

## Article

# Baseline Methods for the Parameter Estimation of the Generalized Pareto Distribution

Jacinto Martín <sup>1,†</sup> , María Isabel Parra <sup>1,†</sup> , Mario Martínez Pizarro <sup>2,†</sup>  and Eva López Sanjuán <sup>2,\*,†</sup> 

<sup>1</sup> Departamento de Matemáticas, Facultad de Ciencias, Universidad de Extremadura, 06006 Badajoz, Spain; jmartin@unex.es (J.M.); mipa@unex.es (M.I.P.)

<sup>2</sup> Departamento de Matemáticas, Centro Universitario de Mérida, Universidad de Extremadura, 06800 Mérida, Spain; mariomp@unex.es

\* Correspondence: etlopez@unex.es

† The authors contributed equally to this work.

**Abstract:** In the parameter estimation of limit extreme value distributions, most employed methods only use some of the available data. Using the peaks-over-threshold method for Generalized Pareto Distribution (GPD), only the observations above a certain threshold are considered; therefore, a big amount of information is wasted. The aim of this work is to make the most of the information provided by the observations in order to improve the accuracy of Bayesian parameter estimation. We present two new Bayesian methods to estimate the parameters of the GPD, taking into account the whole data set from the baseline distribution and the existing relations between the baseline and the limit GPD parameters in order to define highly informative priors. We make a comparison between the Bayesian Metropolis–Hastings algorithm with data over the threshold and the new methods when the baseline distribution is a stable distribution, whose properties assure we can reduce the problem to study standard distributions and also allow us to propose new estimators for the parameters of the tail distribution. Specifically, three cases of stable distributions were considered: Normal, Lévy and Cauchy distributions, as main examples of the different behaviors of the tails of a distribution. Nevertheless, the methods would be applicable to many other baseline distributions through finding relations between baseline and GPD parameters via studies of simulations. To illustrate this situation, we study the application of the methods with real data of air pollution in Badajoz (Spain), whose baseline distribution fits a Gamma, and show that the baseline methods improve estimates compared to the Bayesian Metropolis–Hastings algorithm.

**Keywords:** Bayesian inference; generalized Pareto distribution; Metropolis–Hastings algorithm; stable distributions; extreme value theory



**Citation:** Martín, J.; Parra, M.I.; Pizarro, M.M.; Sanjuán, E.L. Baseline Methods for the Parameter Estimation of the Generalized Pareto Distribution. *Entropy* **2022**, *24*, 178. <https://doi.org/10.3390/e24020178>

Academic Editor: Antonio M. Scarfone

Received: 17 November 2021

Accepted: 21 January 2022

Published: 25 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Extreme value theory (EVT) is a set of statistical tools employed for modeling and predicting the occurrence of rare events outside the range of available data. It has been widely used to study events that are more extreme than any previously observed, e.g., in disciplines such as climatology: extreme events of temperature [1–4], precipitations [5–10], and solar climatology [11–13]; finance and insurance: applications to risk theory [14–20]; and engineering: design for modern buildings [21].

There are two approaches for modeling an extreme value problem. The first one is the block maxima method: dividing the sample space into blocks of equal size, the maxima values of each block follow, under a certain domain of attraction conditions, approximately a Generalized Extreme Value (GEV) distribution [22]. The second way to deal with an extreme value data set attempts to make use of the available information about the upper tail of the distribution than just the block maxima. The so-called *peaks-over-threshold* (POT) method is due to hydrologists trying to model floods. Loosely speaking, References [23,24] showed that when we consider the distribution of data above a certain threshold  $u$ , it can

usually be approximated by a properly scaled Generalized Pareto distribution (GPD) as  $u$  tends to the endpoint of the distribution. This is the point of view considered in this article.

Due to its importance, several methods have been proposed for estimating the shape and scale parameters of the GPD. Classical methods include the method of moments, probability weighted moments, maximum likelihood and others. An exhaustive review of them can be consulted in [25,26]. Other researchers have proposed generalizations of the GPD: Reference [27] proposed a three-parameter Pareto distribution and employed POT to make estimations of Value at Risk and [28] introduced an extension of GPD and performed parametric estimation. However, classical methods might be inappropriate in certain situations, as explained in [26]. That is why Bayesian inference could be advisable. There are not many approaches to the GPD parameter estimation through Bayesian techniques. We can cite [29], who recommended the use of conjugate prior distributions; Reference [30] estimated the shape parameter when it is positive, and the computation of the posterior distribution was implemented via the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling; Reference [31] employed Gamma and Pareto priors via MCMC; Reference [32] proposed a Bayesian mixture model, where the threshold was also a random parameter; Reference [33] employed Jeffrey's prior and Metropolis–Hastings (MH) method and [34] employed the GPD distribution itself as the prior density.

In this paper, we aim at seizing all the available information coming from data in order to estimate the parameters of the GPD in a way as accurate as possible. A similar idea was also implemented in [35], for the estimation of the parameters of the Gumbel distribution when the baseline distribution is Gumbel, Exponential or Normal.

We will take into account all the data of the baseline distribution and study the relation between the baseline parameters and the parameters of the limit GPD in order to incorporate such relation into the sketch of new methods to make estimations. Concretely, we propose two methods and compare them with the classical MH method for data over the threshold. In addition, we will analyze four particular examples of underlying distribution: Exponential, Lévy, Cauchy and Normal, and make a special study of stable distributions.

## 2. Generalized Pareto Distribution and Its Relation with Extreme Value Theory

Let  $X$  be a random variable with distribution function  $F$ . Define  $u$  as the threshold value and let  $X_u = X - u | X > u$  be the random variable with distribution function

$$F_{X_u}(x) = P[X_u \leq x] = P[X \leq x + u | X > u] = \frac{F(x + u) - F(u)}{1 - F(u)},$$

for  $0 \leq x \leq x_F - u$ , being  $x_F$  the right endpoint of  $F$ , that is,  $x_F := \sup\{x : F(x) < 1\}$ .

Notice that  $X_u$  is the random variable that we obtain when we consider the distribution of data above the threshold, which we usually call the tail distribution.

Given a random variable  $X$ , we say that it follows a Generalized Pareto Distribution (GPD) when its distribution function is

$$G(x|\zeta, \sigma) = \begin{cases} 1 - \left(1 + \frac{\zeta x}{\sigma}\right)^{-1/\zeta}, & \zeta \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & \zeta = 0 \end{cases} \quad (1)$$

where  $\sigma > 0$  and  $\zeta \in \mathbb{R}$  are the scale and shape parameters, respectively. Equation (1) is valid when  $x \geq 0$  for  $\zeta \geq 0$ , and for  $0 \leq x \leq -\sigma/\zeta$  for  $\zeta < 0$ .

The fundamental result that connects EVT and GPD distribution belongs to [23,24], and it establishes that under certain mild conditions, for a random variable  $X$ , the distribution of  $X_u$  for a sufficiently high threshold  $u$  follows a properly scaled Generalized Pareto distribution (GPD).

We will call the distribution function of  $X$ ,  $F$ , the baseline distribution of the GPD. The parameters  $\zeta$  and  $\sigma$  will depend on the value of the threshold  $u$ , and on the baseline distribution. For example,  $\zeta$  is determined by the shape of the upper tail of the baseline distribution  $F$ . Positive values of the shape parameter correspond to heavy tails, while

negative ones come from light tails. The special case  $\zeta = 0$  will appear when the upper tail of the distribution tends to an exponential distribution of parameter  $1/\sigma$ .

In this work, we will consider several types of baseline distributions for the GPD. Traditionally, estimation for the parameters of the limit GPD in Extreme Value Theory has been made taking into account only values above the threshold, but this information might be scarce. We propose a new strategy consisting in seizing all the data from the baseline distribution. We will show how this strategy can produce accurate estimations for the parameters of the GPD.

In the case when the baseline distribution is an exponential distribution with parameter  $\lambda$ , with distribution function  $F(x) = 1 - e^{-\lambda x}$ ,  $x \geq 0$ , for every  $u \geq 0$ ,

$$\begin{aligned} F_{X_u}(x) &= \frac{1 - e^{-\lambda(x+u)} - (1 - e^{-\lambda u})}{e^{-\lambda u}} \\ &= 1 - e^{-\lambda x} = F(x), \quad \forall x \geq 0. \end{aligned}$$

Consequently, the tail distribution  $X_u$  is the same as the underlying distribution in the exponential case. Therefore, we must employ all the available data to estimate the parameter  $\lambda = 1/\sigma$  in the definition of the GPD (1).

The case when  $\zeta \neq 0$  is different. In this paper, we will consider some of the most employed distributions as underlying distributions: normal distribution, which has light tails ( $\zeta < 0$ ); and the Cauchy and Lévy distributions, which lead to heavy tails ( $\zeta > 0$ ). Those distributions are stable; therefore, they have additional properties that will be helpful to estimate the parameters of the GPD. We will study such properties below.

With respect to the threshold, it can be settled as a known value, which has a physical meaning, depending on the characteristics of the data, or it can be defined as an upper order statistic. It is generally defined as a  $p$ -quantile of the underlying distribution  $q_p$ , for appropriate values of  $p$ .

### 3. Stable Distributions

Stable distributions are a rich class of probability distributions characterized by [36] and they have been proposed as a model for many types of physical and economic systems because of their interesting properties. They are the attractors of sums of independent, identically distributed distributions whether or not the mean or variance is finite. Good references to study them are [16] or [37].

Let  $Z$  be a random variable with parameters defined by its characteristic function:

$$E[e^{itZ}] = \begin{cases} \exp\left\{-|t|^\alpha \left(1 - i\beta \tan \frac{\pi\alpha}{2} \operatorname{sign}(t)\right)\right\}, & \text{if } \alpha \neq 1 \\ \exp\left\{-|t| \left(1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \log |t|\right)\right\}, & \text{if } \alpha = 1 \end{cases} \quad (2)$$

where the parameter  $\alpha \in (0, 2]$  is called the index of stability, and  $\beta \in [-1, 1]$  is the skewness parameter. When  $\beta = 0$ , the distribution is symmetric.

A random variable  $X$  is said to follow a stable distribution with parameters  $a > 0$  and  $b \in \mathbb{R}$  if it satisfies that

$$X = aZ + b. \quad (3)$$

Generally, densities can be expressed only by complicated special functions, but there are three special cases of stable distributions that have probability density functions, which can be expressed analytically:

- When  $\alpha = 1/2$  and  $\beta = 1$ , we obtain Lévy distributions,  $Z \sim \mathcal{L}(0, 1)$ . If  $X \sim \mathcal{L}(\gamma, \delta)$ , then  $a = \delta$  and  $b = \gamma$  in (3).
- When  $\alpha = 1$  and  $\beta = 0$ , we obtain the family of Cauchy distributions,  $Z \sim \mathcal{C}(0, 1)$ . If  $X \sim \mathcal{C}(\gamma, \delta)$ , then  $a = \delta$  and  $b = \gamma$  in (3).
- When  $\alpha = 2$  and  $\beta = 0$ , we obtain the normal distribution,  $N(0, \sqrt{2})$ . If  $X \sim N(\mu, \sigma)$ , then  $a = \sigma$  and  $b = \mu$  in (3). As usual, in this case we will denote  $Z \sim N(0, 1)$ .

In particular, as we can standardize the stable distributions of a family, the  $p$ -quantiles  $q_p$  for a stable distribution  $X$ , can be expressed in terms of the  $p$ -quantiles  $z_p$  of the standard distribution  $Z$ , as

$$q_p = az_p + b. \tag{4}$$

Let us assume that  $X$  follows a stable distribution with parameters  $a$  and  $b$  for Equation (3), and fix  $u = q_p$  as the threshold for the problem of extreme value. Then, for  $u$  large enough,  $z_p$  from (4) is also large, and, consequently, denoting  $u_Z = z_p$ , Theorem 1 guarantees that  $Z_{u_Z} \sim GPD(\xi_Z, \sigma_Z)$ . Therefore, as

$$\frac{X_u}{a} = Z_{u_Z} \tag{5}$$

and

$$\begin{aligned} F_{X_u}(x) &= P(X_u \leq x) = P(X_u/a \leq x/a) \\ &\approx G\left(\frac{x}{a} \mid \xi_Z, \sigma_Z\right) = 1 - \left(1 + \frac{\xi_Z x}{\sigma_Z a}\right)^{-1/\xi_Z}, \end{aligned} \tag{6}$$

then

$$X_u \sim GPD(\xi_Z, a\sigma_Z). \tag{7}$$

The parameter  $\xi_Z$  will remain constant for all the random variables of the same stable family, whatever the parameters of the baseline variable are, while the scale parameter is obtained through the product of the standardization parameter  $a$  and the scale parameter  $\sigma_Z$  for the GPD limiting distribution of  $Z_{u_Z}$

$$\xi = \xi_Z, \sigma = a\sigma_Z. \tag{8}$$

In the case when the underlying distribution is a Cauchy or a Lévy, or any stable distribution  $X$  with the index  $0 < \alpha < 2$ , the tail of the distribution is considered to be "heavy", therefore it leads to a GPD where  $\xi > 0$ .

Stable distributions with index of stability  $\alpha \neq 2$ , (all of them except normal distribution) also verify an interesting property. As we can see in [36], given the standard distribution  $Z$ , its survival function  $\bar{F}$  can be approximated by:

$$\bar{F}(x) \sim (1 + \beta)C_\alpha x^{-\alpha}, \quad x \rightarrow \infty \tag{9}$$

where  $C_\alpha = \frac{1}{\pi} \Gamma(\alpha) \sin\left(\frac{\alpha\pi}{2}\right)$ .

From this approximation, we can infer that the shape of the tail of the distribution will only depend on the index of stability  $\alpha$ . Therefore, if we consider the GPD that models  $Z_{u_Z}$ , the shape parameter  $\xi_Z$  will be fixed. We will estimate it through simulation.

**Proposition 1.** *When the baseline distribution is a standard stable distribution  $Z$  with  $\alpha < 2$ , the relation between the parameters of  $Z$  and the parameters of the GPD that models the distribution above the  $p$ -quantile of  $Z$ ,  $u_Z$ , is:*

$$\hat{\xi}_Z = \frac{1}{\alpha}, \quad \hat{\sigma}_Z = \frac{1}{\alpha} \left( \frac{C_\alpha(1 + \beta)}{1 - p} \right)^{1/\alpha} \tag{10}$$

**Proof.** From Theorem 1, for  $u_Z$  that is big enough,

$$\bar{F}_{u_Z}(x) \sim \left(1 + \frac{\xi_Z}{\sigma_Z} x\right)^{-1/\xi_Z}$$

and by Proposition 1 (9), also for big values of  $u_Z$

$$\bar{F}_{u_Z}(x) \sim \frac{(1 + \beta)C_\alpha x^{-\alpha}}{\bar{F}_Z(u_Z)} = \frac{(1 + \beta)C_\alpha x^{-\alpha}}{1 - p}.$$

Then, making equal both expressions,

$$\left(1 + \frac{\xi_Z}{\sigma_Z} x\right)^{-1/\xi_Z} = \left(\left[\frac{(1 + \beta)C_\alpha}{1 - p}\right]^{-1/\alpha} x\right)^{-\alpha}$$

Therefore, we can take  $\hat{\xi}_Z = 1/\alpha$  as an estimator for  $\xi_Z$  (notice that the shape of the tail of the distribution depends only on the value of  $\alpha$ ). □

Substituting  $\xi_Z$  by  $1/\alpha$ , we have

$$1 + \frac{1}{\alpha\sigma_Z} x = \left[\frac{(1 + \beta)C_\alpha}{1 - p}\right]^{-1/\alpha} x,$$

so

$$1 = \left(\left[\frac{(1 + \beta)C_\alpha}{1 - p}\right]^{-1/\alpha} - \frac{1}{x}\right)\alpha\sigma_Z \sim \left[\frac{(1 + \beta)C_\alpha}{1 - p}\right]^{-1/\alpha} \alpha\sigma_Z$$

as  $1/x$  can be negligible. Therefore, we define

$$\hat{\sigma}_Z = \frac{1}{\alpha} \left(\frac{C_\alpha(1 + \beta)}{1 - p}\right)^{1/\alpha}.$$

In Section 5, we will assure the accuracy of these estimators through an extensive simulation study.

#### 4. Metropolis–Hastings (MH) Method

In this section, we will explain how to apply the Markov chain Monte Carlo (MCMC) method through the Metropolis–Hastings (MH) algorithm to make the estimations for stable distributions. We have to distinguish between light tails ( $\xi < 0$ ) and heavy tails ( $\xi > 0$ ). Let us assume  $X_u \sim GPD(\xi, \sigma)$  and that we dispose of  $m$  values.

Let  $\mathbf{x} = (x^1, \dots, x^n)$  be a sample of  $n$  values from  $X$  and  $\mathbf{x}_u = (x_u^1, \dots, x_u^m)$  be a sample of  $m$  values from  $X_u$ .

##### 4.1. Light Tails $\xi < 0$

Take  $k = -\xi$ , and  $\delta = -\frac{\sigma}{\xi}$ , so  $X_u \sim GPD(-k, k\delta)$ , with the likelihood function

$$L(k, \delta | \mathbf{x}_u) = k^{-m} \delta^{-m} \prod_{i=1}^m \left(1 - \frac{x_u^i}{\delta}\right)^{-1+1/k}$$

Considering  $\Gamma(a_0, b_0)$  and  $\Gamma(a_1, b_1)$  as prior distributions for both parameters. Then, the MH algorithm is applied:

1. Draw a starting sample  $(k^{(0)}, \delta^{(0)})$
2. For  $j = 0, 1, \dots$ 
  - Sample candidates  $k^*, \delta^*$  from the proposal distributions

$$k^* \sim \mathcal{N}(k^{(j)}, \nu_k), \delta^* \sim \mathcal{N}(\delta^{(j)}, \nu_\delta)$$

- Calculate the ratios

$$r_k = \frac{\pi(k^* | \delta^{(j)}, \mathbf{x}_u)}{\pi(k^{(j)} | \delta^{(j)}, \mathbf{x}_u)}, r_\delta = \frac{\pi(\delta^* | k^{(j)}, \mathbf{x}_u)}{\pi(\delta^{(j)} | k^{(j)}, \mathbf{x}_u)}$$

- Set

$$k^{(j+1)} = \begin{cases} k^*, & \text{with probability } \min\{1, r_k\} \\ k^{(j)}, & \text{otherwise} \end{cases}$$

$$\delta^{(j+1)} = \begin{cases} \delta^*, & \text{with probability } \min\{1, r_\delta\} \\ \delta^{(j)}, & \text{otherwise} \end{cases}$$

3. Iterate the former procedure.

Notice that

$$\begin{aligned} r_k &= \left(\frac{k^*}{k^{(j)}}\right)^{a_0-m-1} \exp\left\{b_0(k^* - k^{(j)}) + \left(\frac{1}{k^*} - \frac{1}{k^{(j)}}\right) \sum_{i=1}^m \ln\left(1 - \frac{x_u^i}{\delta^{(j)}}\right)\right\} \\ r_\delta &= \left(\frac{\delta^*}{\delta^{(j)}}\right)^{a_1-m-1} \exp\left\{b_1(\delta^* - \delta^{(j)}) + \left(\frac{1}{k^{(j)}} - 1\right) \sum_{i=1}^m \ln\left(1 - \frac{x_u^i}{\delta^*}\right)\right. \\ &\quad \left. - \left(\frac{1}{k^{(j)}} - 1\right) \sum_{i=1}^m \ln\left(1 - \frac{x_u^i}{\delta^{(j)}}\right)\right\} \end{aligned}$$

Finally, we obtain estimations for  $\xi$  and  $\sigma$  from  $\xi = -k$  and  $\sigma = k\delta$ .

#### 4.2. Heavy Tails $\xi > 0$

In this case, the likelihood function is

$$L(\xi, \sigma | \mathbf{x}_u) = \sigma^{-m} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)}$$

Taking a type I Pareto  $(a_0, b_0)$  as prior distribution for  $\xi$  and  $\text{Inv}\Gamma(a_1, b_1)$  for  $\sigma$ ,

$$\begin{aligned} \pi(\xi) &\propto \xi^{-(a_0+1)}, \text{ with } \xi > b_0 \\ \pi(\sigma) &\propto \sigma^{-(a_1+1)} \exp\left\{-\frac{b_1}{\sigma}\right\} \end{aligned}$$

Posterior conditional distributions are

$$\begin{aligned} \pi(\xi | \sigma, \mathbf{x}_u) &\propto \xi^{-(a_0+1)} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)} \\ \pi(\sigma | \xi, \mathbf{x}_u) &\propto \sigma^{-(m+a_1+1)} \exp\left\{-\frac{b_1}{\sigma}\right\} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)} \end{aligned}$$

Then, MH algorithm is applied, as in the previous case.

Notice that

$$\begin{aligned} r_\xi &= \left(\frac{\xi^{(j)}}{\xi^*}\right)^{a_0+1} \exp\left\{\left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^{(j)}}\right) - \left(1 + \frac{1}{\xi^*}\right) \sum_{i=1}^m \ln\left(1 + \xi^* \frac{x_u^i}{\sigma^{(j)}}\right)\right\} \\ r_\sigma &= \left(\frac{\sigma^{(j)}}{\sigma^*}\right)^{m+a_1+1} \exp\left\{b_1\left(\frac{1}{\sigma^{(j)}} - \frac{1}{\sigma^*}\right) + \left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^{(j)}}\right)\right. \\ &\quad \left. - \left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^*}\right)\right\} \end{aligned}$$

## 5. Baseline MH Method (BMH)

In this section, we will introduce Baseline Metropolis–Hastings (BMH) method, designed according to the objectives of seizing all the available data from the baseline distribution and making use of the existing relations between the baseline parameters and the limit GPD parameters. The method consists of:

1. Applying the MH algorithm to estimate the parameters for the baseline distribution  $\theta$ .
2. Making use of the relations between the baseline parameters  $\theta$  and the GPD parameters  $\zeta$  and  $\sigma$  to compute estimations for  $\zeta$  and  $\sigma$ .

In the case of stable distributions, these relations have been explained in previous sections and are given by (8). We will detail below the application of the BMH method for the three selected stable distributions.

For the rest of the baseline distributions, the strategy to find such relations would be made thorough studies of simulation, in order to establish correspondences between the baseline parameters and the tail GPD parameters. At the moment, there are no studies in the literature about this subject; therefore, it would be interesting to perform them in future research.

Then, in the case of stable distributions, the application of BMH would be:

1. Apply the MH algorithm to estimate scale parameter  $a$  from the stable baseline distribution.
2. Make use of the relation (8) to compute estimations for  $\zeta$  and  $\sigma$ , using estimators for  $\zeta_Z$  and  $\sigma_Z$  that we detail below.

### 5.1. Estimations for $\zeta_Z$ and $\sigma_Z$

In order to provide good estimations for  $\zeta_Z$  and  $\sigma_Z$ , we made a thorough simulation study for the three baseline distributions we have considered: Lévy, Cauchy, and Normal distribution. We took values for  $p \in [0.990, 0.995]$  with increments of 0.001, and set the threshold  $u_Z = q_p$ . For each distribution, and for each value of the threshold,  $m = 1000$  values from  $Z_{u_Z}$  were generated. This sequence was repeated 100 times. Therefore, we obtained 100 point estimations for each  $p$ .

To guarantee the convergence of the MCMC algorithm, we must be sure that the posterior distribution has been reached. These proceedings were made using library coda [38] for R software, taking 10,000 values for the burn-in period, 25 values for the thinning, and selecting initial values for each sample. Finally, to obtain the posterior distribution for each parameter, a Markov chain of length 10,000 was obtained and we considered the mean as the estimator. The results of the simulation study are shown in Figure 1.

#### 5.1.1. Lévy and Cauchy Baseline Distribution

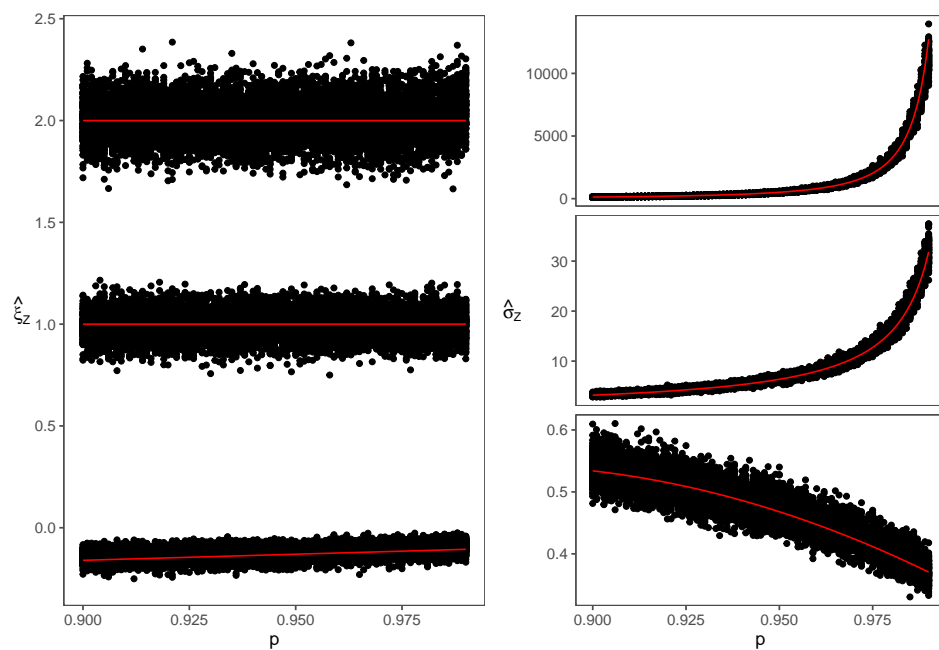
By Proposition 1 (10), we had estimations for  $\zeta_Z$  and  $\sigma_Z$ . Concretely, these are

$$\hat{\zeta}_Z = 2, \quad \hat{\sigma}_Z = \frac{4}{\pi}(1-p)^{-2}, \quad (11)$$

for the Lévy distribution and,

$$\hat{\zeta}_Z = 1, \quad \hat{\sigma}_Z = \frac{1}{\pi}(1-p)^{-1} \quad (12)$$

for the Cauchy distribution.



**Figure 1.** Estimations for  $\zeta_Z$  (left) and  $\sigma_Z$  (right) for Lévy (upper charts), Cauchy (middle charts), Normal (lower charts) and estimators from Equations (11)–(13) plotted in red for  $p \in [0.900, 0.995]$ .

5.1.2. Normal Baseline Distribution

In the case of the normal distribution, property (10) is not verified. In this case, the adjustment from the simulation study is:

$$\hat{\zeta}_Z = -0.7 + 0.61p, \quad \hat{\sigma}_Z = 0.34 + 3.18(1 - p) - 12.4(1 - p)^2. \tag{13}$$

Now, we will estimate the scale parameter  $a$  of the baseline distribution  $X$ . Notice that parameter  $b$  for the stable distribution  $X$  defined in (3) does not have any influence on the estimation of the parameters of the GPD, as we have shown before. Consequently, we will assume  $b = 0$  from now on.

5.2. Lévy Distribution

Likelihood function for Lévy distribution is:

$$L(a|\mathbf{x}) \propto a^{n/2} \exp\left\{-\frac{a}{2} \sum_{i=1}^n \frac{1}{x^i}\right\}$$

Taking a prior distribution  $\Gamma(a_0, b_0)$  for  $a$ , and making use of (8) and (11), we obtain estimations  $\hat{\zeta}$  and  $\hat{\sigma}$ .

5.3. Cauchy Distribution

Likelihood function for Cauchy is:

$$L(a|\mathbf{x}) \propto a^{-n} \prod_{i=1}^n \left(1 + \left(\frac{x^i}{a}\right)^2\right)^{-1}$$

Taking  $\Gamma(a_0, b_0)$  as prior distribution and, making use of (8) and (12), we obtain estimations  $\hat{\zeta}$  and  $\hat{\sigma}$ .

5.4. Normal Distribution

In this case, we consider  $\text{Inv}\Gamma(a_0, b_0)$  as prior distribution for  $a^2$ , and making use of (8) and (13), we obtain estimations  $\hat{\zeta}$  and  $\hat{\sigma}$ .



### 6. Informative Priors Baseline MH (IPBMH)

Finally, we propose an MH method to estimate the parameters  $\xi$  and  $\sigma$ , where highly informative priors are employed, seizing the estimations computed before. Employing estimations of  $\xi_Z, \sigma_Z$  and  $a$  obtained in previous sections, we can settle priors for  $\xi$  and  $\sigma$ , which are very informative. Notice that this way of proceeding keeps the original idea of Extreme Value Theory, granting more weight to tail values because they are employed twice: to compute estimations for  $\xi_Z$  and  $\sigma_Z$  and also through the likelihood function. As we commented before, this method could also be implemented for other baseline distributions once we have found relations between baseline and GPD parameters.

For stable distributions, highly informative priors are

$$\xi \sim N(\xi_Z, b_1), \sigma \sim N(a \cdot \sigma_Z, b_2)$$

Then, for the three distributions studied,  $a$  is estimated through BMH and  $\xi_Z, \sigma_Z$  are estimations computed through (11)–(13). In addition,

- $b_1$  is constant, being 0.03, 0.065 and 0.1 for Normal, Cauchy and Lévy baseline distributions, respectively.
- $b_2 = \exp\{c_1 p^2 + c_2 p + c_3\}$ , where values are given in Table 1.

**Table 1.** Values for  $c_1, c_2, c_3$  for the three baseline distributions.

Distribution	$c_1$	$c_2$	$c_3$
Lévy	500.2	−900.9	408.2
Cauchy	323.57	−588.51	266.13
Normal	−46.24	83.55	−41.58

The Joint posterior distribution is

$$\pi(\xi, \sigma | \mathbf{x}_u) \propto \sigma^{-m} \exp\left\{-\frac{1}{2b_1^2}(\xi - \xi_Z)^2 - \frac{1}{2b_2^2}(\sigma - a \cdot \sigma_Z)^2\right\} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)}$$

and marginal distributions are

$$\begin{aligned} \pi(\xi | \sigma, \mathbf{x}_u) &\propto \exp\left\{-\frac{1}{2b_1^2}(\xi - \xi_Z)^2\right\} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)} \\ \pi(\sigma | \xi, \mathbf{x}_u) &\propto \sigma^{-m} \exp\left\{-\frac{1}{2b_2^2}(\sigma - a \cdot \sigma_Z)^2\right\} \prod_{i=1}^m \left(1 + \xi \frac{x_u^i}{\sigma}\right)^{-(1+1/\xi)} \end{aligned}$$

Then, we apply the MH algorithm with

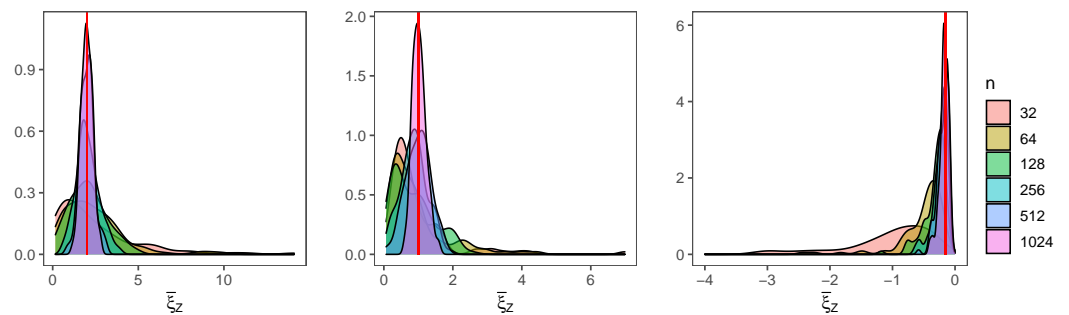
$$\begin{aligned} r_\xi &= \exp\left\{\frac{1}{2b_1^2} \left( (\xi^{(j)} - \xi_Z)^2 - (\xi^* - \xi_Z)^2 \right) \right. \\ &\quad \left. - \left(1 + \frac{1}{\xi^*}\right) \sum_{i=1}^m \ln\left(1 + \xi^* \frac{x_u^i}{\sigma^{(j)}}\right) + \left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^{(j)}}\right) \right\} \\ r_\sigma &= \left(\frac{\sigma^{(j)}}{\sigma^*}\right)^m \exp\left\{\frac{1}{2b_2^2} \left( (\sigma^{(j)} - a \cdot \sigma_Z)^2 - (\sigma^* - a \cdot \sigma_Z)^2 \right) \right. \\ &\quad \left. - \left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^*}\right) + \left(1 + \frac{1}{\xi^{(j)}}\right) \sum_{i=1}^m \ln\left(1 + \xi^{(j)} \frac{x_u^i}{\sigma^{(j)}}\right) \right\} \end{aligned}$$

### 7. Simulation Study

Now, we will develop a thorough simulation study in order to compare the accuracy of the three MH methods of estimation: MH, BMH and IPBMH.

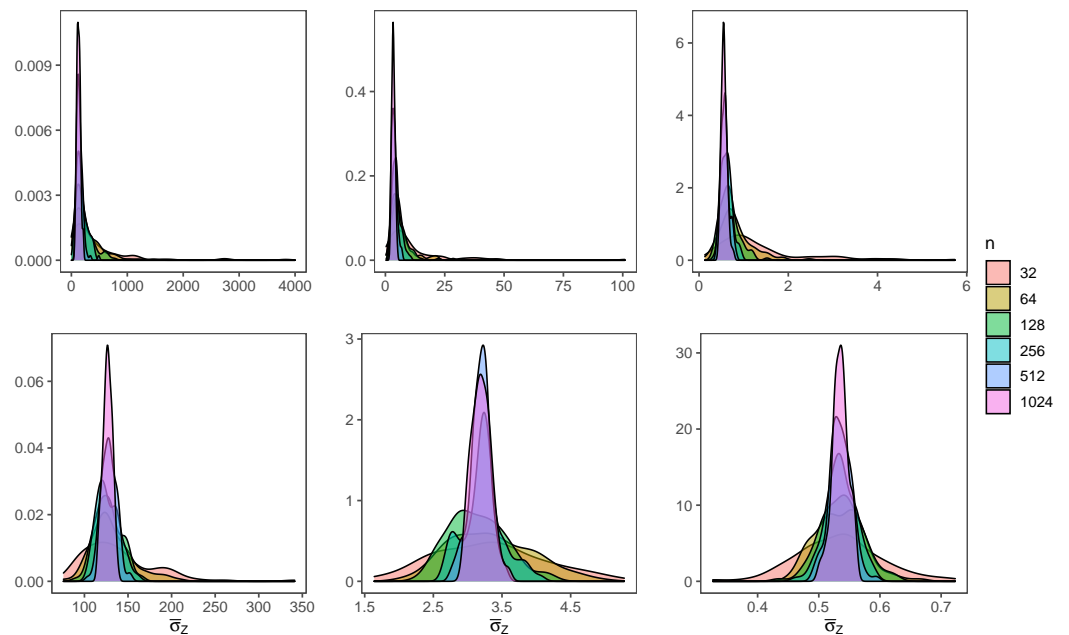
We fixed  $p = 0.9$  and the threshold  $u = q_p$ . Furthermore, we take  $b = 0$ . We sample  $n = 2^i$ , with  $i = 5, \dots, 10$  values from the three baseline distributions considered, with scale  $a = 2^j$ ,  $j = -2, -1, 0, 1, 2$ . We obtained an MCMC with length 10,000, taking 10,000 values for the burn-in period, 25 values for the thinning and selecting initial values for each sample. Finally, this sequence was repeated 100 times and we considered the mean as the estimator.

In Figure 2, we can see the posterior distribution of the parameters  $\zeta_Z$  and  $\sigma_Z$  for the different sample sizes  $n$ , when the baseline distribution is  $\mathcal{L}(0,1)$  (left),  $\mathcal{C}(0,1)$  (middle) and  $\mathcal{N}(0,1)$  (right), for the methods MH and BMH. For the first one, the distribution is right skewed, although skewness becomes smaller as  $n$  increases. BMH offers a point estimation, plotted as a vertical red line.



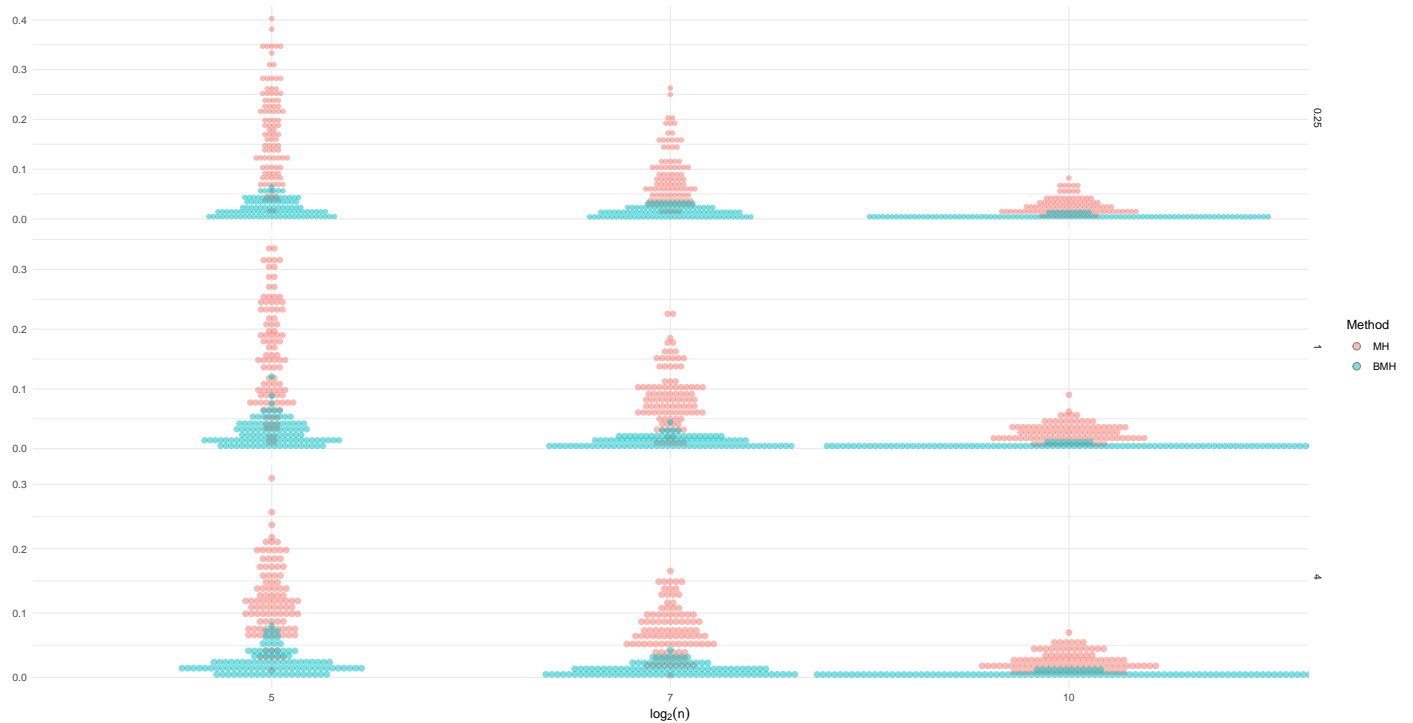
**Figure 2.** Posterior distribution of  $\zeta_Z$  for the different values of  $n$ , when the baseline distribution is  $\mathcal{L}(0,1)$  (left),  $\mathcal{C}(0,1)$  (middle) and  $\mathcal{N}(0,1)$  (right), for the MH method. The estimation of BMH is plotted as a vertical red line.

In Figure 3, we can see the posterior distribution for  $\sigma_Z$ , for both methods. MH (upper charts) shows much more skewness, such as in the case of  $\zeta_Z$ , while estimations from BMH (lower charts) are less skewed and dispersed.

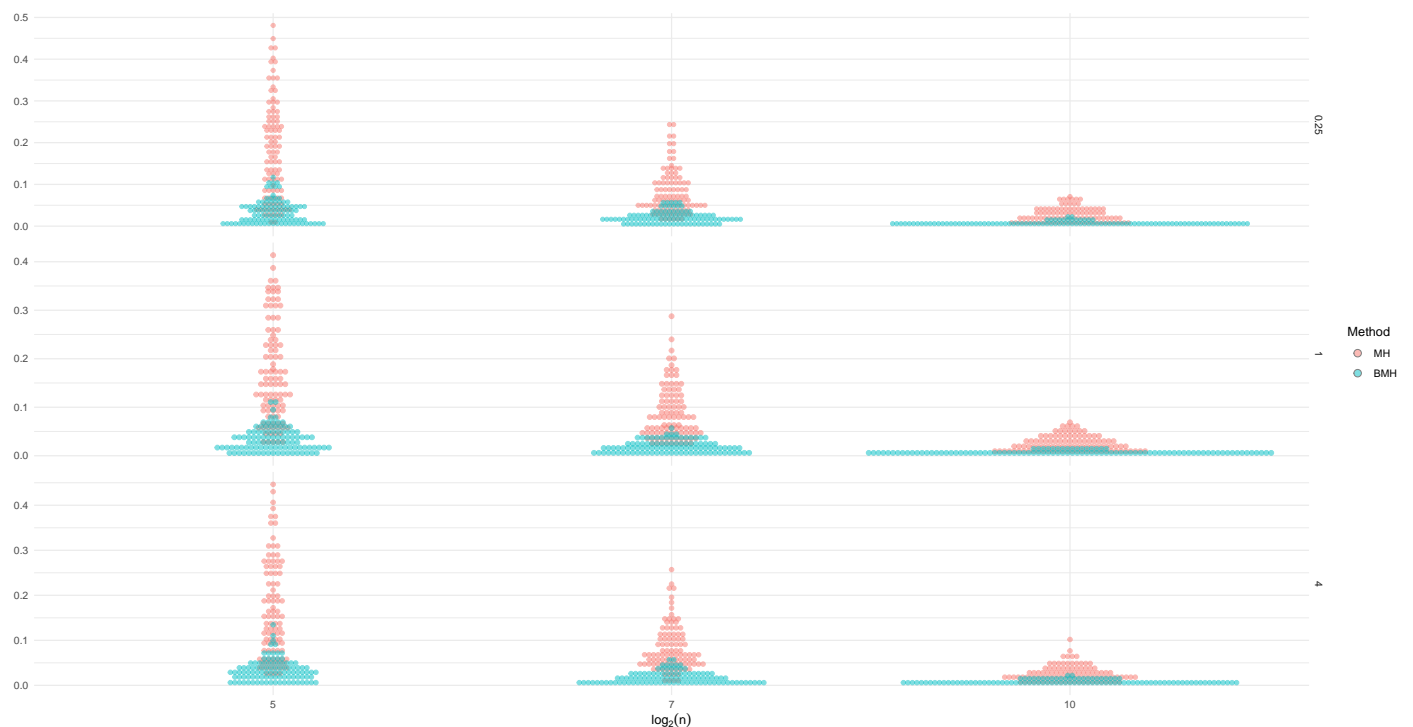


**Figure 3.** Posterior distribution of  $\sigma_Z$  for the different values of  $n$ , when the baseline distribution is  $\mathcal{L}(0,1)$  (left),  $\mathcal{C}(0,1)$  (middle) and  $\mathcal{N}(0,1)$  (right), for the method MH (upper charts) and BMH (lower charts).

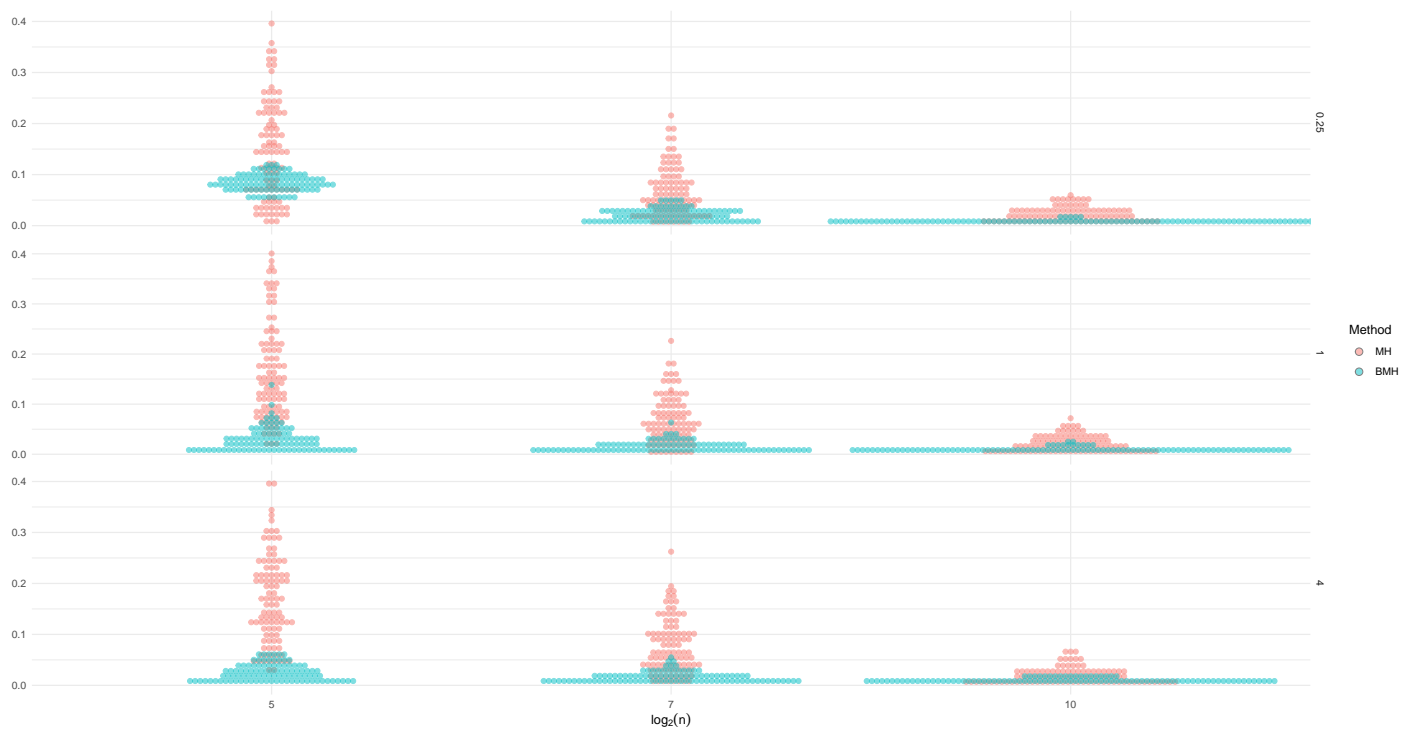
Now, we will compare mean absolute errors (MAE) for MH and BMH when  $a = 0.25, 1, 4$ , and for sample sizes  $n = 2^5, 2^7, 2^9$ . In Figures 4–6, we can see how BMH provides smaller errors than MH.



**Figure 4.** MAE for MH and BMH when  $a = 0.25, 1, 4$ , and for sample sizes  $n = 2^5, 2^7, 2^9$ , for the baseline distribution  $\mathcal{L}(0, 1)$ .



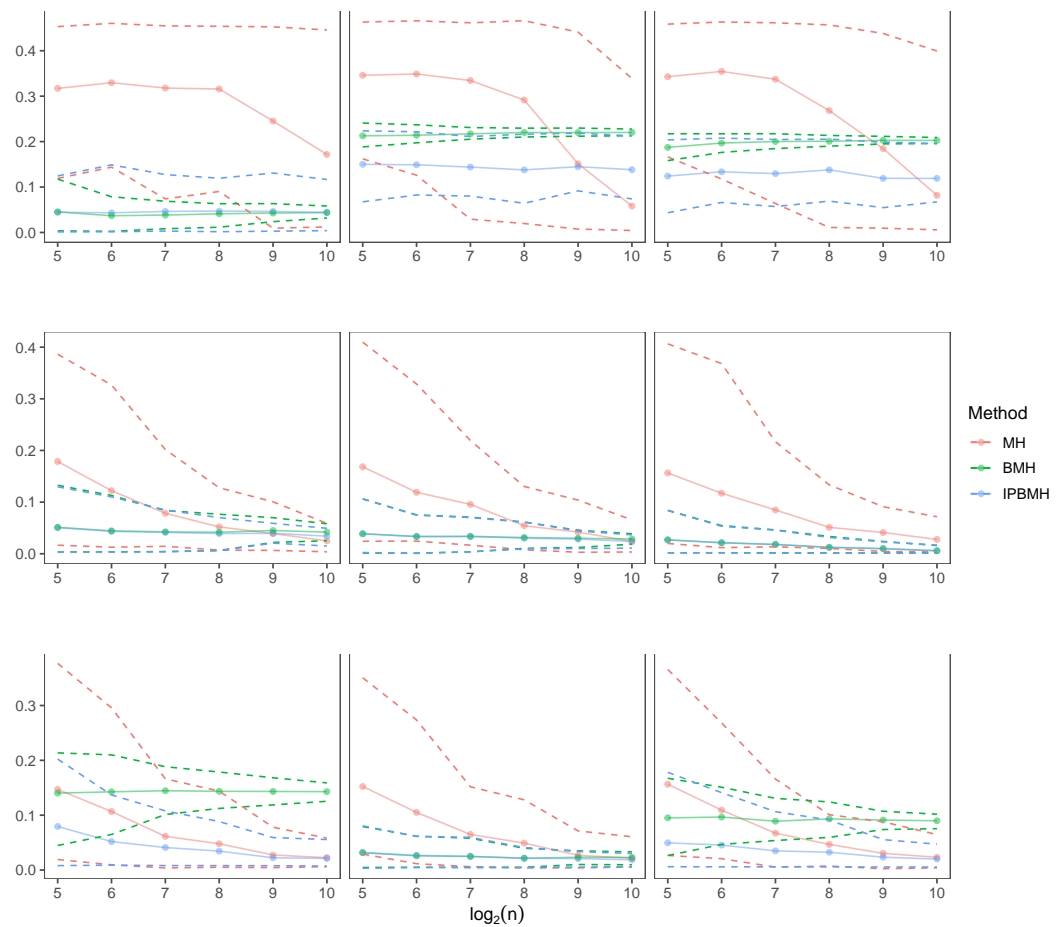
**Figure 5.** MAE for MH and BMH when  $a = 0.25, 1, 4$ , and for sample sizes  $n = 2^5, 2^7, 2^9$ , for the baseline distribution  $\mathcal{C}(0, 1)$ .



**Figure 6.** MAE for MH and BMH when  $a = 0.25, 1, 4$ , and for sample sizes  $n = 2^5, 2^7, 2^9$ , for the baseline distribution  $\mathcal{N}(0, 1)$ .

Clearly, BMH provides more accurate estimations for the parameters of the GPD when the baseline distribution is stable and known. However, in practical situations, we might not know which is the baseline distribution, or data could better fit a mixture of distributions rather than a simple one. In these situations, the use of highly informative priors, built with the information available from all the data, could be advisable. To develop this idea, we simulated from different mixtures and computed values of MAE for the three methods, finding that IPBMH is the method that shows the best behavior when data differ from the simple distributions.

In Figure 7, we can see MAE for the three methods, in the case of the mixtures employed ( $\alpha = 0.5$ ):  $\alpha F(0, 1/2) + (1 - \alpha)F(0, 2)$  (left charts),  $\alpha F(0, 1) + (1 - \alpha)F(1, 1)$  (middle charts) and  $\alpha F(0, 1) + (1 - \alpha)F(1, 2)$  (right charts), for the three baseline distributions considered (Lévy, Cauchy, Normal). In general, IPBMH is the most advisable method, especially for the case when data are scarce. Notice that when data approaches one of the pure stable distributions, for example, in the case of Cauchy mixtures (which are still quite similar to a simple Cauchy) and the second mixture for the Normal distribution, BMH and IPBMH show very similar results. However, when the mixtures differ significantly from the simple distribution, the method IPBMH and MH are more advisable than BMH.

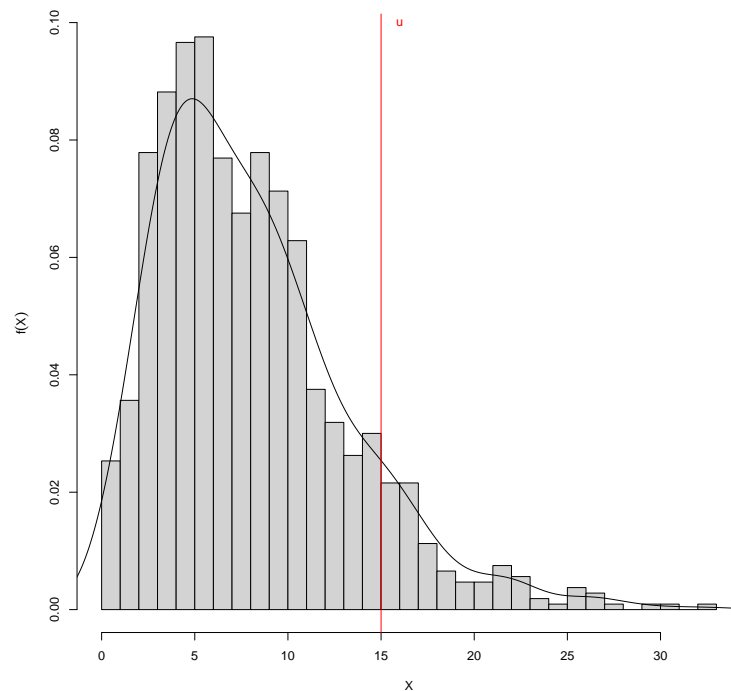


**Figure 7.** MAE and 2.5%, 97.5% quantiles for the three methods, for  $\alpha F(0, 1/2) + (1 - \alpha)F(0, 2)$  (left charts),  $\alpha F(0, 1) + (1 - \alpha)F(1, 1)$  (middle charts) and  $\alpha F(0, 1) + (1 - \alpha)F(1, 2)$  (right charts), for Lévy (upper charts), Cauchy (medium charts), Normal (lower charts).

### 8. An Application: PM 2.5 in Badajoz (Spain) during the Period 2011–2020

As we mentioned before, both baseline methods can be generalized for other baseline distributions by studying the relations between the parameters of the baseline distributions and the parameters of the limit GPD. We show an example, employing real data whose baseline distribution can be fitted by a Gamma distribution.

Data from measurements of the levels for diverse air pollutants in many municipalities in Spain are publicly available on the website <https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/>, accessed on 15 January 2022. Particulate matter is a mixture of solid particles and liquid droplets that can be inhaled and cause serious health problems. In particular, we have selected particulate matter less than 2.5 micrometers in diameter, known as PM 2.5, because it is considered to be especially dangerous to human health. In this context, studying the tail distribution above a certain threshold is essential because the World Health Organization recommends not to exceed an average daily value of  $25 \mu\text{g}/\text{m}^3$  and not to exceed an average annual value of  $10 \mu\text{g}/\text{m}^3$ . We studied levels of PM 2.5 measured in  $\mu\text{g}/\text{m}^3$  for the last ten years available, 2011–2020, from Badajoz. There are  $n = 1066$  observations, and, as can be seen in Figure 8, data can be fitted by a Gamma distribution. As in the previous simulations,  $p = 0.90$  and the threshold  $u = q_p$ , resulting  $u = 15 \mu\text{g}/\text{m}^3$ .

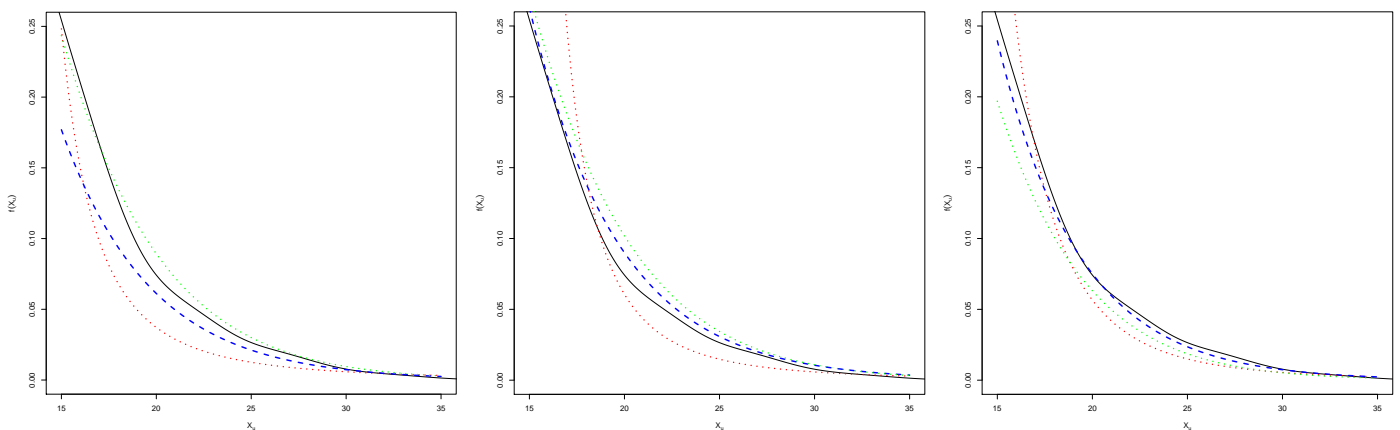


**Figure 8.** Histogram of PM 2.5 (in  $\mu\text{g}/\text{m}^3$ ) in Badajoz for 2011–2020, threshold  $u = q_p = 15$ ,  $p = 0.90$  and density (black curve).

Making a simulation study, similar to the previous ones made for Normal, Lévy and Cauchy, we obtained the following relation between the parameters of the baseline distribution  $\Gamma(\alpha, \beta)$  and the parameters of limit GPD( $\xi, \sigma$ ):

$$\hat{\xi} = 0, \hat{\sigma} = \frac{1}{\beta}(1 + 0.22 \log_2 \alpha) \tag{14}$$

Then, we randomly selected 50 data and applied the three methods MHM, BDM and IPBDM to fit the tail data. This proceeding was repeated many times, and we found three possible behaviors, as shown in Figure 9. In the first case (left chart), there is an example of the most usual situation: IPBDM offers intermediate estimations, between MHM and BDM. When there are scarce tail data (middle chart), MH differs significantly from the real density, while BDM and IPBDM provide better estimations. Finally, in the right chart, a common situation is shown in which IPBDM clearly offers the best estimations.



**Figure 9.** Tail distribution of data, density (black curve), estimations for MHM (red colored), BDM (green colored) and IPBDM (blue colored).

## 9. Conclusions

1. In the parameter estimation of GPD, usual EVT methods waste a lot of information. We have proposed two MH methods that make the most of all the information available from the data set. They are based on making use of the existing relations between baseline and GPD parameters, through informative priors.
2. When considering the GPD coming from stable baseline distributions, we employed singular properties of stable distributions to simplify the problem (reducing to standard cases) and to provide estimators for the parameters of the GPD.
3. We have achieved very accurate estimations for the parameters of the GPD when the baseline distribution is Cauchy, Normal or Lévy, making use of the properties of stable distributions and MH methods.
4. We have studied the goodness of the estimations for classical MH method and BMH when the baseline distribution is standard Cauchy, Normal or Lévy. Clearly, BMH provides more accurate estimations than MH.
5. In most real situations, data do not fit a simple distribution. We simulated some examples of mixtures of stable distributions and showed that IPBMH provides more accurate estimations than the other methods.
6. These proposals could be generalized for other baseline distributions by studying the relations between the parameters of the baseline distributions and the parameters of the limit GPD. We provide an application with real data to illustrate this situation.

**Author Contributions:** Investigation, J.M., M.I.P., M.M.P. and E.L.S.; Methodology, J.M., M.I.P., M.M.P. and E.L.S.; Writing—review & editing, J.M., M.I.P., M.M.P. and E.L.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Junta de Extremadura, Consejería de Economía, Ciencia y Agenda Digital GR21057, partially supported by Fondo Europeo de Desarrollo Regional (FEDER).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Nogaj, M.; Yiou, P.; Parey, S.; Malek, F.; Naveau, P. Amplitude and frequency of temperature extremes over the North Atlantic region. *Geophys. Res. Lett.* **2006**, *33*. [[CrossRef](#)]
2. Coelho, C.A.S.; Ferro, C.A.T.; Stephenson, D.B.; Steinskog, D.J. Methods for Exploring Spatial and Temporal Variability of Extreme Events in Climate Data. *J. Clim.* **2008**, *21*, 2072–2092. [[CrossRef](#)]
3. Acero, F.J.; Fernández-Fernández, M.I.; Carrasco, V.M.S.; Parey, S.; Hoang, T.T.H.; Dacunha-Castelle, D.; García, J.A. Changes in heat wave characteristics over Extremadura (SW Spain). *Theor. Appl. Climatol.* **2017**, *133*, 605–617. [[CrossRef](#)]
4. García, J.A.; Pizarro, M.M.; Acero, F.J.; Parra, M.I. A Bayesian Hierarchical Spatial Copula Model: An Application to Extreme Temperatures in Extremadura (Spain). *Atmosphere* **2021**, *12*, 897. [[CrossRef](#)]
5. García, J.; Gallego, M.C.; Serrano, A.; Vaquero, J. Trends in Block-Seasonal Extreme Rainfall over the Iberian Peninsula in the Second Half of the Twentieth Century. *J. Clim.* **2007**, *20*, 113–130. [[CrossRef](#)]
6. Re, M.; Barros, V.R. Extreme rainfalls in SE South America. *Clim. Chang.* **2009**, *96*, 119–136. [[CrossRef](#)]
7. Acero, F.J.; Gallego, M.C.; García, J.A. Multi-day rainfall trends over the Iberian Peninsula. *Theor. Appl. Climatol.* **2011**, *108*, 411–423. [[CrossRef](#)]
8. Acero, F.J.; García, J.A.; Gallego, M.C. Peaks-over-Threshold Study of Trends in Extreme Rainfall over the Iberian Peninsula. *J. Clim.* **2011**, *24*, 1089–1105. [[CrossRef](#)]
9. Acero, F.J.; Parey, S.; Hoang, T.T.H.; Dacunha-Castelle, D.; García, J.A.; Gallego, M.C. Non-stationary future return levels for extreme rainfall over Extremadura (southwestern Iberian Peninsula). *Hydrol. Sci. J.* **2017**, *62*, 1394–1411. [[CrossRef](#)]
10. Wi, S.; Valdés, J.B.; Steinschneider, S.; Kim, T.W. Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stoch. Environ. Res. Risk Assess.* **2015**, *30*, 583–606. [[CrossRef](#)]
11. Ramos, A.A. Extreme value theory and the solar cycle. *Astron. Astrophys.* **2007**, *472*, 293–298. [[CrossRef](#)]

12. Acero, F.J.; Carrasco, V.M.S.; Gallego, M.C.; García, J.A.; Vaquero, J.M. Extreme Value Theory and the New Sunspot Number Series. *Astrophys. J.* **2017**, *839*, 98. [[CrossRef](#)]
13. Acero, F.J.; Gallego, M.C.; García, J.A.; Usoskin, I.G.; Vaquero, J.M. Extreme Value Theory Applied to the Millennial Sunspot Number Series. *Astrophys. J.* **2018**, *853*, 80. [[CrossRef](#)]
14. Longin, F.M. Value at Risk and Extreme Values. *IFAC Proc. Vol.* **1998**, *31*, 45–49. [[CrossRef](#)]
15. Singh, A.K.; Allen, D.E.; Powell, R.J. Value at risk estimation using extreme value theory. In Proceedings of the 19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011.
16. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*; Springer: Berlin, Germany; New York, NY, USA, 1997.
17. Trzpiot, G.; Majewska, J. Estimation of Value at Risk: Extreme value and robust approaches. *Oper. Res. Decis.* **2010**, *20*, 131–143.
18. Magnou, G. An Application of Extreme Value Theory for Measuring Financial Risk in the Uruguayan Pension Fund. *Compend. Cuad. Econ. Adm.* **2017**, *4*, 1–19.
19. Gilli, M.; Këllezzi, E. An application of extreme value theory for measuring financial risk. *Comput. Econ.* **2006**, *27*, 207–228. [[CrossRef](#)]
20. Bali, T.G. A generalized extreme value approach to financial risk measurement. *J. Money Credit Bank.* **2007**, *39*, 1613–1649. [[CrossRef](#)]
21. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.M. *Extreme Value and Related Models with Applications in Engineering and Science*; Wiley-Interscience: Hoboken, NJ, USA, 2004.
22. Gumbel, E.J. *Statistics of Extremes (Dover Books on Mathematics)*; Dover Publications: Mineola, NY, USA, 2012.
23. Balkema, A.A.; De Haan, L. Residual life time at great age. *Ann. Probab.* **1974**, *2*, 792–804. [[CrossRef](#)]
24. Pickands, J. Statistical inference using extreme order statistics. *Ann. Stat.* **1975**, *3*, 119–131.
25. De Zea Bermudez, P.; Kotz, S. Parameter estimation of the generalized Pareto distribution—Part I. *J. Stat. Plan. Inference* **2010**, *140*, 1353–1373. [[CrossRef](#)]
26. De Zea Bermudez, P.; Kotz, S. Parameter estimation of the generalized Pareto distribution—Part II. *J. Stat. Plan. Inference* **2010**, *140*, 1374–1388. [[CrossRef](#)]
27. Korkmaz, M.Ç.; Altun, E.; Yousof, H.M.; Afify, A.Z.; Nadarajah, S. The Burr X Pareto Distribution: Properties, Applications and VaR Estimation. *J. Risk Financ. Manag.* **2018**, *11*, 1. [[CrossRef](#)]
28. Ihtisham, S.; Khalil, A.; Manzoor, S.; Khan, S.A.; Ali, A. Alpha-Power Pareto distribution: Its properties and applications. *PLoS ONE* **2019**, *14*, e0218027. [[CrossRef](#)]
29. Arnold, B.C.; Press, S.J. Bayesian estimation and prediction for Pareto data. *J. Am. Stat. Assoc.* **1989**, *84*, 1079–1084. [[CrossRef](#)]
30. Diebolt, J.; El-Aroui, M.A.; Garrido, M.; Girard, S. Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling. *Extremes* **2005**, *8*, 57–78. [[CrossRef](#)]
31. De Zea Bermudez, P.; Turkman, M.A. Bayesian approach to parameter estimation of the generalized Pareto distribution. *Test* **2003**, *12*, 259–277. [[CrossRef](#)]
32. Behrens, C.N.; Lopes, H.F.; Gamerman, D. Bayesian analysis of extreme events with threshold estimation. *Stat. Model.* **2004**, *4*, 227–244. [[CrossRef](#)]
33. Castellanos, M.E.; Cabras, S. A default Bayesian procedure for the generalized Pareto distribution. *J. Stat. Plan. Inference* **2007**, *137*, 473–483. [[CrossRef](#)]
34. Zhang, J.; Stephens, M.A. A new and efficient estimation method for the generalized Pareto distribution. *Technometrics* **2009**, *51*, 316–325. [[CrossRef](#)]
35. Martín, J.; Parra, M.I.; Pizarro, M.M.; Sanjuán, E.L. Baseline Methods for Bayesian Inference in Gumbel distribution. *Entropy* **2020**, *22*, 1267. [[CrossRef](#)]
36. Lévy, P. *Calcul des Probabilités*; Gauthier Villars: Paris, France, 1925.
37. Nolan, J.P. *Stable Distributions—Models for Heavy Tailed Data*; Birkhauser: Boston, MA, USA, 2018; Chapter 1; Available online: <http://fs2.american.edu/jpnolan/www/stable/stable.html> (accessed on 15 January 2022).
38. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **2006**, *6*, 7–11.