Monotone trends in the distribution of climate extremes

Roth, Martin; Jongbloed, Geurt; Buishand, Adri

DOI
10.1007/s00704-018-2546-x

Publication date
2018

Document Version
Final published version

Published in
Theoretical and Applied Climatology

Citation (APA)
Roth, M., Jongbloed, G., & Buishand, A. (2018). Monotone trends in the distribution of climate extremes. *Theoretical and Applied Climatology*, *136 (2019)*(3-4), 1175–1184. https://doi.org/10.1007/s00704-018-2546-x

Noname manuscript No.
(will be inserted by the editor)

MONOTONE TRENDS IN THE DISTRIBUTION OF CLIMATE EXTREMES

Martin Roth · Geurt Jongbloed · Adri Buishand

Received: date / Accepted: date

Abstract The Generalized Pareto Distribution (GPD) is often used in the statistical analysis of climate extremes. For a sample of independent and identically distributed observations, the parameters of the 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. In a simulation experiment we show that the Iterative Convex Minorant (ICM) algorithm is much faster than the Projected Gradient (PG) algorithm. We apply the approach to the daily maxima of the Central England Temperature (CET) data. A clear positive trend in the GPD scale parameter is found, leading to an increase in the 100-year return level from about 31 degrees in the 1880s to 34 degrees in 2015.

Keywords nonparametric estimation · isotonic regression · peaks-over-threshold · GPD · Central England temperature

1 Introduction

Statistical modelling of climate extremes is important for many branches of modern society. Examples include insurance, heat stress, and the planning of critical

M. Roth and A. Buishand
RWDW, Royal Netherlands Meteorological Institute,
De Bilt, The Netherlands
Tel.: +49-8321-7878655
E-mail: martin@roth-mail.org
G. Jongbloed
Delft Institute of Applied Mathematics, Delft University of Technology,
Delft, The Netherlands.

infrastructure such as dams 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 indentically distributed excesses over a threshold u can be approximated by a GPD, if u is sufficiently high (Reiss and Thomas, 2007). 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

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

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 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 (Embrechts et al., 1997). The restriction $\xi > -0.5$ does not pose a severe restriction in our setting, as climate data exhibit shape parameters in the interval $(-0.5, 0.5)$. For instance, for daily rainfall slightly positive values of ξ up to about 0.3 are usually found (Roth et al., 2012; Langousis et al., 2016; Carreau et al., 2017), whereas for daily maximum temperatures ξ tends to be negative but not less than -0.5 (Lucio et al., 2010) 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 (Embrechts et al., 1997). Hosking and Wallis (1987) show 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.

The characteristics of climate extremes may vary over time and, hence, the GPD parameters may be no longer constant. It is often assumed that these parameters vary (log-) linearly with time or a time-dependent covariate (e.g. Coelho et al., 2008; Beguería et al., 2010; Kyselý et al., 2010; Acero et al., 2011; Roth et al., 2012; Van de Vyver, 2012; Tramblay et al., 2013). A number of these authors also studied quadratic changes in GPD parameters. The shape parameter can often be kept constant.

Another current in the literature focuses on non-parametric, smooth trends in the GPD parameters (e.g. Hall and Tajvidi, 2000; Chavez-Demoulin and Davison, 2005).

In the present paper we leave the strong restriction of (log-)linearity, but keep more structure than in the above mentioned non-parameteric trend studies by im-

posing monotonicity on the scale estimate. For climate-change studies such a setting is of interest, when an increasing or decreasing trend is expected that is non-linear. This might, for one instance, be the case for temperature extremes as a result of the increased atmospheric greenhous gas concentrations. This trend can be described using a monotone function of time, other covariates might be considered as well. For the rest of the paper the shape parameter is assumed to be constant.

In section 2 we introduce the estimation method, using two different algorithms. Section 3 presents a simulation experiment of the proposed estimators. In section 4 the proposed approach is applied to the daily maxima of the Central England Temperature (CET) data, which are available from 1878 onwards. The conclusion is given in section 5.

2 Method

2.1 Maximum Likelihood estimation

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

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

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

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

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

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

yielding (for $\xi \neq 0$)

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

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

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 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.

The Lemma in the appendix shows that $\ell_p(\xi)$ is well defined. However, the function $\sigma \mapsto \ell(\xi, \sigma)$ is not concave for $\xi \neq 0$, as shown in the appendix. 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 ξ to maximize the full log likelihood ℓ.

2.2 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 Eq. 3, the negative log likelihood is minimized. The case $\xi = 0$ is special in this respect. As can be seen in Section 1.5 in Robertson et al. (1988), the optimization problem for $\xi = 0$ is a special case of the so-called Gamma extremum problem. The solution of this problem is given by

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

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

$$\begin{aligned} \mathrm{pr}(\mathbf{y}) &= \arg\min\{||\mathbf{x} - \mathbf{y}||_2 : \mathbf{x} \in C\} \\ &= \arg\min_{\mathbf{x} \in C} \frac{1}{2} \sum_{i=1}^{n} (y_i - x_i)^2. \end{aligned} \tag{7}$$

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

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

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 $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$ is in C. The projection gives the (un-weighted) least squares isotonic regression of $\mathbf{y} = (y_1, \ldots, y_n)$ (Robertson et al., 1988, Lemma 1.2.1).

For $\xi \neq 0$ 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 $\ell_p(\xi)$ for values $\xi \neq 0$, an iterative algorithm is needed. A possible algorithm that can be used in this setting, is the Projected Gradient (PG) algorithm, developed

independently by Goldstein (1964) and Levitin and Polyak (1966) for minimizing a continuously differentiable function on a convex subset of $\mathbb{R}^n$.

For $f(\sigma) = -\ell(\xi, \sigma)$ and a given initial starting value σ_0 the PG algorithm is defined by

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

where $a_k > 0$ 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

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

where $s > 0, \beta \in (0,1)$, and $\mu \in (0,1)$ are given scalars and $< \cdot, \cdot >$ denotes the standard scalar product. In our implementation we set $s = 1$, $\beta = 0.5$, and $\mu = 1\mathrm{e}{-4}$. Bertsekas (1976) and Gafni and Bertsekas (1982) 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 Jongbloed (1998) and for instance used in Roth et al. (2015) to estimate monotone trends in high daily precipitation quantiles. The ICM algorithm can incorporate positive weights, using the weighted projection

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

where 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 W^k with positive weights w_i^k, one can define one step in the ICM algorithm by:

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

The scaling constant α_k can again be chosen as in Eq. 11. If the Hessian H has positive diagonal entries, these are a natural choice for the weight matrix W at each step. However, in our case this condition is not fulfilled. After experimenting with different weights, setting $\mathrm{W} = \mathrm{diag}(|\mathrm{H}|)$, i.e. the diagonal matrix consisting of the absolute values of the diagonal entries of the Hessian (Blobel and Lohrmann, 1998), worked quite well.

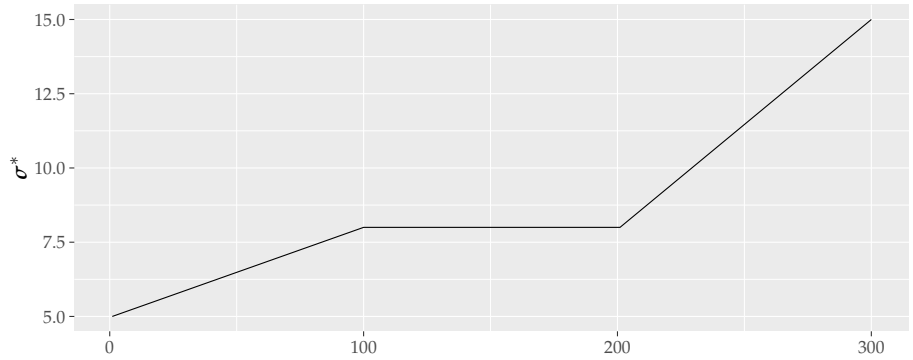

Fig. 1 The scale parameter vector σ^* used in the simulation.

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 (using 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.

3 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 Fig. 1. For the implementation of the algorithms we use the expressions for the needed partial derivatives as given in the Appendix.

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

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

where $\ell_p(\tilde{\xi})$ is the profile log likelihood for shape parameter $\tilde{\xi}$ and $\ell(\tilde{\xi}, \sigma_k)$ the log likelihood for the k-th iterate. Fig. 2 shows an example of such a plot for both shape parameters. The ICM algorithm needs only 13 (18) iterations, while the PG algo-

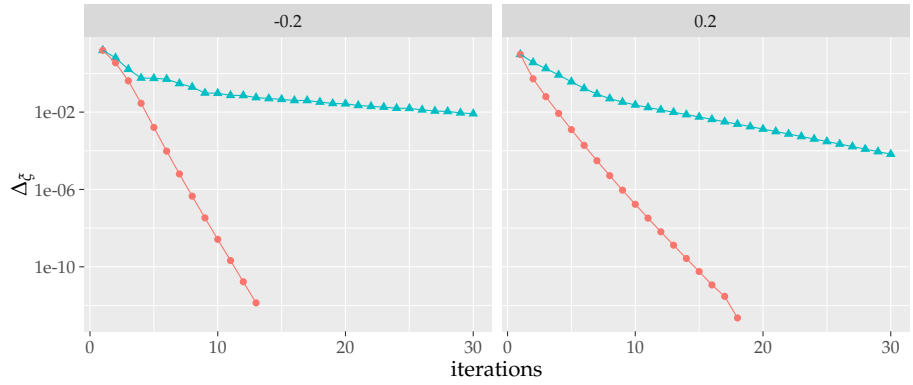

Fig. 2 Trace of the deviance $\Delta_{\tilde{\xi}}$ based on the ICM algorithm (red dots) and the PG algorithm (blue triangles) for $\tilde{\xi} = -0.2$ (left) and $\tilde{\xi} = 0.2$ (right).

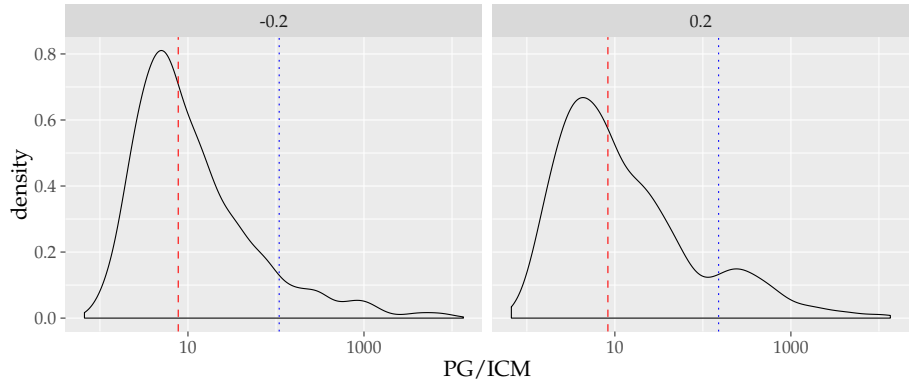

Fig. 3 Density of the ratio between the computation time of the PG and the ICM approach for $\tilde{\xi} = -0.2$ and $\tilde{\xi} = 0.2$ based on 1100 simulations. The vertical line indicates the median (dashed red) and mean (dotted blue) ratio.

rithm uses 286 (90) for $\tilde{\xi} = -0.2$ (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 Fig. 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 Eq. 7 is the solution. Then, we compute $\ell_p(\xi)$ for $\xi \in (0, 0.5)$ incrementally moving from 0

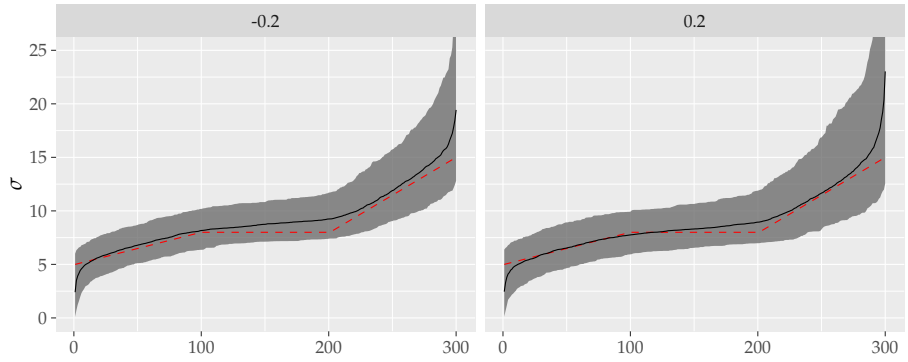

Fig. 4 Median (black line) of the scale estimates and 95% confidence band (grey area) for the scale parameter based on 1100 bootstrap samples with scale parameter vector σ^* (dashed red line) and shape parameter $\xi = -0.2$ (left) or 0.2 (right).

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.

Fig. 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 isotonic setting and known as the spiking problem (Woodroofe and Sun, 1993). Fig. 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. Zhang and Stephens, 2009).

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 (Obeysekera and Salas, 2013; Schendel and Thongwichian, 2015) and those based on the asymptotic normality of $\hat{\xi}$ (Coles, 2001). Murphy and Van der Vaart (2000) 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

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

converges to a χ^2_1 distribution. Hence, by this it can be deduced that

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

with c_α being the $(1-\alpha)$ quantile of the χ^2_1 distribution, constitutes a $(1-\alpha)$ asymptotic confidence interval for the shape parameter. Fig. 6 shows the 95% profile likelihood asymptotic confidence interval for one realization of the simulation.

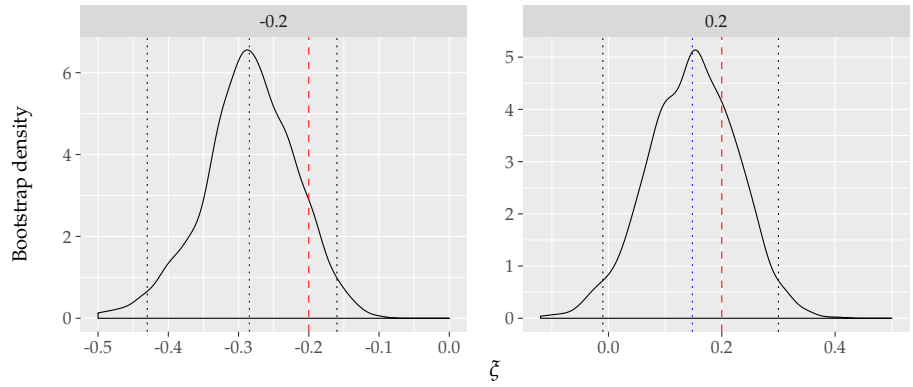

Fig. 5 Density of the estimated shape parameter based on a parametric bootstap. The dashed red line marks the true shape parameter, the blue dotted line the mean estimate, and the black dotted lines mark the 95% bootstrap percentile confidence interval.

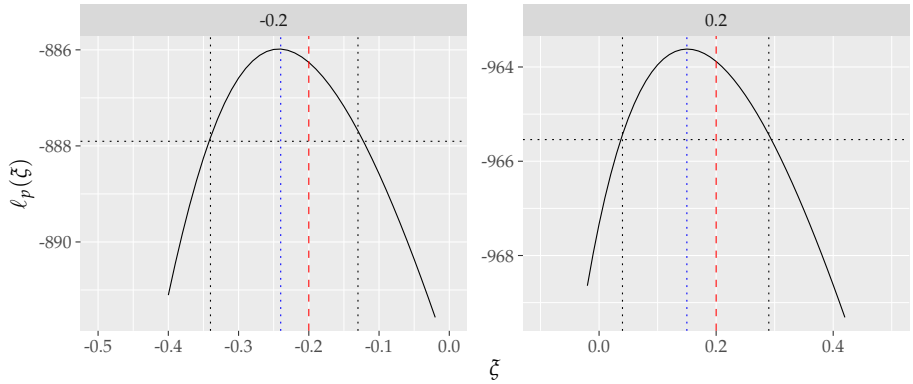

Fig. 6 Profile likelihood with 95% confidence interval for the shape parameter for the simulated data. The dashed red line marks the true shape parameter, the blue dotted line the estimate, and the black dotted lines mark the asymptotic 95% confidence interval.

4 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 beteen the Lancashire plains, London and Herefordshire in the West Midlands (Parker et al., 1992; Parker and Horton, 2005). In the context of extreme value analysis of non-stationary time series, Davison and Ramesh (2000) considered the r-largest values in each year, but for the daily mean temperatures of this time series from 1772. Padoan and Wand (2008) examined the annual maxima of the daily mean temperatures from 1878.

Fig. 7 shows the annual maxima of the series for the period 1878 to 2015. The smooth trend in this figure is obtained using loess (Cleveland, 1979). Apart from a

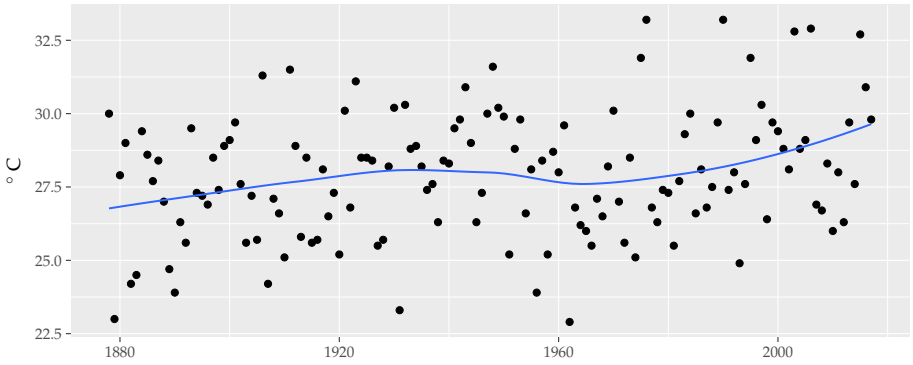

Fig. 7 CET annual maxima.

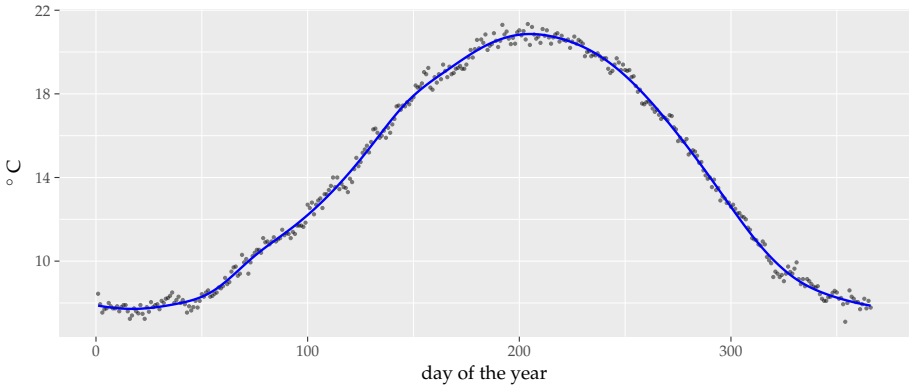

Fig. 8 Annual cycle of the 0.6 quantile of the daily maximum temperature. The points mark the daily values and the blue line indicates the cyclic GAM fit.

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 Padoan and Wand (2008) or the *r*-largest value approach Davison and Ramesh (2000) we consider in our application all peaks over a high threshold. In order to ensure independent peaks, we first decluster the data. In the first step of the declustering procedure, we determine the 0.6 quantile for each day and smooth these quantiles using a cyclic GAM model (Rigby and Stasinopoulos, 2005), see Fig. 8. We consider two blocks as independent, when there are at least 4 days below this quantile. From these blocks we take the maximum.

In the following we consider only peaks exceeding 24 degrees, which yields on average 2.86 peaks per year. Fig. 9 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. Fig. 10 shows the peak values together with the 0.5, 0.75, and 0.975 linear regression quan-

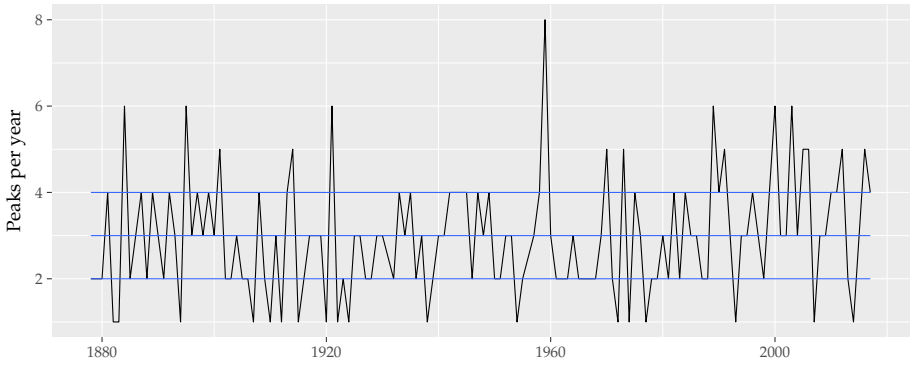

Fig. 9 Number of peaks per year in the CET data. The blue lines indicate the linear 0.25, 0.5, and 0.75 regression quantile.

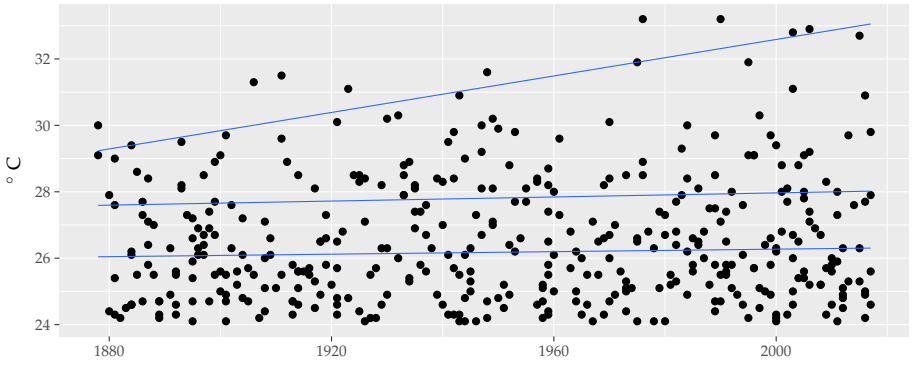

Fig. 10 Peak values in the CET data with the linear 0.5, 0.75, and 0.975 regression quantile.

tiles. While the median and the 0.75 regression quantile show a small positive trend, the 0.975 regression quantile exhibits a clear positive trend. This seems to justify the use of a constant threshold (which dominates the trend in the lower and middle parts of the distribution) and a montonically increasing scale parameter (which is more influential in the upper part of the distribution). A notable point in Fig. 10 is that there are 5 peaks exceeding $32^\circ C$ after 1970 and no peaks of this magnitude before that year, which partly explains the rather strong positive trend in the 0.975-quantile.

Fig. 11 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 Padoan and Wand (2008) for the annual maximum daily mean temperature. 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 Fig. 12. The scale parameter increases steadily

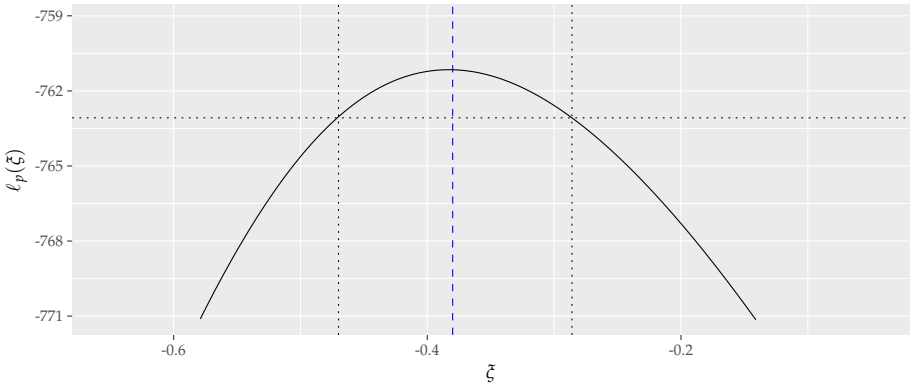

Fig. 11 Profile likelihood for the CET data with 95% confidence interval for the shape parameter. The dashed blue line marks the final likelihood estimate of the shape parameter.

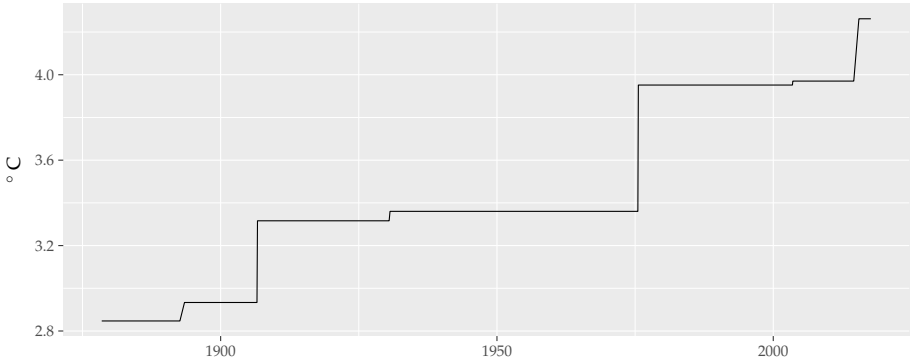

Fig. 12 Trimmed scale estimate for the CET data.

about 1° C over the period of the record. This trend might, however, still be modelled linearly, in line with the conclusions of Davison and Ramesh (2000) for the extremes of the daily mean temperature.

Fig. 13 shows the same quantiles as in Fig. 10, but adds these quantiles as modeled by the GPD distribution with isotonic scale parameter. Moreover, it shows the 100-year return level, which corresponds here with the 0.9965 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. The 100-year return level shows an increase of about 3° C since the 1880s as a result of the trend in the GPD scale parameter. Fig. 14 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 (Davison and Hinkley, 1997). Overall the fit seems to be good. In particular, the plot does not suggest a larger value for the shape

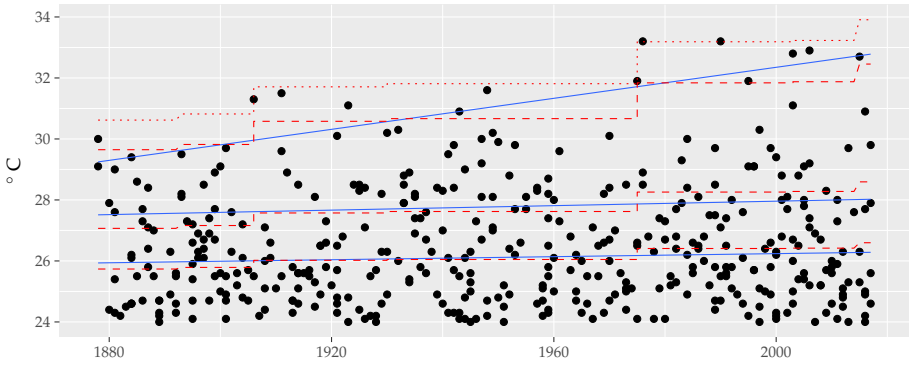

Fig. 13 Same as Figure 10 with the 0.5, 0.75, and 0.975 quantile modelled by the GPD in red (dashed lines). The dotted red line on top indicates the 100-year return level.

Fig. 14 QQ plot for the rescaled empirical quantiles of the CET data (solid black line) with a uniform 95% confidence band.

parameter, i.e., a longer upper tail, as found by Padoan and Wand (2008) for the maxima of the daily mean temperatures.

5 Conclusion and further directions

We have developed a two-stage procedure to find the ML estimates for independent observations from GPD distributions with common shape parameter ξ and an increasing trend in the scale parameter vector σ, which is useful to describe nonlinear increasing trends in climate extremes. The first step is to compute the profile (log) likelihood for fixed values of ξ. For $\xi = 0$, there is an exact algorithm to compute this. For $\xi \neq 0$ and $\xi > -0.5$, 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. A clear positive trend in the GPD scale parameter is found, leading to an increase of about 3 degrees in the 100-year return level between the 1880s and the present.

The algorithms are available via the R package gpdIcm (`https://github.com/MartinRoth/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 montone 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.

For the CET data a constant threshold could be taken. Other applications may require a monotone trend in the threshold as well, which could be estimated for instance with the method described in Roth et al. (2015). In the case that the shape parameter cannot be assumed constant, but a linear or other simple parametric model seems plausible, one can extend the grid search for the profile likelihood approach correspondingly. It should be noted that the power of detecting a trend in the shape parameter is low for small and moderate sample sizes if $\xi > 0$, see also Naveau et al. (2014). Moreover, a trend in the shape parameter is sensitive to outliers (Roth et al., 2012).

Appendix

Lemma *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 $\sigma \downarrow 0$,

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

and for $\sigma \to \infty$,

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

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 $\xi = 0$, $\ln\left[g_{0,\sigma}(y)\right] = -\ln(\sigma) - y/\sigma$, 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

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

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

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

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

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

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

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

Acero FJ, García JA, Gallego MC (2011) Peaks-over-threshold study of trends in extreme rainfall over the Iberian peninsula. Journal of Climate 24:1089–1105, doi: 10.1175/2010JCLI3627.1

Beguería S, Angulo-Martínez M, Vicente-Serrano SM, López-Moreno JI, El-Kenawy A (2010) Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. International Journal of Climatology 31(14):2102–2114, doi: 10.1002/joc.2218

Bertsekas D (1976) On the Goldstein-Levitin-Polyak gradient projection method. IEEE Transactions on Automatic Control 21(2):174–184, doi: 10.1109/TAC.1976.1101194

Blobel V, Lohrmann E (1998) Statistische und numerische Methoden der Datenanalyse. Teubner, Stuttgart, Leipzig, URL `http://www-library.desy.de/elbook.html`, e-Version from 2012

Carreau J, Naveau P, Neppel L (2017) Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. Water Resources Research 53:4407–4426, doi: 10.1002/2017WR020758

Chavez-Demoulin V, Davison AC (2005) Generalized additive modelling of sample extremes. Applied Statistics 54(1):207–222, doi: 10.1111/j.1467-9876.2005.00479.x

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74:829–836, doi: 10.1080/01621459.1979.10481038

Coelho CAS, Ferro CAT, Stephenson DB, Steinskog DJ (2008) Methods for exploring spatial and temporal variability of extreme events in climate data. Journal of Climate 21(10):2072–2092, doi: 10.1175/2007JCLI1781.1

Coles SG (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, London, UK

Davison AC, Hinkley DV (1997) Bootstrap Methods and their Application. Cambridge University Press, Cambridge, UK

Davison AC, Ramesh NI (2000) Local likelihood smoothing of sample extremes. Journal of the Royal Statistical Society B 62(1):191–208, doi: 10.1111/1467-9868.00228

Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling Extremal Events. Springer, Berlin

Gafni EM, Bertsekas DP (1982) Convergence of a gradient projection method. LIDS-P 1201, Massachusetts Institute of Technology, Cambridge, Massachusetts

Goldstein AA (1964) Convex programming in Hilbert space. Bulletin of the American Mathematical Society 70:709–710, doi: 10.1090/S0002-9904-1964-11178-2

Hall P, Tajvidi N (2000) Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. Statistical Science 15(2):153–167, doi: 10.2307/2676729

Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3):339–349, doi: 10.1080/00401706.1987.10488243

Jongbloed G (1998) The iterative convex minorant algorithm for nonparametric estimation. Journal of Computational and Graphical Statistics 7(3):310–321, doi: 10.1080/10618600.1998.10474778

Kyselý J, Picek J, Beranová R (2010) Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. Global and Planetary Change 72(1-2):55–68, doi: 10.1016/j.gloplacha.2010.03.006

Langousis A, Mamalakis A, Puliga M, Deidda R (2016) Threshold detection for the generalized pareto distribution: Review of representative methods and application to the noaa ncdc daily rainfall database. Water Resources Research 52:2659–2681, doi: 10.1002/2015WR018502

Levitin ES, Polyak BT (1966) Constrained minimization methods. USSR Computational Mathematics and Mathematical Physics 6(5):1–50

Lucio PS, Silva AM, Serrano AI (2010) Changes in occurrences of temperature extremes in continental portugal: a stochastic approach. Meteorological Applica-

tions 17:404–418, doi: 10.1002/met.171

Murphy SA, Van der Vaart AW (2000) On profile likelihood. Journal of the American Statistical Association 95(450):449–465

Naveau P, Guillou A, Rietsch T (2014) A non-parametric entropy-based approach to detect changes in climate extremes. Journal of the Royal Statistical Society B 76(5):861–884, doi: 10.1111/rssb.12058

Obeysekera J, Salas JD (2013) Quantifying the uncertainty of design floods under nonstationary conditions. Journal of Hydrologic Engineering 19(7):1438–1446, doi: 10.1061/(ASCE)HE.1943-5584.0000931

Padoan SA, Wand MP (2008) Mixed model-based additive models for sample extremes. Statistics and Probability Letters 78(17):2850–2858, doi: 10.1016/j.spl.2008.04.009

Parker DE, Horton B (2005) Uncertainties in Central England temperature 1878–2003 and some improvements to the maximum and minimum series. International Journal of Climatology 25:1173–1188, doi: 10.1002/joc.1190

Parker DE, Legg TP, Folland CK (1992) A new daily Central England temperature series 1772–1991. International Journal of Climatology 12:317–342

Reiss RD, Thomas M (2007) Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields, 3rd edn. Birkhäuser, Basel

Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. Applied Statistics 54:507–554, doi: 10.1111/j.1467-9876.2005.00510.x

Robertson T, Wright FT, Dykstra RL (1988) Order Restricted Statistical Inference. Wiley, Chichester, UK

Roth M, Buishand TA, Jongbloed G, Klein Tank AMG, van Zanten JH (2012) A regional peaks-over-threshold model in a nonstationary climate. Water Resources Research 48(11):W11,533, doi: 10.1029/2012WR012214

Roth M, Buishand TA, Jongbloed G (2015) Trends in moderate rainfall extremes: A regional montone regression approach . Journal of Climate 28(22):8760–8769, doi: 10.1175/JCLI-D-14-00685.1

Schendel T, Thongwichian R (2015) Flood frequency analysis: Confidence interval estimation by test inversion bootstrapping. Advances in Water Resources 83:1–9

Tramblay Y, Neppel L, Carreau J, Najib K (2013) Non-stationary frequency analysis of heavy rainfall events in southern France. Hydrological Sciences Journal 58(2):280–294, doi: 10.1080/02626667.2012.754988

Van de Vyver H (2012) Evolution of extreme temperatures in Belgium since the 1950s. Theoretical and Applied Climatology 107(1):113–129, doi: 10.1007/s00704-011-0456-2

Woodroofe M, Sun J (1993) A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. Statistica Sinica 3(2):501–515

Zhang J, Stephens MA (2009) A new and efficient estimation method for the generalized Pareto distribution. Technometrics 51(3):316–325, doi: 10.1198/tech.2009.08017

Acronyms

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