

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–14

Monotone trends in the **GPD** scale parameter

M. Roth^{a*}, G. Jongbloed^b, and T. A. Buishand^a

^a*Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, De Bilt,
The Netherlands.*

^b*Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, Delft,
The Netherlands.*

(Received 00 Month 20XX; accepted 00 Month 20XX)

For a sample of independent and identically distributed observations, the parameters of the Generalized Pareto Distribution (**GPD**) can be estimated by the Maximum Likelihood (**ML**) method. In this paper, we drop the assumption of identically distributed random variables. We consider independent observations from **GPD** distributions having a common shape parameter but possibly an increasing trend in the scale parameter. Such a model, with increasing scale parameter, can be used to describe a trend in the observed extremes as time progresses. Estimating an increasing trend in a distribution parameter is common in the field of isotonic regression. We use ideas and tools from that area to compute **ML** estimates of the **GPD** parameters. We also study these estimates in a simulation experiment. Moreover, we apply the approach to the Central England Temperature (**CET**) data.

Keywords: nonparametric estimation; isotonic regression; peaks-over-threshold; **GPD**; Central England temperature

1. Introduction

Modeling extremes is crucial in many branches of modern society. Examples include finance, insurance, and the planning of critical infrastructure such as dikes or sewer systems. Often the Generalized Pareto Distribution (**GPD**) is used to model the tail of the distribution, which is justified by the *Pickands–Balkema–De Haan* theorem. It states that, under certain regularity conditions, the distribution of independent and identically distributed excesses over a threshold u can be approximated by a **GPD**, if u is sufficiently high [19]. We consider the two-parameter **GPD** with $\xi \in \mathbb{R}$ and $\sigma > 0$ denoting the shape and scale parameter, respectively. Its cumulative distribution function is given by

$$G_{\xi,\sigma}(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, \quad (1)$$

with support $y \geq 0$ for $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ for $\xi < 0$. For $\xi = 0$ the **GPD** reduces to the exponential distribution with scale parameter σ . The density of the **GPD** in the

*Corresponding author. Email: roth@knmi.nl

case $\xi \neq 0$ is given by

$$g_{\xi,\sigma}(y) = \frac{1}{\sigma} \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}-1} \tag{2}$$

on its support.

For $\xi > -0.5$, parameter estimates can be obtained using the Maximum Likelihood (ML) approach [7]. The restriction $\xi > -0.5$ does not pose a severe restriction in our setting, as applications in hydrology and more generally in environmental studies usually exhibit shape parameters in the interval $(-0.5, 0.5)$ [11]. Therefore, we restrict ourselves to the case $\xi > -0.5$. The likelihood equations can only be solved numerically, which is usually done by the Newton-Raphson approach or variants including gradient descent steps [7]. [11] shows that for small sample sizes, the probability weighted moment estimators and moment estimators have generally smaller root mean squared error than the ML estimators for $\xi \in [0, 0.4]$ and $\xi \in [-0.2, 0.2]$, respectively. A drawback of these approaches is their lack of flexibility compared to the ML method, which is necessary when it comes to the inclusion of trends.

In many applications, there are reasons to expect a monotone trend in the behavior of extremes. For insurance and infrastructure planning, climate change may lead to such a monotone trend. The change can be described using a monotone function of time, other covariates might be considered as well. In this paper we consider the problem of estimating a nondecreasing trend in the scale parameter of independent observations from GPD distributions. The proposed approach is applied to the daily maxima of the Central England Temperature (CET) data, which are available from 1878 onwards.

2. Maximum Likelihood estimation

Suppose that Y_1, \dots, Y_n are independent random variables, such that $Y_i \sim G_{\xi,\sigma_i}$ for some common shape parameter $\xi > -0.5$ and $0 < \sigma_1 \leq \dots \leq \sigma_n$. We want to estimate the parameter ξ and the vector of scale parameters $\sigma \in C$, where

$$C = \{ \sigma = (\sigma_1, \dots, \sigma_n) \in (0, \infty)^n : \sigma_1 \leq \dots \leq \sigma_n \}. \tag{3}$$

Thus, for this purpose we consider the ML approach. Based on observed values $\mathbf{y} = (y_1, y_2, \dots, y_n)$, the log likelihood for ξ and σ is given by

$$\ell(\xi, \sigma) = \sum_{i=1}^n \ln [g_{\xi,\sigma_i}(y_i)], \tag{4}$$

where $g_{\xi,\sigma}$ is the density of the GPD as given in (2). Note that

$$\begin{aligned} \ln [g_{\xi,\sigma}(y)] &= \ln \left[\frac{1}{\sigma} \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}-1} \right] \\ &= \frac{1}{\xi} \left[\ln(\sigma) - \ln(\sigma + \xi y) \right] - \ln(\sigma + \xi y), \end{aligned} \tag{5}$$

yielding (for $\xi \neq 0$)

$$\ell(\xi, \sigma) = \sum_{i=1}^n \left(\frac{1}{\xi} [\ln(\sigma_i) - \ln(\sigma_i + \xi y_i)] - \ln(\sigma_i + \xi y_i) \right).$$

The maximizing argument $(\hat{\xi}, \hat{\sigma})$ of the log likelihood in Eq. (4) is the **ML** estimator for ξ and σ .

One way to maximize ℓ over $(-0.5, \infty) \times C$ is using the profile (log) likelihood in a two-step procedure. In this approach, for a fine grid of possible ξ -values, the profile likelihood is constructed, i.e.

$$\ell_p(\xi) = \max_{\sigma \in C} \ell(\xi, \sigma). \tag{6}$$

For each ξ , the log likelihood ℓ is maximized over σ . As ξ is one-dimensional, this profile likelihood can be visualized naturally. In the second step, one searches for the ξ maximizing $\ell_p(\xi)$. Together, with the corresponding σ , this defines the **ML** estimate. Of course, in order for this to be applicable, a method is needed to actually compute the profile likelihood, i.e., to maximize ℓ over C for fixed ξ .

LEMMA 2.1 *For each $\xi > -0.5$, there exists a $\sigma_\xi \in C$ such that*

$$\ell(\sigma_\xi, \xi) \geq \ell(\sigma, \xi) \text{ for all } \sigma \in C.$$

Consequently, ℓ_p given in (6) is well defined.

Proof. Fix $\xi > 0$ and note that $\sigma \mapsto \ell(\xi, \sigma)$ is continuous on C . Moreover, note that by (5), for $y > 0$ fixed and $\sigma \downarrow 0$,

$$\ln [g_{\xi, \sigma}(y)] \sim \frac{1}{\xi} \ln(\sigma) \rightarrow -\infty$$

and for $\sigma \rightarrow \infty$,

$$\ln [g_{\xi, \sigma}(y)] \sim -\ln(\sigma) \rightarrow -\infty.$$

Therefore, in maximizing $\sigma \mapsto \ell(\xi, \sigma)$ over C , attention can be restricted to a compact subset of C , namely $\sigma \in C$ for which $\delta \leq \sigma_1 \leq \sigma_n \leq 1/\delta$ for some small $\delta > 0$. This ensures the existence of σ_ξ .

For $\xi = 0$, $\ln [g_{0, \sigma}(y)] = -\ln(\sigma) - y/\sigma$, leading to the same conclusion. In the case $\xi \in (-0.5, 0)$, the restriction $y \leq -\sigma/\xi$ implies that $\sigma \geq -\xi y$. For $\sigma \downarrow -\xi y$ we obtain

$$\ln [g_{\xi, \sigma}(y)] \sim \left(-\frac{1}{\xi} - 1\right) \ln(\sigma + \xi y) \rightarrow -\infty,$$

due to the fact that $(-\frac{1}{\xi} - 1) > 0$ for $\xi \in (-0.5, 0)$. For $\sigma \rightarrow \infty$ we obtain as before

$$\ln [g_{\xi, \sigma}(y)] \sim -\ln(\sigma) \rightarrow -\infty.$$

Thus, attention can be restricted again to a compact subset of C , namely for some small $\delta > 0$

$$\cup_{i=1}^n \{\sigma \in C : -\zeta y_i + \delta \leq \sigma_i \leq 1/\delta\}.$$

On this set, $\sigma \mapsto \ell(\zeta, \sigma)$ is continuous and hence $\ell_p(\zeta)$ is well defined. ■

It is interesting to note that for $\zeta \neq 0$, the function $\sigma \mapsto \ell(\zeta, \sigma)$ is not concave, see the appendix to this article. Therefore, optimization algorithms that need this property cannot be used. In the next section, we will address the problem of computing the function ℓ_p and maximizing this in ζ to maximize the full log likelihood ℓ .

3. Computing the profile log likelihood

In this section two methods are presented to compute ℓ_p . Rather than maximizing the log likelihood over the cone C in \mathbb{R}^n , defined in (3), the negative log likelihood is minimized. The case $\zeta = 0$ is special in this respect. As can be seen in Section 1.5 in [20], the optimization problem for $\zeta = 0$ is a special case of the so-called Gamma extremum problem. The solution of this problem is given by

$$\hat{\sigma} = \text{pr}(\mathbf{y}),$$

where pr is the projection operator from \mathbb{R}^n onto C , defined by

$$\text{pr}(\mathbf{y}) = \arg \min \{\|\mathbf{x} - \mathbf{y}\|_2 : \mathbf{x} \in C\} = \arg \min_{\mathbf{x} \in C} \frac{1}{2} \sum_{i=1}^n (y_i - x_i)^2. \quad (7)$$

An elegant way to obtain the projection $\text{pr}(\mathbf{y})$ explicitly is via the derivative of the greatest convex minorant of a diagram of points. More specifically, defining $P_0 = (0, 0)$ and

$$P_j = \left(j, \sum_{i=1}^j y_i \right), \quad 1 \leq j \leq n, \quad (8)$$

one can construct the greatest convex function lying entirely below the diagram of points. Then taking the left derivative of this function at j , gives $\hat{\sigma}_j$. By construction, the vector $\hat{\sigma} = (\hat{\sigma}_1, \dots, \hat{\sigma}_n)$ is in C . The projection gives the (un-weighted) least squares isotonic regression of $\mathbf{y} = (y_1, \dots, y_n)$ [20, Lemma 1.2.1].

For $\zeta \neq 0$ such a connection between the **ML** estimator of ordered scale parameters in a **GPD** model and plain isotonic regression does not exist. In order to compute $\ell_p(\zeta)$ for values $\zeta \neq 0$, an iterative algorithm is needed. A possible algorithm that can be used in this setting, is the Projected Gradient (**PG**) algorithm, developed independently by [9, 13] for minimizing a continuously differentiable function on a convex subset of \mathbb{R}^n . For $f(\sigma) = -\ell(\zeta, \sigma)$ and a given initial starting value σ_0 the **PG** algorithm is defined

by

$$\sigma_{k+1} = \text{pr} [\sigma_k - \alpha_k \nabla f(\sigma_k)], \tag{9}$$

where $\alpha_k > 0$ is the step size. At each step it has to be ensured, that the new iterate lies within the support of the **GPD**. The following Goldstein-Armijo type choice for the step size is considered:

$$\alpha_k = \beta^{m_k} s, \tag{10}$$

with m_k the smallest integer, such that

$$f(\sigma_{k+1}) \leq f(\sigma_k) - \mu \cdot \langle \nabla f(\sigma_k), \sigma_{k+1} - \sigma_k \rangle, \tag{11}$$

where $s > 0, \beta \in (0, 1)$, and $\mu \in (0, 1)$ are given scalars and $\langle \cdot, \cdot \rangle$ denotes the standard scalar product. In our implementation we set $s = 1, \beta = 0.5$, and $\mu = 1e-4$. [1, 8] showed that in this setting every limit point of $\{\sigma_k\}$ is stationary. If such a limit point exist we take this as the scale estimate $\hat{\sigma}$.

An alternative algorithm that can be used to compute $\ell_p(\xi)$, under the monotonicity constraint, is the Iterative Convex Minorant (**ICM**) algorithm studied by [12] and for instance used in [21] to estimate monotone trends in high daily precipitation quantiles. The **ICM** algorithm can incorporate positive weights, using the weighted projection

$$\text{pr}_{\mathbf{W}}(\mathbf{y}) = \arg \min_{x \in C} \frac{1}{2} \sum_{i=1}^n (y_i - x_i)^2 w_i,$$

where \mathbf{W} is a diagonal matrix with positive diagonal entries w_i . This projection can be obtained explicitly as before from the following point diagram, $P_0 = (0, 0)$ and

$$P_j = \left(\sum_{i=1}^j w_i, \sum_{i=1}^j y_i \cdot w_i \right).$$

For a weight matrix \mathbf{W}^k with positive weights w_i^k , one can define one step in the **ICM** algorithm by:

$$\sigma_{k+1} = \sigma_k + \alpha_k \left(\text{pr}_{\mathbf{W}^k} \left[\sigma_k - (\mathbf{W}^k)^{-1} \nabla f(\sigma_k) \right] - \sigma_k \right). \tag{12}$$

The scaling constant α_k can again be chosen as in (11). If the Hessian \mathbf{H} has positive diagonal entries, these are a natural choice for the weight matrix \mathbf{W} at each step. However, in our case this condition is not fulfilled. After experimenting with different weights, setting $\mathbf{W} = \text{diag}(|\mathbf{H}|)$, i.e. the diagonal matrix consisting of the absolute values of the diagonal entries of the Hessian, worked quite well [2].

The name of the **ICM** algorithm stems from the computation of iterative projections via the greatest convex minorant of a point diagram. Note the geometric difference between the **PG** algorithm and the **ICM** algorithm. In the **PG** algorithm, in principle a whole line segment connecting the current iterate σ_k and $\sigma_k - \nabla f(\sigma_k)$ is projected (us-

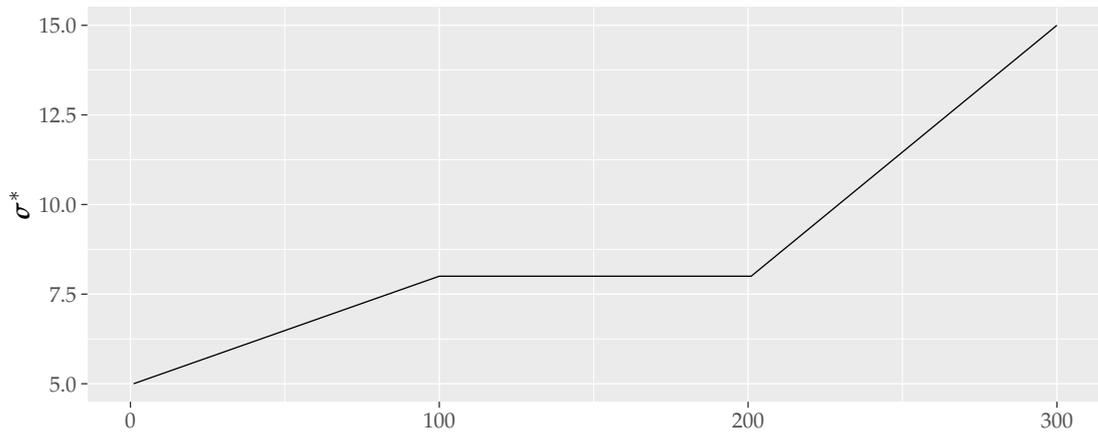


Figure 1. The scale parameter vector σ^* used in the simulation.

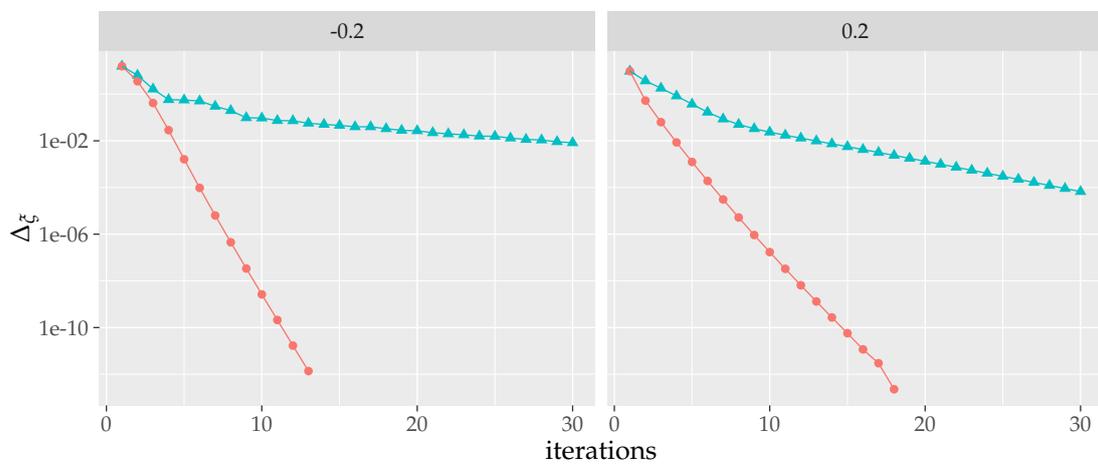


Figure 2. Trace of the deviance Δ_{ξ} based on the **ICM** algorithm (red dots) and the **PG** algorithm (blue triangles) for $\xi = -0.2$ (left) and $\xi = 0.2$ (right).

ing multiple projections), leaving a trace on the cone C that is in general not a line segment, but a ‘broken line’. The **ICM** algorithm just takes the point $\sigma_k - (W^k)^{-1} \nabla f(\sigma_k)$ and projects it on C . Then a new iterate is chosen from the line segment connecting σ_k and this projection, a line that lies completely within C due to convexity of C . Therefore, one can assume that one iteration of the **ICM** algorithm is faster than one of the **PG** algorithm.

Having two algorithms that can be used to compute the profile (log) likelihood function ℓ_p on a grid of ξ -values, the next step is to plot it on such a grid and find its maximum.

4. Simulation study

We carried out a small simulation experiment using the values -0.2 and 0.2 for the shape parameter. The used scale parameter vector σ^* is shown in Figure 1. For the implementation of the algorithms we use the expressions for the needed partial derivatives as given in the Appendix.

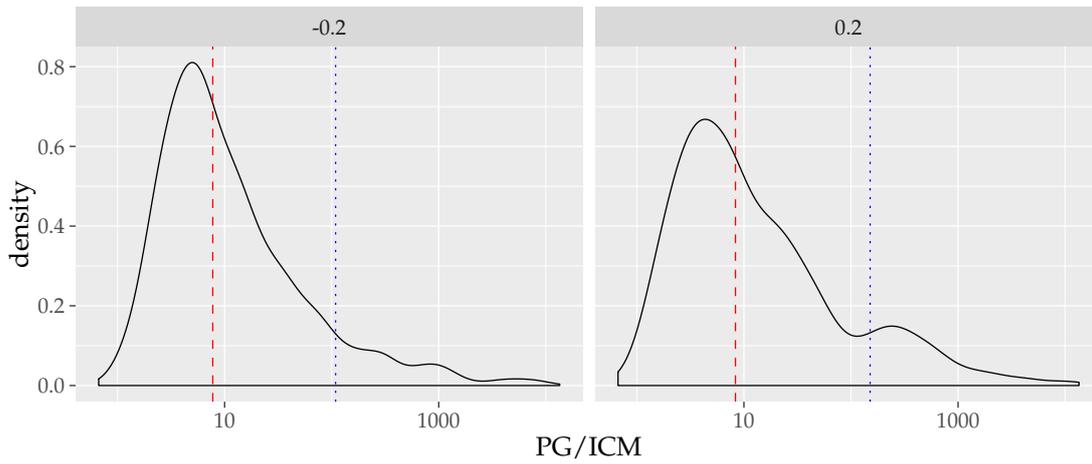


Figure 3. Density of the ratio between the computation time of the **PG** and the **ICM** approach for $\xi = -0.2$ and $\xi = 0.2$ based on 1100 simulations. The vertical line indicates the median (dashed red) and mean (dotted blue) ratio.

First we compare the speed of the two algorithms. Because we use a profile likelihood approach we assume that the shape parameter is known. Moreover, we use σ^* as the starting value for the two algorithms. The **ICM** algorithm needs less iterations to converge. This can be visualized by plotting the deviance measure

$$\Delta_{\xi} := 2 (\ell_p(\xi) - \ell(\xi, \sigma_k)) ,$$

where $\ell_p(\xi)$ is the profile log likelihood for shape parameter ξ and $\ell(\xi, \sigma_k)$ the log likelihood for the k -th iterate. Figure 2 shows an example of such a plot for both shape parameters. The **ICM** algorithm needs only 13 (18) iterations, while the **PG** algorithm uses 286 (90) for $\xi = -0.2$ (0.2). In our simulations the standard number of maximal repetitions, i.e. 10^5 , is sometimes not enough for the **PG** algorithm to converge. With the **ICM** algorithm no problems were observed as the typically needed number of simulations is well below. Although the **PG** algorithm is fully implemented in C++ and the **ICM** algorithm mostly in R, the fact that the **ICM** algorithm uses less and faster iterations has a drastic effect on the computation time, as shown in Figure 3. Only simulations where both approaches converge are shown. In less than 0.5% of the simulations the **PG** algorithm is faster. In all other simulations the **ICM** algorithm is considerably faster, the median of the ratio of the computation time is about 8 and the average is larger than 100.

We now drop the assumption of a known shape parameter. For the computation of the profile likelihood we start at $\xi = 0$, where $\text{pr}(\mathbf{y})$ as defined in (7) is the solution. Then, we compute $\ell_p(\xi)$ for $\xi \in (0, 0.5)$ incrementally moving from 0 towards 0.5, at each step taking the solution of the previous step as starting value. The interval $(0, -0.5)$ is treated correspondingly. The overall restriction to the interval $(-0.5, 0.5)$ is due to the restriction on the **ML** approach and the typical value of the shape parameter in environmental applications, see Section 1.

Figure 4 shows the point-wise median of the scale estimates from 1100 simulations, together with the area between the point-wise 5 and 95 percentiles. The sampling distribution of the estimate is getting more biased at both ends. At the start the bias is negative and at the end the bias is positive. This phenomenon is quite common in the

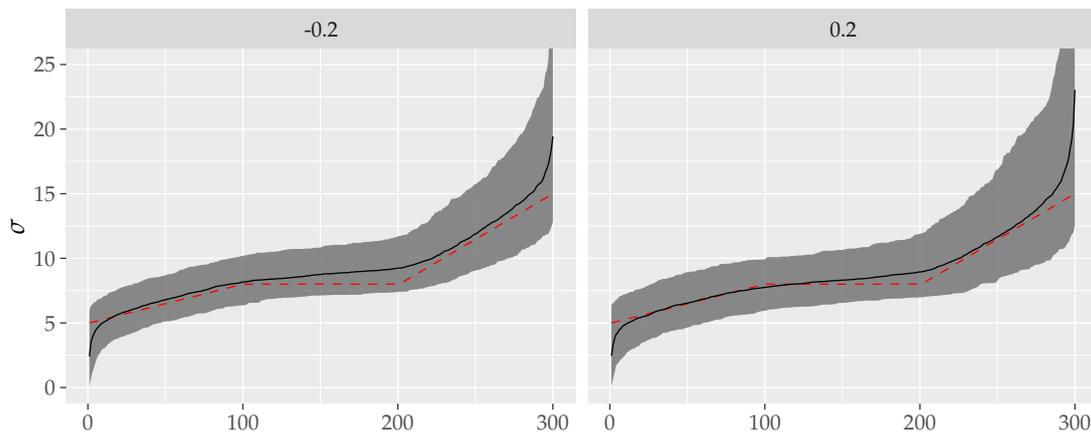


Figure 4. Median (black line) of the scale estimates and 95% confidence band (grey area) for the scale parameter based on 1100 bootstrap samples with scale parameter vector σ^* (dashed red line) and shape parameter $\xi = -0.2$ (left) or 0.2 (right).

isotonic setting and known as the spiking problem [23]. Figure 5 shows the corresponding bootstrap density of the estimated shape parameter. There is apparently a negative bias, which is in line with the literature on the classical setting of extreme value theory, e.g. [24].

Using the profile likelihood approach, one obtains immediately asymptotic profile likelihood confidence intervals for the shape parameter, which are often assumed to be more accurate than bootstrap confidence intervals [15, 22] and those based on the asymptotic normality of $\hat{\xi}$ [4]. [14] justify the use of the profile likelihood confidence interval for semiparametric models. The profile likelihood confidence interval is based on the fact, that the profile deviance

$$D_p(\xi) = 2 (\ell(\hat{\xi}, \hat{\sigma}) - \ell_p(\xi))$$

converges to a χ_1^2 distribution. Hence, by this it can be deduced that

$$C_\alpha = \{\xi : D_p(\xi) \leq c_\alpha\},$$

with c_α being the $(1 - \alpha)$ quantile of the χ_1^2 distribution, constitutes a $(1 - \alpha)$ asymptotic confidence interval for the shape parameter. Figure 6 shows the 95% profile likelihood asymptotic confidence interval for one realization of the simulation.

5. Application

In the following we consider the daily maximum temperatures of the CET data set, which are available from 1878 onwards from the Hadley Centre (<http://www.metoffice.gov.uk/hadobs/hadcet/>). The CET series is a constructed data set, representative of the temperature in Central England, i.e. the area between the Lancashire plains, London and Herefordshire in the West Midlands [17, 18]. In the context of extreme value analysis of non-stationary time series, [16] examined the annual maxima of this data set. [6] considered an r -largest values approach, but for the daily mean

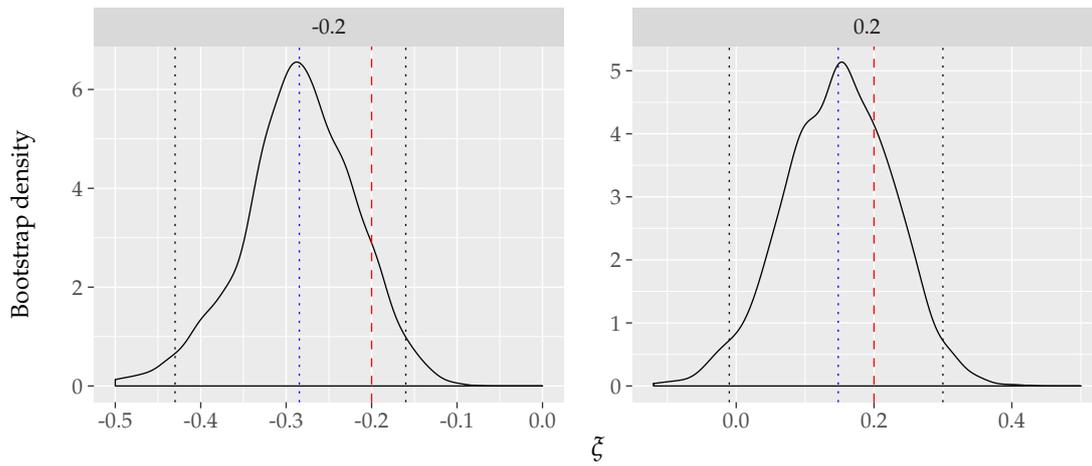


Figure 5. Density of the estimated shape parameter based on a parametric bootstrap. The dashed red line marks the true shape parameter, the blue dotted line the mean estimate, and the black dotted lines mark the 95% bootstrap percentile confidence interval.

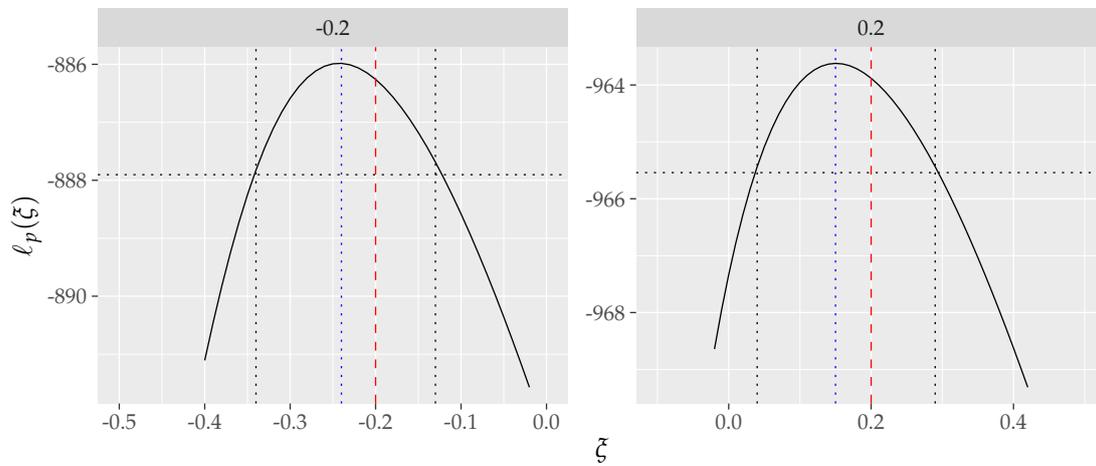


Figure 6. Profile likelihood with 95% confidence interval for the shape parameter for the simulated data. The dashed red line marks the true shape parameter, the blue dotted line the estimate, and the black dotted lines mark the asymptotic 95% confidence interval.

temperatures of the **CET** series, which is available from 1772.

Figure 7 shows the annual maxima of the series for the period 1878 to 2015. The smooth trend in this figure is obtained using loess [3]. Apart from a small trough around 1960, the mean annual maximum seems to increase throughout the series. Therefore, a monotone estimation approach looks promising.

Instead of following the annual maximum approach [16] or the r -largest value approach [6] we consider in our application all peaks over a high threshold. In order to ensure independent peaks, we consider the same declustering method as [6]. That is, we first determine temporal blocks, which are separated by at least 4 days below 16 degrees. From these blocks we take the maximum. In the following we consider only peaks exceeding 18 degrees, which yields on average 5.04 peaks per year. Figure 8 shows the number of peaks per year, together with the 0.25, 0.5, and 0.75 linear regression quantile. It is apparent, that apart from the internal variation there is no trend in the number of peaks per year. Figure 9 shows the peak values together with the 0.5, 0.75, and 0.975 linear regression quantiles. While the median shows a slight decrease,

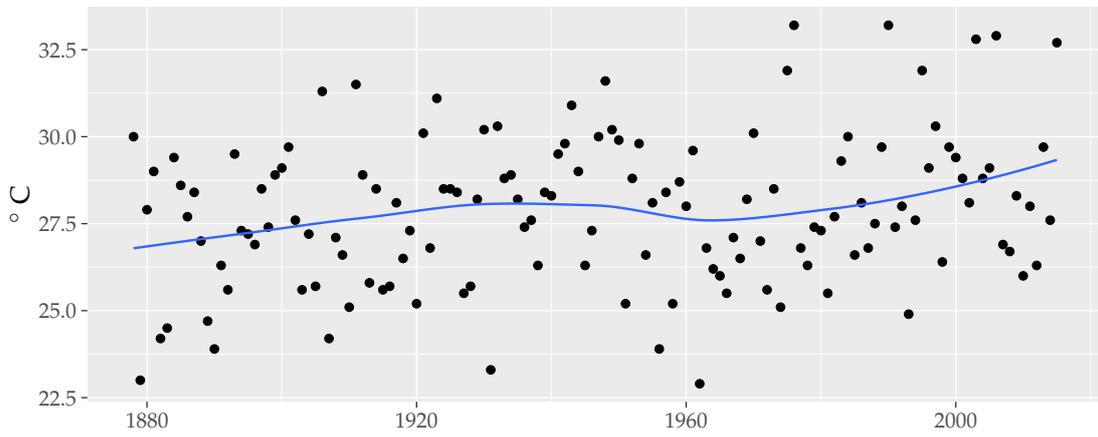


Figure 7. CET annual maxima.

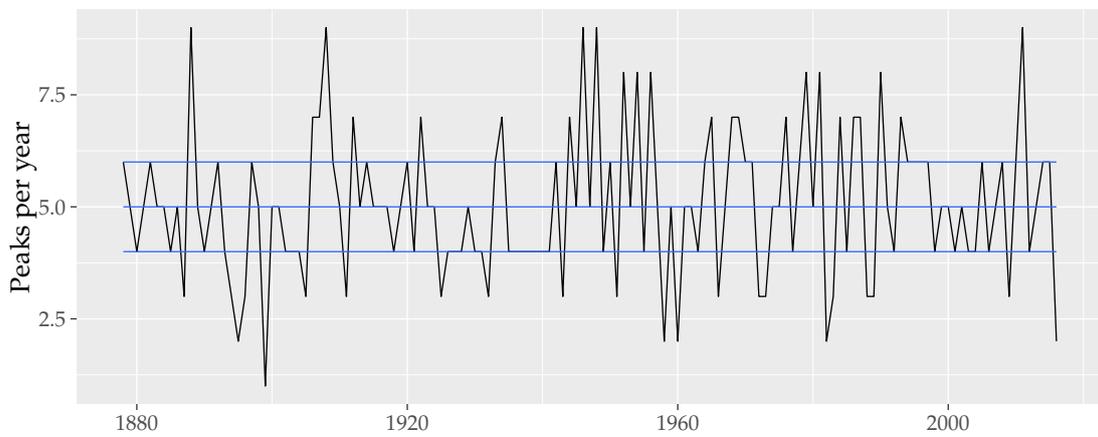


Figure 8. Number of peaks per year in the CET data. The blue lines indicate the linear 0.25, 0.5, and 0.75 regression quantile.

the 0.75 regression quantile has a small positive trend. However, more interesting is the clear positive trend in the 0.975 regression quantile. This seems to justify the use of a constant threshold (which is more influential in the lower part of the distribution) and a monotonically increasing scale parameter (which is more influential in the upper part of the distribution).

Figure 10 shows the obtained profile likelihood confidence interval for the shape parameter. The ML estimate of the shape parameter -0.38 is relatively small compared to the estimate -0.11 given by [16]. The corresponding scale estimate is trimmed at the ends in order to minimize the effect of the spiking. The trimming is achieved by replacing the first (last) 1% of the scale vector entries by the lower (upper) first percentile of the vector entries, see Figure 11. The estimate seems to be in line with the increased trend in recent years detected by [16]. However, overall the trend might be still modeled linearly, in line with the conclusions of [6] for the extremes of the daily mean temperature.

Figure 12 shows the same quantiles as in Figure 9, but adds these quantiles as modeled by the GPD distribution with isotonic scale parameter. Moreover, it shows the

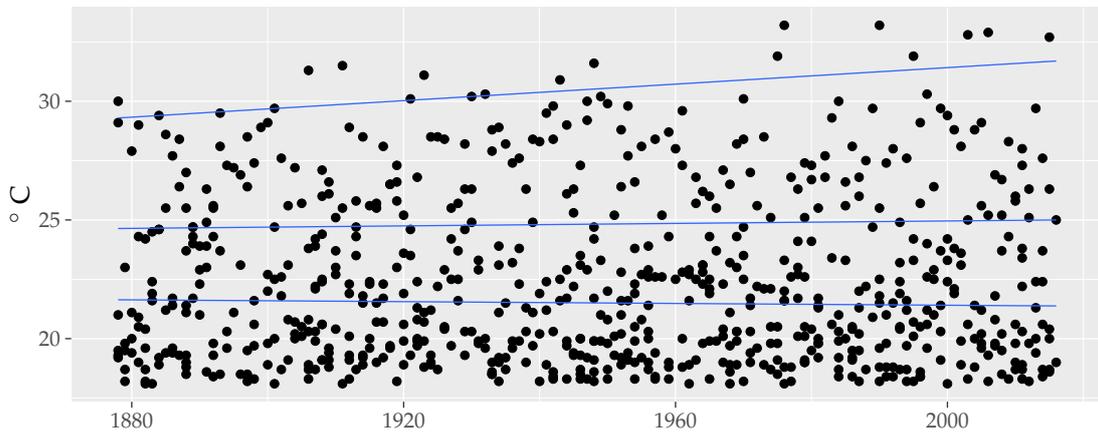


Figure 9. Peak values in the **CET** data with the linear 0.5, 0.75, and 0.975 regression quantile.

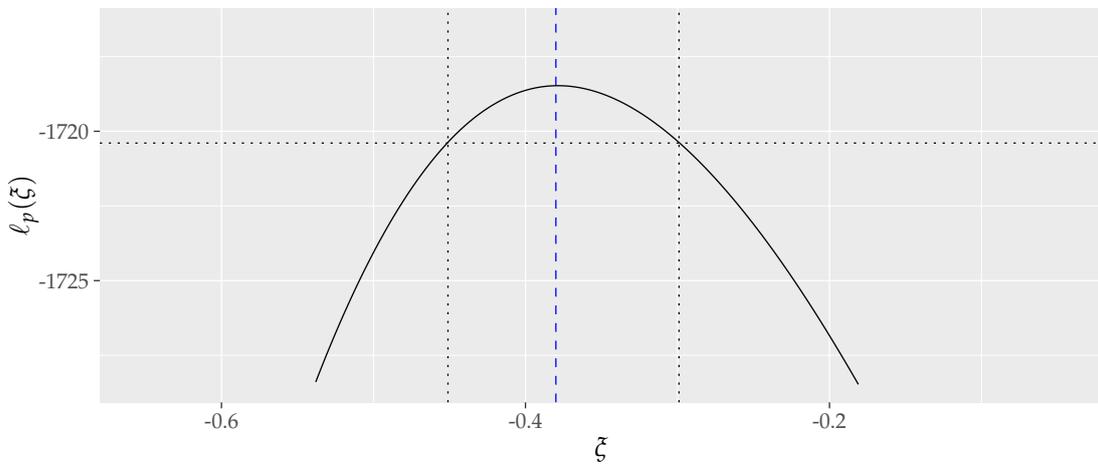


Figure 10. Profile likelihood for the **CET** data with 95% confidence interval for the shape parameter. The dashed blue line marks the final likelihood estimate of the shape parameter.

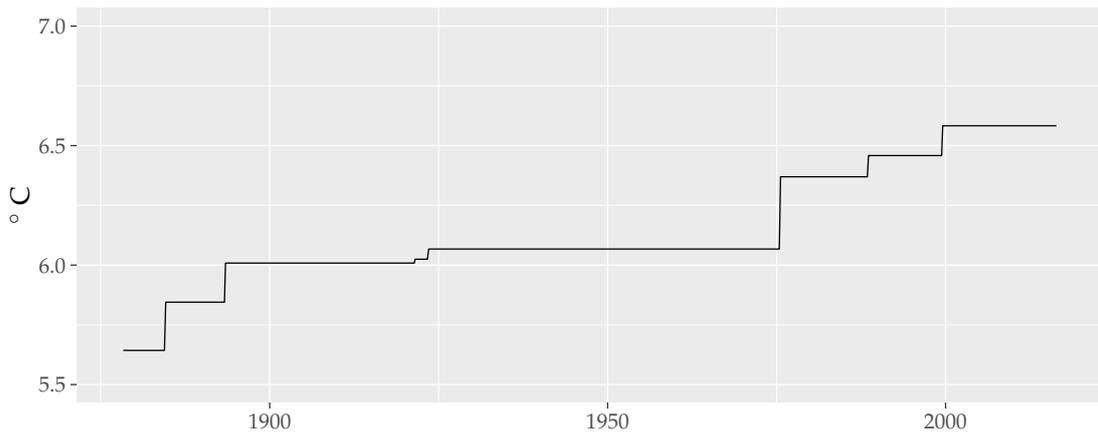


Figure 11. Trimmed scale estimate for the **CET** data.

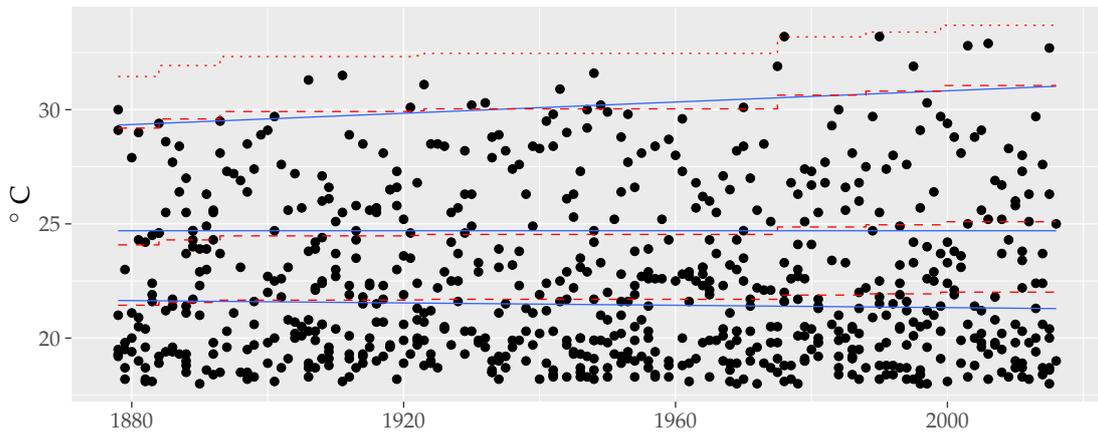


Figure 12. Same as Figure 9 with the 0.5, 0.75, and 0.975 quantile modelled by the GPD in red (dashed lines). The dotted red line on top indicates the 100-year return level.

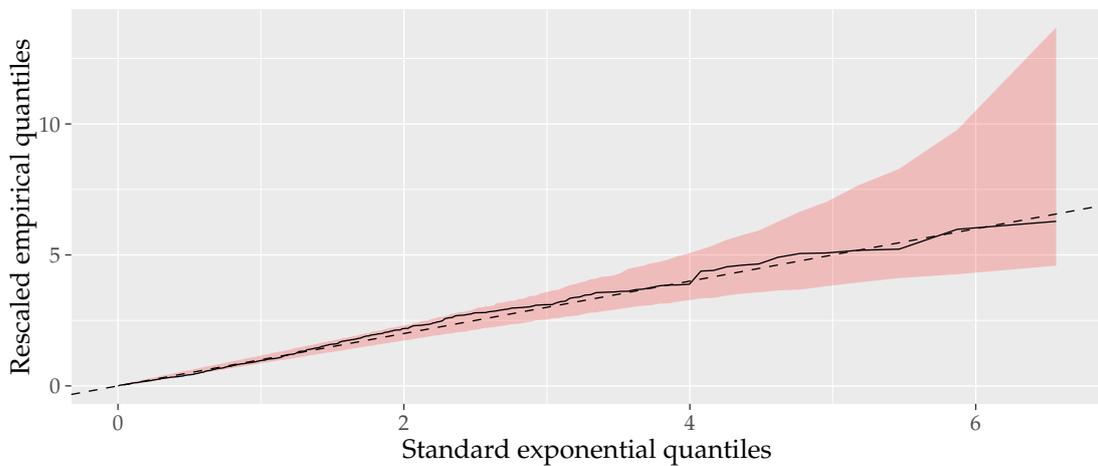


Figure 13. QQ plot for the rescaled empirical quantiles of the CET data (solid black line) with a uniform 95% confidence band.

100-year return level, which corresponds here with the 0.998 quantile and is exceeded on average once in 100 years. This extrapolation is made simple by the GPD approach and demonstrates the advantage over an ordinary quantile regression approach, where these extreme quantiles are less reliable. Figure 13 shows a quantile-quantile plot after rescaling the residuals to a standard exponential distribution with a uniform 95% confidence band, obtained by a parametric bootstrap [5]. Overall the fit seems to be quite good.

6. Conclusion and further directions

We have developed a two-stage procedure to find the ML for independent observations from GPD distributions with common shape parameter ζ and an increasing trend in the scale parameter vector σ . The first step is to compute the profile (log) likelihood for fixed values of ζ . For $\zeta = 0$, there is an exact algorithm to compute this. For $\zeta \neq 0$ and $\zeta > -0.5$, we describe and test two iterative algorithms, the PG algorithm and the ICM algorithm. The ICM algorithm needs less iterations than the PG algorithm and the

iterations are also faster in the **ICM** algorithm. In the second step the profile likelihood is maximized over a grid of shape parameters in order to obtain the **ML** estimates. The **ICM** algorithm is used to obtain the **GPD** parameters in a peaks-over-threshold model, with increasing trend in the scale parameter, for the daily maximum temperatures in the **CET** data set.

The algorithms are available via the R package `gpdIcm`. These make it possible to perform significance tests for the null hypothesis that the scale parameters are equal against the alternative that these are increasing. Moreover, testing the null hypothesis that the scale parameter is linearly increasing against a monotone alternative becomes viable. In the present example, however, the use of linear modeling seems adequate. Likelihood ratio tests, but also permutation-based tests can be studied using the algorithms described in this paper.

Appendix

Consider the first (partial) derivative

$$\frac{\partial \ln g_{\xi, \sigma}(y)}{\partial \sigma} = \frac{y - \sigma}{\sigma(\sigma + \xi y)}.$$

This shows that $\sigma \mapsto \ln g_{\xi, \sigma}(y)$ is unimodal with maximum $\sigma = y$ for fixed ξ . The second derivative is given by

$$\frac{\partial^2 \ln g_{\xi, \sigma}(y)}{\partial \sigma^2} = \frac{(\sigma - y)^2 - (\xi + 1)y^2}{\sigma^2(\sigma + \xi y)^2}.$$

It follows that

$$\frac{\partial^2 \ln g_{\xi, \sigma}(y)}{\partial \sigma^2} = 0 \iff \sigma = y(1 \pm \sqrt{1 + \xi}).$$

This shows that the second derivative exhibits in general at least one change of sign. Thus, the log likelihood is not concave for $\xi \neq 0$.

References

- [1] D. Bertsekas, *On the Goldstein-Levitin-Polyak gradient projection method*, IEEE Transactions on Automatic Control 21 (1976), pp. 174–184.
- [2] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Teubner, Stuttgart, Leipzig, 1998, Available at <http://www-library.desy.de/elbook.html>, e-Version from 2012.
- [3] W.S. Cleveland, *Robust locally weighted regression and smoothing scatterplots*, Journal of the American Statistical Association 74 (1979), pp. 829–836.
- [4] S.G. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, UK, 2001.
- [5] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, UK, 1997.
- [6] A.C. Davison and N.I. Ramesh, *Local likelihood smoothing of sample extremes*, Journal of the Royal Statistical Society B 62 (2000), pp. 191–208.
- [7] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events*, Springer, Berlin, 1997.

- [8] E.M. Gafni and D.P. Bertsekas, *Convergence of a gradient projection method*, LIDS-P 1201, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, Massachusetts, 1982.
- [9] A.A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society 70 (1964), pp. 709–710.
- [10] J.R.M. Hosking, *Algorithm AS 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution*, Applied Statistics 34 (1985), pp. 301–310.
- [11] J.R.M. Hosking and J.R. Wallis, *Parameter and quantile estimation for the generalized Pareto distribution*, Technometrics 29 (1987), pp. 339–349.
- [12] G. Jongbloed, *The iterative convex minorant algorithm for nonparametric estimation*, Journal of Computational and Graphical Statistics 7 (1998), pp. 310–321.
- [13] E.S. Levitin and B.T. Polyak, *Constrained minimization methods*, USSR Computational Mathematics and Mathematical Physics 6 (1966), pp. 1–50.
- [14] S.A. Murphy and A.W. van der Vaart, *On profile likelihood*, Journal of the American Statistical Association 95 (2000), pp. 449–465.
- [15] J. Obeysekera and J.D. Salas, *Quantifying the uncertainty of design floods under nonstationary conditions*, Journal of Hydrologic Engineering 19 (2013), pp. 1438–1446.
- [16] S.A. Padoan and M.P. Wand, *Mixed model-based additive models for sample extremes*, Statistics and Probability Letters 78 (2008), pp. 2850–2858.
- [17] D.E. Parker and B. Horton, *Uncertainties in Central England temperature 1878–2003 and some improvements to the maximum and minimum series*, International Journal of Climatology 25 (2005), pp. 1173–1188.
- [18] D.E. Parker, T.P. Legg, and C.K. Folland, *A new daily Central England temperature series 1772–1991*, International Journal of Climatology 12 (1992), pp. 317–342.
- [19] R.D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd ed., Birkhäuser, Basel, 2007.
- [20] T. Robertson, F.T. Wright, and R.L. Dykstra, *Order Restricted Statistical Inference*, Wiley, Chichester, UK, 1988.
- [21] M. Roth, T.A. Buishand, and G. Jongbloed, *Trends in moderate rainfall extremes: A regional monotone regression approach*, Journal of Climate 28 (2015), pp. 8760–8769.
- [22] T. Schendel and R. Thongwichian, *Flood frequency analysis: Confidence interval estimation by test inversion bootstrapping*, Advances in Water Resources 83 (2015), pp. 1–9.
- [23] M. Woodroffe and J. Sun, *A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing*, Statistica Sinica 3 (1993), pp. 501–515.
- [24] J. Zhang and M.A. Stephens, *A new and efficient estimation method for the generalized Pareto distribution*, Technometrics 51 (2009), pp. 316–325.

Acronyms

CET	Central England Temperature
GPD	Generalized Pareto Distribution
ICM	Iterative Convex Minorant
ML	Maximum Likelihood
PG	Projected Gradient