

# REFERENCE PRIORS FOR THE GENERALIZED EXTREME VALUE DISTRIBUTION

Likun Zhang and Benjamin A. Shaby

*Lawrence Berkeley National Laboratory and Colorado State University*

*Abstract:* We derive a collection of reference prior distributions for Bayesian analysis under the three-parameter generalized extreme value (GEV) distribution. These priors are based on an established formal definition of non-informativeness. They depend on the ordering of the three parameters, and we show that the GEV is unusual in that some orderings fail to yield proper posteriors for any sample size. We also consider a reparametrization that explicitly regards return level estimation, which is the most common goal of GEV analysis, to be the most important inferential task. We investigate the properties of the derived priors using simulation and apply them to an analysis of a fire threat index in California.

*Key words and phrases:* Noninformative priors, objective Bayes, posterior normality.

## 1. Introduction

For Bayesian analysis under the three-parameter generalized extreme value (GEV) model, a formal notion of non-informativeness of the prior distri-

---

bution can be achieved using reference analysis. We derive reference priors under the standard parametrization of the GEV, showing that the resulting posterior distributions are improper for some, but not all, orderings of the parameters, and furthermore that re-parametrizing to prioritize inference on a high quantile results in the same behavior as the standard parametrization. Through simulation, we compare the performance of reference priors to two previously-recommended priors: an alternative rule-based non-informative prior and a prior based on domain knowledge, finding none to be uniformly most desirable. The tradeoffs are clearly displayed in our analysis of the extremes of a fire threat index observed in California. In the absence of specific domain knowledge about the tail of the process under investigation, particularly when estimation about a high quantile is the goal of the analysis, the reference prior that we describe here which prioritizes inference on that high quantile might be considered a good default option. We emphasize that the purpose of this paper is not to suggest a prior that is superior to existing priors in terms of its estimation performance, but rather derive the prior that satisfies the notion of noninformativeness in the reference analysis sense, particularly when return level estimation is the primary goal.

---

In Bayesian inference, one typically proposes a parametric model

$$\mathcal{M} = \{p(y|\boldsymbol{\theta}) : y \in \mathcal{Y}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

in which the conditional probability density  $p(y|\boldsymbol{\theta})$  is assumed to be an appropriate characterization of the true underlying mechanism of how the observed data are generated. Then analysis is performed using the available information to infer the parameters that describe the model. Naturally, the choice of model involves a certain amount of subjectivity. In practice, it may be desirable to perform Bayesian analysis that is *objective* in the sense that that it only depends on the assumed model and the available data, and excludes personal beliefs about the model parameters. To achieve this, it is necessary to adopt structural rules that formalize what it means for prior distributions to be non-informative.

The flat prior which assigns equal probability across the support of the parameters, as justified by Jakob Bernoulli’s principle of insufficient reason (Stigler, 1986, p. 135), is a straightforward but naïve solution. This seemingly non-informative prior suffers from multiple pathologies, often producing marginalization paradoxes (Dawid et al., 1973) or implicitly containing large amounts of information that could dominate the analysis. Its casual use is therefore discouraged.

In view of the limitations of using constant priors, Jeffreys (1961) for-

---

mulated a rule for selecting priors, which is taken to be

$$\pi_J(\boldsymbol{\theta}) \propto \det(I(\boldsymbol{\theta}))^{1/2}, \quad (1.1)$$

where  $I(\boldsymbol{\theta})$  is the Fisher information matrix. An attractive property of Jeffreys's rule is that it is invariant under reparametrization of the parameter  $\boldsymbol{\theta}$ . Jeffreys's rule chooses priors by convention, rather than as a unifying representation of ignorance (Kass and Wasserman, 1996), which is also true for several subsequent efforts to construct rules for selecting priors. Maximum entropy priors are another well-studied type of noninformative prior. Entropy of  $\pi$  captures the amount of uncertainty implied by  $\pi$ , and the prior with larger entropy is considered to be less informative, which leads to the selection of prior that maximizes the entropy (Jaynes, 1982). See Kass and Wasserman (1996) for a complete review and critique of Jeffreys's rule and maximum entropy.

Here we consider the class of reference priors suggested by Bernardo (1979), which has proven to be very successful in many settings, including exponential regression (Ye and Berger, 1991), multinomial models (Berger and Bernardo, 1992), and auto-regressive time series models (Berger and Yang, 1994). For a collection of regular priors  $\mathcal{P}$ , the amount of *missing information* about the (univariate, for the moment) parameter  $\theta$ , which could potentially be obtained by repeatedly sampling from the assumed

---

model  $\mathcal{M}$ , is measured for each  $\pi \in \mathcal{P}$ . The reference prior is defined to be the prior  $\pi^\theta = \pi(\theta | \mathcal{M}, \mathcal{P})$  that maximizes the missing information within the class of candidate priors  $\mathcal{P}$ , which assures that the information from the available data will not be dominated by prior beliefs. When there are multiple parameters involved in  $\mathcal{M}$ , the reference prior will then be developed via a stepwise procedure.

In this study, we are interested in finding the reference priors for the family of GEV distributions, whose distribution function can be parametrized by  $\boldsymbol{\theta} = (\mu, \tau, \xi)$ :

$$P(y | \boldsymbol{\theta}) = \begin{cases} \exp \left\{ - \left[ 1 + \xi \left( \frac{y - \mu}{\tau} \right) \right]^{-1/\xi} \right\}, & \xi \neq 0, \\ \exp \left\{ - \exp \left[ - \frac{y - \mu}{\tau} \right] \right\}, & \xi = 0, \end{cases}$$

for  $1 + \xi(y - \mu)/\tau > 0$  when  $\xi \neq 0$ , where the scale parameter  $\tau > 0$ , location parameter  $\mu \in \mathbb{R}$ , and shape parameter  $\xi \in \mathbb{R}$ . The GEV is an important class of distributions because it arises as the limiting distribution of re-normalized maxima taken over increasingly large samples of random variables. It is therefore considered the standard tool for analyzing the far right tail of univariate processes. However, the support of GEV distribution is dependent on its parameter, which makes it challenging to derive the common asymptotic properties of likelihood-based estimators. Frequentist asymptotic results have only been recently established for the local maxi-

---

mum likelihood estimator (MLE) found on a pre-determined compact subset of  $\Theta$ ; see Dombry (2015) and Bücher and Segers (2017). Zhang and Shaby (2021b) further showed that the local MLE found on a compact set is actually the unique and global maximizer of the GEV likelihood function when  $n$  is sufficiently large.

Nevertheless, it is difficult to examine the joint likelihood function over entire parameter space  $\Theta$  when applying Bayesian methods. Relatively few systematic explorations of prior specifications have appeared, most of which are proposed explicitly for the sub-families of the GEV distributions. Ramos et al. (2018) established two reference priors specific to Fréchet distribution. Ho (2010) investigated noninformative matching prior, and Eugenia Castellanos and Cabras (2007) studied a Jeffreys prior for the parameters of the generalized Pareto distribution, which is closely related to the GEV. Sun (1997) derived reference and matching priors for the two-parameter Weibull distribution. For the GEV likelihoods, Northrop and Attalides (2016) extensively discussed posterior propriety when paired with the Jeffreys prior (while holding  $\mu$  fixed), the maximal data information (MDI) prior (Zellner, 1971), and independent uniform priors. This paper conducts an analogous investigation to Northrop and Attalides (2016), but with reference priors. Beranger et al. (2019) focused on the estimation of

---

returns levels in a Bayesian framework with the prior  $\pi(\boldsymbol{\theta}) \propto 1/\tau$ .

To derive the reference priors via a conditioning argument described in Bernardo (2005), we utilize the large sample Bayesian results from Zhang and Shaby (2021a). Given an independent and identically distributed sequence of observations, they formally established the asymptotic posterior normality for the family of GEV distributions as seen in the Bernstein-von Mises theorem—the posterior distribution of the GEV parameter vector, paired with a class of priors which factorizes as  $\pi(\boldsymbol{\theta}) \propto g(\xi)/\tau$ , converges to a normal distribution centered at the true parameter. The tail heaviness of  $g(\xi)$  when  $\xi \rightarrow \infty$  was controlled to obtain posterior propriety and asymptotic normality. This class of priors is commonly seen in location-scale models, although the asymptotic posterior normality is ensured as long as the conditioning argument from Bernardo (2005) is valid for any one prior.

More importantly, since the reference prior depends on the ordering of the parameter vector, we investigate the properties of the resultant posteriors under different orderings. In particular, the reference technique provides no guarantees that it will give priors that correspond to proper posteriors, although only a few cases are known of models that satisfy the assumptions of the standard reference technique and fail to yield proper posteriors (Berger et al., 2001; Ramos et al., 2017). We therefore check posterior

---

propriety under different orderings of  $\boldsymbol{\theta} = (\mu, \tau, \xi)$  and find, somewhat surprisingly, that some are not proper.

## 2. Reference priors for GEV distribution

### 2.1 Formal definitions of reference priors

We begin by looking at model  $\mathcal{M}$  with univariate parameter  $\theta$ . To measure the missing information of  $\pi \in \mathcal{P}$  that could be obtained through one sample generated from the model  $\mathcal{M}$ , Bernardo (2005) calculates the Kullback-Leibler distance of the joint density  $p(y, \theta) = p(y | \theta)\pi(\theta)$  from  $p(y)\pi(\theta)$ , where  $p(y) = \int_{\Theta} p(y | \theta)\pi(\theta)d\theta$ . We denote this distance by  $I\{\pi | \mathcal{M}\}$ .

For  $n$  conditionally independent observations given  $\theta$   $\{y_1, \dots, y_n\}$ , we denote the corresponding multivariate model by  $\mathcal{M}^n = \{\prod_{i=1}^n p(y_i | \theta) : y_i \in \mathcal{X}, \theta \in \Theta\}$ . As  $n \rightarrow \infty$ ,  $I\{\pi | \mathcal{M}^n\}$  will become an accurate measure of the missing information about  $\theta$  with respect to the prior  $\pi$ . The reference prior  $\pi^\theta$  is defined as the prior function such that, for some increasing sequence  $\{\Theta_i\}$  with  $\lim_{i \rightarrow \infty} \Theta_i = \Theta$  and  $\int_{\Theta_i} \pi^\theta(\theta)d\theta < \infty$ ,

$$\lim_{n \rightarrow \infty} [I\{\pi_i^\theta | \mathcal{M}^n\} - I\{\pi_i | \mathcal{M}^n\}] \geq 0, \quad \forall \Theta_i, \quad \forall \pi \in \mathcal{P} \quad (2.2)$$

where  $\pi_i^\theta$  and  $\pi_i$  are renormalized versions of  $\pi^\theta$  and  $\pi$  restricted on  $\Theta_i$ .

If  $\Theta$  is a finite parameter space, (2.2) yields the maximum entropy and

## 2.1 Formal definitions of reference priors

---

the reference prior is the uniform distribution. If  $\Theta$  is a continuous parameter space, the reference prior can be more complex, and it might be difficult to write it down in explicit form, depending on the regularity conditions imposed on  $\mathcal{P}$ . However, if the posterior distribution of the parameter is asymptotically normal with standard deviation  $s(\tilde{\theta}_n)/\sqrt{n}$ , where  $\tilde{\theta}_n$  is a consistent estimator of  $\theta$ , then the reference prior is proportional to  $s(\theta)^{-1}$ , given it is a permissible prior.

The extension of the reference prior to the case of several parameters in the model  $\mathcal{M}$  is achieved through reducing the multiple parameter problem to a sequential application of the established procedure for the single parameter case. We first assume an ordering of inferential priority  $\{\theta_1, \dots, \theta_m\}$ , with  $\theta_1$  being the most important. Conditioning on all the more ‘important’ parameters, we calculate the reference prior for the nuisance parameter  $\theta_m$ , and then move the nuisance parameter out of the model via integrating the product of this prior and the model density. This process repeats until only the most important parameter  $\theta_1$  is left in the model. In the end, the product of  $m$  conditional reference priors yields the reference prior under the particular ordering. Generally different orderings produce different priors.

The aforementioned procedure seems formidable. Fortunately, under asymptotic posterior normality, reference priors can be easily obtained in

---

## 2.1 Formal definitions of reference priors

---

terms of the corresponding Fisher information matrix.

**Lemma 1.** *(Bernardo (2005), Theorem 14) Let  $\mathcal{P}_0$  be the class of all continuous priors with support  $\Theta$ , and  $\mathcal{M} = \{p(y|\boldsymbol{\theta}); y \in \mathcal{Y}, \boldsymbol{\theta} \in \Theta = \prod_{j=1}^m \Theta_j\}$  be the assumed model. From any one prior in  $\mathcal{P}_0$ , if the posterior density  $\pi(\boldsymbol{\theta} | y_1, \dots, y_n)$  is asymptotically normal with covariance  $V(\tilde{\boldsymbol{\theta}}_n)/n$ , where  $\tilde{\boldsymbol{\theta}}_n$  is a consistent estimator of  $\boldsymbol{\theta}$ , let  $H_j$  be the inverse of the upper  $j \times j$  submatrix of  $V$  and  $h_{jj}(\boldsymbol{\theta})$  be the bottom right element of  $H_j$ . Then the reference prior, corresponding to the ordering  $\{\theta_1, \dots, \theta_m\}$ , is*

$$\pi(\boldsymbol{\theta} | \mathcal{M}, \mathcal{P}_0) = \pi(\theta_m | \theta_1, \dots, \theta_{m-1}) \times \dots \times \pi(\theta_2 | \theta_1) \pi(\theta_1),$$

where  $\pi(\theta_m | \theta_1, \dots, \theta_{m-1}) = h_{mm}^{1/2}(\boldsymbol{\theta})$ , and for  $i = 1, \dots, m-1$ ,

$$\pi(\theta_j | \theta_1, \dots, \theta_{j-1}) \propto \exp \left[ \int_{\Theta^{j+1}} \prod_{l=j+1}^m \pi(\theta_l | \theta_1, \dots, \theta_{l-1}) \log \{h_{jj}^{1/2}(\boldsymbol{\theta})\} d\boldsymbol{\theta}^{j+1} \right]$$

with  $\boldsymbol{\theta}^{j+1} = \{\theta_{j+1}, \dots, \theta_m\}$ . Moreover, if  $\Theta_j$  does not depend on  $\{\theta_1, \dots, \theta_{j-1}\}$ ,

and the functions  $h_{jj}(\boldsymbol{\theta})$  can be factorized in the form

$$h_{jj}^{1/2}(\boldsymbol{\theta}) \propto f_j(\theta_j) g_j(\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m), \quad j = 1, \dots, m,$$

then the reference prior is simply  $\pi^\theta(\boldsymbol{\theta}) = \prod_{j=1}^m f_j(\theta_j)$ .

Because of the irregularity of the GEV likelihood function, it is not obvious that it is safe to derive reference priors using Lemma 1, which assumes posterior asymptotic normality. However, Zhang and Shaby (2021a)

---

## 2.2 Fisher information matrix for GEV distribution

have formally established posterior asymptotic normality with covariance  $I^{-1}(\hat{\boldsymbol{\theta}}_n)/n$  for independent GEV sequences for  $\boldsymbol{\Theta} = \{(\tau, \mu, \xi) : \tau > 0, \xi > -1/2\}$ , in which  $\hat{\boldsymbol{\theta}}_n$  is the local MLE with strong consistency. We proceed using the reference prior algorithm described in Lemma 1.

## 2.2 Fisher information matrix for GEV distribution

The score function of the GEV log-likelihood and the Fisher information matrix have been derived by Prescott and Walden (1980). The log-likelihood can be written as:

$$l(\boldsymbol{\theta}; y) = -\log \tau - \left(\frac{1}{\xi} + 1\right) \log \left\{1 + \xi \left(\frac{y - \mu}{\tau}\right)\right\} - \left\{1 + \xi \left(\frac{y - \mu}{\tau}\right)\right\}^{-1/\xi}.$$

The Fisher information matrix is defined as the variances of the score functions, whose exact form can be found in the Appendix S1. To apply Lemma 1, we need to calculate the determinant of the Fisher information matrix.

**Proposition 1.** *One can verify that*

$$|I(\boldsymbol{\theta})| = \frac{1}{\tau^4 \xi^4} \left[ \frac{\pi^2}{6} \{p - \Gamma^2(\xi + 2)\} - \{q - s\Gamma(\xi + 2)\}^2 \right], \quad (2.3)$$

where  $p = (1 + \xi)^2 \Gamma(2\xi + 1)$ ,  $\xi q = \xi(1 + \xi) \Gamma'(\xi + 1) + (1 + \xi)^2 \Gamma(1 + \xi)$ ,  $s = 1 - \gamma + \frac{1}{\xi}$  and  $\gamma$  is the Euler–Mascheroni constant.

### 2.3 Calculate the reference priors

Zhang and Shaby (2021a) established  $V^{-1}(\boldsymbol{\theta}) = H(\boldsymbol{\theta}) = I(\boldsymbol{\theta})$  as needed in Lemma 1. We now calculate the reference priors under all orderings of  $\boldsymbol{\theta} = (\tau, \mu, \xi)$ .

**Proposition 2.** *Let  $\mathcal{P}_0$  be the class of all continuous priors with support  $\Theta = (0, \infty) \times \mathbb{R} \times (-1/2, \infty)$ . Denote the upper  $j \times j$  submatrix of  $I^{-1}(\boldsymbol{\theta})$  by  $V_j$ ,  $H_j = V_j^{-1}$ , and  $h_{jj}(\boldsymbol{\theta})$  is the lower right element of  $H_j$ .*

(A) *Under the ordered parametrizations  $(\xi, \tau, \mu)$  and  $(\xi, \mu, \tau)$ ,*

$$\pi(\xi, \tau, \mu | \mathcal{P}_0) \propto \frac{1}{\tau} h_{11}^{1/2}(\xi) = \frac{1}{\tau |\xi|} \left[ \frac{\pi^2}{6} - \frac{\{q - s\Gamma(\xi + 2)\}^2}{p - \Gamma^2(\xi + 2)} \right]^{1/2}. \quad (2.4)$$

*When  $\xi \rightarrow 0$ ,  $h_{11}(\xi) = 11\pi^4/360 - 6\zeta(3)/\pi^2 + o(1) \approx 2.098 + o(1)$ ,*

*where  $\zeta(3)$  is the Apéry's constant.*

*When  $\xi \rightarrow \infty$ ,  $h_{11}(\xi) = \pi^2/(6\xi^2) + o(1/\xi^3)$ .*

*When  $\xi \rightarrow -1/2$ ,  $h_{11}(\xi) = 2\pi^2/3 + O(2\xi + 1)$ .*

(B) *Under the ordered parametrizations  $(\mu, \tau, \xi)$  and  $(\tau, \mu, \xi)$ ,*

$$\pi(\mu, \tau, \xi | \mathcal{P}_0) \propto \frac{1}{\tau} h_{33}^{1/2}(\xi) = \frac{1}{\tau |\xi|} \left[ \frac{\pi^2}{6} + s^2 - \frac{2q}{\xi} + \frac{p}{\xi^2} \right]^{1/2}. \quad (2.5)$$

*When  $\xi \rightarrow 0$ ,  $h_{33}(\xi) = \Gamma^{(2)}(1) + \Gamma^{(3)}(1) + \Gamma^{(4)}(1)/4 + o(1) \approx 2.424 + o(1)$ .*

*When  $\xi \rightarrow \infty$ ,  $h_{33}(\xi) = \{\Gamma(2\xi + 1) - 2\Gamma(\xi + 1)\psi(\xi + 1)\}/\xi^2 + o(1)$ ,*

*where  $\psi$  denotes the digamma function.*

### 2.3 Calculate the reference priors

---

When  $\xi \rightarrow -1/2$ ,  $h_{33}(\xi) = 4/(2\xi + 1) + O(1)$ .

(C) Under the the ordered parametrization  $(\mu, \xi, \tau)$ ,

$$\pi(\mu, \xi, \tau | \mathcal{P}_0) \propto \frac{1}{\tau} h_{22}^{1/2}(\xi) = \frac{1}{\tau |\xi|} \left[ \frac{\pi^2}{6} + (1 - \gamma)^2 - \frac{\{\Gamma(\xi + 2)/\xi - q + 1 - \gamma\}^2}{1 + p - 2\Gamma(\xi + 2)} \right]^{1/2}. \quad (2.6)$$

When  $\xi \rightarrow 0$ ,  $h_{22}(\xi) \approx 2.363 + o(1)$ .

When  $\xi \rightarrow \infty$ ,  $h_{22}(\xi) = \{\pi^2/6 + (1 - \gamma)^2\}/\xi^2 + o(1/\xi^2)$ .

When  $\xi \rightarrow -1/2$ ,  $h_{22}(\xi) = 2\pi^2/3 + 4(1 - \gamma)^2 + O(2\xi + 1)$ .

(D) Under the ordered parametrization  $(\tau, \xi, \mu)$ ,

$$\pi(\tau, \xi, \mu | \mathcal{P}_0) \propto \frac{1}{\tau} h_{22}^{1/2}(\xi) = \frac{1}{\tau |\xi|} \left[ \frac{\pi^2}{6} + s^2 - \frac{q^2}{p} \right]^{1/2}. \quad (2.7)$$

When  $\xi \rightarrow 0$ ,  $h_{22}(\xi) \approx \pi^2(1 - \gamma)^2/6 + 2(\gamma - 1)\zeta(3) + 11\pi^4/360 + o(1) \approx 2.254 + o(1)$ .

When  $\xi \rightarrow \infty$ ,  $h_{22}(\xi) = \{\pi^2/6 + (1 - \gamma)^2\}/\xi^2 + o(1/\xi^2)$ .

When  $\xi \rightarrow -1/2$ ,  $h_{22}(\xi) = 2\pi^2/3 + 4(1 + \gamma)^2 + O(2\xi + 1)$ .

*Proof.* The proof of Proposition 2 is given in Appendix S1. The asymptotic behavior of  $h_{jj}$  when  $\xi$  approaches 0, 1 or  $-1/2$  will be used later on to establish propriety/impropriety of the posterior distributions, formally verifying the conjectures formulated therein.  $\square$

---

### 2.3 Calculate the reference priors

Proposition 2(A) applies to the case where the shape parameter  $\xi$  is considered the most important for inference. This is probably the most common use case under the standard parametrization, as the shape parameter plays the critical role of controlling the thickness of the right tail. Understanding the value of  $\xi$  for a data generating process under analysis is therefore an inferential task with clear and important ramifications.

Proposition 2(B) applies to the orderings where the shape parameter  $\xi$  is of least importance. These include the ordering that corresponds to the conventional notation  $\boldsymbol{\theta} = (\mu, \tau, \xi)$ , which makes it tempting to consider this ordering as somehow canonical. However, this convention is arbitrary with respect to the inferential importance of the parameters, and it is hard to come up with a clear scenario where either ordering referred to in Proposition 2(B) would be preferred. Similarly, the orderings to which Propositions 2(C) and 2(D) apply, those which consider the shape parameter to be of middle importance, may not be commonly applicable in practice.

In all orderings, we see that the three parameters are independent of each other in the reference prior. Furthermore, the prior for the location parameter  $\mu$  is flat, and the prior for the scale parameter  $\tau$  is proportional to  $1/\tau$ . This is typically the case for reference priors for location-scale families, and differs from the Jeffreys prior, which usually has the scale parameter

---

### 2.3 Calculate the reference priors

proportional to  $1/\tau^2$ .

The left panel of Figure 1 shows the prior function for  $\xi$  under the orderings  $(\xi, \tau, \mu)$  and  $(\xi, \mu, \tau)$ . We can see that it decreases at rate  $1/\xi$ , which of course means the prior is improper, and that it converges to its limiting form in its right tail fairly quickly. The right panel of Figure 1 shows the prior function for  $\xi$  under the orderings  $(\mu, \tau, \xi)$  and  $(\tau, \mu, \xi)$ . Its behavior is qualitatively very different, increasing very quickly in  $\xi$ , suggesting that it may not yield a proper posterior. Here again, we see good correspondence between the limiting form and the exact function, this time in the left limit as  $\xi \rightarrow -1/2$ .

When the importance of  $\xi$  is in the middle, the reference priors (2.6) and (2.7) behave similarly to (2.4) under the ordering  $(\xi, \mu, \tau)$ , and have finite limits when  $\xi \rightarrow -1/2$ . Though appearing to be different on the left, they possess the exact same tail  $\sqrt{\pi^2/6 + (1 - \gamma)^2}/\xi$ . See Figure 2. To differentiate between the two different  $h_{22}(\xi)$  appearing in (2.6) and (2.7), we henceforth designate  $h_{22,1}(\xi)$  for the ordered parametrization  $(\mu, \xi, \tau)$ , and  $h_{22,2}(\xi)$  for  $(\tau, \xi, \mu)$ .

## 2.4 Parametrization under return level

---

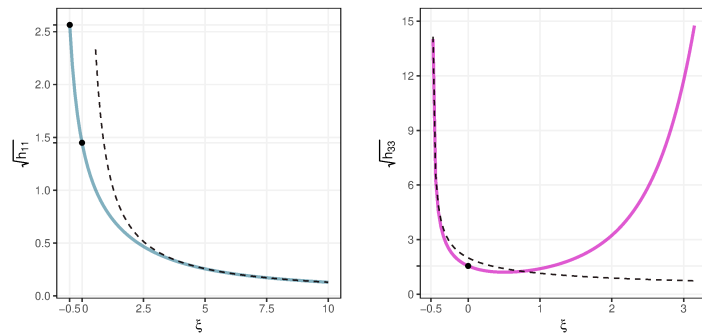


Figure 1: **Left.**  $h_{11}^{1/2}(\xi)$  from  $-1/2$  to 10 under parametrization  $(\xi, \tau, \mu)$  or  $(\xi, \mu, \tau)$ . When  $\xi \rightarrow 0$ ,  $h_{11}^{1/2}(\xi) \rightarrow \sqrt{2.098}$ . When  $\xi \rightarrow -1/2$ ,  $h_{11}^{1/2}(\xi) \rightarrow \sqrt{2\pi^2/3}$ . For  $\xi \in (4, 10)$ ,  $h_{11}^{1/2}(\xi)$  and  $\pi/(\sqrt{6}\xi)$  (dashed curve) are already getting very close, which verifies (A). **Right.**  $h_{33}^{1/2}(\xi)$  from  $-1/2$  to 3 under parametrization  $(\mu, \tau, \xi)$  or  $(\tau, \mu, \xi)$ . When  $\xi \rightarrow 0$ ,  $h_{33}^{1/2}(\xi) \approx \sqrt{2.424}$ . For  $\xi \rightarrow -1/2$ , we compare  $h_{33}^{1/2}(\xi)$  to  $2/\sqrt{2\xi+1}$  (dashed curve).

## 2.4 Parametrization under return level

Most often, the quantity of ultimate interest in a GEV analysis is a high quantile, usually referred to in this context as a *return level*. The GEV is the limiting distribution of re-normalized maxima, but in practice observations always must be divided into blocks of finite size, and the collection of maximum values of each block is analyzed. Then the GEV is fit to the block maxima under the approximation that the limiting distribution is a good representation of the maxima over finite blocks. The return level,

## 2.4 Parametrization under return level

---

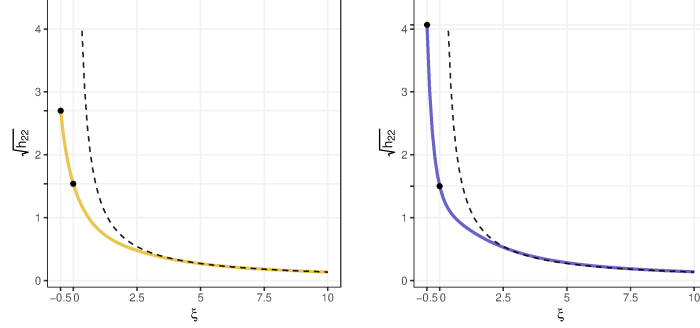


Figure 2: **Left.**  $h_{22}^{1/2}(\xi)$  from  $-1/2$  to  $10$  under parametrization  $(\mu, \xi, \tau)$ . When  $\xi \rightarrow 0$ ,  $h_{22}^{1/2}(\xi) \approx \sqrt{2.363}$ . When  $\xi \rightarrow -1/2$ ,  $h_{22}^{1/2}(\xi) \rightarrow \sqrt{2\pi^2/3 + 4(1-\gamma)^2}$ . **Right.**  $h_{22}^{1/2}(\xi)$  from  $-1/2$  to  $10$  under parametrization  $(\tau, \xi, \mu)$ . When  $\xi \rightarrow 0$ ,  $h_{22}^{1/2}(\xi) \approx \sqrt{2.254}$ . When  $\xi \rightarrow -1/2$ ,  $h_{22}^{1/2}(\xi) \rightarrow \sqrt{2\pi^2/3 + 4(1-\gamma)^2}$ . For  $\xi \rightarrow \infty$ , we compare  $h_{22}^{1/2}(\xi)$  to  $\sqrt{\pi^2/6 + (1-\gamma)^2}/\xi$  under both parametrizations (dashed curves).

which we denote here as  $\mu_T$ , is the value that is exceeded on average in one block out of every  $T$  blocks. For example, a natural blocking scheme for environmental data is often to take yearly maxima, so that  $\mu_T$  is the  $T$ -year return level, which would be exceeded on average once every  $T$  years. The return level (for better or worse) is the standard language by which the magnitudes of events of a specified “rareness” are communicated. Consequently, the return level is also often the basis for practices like government regulations and engineering standards.

---

## 2.4 Parametrization under return level

Since estimating a return level is overwhelmingly the most common end goal of GEV analysis, here we derive reference priors corresponding to the case where inference on the return level is explicitly prioritized. To do this, we simply change from the  $\boldsymbol{\theta} = (\mu, \tau, \xi)$  parametrization to the  $\boldsymbol{\phi} = (\mu_T, \tau, \xi)$  parametrization, which simply requires the transformation

$$\mu = \mu(\boldsymbol{\phi}) = \mu_T - \frac{\tau}{\xi} \left\{ \log^{-\xi} \left( \frac{T}{T-1} \right) - 1 \right\}.$$

We will use the fact that reference priors are coherent under monotone transformations of each parameter in the sense that  $\pi^{\boldsymbol{\phi}}(\boldsymbol{\phi}) = \pi^{\boldsymbol{\theta}}[\boldsymbol{\theta}(\boldsymbol{\phi})]|J(\boldsymbol{\phi})|$ , where  $J(\boldsymbol{\phi})$  is the Jacobian of the inverse transformation  $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\phi})$  (Bernardo, 2005). Under the transformation from  $(\mu_T, \tau, \xi)$  to the  $(\mu, \tau, \xi)$ ,  $|J(\boldsymbol{\phi})| = 1$ , and thus the reference priors under various ordered parametrization are

$$\begin{aligned} \pi(\xi, \tau, \mu_T | \mathcal{P}_0) &= \frac{1}{\tau} h_{11}^{1/2}(\xi), \quad \pi(\mu_T, \tau, \xi | \mathcal{P}_0) = \frac{1}{\tau} h_{33}^{1/2}(\xi), \\ \pi(\mu_T, \xi, \tau | \mathcal{P}_0) &= \frac{1}{\tau} h_{22,1}^{1/2}(\xi), \quad \pi(\tau, \xi, \mu_T | \mathcal{P}_0) = \frac{1}{\tau} h_{22,2}^{1/2}(\xi), \end{aligned} \tag{2.8}$$

which indicates that  $\boldsymbol{\phi} = (\mu_T, \tau, \xi)$  behaves, for the purpose of reference analysis, identically to the standard location parametrization. Hence the reference priors under the common scenario where the return level has the most important inferential priority are those described by (2.5) and (2.6).

## 2.5 The propriety of the posterior

Reference priors generated using the procedure in Lemma 1 are not guaranteed to be permissible in the sense that they result in proper posteriors.

We therefore now examine whether the reference priors  $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta} | \mathcal{P}_0)$ , associated with all ordered parametrizations of  $\boldsymbol{\theta}$ , are permissible.

**Theorem 1.** *Let data consist of i.i.d observations of size  $n$ ,  $\mathbf{y}_n = \{y_1, \dots, y_n\}$ , from  $GEV(\boldsymbol{\theta}_0)$ , where  $\boldsymbol{\theta}_0 = (\xi_0, \tau_0, \mu_0)$ . For reference prior function (2.4) and sample size  $n \geq 4$ , the normalizing constant for the posterior  $C_n = \int_{\Theta_n} p(\mathbf{y}_n | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} < \infty$ , in which  $p(\mathbf{y}_n | \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i | \boldsymbol{\theta})$ ,  $\Theta_n = \{\boldsymbol{\theta} : 1 + \xi \left( \frac{y_i - \mu}{\tau} \right) > 0, i = 1, \dots, n\}$ .*

Moreover, for any  $n \geq 4$ ,

$$C_n \leq \frac{6(n-1)^{-n+2}}{(n-2) \prod_{j=2}^n \delta_j} + \frac{3\Gamma(n)\Gamma(n-1)e}{(n-1)(n-3)\delta_n^{n-1}}, \quad (2.9)$$

where  $\delta_j = \delta_j(n) = y_{(j)} - y_{(1)}$  and  $y_{(1)} < \dots < y_{(n)}$  are the order statistics.

**Corollary 1.** *Since the reference priors under ordered parametrization  $(\mu, \xi, \tau)$  and  $(\tau, \xi, \mu)$  have the same tail properties as (2.4), the posterior obtained from (2.6) or (2.7) is assured to be proper when sample size  $n \geq 4$ .*

Furthermore, for both (2.6) and (2.7),

$$C_n \leq \frac{6(n-1)^{-n+2}}{(n-2) \prod_{j=2}^n \delta_j} + \frac{5\Gamma(n)\Gamma(n-1)e}{(n-1)(n-3)\delta_n^{n-1}}$$

---

for any  $n \geq 4$ ,  $\delta_j = \delta_j(n) = y_{(j)} - y_{(1)}$ .

**Theorem 2.** *Under the same assumptions as Theorem 1, for reference prior (2.5) under the order that  $\xi$  is least preferred, there is no sample size  $n > 0$  for which the corresponding posterior is proper.*

The proof of these results can be found Appendix S2. They tell us that as long as the data consist of more than 4 block maxima, the reference priors based on the orderings where the shape parameter is either the most important or the second most important for inference will yield proper posteriors. In contrast, the reference priors based on the orderings where the shape parameter is the least important for inference will always fail to yield a proper posterior. Hence, this prior should never be used and will not be considered further. By (2.8), similar posterior propriety statements for different orderings can be made under the parameterization with the return level  $\mu_T$ .

### 3. Simulations

To assess the performance of the reference priors derived above, we conduct a small simulation study that mimics applied settings that are frequently encountered in the analysis of environmental extremes. We compare the reference priors to the MDI prior, a competing rule-based prior that was

---

recently suggested for the GEV, and to a beta prior that has been recommended in the applied literature. We do not consider the Jeffreys rule prior (1.1) as it fails to yield a proper posterior; see Appendix S2. We make comparisons across a range of performance metrics.

To simulate the data, we fix  $\mu = 0$ ,  $\tau = 1$ , and consider  $\xi = 0.15$  (typical of annual rainfall maxima),  $\xi = -0.2$  (typical of annual temperature maxima). We also simulate data from a GEV with  $\xi = 1$ , which is very heavy-tailed and not typically seen in environmental data, but might be of interest for applied work in other domains. For each parameter setting, we simulate 1,000 data sets, each of which consists of sample size  $n = 50$ . For each data set, we obtain draws from posterior densities based on reference priors under ordered parametrizations  $(\xi, \mu, \tau)$ ,  $(\mu, \xi, \tau)$  and  $(\tau, \xi, \mu)$ , a Beta(6, 9) prior which appears in Martins and Stedinger (2000) as a recommendation for hydrological data (except in the  $\xi = 1$  case, as  $\xi = 1$  is not in the support of the Beta prior), and the MDI prior which uses the negative entropy of  $p(y | \boldsymbol{\theta})$  (Zellner, 1971):

$$\pi_{MDI}(\boldsymbol{\theta}) \propto \exp \left\{ \int_{\mathcal{Y}} p(y | \boldsymbol{\theta}) \log p(y | \boldsymbol{\theta}) dy \right\} = \frac{1}{\tau} e^{-\gamma(1+\xi)-1},$$

which has a much lighter tail of  $\xi$  compared to the three reference priors under consideration.

To directly sample from the posterior densities of the GEV parameters,

Table 1: Bias and RMSE of the posterior means under five different priors.

Posterior estimators	Bias					RMSE				
	Beta(6, 9)	$h_{11}^{1/2}(\xi)$	$h_{22,1}^{1/2}(\xi)$	$h_{22,2}^{1/2}(\xi)$	MDI	Beta(6, 9)	$h_{11}^{1/2}(\xi)$	$h_{22,1}^{1/2}(\xi)$	$h_{22,2}^{1/2}(\xi)$	MDI
$\xi = -0.2$										
$\xi$	0.0398	-0.0015	-0.0025	-0.0083	-0.0032	0.0686	0.0990	0.1001	0.0999	0.0983
$\mu_{50}$	0.1937	0.1378	0.1322	0.1173	0.1512	0.3856	0.4432	0.4472	0.4301	0.4496
$\mu_{100}$	0.2728	0.1952	0.1885	0.1655	0.2148	0.4833	0.5878	0.5931	0.5674	0.5981
$\xi = 0.15$										
$\xi$	-0.1051	0.0055	0.0045	0.0036	0.0075	0.1256	0.1223	0.1221	0.1256	0.1228
$\mu_{50}$	-0.6402	0.7561	0.7399	0.7454	0.7725	1.0613	1.8915	1.8983	1.9513	1.9448
$\mu_{100}$	-0.9955	1.3050	1.2780	1.2903	1.3310	1.4862	3.0631	3.0769	3.1990	3.1828
$\xi = 1$										
$\xi$	-	0.0204	0.0196	0.0236	0.0149	-	0.1912	0.1924	0.1925	0.1893
$\mu_{50}$	-	45.854	45.689	46.920	42.545	-	95.564	96.198	99.675	90.013
$\mu_{100}$	-	156.33	155.43	160.07	144.55	-	336.71	339.35	367.08	318.24

one can use the `revdbayes` package (Northrop, 2020) in R which permits user-specified priors. Alternatively, one could easily obtain the posterior samples using the random walk Metropolis (RWM) algorithm, as we do here. In the subsequent analyses, we run the MCMC chain for each replicate, and each chain has 10,000 iterations. We discard a burn-in period of 5,000 iterations, and then thin the results by a factor of 10. Using posterior means as estimators, the averaged bias, root mean squared error (RMSE) of the shape parameter  $\xi$ , 50-year and 100-year return levels are calculated and shown in Table 1.

The reference prior under ordering  $(\xi, \mu, \tau)$  and MDI prior have similar performance. The two reference priors with  $\xi$  being the second highest pri-

---

ority have outcomes that are indistinguishable from each other and slightly worse than the MDI and reference prior that prioritizes  $\xi$ . For  $\xi = -0.2$  and  $\xi = 0.15$ , the beta prior yields smaller RMSE for point estimates of almost all parameters and return levels. In general, the beta prior trades a bit of bias for greatly reduced variance, as is typical for a generally well-designed informative prior.

We also investigate the calibration characteristics of the posteriors generated from the various priors. Bayesian calibration, in its simplest form, says that credible intervals are accurate reflections of uncertainty when they cover the true parameter at their nominal rates (see Box, 1980, e.g.); essentially, they should behave like confidence intervals. Table 2 shows the empirical coverage rates of 95% and 99% credible intervals obtained under each prior and data-generating scenario. It shows that the rule-based priors produce almost perfectly calibrated intervals, while the beta prior produces intervals that are well-calibrated in the  $\xi = -0.2$  case but too small in the  $\xi = 0.15$  case, giving an unrealistically confident posterior distribution.

We also apply two proper scoring rules (Gneiting and Raftery, 2007), interval scores and quantile scores, to assess the quality of the posterior distributions under five different priors, which we obtained for each replicate using MCMC. For both scores, higher scores indicate better performance.

Table 2: Empirical coverage rates of  $(1 - \alpha) \times 100\%$  credible intervals.

Coverage rates (%)	$\alpha = 0.05$					$\alpha = 0.01$				
	Beta(6, 9)	$h_{11}^{1/2}(\xi)$	$h_{22,1}^{1/2}(\xi)$	$h_{22,2}^{1/2}(\xi)$	MDI	Beta(6, 9)	$h_{11}^{1/2}(\xi)$	$h_{22,1}^{1/2}(\xi)$	$h_{22,2}^{1/2}(\xi)$	MDI
$\xi = -0.2$										
$\xi$	97.1	95.2	95.6	95.9	95.9	99.5	99.0	99.0	98.8	99.1
$\mu_{50}$	95.9	94.3	94.4	94.5	94.9	99.4	99.0	99.0	99.0	99.0
$\mu_{100}$	95.3	93.9	93.7	94.2	94.1	99.3	98.9	98.8	98.9	99.0
$\xi = 0.15$										
$\xi$	81.3	94.9	94.6	94.6	94.6	95.1	98.6	98.6	98.5	98.8
$\mu_{50}$	86.9	94.7	94.4	94.1	94.5	96.5	99.0	99.2	98.8	98.9
$\mu_{100}$	85.5	94.2	94.5	94.3	94.6	95.9	98.9	98.9	98.9	99.0
$\xi = 1$										
$\xi$	-	95.6	95.7	95.5	95.6	-	98.8	98.6	98.8	98.8
$\mu_{50}$	-	95.6	95.3	95.3	95.4	-	99.1	99.0	99.1	99.1
$\mu_{100}$	-	95.3	95.2	95.3	95.4	-	99.1	99.2	99.0	99.1

The interval score considers both coverage and width of the  $(1 - \alpha) \times 100\%$  posterior credible interval for a given parameter  $\theta$ :

$$S_{\text{int}}(l, u; \theta_0) = -(u - l) - \frac{2}{\alpha}(l - \theta_0)I\{\theta_0 < l\} - \frac{2}{\alpha}(\theta_0 - u)I\{\theta_0 > u\},$$

where  $[l, u]$  are the lower and upper bounds that are the posterior quantiles at level  $\alpha/2$  and  $1 - \alpha/2$ , and  $\theta_0$  is the true value of  $\theta$  that generates the samples. The quantile score adopts a scoring rule that is similar to the check function proposed by Koenker and Bassett (1978):

$$S(r; x) = (x - r)(I\{x \leq r\} - p),$$

where  $x$  is the true  $p$ th quantile,  $r$  is the  $p$ th posterior quantile. It is specifically designed to evaluate quantile estimates, so it is more appropriate than RMSE in assessing the quality of return level analyses. In our analyses,

---

we calculate the quantile scores for 50-year and 100-year return levels, which correspond to 0.02th and 0.01th quantiles. Figure 3 summarizes both scores for 1,000 replicates using boxplots.

The reference priors and the MDI prior yield comparable scores across all scenarios. Due to larger widths of the credible intervals produced by the non-informative priors, the beta prior displays an obvious lead in interval scores, but there are many outlier scores beneath the whiskers that spread the interquartile ranges for the other priors. In addition, this performance advantage of the beta prior becomes less pronounced for the quantile scores. For  $\xi = -0.2$ , the beta prior may be slightly worse than the noninformative priors at estimating the return levels. This particular beta prior is specifically designed with hydrological data in mind, so it is not surprising that it performs better when the shape parameter is slightly positive, rather than slightly negative.

Figure 3 also shows that scores systematically decrease as the true  $\xi$  increases. To better understand how different values of  $\xi$  and sample size  $n$  affect the outcomes from each prior, we perform additional simulations, varying  $\xi$  from -0.4 to 0.4 (with  $n = 50$ ), and varying the sample size 15 to 155 (with  $\xi = 0.2$ ). For each parameter setting, we generate 1,000 replicates and run the MCMC chain for 10,000 iterations in the same way as before.

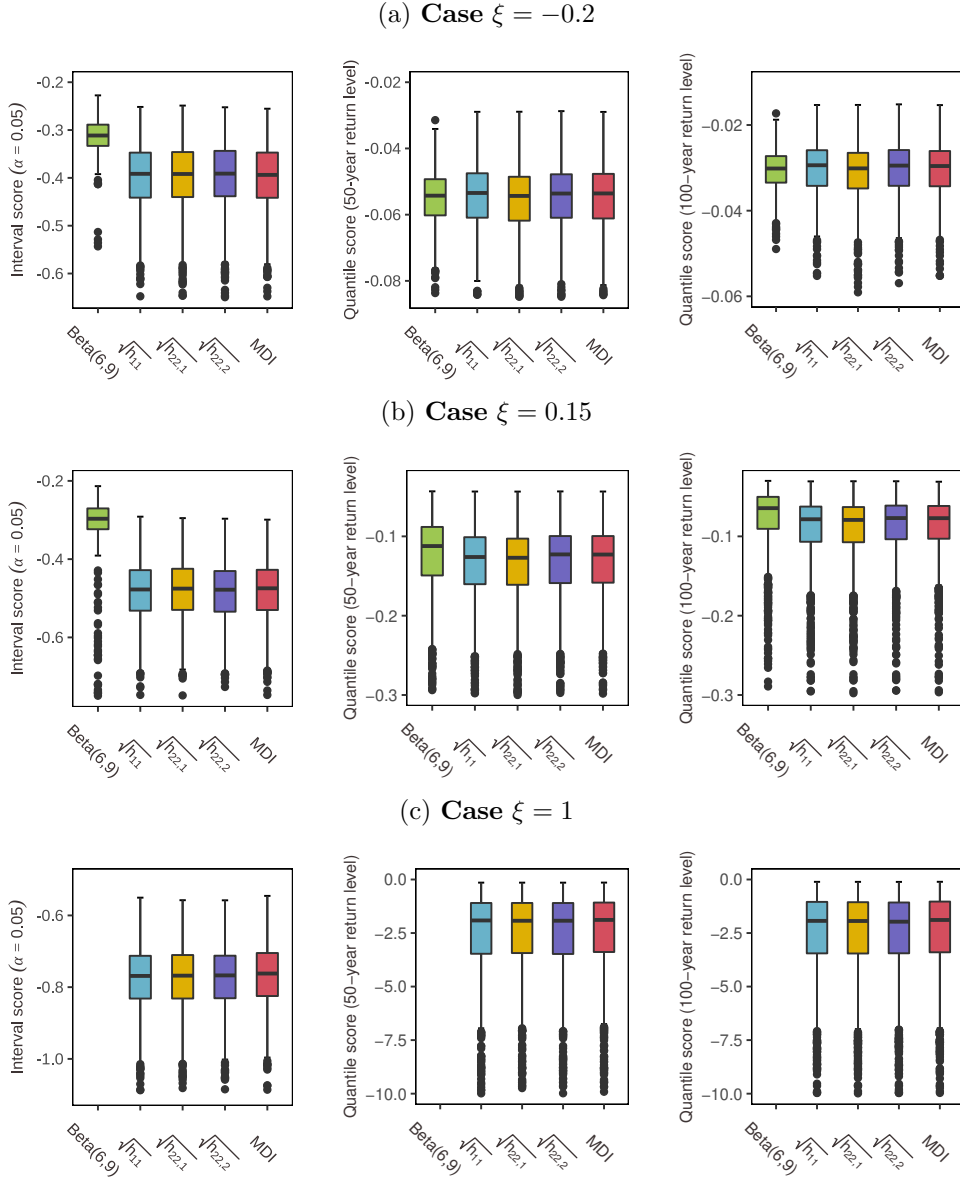


Figure 3: Comparisons of the interval scores ( $\alpha = 0.05$ ) for the posterior credible intervals for  $\xi$ , and quantile scores (50-year and 100-year return levels) among different posterior distributions under five priors. Each score is calculated 1,000 times, once for each replicate consisting of  $n = 50$  samples. Higher values indicate better predictive quality in all panels.

---

The bias, RMSE and averaged interval scores of the 50-year and 100-year return level estimators are compared in Figure 4 and 5. The performances of the MDI prior and reference priors are almost identical.

Since larger  $\xi$  is associated with more extreme quantiles and thus more unstable estimation, the bias and RMSE in Figure 4 increase with  $\xi$  for all priors. Interestingly, the non-informative priors tend to overestimate the return levels, while the beta prior tends to underestimate the return levels. When  $-0.3 < \xi < 0.3$ , the beta prior produces smaller RMSE and higher interval scores, demonstrating the advantage of prior knowledge. However, when  $|\xi| > 0.3$ , the non-informative priors overtake the beta prior in terms of interval scores, indicating better coverage of the credible intervals. When the shape parameter is very small ( $\xi < -0.3$ ), the beta prior yields higher RMSE than the noninformative priors.

By definition, the reference prior technique assumes weak initial knowledge and maximizes the missing information that the data provides; but perfect knowledge is only attained asymptotically when  $n \rightarrow \infty$ . Figure 5 demonstrates that as  $n$  becomes larger, all metrics improve and then stabilize, and the information from the data outweighs the prior knowledge. For this combination of beta prior and data-generating model, it takes a fairly large sample size ( $n \geq 120$ ) for the reference and MDI priors to achieve

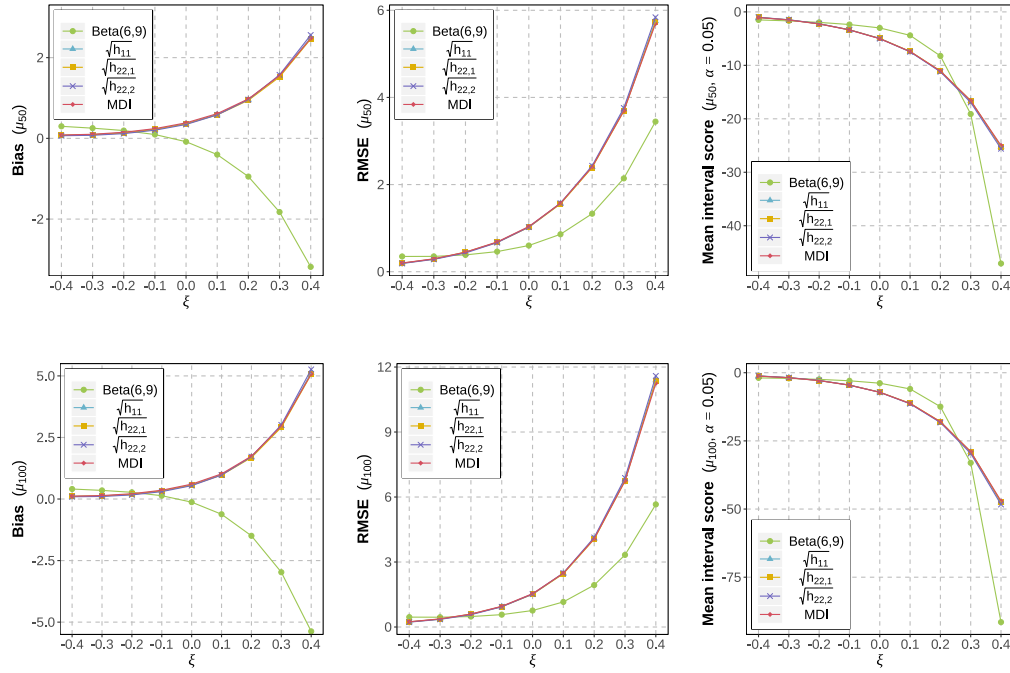


Figure 4: Bias, RMSE, and averaged interval scores of the 50-year return level (top row) and 100-year return level (bottom row) estimators under different true  $\xi$  values, where sample size  $n = 50$  and number of simulations  $N = 1,000$ .

similar performance with respect to RMSE and interval scores. With large sample sizes, the bias is considerably greater for the beta prior.

#### 4. Data analysis

Dry and warm weather conditions continue to pose a high risk of devastating wildfires in California, with dried up and dead vegetation from the

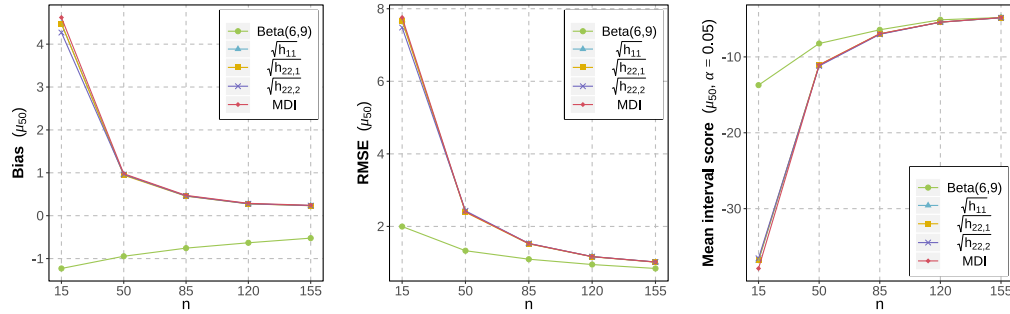


Figure 5: Bias, RMSE, and averaged interval scores of the 50-year return level under different sample sizes, in which true  $\xi = 0.2$  and number of simulations  $N = 1,000$ . The same plots for the 100-year return level behave similarly and are omitted.

2011–2017 drought further increasing risk by acting as kindling. To study the tail behavior the fire risk, we consider the yearly maximum of the Fosberg Fire Weather Index (FFWI) from 1973 to 2018 at four monitoring stations (Dunn et al., 2012). These are the stations that are closest to four deadly wildfires that happened recently, in 2018–2019; see the locations of the wildfires and monitoring stations in Figure 6. The FFWI quantifies the potential wildfire threat by calculating a single number summary from temperature, wind speed, and relative humidity; larger index values reflect greater risk of rapid drying and high flame lengths (Fosberg, 1978). FFWI does not account for human activities and fuel sources, such as changes in land management practices and incursion of invasive species that may

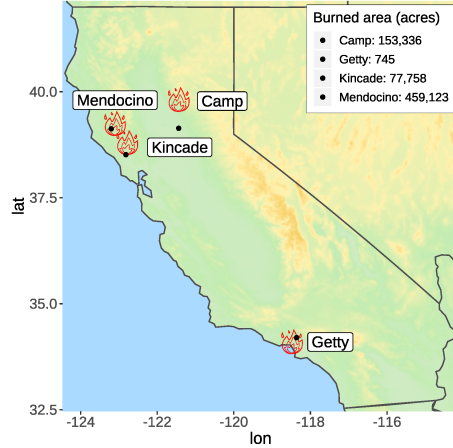


Figure 6: The locations of four notable wildfires in California that happened in 2018–2019. The black dots are the closest available monitoring fire stations to each wildfire respectively.

provide additional fuel. Studying the tail behavior of FFWI will provide useful measures on potential weather impacts on fire threat.

First we analyze the 46 annual maxima of FFWI at each fire station, and run one MCMC chain for 10,000 iterations using each prior listed in Section 3. To obtain starting values for MCMC, we conduct a rough grid search to find the parameters that approximately maximize the un-normalized posterior  $p(\mathbf{x} | \theta)\pi(\theta)$ , when  $\pi(\theta)$  is the prior of choice. For each MCMC chain, we discard a burn-in period of 5,000 iterations and thin the chain by 10 steps. Figure 7 reports the posterior means and 95% posterior credible intervals for all three parameters. We can see that the credible intervals

---

obtained using different priors exhibit close resemblance for  $\mu$  and  $\tau$  across all stations, and posterior means are almost identical. As expected, the beta prior generates slightly narrower credible intervals for the shape parameter  $\xi$ , the value of which is confined between  $-0.25$  and  $0.15$  for Mendocino, Getty, and Camp. However, there is an evident disparity amongst the estimates of  $\xi$  for Kincade, which suggest values higher than  $0.7$  from the MDI and the reference priors. This estimate could be suspicious because the fire station near Mendocino is less than  $140km$  away and has much lower  $\xi$  values than the one near Kincade, although their weather patterns may still differ in potentially important respects, despite their close proximity.

In the above analysis, we have assumed that there has been no trend in the distribution of annual maximum FFWI values. To examine whether this assumption is plausible, we look for time trends in 50-year return levels. We collect the annual maxima in backward 20-year sliding windows from 1992-2018 and perform similar Bayesian analyses as described previously, for each 20-year window, for each station. Since the performance for the non-informative priors are indistinguishable, we only use the beta prior and the reference prior under the ordering  $(\mu, \xi, \tau)$ , which corresponds to the reference prior that prioritizes inference on the return level. Figure 8 shows the posterior mean and 95% credible intervals of the 50-year return levels

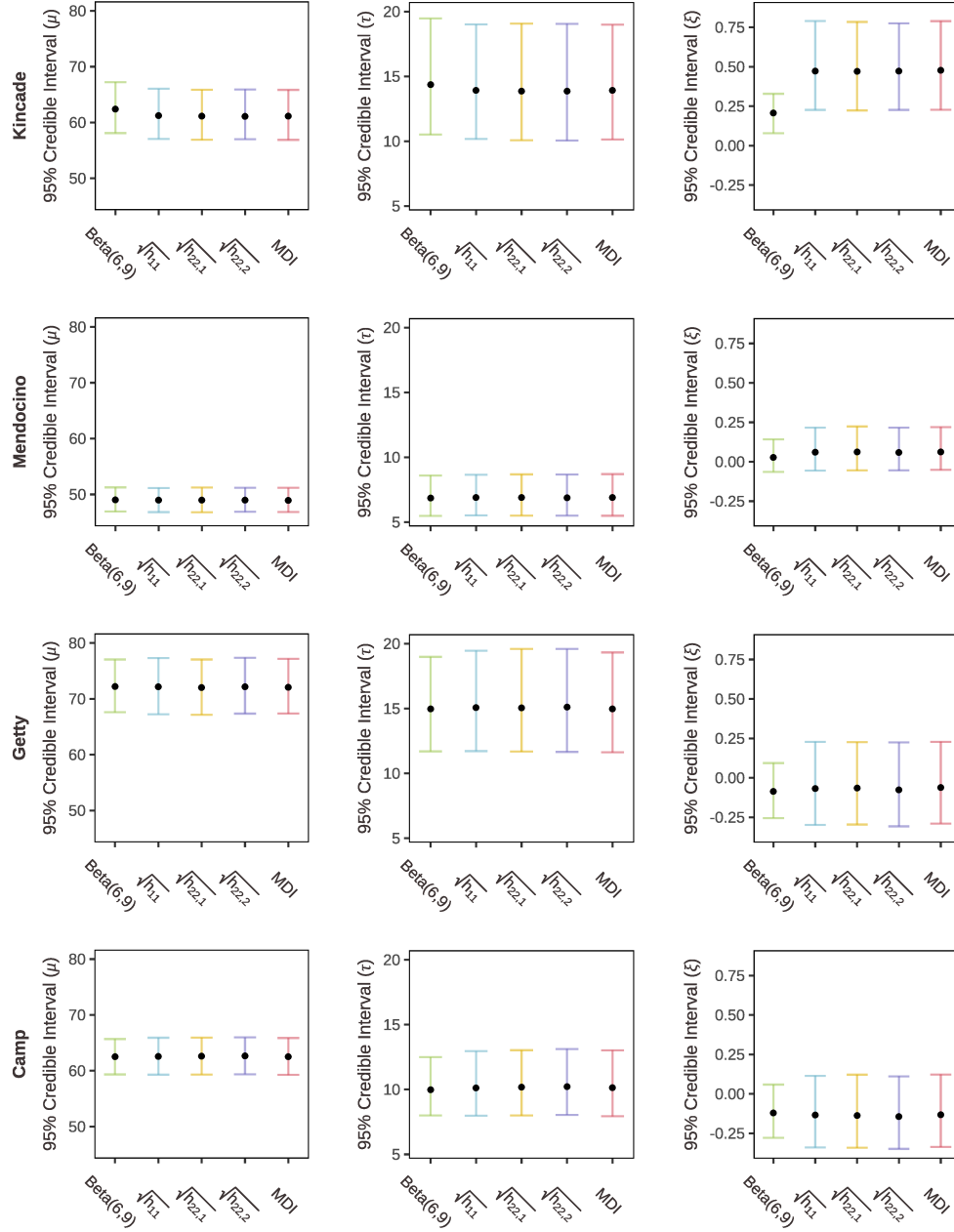


Figure 7: Posterior means and 95% credible intervals from the posterior distributions for  $\mu$ ,  $\tau$  and  $\xi$ , where each row shows results from one monitoring station. For better comparison, the scale limits for the y axis are set to be the same for each parameter across the rows.

---

calculated using the parameters of each MCMC iteration. For Kincade and Mendocino, the return levels are much higher for the 20-year windows that stop in 1991-1998 when using the reference prior, but the estimates from the two methods coincide well after around 2000. We observe a possible slight increase in return levels for Getty and Camp. Aside from that, there is no obvious trend in the return levels, suggesting that our assumption of constant GEV parameters is adequate. It also may suggest that the recent surge in the scale and frequency of wildfires in California may have more to do with fuel availability, human activity, and land use than changes in weather factors.

## **5. Discussion**

In this paper, we used the procedure in Bernardo (2005) to derive reference priors for the family of generalized extreme value distributions. We found that when the primary inferential task is to estimate a return level, the most common use case of the GEV, the reference priors are identical to those under the standard parametrization. Furthermore, for different orderings of the parameters, we completed the specification of these priors by filling the limits at discontinuity points and deriving tail approximations. To answer the question of posterior propriety, we provided upper bounds

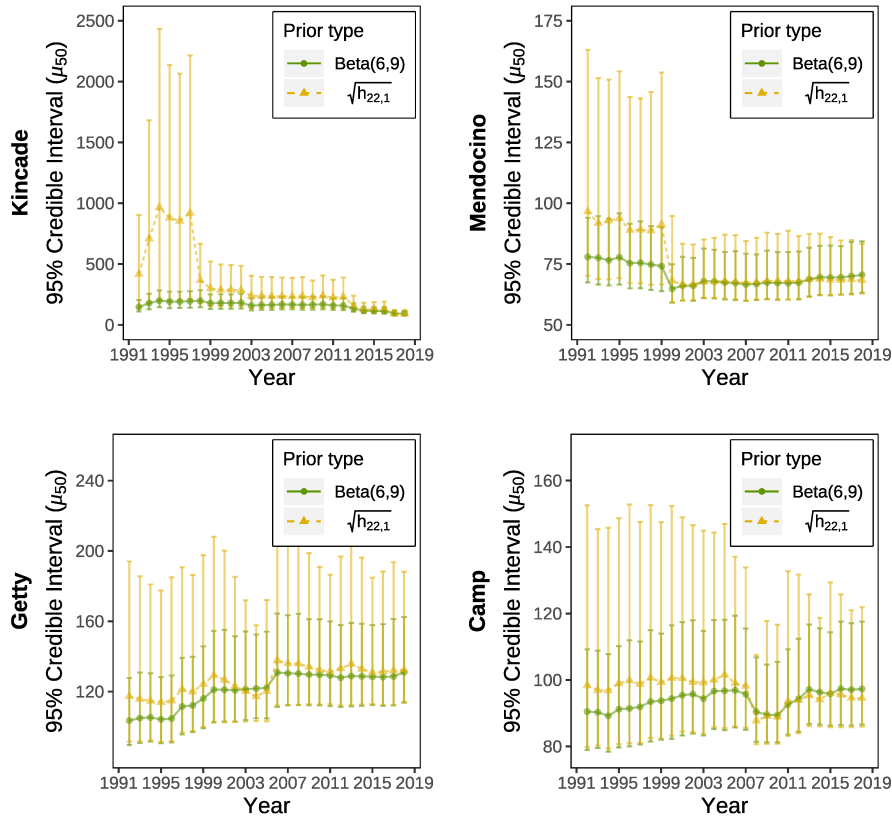


Figure 8: Posterior means and 95% credible intervals for the 50-year return levels calculated from the posterior distributions of every MCMC iteration. For each station, the MCMC analyses was performed using annual maxima in a sliding window that trails back 20 years.

---

for the posterior normalizing constants under the orderings that give permissible reference priors, and found that the posterior is improper under reference priors that treat  $\xi$  as least important of the three parameters. This is a surprising result because examples of reference priors that are not permissible are very rare in the literature.

We performed a series of simulations to compare the performance of the reference priors to other priors suggested in the literature, under varying values of the shape parameter  $\xi$  and sample size  $n$ . For  $\xi$  within a reasonable range, the beta prior seems to have the best performance; when  $|\xi| < 0.3$  it has the best RMSE and interval scores for the estimation of  $\xi$  and return levels, even while its bias is consistently worse than the non-informative priors we tested, due to its reduced variance. Nonetheless, the superiority of the informative beta prior is less obvious as  $n$  grows. When  $n \geq 120$ , it may be better to use the MDI prior or one of the reference priors, as the bias is lower, and RMSE and interval scores are comparable to that of the beta prior. When  $|\xi| > 0.3$ , using this particular beta prior is not advised, especially when the goal is to accurately estimate a return level.

When outside information about the shape parameter  $\xi$  is available, we conclude, unsurprisingly, that using a well-designed informative prior can improve results by stabilizing estimation of the shape parameter  $\xi$ ,

---

especially when the number of observations is small. This is feasible for well-studied variables like precipitation and temperature, but it is somewhat dangerous for less common variables like FFWI, as it is much less clear how to design reasonable informative priors, and estimation performance might suffer due to poor choices.

In contrast, an advantage of using rule-based non-informative priors is that they have stable and reliable performance without requiring prior knowledge of the data-generating process. The rule-based priors we investigated here performed indistinguishably. Without clear empirical guidance based on performance characteristics, we suggest choosing from among the non-informative priors based on the underlying principles from which they derive. If the MDI principle is appealing, then that is a fine choice. If the missing information idea is more attractive, then one of the reference priors might be preferable. If the goal of the analysis is estimation of a return level, then the reference prior corresponding to the parameter ordering  $(\mu_T, \xi, \sigma)$  seems like the natural one to choose.

## **Supplementary Materials**

The Supplementary Material provides details on the construction of the reference priors for the family of generalized extreme value distribution.

Also, it contains the proof for the posterior propriety of the reference priors, and some diagnostics of the MCMC algorithm used in the data analysis.

## References

- Beranger, B., S. A. Padoan, and S. A. Sisson (2019). Estimation and uncertainty quantification for extreme quantile regions. *Extremes*, 1–27.
- Berger, J. O. and J. M. Bernardo (1992). Ordered group reference priors with application to the multinomial problem. *Biometrika* 79(1), 25–37.
- Berger, J. O., V. De Oliveira, and B. Sansó (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* 96(456), 1361–1374.
- Berger, J. O. and R.-y. Yang (1994). Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory* 10(3-4), 461–482.
- Bernardo, J.-M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B* 41(2), 113–147. With discussion.
- Bernardo, J. M. (2005). Reference analysis. In *Bayesian thinking: modeling and computation*, Volume 25 of *Handbook of Statist.*, pp. 17–90. Elsevier/North-Holland, Amsterdam.
- Box, G. E. P. (1980). Sampling and Bayes’ inference in scientific modelling and robustness. *J. Roy. Statist. Soc. Ser. A* 143(4), 383–430. With discussion.
- Bücher, A. and J. Segers (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* 20(4), 839–872.

## REFERENCES

---

- Dawid, A. P., M. Stone, and J. V. Zidek (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* 35, 189–233.
- Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* 21(1), 420–436.
- Dunn, R. J. H., K. M. Willett, P. W. Thorne, E. V. Woolley, I. Durre, A. Dai, D. E. Parker, and R. E. Vose (2012). HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *arXiv preprint arXiv:1210.7191*.
- Eugenia Castellanos, M. and S. Cabras (2007). A default Bayesian procedure for the generalized Pareto distribution. *J. Statist. Plann. Inference* 137(2), 473–483.
- Fosberg, M. A. (1978). Weather in wildland fire management: the fire weather index. *US For Serv Reprints of articles by FS employees*.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* 102(477), 359–378.
- Ho, K.-W. (2010). A matching prior for extreme quantile estimation of the generalized Pareto distribution. *J. Statist. Plann. Inference* 140(6), 1513–1518.
- Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE* 70(9), 939–952.
- Jeffreys, H. (1961). *Theory of probability*. Third edition. Clarendon Press, Oxford.
- Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. *J.*

## REFERENCES

---

- Amer. Statist. Assoc.* 91(435), 1343–1370.
- Koenker, R. and G. Bassett, Jr. (1978). Regression quantiles. *Econometrica* 46(1), 33–50.
- Martins, E. S. and J. R. Stedinger (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research* 36(3), 737–744.
- Northrop, P. J. (2020). *revdbayes: Ratio-of-Uniforms Sampling for Bayesian Extreme Value Analysis*. R package version 1.3.9.
- Northrop, P. J. and N. Attalides (2016). Posterior propriety in Bayesian extreme value analyses using reference priors. *Statist. Sinica* 26(2), 721–743.
- Prescott, P. and A. T. Walden (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* 67(3), 723–724.
- Ramos, P. L., J. A. Achcar, F. A. Moala, E. Ramos, and F. Louzada (2017). Bayesian analysis of the generalized gamma distribution using non-informative priors. *Statistics* 51(4), 824–843.
- Ramos, P. L., F. Louzada, E. Ramos, and S. Dey (2018). The frechet distribution: Estimation and application an overview. *arXiv preprint arXiv:1801.05327*.
- Stigler, S. M. (1986). *The history of statistics*. The Belknap Press of Harvard University Press, Cambridge, MA. The measurement of uncertainty before 1900.
- Sun, D. (1997). A note on noninformative priors for Weibull distributions. *J. Statist. Plann. Inference* 61(2), 319–338.

## REFERENCES

---

- Ye, K. and J. O. Berger (1991). Noninformative priors for inferences in exponential regression models. *Biometrika* 78(3), 645–656.
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. John Wiley & Sons, Inc., New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.
- Zhang, L. and B. Shaby (2021a). Asymptotic posterior normality of the generalized extreme value distribution. *arXiv preprint arXiv:2103.05747*.
- Zhang, L. and B. Shaby (2021b). Uniqueness and global optimality of the maximum likelihood estimator for the generalized extreme value distribution. *Biometrika To appear*.

Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory

E-mail: likunz@lbl.gov

Department of Statistics, Colorado State University

E-mail: bshaby@colostate.edu