



UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2014:33

Peaks-Over-Threshold Modelling of Environmental Data

Esther Bommier

Examensarbete i matematik, 30 hp
Handledare och examinator: Jesper Rydén
September 2014

A large, faint watermark of the Uppsala University seal is visible in the bottom right corner of the page. The seal features a sun with rays in the center, surrounded by the Latin text "ALMA MATER UPPSALAENSIS" and "VERITAS".

Department of Mathematics
Uppsala University

Contents

1	Introduction	1
1.1	Background	1
1.2	Outline	2
2	The Peaks-Over-Threshold Method	3
2.1	Extreme Value Analysis	3
2.2	The Block Maxima Approach	4
2.3	The Generalized Pareto Distribution	4
2.4	Estimation of the Parameters	6
2.5	Return Levels	6
2.6	Declustering	8
3	Choice of Threshold	9
3.1	The Mean Residual Life Plot	9
3.2	The Parameter Stability Plot	10
3.3	The Dispersion Index Plot	10
3.4	Rules of Thumb	11
3.5	A Multiple-Threshold Model	11
4	Non-Stationary Extremes	15
4.1	Trend Variation	15
4.2	Model Choice	15
5	Extremal Mixture Models	16
5.1	Parametric Mixture Models	16
5.2	Semi-parametric Mixture Models	17
5.3	Non-parametric Mixture Models	17
6	Application: Temperature Extremes in Uppsala	18
6.1	Dataset	18
6.2	Choice of Threshold	20
6.3	Declustering	23
6.4	Non-stationarity: Modelling	25
6.5	Extremal Mixture Models	25

6.6	Estimation of the GPD Parameters	26
6.7	Return Levels and Return Periods	27
6.8	Comparison with the Block Maxima Method	29
6.9	Analysis of the Daily Minima in Uppsala	30
7	Conclusion	33
	Bibliography	34

1 Introduction

1.1 Background

The scope of this study is to analyse temperatures records in Uppsala, Sweden, during the time period 1840-2012, using extreme value theory.

Extreme value theory is used for describing the distribution of rare events, especially in financial, insurance, hydrology or environmental applications, where the risk of extreme events is of interest (Reiss et al., 2001). Where inference about extremes can be challenging due to the scarcity of data, extreme value models are used to study the behaviour of the tail of the distribution. A first approach is the block maxima method, which has already been applied on the temperatures in Uppsala for the time period 1840-2001, see Rydén (2011).

Another very popular extreme value model is the Generalized Pareto Distribution, which can give a good model for the upper tail, providing reliable extrapolation for exceedances over a sufficiently high threshold. This is called the Peaks-Over-Threshold (POT) method. In application of such a model the choice of threshold is crucial, as it defines which part of the data can be considered as extreme, or more formally where the asymptotically justified extreme value models will provide an adequate approximation to the tail of the distribution.

However, it is common to observe non-stationarity in the context of environmental data, due to seasonal variations and long-term trends. Hence, the series of the temperatures of Uppsala cannot be considered stationary and another approach accounting for time-varying parameters needs to be developed.

In the last decade, extremal mixture models have been established (see the review by Scarrott and MacDonald 2012), which have the advantage of capturing not only the tail distribution above the threshold, but also the bulk (data below the threshold) distribution. Moreover, instead of applying a model with a fixed threshold method, these models treat the threshold as a parameter to be estimated. In this way, the uncertainty due to the selection of the threshold can be taken into account through the inference method.

1.2 Outline

The outline of the present report is as follows. First, the theory of the POT method is introduced. As the choice of threshold is crucial, several methods for the selection of the threshold are presented. Then, an approach for non-stationary extremes is presented. In another section, some extremal mixture models are described. Finally, all these approaches are applied to a real dataset, consisting of the temperatures in Uppsala.

2 The Peaks-Over-Threshold Method

2.1 Extreme Value Analysis

The purpose of extreme value analysis is to model the risk of extreme, rare events, by finding reliable estimates of the frequency of these events. For example, for the design of a breakwater, a coastal engineer would seek to estimate the 50-year wave and design the structure accordingly. Extreme value analysis is based on the asymptotic behaviour of observed extremes. A theoretical definition and more detailed explanation can be found in Coles (2001). The following sections are adapted from this work.

Let X_1, \dots, X_n be n independent and identically distributed random variables, with distribution function F ($F(x) = \mathbf{P}\{X \leq x\}$). Define the maximum order statistic

$$M_n = \max \{X_1, \dots, X_n\} \quad (2.1)$$

For example, if n is the number of observations in a year, then M_n corresponds to the annual maximum. Its distribution function is given by F^n and for all $x < \sup \{x; F(x) < 1\}$, $F^n(x) \rightarrow 0$ as $n \rightarrow \infty$. The Gnedenko theorem or extreme value theorem (1943) for maxima is an analogous of the central limit theorem for the mean. Normalizing series $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ need to be introduced, and a law for $M_n^* = \frac{M_n - b_n}{a_n}$ can be obtained.

Theorem 1 (Gnedenko). *Under certain regularity conditions on the distribution function F , there exist $\xi \in \mathbb{R}$ and two normalizing series $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ such that for all $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} = H_\xi(x),$$

with, if $\xi > 0$,

$$H_\xi(x) = \begin{cases} 0 & \text{if } x \leq \mu \\ \exp \left\{ - \left(\frac{x - \mu}{a} \right)^{-1/\xi} \right\} & \text{if } x > \mu \end{cases}$$

if $\xi < 0$,

$$H_\xi(x) = \begin{cases} \exp \left\{ - \left[- \left(\frac{x - \mu}{a} \right) \right]^{1/\xi} \right\} & \text{if } x < \mu \\ 1 & \text{if } x \geq \mu \end{cases}$$

if $\xi = 0$,

$$H_0(x) = \exp \left\{ - \exp \left[- \left(\frac{x-b}{a} \right) \right] \right\} \text{ for all } x \in \mathbb{R}$$

Remarks:

- The distribution function H_ξ is called extreme value distribution. We then say that F is in the domain of attraction of H_ξ .
- These distributions are indexed by a parameter ξ called the shape parameter, and three domains of attraction should be distinguished, depending on the sign of ξ : F is in the Fréchet domain of attraction if $\xi > 0$, the Gumbel domain of attraction if $\xi = 0$ or the Weibull domain of attraction if $\xi < 0$.

Definition 1. *The Generalized Extreme Value (GEV) distribution is a parametrization of these three distributions into one formula:*

$$H_\xi(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (2.2)$$

defined on the set $\{x; 1 + \xi(x - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$ (location parameter), $\sigma > 0$ (scale parameter) and $-\infty < \xi < \infty$ (shape parameter).

2.2 The Block Maxima Approach

One approach to study extremes is the block maxima method, which considers the distribution of the maximum order statistic defined in (2.1).

A GEV distribution is then fitted to the series of extremal observations.

But this block maxima approach is wasteful of data as only one data point in each block is taken. The second highest value in one block may be larger than the highest of another block and this is generally not accounted for.

The peaks-over-threshold (POT) method is a way to avoid this drawback as it uses the data more efficiently: the idea is to consider several large values instead of only the largest one. This approach consists in using observations which are above a given threshold, called *exceedances*.

2.3 The Generalized Pareto Distribution

The POT method has been used in many fields to identify extremal events such as loads, wave heights, floods, wind velocities, insurance claims, etc. This approach provides a model for independent exceedances over a high threshold. It is clear that this method needs the determinations of a threshold which is neither too high (to get enough observations) nor too low (not to take into account non-extreme values). In the following, the threshold will be denoted u .

Definition 2. Let X be a random variable with distribution function F and right endpoint $x_F = \sup\{x; F(x) < 1\}$. For all $u < x_F$, the function

$$F_u(x) = P\{X - u \leq x | X > u\}, \quad x \geq 0$$

is called the distribution function of exceedances above threshold u .

Remark: By the conditional probabilities, F_u can also be defined as

$$F_u(x) = \begin{cases} \frac{F(u+x) - F(u)}{1 - F(u)} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

Let $Y = X - u$ for $X > u$ and for n observed variables X_1, \dots, X_n , we can write $Y_j = X_i - u$ such that i is the index of the j th exceedance, $j = 1, \dots, n_u$. The distribution of the exceedances (Y_1, \dots, Y_{n_u}) can be approximated by a Generalized Pareto Distribution (Pickands, 1975).

Definition 3. Let σ_u be a strictly positive function and $\xi \in \mathbb{R}$. The Generalized Pareto Distribution (GPD) is given by

$$G_{\xi, \sigma_u}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right) & \text{if } \xi = 0 \end{cases} \quad (2.3)$$

where $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\frac{\sigma_u}{\xi}$ if $\xi < 0$.

Theorem 2. If F is in one of the three domains of attraction of the GEV (Fréchet, Gumbel or Weibull), then there exist a strictly positive function σ_u and $\xi \in \mathbb{R}$ such that

$$\lim_{u \uparrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - G_{\xi, \sigma_u}(y)| = 0,$$

where G_{ξ, σ_u} is the GPD and F_u is the distribution function of the exceedances above u .

Thus, for u large enough, the distribution of exceedances above u is approximated by a GPD:

$$F_u \approx G_{\xi, \sigma_u}$$

The parameters of the GPD are uniquely determined by those of the GEV: the shape parameter ξ is equal to that of the corresponding GEV and the scale parameter σ_u is a function of the GEV location and shape parameters:

$$\sigma_u = \sigma + \xi(u - \mu) \quad (2.4)$$

The notation σ_u was used to distinguish the scale parameter from the corresponding parameter of the GEV and to emphasise its dependence on u . From now on, this distinction will be dropped out for notational convenience.

As with the GEV, the shape parameter of the GPD is dominating in determining the behaviour of the tail. If $\xi < 0$, the distribution of excesses has an upper bound at $u - \frac{\tilde{\sigma}}{\xi}$, while on the contrary if $\xi > 0$, the distribution has no upper limit.

Note: The expected value of the GPD with parameters σ and ξ is

$$E(Y) = \frac{\sigma}{1 - \xi}, \text{ provided } \xi < 1 \quad (2.5)$$

When $\xi \geq 1$, the mean is infinite.

2.4 Estimation of the Parameters

Several numerical methods for estimating σ and ξ have been proposed (Pickands estimator, Probability Weighted Moments method, Moment method, Maximum Likelihood method), but the only one which combines theoretical efficiency, gives a general basis for inference and extends straightforwardly to the models incorporating non-stationarity and covariate dependence is the maximum likelihood method (Anderson et al., 2001).

Let y_1, \dots, y_{n_u} be a sequence of n_u exceedances of a threshold u . For $\xi \neq 0$, the log-likelihood is derived from (2.3) as

$$l(\sigma, \xi) = -n_u \log \sigma - (1 + 1/\xi) \sum_{i=1}^{n_u} \log \left(1 + \xi \frac{y_i}{\sigma} \right) \quad (2.6)$$

provided $(1 + \xi y_i/\sigma) > 0$ for $i = 1, \dots, n_u$; otherwise $l(\sigma, \xi) = \infty$.

When $\xi = 0$, the log-likelihood can be derived in a similar way:

$$l(\sigma) = -n_u \log \sigma - (1/\sigma) \sum_{i=1}^{n_u} y_i$$

However, the maximum likelihood estimator (MLE) is not always valid and regularity conditions do not always exist. The MLE is valid for $\xi > -1$, but the asymptotically normal properties of the MLE is only valid for $\xi > -1/2$. When $\xi < -1$, maximum likelihood estimators generally do not exist (see Smith (1985) for more details).

2.5 Return Levels

It is often of interest to evaluate return levels, for example the N -year return level x_N , which is exceeded once every N years.

Suppose that the GPD with parameters σ and ξ is a suitable model for exceedances, then the probability of the exceedances of a variable X over a suitable high threshold u can be written as

$$\mathbf{P} \{X > x | X > u\} = \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

provided that $x > u$ and $\xi \neq 0$. Let $\zeta_u = \mathbf{P}\{X > u\}$, i.e. ζ_u is the probability of the occurrence of an exceedance of a high threshold u , then

$$\mathbf{P}\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}$$

The subscript u in ζ_u emphasizes the fact that this value depends on the choice of threshold u .

Then, the level x_m that is exceeded on average once every m observations will be obtained from solving

$$\zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi} = \frac{1}{m}$$

Rearranging yields

$$x_m = u + \frac{\sigma}{\xi} \left[(m\zeta_u)^\xi - 1 \right] \quad (2.7)$$

which is valid for m large to ensure that $x > u$.

For $\xi = 0$, applying the same procedure to the corresponding equation in (2.3) leads to

$$x_m = u + \sigma \log(m\zeta_u) \quad (2.8)$$

x_m is called the m -observation return level. If we are interested in the N -year return level, let n_y be the number of observations per year, then $m = N \times n_y$. Hence, the N -year return level is

$$x_N = \begin{cases} u + \frac{\sigma}{\xi} \left[(Nn_y\zeta_u)^\xi - 1 \right] & \text{if } \xi \neq 0 \\ u + \sigma \log(Nn_y\zeta_u) & \text{if } \xi = 0 \end{cases} \quad (2.9)$$

This equation suggests that in order to determine the N -year return level, three parameters need to be fitted, σ , ξ and ζ_u . If we assume that the exceedances of a high threshold u are rare events, ζ_u could be expected to follow a Poisson distribution. Here, we deviate slightly from Coles (2001) who suggests that the number of exceedances of u follows the binomial distribution $\text{Bin}(n, \zeta_u)$. The Poisson distribution is characterised by the parameter λ , which is the mean of threshold exceedances per unit time. Then, ζ_u can be estimated as

$$\hat{\zeta}_u = \frac{\hat{\lambda}}{n_y},$$

where an unbiased estimate of λ is given by

$$\hat{\lambda} = \frac{n_u}{M}$$

with n_u the number of exceedances over the selected threshold u and M the number of years of records. Reformulating equation (2.9) in terms of λ , we get

$$x_N = \begin{cases} u + \frac{\sigma}{\xi} \left[(\lambda N)^\xi - 1 \right] & \text{if } \xi \neq 0 \\ u + \sigma \log(\lambda N) & \text{if } \xi = 0 \end{cases} \quad (2.10)$$

σ and ξ are estimated using the maximum likelihood method.

2.6 Declustering

As mentioned previously, the POT method requires the exceedances to be mutually independent. However, for the temperatures data, threshold exceedances are seen to occur in groups: an extremely warm day is likely to be followed by another. In order to deal with dependent variables and be able to apply the POT method, a commonly used technique is declustering, which filters the dependent observations to obtain a set of threshold excesses that are approximately independent.

The clusters are defined as follows. First, a threshold is fixed and clusters of exceedances are consecutive exceedances of this threshold. Then, a run length (minimum separation) r is set between each cluster: a cluster is terminated whenever the separation between two threshold exceedances is greater than the run length r .

The maximum excess in each cluster is identified, and these cluster maxima can be assumed to be independent. Then, the GPD can be fitted to the cluster maxima.

3 Choice of Threshold

The choice of the threshold is not straightforward; indeed, a compromise has to be found: a high threshold value reduces the bias as this satisfies the convergence towards the extreme value theory but however increases the variance for the estimators of the parameters of the GPD, as there will be fewer data from which to estimate parameters. A low threshold value results in the opposite i.e. a high bias but a low variance of the estimators, there is more data with which to estimate the parameters.

3.1 The Mean Residual Life Plot

Davison and Smith (1990) suggest a graphical method for the selection of the threshold. This method is based on the mean of the GPD. Suppose the GPD is valid as a model for the excesses of a threshold u_0 generated by a series X_1, \dots, X_n . By equation (2.5),

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi},$$

provided $\xi < 1$ and where σ_{u_0} is the GPD scale parameter for exceedances over threshold u_0 .

The threshold stability property of the GPD means that if the GPD is a valid model for excesses over some threshold u_0 , then it is valid for excesses over all thresholds $u > u_0$. Hence, for $u > u_0$,

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}, \quad (3.1)$$

by (2.4). Thus, for all $u > u_0$, $E(X - u | X > u)$ is a linear function of u . Furthermore, $E(X - u | X > u)$ is the mean of the excesses of the threshold u , and can be estimated by the sample mean of the threshold excesses. This leads to the mean residual life plot defined by the locus of points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right); u < x_{\max} \right\},$$

where $x_{(1)}, \dots, x_{(n_u)}$ consist of the n_u observations that exceed u , and x_{\max} is the largest of the X_i .

It means that for a range of thresholds u , we identify the corresponding mean threshold excess, then plot this mean threshold excess against u , and look for the value u_0 above

which we can see linearity in the plot. Indeed, if the GPD assumption is correct, then the plot should follow a straight line with intercept $\frac{\sigma_u}{1-\xi}$ and slope $\frac{\xi}{1-\xi}$ (before it becomes unstable due to the few very high data points). Confidence intervals can be added to this plot as the empirical mean can be supposed to be normally distributed (Central Limit Theorem). However, normality doesn't hold anymore for high thresholds as there are less and less excesses. Moreover, by construction, this plot always converge to the point $(x_{\max}, 0)$.

A drawback of such a plot is that its interpretation can be challenging.

3.2 The Parameter Stability Plot

Another graphical method which is widely used to determine the threshold u is the parameter stability plot. The idea of this plot is that if the exceedances of a high threshold u_0 follow a GPD with parameters ξ and σ_{u_0} , then for any threshold u such that $u > u_0$, the exceedances still follow a GPD with shape parameter $\xi_u = \xi$ and scale parameter $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$.

Let

$$\sigma^* = \sigma_u - \xi_u u$$

This new parametrisation does not depend on u any longer, given that u_0 is a reasonably high threshold.

The plot is defined by the locus of points

$$\{(u, \sigma^*); u < x_{\max}\} \text{ and } \{(u, \xi_u); u < x_{\max}\},$$

where x_{\max} is the maximum of the observations.

Thus, estimates of σ^* and ξ_u are constant for all $u > u_0$ if u_0 is a suitable threshold for the asymptotic approximation. The threshold should be chosen at the value where the shape and scale parameters remain constant.

3.3 The Dispersion Index Plot

Ribatet (2006) presented another plot to determine the threshold, namely the dispersion index plot, which is particularly useful when dealing with time series. This plot relies on the fact that the data is generated by a Poisson process. Indeed, according to the extreme value theory, the exceedances above a high threshold should be Poisson distributed.

Let X be a Poisson distributed random variable with parameter λ . Then,

$$\mathbb{P}\{X = k\} = \exp(-\lambda) \frac{\lambda^k}{k!}, \quad k \in \mathbb{N},$$

and $\mathbb{E}[X] = \text{Var}[X]$. The idea is to use the dispersion index DI introduced by Cunnane (1979), defined by

$$DI = \frac{\sigma^2}{\mu},$$

where σ^2 is the intensity of the Poisson process and λ is the mean number of events in a block (a year in our case). One can test if the ratio DI differs from 1: if DI is significantly close to 1, the sample can be modelled by a Poisson process and the correspondent threshold is not rejected.

However, for this assumption of Poisson distribution to be correct, the extreme events should be independent, and thus it is necessary to proceed through a declustering (see section 2.6) when dealing with time series to make sure that independence between events is preserved.

3.4 Rules of Thumb

Another procedure consists in choosing one of the sample points as a threshold: the choice is practically equivalent to estimation of the k th upper order statistic X_{n-k+1} from the ordered sequence $X_{(1)}, \dots, X_{(n)}$. Frequently used is the 90% quantile, but this is inappropriate from a theoretical point of view (Scarrott and MacDonald, 2012). Other rules can be used, such as the square root $k = \sqrt{n}$ or the rule $k = n^{2/3}/\log(\log(n))$.

3.5 A Multiple-Threshold Model

The selection of a threshold from the parameter stability plot is problematic: the estimates at two different thresholds are strongly dependent. What needs to be tested is the null hypothesis $H_0: \xi_u = \xi_{u_0}$ for all $u \leq u_0$ for some u_0 , where ξ_u is the value of the shape parameter at threshold u . What is presented here is a discretized version of this hypothesis which can be tested more objectively.

Wadsworth and Tawn (2012) developed a model in which the shape parameter is modelled as a piecewise constant function of the threshold. In doing so they suppose that they can make the approximation above a lower threshold $v < u$. The jump point in this representation will be the threshold u :

$$\xi(x) = \begin{cases} \xi_{vu}, & v < x < u \\ \xi_u, & x > u \end{cases}$$

To attain a better approximation to $\xi(x)$, Northrop and Coleman (2014) extends the piecewise constant representation to an arbitrary number m of thresholds $u_1 < \dots < u_m$:

$$\xi(x) = \begin{cases} \xi_i, & u_i < x < u_{i+1}, \text{ for } i = 1, \dots, m-1 \\ \xi_u, & x > u_m \end{cases} \quad (3.2)$$

Let $v_j = u_j - u_1$ for $j = 1, \dots, m$ and $w_j = u_{j+1} - u_j = v_{j+1} - v_j$ for $j = 1, \dots, m-1$. Let Y denote an exceedance of u_1 . For $j = 1, \dots, m$, the conditional density of $(Y - v_j)|v_j < Y < v_{j+1}$ is given by

$$f_{(Y-v_j)|v_j < Y < v_{j+1}}(y - v_j) = \frac{\mathbf{P}\{Y > v_j\}}{\mathbf{P}\{v_j < Y < v_{j+1}\}} f_j(y - v_j), \quad 0 < y - v_j < w_j,$$

where $f_j(y - v_j) = \frac{1}{\sigma_j} \left[1 + \xi_j \frac{y - v_j}{\sigma_j} \right]^{-(1+1/\xi_j)}$ is a GP density function for $Y - v_j$ with parameters σ_j and ξ_j .

We have

$$\begin{aligned} \frac{\mathbf{P}\{v_j < Y < v_{j+1}\}}{\mathbf{P}\{Y > v_j\}} &= \frac{\mathbf{P}\{Y < v_{j+1}\} \cap \{Y > v_j\}}{\mathbf{P}\{Y > v_j\}} \\ &= \mathbf{P}\{Y < v_{j+1} | Y > v_j\} \\ &= \mathbf{P}\{Y - v_j < w_j | Y > v_j\} \\ &= \int_{v_j}^{w_j+v_j} f_j(t - v_j) dt \\ &= F_j(w_j), \end{aligned}$$

where F_j is the cumulative distribution function of the GPD with parameters σ_j and ξ_j .

Let $p_j = \mathbf{P}\{Y > v_j\}$. Then $p_1 = 1$ and

$$\begin{aligned} 1 - p_j = \mathbf{P}\{Y \leq v_j\} &= \mathbf{P}\left\{ \bigcup_{i=1}^{j-1} \{Y < v_{i+1} | Y > v_i\} \right\} \\ &= \mathbf{P}\left\{ \bigcup_{i=1}^{j-1} \{Y - v_i < w_i | Y > v_i\} \right\} \end{aligned}$$

Thus,

$$\begin{aligned} p_j &= \mathbf{P}\left\{ \left(\bigcup_{i=1}^{j-1} \{Y - v_i < w_i | Y > v_i\} \right)^c \right\} \\ &= \mathbf{P}\left\{ \bigcap_{i=1}^{j-1} \{Y - v_i < w_i | Y > v_i\}^c \right\} \\ &= \prod_{i=1}^{j-1} (1 - F_i(w_i)) \\ &= \prod_{i=1}^{j-1} \left[1 + \frac{\xi_i w_i}{\sigma_i} \right]^{-1/\xi_i}, \quad j = 2, \dots, m, \end{aligned}$$

where A^c denotes the complement of the set A .

Also, $P\{v_j < Y < v_{j+1}\} = p_j - p_{j+1} = p_j F_j(w_j)$ and an equivalent formulation is

$$\begin{aligned} f(y) &= \prod_{j=1}^m \{P\{v_j < Y < v_{j+1}\} f(Y - v_j | v_j < Y < v_{j+1})\}^{I_j} \\ &= \prod_{j=1}^m \{p_j F_j(w_j) f(Y - v_j | v_j < Y < v_{j+1})\}^{I_j} \\ &= \prod_{j=1}^m \{p_j f_j(y)\}^{I_j}, \end{aligned} \quad (3.3)$$

where $I_j = I(v_j < y < v_{j+1})$ and $I(A)$ is the indicator function of the set A .

The shape parameter is then modelled as a piecewise constant function $\xi(y)$ with jump points at v_j , $j = 2, \dots, m$. To avoid any discontinuity in $f(y)$, the scale parameter is set at $\sigma_{j+1} = \sigma_j + \xi_j w_j$ for $j = 1, \dots, m-1$, so that $\sigma_j = \sigma_1 + \sum_{i=1}^{j-1} \xi_i w_i$. The parameters of the model are $\boldsymbol{\theta} = (\sigma_1, \xi_1, \dots, \xi_m)$.

Now, let $\mathbf{y} = (y_1, \dots, y_n)$ be n exceedances of threshold u_1 from density (3.3). The likelihood function is

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \boldsymbol{\theta}) &= \prod_{i=1}^n \prod_{j=1}^m \{p_j f_j(y_i)\}^{I_{ij}} \\ &= \prod_{i=1}^n \prod_{j=1}^m \left\{ \frac{p_j}{\sigma_j} \left(1 + \xi_j \frac{y_i}{\sigma_j}\right)^{-1+1/\xi_j} \right\}^{I_{ij}}, \end{aligned}$$

where $I_{ij} = I(v_j < y_i < v_{j+1})$.

The log-likelihood function is then

$$l(\mathbf{y}; \boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^m I_{ij} \left\{ \log p_j - \log \sigma_j - \left(1 + \frac{1}{\xi_j}\right) \log \left[1 + \xi_j \frac{y_i}{\sigma_j}\right] \right\} \quad (3.4)$$

Consider threshold u_1 . The idea is to test whether a common GP model can be fitted on all intervals $(v_k; v_{k+1})$, $k = 1, \dots, m$, that is to say, to test $H_0 : \xi_1 = \dots = \xi_m$. If H_0 is rejected, then a threshold greater than u_1 should be chosen.

Let us denote by $\tilde{\sigma}_1$ and $\tilde{\xi}_1$ the restricted MLEs of σ_1 and ξ_1 under the null hypothesis, and $\hat{\sigma}_1, \hat{\xi}_j$, $j = 1, \dots, m$ the unrestricted MLEs. Northrop and Coleman (2014) suggest a reparametrization $\boldsymbol{\theta} = (\sigma_1, \phi_1, \dots, \phi_m)$, where $\phi_j = \xi_j / \sigma_j$, $j = 1, \dots, m$. Let $\widehat{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}$ denote the restricted and unrestricted MLEs of $\boldsymbol{\theta}$ respectively.

Consider now the likelihood ratio

$$W = 2 \left\{ l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_0) \right\}$$

and the score statistic

$$S = U(\widehat{\boldsymbol{\theta}}_0)^T i^{-1}(\widehat{\boldsymbol{\theta}}_0) U(\widehat{\boldsymbol{\theta}}_0),$$

where $U(\boldsymbol{\theta})$ is the score function ($U_i = \partial l(\boldsymbol{\theta})/\partial \theta_i$), $i(\boldsymbol{\theta})$ is the expected information matrix ($i(\boldsymbol{\theta})_{ij} = \mathbf{E} \{ \partial^2 l(\boldsymbol{\theta})/\partial \theta_i \partial \theta_j \}$) and A^T is the transpose of the matrix A . A derivation of these expressions is given in Northrop and Coleman (2014). Provided that $\xi_m > -1/2$ (Smith, 1985), the asymptotic null distribution of W and S is χ_{m-1}^2 .

The statistic S has the advantage over W that it requires only a fit of the null model at the threshold of interest, while the calculation of W needs m shape parameters to be estimated, which is more time-consuming.

Suppose now that the lowest threshold considered is u_i for some $i = 1, \dots, m-1$, so that the set of thresholds is (u_i, \dots, u_m) . The null hypothesis becomes $H_0^{(i)} : \xi_i = \dots = \xi_m$ and the asymptotic null distribution of W and S is χ_{m-i}^2 .

p -values of the tests can be calculated and we can use them to select the threshold. A small p -value suggests that $H_0^{(i)}$ is not true, in other words, that u_i is not high enough, while a large p -value suggests that u_i might be high enough. For instance, if we set the confidence rate at 5%, H_0 is rejected whenever the p -value is smaller than 0.05.

The choice of threshold is then made as follows: we perform tests with lowest thresholds u_1, \dots, u_{m-1} and we select the lowest threshold u_k with the property that the null hypothesis is not rejected at it and at all the higher thresholds considered (u_{k+1}, \dots, u_m).

4 Non-Stationary Extremes

The POT method explained previously is only valid for extremes for which we can assume stationarity. The idea is now to describe non-stationarity, as it is the case for temperature series (such a series shows a strong seasonal pattern, see figure 6.2), by allowing the GPD parameters to depend on time or other covariates.

In the literature, various approaches have been developed to deal with non-stationarity arising from seasonality.

4.1 Trend Variation

To check for a linear trend, we use a model \mathcal{M}_1 with the following form for $\sigma(t)$ in the GPD:

$$\sigma(t) = \exp(\alpha + \beta t)$$

The exponential is used to ensure the non-negativity of $\sigma(t)$.

4.2 Model Choice

In order to see whether the non-stationary model is worthwhile, that is to say, if the non-stationary model provides an improvement over the simple model, we use the deviance statistic: with models $\mathcal{M}_0 \subset \mathcal{M}_1$, the deviance statistic is defined as

$$D = 2 \{l_1(\mathcal{M}_1) - l_0(\mathcal{M}_0)\} \tag{4.1}$$

where $l_1(\mathcal{M}_1)$ and $l_0(\mathcal{M}_0)$ are the maximized log-likelihoods under models \mathcal{M}_1 and \mathcal{M}_0 respectively.

The asymptotic distribution of D is given by the χ_k^2 distribution with k degrees of freedom, where k is the difference in dimensionality of \mathcal{M}_1 and \mathcal{M}_0 . Thus, calculated values of D can be compared to critical values from χ_k^2 , and large values of D suggest that model \mathcal{M}_1 explains more of the variation in the data than \mathcal{M}_0 .

For example, to test for existence of trend in σ , that is whether $\beta = 0$, twice the difference in log-likelihood for a model containing β and one not containing β would be compared to a quantile from the χ_1^2 distribution.

5 Extremal Mixture Models

Traditionally, the threshold is selected before fitting (see section 3 for different means of selection), giving the so-called *fixed threshold approach*. The disadvantage of these fixed threshold approaches is that once the threshold has been chosen it is treated as fixed, so the associated subjectivity and/or uncertainty is ignored in subsequent inference (Scarrott and MacDonald, 2012). To overcome this problem, the idea is to compare the GPD parameters or likelihood at different choices of threshold. However, a different threshold means a different size of the sample. In order to avoid this, extremal mixture models have been developed, which can estimate the entire distribution function. Then the sample size is fixed for each threshold considered.

A literature review on the different mixture models available, with their advantages and disadvantages, is provided in Hu (2013). In the present study, we will focus only on the models implemented in the R package `evmix` created by Hu and Scarrott (2013).

The mixture models have been classified by the type of bulk distribution (distribution of the data below the threshold) models: parametric, semi-parametric and non-parametric.

5.1 Parametric Mixture Models

One of the simplest extreme value mixture models has been developed by Behrens et al. (2004). It combines a parametric form for the bulk distribution and a GPD for the distribution above the threshold. The threshold is considered as a parameter to be estimated and the two distributions are spliced at this point. The bulk distribution can be any parametric distribution, such as normal, Weibull, gamma, log-normal, beta.

Zhao et al. (2010) developed a two tail mixture model, where the bulk is described by a normal distribution and the lower and upper tails are described by a GPD. One of the drawbacks of this model is that it has eight parameters to be estimated.

Frigessi et al. (2002) suggested a dynamically weighted model which uses the Weibull distribution to describe the bulk and a GPD for the tail. It uses a continuous transition function to connect the bulk to the tail model: the mixing function is a Cauchy distribution. That makes the mixture density continuous. This model is designed for positive data set

only and the full data set is used for the inference to determine the GPD parameters.

Carreau and Bengio (2009) proposed an hybrid Pareto model by splicing a normal distribution with a GPD and set continuity constraint on the density and on its first derivative at the threshold. The five parameters are then reduced to three. However, this model performs very poorly in practice (Scarrott and MacDonald, 2012), because these two constraints are rather strong.

5.2 Semi-parametric Mixture Models

do Nascimento et al. (2011) considered a semi-parametric Bayesian approach by including a weighted mixture of gamma densities for the bulk distribution and GPD as the tail distribution. The main advantage of such a model over the parametric ones is that it provides more flexibility for the bulk model. Here again, the data set should be restricted to positive values.

5.3 Non-parametric Mixture Models

MacDonald et al. (2011) developed a non-parametric kernel density estimator. Kernel density estimation is a method of estimating the probability distribution of a random variable based on a random sample. In contrast to a histogram, kernel density estimation produces a smooth estimate. More information about kernel density estimation can be found in Wand and Jones (1995) or Silverman (1986). So, here, the observations below the threshold are assumed to follow a non-parametric density, and the observations above the threshold are assumed to follow a GPD.

MacDonald et al. (2011) also introduced a two-tailed mixture model, constructed by splicing the standard kernel density estimator with two extreme value tail models.

The main advantage of the non-parametric approach is that the bulk fit is robust to the tail fit, but a major drawback of this method is the computational complexity.

6 Application: Temperature Extremes in Uppsala

6.1 Dataset

The data we propose to work with is the long series of temperature observations in Uppsala, Sweden ($59^{\circ}51'19''\text{N}$ / $17^{\circ}37'55''\text{E}$), during the time period 1840-2012. Before 1839, two observations of the temperature were made per day, one around sunrise and one during the afternoon. After 1839, the invention of a max-and-min thermometer allowed the recording of daily minima and maxima, which makes the data more reliable. Then we have two values per day, a minimum and a maximum. A complete description of this data (historical background, data analysis) can be found in Bergström and Moberg (2002).

We will first perform our analysis on the daily maxima series, and then on the daily minima series. A box plot of the monthly temperatures has been performed, showing the variability of the temperatures for each month (see figure 6.1).

In the box plot, we can see that the variability of maxima is varying over the year: there seems to be more variability in the winter months (for example the difference between the highest and the smallest maximum in January is 35.3°C , and "only" 24.4°C in September). It might then be more difficult to estimate extremes for the months where the spreading is large, so our examples will be essentially presented through results for the spring or summer months, choosing the ones that are the most relevant.

In order to be able to apply the POT method, it has been seen previously that the observations need to be independent identically distributed. However, one can observe that the temperatures follow a seasonal pattern: the temperature increases from January until it reaches a peak during summer and then decreases until it reaches a minimum at the end of the year (during winter). As an example, we have plotted the daily maxima for the years 1850 and 1851 (see figure 6.2).

To avoid these seasonal variations, a first common approach in an analysis is to consider different series for each month of the year. In our analysis, the temperatures of January from 1840 to 2012, the temperatures of February from 1840 to 2012, etc. A small trend (an increase in temperature through the years) could be observed for certain months, but it was not taken into account. Then, the selection of a fixed threshold for each month was

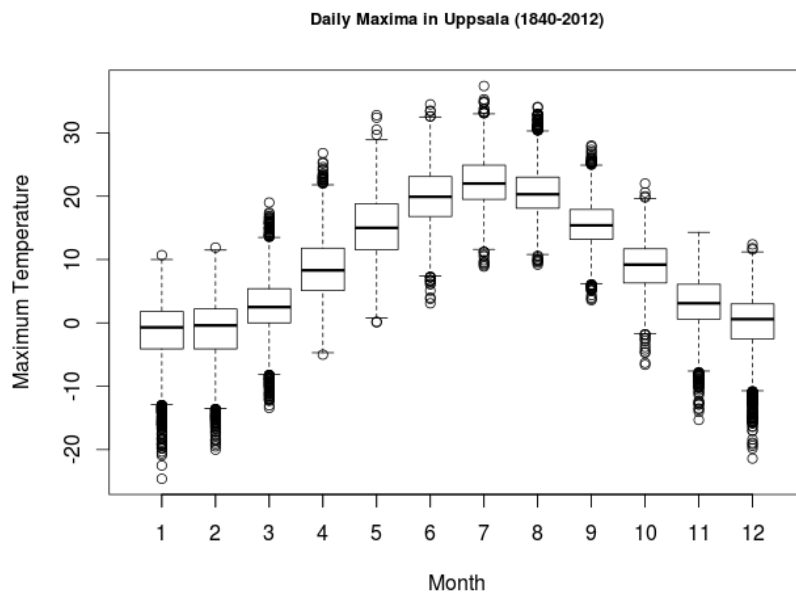


Figure 6.1: Box plot

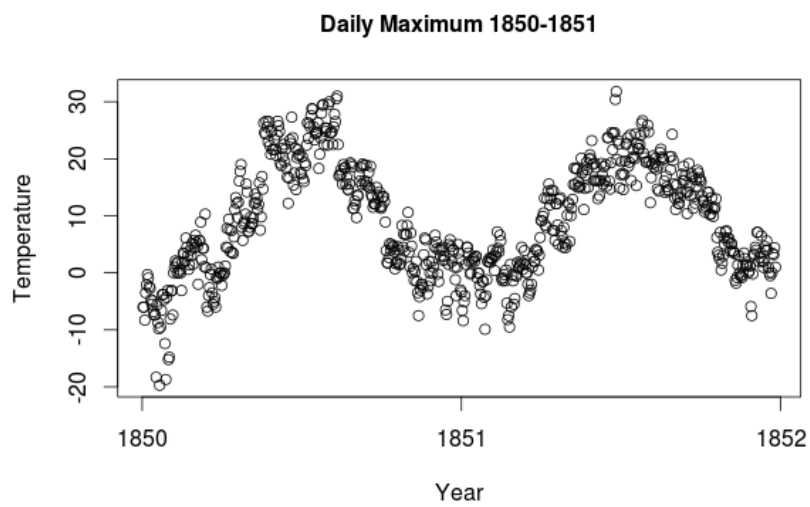


Figure 6.2: Maximum daily temperature records from January 1850 to December 1851

made possible.

Month	90% quantile	$k = \sqrt{n}$	$k = n^{2/3}/\log(\log(n))$
January	4.0	7.2	3.5
February	4.6	7.8	4.1
March	8.5	12.9	7.7
April	15.3	20.2	14.5
May	22.0	26.3	21.4
June	26.0	29.5	25.4
July	27.7	31.3	27.1
August	25.8	29.9	25.2
September	20.3	24.0	19.8
October	13.7	16.8	13.3
November	8.0	11.1	7.7
December	5.0	8.1	4.7

Table 6.1: Monthly threshold calculated with different rules (see section 3.4)

6.2 Choice of Threshold

The choice of threshold is very crucial when one wants to apply the POT method and estimate the GPD parameters. Indeed the threshold should not be too high in order to have enough data to deal with, but neither too low. All the different methods presented in section 3 have been used, and it is then possible to compare the different values of thresholds obtained.

First, the three rules introduced in section 3.4 have been applied to each month series. Table 6.1 summarizes the values of the thresholds. We can see that the rule $k = \sqrt{n}$ gives relatively high thresholds compared to the commonly used rule of the 90% quantile. On the contrary, the rule $k = n^{2/3}/\log(\log(n))$ gives lower but almost the same threshold than the 90% quantile.

Mean residual life plots have been performed but they were difficult to interpret. It was quite hard to choose a threshold from such a plot, as the graph is approximately linear from a very small threshold (see figure 6.3 for an example, for the June daily maxima).

The parameter stability plots involves plotting the modified scale parameter and the shape parameter against the threshold u for a range of threshold which has been chosen to go from the 80% quantile to the 99% quantile. The parameter estimates should be stable (i.e. constant) above the threshold at which the GPD model becomes valid. In practise, the mean excess values and parameter estimates are calculated from relatively small amounts of data, so the plots will not look either linear or constant even when the GPD model becomes valid. Confidence intervals are included, so that we can evaluate whether the plots look linear or constant once we have accounted for the effects of estimation uncertainty.



Figure 6.3: Mean Residual Life Plot for the June maxima series

This is not easy to do by eye, and the interpretation of these plots often requires a good deal of subjective judgement. An example, for the August daily maxima, is given in figure 6.4.

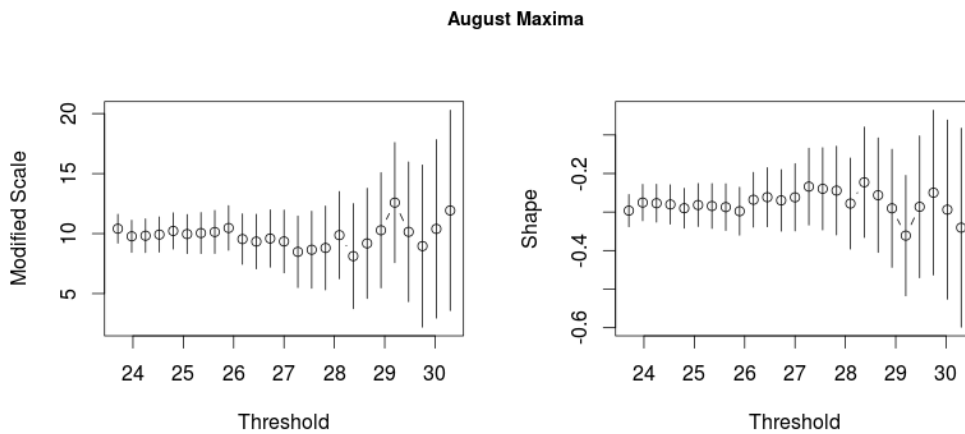


Figure 6.4: Parameter stability plot for a GPD fit to the August maxima series

Another approach which has been considered for the threshold selection is the dispersion index plot, since it is particularly useful when dealing with time series. As an example, the dispersion index plot has been performed for the August maxima data (see figure 6.5). As it is a highly correlated time series, extreme events have been extracted

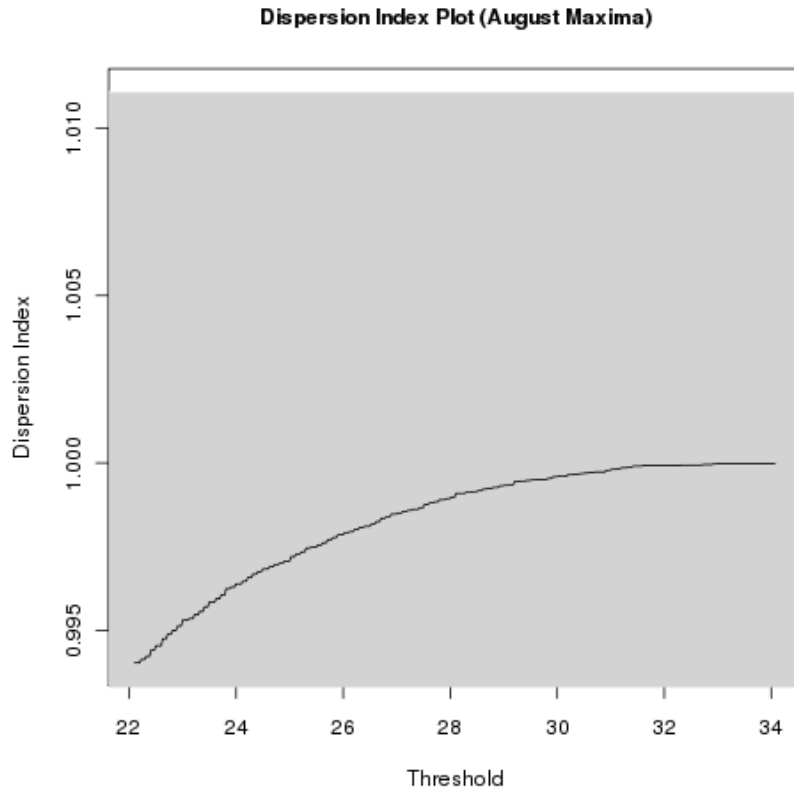


Figure 6.5: Dispersion index plot for the August maxima series

using the declustering method with a relatively low threshold (less than the 70% quantile). Indeed, the dispersion index plot is a mean for the selection of a threshold, but the declustering of the data already needs the choice of a threshold. Then it was quite unclear how to use the dispersion index plot, and which threshold should be used for the declustering.

The dispersion index plot shows the same pattern for every monthly data. It is really smooth and gets closer and closer to one, but it is quite difficult to decide which value of the dispersion index should be chosen for the selection of the threshold. Indeed, the closer the dispersion index gets to one, the higher the threshold. But in our case, we can observe that the value of the dispersion index is already very close to 1 for a relatively low threshold (22°C in August).

Then we applied the multiple-threshold diagnostic. To do this, we used $m = 18$ thresholds from the 85% quantile for each month and we increased the thresholds in steps of 0.25 (for example, the set of thresholds for January is $(u_1, \dots, u_{18}) = (3.10, 3.35, \dots, 7.35)$). We calculated the p -values using the likelihood ratio test and the score test (see section 3.5) based on the χ_{m-i}^2 null distribution, $i = 1, \dots, m - 1$. Both the likelihood ratio test and

Month	Threshold
January	4.85
February	7.45
March	8.35
April	16.65
May	24.3
June	27.15
July	29.1
August	28.7
September	21.65
October	15.65
November	10.05
December	4.45

Table 6.2: Monthly threshold selected from the multiple threshold diagnostic plot

the score test gave the same value for the lowest threshold to be selected, and the results are presented in table 6.2. An example (for the April daily maxima) of the plot obtained is given in figure 6.6. The plain circles represent the p -values for the test using the score test statistic S and the triangles represent the p -values for test involving the likelihood ratio W .

We can observe that the threshold selected with this method are higher than the 90% quantile. However, with this method it was a lot easier to select the threshold than with all the graphical methods presented previously. Indeed it requires less subjectivity and experience to detect the good value of a threshold, since one only has to select the value of the threshold at which all the p -values for the higher thresholds considered is more than 0.05.

6.3 Declustering

As we have seen previously, it is necessary for the temperature data to perform a declustering in order to estimate the parameters σ and ξ . A run period of five days was chosen, and the threshold was set to the 90% quantile, as it seems to be quite accurate. The estimates of the parameters are shown in table 6.3.

We can observe that the shape parameter ξ , which defines the behaviour of the tail of the GPD, shows relatively little variation. Its value is always negative, considering both the non-declustered and declustered data, implying strong evidence for an unbounded tail in each month. For the non-declustered data, its value is comprised between -0.299 and -0.167; for the declustered data, its value is comprised between -0.422 and -0.203. The value of the shape parameter is higher when the data has not been declustered yet.

As for the scale parameter σ , which measures the scale or "variability" of the distri-

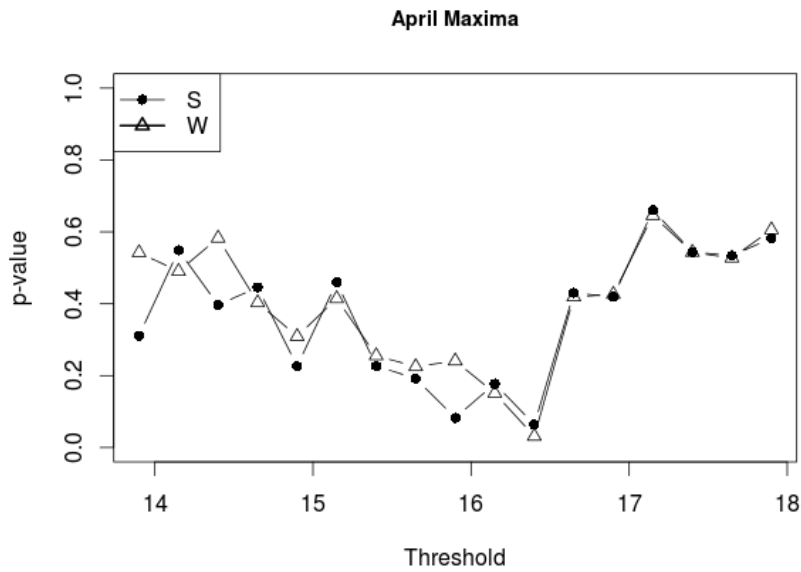


Figure 6.6: Multiple threshold diagnostic plot for the threshold set $(13.90, \dots, 18.15)$, to determine a threshold for the April daily maxima series

Estimate	Without declustering		With declustering	
	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\xi}$
January	1.829	-0.187	2.498	-0.305
February	2.067	-0.207	2.809	-0.330
March	2.940	-0.238	3.936	-0.332
April	3.152	-0.235	4.072	-0.309
May	2.748	-0.222	3.454	-0.292
June	2.372	-0.247	3.179	-0.337
July	2.141	-0.178	2.537	-0.203
August	2.792	-0.299	3.563	-0.397
September	2.396	-0.270	3.519	-0.422
October	1.789	-0.167	2.573	-0.265
November	1.875	-0.252	2.525	-0.359
December	2.074	-0.234	2.611	-0.308

Table 6.3: Parameter estimates $\hat{\sigma}$ and $\hat{\xi}$ without and with declustering, with threshold set at the 90% quantile

Estimate	Model \mathcal{M}_0		Model \mathcal{M}_1			Deviance D
	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\xi}$	
January	1.829	-0.187	0.419	6.461×10^{-5}	-0.216	7.887
February	2.067	-0.207	0.430	1.009×10^{-4}	-0.204	16.094
March	2.939	-0.238	0.818	1.031×10^{-4}	-0.306	24.171
April	3.152	-0.235	1.070	3.181×10^{-5}	-0.249	2.357
May	2.748	-0.222	1.341	-8.815×10^{-5}	-0.328	18.700
June	2.371	-0.247	0.974	-4.741×10^{-5}	-0.241	3.674
July	2.142	-0.178	0.836	-2.752×10^{-5}	-0.180	1.289
August	2.791	-0.299	1.057	-1.424×10^{-5}	-0.291	0.602
September	2.397	-0.271	0.919	-2.131×10^{-5}	-0.266	1.309
October	1.789	-0.167	0.598	-6.297×10^{-6}	-0.165	0.078
November	1.875	-0.252	0.483	4.632×10^{-5}	-0.254	5.260
December	2.074	-0.234	0.492	8.394×10^{-5}	-0.264	15.773

Table 6.4: Comparison of models 0 and 1 to see if there is a linear trend in σ

bution, it also shows little variation, varying from 1.789 to 3.152 for the non-declustered data and from 2.525 to 4.072 for the declustered data. The value of the scale parameter is higher when the data has been declustered.

6.4 Non-stationarity: Modelling

We want to see if the trend observed is significant. To do so, we compare two models, one simple model with parameters fixed over time, and one non-stationary model allowing the scale parameter σ to vary over time. Then the two models considered are as follows:

- Model \mathcal{M}_0 : σ and ξ do not depend on time
- Model \mathcal{M}_1 : $\sigma(t) = \exp(\alpha + \beta t)$ but ξ does not depend on time

The threshold chosen for each month series was set to the 90% quantile.

The deviance statistic D is used to test model 1 against model 0: if $D > \chi_1^2(0.05) = 3.841$ (for a 95% level of significance), model 1 provides a significant improvement in fit over model 0 and there is a linear trend in the parameter σ . The results in table 6.4 show that there is no linear trend in the log-scale parameter for the following months: April, June, July, August, September and October; in other words, there is a trend in the log-scale parameter for the winter months.

6.5 Extremal Mixture Models

These models seem interesting in the sense that they could give us another mean of selecting the threshold: the threshold being a parameter to be estimated, the values of

Estimate	Model 1		Model 2	
	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\xi}$
January	2.498	-0.305	2.045	-0.248
February	2.809	-0.330	1.588	-0.236
March	3.936	-0.332	4.027	-0.336
April	4.072	-0.309	3.483	-0.283
May	3.454	-0.292	2.242	-0.204
June	3.179	-0.337	2.600	-0.299
July	2.537	-0.203	2.337	-0.218
August	3.563	-0.397	2.082	-0.315
September	3.519	-0.422	2.922	-0.418
October	2.573	-0.265	1.745	-0.184
November	2.525	-0.359	1.970	-0.424
December	2.611	-0.308	2.809	-0.311

Table 6.5: Comparison of the estimates of the GPD parameters, for different thresholds

thresholds obtained by these methods could be different than the ones we got previously (by diagnostic plots or rules of thumb).

We tried to apply several models to our data, starting with the one developed by Behrens et al. (2004). In this attempt, we tried to compare the fit of a gamma distribution for the bulk and a GPD for the tail to the fit of a Weibull distribution for the bulk and a GPD for the tail.

However, in each case, the threshold obtained was very close to the 90% quantile. This is due to the fact that the 90% quantile is used as an initial value for the optimisation. So if there is a local mode around this value, then the optimisation will get stuck.

To avoid this, the idea was to use the profile likelihood approach to maximise the likelihood over the threshold first (by setting a range of thresholds to be considered). However, the threshold which maximized the likelihood gave a proportion of data above the threshold of more than 30%, which is too high.

We tried almost all the models present in the `evmix` package, and the results were always the same: the best threshold was very close to the 90% quantile. Thus these results were unconvincing and we did not go deeper into this analysis.

6.6 Estimation of the GPD Parameters

The POT series are created by declustering peaks above the 90% quantile (defined as model 1) and above the threshold selected by the multiple-threshold diagnostic plot (defined as model 2) (see table 6.2). The estimates of the GPD parameters are displayed in table 6.5.

We can see that the shape parameter is always negative, whatever threshold is chosen. Moreover, the shape parameter seems to be smaller (or greater in absolute value) when the threshold is set to the 90% quantile, except for certain months: March, July, December

(but the difference is very small) and November (where the difference is more significant). So the lower the threshold, the smaller the shape parameter.

The scale parameter seems to be much larger when the threshold is set to the 90% quantile, except for March and December, so the lower the threshold, the larger the scale parameter.

6.7 Return Levels and Return Periods

Table 6.6 shows the return levels for two different thresholds: u_1 set at the 90% quantile and u_2 set at the threshold determined by the multiple-threshold diagnostic plot (see table 6.6). The return levels were calculated for different return periods of 5, 10, 50, 100 and 1000 years.

	u_1					Maximum	u_2				
	x_5	x_{10}	x_{50}	x_{100}	x_{1000}		x_5	x_{10}	x_{50}	x_{100}	x_{1000}
January	10.51	10.83	11.36	11.52	11.86	10.7	10.53	10.93	11.64	11.87	12.40
February	11.56	11.88	12.38	12.53	12.84	11.9	11.29	11.72	12.50	12.75	13.34
March	18.23	18.67	19.37	19.57	20.00	19	18.29	18.72	19.39	19.59	19.99
April	25.82	26.33	27.17	27.42	27.95	26.8	25.77	26.34	27.29	27.59	28.24
May	31.25	31.73	32.52	32.76	33.29	32.8	30.77	31.36	32.46	32.83	33.74
June	33.79	34.14	34.68	34.84	35.16	34.5	33.71	34.11	34.77	34.97	35.40
July	35.83	36.40	37.46	37.81	38.70	37.4	35.72	36.30	37.34	37.68	38.52
August	33.63	33.90	34.32	34.43	34.64	34.1	33.33	33.72	34.35	34.54	34.94
September	27.70	27.94	28.28	28.37	28.54	28	27.60	27.86	28.24	28.35	28.53
October	20.97	21.38	22.09	22.31	22.82	22	20.58	21.12	22.15	22.50	23.41
November	13.95	14.19	14.56	14.67	14.88	14.3	13.80	14.03	14.36	14.45	14.60
December	11.77	12.10	12.64	12.80	13.15	12.4	11.87	12.18	12.70	12.85	13.18

Table 6.6: Comparison of return levels, for different thresholds

	Location μ	Scale σ	Shape ξ
January	-2.841	5.024	-0.370
February	-2.751	5.139	-0.349
March	0.922	4.591	-0.248
April	6.694	4.590	-0.184
May	13.218	5.160	-0.255
June	18.266	4.647	-0.279
July	20.718	3.866	-0.222
August	19.140	3.566	-0.184
September	14.201	3.560	-0.237
October	7.539	3.906	-0.268
November	1.749	4.196	-0.332
December	-1.539	4.699	-0.336

Table 6.7: Estimation of the GEV parameters

The results are shown in table 6.6. We can see that the return levels estimated are almost the same whatever threshold we chose. It is also interesting to note that in most cases (all months but two) the maximum of daily maxima is comprised between x_{10} , the return level corresponding to the return period of 10 years, and x_{50} , the return level corresponding to the return period of 50 years. Having a deeper look at the data, we can see that the maximum is reached only once between 1840 and 2012 (a 172-year time period), and the second maximum is much lower than the maximum (and this for all months). This means that the estimation of the return values is slightly overestimated.

6.8 Comparison with the Block Maxima Method

Another approach, the block maxima method, was applied to compare the results from the estimation of the GEV parameters with those from the GPD previously found. We adapted this method to our twelve time series (the maxima temperatures for each month) in the sense that the blocks we built consist of the temperatures of one month for each year of the time period 1840-2012. This gives us 172 blocks of size the length of the month (number of days in the month), so the GEV function will be estimated with 172 values for each month.

The results are given in table 6.7. We can see that the scale parameter is much higher in the estimation of the GEV parameters than in the GPD case, and the shape parameter is more negative in the GEV case than in the GPD. This means that the curve of return levels vs. return periods (with a logarithmic scale for the return periods) will be more concave than in the GPD case. This leads to significantly higher return period estimates for high temperatures with the GEV model.

Month	u_1	u_2
January	16.2	17.25
February	17.3	15.10
March	13.6	13.95
April	5.5	7.80
May	1.4	3.75
June	-3.9	2.00
July	-7.6	-4.55
August	-6.1	-7.00
September	-1.2	1.15
October	3.1	4.45
November	9.1	9.85
December	14.3	13.85

Table 6.8: Monthly threshold for the monthly series of the negated daily minima in Uppsala

6.9 Analysis of the Daily Minima in Uppsala

In this section, we will perform the same analysis on the daily minimum temperatures of Uppsala. The temperature series has also been cut into monthly series, and the temperatures have been negated in order to be able to apply the POT method.

The choice of threshold has been made only by means of the multiple-threshold diagnostic plot (see section 3.5) and determination of the 90% quantile. Indeed, the other graphical methods (mean residual life plot, parameter stability plot and dispersion index plot) require too much subjectivity and it is hard to determine a threshold from the plots. The thresholds for each months are presented in table 6.8. In the table, u_1 is the threshold corresponding to the 90% quantile and u_2 is the threshold determined by the multiple-threshold diagnostic plot. Here again, we used 18 thresholds from the 85% quantile and the step between the thresholds tested was set to 0.25°C. We can observe that the threshold determined by the 90% quantile is most of the time lower than the one determined by the multiple-threshold diagnostic plot (except for the months of February, August and December).

Then, the estimation of the GPD parameters was made. To do this, a declustering of the data was performed, using the same run length of 5 days. Table 6.9 shows the results of the estimation of the GPD parameters.

We can see that the shape parameter is always negative, whatever threshold is chosen, except for the June series when using the threshold u_2 . The shape parameters estimated for both u_1 and u_2 are very close, there is not a great difference between them (less than 0.05 in most cases). We cannot say if one threshold gives a higher or lower shape parameter.

Estimate	Model 1		Model 2	
	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\xi}$
January	5.25	-0.16	5.04	-0.16
February	6.90	-0.47	8.09	-0.48
March	5.77	-0.28	5.43	-0.26
April	3.30	-0.08	2.86	-0.04
May	2.61	-0.30	1.83	-0.29
June	2.65	-0.15	2.24	0.04
July	2.56	-0.25	1.50	-0.09
August	3.06	-0.41	3.49	-0.42
September	3.06	-0.46	1.75	-0.37
October	2.75	-0.07	2.45	-0.02
November	5.68	-0.35	5.18	-0.32
December	6.43	-0.43	6.68	-0.44

Table 6.9: Estimation of the GPD parameters for the negated daily minima in Uppsala, after declustering, for thresholds u_1 and u_2

The scale parameter seems to be much larger when the threshold is set to the 90% quantile, like in the maxima case, except for February, August and December (the same months where the threshold u_1 was larger than the threshold u_2), so the lower the threshold, the larger the scale parameter.

As the aim of extreme value analysis is to determine return levels of given return periods, these values were estimated for the daily minima monthly series too. Table 6.10 shows the return levels for different return periods of 10, 50 and 100 years, for the two different thresholds u_1 and u_2 defined above.

We can see that the return levels estimated are almost the same whatever threshold we chose (the difference is about 0.1 in most cases). Moreover, most of the time, the maximum of the negated daily minima (or minimum) is comprised between x_{10} , the return level corresponding to the return period of 10 years, and x_{100} , the return level corresponding to the return period of 100 years. Having a deeper look at the data, we can observe that the minimum is reached only once between 1840 and 2012 (a 172-year time period) - except in June where the minimum -9°C is reached twice - and the second minimum is much larger than the minimum - except in November where the minimum is -23.5°C and the second minimum is -23.4°C . This means that the calculation return levels slightly overestimates their value.

In the same manner than for the daily maxima, we adapted the GEV model to our data, in order to compare the estimated GEV parameters to those determined with the POT method. The scale parameter calculated in the GEV case was sometimes higher, sometimes lower than in the GPD case, but always quite close to the value in the GPD

	u_1			Maximum	u_2		
	x_{10}	x_{50}	x_{100}		x_{10}	x_{50}	x_{100}
January	36.4	39.4	40.4	39.5	36.2	39.2	40.2
February	31.0	31.5	31.6	30.9	31.1	31.6	31.7
March	30.4	31.8	32.3	32.1	30.3	31.9	32.4
April	20.9	24.0	25.2	22.5	20.5	24.1	25.5
May	8.5	9.1	9.3	8.5	8.4	9.0	9.2
June	6.4	8.0	8.6	9.0	6.0	10.0	11.8
July	0.3	1.1	1.4	1.0	0.2	1.7	2.3
August	0.7	1.0	1.1	0.9	0.7	1.0	1.1
September	5.1	5.3	5.4	5.2	5.0	5.4	5.5
October	16.3	19.0	20.1	15.6	17.0	20.6	22.2
November	23.4	24.3	24.5	23.5	23.4	24.4	24.7
December	27.9	28.5	28.7	28.2	28.0	28.5	28.7

Table 6.10: Return levels of different return periods, for thresholds u_1 and u_2

case. The shape parameter was always negative, but most of the time "less negative" (or smaller in absolute value) than the one estimated in the GPD case. Thus, the curve of return levels vs. return periods (with a logarithmic scale for the return periods) is less concave than in the GPD case. This leads to lower return period estimates for cold temperatures with the GEV model.

7 Conclusion

The aim of this work was to study the series of temperatures of Uppsala by using the peaks-over-threshold method.

A particular interest was given to the choice of the threshold to perform such an analysis. Indeed, the choice of threshold affects the estimation of the GPD parameters. Different techniques were used to help the choice of the threshold: graphical diagnostics, rules of thumbs and statistical tests. Rules of thumb are easy to use since they choose one particular value from the series as the threshold. Graphical diagnostics consist in analysing different plots, but are often very difficult to interpret as they usually require a lot of subjectivity. Finally, a new method using statistical tests was used, which is more demanding computationally, but gave results easier to interpret, as the comparison of p -values is more straightforward. Moreover, this method provided very reasonable thresholds when compared to the frequently used 90% quantile.

Concerning the data, a strong seasonal variation is present in the temperature series, as temperatures fluctuate along the year. To perform our study, we decided to avoid this seasonal variations and cut the data into different series for each month. A small trend was found for certain months, but was not taken into account when estimating the parameters.

Comparing the results of the estimation of the GPD parameters with those of the GEV for the same time series gave different values, which led to different estimations of return periods: for the maxima, estimates of the return periods are larger with the GEV model than with the POT method, while for the minima, the GEV model gives lower return period estimates than the POT method.

For further work, it could be possible to use a time-varying threshold and perform the analysis on the whole time series. This would take into account the effect of seasonality. Another idea would be to perform a deeper analysis using the mixture models.

Bibliography

- C. Anderson, D. Carter, and P. Cotton. Wave climate variability and impact on offshore design extremes. Technical report, March 2001. URL <http://clive-anderson.staff.shef.ac.uk/waves.pdf>.
- C. N. Behrens, H. F. Lopes, and D. Gamerman. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244, 2004.
- H. Bergström and A. Moberg. Daily air temperature and pressure series for Uppsala (1722–1998). In *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources*, pages 213–252. Springer, 2002.
- J. Carreau and Y. Bengio. A hybrid pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76, 2009.
- S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.
- C. Cunnane. A note on the poisson assumption in partial duration series models. *Water Resources Research*, 15(2):489–494, 1979.
- A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442, 1990.
- F. F. do Nascimento, D. Gamerman, and H. F. Lopes. A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, 22(2):661–675, 2011.
- A. Frigessi, O. Haug, and H. Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235, 2002.
- Y. Hu. Extreme value mixture modelling with simulation study and applications in finance and insurance, 2013.
- Y. Hu and C. Scarrott. Extreme value mixture modelling, threshold estimation and boundary corrected kernel density estimation, 2013. URL <http://www.math.canterbury.ac.nz/~c.scarrott/evmix>. Available on CRAN.
- A. MacDonald, C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell. A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157, 2011.

- P. J. Northrop and C. L. Coleman. Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2):289–303, 2014.
- J. Pickands. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.
- R.-D. Reiss, M. Thomas, and R. Reiss. *Statistical analysis of extreme values*, volume 2. Springer, 2001.
- M. A. Ribatet. A user’s guide to the pot package (version 1.4). August 2006. URL <http://cran.r-project.org/>.
- J. Rydén. Statistical analysis of temperature extremes in long-time series from Uppsala. *Theoretical and applied climatology*, 105(1-2):193–197, 2011.
- C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60, 2012.
- B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.
- R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.
- J. Wadsworth and J. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):543–567, 2012.
- M. P. Wand and M. C. Jones. *Kernel smoothing*. Monographs on statistics and applied probability. Chapman & Hall/CRC, Boca Raton (Fla.), London, New York, 1995. ISBN 0-412-55270-1. URL <http://opac.inria.fr/record=b1093728>.
- X. Zhao, C. Scarrott, L. Oxley, and M. Reale. Extreme value modelling for forecasting market crisis impacts. *Applied Financial Economics*, 20(1-2):63–72, 2010.