- Casson, E., & Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1(4), 449–468.
- Castro-Camilo, D., Huser, R., & Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3), 517–534.
- Coles, S. G., & Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(4), 463–478.
- de Melo Mendes, B. V., & Lopes, H. F. (2004). Data driven estimates for mixtures. *Computational Statistics & Data Analysis*, 47(3), 583–598.
- Do Nascimento, F. F., Gamerman, D., & Lopes, H. F. (2011). Regression models for exceedance data via the full likelihood. *Environmental and ecological statistics*, 18(3), 495–512.
- El Adlouni, S., & Ouarda, T. B. (2009). Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo. *Water Resources Research*, 45(6). <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008WR007575>
- El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43(3). <https://agupubs.onlinelibrary.wiley.com/toc/19447973/2007/43/3>
- Frigessi, A., Haug, O., & Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3), 219–235.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 243–268.
- Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S., & Moyer, E. J. (2016). Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(1), 79–103.
- Jóhannesson, A. V., Siebert, S., Huser, R., Bakka, H., & Hrafnelsson, B. (2021). Approximate Bayesian inference for analysis of spatio-temporal flood frequency data. *arXiv preprint:1907.04763*.
- Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman & Hall/CRC Press.
- Leonelli, M., & Gamerman, D. (2020). Semiparametric bivariate modelling with flexible extremal dependence. *Statistics and Computing*, 30(2), 221–236.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1–25.
- Muraleedharan, G., Soares, C. G., & Lucas, C. (2011). *Characteristic and moment generating functions of generalised extreme value distribution (GEV)*. In L. L. Wright (Ed.), *Sea level rise, coastal engineering, shorelines and tides* (pp. 269–276). Nova Science Publishers.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313.
- Opitz, T., Huser, R., Bakka, H., & Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21(3), 441–462.

- Reynkens, T., Verbelen, R., Beirlant, J., & Antonio, K. (2017). Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. *Insurance: Mathematics and Economics*, 77, 65–77.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319–392.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4, 395–421.
- Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, 10(1), 33–60.
- Schmiedt, A. B. (2016). Domains of attraction of asymptotic distributions of extreme generalized order statistics. *Communications in Statistics-Theory and Methods*, 45(7), 2089–2104.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1), 1–28.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90.
- Stein, M. (2017). Should annual maximum temperatures follow a generalized extreme value distribution? *Biometrika*, 104(1), 1–16.
- Stuart, A., Ord, J. K., & Arnold, S. F. (2004). *Kendall's advanced theory of statistics: Classical inference and the linear model* (Vol. 2A). Wiley.
- Vettori, S., Huser, R., & Genton, M. G. (2019). Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3), 831–841.
- Vettori, S., Huser, R., Segers, J., & Genton, M. G. (2020). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *Journal of Computational and Graphical Statistics*, 29, 174–190.
- Wang, X., Ryan, Y. Y., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman & Hall/CRC Press.
- Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 26(11), 3904–3918.
- Yee, T. W., & Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10(1-2), 1–19.
- Zhong, P., Huser, R., & Opitz, T. (2022). Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude. *The Annals of Applied Statistics*, 16(1), 272–299.

**How to cite this article:** Castro-Camilo, D., Huser, R., & Rue, H. (2022). Practical strategies for generalized extreme value-based regression models for extremes. *Environmetrics*, 33(6), e2742. <https://doi.org/10.1002/env.2742>

## APPENDIX

Here, we provide details on the computation of the density of  $H$  in (1) and its first derivative. For ease of notation, we denote  $p(x) = p(x; a, b)$ .

### A.1 The bGEV density

For  $x \in \mathbb{R}$  we can write

$$\begin{aligned}
 H(x) &= \exp\{p(x) \log F(x) + (1 - p(x)) \log G(x)\}, \\
 F(x) &= \exp\{-t_1(x)\}, \quad t_1(x) = \{\max(0, z_1(x))\}^{-1/\xi}, \quad z_1(x) = \left\{ \frac{x - q_\alpha}{s_\beta(\ell_{1-\beta/2, \xi} - \ell_{\beta/2, \xi})^{-1}} + \ell_{\alpha, \xi} \right\}_+, \\
 G(x) &= \exp\{-t_2(x)\}, \quad t_2(x) = \exp(-z_2(x)), \quad z_2(x) = \frac{x - \tilde{q}_\alpha}{\tilde{s}_\beta(\ell_{\beta/2} - \ell_{1-\beta/2})^{-1}} - \ell_\alpha.
 \end{aligned}$$

The density for  $H$ ,  $h$ , can then be easily derived:

$$h(x) = H(x) \left\{ p'(x) \log F(x) + p(x) \frac{f(x)}{F(x)} - p'(x) \log G(x) + (1 - p(x)) \frac{g(x)}{G(x)} \right\},$$

where  $p'(x) = \frac{1}{(b-a)} f_{\text{Beta}}\left(\frac{x}{(b-a)} | c_1, c_2\right)$  and  $f_{\text{Beta}}$  is the density of the Beta distribution with shape parameters  $c_1$  and  $c_2$ . Also,

$$f(x) = \frac{1}{\xi} F(x) z_1(x)^{-(1+1/\xi)} z_1'(x), \quad g(x) = G(x) t_2(x) z_2'(x),$$

with  $z'_1(x) = (\ell_{1-\beta/2,\xi} - \ell_{\beta/2,\xi})/s_\beta$  and  $z'_2(x) = (\ell_{\beta/2} - \ell_{1-\beta/2})/\tilde{s}_\beta$ . Note that  $f$  and  $z'_1$  are defined for  $x > q_\alpha - s_\beta(\ell_{1-\beta/2,\xi} - \ell_{\beta/2,\xi})^{-1}\ell_{\alpha,\xi}$ .

## A.2 First and second derivatives of the bGEV density

Let

$$m(x) = p'(x) \log F(x) + p(x) \frac{f(x)}{F(x)} - p'(x) \log G(x) + (1-p(x)) \frac{g(x)}{G(x)}.$$

The we can write  $h(x) = H(x)m(x)$  and

$$\begin{aligned} h'(x) &= h(x)m(x) + H(x)m'(x), \\ h''(x) &= h'(x)m(x) + 2h(x)m'(x) + Hm''(x), \end{aligned}$$

with

$$\begin{aligned} m' &= p'' \{ \log F - \log G \} + 2p' \left\{ \frac{f}{F} - \frac{g}{G} \right\} + p \left( \frac{f'F - f^2}{F^2} \right) + (1-p) \left( \frac{g'G - g^2}{G^2} \right), \\ m'' &= p''' \{ \log F - \log G \} + 3p'' \left\{ \frac{f}{F} - \frac{g}{G} \right\} + 3p' \left\{ \frac{f'F - f^2}{F^2} - \frac{g'G - g^2}{G^2} \right\} \\ &\quad + p \left\{ \frac{f''F^2 - 3ff'F + 2f^3}{F^3} \right\} + (1-p) \left\{ \frac{g''G^2 - 3gg'G + 2g^3}{G^3} \right\}. \end{aligned}$$

Note that we removed the  $x$  argument for ease of notation. Note also that

$$p''(x) = \frac{1}{(b-a)^2} f'_{\text{Beta}} \left( \frac{x-a}{b-a} | c_1, c_2 \right), \quad p'''(x) = \frac{1}{(b-a)^3} f''_{\text{Beta}} \left( \frac{x-a}{b-a} | c_1, c_2 \right),$$

where

$$\begin{aligned} f'_{\text{Beta}}(y|c_1, c_2) &= \frac{(c_1 - 1)y^{c_1-2}(1-y)^{c_2-1} - (c_2 - 1)y^{c_1-1}(1-y)^{c_2-2}}{\beta(c_1, c_2)}, \\ f''_{\text{Beta}}(y|c_1, c_2) &= \frac{(c_1 - 1)\{(c_1 - 2)y^{c_1-3}(1-y)^{c_2-1} - (c_2 - 1)y^{c_1-2}(1-y)^{c_2-2}\}}{\beta(c_1, c_2)} \\ &\quad - \frac{(c_2 - 1)\{(c_1 - 1)y^{c_1-2}(1-y)^{c_2-2} - (c_2 - 2)y^{c_1-1}(1-y)^{c_2-3}\}}{\beta(c_1, c_2)}. \end{aligned}$$

Finally, the first two derivatives of  $f$  and  $g$  can be expressed as

$$\begin{aligned} f'(x) &= \frac{z_1(x)^{-(2+1/\xi)} z'_1(x)}{\xi} \{ f(x)z_1(x) - (1+1/\xi)F(x) \}, \\ f''(x) &= \frac{z_1(x)^{-(3+1/\xi)} z'_1(x)}{\xi} [ -(2+1/\xi) \{ f(x)z_1(x) - (1+1/\xi)F(x) \} + z_1(x) \{ f'(x)z_1(x) + f(x)(z'_1(x) - 1 - 1/\xi) \} ], \\ g'(x) &= g(x)t_2(x)z'_2(x) - G(x)t_2(x)(z'_2(x))^2, \\ g''(x) &= g'(x)t_2(x)z'_2(x) - 2g(x)t_2(x)(z'_2(x))^2 - G(x)t_2(x)(z'_2(x))^3. \end{aligned}$$