Environmental and Ecological Statistics (2022) 29:827–847
https://doi.org/10.1007/s10651-022-00548-1

Some efficient closed-form estimators of the parameters of the generalized Pareto distribution

Steven G. From[1] · **Suthakaran Ratnasingam**[2]

Received: 8 September 2021 / Revised: 4 August 2022 / Accepted: 16 August 2022 /
Published online: 12 September 2022

Abstract
In this paper, we consider several families of closed-form estimators of the two parameters of the Generalized Pareto Distribution (GPD). These estimators are easy to compute and have high efficiency when compared to previously proposed methods. We also consider some estimators which are not of closed-form. All methods are based on certain order statistics. The proposed procedures are best for extreme values of the shape parameters and sample sizes of 100 or larger. Monte Carlo simulations are conducted to investigate the performance of the proposed parameter estimation procedures. Our findings suggest that the proposed estimation methods are competitive compared to the existing methods. We provide a real data application to illustrate the utilization of the proposed methods in estimating the GPD parameters.

Keywords Elemental percentile method · Maximum likelihood · Method of moments · Modified Cramér–Von Misses method · Order statistics · Probability weighted moments · Product of spacings

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be random sample from the two-parameter Generalized Pareto Distribution (GPD) with the location parameter μ is assumed to be equal to zero. The cumulative distribution function (cdf) is given by

Handling Editor: Luiz Duczmal

✉ Suthakaran Ratnasingam
suthakaran.ratnasingam@csusb.edu

[1] Department of Mathematics, University of Nebraska at Omaha, Omaha, NE 68182, USA

[2] Department of Mathematics, California State University, San Bernardino, San Bernardino, CA 92407, USA

$$F(x;k,\sigma) = \begin{cases} 1 - \left(1 - \dfrac{kx}{\sigma}\right)^{1/k} & \text{if } k \neq 0,\ \sigma > 0, \\ 1 - \exp\left(-\dfrac{x}{\sigma}\right) & \text{if } k = 0,\ \sigma > 0, \end{cases} \tag{1}$$

where k and σ are the shape and scale parameters respectively. The range of x is $0 \leq x < \infty$ for $k \leq 0$ and $0 < x < \delta$, for $k > 0$, where $\delta = \sigma/k$. The cdf is zero for any value of $x < 0$. The corresponding probability density function (pdf) is

$$f(x;k,\sigma) = \begin{cases} \dfrac{1}{\sigma}\left(1 - \dfrac{kx}{\sigma}\right)^{1/k-1} & \text{if } k \neq 0,\ \sigma > 0, \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{x}{\sigma}\right) & \text{if } k = 0,\ \sigma > 0. \end{cases} \tag{2}$$

The shapes of the GPD for various values of k are illustrated in Fig. 1. The quantile function $F^{-1}(p;k,\sigma) = x$, the inverse function of $p = F(x;k,\sigma)$ is:

$$x = F^{-1}(p;k,\sigma) = \begin{cases} \dfrac{\sigma}{k}\left(1 - (1-p)^k\right) & \text{if } k \neq 0,\ \sigma > 0, \\ -\sigma \ln(1-p) & \text{if } k = 0,\ \sigma > 0. \end{cases} \tag{3}$$

The problem considered here is the estimation of the GPD parameters k and σ. The GPD has been used in applications that involve the modeling of exceedances over a threshold, the distribution of extreme value statistics (such as the largest or smallest values in a random sample), and is closely related to families of generalized extreme value distributions, for example, Galambos (1981, 1984). It has many practical

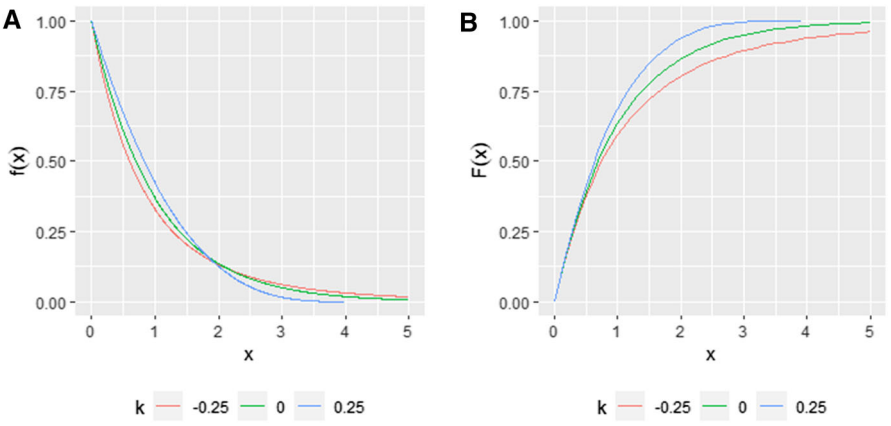

Fig. 1 The pdf (**A**) and cdf (**B**) of the GPD for various values of the shape parameter k

applications in engineering design and hydrology, for example, Smith (1984), Hosking et al. (1985), Van Montfort and Witter (1986), Castillo (1988), Castillo (1994), and Walshaw (1990).

Most previous papers focused attention on the case $-1/2 < k < 1/2$ since many practical applications involve this range of values for k. In addition, the large sample theory is applicable for this range. However, more extreme values of k are also of interest, as pointed out by Castillo and Hadi (1997). Cheng and Chou (2000a, b) considered the estimation of σ when k is known. They obtain the BLUE (best linear unbiased estimator) and ABLUE (asymptotically best linear unbiased estimator) of σ when k is known.

In this paper, both k and σ are unknown parameters to be estimated. The GPD was first discussed by Pickands (1975). They proposed estimators which are special cases of the estimators considered by Castillo and Hadi (1997). Related papers are Castillo and Hadi (1995a, b). The Maximum likelihood (MLE) estimation of the GPD has been discussed by DuMouchel (1983), Davison (1984), Smith (1984, 1985), Hosking and Wallis (1987), and Grimshaw (1993). Hosking and Wallis (1987) compared the MLE to method of moments (MOM) estimation and probability weighted moments estimation (PWM). These two estimation methods produce closed-form estimators of k and σ. The MLE, MOM, and PWM estimation methods all have serious drawbacks. The likelihood function for the MLE can be made infinite for $k > 1$ and Cramér's regularity conditions are not satisfied for all k values. In addition, the MLE procedure may not converge. Even if the MLE exists, Hosking and Wallis (1987) demonstrated that the MLE does not manifest its best asymptotically normal property unless $n > 500$ and $-1/2 < k < 1/2$. Since $\mathrm{Var}(X_i) = \infty$ for $k \leq -1/2$, the MOM and PWM do not always exist.

In response to these drawbacks of the MLE, MOM, and PWM methods, Castillo and Hadi (1997) proposed the elemental percentile method (EPM). Let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ denote the order statistics corresponding to the random sample $X_1, X_2, \ldots, X_n$. Let $P_{i,n} = i/(n+1), i = 1, 2, \ldots, n$. By choosing any two order statistics $X_{i,n}$ and $X_{j,n}$ for $i < j$, the EPM method solves for k and σ in the following two equations.

$$F(X_{i,n}; k, \sigma) = P_{i,n}, \tag{4}$$

and

$$F(X_{j,n}; k, \sigma) = P_{j,n}. \tag{5}$$

In general, system (4)–(5) has no closed-form solution. Theorem 1 of Castillo and Hadi (1997) details how to solve for k and σ in (4)–(5). Let $\delta = \sigma/k$. Use the method of bisection to solve for δ in the equation

$$C_i \ln(1 - X_{j,n}/\delta) = C_j \ln(1 - X_{i,n}/\delta), \tag{6}$$

where $C_i = \ln(1 - P_{i,n})$. If $X_{i,n} < \dfrac{C_i X_{j,n}}{C_j}$, we search for a root on the interval $(\delta_0, 0)$, where

$$\delta_0 = \frac{X_{i,n} X_{j,n} (C_j - C_i)}{\left(C_j X_{i,n} - C_i X_{j,n}\right)}. \tag{7}$$

If $X_{i,n} > \dfrac{C_i X_{j,n}}{C_j}$, then we search for root on the interval $(X_{j,n}, \delta_0)$. Let $\hat{\delta}_{i,j}$ denote this root or solution. Then the estimators of k and σ are

$$\hat{k}_{i,j} = \frac{\ln\left(1 - X_{i,n}/\hat{\delta}_{i,j}\right)}{C_i}.$$

and

$$\hat{\sigma}_{i,j} = \hat{k}_{i,j}\,\hat{\delta}_{i,j}.$$

Algorithm 1 of Castillo and Hadi (1997) gives the details for defining $\hat{k}_{i,j}$ and $\hat{\sigma}_{i,j}$ in the case where δ_0 is not defined $\left(C_j X_{i,n} = C_i X_{j,n}\right)$ in (6). Several methods for selecting the values of i and j were considered by these authors. They reported their results for the case $i = 1, 2, \ldots, n-1$ and $j = n$. Thus, they obtained $\hat{k}_{1,n}, \hat{k}_{2,n}, \ldots, \hat{k}_{n-1,n}$ and $\hat{\sigma}_{1,n}, \hat{\sigma}_{2,n}, \ldots, \hat{\sigma}_{n-1,n}$. The EPM estimates of k and σ are

$$\hat{k}_{EPM} = \text{median}\left\{\hat{k}_{1,n}, \hat{k}_{2,n}, \ldots, \hat{k}_{n-1,n}\right\},$$

and

$$\hat{\sigma}_{EPM} = \text{median}\left\{\hat{\sigma}_{1,n}, \hat{\sigma}_{2,n}, \ldots, \hat{\sigma}_{n-1,n}\right\}.$$

The choice $j = n$ guarantees estimates of k and σ that are consistent with the observed data $X_1, \ldots, X_n$, that is, for $k > 0$,

$$\frac{\hat{k}_{i,j}\, X_{j,n}}{\hat{\sigma}_{i,j}} \leq 1,$$

holds for $j = n$, $i = 1, 2, \ldots, n-1$. Thus, the EMP estimators do not suffer a drawback of both MOM and PWM estimates. Namely, Castillo and Hadi (1997) showed that, even if $n = 100$, the MOM and PWM methods can produce estimates of k and σ which yield a range of values of X_i that are not consistent with the data, in almost half of all simulated data sets for some choices of k. The choice $i = n/2$ and $j = 3n/4$ leads to an estimator discussed by Pickands (1975). These are of closed-form and are given by

$$\hat{k}_p = \frac{1}{\ln(2)} \ln\left(\frac{X_{n/2,n}}{X_{3n/4,n} - X_{n/2,n}}\right). \tag{8}$$

$$\hat{\sigma}_p = \hat{k}_p \left(\frac{\left(X_{n/2,n}\right)^2}{2X_{n/2,n} - X_{3n/4,n}}\right). \tag{9}$$

Further, Castillo and Hadi (1997) compared the EPM estimators to the MOM and PWM estimators on the basis of simulated root mean square error (RMSE). They found that the EPM estimators are the best for extreme values of k, that is, for $k < -0.4$ and $k > 0.4$. For $-0.4 \leq k \leq 0.4$, the MOM and PWM perform better. They found that the MOM estimates have the smallest RMSE for $0 < k \leq 0.4$ and that PWM is best

for $-0.4 \le k \le 0$. See Hosking and Wallis (1987) or Castillo and Hadi (1997) for formulas to compute the MOM and PWM estimators. The MOM and PWM estimators are especially poor for extremely large and negative values of k ($k = -1$ and $k = -2$ in Table 2 of Castillo and Hadi (1997)). There is no finite mean and variance for these k values. As a result, the MOM and PWM estimators are particularly weak for very large and negative values of k.

2 Proposed Estimators

In this section, we describe some new estimators of the GPD parameters. We consider both estimators of closed-form and estimators which are not of closed-form, but chosen to minimize/maximize a certain function of the order statistics. The first four methods (M1, M2, M3, and QM below) produce closed-form estimators. The last two methods (POS and LCVM) produce estimators minimizing/maximizing a certain function of the order statistics.

2.1 Method 1 (M1)

This method uses the EPM method but selects the two order statistics differently from any of the selection schemes given under Algorithm 2 of Castillo and Hadi (1997). We select the two order statistics to guarantee a closed-form solution to (4) and (5). We do this several times before taking the median. Also, we do not use nearly as many $\hat{k}_{i,j}$ and $\hat{\sigma}_{i,j}$ values. In fact, regardless of n, the sample size, we never use more than 5 or 10 order statistics.

Let $0 < Q < 1$, and let $P = 1 - (1 - Q)^{1/2}$. Then $0 < P < Q < 1$. Let $n(1) = \text{nint}\big((n+1)P\big)$ and $n(2) = \text{nint}\big((n+1)Q\big)$, where 'nint' stands for nearest integer. Apply the EPM method and solve for k, σ in the equations $F(X_{n(1),n}; k, \sigma) = P$ and $F(X_{n(2),n}; k, \sigma) = Q$. We obtain $k = k_{P,Q}$ and $\sigma = \sigma_{P,Q}$, where

$$k_{P,Q} = \frac{\ln\left(X_{n(2),n}/X_{n(1),n} - 1\right)}{\ln(1-P)},$$
$$\sigma_{P,Q} = \delta_{P,Q} \cdot k_{P,Q}, \quad \text{and}$$
$$\delta_{P,Q} = \frac{\left(X_{n(1),n}\right)^2}{\left(2X_{n(1),n} - X_{n(2),n}\right)}.$$

If $P = 1/2$, $Q = 3/4$, we obtain (Pickands 1975) estimators given by (8)–(9), as a special case. To define the Method 1 (M1) estimators of k and σ, we compute $k_{P,Q}$ and $\sigma_{P,Q}$ for m different ordered pairs $(P_1, Q_1), (P_2, Q_2), \ldots, (P_m, Q_m)$ satisfying $0 < Q_i < 1$ and $P_i = 1 - (1 - Q_i)^{1/2}$, $i = 1, 2, \ldots, m$. We shall discuss later how

to choose the Q_i values. Let $\hat{k}_1^{(i)} = k_{P_i,Q_i}$ and $\hat{\sigma}_1^{(i)} = \sigma_{P_i,Q_i}$. Let

$$\hat{k}_1^{(0)} = \text{median}\{\hat{k}_1^{(1)}, \ldots, \hat{k}_1^{(m)}\}, \text{ and}$$
$$\hat{\sigma}_1^{(0)} = \text{median}\{\hat{\sigma}_1^{(1)}, \ldots, \hat{\sigma}_1^{(m)}\}.$$

If these are consistent with the data, we are done. Otherwise, we must 'adjust' these estimators to be consistent with the range of the data, if $\hat{k}_1^{(0)} > 0$. We choose $m = 5$ with $Q_1 = 0.50$, $Q_2 = 0.60$, $Q_3 = 0.75$, $Q_4 = 0.85$ and $Q_5 = n/(n+1)$. There is nothing special about these values. Numerous numerical studies have found that any selection scheme for choosing Q_i values which picks more or less equally-spaced values in the interval [0.50, 1] works well. Note that (P_3, Q_3) corresponds to Pickands (1975) estimator (since $P_3 = 1/2$), and that (P_5, Q_5) is a value (or close to a value) used by the EPM method, and guarantees that at least one pair (P_i, Q_i) will produce estimators consistent with the data. Let

$$W_1 = \frac{\hat{k}_1^{(0)} X_{n(m),n}}{\hat{\sigma}_1^{(0)}} = \frac{\hat{k}_1^{(0)} X_{n,n}}{\hat{\sigma}_1^{(0)}}.$$

The M1 estimator of k is

$$\hat{k}_1 = \begin{cases} \hat{k}_1^{(0)} & \text{if } W_1 < 1, \\ \hat{k}_1^{(m)} & \text{if } W_1 \geq 1. \end{cases}$$

The M1 estimator of σ is

$$\hat{\sigma}_1 = \begin{cases} \hat{\sigma}_1^{(0)} & \text{if } W_1 < 1, \\ \hat{\sigma}_1^{(m)} & \text{if } W_1 \geq 1. \end{cases}$$

Thus, $\hat{k}_1^{(m)} = \hat{k}_1^{(5)}$ based upon the largest order statistics 'adjust' the estimator, if necessary ($W_1 \geq 1$), to be consistent with the data. Thus, M1 produces m different closed-form estimates of k and σ for each selection of $Q_1, \ldots, Q_m$.

2.2 Method 2 (M2)

This method is the same as M1 except we use only the m largest order statistics. For large positive values of k, this will produce more efficient estimators than M1, as will be seen later. It is believed that the EPM method of Castillo and Hadi (1997) utilizes the largest order statistics too much for $k < 0$, but this is an advantage for $k > 0$, hence the use of the largest order statistics only by M2. Again, we choose $m = 5$. Let $Q_i = (n - m + i)/(n+1)$, $i = 1, 2, \ldots, m$, $P_i = 1 - (1 - Q_i)^{1/2}$. Note that $Q_m = n/(n+1)$, so that again, at least one pair (P_i, Q_i) will produce estimates

consistent with the data. Let

$$\hat{k}_2^{(0)} = \text{median}\{\hat{k}_2^{(1)}, \ldots, \hat{k}_2^{(m)}\}, \quad \text{and}$$
$$\hat{\sigma}_2^{(0)} = \text{median}\{\hat{\sigma}_2^{(1)}, \ldots, \hat{\sigma}_2^{(m)}\}.$$

where $\hat{k}_2^{(i)} = k_{P_i,Q_i}$, $\hat{\sigma}_2^{(i)} = \sigma_{P_i,Q_i}$. Let

$$W_2 = \frac{\hat{k}_2^{(0)} X_{n,n}}{\hat{\sigma}_2^{(0)}}.$$

The M2 estimator of k is

$$\hat{k}_2 = \begin{cases} \hat{k}_2^{(0)} & \text{if } W_2 < 1, \\ \hat{k}_2^{(m)} & \text{if } W_2 \geq 1, \end{cases}$$

The M2 estimator of σ is

$$\hat{\sigma}_2 = \begin{cases} \hat{\sigma}_2^{(0)} & \text{if } W_2 < 1, \\ \hat{\sigma}_2^{(m)} & \text{if } W_2 \geq 1. \end{cases}$$

2.3 Method 3 (M3)

This method is a hybrid method based on M1 and M2 estimates. It can be seen that, overall, this method is the most efficient of the closed-form estimation methods considered, over a wide range of k values. The M3 estimator of k is the average of the M1 and M2 estimators of k

$$\hat{k}_3 = \frac{1}{2}(\hat{k}_1 + \hat{k}_2). \tag{10}$$

Let

$$\hat{\sigma}_3^{(0)} = \begin{cases} \hat{\sigma}_1 & \text{if } \hat{k}_1 \leq \frac{1}{4}, \\ \frac{1}{2}(\hat{\sigma}_1 + \hat{\sigma}_2) & \text{if } \hat{k}_1 > \frac{1}{4}. \end{cases} \tag{11}$$

Let

$$W_3 = \frac{\hat{k}_3 X_{n,n}}{\hat{\sigma}_3^{(0)}}.$$

The M3 estimator of σ is

$$\hat{\sigma}_3 = \begin{cases} \hat{\sigma}_3^{(0)} & \text{if } W_3 < 1, \\ \frac{1}{2}(\hat{\sigma}_1 + \hat{\sigma}_2) & \text{if } W_3 \geq 1. \end{cases} \tag{12}$$

Thus, the adjustment of the estimator of σ is made only if $\hat{k}_1 \leq 1/4$ and $W_3 < 1$. It should be noted that the substitution of $\hat{k}_3$ for $\hat{k}_1$ in (11) produces biases and MSEs that

are very similar to (11) in most cases, with the exception that this substitution produces slightly higher biases and MSEs than (11) for large negative values of k. Many other averaging schemes were tried, but none consistently produced lower biases/MSEs for all values of k and σ, and for all n. There are undoubtedly better rules, but the one presented here is very efficient as well as easy to program and compute.

2.4 Method QM

The fourth closed-form estimator method is also based on the EPM, but in a uniquely different way. The relationship between P_i and Q_i is different. Again, we choose $0 < Q_1 < Q_2 < \cdots < Q_m < 1$ with $Q_m = n/(n+1)$. Instead, we now require

$$P_i = 1 - (1 - Q_i)^{1/3}, \quad i = 1, 2, \ldots, m. \tag{13}$$

The cube root of $(1 - Q_i)$ is used instead of the square root. This will lead to estimates that are the roots of quadratic equations. For this reason, we call the following method as the quadratic method (QM). We use the same basic procedure as discussed in Method M1, except we require P_i and Q_i to satisfy (13). To obtain m estimates of k and σ, we substitute P_i for P and Q_i for Q in the following equations. Let $0 < Q < 1$, $P = 1 - (1 - Q)^{1/3}$. Let $n(1) = \text{nint}\{(n+1)P\}$ and $n(2) = \text{nint}\{(n+1)Q\}$. We solve the EPM equations for k and σ, as done in (4) and (5): $F(X_{n(1),n}; k, \sigma) = P$ and $F(X_{n(2),n}; k, \sigma) = Q$. Let $Y_1 = X_{n(1),n}$, $Y_2 = X_{n(2),n}$, $\delta = \sigma/k$. Then

$$\ln\left(1 - \frac{Y_1}{\delta}\right) = k \ln(1 - P), \tag{14}$$

and

$$\ln\left(1 - \frac{Y_2}{\delta}\right) = 3k \ln(1 - P). \tag{15}$$

Exponentiating and simplifying, we obtain

$$1 - \frac{Y_2}{\delta} = \left(1 - \frac{Y_1}{\delta}\right)^3$$

or

$$(-3Y_1 + Y_2)\,\delta^2 + (3Y_1^2)\,\delta - Y_1^3 = 0, \tag{16}$$

with roots

$$\delta = \delta_1 = \frac{-3Y_1^2 + \sqrt{4Y_1^3 Y_2 - 3Y_1^4}}{2(Y_2 - 3Y_1)}, \tag{17}$$

and

$$\delta = \delta_2 = \frac{-3Y_1^2 - \sqrt{4Y_1^3 Y_2 - 3Y_1^4}}{2(Y_2 - 3Y_1)}. \tag{18}$$

To determine whether we use (17) or (18) we must specify two cases.

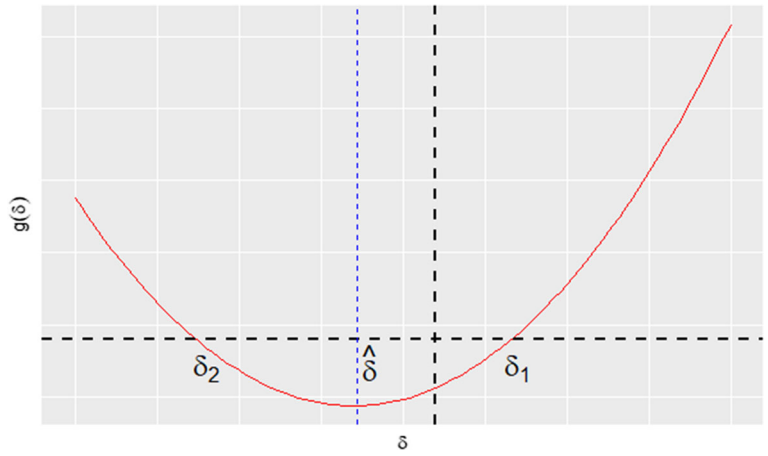

Fig. 2 Case 1

Case 1

$Y_2 - 3Y_1 > 0$. Then $\delta_1 > 0$ and $\delta_2 < 0$. Let $c_1 = \ln(1-P)$ and $c_2 = \ln(1-Q)$. Then $3c_1/c_2 = 1$, since $P = 1-(1-Q)^{1/3}$. Also, $Y_2 > 3Y_1$ gives $Y_1 < c_1 Y_2/c_2$. Thus, the EPM method chooses the root δ in (16) to satisfy $\delta \in (\delta_0, 0)$, which means we must choose $\delta = \delta_2$, the negative root. Let $g(\delta) = (Y_2 - 3Y_1)\delta^2 + (3Y_1^2)\delta - Y_1^3$. Let $\hat{\delta} = -3Y_1^2/\left[2(Y_2 - 3Y_1\right]$. For Case 1, a graph of $g(\delta)$ vs δ is helpful and it is sketched in Fig. 2.

Case 2

$Y_2 - 3Y_1 < 0$. First of all, let's show that $\delta_1 \le Y_2$. By the EPM method, we will then have that $\delta_2 > Y_2$, since δ_2 is the only other root. After some algebra, $\delta_1 \le Y_2$ holds iff

$$-3 + \sqrt{4w-3} \ge 2w(w-3), \quad 1 \le w \le 3, \tag{19}$$

where $w = Y_2/Y_1$. Let $h(w) = -3 + \sqrt{4w-3} - 2w(w-3)$ for $w \in [1,3]$. Note that $h(w)$ is a concave function since its second derivative is negative. That means $h(w)$ has a unique maximum, which is at 7/4. Since $h(1) = 2 < h(7/4)$, and $h(3) = 0$, we get $h(w) \ge h(1) = 2$ for $w \in [1, 7/4]$, and $h(w) \ge h(3) = 0$ for $w \in [7/4, 3]$. Hence, $h(w)$ is non-negative for $w \in [1,3]$. So we get $-3 + \sqrt{4w-3} \ge 2w(w-3)$, $1 \le w \le 3$, as desired. Thus, $\delta_1 \le Y_2$ holds. Thus, $\delta_2 > Y_2$ must hold and we choose $\delta = \delta_2$ again. A graph of $g(\delta)$ for Case 2 is shown in Fig. 3.

Cases 1 and 2 above establish that the estimator of δ is

$$\delta^*_{P,Q} = \frac{-3Y_1^2 - \sqrt{4Y_1^3 Y_2 - 3Y_1^4}}{2(Y_2 - 3Y_1)}. \tag{20}$$

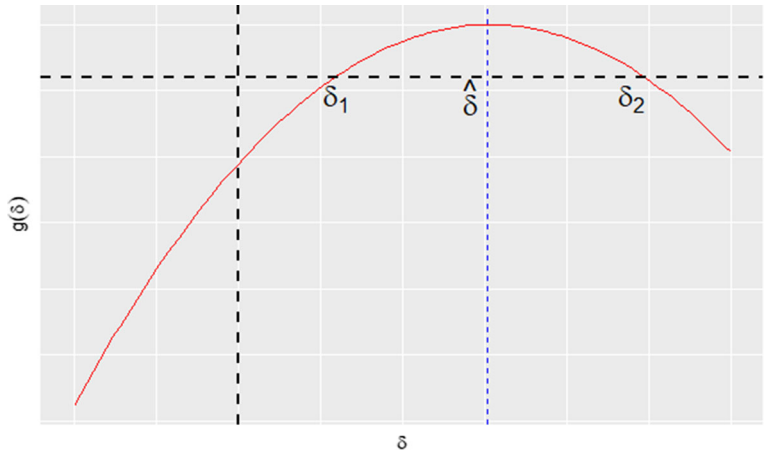

Fig. 3 Case 2

The discriminant $\left(4Y_1^3Y_2 - 3Y_1^4\right)$ will never be negative, since $4Y_1^3Y_2 > 4Y_1^4 > 3Y_1^4 > 0$. Thus, with probability one, $\delta^*_{P,Q}$ will exist, since $P\left(Y_2 - 3Y_1 = 0\right) = 0$, for each choice of $(P_i, Q_i),\ i = 1, \ldots, m$. After obtaining $\delta^*_{P,Q}$, we obtain $k^*_{P,Q}$ from either (14) or (15), from which we obtain $\sigma^*_{P,Q}$ as $\sigma^*_{P,Q} = k^*_{P,Q}\,\delta^*_{P,Q}$.

We use the same choices for Q_i as in Method M1. Let $m = 5$, $Q_1 = 0.50$, $Q_2 = 0.60$, $Q_3 = 0.75$, $Q_4 = 0.85$, $Q_5 = n/(n+1)$. Let $\hat{k}^*_i = k^*_{P_i,Q_i}$, $\hat{\sigma}^*_i = \sigma^*_{P_i,Q_i}$. Let

$$\hat{k}_4^{(0)} = \text{median}\left\{\hat{k}^*_1, \ldots, \hat{k}^*_m\right\}, \quad \text{and}$$
$$\hat{\sigma}_4^{(0)} = \text{median}\left\{\hat{\sigma}^*_1, \ldots, \hat{\sigma}^*_m\right\}.$$

Let $W_4 = \dfrac{\hat{k}_4^{(0)} X_{n,n}}{\hat{\sigma}_4^{(0)}}$. The QM estimates of k and σ are:

$$\hat{k}_4 = \begin{cases} \hat{k}_4^{(0)} & \text{if } W_4 < 1, \\ \hat{k}^*_m & \text{if } W_4 \geq 1, \end{cases}$$
$$\hat{\sigma}_4 = \begin{cases} \hat{\sigma}_4^{(0)} & \text{if } W_4 < 1, \\ \hat{\sigma}^*_m & \text{if } W_4 \geq 1. \end{cases}$$

The M2 version of QM was also considered but did not perform better than M2 itself, so we do not consider the use of the m largest order statistics in this case.
Next, we consider two estimation methods that do not lead to closed-form estimates of k and σ, but compare favorably to the other estimators discussed here.

2.5 Method POS

This method is based on a certain variation of a product of spacings (POS). Choose k and σ which maximize the product of spacings (uniform distribution on (0,1) spacings):

$$D_1 = \prod_{i=1}^{n+1} \Big(A_i - A_{i-1} \Big), \tag{21}$$

where $A_i = F(X_{i,n}; k, \sigma),\ i = 1, \ldots, n-1,\ A_0 = 0$ and $A_{n+1} = 1$. The quantities $A_i - A_{i-1}$ in (21) have the distribution of the spacings from a uniform distribution on (0, 1), if k and σ are the true values of the parameters. To avoid computational difficulties, it is more convenient to maximize $\ln(D_1)$, or, equivalently, minimize minus $\ln(D_1)$. It is worth noting that if values of k and σ which are inconsistent with the sample data are substituted into D_1, then $D_1 = 0$, with $\ln(D_1) = -\infty$, producing a minimum, rather than a maximum value. Thus, special care must be taken in any minimization or maximization routine to avoid values of k and σ which are infeasible (inconsistent with the data). In the numerical studies of the next section, we employ the simulated annealing algorithm (SANN; Belisle 1992) to obtain estimates. The optimization problem based on the SANN algorithm requires only function values. Further, it works effectively with functions that are not differentiable. This avoids having to compute the first and second derivatives of D_1 with respect to k and σ, which will not always exist for $k > 0$. The M1 estimators were used as initial estimates of k and σ. The numerical optimization can be easily done using the function *optim()* from the *base* package in *R*. All these points are applicable for the computation of the LCVM estimates below.

2.6 Method LCVM

This method minimizes a logarithmic version of the well-known Cramér–Von Mises distance measure, hence, we call it the Logarithmic Cramér–Von Mises method (LCVM). Choose k and σ minimizing

$$D_2 = \sum_{i=1}^{n} \Big(B_i - C_i \Big)^2, \tag{22}$$

where $B_i = \ln\big(F(X_{i,n}, k, \sigma)\big), C_i = \ln\big(i/(n+1)\big),\ i = 1, 2, \ldots, n$. Again, the M1 method provides the initial estimates of k and σ. It is worth noting that if the current values of k and σ are infeasible, we replace D_2 with a very large positive number. Alternatively, a penalty function approach could be used. It should be noted that the usual minimum distance estimators, such as Cramér–Von Mises, Kolmogorov–Simirnov and Anderson–Darling estimators were also considered, but did not perform well as the POS and LCVM estimators. Distance measures based upon Shannon's entropy (of various orders) and Gini diversity were also considered, but they did not have lower biases and MSEs than POS or LCVM and were often significantly worse.

3 Numerical Comparisons

In this section, we present the estimated root mean squared error (RMSE) of the estimators of k and σ for all seven methods: EPM of Castillo and Hadi (1997), M1, M2, M3, QM, POS, and LCVM. We use the inverse cdf method to generate all pseudo-random variates based on equation (3). In order to obtain seven sets of estimators of k and σ, we generate 2000 random samples for each combination of n, k and σ. Without loss of generality, we let $\sigma = 1$ throughout the simulation study. We use many of the same combinations of n, k, and σ as Castillo and Hadi (1997), but we also include a few additional cases. Tables 1 and 2 present the RMSEs for the estimation of k and σ respectively. Further, we have developed an *R* package, **EfficientClosedGPD**, which can be found at https://github.com/suthakaranr/EfficientClosedGPD in which all six methods are implemented to allow readers to estimate parameters of the Generalized Pareto Distribution.

From Tables 1 and 2, several observations can be made:

- For the estimation of k, the M3 estimator outperforms the EPM estimator of Castillo and Hadi (1997), especially for $k < 0$ and $n \geq 50$. However, for $k \geq 0$ and $n \geq 50$, the EPM estimator performs better than the M3 estimator. Thus, the M3 method is recommended for extreme values of k and $n \geq 50$, being highly efficient and of closed-form.
- For small n (for example, $n = 15$), the EPM method performs significantly better than M3 in the estimation of k.
- For the estimation of k and σ, POS performs slightly better or as well as EPM, and is significantly better for $n \geq 50$.
- For the estimation of σ, EPM performs slightly better or as well as M3, and is significantly better for small n.
- For extreme negative values of k, POS and LCVM perform the best. However, the performance of QM is by far the best for $k \leq -3$; see Tables 4 and 5. However, QM performs poorly for $k > -1$.
- For the estimation of k, the LCVM performs significantly better than EPM for $n \geq 50$. For small n (for example, $n = 15$), the EPM method performs better than LCVM.
- For the estimation of σ, EPM performs slightly better than LCVM, however, LCVM requires only one call to a minimization routine.
- When comparing the RMSE of M3 to the RMSEs of M1 and M2, we can see that for $k < 0$, in almost every case, the hybrid method M3 beats both M1 and M2. For the estimation of σ, for $n \geq 50$, M1 is better than M2 for $k \leq -1/4$, while the opposite is true for $k > -1/4$. This motivated the M3 method, and is the reason $\hat{\sigma}_3^{(0)}$ given by (11) earlier was defined the way that it was.
- Despite the fact that the biases are not presented, the EPM method generally has slightly smaller biases. In most cases, the bias comparisons reflect the RMSE comparisons. In cases where EPM and M3 have equal or nearly equal RMSEs, EPM usually has slightly smaller bias. Otherwise, no discernible pattern of biases was found.

Table 1 RMSE of k estimators

		Method						
n	k	EPM	M1	M2	M3	QM	POS	LCVM
15	−2.00	1.1665	1.2063	1.2038	1.0454	1.2372	0.9834	1.0949
	−1.00	0.8082	1.0353	1.0272	0.8856	1.2411	0.7108	0.8575
	−0.50	0.6889	1.1028	1.0886	0.9517	1.3836	0.6005	0.7564
	−0.25	0.6231	1.1357	1.1380	0.9108	1.4924	0.5337	0.7136
	0.00	0.6007	1.3402	1.1744	1.0076	1.5918	0.4773	0.6916
	0.25	0.5842	1.3607	1.1997	1.0896	1.6904	0.4431	0.6500
	0.50	0.5902	1.3853	1.2601	1.1632	1.7861	0.4004	0.6327
	1.00	0.6326	1.5957	1.3849	1.2837	2.0017	0.4017	0.6905
	2.00	0.8632	1.8682	1.6753	1.5114	2.3761	0.5577	0.8802
	3.00	1.1253	2.2132	1.9848	1.9268	2.8551	0.9799	1.0844
50	−2.00	0.6304	0.6223	0.7641	0.5442	0.6043	0.4686	0.5474
	−1.00	0.4542	0.4571	0.4393	0.3596	0.4759	0.3350	0.4278
	−0.50	0.3793	0.4355	0.3272	0.2961	0.5127	0.2735	0.3486
	−0.25	0.3501	0.4783	0.2962	0.3023	0.5998	0.2370	0.3308
	0.00	0.3114	0.5278	0.2781	0.3267	0.6748	0.2140	0.2985
	0.25	0.3016	0.6088	0.2725	0.3374	0.7407	0.1829	0.2891
	0.50	0.3115	0.6674	0.2666	0.3689	0.8030	0.1724	0.2710
	1.00	0.3284	0.7390	0.3142	0.4407	0.8842	0.1748	0.2939
	2.00	0.4458	0.8139	0.4750	0.5479	1.0387	0.2868	0.4108
	3.00	0.5958	0.9930	0.6478	0.7147	1.2283	0.5666	0.5633
100	−2.00	0.4294	0.4352	0.6147	0.4047	0.4349	0.3226	0.3926
	−1.00	0.3139	0.3215	0.3575	0.2527	0.3374	0.2254	0.2995
	−0.50	0.2617	0.2959	0.2459	0.2093	0.3480	0.1796	0.2488
	−0.25	0.2451	0.3316	0.2029	0.2005	0.3934	0.1614	0.2212
	0.00	0.2267	0.4182	0.1746	0.2230	0.4678	0.1350	0.2037
	0.25	0.2109	0.4771	0.1561	0.2437	0.5388	0.1113	0.1906
	0.50	0.2135	0.5306	0.1596	0.2744	0.5905	0.1035	0.1918
	1.00	0.2323	0.5777	0.1790	0.3106	0.6475	0.1188	0.2052
	2.00	0.3184	0.6702	0.2919	0.3816	0.7288	0.2114	0.2791
	3.00	0.4253	0.7650	0.4140	0.4914	0.8420	0.4043	0.4226
200	−2.00	0.3247	0.3180	0.5119	0.3388	0.3061	0.2180	0.2680
	−1.00	0.2350	0.2294	0.2910	0.2054	0.2402	0.1512	0.2038
	−0.50	0.1961	0.1986	0.1939	0.1491	0.2186	0.1189	0.1817
	−0.25	0.1758	0.2081	0.1497	0.1305	0.2512	0.1025	0.1670

Table 1 continued

		Method						
n	k	EPM	M1	M2	M3	QM	POS	LCVM
	0.00	0.1692	0.2675	0.1198	0.1411	0.3213	0.0864	0.1462
	0.25	0.1479	0.3061	0.1026	0.1611	0.3728	0.0760	0.1305
	0.50	0.1492	0.3510	0.0959	0.1779	0.4133	0.0665	0.1310
	1.00	0.1648	0.3970	0.1184	0.2169	0.4616	0.0837	0.1379
	2.00	0.2216	0.4160	0.2123	0.2619	0.5164	0.1450	0.1838
	3.00	0.2916	0.4945	0.3030	0.3315	0.5962	0.3086	0.2865

- Table 3 compares M1 and M3 estimators to the estimators of Pickands (1975), see (8)–(9): We see that M3 significantly improves upon the estimator of Pickands, especially for $k < 0$, and uses the Pickands estimators.
- None of the methods in this paper are outlier robust, since all methods use the largest order statistics. It should be mentioned that the choice $Q_1 = 0.5$, $Q_2 = 0.6$, $Q_3 = 0.75$, $Q_4 = 0.8$ and $Q_5 = 0.85$ performs well in methods M1, M3, and QM, but is about 10% less efficient than the Q_i values used for this paper. Thus, these closed-form estimators are easily adapted to the possible undue influence of outliers. Tables 4 and 5 below give the RMSE for $k = -3$ and $k = -5$ for the four closed-form methods (M1, M2, M3, ad QM) for completeness. As we mentioned earlier, these results are based on 2000 simulated data sets. No computational problems were experienced for these four methods, even for $k = -5, -10$, although MSEs rapidly increase, since the GPD is extremely heavy-tailed for every large and negative values of k.

According to Tables 4 and 5, we see that M1 and QM are the best for $k \leq -3$, especially for estimating σ. These values of k correspond to distributions having decreasing failure rate (DFR), that is, the failure rate functions

$$r(x; k, \sigma) = \frac{f(x; k, \sigma)}{1 - F(x; k, \sigma)},$$

where $f(\cdot)$ and $F(\cdot)$ are given by (2) and (1), is decreasing in x. Such distributions have 'fractal-like' behavior and are being used in a growing number of fields. For these types of distributions, M1 and QM are the best.

4 Application

In this section, the proposed methods have been applied to a real-world data set to evaluate the performance of the parameter estimation procedures. We fit the GPD to the Bilbao waves data used in Castillo and Hadi (1997). The data measures zero-crossing hourly mean periods (in seconds) of the sea waves in a Bilbao buoy in January 1997. One purpose of the data is to investigate the influence of periods on beach

Table 2 RMSE of σ estimators

		Method						
n	k	EPM	M1	M2	M3	QM	POS	LCVM
15	−2.00	1.2065	1.2218	2.6398	1.0861	1.5205	1.4194	0.8241
	−1.00	0.7599	1.0441	1.6076	0.8467	1.1987	0.9895	0.6903
	−0.50	0.6503	0.9963	1.2765	1.0139	1.1501	0.8185	0.6238
	−0.25	0.5625	0.9269	1.1920	0.8331	1.1525	0.6745	0.6108
	0.00	0.5273	0.9384	1.1751	0.8644	1.1245	0.5963	0.5854
	0.25	0.5116	0.8804	1.0526	0.8740	1.1201	0.5384	0.5659
	0.50	0.4722	0.8805	1.0365	0.8265	1.0982	0.4495	0.5550
	1.00	0.4342	0.8094	0.8618	0.7761	1.0449	0.3445	0.5314
	2.00	0.3869	0.7149	0.7307	0.6594	0.9310	0.2898	0.5412
	3.00	0.3682	0.6271	0.6482	0.6051	0.8601	0.3301	0.4645
50	−2.00	0.4425	0.4646	1.4939	0.4414	0.4329	0.4660	0.4840
	−1.00	0.3562	0.3595	0.6054	0.3679	0.3572	0.3535	0.4562
	−0.50	0.3116	0.3448	0.4404	0.2999	0.3645	0.3152	0.4055
	−0.25	0.2772	0.3383	0.3937	0.2915	0.3892	0.2901	0.3905
	0.00	0.2687	0.3771	0.3677	0.2786	0.4123	0.2651	0.3702
	0.25	0.2605	0.3861	0.3272	0.2826	0.4198	0.2314	0.3675
	0.50	0.2387	0.3782	0.3015	0.2723	0.4249	0.2062	0.3514
	1.00	0.2224	0.3660	0.2650	0.2632	0.4116	0.1676	0.3472
	2.00	0.2023	0.3245	0.2295	0.2423	0.3802	0.1467	0.3225
	3.00	0.1928	0.2978	0.2148	0.2196	0.3568	0.1866	0.2463
100	−2.00	0.3185	0.3389	1.1425	0.3333	0.2982	0.2856	0.3513
	−1.00	0.2529	0.2541	0.4988	0.2582	0.2450	0.2311	0.3153
	−0.50	0.2262	0.2405	0.3293	0.2220	0.2408	0.1982	0.2990
	−0.25	0.2082	0.2514	0.2954	0.2090	0.2494	0.1861	0.2798
	0.00	0.2017	0.2674	0.2415	0.1956	0.2651	0.1651	0.2634
	0.25	0.1853	0.2780	0.2037	0.1942	0.2798	0.1445	0.2515
	0.50	0.1729	0.2860	0.1936	0.1894	0.2856	0.1307	0.2520
	1.00	0.1602	0.2599	0.1636	0.1784	0.2778	0.1185	0.2403
	2.00	0.1449	0.2339	0.1463	0.1583	0.2554	0.1034	0.2252
	3.00	0.1315	0.2138	0.1446	0.1528	0.2391	0.1407	0.1711
200	−2.00	0.2216	0.2186	1.0607	0.2217	0.2030	0.1906	0.2469
	−1.00	0.1790	0.1766	0.4359	0.1711	0.1705	0.1549	0.2283
	−0.50	0.1617	0.1634	0.2958	0.1613	0.1591	0.1328	0.2091
	−0.25	0.1463	0.1550	0.2285	0.1485	0.1644	0.1229	0.1988

Table 2 continued

		Method						
n	k	EPM	M1	M2	M3	QM	POS	LCVM
	0.00	0.1366	0.1594	0.1710	0.1387	0.1810	0.1092	0.1828
	0.25	0.1260	0.1766	0.1474	0.1276	0.1894	0.1004	0.1771
	0.50	0.1183	0.1829	0.1317	0.1226	0.1928	0.0892	0.1782
	1.00	0.1117	0.1813	0.1123	0.1151	0.1912	0.0841	0.1563
	2.00	0.1005	0.1619	0.1035	0.1111	0.1764	0.0722	0.1540
	3.00	0.0957	0.1455	0.1043	0.0996	0.1663	0.1015	0.1192

Table 3 RMSE values for k and σ estimators

		Estimator of k			Estimator of σ		
n	k	Pickands	M1	M3	Pickands	M1	M3
100	−2	0.5670	0.4399	0.4146	0.4347	0.3325	0.3325
50	−1	0.6194	0.4633	0.3609	0.4558	0.3692	0.3570
15	−0.5	1.0448	1.1456	0.9665	0.9437	0.9687	0.9126
200	0	0.2613	0.2638	0.1433	0.1707	0.1704	0.1323
50	0.5	0.5034	0.6657	0.3708	0.3102	0.3805	0.2767
100	1	0.3588	0.5821	0.3133	0.1998	0.2649	0.1742
200	2	0.2878	0.4293	0.2576	0.1220	0.1554	0.1057

Table 4 Estimators for k

		Method			
k	n	M1	M2	M3	QM
−3	15	1.5181	1.5385	1.3759	1.4144
	50	0.8152	1.0987	0.7984	0.7727
	100	0.5584	0.9079	0.5874	0.5293
	200	0.3931	0.7644	0.4629	0.3783
−5	15	2.2590	2.3296	2.0876	2.0514
	50	1.2248	1.8065	1.2689	1.1327
	100	0.8232	1.5052	0.9401	0.7711
	200	0.5820	1.2711	0.7481	0.5487

morphodynamics and other problems related to the right tail. Only the 197 observations with periods above 7s were taken into consideration. In order to make a fair comparison with the EPM method developed by Castillo and Hadi (1997), we model these data by the GPD with thresholds at $u = 7,\ 7.5,\ 8,\ 8.5,\ 9$, and 9.5. Similar to Castillo and Hadi (1997), the goodness-of-fit of each estimation procedure is assessed by the

Table 5 Estimators for σ

		Method			
k	n	M1	M2	M3	QM
−3	15	1.7951	5.4333	2.0311	1.7962
	50	0.6066	3.2935	0.6029	0.5341
	100	0.4259	2.6281	0.4233	0.3591
	200	0.2744	2.8588	0.2803	0.2367
−5	15	5.4715	45.4132	9.4836	5.0573
	50	1.1307	36.0548	1.1222	0.8371
	100	0.7217	19.6948	0.6905	0.5276
	200	0.4229	24.4394	0.4342	0.3225

average scaled absolute error (ASAE). The ASAE is defined below.

$$\text{ASAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_{i,n} - \hat{x}_{i,n}|}{\left(x_{n,n} - x_{1,n}\right)}, \tag{23}$$

where

$$\hat{x}_{i,n} = \frac{\hat{\sigma}}{\hat{k}}\left[1 - \left(1 - p_{i,n}\right)^{\hat{k}}\right].$$

In general, the best method is the one with the minimum ASAE value. Let u be a given threshold. If X is a GPD(k, σ), then $X - u$ given that $X > u$ for any u is a GPD$(k, \sigma - ku)$. Let m be a number of exceedances over the threshold value u. First, we compute ASAEs for all seven methods at various threshold levels. The results are summarized in Table 6. For instance, when the threshold $u = 7$, the ASAE for the EPM is 0.0317. It is evident that methods M1, M2, and M3 produce smaller ASAEs. When the threshold $u \geq 8$, the LCVM method yields smaller ASAEs than the EMP method. Next, we compute the standard errors through bootstrap samples for all seven estimation methods at various threshold values of u. The results are based on 1000 iterations and they are summarized in Tables 7 and 8. For example, when the threshold $u = 7$, the EPM method estimate of k is 0.8612 with a standard error of 0.0051. The M1 estimates are similar to the EPM estimates except for the threshold value $u = 9$. Further, we fit the GPD to the Bilbao waves data with $u = 0$. The estimated GPD parameters based on all seven methods and corresponding ASAEs are given in Table 9. According to Table 9, the method M1 yields smallest ASAEs, followed by QM. The method QM has smaller ASAEs than the EPM method. We observe that the methods POS, LCVM, and EPM have comparable ASAEs. In comparison to the existing EPM method, our proposed estimation methods including M1, M3, QM, POS, and LCVM, provide the lowest ASAE values. Moreover, the estimated densities based on all seven methods for the sea waves data are sketched in Fig. 4.

Table 6 ASAEs for seven estimations methods

u	m	M1	M2	M3	QM	POS	LCVM	EPM
7	179	0.0317	0.0241	0.0265	0.0217	0.0317	0.0584	0.0317
7.5	154	0.0134	0.0197	0.0136	0.0139	0.1085	0.0214	0.0134
8	106	0.0210	0.0392	0.0290	0.0176	0.0210	0.0131	0.0210
8.5	69	0.0480	0.0348	0.0386	0.0194	0.0480	0.0471	0.0480
9	41	0.0545	0.0432	0.0478	0.0640	0.1903	0.0341	0.0666
9.5	17	0.0754	0.1070	0.0768	0.0628	0.1227	0.0742	0.0754

Table 7 Estimators for k at various thresholds u and corresponding standard errors in parentheses

u	m	M1	M2	M3	QM	POS	LCVM	EPM
7	179	0.8612	0.7494	0.8053	0.8276	0.8612	1.0328	0.8612
		(0.0126)	(0.0035)	(0.0063)	(0.0185)	(0.0020)	(0.0094)	(0.0051)
7.5	154	0.5539	0.7028	0.6284	0.5997	0.4346	0.7332	0.5539
		(0.0123)	(0.0040)	(0.0064)	(0.0136	(0.0021)	(0.0108)	(0.0055)
8	106	0.5848	0.8791	0.7319	0.7636	0.5848	0.6616	0.5848
		(0.0152)	(0.0052)	(0.0083)	(0.0185)	(0.0027)	(0.0149)	(0.0066)
8.5	69	0.6738	0.9724	0.8231	0.8156	0.6738	0.2977	0.6738
		(0.0196)	(0.0074)	(0.0112)	(0.0233)	(0.0034)	(0.0305)	(0.0083)
9	41	1.2333	1.0550	1.1442	0.5599	0.4113	0.7562	1.0550
		(0.0301)	(0.0113)	(0.0165)	(0.0291)	(0.0056)	(0.0453)	(0.0120)
9.5	17	1.1224	1.5129	1.8177	1.2731	0.9163	1.4882	1.1224
		(0.0497)	(0.0373)	(0.0349)	(0.0462)	(0.0089)	(0.0654)	(0.0182)

Table 8 Estimators for σ at various thresholds u and corresponding standard errors in parentheses

u	m	M1	M2	M3	QM	POS	LCVM	EPM
7	179	2.5264	2.2586	2.3925	3.9495	2.5264	2.9964	2.5264
		(0.0147)	(0.0088)	(0.0089)	(0.0214)	(0.0174)	(0.0148)	(0.0092)
7.5	154	1.5493	1.7744	1.6619	1.5937	1.0827	1.7609	1.5493
		(0.0100)	(0.0080)	(0.0071)	(0.0103)	(0.0183)	(0.0098)	(0.0065)
8	106	1.3872	1.7133	1.5502	1.4930	1.3872	1.4031	1.3872
		(0.0110)	(0.0087)	(0.0082)	(0.0130)	(0.0224)	(0.0097)	(0.0070)
8.5	69	1.2098	1.3771	1.2935	1.2082	1.2098	0.9258	1.2098
		(0.0119)	(0.0089)	(0.0086)	(0.0128)	(0.0255)	(0.0086)	(0.0076)
9	41	1.0179	0.9682	0.9931	0.7631	0.3926	0.7868	0.8155
		(0.0139)	(0.0085)	(0.0089)	(0.0116)	(0.0266)	(0.0085)	(0.0065)
9.5	17	0.4955	0.7853	0.6404	0.5224	0.3710	0.5975	0.4955
		(0.0122)	(0.0111)	(0.0102)	(0.0109)	(0.0407)	(0.0091)	(0.0059)

Table 9 Estimated values of GPD distribution parameters by M1, M2, M3, QM, POS, LCVM and EPM methods with threshold value $u = 0$

Methods	Parameters		ASAE
	σ	k	
M1	72.3776	8.8068	0.3425
M2	15.1893	1.5438	0.8332
M3	43.7834	5.1753	0.4335
QM	97.0948	11.8376	0.2998
POS	73.5043	7.4222	0.4784
LCVM	61.6182	6.2210	0.4795
EPM	31.2768	3.4700	0.4872

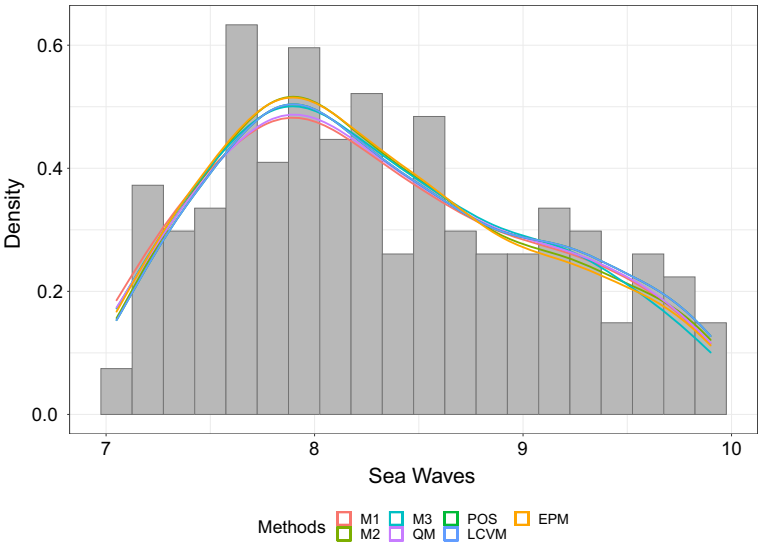

Fig. 4 Estimated densities for all seven methods for zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao Buoy with threshold $u = 0$

5 Conclusions

In this paper, efficient new estimators of the GPD parameters have been proposed. Some of these are very easy to compute and of closed-form, even for very large and negative values of k. If ease of computation is as important as efficiency, then the closed-form estimators M1, M3, and QM are recommended. Simulations results show that for the estimation of k, for a large sample ($n \geq 50$), both POS and LCVM methods perform as well or significantly better than the EPM method. Furthermore, for the estimation of σ, the POS method performs significantly better than the EPM method when the sample size is $n \geq 100$. We recommend M1 and QM methods for negative $k(\leq -3)$ values especially for estimating σ. Our proposed estimations methods are applied to a real data set to illustrate the estimating procedure. Bootstrap confidence intervals are very easy to find for the closed-form estimators and are a

way to obtain confidence intervals for the GPD parameters if desired. We have also developed an *R* package, **EfficientClosedGPD**, that allows readers to estimate GPD parameters using the proposed methods.

Acknowledgements The authors sincerely thank the Associate Editor and the referee for their comments which resulted in this improved version of the work.

Appendix

The proposed efficient new estimators of the GPD parameters are implemented as an *R* package called **EfficientClosedGPD**, freely available on GitHub. For instance, the GPD parameters based on the methods M1, M2, M3, QM, POS, and LCVM can be obtained as follows.

```
rm(list = ls())
library(devtools) # Make sure that the devtools library is loaded
install_github(''suthakaranr/EfficientClosedGPD'')
library(EfficientClosedGPD)  # Load the package
set.seed(650)
x = rgpd2(40, 2, 2) # Generate sample
Method1(x) # Method M1
Method2(x) # Method M2
Method3(x) # Method M3
MethodQM(x) # Method QM
MethodPOS(x) # Method POS
MethodLCVM(x) # Method LCVM
```

References

Belisle CJP (1992) Convergence theorems for a class of simulated annealing algorithms on Rd. J Appl Probab 29:885–895

Castillo E (1988) Extreme value theory in engineering. Academic Press, New York

Castillo E (1994) Extremes in engineering applications. In: Galambos J et al (ed) Proceedings of the conference on extreme value theory and its applications. Kluwer, Gaithersburg, pp 15–42

Castillo E, Hadi AS (1995a) A method for estimating parameters and quantiles of distributions of continuous random variables. Comput Stat Data Anal 20:421–439

Castillo E, Hadi AS (1995b) Modeling lifetime data with applications to fatigue models. J Am Stat Assoc 90:1041–1054

Castillo E, Hadi AS (1997) Fitting the generalized pareto distribution to data. J Am Stat Assoc 92(440):1609–1620

Cheng S, Chou CH (2000a) On the ABLUE of the scale parameter of the generalized Pareto distribution. Tamakang J Math 31(4):317–330

Cheng S, Chou CH (2000b) On the BLUE of the scale parameter of the generalized Pareto distribution. Tamakang J Math 31(3):165–172

Davison A C (1984) Modelling excesses over high thresholds, with an application. In: Tiago de oliveira J (ed) Statistical extremes and applications, Dordrecht, Reidel, pp 461–482

DuMouchel W (1983) Estimating the stable index σ in order to measure tail thickness. Ann Stat 11:1019–1036

Galambos J (1981) Extreme value theory in applied probability. Math Sci 6:13–26

Galambos J (1984) Introduction, order statistics, exceedances. law of large numbers. In: Statistical extremes and applications, NATO ASI Series, Dordrecht, Reidel
Grimshaw SD (1993) Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35:185–191
Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3):339–349
Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme-value distribution by the method of probability weighted moments. Technometrics 27:251–261
Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3:119–131
Smith R L (1984) Threshold methods for sample extremes. In: Tiago de oliveira J (ed) Statistical extremes and applications, Dordrecht, Reidel, pp 621–638
Smith RL (1985) Maximum likelihood estimation in a class of nonregular cases. Biometrika 72:67–90
Van Montfort MAJ, Witter JV (1986) The generalized pareto distribution applied to rainfall depths. Hydrol Sci J 31:151–162
Walshaw D (1990) Discussion of "models for exceedances over high thresholds. by A. C. Davison and R. L. Smith. J R Stat Soc Ser B 52:393–442