Statistica Sinica Preprint No: SS-2021-0258	
Title	Reference Priors for the Generalized Extreme Value Distribution
Manuscript ID	SS-2021-0258
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0258

Complete List of Authors	Likun Zhang and Benjamin A. Shaby
Corresponding Author	Likun Zhang
E-mail	likunz@lbl.gov

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

Statistica Sinica

REFERENCE PRIORS FOR THE GENERALIZED EXTREME VALUE DISTRIBUTION

Likun Zhang and Benjamin A. Shaby

University of Missouri and Colorado State University

Abstract: We derive a collection of reference prior distributions for a Bayesian analysis under the three-parameter generalized extreme value (GEV) distribution. These priors are based on an established formal definition of noninformativeness. 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 a return level estimation, which is the most common goal of a GEV analysis, to be the most important inferential task. We investigate the properties of the derived priors using a simulation, and apply the priors to an analysis of a fire threat index in California.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

1. Introduction

For a Bayesian analysis under the three-parameter generalized extreme value (GEV) model, a formal notion of the noninformativeness of the prior distribution can be achieved using a 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. We further show that re-parametrizing to prioritize an inference on a high quantile results in the same behavior as the standard parametrization. Using a simulation, we compare the performance of the reference priors with that of two previously recommended priors: an alternative rule-based noninformative prior, and a prior based on domain knowledge, finding none to be uniformly most desirable. The tradeoffs are evident 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 an estimation about a high quantile is the goal of the analysis, the reference prior described here, which prioritizes an inference on that high quantile, might be considered a good default option. We emphasize that our purpose is not to suggest a prior that is superior to existing priors in terms of its estimation performance, but rather to derive a prior that satisfies the notion of noninformativeness

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

in the reference analysis sense, particularly when a return level estimation is the primary goal.

In a 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, an 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 a Bayesian analysis that is *objective*, in the sense that it depends only 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 noninformative.

The flat prior, which assigns an 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 noninformative 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

Its casual use is therefore discouraged.

In view of the limitations of using constant priors, Jeffreys (1961) formulated a rule for selecting priors,

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

where $I(\boldsymbol{\theta})$ is the Fisher information matrix. An attractive property of Jeffreys's rule is that it is invariant under a 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 of several subsequent efforts to construct rules for selecting priors. Maximum entropy priors are another well-studied type of noninformative prior. The entropy of π captures the amount of uncertainty implied by π, and a prior with larger entropy is considered to be less informative. This leads to selecting the 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*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

information about the (univariate, for the moment) parameter θ, 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 as the prior $\pi^\theta = \pi(\theta \,|\, \mathcal{M}, \mathcal{P})$ that maximizes the missing information within the class of candidate priors $\mathcal{P}$, which ensures that the information from the available data is not dominated by prior beliefs. When there are multiple parameters involved in $\mathcal{M}$, the reference prior is then developed using a stepwise procedure.

In this study, we are interested in finding reference priors for the family of GEV distributions, the distribution function of which 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 a univariate process. However, the support of a GEV distribution is dependent on its parameter, which makes it challenging to derive the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

common asymptotic properties of likelihood-based estimators. Frequentist asymptotic results have only recently been established for the local maximum likelihood estimator (MLE) found on a predetermined compact subset of $\boldsymbol{\Theta}$; 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 the entire parameter space $\boldsymbol{\Theta}$ when applying Bayesian methods. There have been relatively few systematic explorations of prior specifications, most of which are proposed explicitly for sub-families of GEV distributions. Ramos et al. (2018) established two reference priors specific to the Fréchet distribution. Ho (2010) investigated a 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 GEV likelihoods, Northrop and Attalides (2016) extensively discussed posterior propriety when paired with the Jeffreys prior (while holding μ fixed), the maximal data information (MDI) prior (Zellner, 1971), and independent uniform priors. We conduct

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

an analogous investigation to that of Northrop and Attalides (2016), but with reference priors. Beranger et al. (2019) focused on estimating the return levels in a Bayesian framework with the prior $\pi(\boldsymbol{\theta}) \propto 1/\tau$.

To derive the reference priors using a conditioning argument described in Bernardo (2005), we use the large-sample Bayesian results from Zhang and Shaby (2021a). Given an independent and identically distributed (i.i.d.) 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 that 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 $\xi \to \infty$ was controlled to obtain the posterior propriety and asymptotic normality. This class of priors is commonly seen in location-scale models, although the asymptotic posterior normality is ensured by a wider class of priors as long as the conditioning argument from Bernardo (2005) is valid for any one prior.

More importantly, because 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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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

2. Reference priors for the GEV distribution

2.1 Formal definitions of reference priors

We begin by looking at model $\mathcal{M}$ with a univariate parameter θ. To measure the missing information of $\pi \in \mathcal{P}$ that could be obtained from 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 θ $\{y_1, \cdots, y_n\}$, 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 $n \to \infty$, $I\{\pi \,|\, \mathcal{M}^n\}$ becomes an accurate measure of the missing information about θ with respect to the prior π. The reference prior π^θ is defined as the prior function such that, for some increasing sequence

$\{\Theta_i\}$ with $\lim_{i\to\infty}\Theta_i = \Theta$ and $\int_{\Theta_i}\pi^\theta(\theta)d\theta < \infty$,

$$\lim_{n\to\infty}[I\{\pi_i^\theta \,|\, \mathcal{M}^n\} - I\{\pi_i \,|\, \mathcal{M}^n\}] \geq 0,\ \forall\Theta_i,\ \forall\pi \in \mathcal{P}, \tag{2.1}$$

where π_i^θ and π_i are renormalized versions of π^θ and π, respectively, restricted on Θ_i.

If Θ is a finite parameter space, (2.1) yields the maximum entropy, and the reference prior is the uniform distribution. If Θ is a continuous parameter space, the reference prior can be more complex, and it might be difficult to express it 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 θ, then the reference prior is proportional to $s(\theta)^{-1}$, given that it is a permissible prior.

Extending the reference prior to the case of several parameters in the model $\mathcal{M}$ is achieved by 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 $\{\theta_1, \cdots, \theta_m\}$, with θ_1 being the most important. Conditioning on all the more "important" parameters, we calculate the reference prior for the nuisance parameter θ_m, and then move the nuisance parameter out of the model by integrating the product of this prior and the model density. This process repeats until

only the most important parameter θ_1 is left in the model. In the end, the product of m conditional reference priors yields the reference prior under the particular ordering. In general different orderings produce different priors.

The aforementioned procedure seems formidable. Fortunately, under asymptotic posterior normality, reference priors can be easily obtained in 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 $\boldsymbol{\Theta}$, and let $\mathcal{M} = \{p(y \,|\, \boldsymbol{\theta}); y \in \mathcal{Y}, \boldsymbol{\theta} \in \boldsymbol{\Theta} = \prod_{j=1}^{m} \Theta_j\}$ be the assumed model. From any one prior in $\mathcal{P}_0$, if the posterior density $\pi(\boldsymbol{\theta} \,|\, y_1, \cdots, y_n)$ 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 let $h_{jj}(\boldsymbol{\theta})$ be the bottom-right element of H_j. Then, the reference prior corresponding to the ordering $\{\theta_1, \cdots, \theta_m\}$ is*

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

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

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

with $\boldsymbol{\theta}^{j+1} = \{\theta_{j+1}, \cdots, \theta_m\}$. Moreover, if Θ_j does not depend on $\{\theta_1, \cdots, \theta_{j-1}\}$, and the functions $h_{jj}(\boldsymbol{\theta})$ can be factorized in the form

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

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

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) 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, the exact form of which can be found in Appendix S1. To apply Lemma 1, we need to calculate the determinant of the Fisher information matrix.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

Proposition 1. *One can verify that*

$$|I(\boldsymbol{\theta})| = \frac{1}{\tau^4\xi^4}\left[\frac{\pi^2}{6}\{p-\Gamma^2(\xi+2)\} - \{q - s\Gamma(\xi+2)\}^2\right], \tag{2.2}$$

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* γ *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* $\boldsymbol{\Theta} = (0,\infty)\times\mathbb{R}\times(-1/2,\infty)$. *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 (ξ,τ,μ) *and* (ξ,μ,τ),

$$\pi(\xi,\tau,\mu\,|\,\mathcal{P}_0) \propto \frac{1}{\tau}h_{11}^{1/2}(\xi) = \frac{1}{\tau|\xi|}\left[\frac{\pi^2}{6} - \frac{\{q-s\Gamma(\xi+2)\}^2}{p-\Gamma^2(\xi+2)}\right]^{1/2}. \tag{2.3}$$

When $\xi\to 0$, $h_{11}(\xi) = 11\pi^4/360 - 6\zeta(3)/\pi^2 + o(1) \approx 2.098 + o(1)$, *where* $\zeta(3)$ *is Apéry's constant.*

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

(B) Under the ordered parametrizations (μ, τ, ξ) *and* (τ, μ, ξ),

$$\pi(\mu, \tau, \xi \,|\, \mathcal{P}_0) \propto \frac{1}{\tau} h_{33}^{1/2}(\xi) = \frac{1}{\tau|\xi|}\left[\frac{\pi^2}{6} + s^2 - \frac{2q}{\xi} + \frac{p}{\xi^2}\right]^{1/2}. \tag{2.4}$$

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

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

where ψ *denotes the digamma function.*

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

(C) Under the the ordered parametrization (μ, ξ, τ),

$$\pi(\mu, \xi, \tau \,|\, \mathcal{P}_0) \propto \frac{1}{\tau} h_{22}^{1/2}(\xi) = \frac{1}{\tau|\xi|}\left[\frac{\pi^2}{6} + (1-\gamma)^2 - \frac{\{\Gamma(\xi+2)/\xi - q + 1 - \gamma\}^2}{1 + p - 2\Gamma(\xi+2)}\right]^{1/2}. \tag{2.5}$$

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

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

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

(D) Under the ordered parametrization (τ, ξ, μ),

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

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

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

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

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

Proposition 2(A) applies when the shape parameter ξ is considered the most important for an inference. This is probably the most common use case under the standard parametrization, because the shape parameter plays the critical role of controlling the thickness of the right tail. Understanding the value of ξ for a data-generating process being analyzed is therefore an inferential task with clear and important ramifications.

Proposition 2(B) applies to orderings in which the shape parameter ξ 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 difficult to come up with a clear scenario in which 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 μ is flat, and the prior for the scale parameter τ 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 a scale parameter proportional to $1/\tau^2$.

The left panel of Figure 1 shows the prior function for ξ under the orderings (ξ, τ, μ) and (ξ, μ, τ). We can see that it decreases at the rate $1/\xi$, which 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 ξ under the orderings (μ, τ, ξ) and (τ, μ, ξ). Its behavior is qualitatively different, increasing very quickly in ξ, 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 $\xi \to -1/2$.

When the importance of ξ is in the middle, the reference priors (2.5) and (2.6) behave similarly to (2.3) under the ordering (ξ, μ, τ), and have finite limits when $\xi \to -1/2$. Though appearing to be different on the left, they possess the same tail $\sqrt{\pi^2/6 + (1-\gamma)^2}/\xi$; see Figure 2. To differentiate between the $h_{22}(\xi)$ appearing in (2.5) and (2.6), we henceforth designate

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

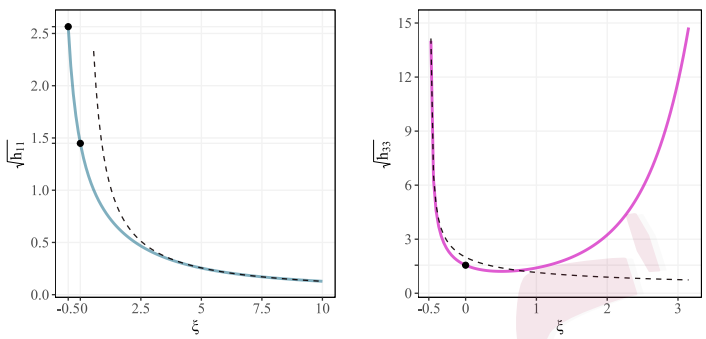

Figure 1: **Left.** $h_{11}^{1/2}(\xi)$ from $-1/2$ to 10 under parametrization (ξ, τ, μ) or (ξ, μ, τ). When $\xi \to 0$, $h_{11}^{1/2}(\xi) \to \sqrt{2.098}$. When $\xi \to -1/2$, $h_{11}^{1/2}(\xi) \to \sqrt{2\pi^2/3}$. For $\xi \in (4, 10)$, $h_{11}^{1/2}(\xi)$ and $\pi/(\sqrt{6}\xi)$ (dashed curve)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

$h_{22,1}(\xi)$ for the ordered parametrization (μ, ξ, τ), and $h_{22,2}(\xi)$ for (τ, ξ, μ).

2.4 Parametrization under a return level

Most often, the quantity of 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. However in practice, observations must always be divided into blocks of finite size, after which the collection of maximum values of each block is analyzed. Then, the GEV

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

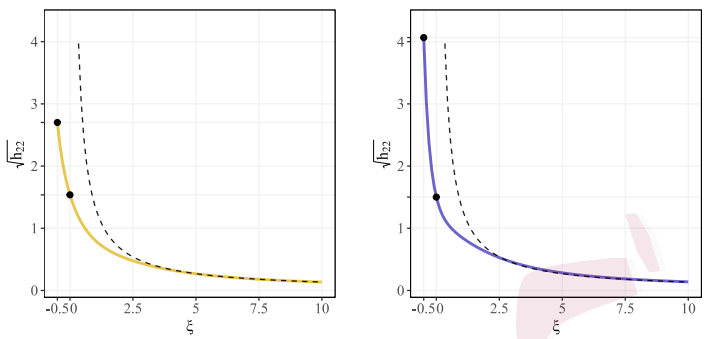

Figure 2: **Left.** $h_{22}^{1/2}(\xi)$ from $-1/2$ to 10 under parametrization (μ, ξ, τ). When $\xi \to 0$, $h_{22}^{1/2}(\xi) \approx \sqrt{2.363}$. When $\xi \to -1/2$, $h_{22}^{1/2}(\xi) \to \sqrt{2\pi^2/3 + 4(1-\gamma)^2}$. **Right.** $h_{22}^{1/2}(\xi)$ from $-1/2$ to 10 under parametrization (τ,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

is fitted to the block maxima under the approximation that the limiting distribution is a good representation of the maxima over finite blocks. The return level, which we denote here as μ_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 μ_T is the T-year return level, which is 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 such as government regulations and engineering standards.

Because estimating a return level is overwhelmingly the most common goal of a GEV analysis, here, we derive reference priors corresponding to the case that explicitly prioritizes an inference on the return level. 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 use the fact that reference priors are coherent under monotone transformations of each parameter in the sense that $\pi^{\phi}(\boldsymbol{\phi}) = \pi^{\boldsymbol{\theta}}[\boldsymbol{\theta}(\boldsymbol{\phi})]|J(\boldsymbol{\phi})|$, where $J(\boldsymbol{\phi})$ is the Jacobian of the inverse transformation $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\phi})$ (Bernardo, 2005). Under the transformation from (μ_T, τ, ξ) to (μ, τ, ξ), $|J(\boldsymbol{\phi})| = 1$, and thus the reference priors under various ordered parametrizations are

$$\begin{aligned} \pi(\xi, \tau, \mu_T \,|\, \mathcal{P}_0) &= \frac{1}{\tau} h_{11}^{1/2}(\xi), \quad \pi(\mu_T, \tau, \xi \,|\, \mathcal{P}_0) = \frac{1}{\tau} h_{33}^{1/2}(\xi), \\ \pi(\mu_T, \xi, \tau \,|\, \mathcal{P}_0) &= \frac{1}{\tau} h_{22,1}^{1/2}(\xi), \quad \pi(\tau, \xi, \mu_T \,|\, \mathcal{P}_0) = \frac{1}{\tau} h_{22,2}^{1/2}(\xi), \end{aligned} \tag{2.7}$$

which indicates that $\boldsymbol{\phi} = (\mu_T, \tau, \xi)$ behaves, for the purpose of a reference analysis, identically to the standard location parametrization. Hence, the reference priors under the common scenario in which the return level has the most important inferential priority are those described by (2.4) and (2.5).

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. Therefore, we 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 the data consist of i.i.d. observations of size n, $\boldsymbol{y}_n = \{y_1, \cdots, y_n\}$, from GEV($\boldsymbol{\theta}_0$), where $\boldsymbol{\theta}_0 = (\xi_0, \tau_0, \mu_0)$. For reference prior function* (2.3) *and sample size $n \geq 4$, the normalizing constant for the posterior $C_n = \int_{\boldsymbol{\Theta}_n} p(\boldsymbol{y}_n \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$, in which $p(\boldsymbol{y}_n \,|\, \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i \,|\, \boldsymbol{\theta})$, $\boldsymbol{\Theta}_n = \{\boldsymbol{\theta} : 1 + \xi\left(\frac{y_i - \mu}{\tau}\right) > 0, i = 1, \cdots, n\}$.*

Moreover, for any $n \geq 4$,

$$C_n \leq \frac{6(n-1)^{-n+2}}{(n-2)\prod_{j=2}^n \delta_j} + \frac{3\Gamma(n)\Gamma(n-1)e}{(n-1)(n-3)\delta_n^{n-1}}, \tag{2.8}$$

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

Corollary 1. *Because the reference priors under the ordered parametrizations (μ, ξ, τ) and (τ, ξ, μ) have the same tail properties as* (2.3)*, the posterior obtained from* (2.5) *or* (2.6) *is ensured to be proper when the sample size $n \geq 4$.*

Furthermore, for both (2.5) *and* (2.6)*,*

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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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* (2.4) *under the order that* ξ *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 four block maxima, the reference priors based on the orderings in which 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 in which 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 (2.7), similar posterior propriety statements for different orderings can be made under the parameterization with the return level μ_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 analyses of environmental extremes. We compare the reference priors with the MDI prior, a competing rule-based prior recently

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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

To simulate the data, we fix $\mu = 0$ and $\tau = 1$, and consider $\xi = 0.15$ (typical of annual rainfall maxima) and $\xi = -0.2$ (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 has a sample size $n = 50$. For each data set, we obtain draws from posterior densities based on reference priors under the ordered parametrizations (ξ, μ, τ), (μ, ξ, τ), and (τ, ξ, μ), a Beta$(6, 9)$ prior that appears in Martins and Stedinger (2000) as a recommendation for hydrological data (except in the case $\xi = 1$, because $\xi = 1$ 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 ξ compared with those of the three reference priors under consideration.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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
					$\xi = -0.2$					
ξ	0.0398	-0.0015	-0.0025	-0.0083	-0.0032	0.0686	0.0990	0.1001	0.0999	0.0983
μ_{50}	0.1937	0.1378	0.1322	0.1173	0.1512	0.3856	0.4432	0.4472	0.4301	0.4496
μ_{100}	0.2728	0.1952	0.1885	0.1655	0.2148	0.4833	0.5878	0.5931	0.5674	0.5981
					$\xi = 0.15$					
ξ	-0.1051	0.0055	0.0045	0.0036	0.0075	0.1256	0.1223	0.1221	0.1256	0.1228
μ_{50}	-0.6402	0.7561	0.7399	0.7454	0.7725	1.0613	1.8915	1.8983	1.9513	1.9448
μ_{100}	-0.9955	1.3050	1.2780	1.2903	1.3310	1.4862	3.0631	3.0769	3.1990	3.1828
					$\xi = 1$					
ξ	-	0.0204	0.0196	0.0236	0.0149	-	0.1912	0.1924	0.1925	0.1893
μ_{50}	-	45.854	45.689	46.920	42.545	-	95.564	96.198	99.675	90.013
μ_{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 can 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 ξ, and 50-year and 100-year return levels are calculated and shown in Table 1.

The reference prior under the ordering (ξ, μ, τ) and the MDI prior perform similarly. The two reference priors with ξ being the second highest

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

priority have outcomes that are indistinguishable from each other, and are slightly worse than the MDI and the reference prior that prioritizes ξ. For $\xi = -0.2$ and $\xi = 0.15$, the beta prior yields a smaller RMSE for the 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. A Bayesian calibration, in its simplest form, states that credible intervals are accurate reflections of uncertainty when they cover the true parameter at their nominal rates (e.g., Box, 1980); essentially, they should behave like confidence intervals. Table 2 shows the empirical coverage rates of the 95% and 99% credible intervals obtained under each prior and data-generating scenario. The results show that the rule-based priors produce almost perfectly calibrated intervals, whereas the beta prior produces intervals that are well calibrated in the $\xi = -0.2$ case, but are 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), namely, interval scores and quantile scores, to assess the quality of the posterior distributions under five different priors, which we obtained for

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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
					$\xi = -0.2$					
ξ	97.1	95.2	95.6	95.9	95.9	99.5	99.0	99.0	98.8	99.1
μ_{50}	95.9	94.3	94.4	94.5	94.9	99.4	99.0	99.0	99.0	99.0
μ_{100}	95.3	93.9	93.7	94.2	94.1	99.3	98.9	98.8	98.9	99.0
					$\xi = 0.15$					
ξ	81.3	94.9	94.6	94.6	94.6	95.1	98.6	98.6	98.5	98.8
μ_{50}	86.9	94.7	94.4	94.1	94.5	96.5	99.0	99.2	98.8	98.9
μ_{100}	85.5	94.2	94.5	94.3	94.6	95.9	98.9	98.9	98.9	99.0
					$\xi = 1$					
ξ	-	95.6	95.7	95.5	95.6	-	98.8	98.6	98.8	98.8
μ_{50}	-	95.6	95.3	95.3	95.4	-	99.1	99.0	99.1	99.1
μ_{100}	-	95.3	95.2	95.3	95.4	-	99.1	99.2	99.0	99.1

each replicate using an MCMC. For both scores, higher scores indicate better performance. The interval score considers both the coverage and the width of the $(1-\alpha)\times 100\%$ posterior credible interval for a given parameter θ:

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

where $[l,u]$ are the lower and upper bounds that are the posterior quantiles at levels $\alpha/2$ and $1-\alpha/2$, and θ_0 is the true value of θ that generates the samples. The quantile score adopts a scoring rule that is similar to the check function proposed by Koenker and Bassett (1978):

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

where x is the true pth quantile, and r is the pth posterior quantile. It is

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

specifically designed to evaluate quantile estimates, so it is more appropriate than the RMSE for 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 the 0.02th and 0.01th quantiles. Figure 3 summarizes both scores for 1,000 replicates using box plots.

The reference priors and the MDI prior yield comparable scores across all scenarios. Owing to the larger widths of the credible intervals produced by the noninformative 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 the scores systematically decrease as the true ξ increases. To better understand how different values of ξ and the sample size n affect the outcomes from each prior, we perform additional simulations, varying ξ from -0.4 to 0.4 (with $n = 50$), and varying the sample size from

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

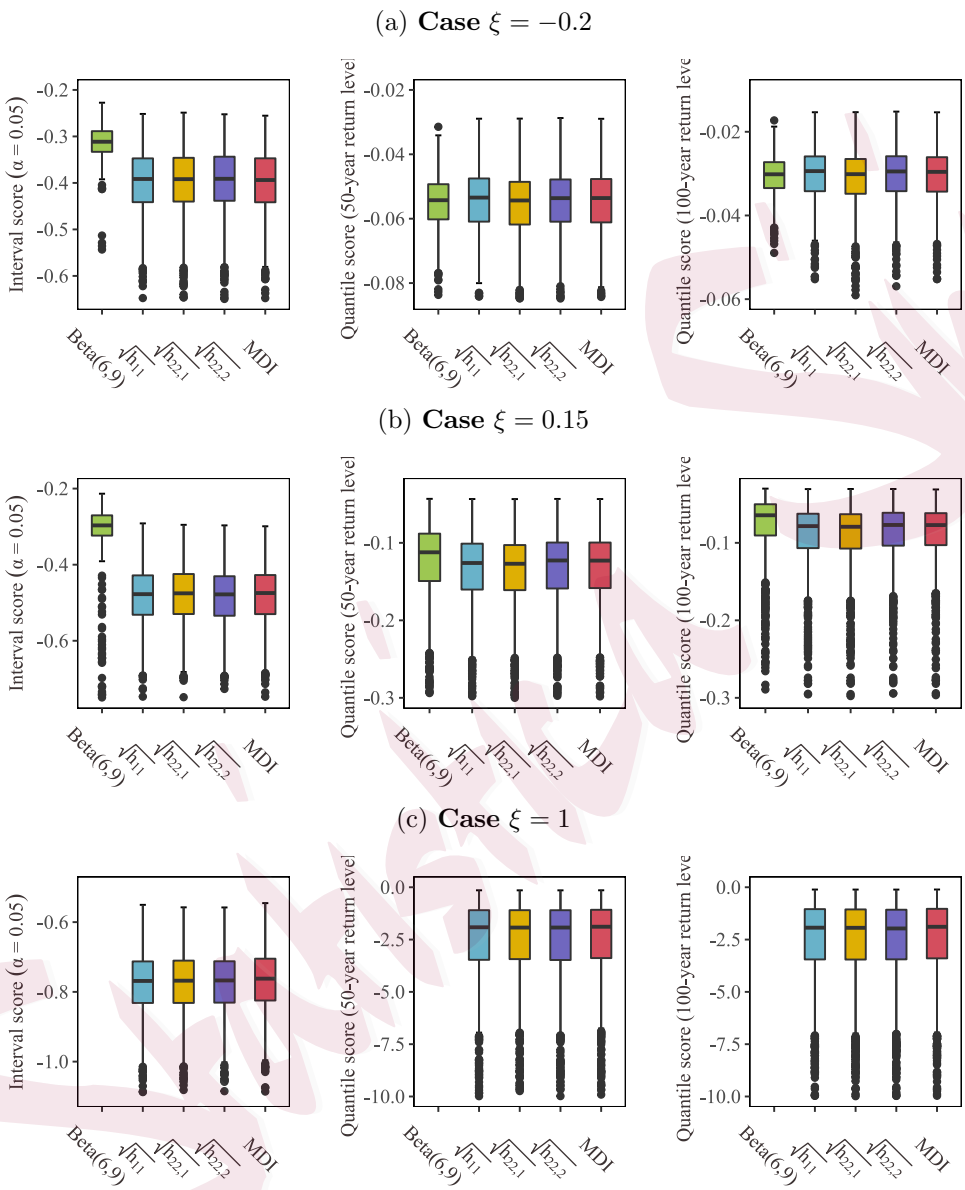

Figure 3: Comparisons of the interval scores ($\alpha = 0.05$) for the posterior credible intervals for ξ and the quantile scores (50-year and 100-year return levels) among different posterior distributions under five priors. Each score is calculated 1,000 times, once for each replicate consisting of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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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

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

By definition, the reference prior technique assumes weak initial knowledge and maximizes the missing information that the data provides; however, perfect knowledge is only attained asymptotically when $n \to \infty$. Figure 5 demonstrates that as n becomes larger, all metrics improve and then stabilize, and the information from the data outweighs the prior knowledge.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

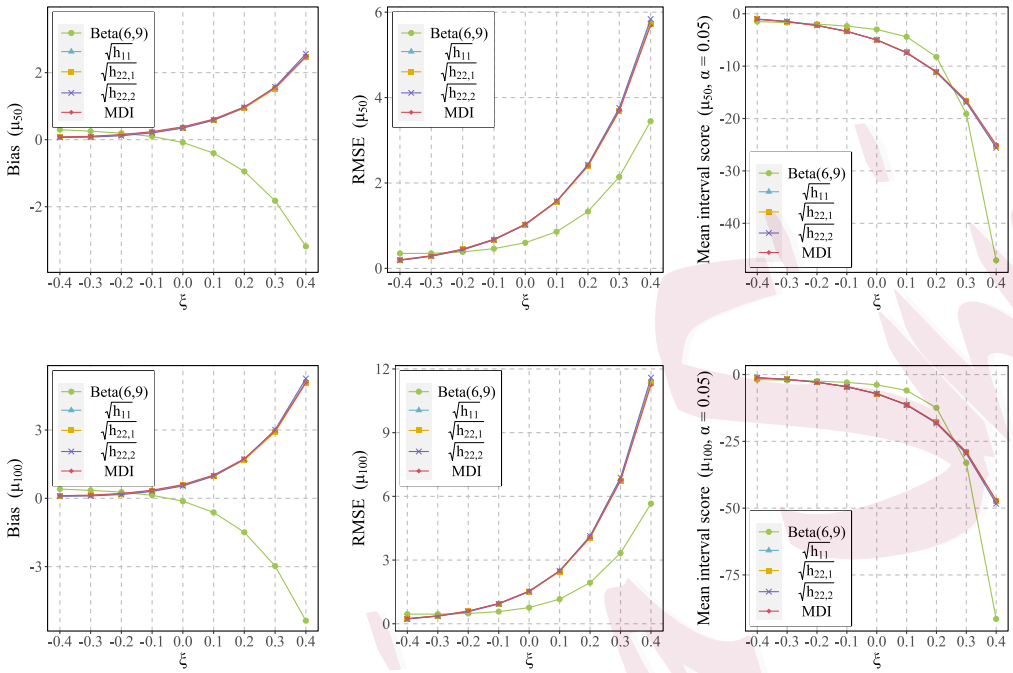

Figure 4: Bias, RMSE, and averaged interval scores of the 50-year return level (top row) and 100-year return level (bottom row) estimators under different true ξ values, where the sample size $n = 50$ and the number of simulations $N = 1,000$.

For this combination of beta prior and data-generating model, it requires a fairly large sample size ($n \geq 120$) for the reference and MDI priors to achieve similar performance with respect to the RMSE and interval scores. With large sample sizes, the bias is considerably greater for the beta prior.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

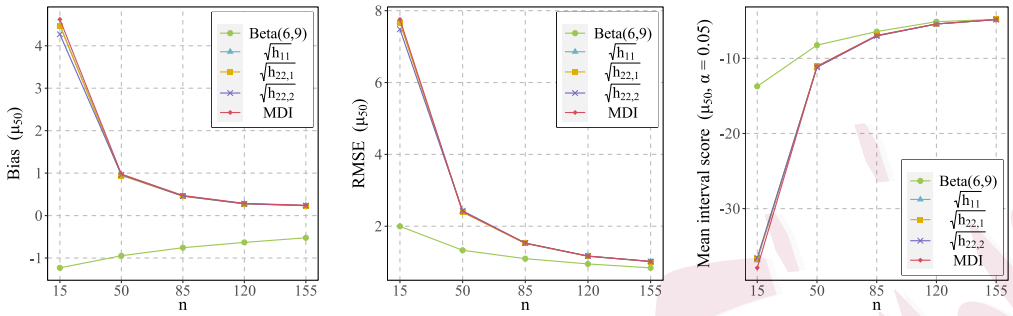

Figure 5: Bias, RMSE, and averaged interval scores of the 50-year return level under different sample sizes, in which the true $\xi = 0.2$ and the number of simulations $N = 1,000$. The same plots for the 100-year return level behave similarly, and thus are omitted.

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 2011–2017 drought further increasing the 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) for the period 1973 to 2018 at four monitoring stations (Dunn et al., 2012). These stations were the closest to four deadly wildfires that occurred in 2018–2019; see Figure 6 for the locations of the wildfires and the monitoring stations. The FFWI quantifies a potential wildfire threat by calculating a single number summary from temperature, wind speed, and relative humidity; here, larger index value reflects a greater

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

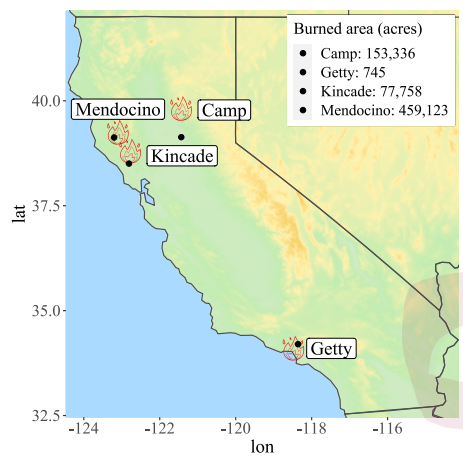

Figure 6: The locations of four notable wildfires in California that happened in 2018–2019. The black dots are the closest available monitoring fire stations to each wildfire.

risk of rapid drying and high flame lengths (Fosberg, 1978). The FFWI does not account for human activities and fuel sources, such as changes in land management practices and the incursion of invasive species that may provide additional fuel. Studying the tail behavior of the FFWI will provide useful measures on the potential effects of weather on the threat of fire.

First, we analyze the 46 annual maxima of the 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 the MCMC, we conduct a rough grid search to find the parameters that approximately maximize the unnormalized posterior $p(\boldsymbol{x} \mid \theta)\pi(\theta)$, when $\pi(\theta)$ is the prior of choice. For each

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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 are similar for μ and τ across all stations, and the posterior means are almost identical. As expected, the beta prior generates slightly narrower credible intervals for the shape parameter ξ, the value of which is confined between -0.25 and 0.15 for Mendocino, Getty, and Camp. However, there is an evident disparity among the estimates of ξ for Kincade, which suggests 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 140 km away, and has much lower ξ values than those of the station near Kincade, although their weather patterns may still differ in potentially important respects, despite their close proximity.

In the above analysis, we have assumed there is no trend in the distribution of the annual maximum FFWI values. To examine whether this assumption is plausible, we look for time trends in the 50-year return levels. We collect the annual maxima in backward 20-year sliding windows from 1992–2018, and perform similar Bayesian analyses to those described previously, for each 20-year window, and for each station. Because the non-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

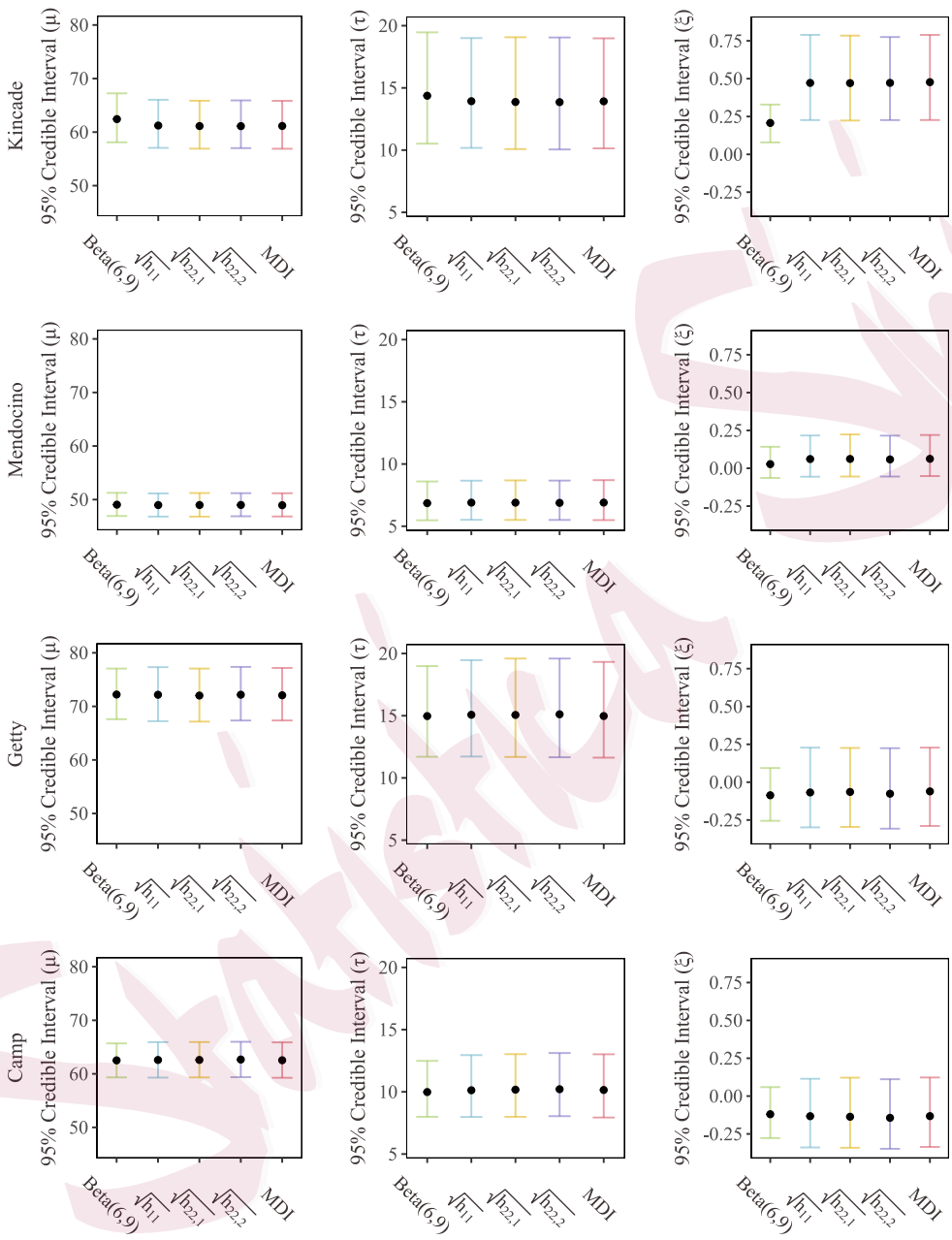

Figure 7: Posterior means and 95% credible intervals from the posterior distributions for μ, τ, and ξ, where each row shows the results from one monitoring station. For better comparison, the scale limits for the y-axis are set to be the same for each parameter across the rows.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

informative priors perform similarly, we use only the beta prior and the reference prior under the ordering (μ, ξ, τ), which corresponds to the reference prior that prioritizes an inference on the return level. Figure 8 shows the posterior mean and 95% credible intervals of the 50-year return levels, 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 the 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 may also suggest that the recent surge in the scale and frequency of wildfires in California has more to do with fuel availability, human activity, and land use than it does with changes in weather factors.

5. Discussion

We have used the procedure in Bernardo (2005) to derive reference priors for the family of GEV 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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

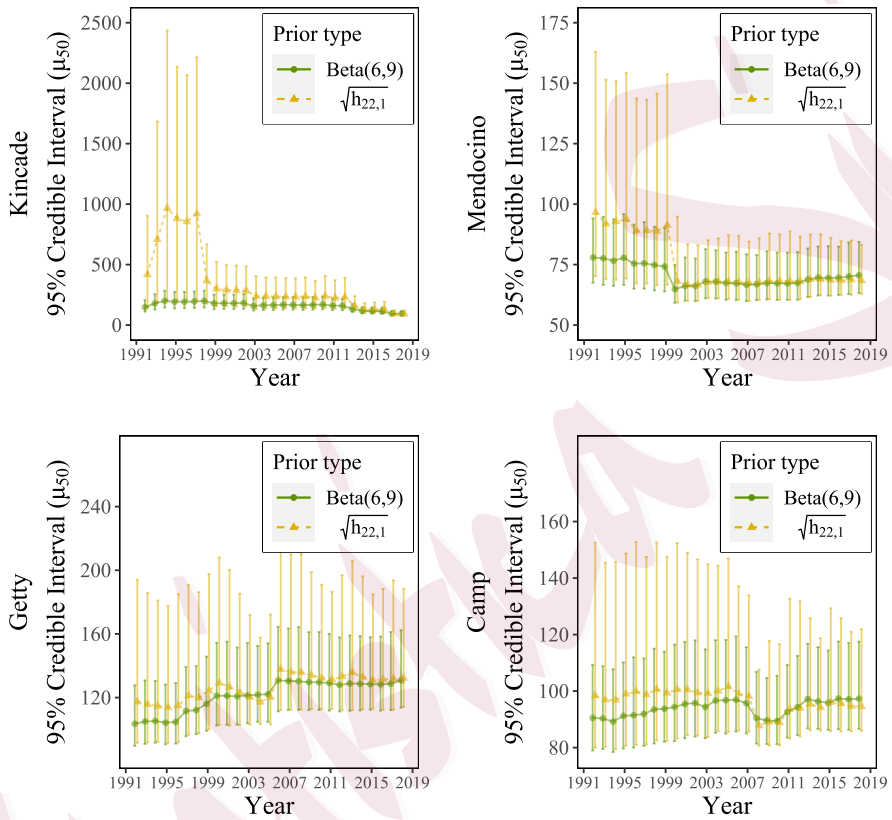

Figure 8: Posterior means and 95% credible intervals for the 50-year return levels calculated from the posterior distributions of every MCMC iteration. For each station, the MCMC analyses were performed using annual maxima in a 20-year sliding window.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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 have provided upper bounds for the posterior normalizing constants under the orderings that give permissible reference priors, finding that the posterior is improper under reference priors that treat ξ as the 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 with that of other priors suggested in the literature, under varying values of the shape parameter ξ and the sample size n. For ξ within a reasonable range, the beta prior seems to perform best; when $|\xi| < 0.3$, it has the best RMSE and interval scores for the estimation of ξ and the return levels, even though its bias is consistently worse than that of the noninformative priors we tested, owing 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, because the bias is lower, and the RMSE and interval scores are comparable to those of the beta prior. When $|\xi| > 0.3$,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0258

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 ξ is available, we conclude, unsurprisingly, that using a well-designed informative prior can improve results by stabilizing the estimation of the shape parameter ξ, especially when the number of observations is small. This is feasible for well-studied variables, such as precipitation and temperature, but it is somewhat dangerous for less common variables, such as the FFWI, because it is much less clear how to design reasonable informative priors, and the estimation performance might suffer as a result of poor choices.

In contrast, an advantage of using rule-based noninformative 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 noninformative priors based on the underlying principles from which they are derived. 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 to estimate a return level, then the reference prior corresponding to the parameter ordering

(μ_T, ξ, σ) seems the natural one to choose.

Supplementary Material

The online Supplementary Material provides details on the construction of the reference priors for the family of GEV distributions. It also contains the proof for the posterior propriety of the reference priors, and some diagnostics for 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.

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. 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.

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