UNIVERSITE CATHOLIQUE DE LOUVAIN

FACULTE DES SCIENCES

ECOLE DE STATISTIQUE, BIOSTATISTIQUE
ET SCIENCES ACTUARIELLES

Temporal analysis of the evolution of extreme values using climatological data

Promoteur :	Johan Segers
Co-promoteur :	Prénom Nom
Lecteurs :	Anna Kirilliouk
	Michel Crucifix

Mémoire présenté en vue de l'obtention du

Master en statistiques, orientation générale

par : **Antoine Pissoort**

Juin 2017

Contents

List of Figures

Acknowledgements

I would first like to thank my thesis supervisor Johan Segers for all his help and his guidance during this year.

I also would like to thank the "Institut Royal de Météorologie" (IRM) of Belgium for his help, his guidance and his provided quality datasets.

Part I

Theoretical Framework

Introduction and Preliminaries

Unlike his counterparts (see for example credit risk analysis, financial applications,...), the extreme value analysis applied on the broad environemental area like here for the meteorological data, has strong impacts on the people lives

An important question is still whether climate changes caused by anthropogenic activities will change the intensity and frequency of extreme events [**?**].

The problem we are here facing in climate change evidence is that of le lack of past data to compare with her

Also, for such an analysis, the number of parameters to take into account is considerable (and tend to infinity)

Can make a parallelism with Chaos Theory and the well-known butterfly effect which have strong applications in weather models

We highly expect the climate change to affect the extreme weather

[extremes in climate change p.347]

[**?**]

"The first myth about climate extremes, which has been purported by researchers in climatology or hydrology, among them prominent names, is that “extremes are defined as rare events” or similar. This myth is debunked by a simple bimodal PDF (Fig. 6.12a). The events sitting in the tails of that distribution are not rare" [Mud14, pp.257]

Until now, studies on climate extremes that consider Europe have usually had a strong national signature , or have had to make use of either a dataset with daily series from a very sparse network of meteorological stations (e.g. eight stations in Moberg et al. (2000)) or standardized data analysis performed by different researchers in different countries along the lines of agreed methodologies (e.g. Brazdil et al., 1996; Heino et al., 1999) [KTWK$^+$02]

During this project, we will try a novel approach, that is to link directly theory and practice and hence present the concepts theoretically and then illustrate by one example retrieved from our application. Even if that could be difficult, we think this approach is advantageous for several reasons :

Extrapolation !!!! See p154 [statistical analysis of extreme book]

Voir effet de l'ilot de chaleur –> urbanisation sur les tempés !

—> artificial warming on cities stations which were not(less) urbanized 100 years ago.

Presentation of the Analysis : Temperatures from Uccle

Comparisons with freely available data A data is freely available on the internet. (`http://lstat.kuleuven.be/Wiley/Data/ecad00045TX.txt` and which was a project initially performed by the KMNI). However, we were reticent to simply analyze these data as we know that it is hard to trust internet's data, even if they come from well-known "authorities". After having made all these comparisons analysis (see start of code...), we remark effectively that there are differences in these two datasets, and hence large errors of measures can easily occur in unofficial data. It confirms the fact that is important to get reliable data if one wants to make reliable analysis. However, these differences tend to be much smaller when considering the "open shelter" version (54% of equal measurements in closed shelter VS 14.4% in the closed case). For this reason, we have confidence that this public datasets is dealing with open shelter temperatures data.

However, for **meteorological considerations**, it is always better to consider temperature's analysis in **closed shelters**. Indeed, thanks to gratefull advices from mr Tricot working for the IRM :

- It can
- It

First analysis

As expected (line 200 code local), we see that there is an upward trend for both (yearly) maxima and minima. However, we remark that this one is less pronounced for minima. That makes sense as (as mentioned above), the global warming is

Statistical tools

[MOTS DEXPLICATIONS SUR TTES LES FORMULES (domain atraction condition, etc...]

Order statistics

First of all, we write the i-th order statistics $X_{(i)}$ which are the statistics ordered by increasing value

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)} \tag{1.1}$$

We adopt this simpler notation by assuming that the number of our observations will be denoted by n.

One order statistics is of particular interest for our purpose, the maximum

$$X_{(n)} := \max_{1 \leq i \leq n} X_i \tag{1.2}$$

for the minimum, $X_{(1)} = \min_{1 \leq i \leq n} X_i$ that we will still use by converse (see section .)

$$X_{(1)} := \min_{1 \leq i \leq n} X_i = -\max_{1 \leq i \leq n} (-X_i) \tag{1.3}$$

We can retrieve the distribution of our statistic of interest $X_{(n)}$

$$\Pr\{X_{(n)} \leq x\} = \Pr\{X_1, ..., X_n \leq x\} \stackrel{(\perp\!\!\!\perp)}{=} \Pr\{X_1 \leq x\}...\Pr\{X_n \leq x\} = F^n(x) \tag{1.4}$$

(write vertically)

where independence ($\perp\!\!\!\perp$) follows from the iid assumption of the sequence $\{X_i\}$.

Miscellaneous

We also define the *survival* or *survivor* function $\bar{F} = 1 - F$ which is widely useful for this kind of (biostatistical) applications.

Finally, we decide to include in Appendix some concepts of convergence (A.), regularly and slowly varying functions (A.), as they will often appear in the text.

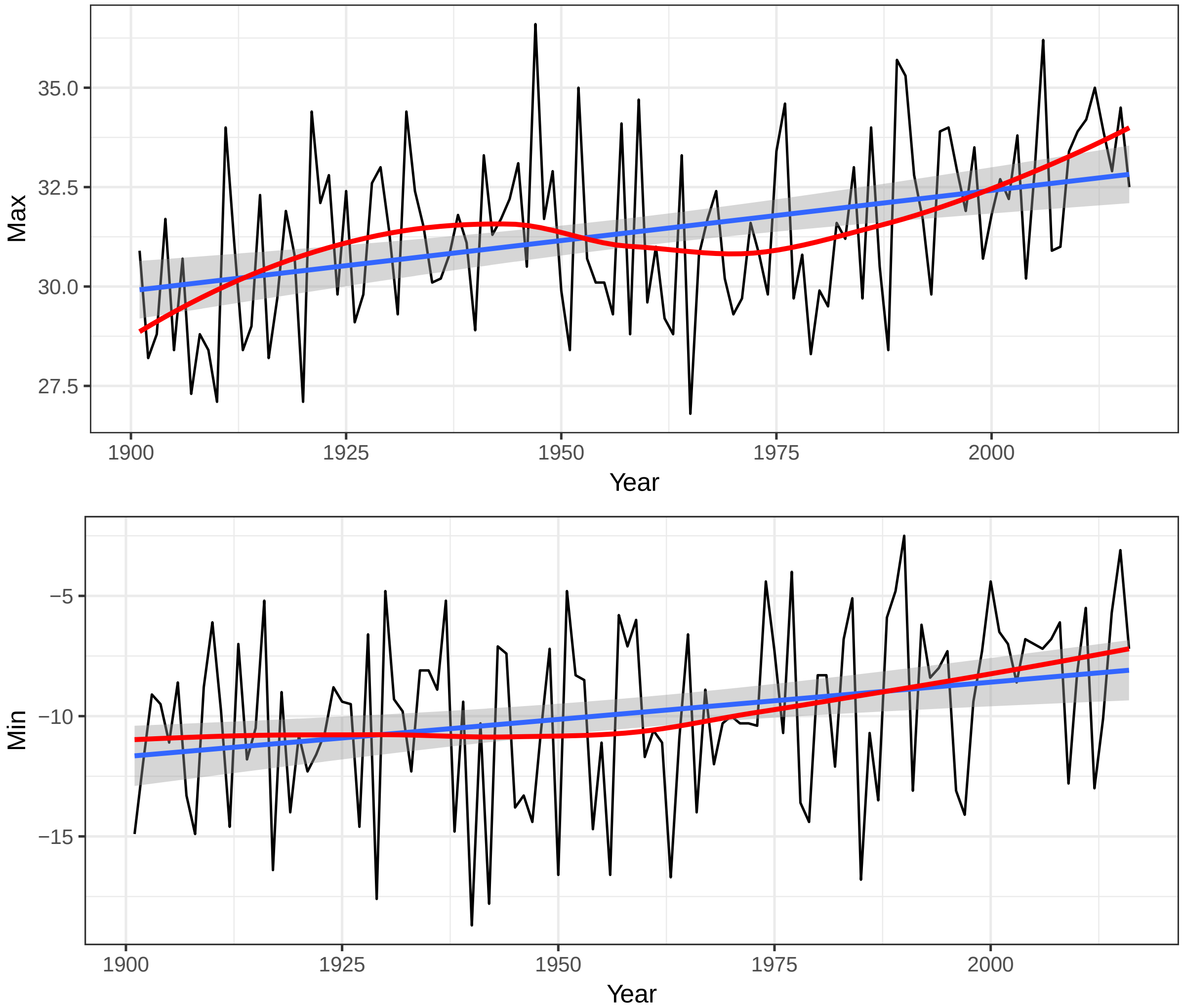

Figure 1.1: First plot representing the **yearly** maxima (above) and minima (below), shaded grey line representing the standard error of the linear trend (regression), red line representing the polynomial nonparametric fit by LOESS.

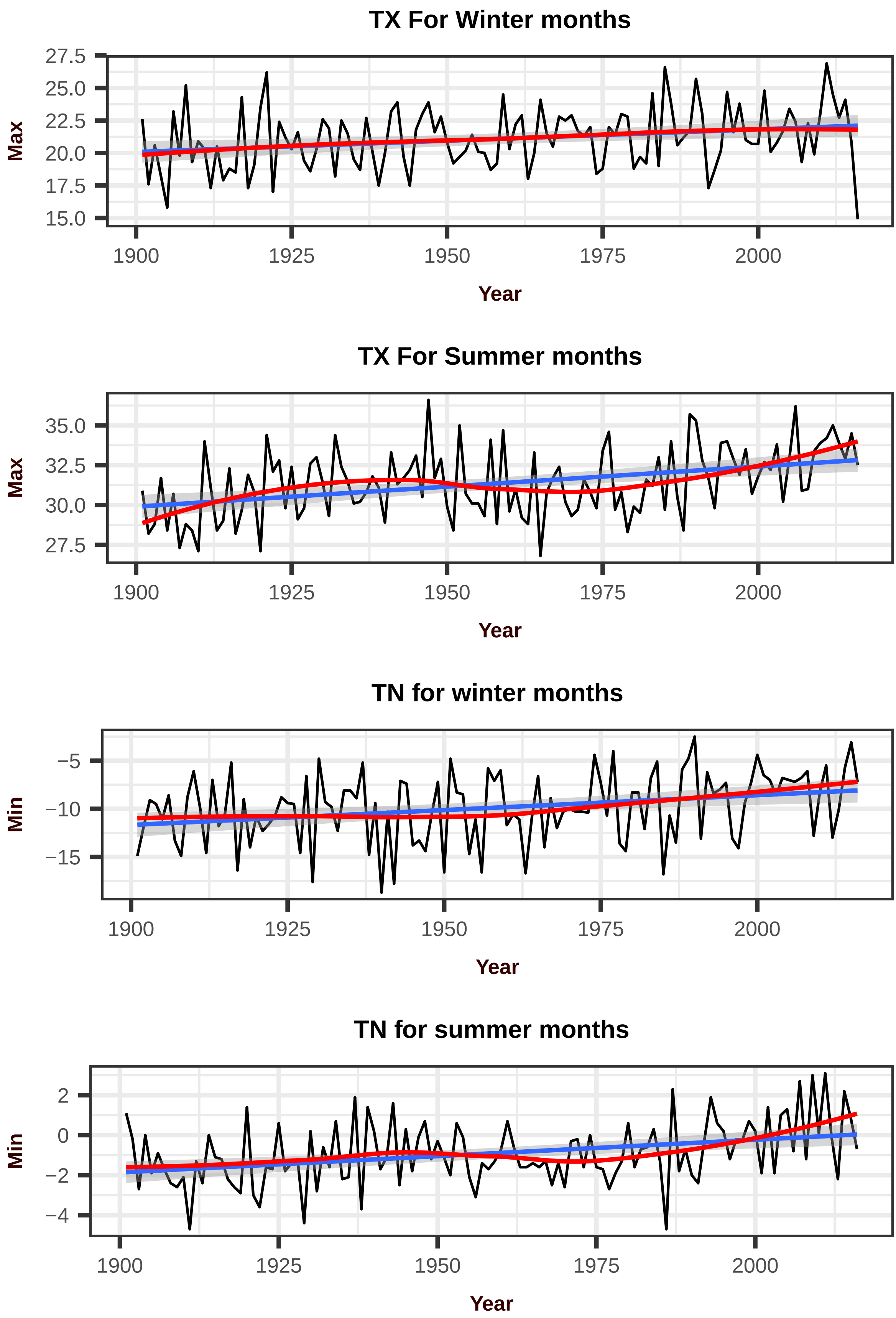

Figure 1.2: First plot representing the **yearly** maxima (above) and minima (below) taking only the summer months (April to September) and winter months (October to March), shaded grey line around the linear trend represents its standard error. See how the polynomial trend (red line) also changes. Obv, TX for smummer and TN for winter are the same series as for the global serie

Extreme Value Theory : Basics

There are two approaches, the block-maxima (section **2.1**) and the peaks-over-threshold approach (section **2.2**) yielding to different extreme value distribution. The former aims at while the latter models the...

Some useful definitions to start with !!

Definition 2.1 (Similar distribution functions)**.** *We say that two distributions functions G and G^* are* **similar** ***or*** *are of the* **same type** *if, for constants $a > 0$ and b*

$$G^*(az+b) = G(z), \qquad \forall x, \tag{2.1}$$

which means that they differ only in location and scale. In the sequel, the concept of *similar* distributions will be useful to derive the three different families of extreme value distributions from other distributions of the *same type.* This is directly linked with max-stable process that we will define...

Definition 2.2 (Max-stability)**.** From [LLR83] or [Res87]*, we say that a distribution G is* **max-stable** *if, for each $n \in \mathbb{N}$*

$$G^n(a_n z + b_n) = G(z), \qquad n = 2, 3, ... \tag{2.2}$$

for some constants $a_n > 0$ and b_n.

In other words, taking powers of G results only in a change of location and scale. **??** This concept will be closely connected with the extremal limit laws in the following ().

Definition 2.3 (Min-stability)**.** *Anageously, from [RT07, pp.23], we say that a distribution function G is* **min-stable** *if*

$$Pr\{X_{(1)} > d_n + c_n z\} = \bar{G}^n(d_n + c_n z) = \bar{G}(z), \tag{2.3}$$

where $c_n = a_n$, $d_n = -b_n$ and $X_{(1)}$ the minimum of the sample of size n, see eq.(1.3).

Principles of stability Behind all the principles about Extreme Value Theory that will be covered during this thesis, will be influenced by the principle of *stability.*

As this will be .. in the following, we think useful to define precisely the concept of *non-degenerate distribution functions.*

Definition 2.4 (Non-degenerate distribution functions)**.** *We say that a distribution function is* **non-degenerate** *if*

We illustrate this by the most common theorem in statistics, the *Central Limit Theorem* (CLT) which plays typically with the empirical mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. We know that (..) $\bar{X}_n$ converges to the true mean μ in probability(?) and thus in distribution, that is to a non-random single point, i.e. to a *degenerate* distribution

$$\Pr\{\bar{X}_n \leq x\} = \begin{cases} 0 & \text{if } x < \mu, \\ 1 & \text{if } x \geq \mu. \end{cases}$$

That is not very useful, in particular for inferential purposes.

For this reason, CLT aims at finding a non-degenerate limiting distribution for $\bar{X}_n$, after allowance for normalization by sequences of constants. We will state it in his most basic form :

Theorem 2.1 (Central Limit Theorem)**.** *Let* $\{X_i\}$ *be a sequence* (or "stochastic process"?!.. Check if written like this is fine ! for the following too) *of n iid random variables with* $E(X_i^2) < \infty$*. Then, as* $n \to \infty$*,*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2),$$

where $\mu = E(X_i)$ *and* $\sigma^2 = V(X) > 0$*. (and d means convergence id distri., see ?)*

Then, by a proper choice of some normalizing constants, μ and $\sqrt{n}$ (as location and scale parameters respectively), we find the non-degenerate Normal distribution in the limit for the empirical mean $\bar{X}_n$.

With the same logic, we find that this is the same for the distribution of $X_{(n)}$

$$\lim_{n\to\infty} \Pr\{X_{(n)} \le x\} = \lim_{n\to\infty} \Pr\{X_i \le x\}^n = \begin{cases} 0 & \text{if} \quad F(x) < 1, \\ 1 & \text{if} \quad F(x) = 1, \end{cases} \tag{2.4}$$

which is also degenerate. This is exactly what Extreme Value Theory aims to achieve for (typically) the maximum order statistics $X_{(n)}$, that is finding a non-degenerate distribution in the limit by means of normalization. This will be the main subject of the next sections.

Block-Maxima

Introduction and extremal types theorem

This theorem, introduced by Fisher and Tippett [FT28], later revised by [Gne43] and finally streamlined by [**?**], is very important for its applications. We remind the distribution of maxima is $\Pr\{X_{(n)} \le x\} = F^n(x)$. It states the following:

Theorem 2.2 (Extremal types theorem)**.** *If the distribution of partial maxima of an independent and identically distributed sequence of random variables with common (unknown) distribution F, say,* $X_{(n)}$*, properly normalized, converges to a non-degenerate (see ...) limiting distribution G, i.e.*

$$\lim_{n\to\infty} Pr\Big\{a_n^{-1}(X_{(n)} - b_n) \le z\Big\} = F^n(a_n z + b_n) = G(z), \qquad \forall x \in \mathbb{R}, \tag{2.5}$$

for some constants $a_n > 0$*,* $b_n \in \mathbb{R}$*.*

[extreme value and cluster ana. of euro... clustering 2011] , meaning that F is said to be in the **domain of attraction**[1] of G, denoted by $F \in D(G)$. This theorem considers an i.i.d. random sample, but it holds true if the original scheme being no longer i.d. still remains independent (we will present the stationary case in section **2.5**). However, even the independent assumption is often poor in practical applications (see for our applications in our case, the temperature....) [**?**] (see application...). with G the *Generalized Extreme Value* (GEV) distribution :

$$G(z) = \exp\left\{ -\left[1 + \xi\Big(\frac{z-\mu}{\sigma}\Big)\right]_+^{-\xi^{-1}} \right\} := G_\xi(z), \tag{2.6}$$

from which we introduce the notation $y_+ = \max(y, 0)$, denoting in the above that $\{z : 1 + \xi\sigma^{-1}(z-\mu) > 0\}$ to ensure the term in the exponential is negative and the distribution

[1] ?? We will more precisely define this concept in the next section.

function converge to 1. We will use this notation in the following. This will define the endpoints of the different distribution functions from the values of the shape parameter, that is $\{\xi > 0;$ $\xi < 0; \xi = 0\}$, more details will be provided in next section . Moreover, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$ with μ, σ and ξ being the three parameters of the model characterizing location, scale and shape respectively. We think important to point out that here, the location parameter μ does not represent the mean as in the classic statistical view, but does represent the "center" of the distribution, and the scale parameter σ is not the standard deviation, but does govern the "size" of the deviations around μ. This can already be pointed out on figures in appendix where we demonstrate for little variations of the parameters.

From [Col01], we introduce an important theorem in Extreme Value Theory and that has many implications. This theorem simply says the following :

Theorem 2.3. *For any distribution function F,*

$$F \text{ is max-stable} \iff F \text{ is GEV}. \tag{2.7}$$

Hence, any distribution functions that are *max-stables* (definition 2.2) are also GEV (theorem**??**), and vice-versa. To gain interesting insights of the implications of this theorem, we think useful to give a proof but only for the "$\Leftarrow$" as the converse requires too much mathematical backgrounds.

$\boxed{Proof}$ If $a_n^{-1}(X_{(n)} - b_n)$ has limit distribution G for large n (2.5), then

$$\Pr\Big\{a_n^{-1}(X_{(n)} - b_n) \leq z\Big\} \approx G(z),$$

and so for any integer k, since nk is large,

$$\Pr\Big\{a_{nk}^{-1}(X_{(n)k} - b_{nk}) \leq z\Big\} \approx G(z), \tag{2.8}$$

But, since $X_{(n)k}$ is the maximum of k variables having identical distribution as $X_{(n)}$,

$$\Pr\Big\{a_{nk}^{-1}(X_{(n)k} - b_{nk}) \leq z\Big\} = \Big[\Pr\Big\{a_{nk}^{-1}(X_{(n)} - b_{nk}) \leq z\Big\}\Big]^k \tag{2.9}$$

giving two expressions for the distribution of M_n, by (2.8) and (2.9) :

$$\Pr\{X_{(n)} \leq z\} \approx G\Big(a_n^{-1}(z - b_n)\Big), \quad \text{and} \quad \Pr\{X_{(n)} \leq z\} \approx G^{1/k}\Big(a_{nk}^{-1}(z - b_{nk})\Big), \tag{2.10}$$

so that G and $G^{1/k}$ are identical apart from location and scale coefficients. Hence, $G$is max-stable and therefore GEV. This gives proof of the **extremal types theorem** in 2.5-2.6. □

Characterization of the GEV distributions

...

The quantity $\xi \in \mathbb{R}$ in (2.6) is called the *extreme value index* (EVI) and is at the center of the analysis in extreme value theory. It determines, in some degree of accuracy, the type of the underlying distribution. Hence, from this general definition of the GEV distribution (2.6), we can directly retrieved three principal classes of EV distributions, from their *standard form*, in the *α-parametrization*, with $\alpha = \xi^{-1}$ (just show in the ξ param. ? for convenience) :

$$\mathbf{I}: G_1(z) = \exp\{-e^{-z}\}, \qquad -\infty < z < \infty \tag{2.11}$$

$$\mathbf{II}: \ G_{2,\alpha}(z) = \begin{cases} 0, & z \leq 0 \\ \exp\{-(z)^{-\alpha}\}, & z > 0, \ \alpha > 0 \end{cases} \tag{2.12}$$

$$\mathbf{III}: \ G_{3,\alpha}(z) = \begin{cases} \exp\big\{-(-z)^{\alpha}\big\}, & z > 0, \ \alpha > 0 \\ 1, & z \geq 0 \end{cases} \tag{2.13}$$

[Faire tableau]

(mettre les indice aux fctions G + le shape parameter est "correct" ????)

The II and III can be reformulated (in the ξ-parametrization) as.....

$$G_{\xi}(z) = \exp\Big\{-(1+\xi z)^{-\xi^{-1}}\Big\}, \qquad \xi \neq 0, \tag{2.14}$$

where we replace z by $(z-\mu)/\sigma$ to add the location and scale parameters μ and σ in order to obtain the three *extreme value distributions*, see for example [RT07, pp.16].

By simply coming back in the ξ-parametrization by using $\xi = \alpha^{-1}$ in the above distribution functions, all these three classes of extreme distributions can be expressed in the same functional form as special cases of this single three-parameter (Actually, there are just location and scale parameters in the type **I** extremal model (2.11) as $\xi \to 0$) distribution (2.6). That is, when $\xi \to 0$ we retrieve the **type I** or *Gumbel* family (2.11) while $\xi > 0$ and $\xi < 0$ leads to the **type II** or *Fréchet* family and to the **type III** or *Weibull* family respectively ((2.12) and (2.13)). Both the Gumbel and Fréchet limiting distributions are unbounded (In fact, the Fréchet distribution has a finite left endpoint in $\mu - \sigma\xi^{-1}$, but this has no really interest here); that is, the upper endpoint tends to $+\infty$ while the Weibull distribution has a finite right endpoint in $\mu - \sigma\xi^{-1}$. In the following, we will define the left and the right endpoint ${}_*x$ and x_*, respectively by :

$${}_*x = \inf\{x : F(x) > 0\}, \quad \text{and} \quad x_* = \sup\{x : F(x) < 1\}.$$

Density We give a representation of the density of these functions by considering the density of the GEV distribution (2.6), that is $g_{\xi}(z) = \frac{dG_{\xi}(z)}{dz}$ (we can assume absolute continuity). This is shown in in figure (**??**) for various shape parameters. We also give in appendix

For the case $\xi \neq 0$,

$$g_{\xi}(z) = \sigma^{-1}\bigg[1+\xi\Big(\frac{z-\mu}{\sigma}\Big)\bigg]_+^{-\xi^{-1}-1} \exp\bigg\{-\bigg[1+\xi\Big(\frac{z-\mu}{\sigma}\Big)\bigg]_+^{-\xi^{-1}}\bigg\}, \tag{2.15}$$

[table + two case for $Xi = 0$ and $xi \neq 0$] [appendix graphs for various location/scale parameter values]

For the case $\xi = 0$, we have

$$g_0(z) = \sigma^{-1}\exp\bigg\{-\Big(\frac{z-\mu}{\sigma}\Big)\bigg\}\exp\bigg\{-\exp\Big[-\Big(\frac{z-\mu}{\sigma}\Big)\Big]\bigg\}, \tag{2.16}$$

Note that the support varies equally as for the distribution functions wrt sign of ξ

?? dernier graphe weibull

In some ways, some people will feel this was unfortunate, because now it is common for people to model and fit the GEV without thinking very clearly about the specific form of their data and distributions [Extremes, distribution, etc..] That is the reason why we think it can be useful to explain in the following some examples of how we can construct such extreme distributions for the three classes in concrete cases (see section **2.1.1**), playing with the appropriate choice of sequences a_n and b_n to retrieve the pertaining distribution family.

Concrete cases : examples of convergence to GEV

This is well not easy to find the sequences in practice. `http://stats.stackexchange.com/questions/105745/extreme-value-theory-show-normal-to-gumbel/105749#105749`

Type I or **Gumbel** distribution $G_1(x)$ can be retrieved by considering, for example, iid exponential distributed sequence $\{X_j\}$ of random variables, that is $X_j \overset{iid}{\sim} Exp(\lambda)$ and consider the largest of these values $X_{(n)}$ as defined earlier. By definition, we know $F(x) = 1 - \exp^{-x}$. Our goal is to find non-random sequences $\{b_n\}$, $\{a_n > 0\}$ such that

$$\lim_{n\to\infty} \Pr\{a_n^{-1}(X_{(n)} - b_n) \leq z\} = G_1(z). \tag{2.17}$$

We can easily find that

$$\begin{aligned} \Pr\{a_n^{-1}(X_{(n)} - b_n) \leq z\} &= \Pr\{X_{(n)} \leq b_n + a_n z\} \\ &= \Big[\Pr\{X_1 \leq b_n + a_n z\}\Big]^n \\ &= \Big[1 - \exp\big\{ -\lambda(b_n + a_n z)\big\}\Big]^n, \end{aligned}$$

from the iid assumption of the random variables and their exponential distribution. Hence, by choosing the sequences $a_n = \lambda^{-1}\log n$ and $b_n = \lambda^{-1}$ and reminding that

$$\begin{aligned} \Big[1 - \exp\big\{ -\lambda(b_n + a_n z)\big\}\Big]^n &= \Big[1 - \frac{1}{n}e^{-z}\Big]^n \\ &\overset{n\to\infty}{\longrightarrow} \exp(-e^{-z}) := G_1(z). \end{aligned}$$

Recall: $\boxed{\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = \exp(x)}$

we find the the so-called standard *Gumbel* distribution in the limit.

We can show the same with iid standard normal random variables, $X_j \overset{iid}{\sim} N(0, 1)$, with sequences $a_n = -\Phi^{-1}(1/n)$ and $b_n = 1/a_n$. (see appendix [extremes, distributions pdf])

Typically, unbounded distributions like the Exponential and Normal (as well as the Gamma, Lognormal, Weibull, etc.) whose tails fall off exponentially or faster will have this same Gumbel limiting distribution for the maxima, and will have medians (and other quantiles) that grow as $n \to \infty$ at the rate of (some power of) $\log n$. This is typical example of light-tailed distribution (i.e., decays exponentially as defined in section **1.1**).

Type II or **Fréchet-Pareto type** distrbution $G_2(x)$

When starting with a sequence $\{X_j\}$ iid random variables (block-maxima?) following a *basic*(or *generalized*, with scale parameter set to 1) Pareto distribution with shape parameter $\alpha \in (0, \infty)$, $X_j \sim Pa(\alpha)$, we have that

$$F(x) = 1 - x^{-\alpha}, \qquad x \in [1, \infty), \tag{2.18}$$

so that we can write, by choosing appropriately $b_n = 0$

$$\begin{aligned} -n\bar{F}(a_n z + b_n) &= -n(a_n z + b_n)^{-\alpha} \\ &= \Big[Q(1-\frac{1}{n})\Big]^{\alpha}(a_n)^{-\alpha}(-z^{-\alpha}), \end{aligned}$$

where $Q(1-\frac{1}{n})$ is the quantile function (see). Hence, it is easy to see that by setting the constant $a_n = Q(1-1/n)$ and keeping $b_n = 0$, we have that

$$\Pr\{a_n^{-1}X_{(n)} \le z\} \to \exp(-z^{-\alpha}),$$

showing that for this particular values of the normalizing constants, we retrieve the Fréchet distribution in the limit from a strict Pareto distribution. The fact that b_n is set to zero can be understood intuitively since for heavy-tailed distribution (see) such as the Pareto distribution, a correction for location is not necessary to obtain non-degenerate limit distribution. [BTV96, pp.51]

[see p.28 memoire other si ft autre exemple]

More generally, we can state the more general following theorem :

Theorem 2.4 (Pareto-type distributions)**.** *?? If for the same choice of normalizing constants as above, that is $a_n = Q(1-\frac{1}{n})$ and $b_n = 0$ and for any $x \in \mathbb{R}$*

$$n[1-F(a_n x)] = \frac{1-F(a_n x)}{1-F(a_n)} \to x^{-\alpha}, \qquad n \to \infty, \tag{2.19}$$

then we obtain the Fréchet distribution in the limit, or written formally "$\bar{F}$ is of Pareto-type*" or, more technically, "$\bar{F}$* is regularly varying with index -α*".*

We let the concepts of **regularly varying functions**, together with **slowly varying functions** be defined in appendix A.1 with some useful theorems and properties, according to [BTV96, p.51-54] and supported by [BGST06, p.49, 77-82].

[BGST06, pp.75] !!!!

Type III or **Weibull** family (?) of distributions $G_3(x)$ are, for example, in the limit of n iid uniform random variables $X_j \sim U[L,R]$ where $L \in \mathbb{R}$ and $R \in \mathbb{R}$ denote respectively the left and the right endpoint of the domain of definition, $R > L$. We have

$$F(x) = \frac{x}{R-L}, \qquad x \in [L,R],$$

which is $= 0$ for $x < L$, $= 1$ for $x > R$. Assuming the general case ($[L,R]$ can be $\neq [0,1]$), we have for the maximum $X_{(n)}$:

$$\begin{aligned} \Pr\{a_n^{-1}(X_{(n)} - b_n) \le z\} &= \Pr\{X_{(n)} \le b_n + a_n z\} \\ &= \Big[1 - \frac{R - b_n - a_n z}{R-L}\Big]^n, \quad \text{if} \quad L \le b_n + a_n z \le Rn, \\ &= (1+\frac{z}{n})^n \to e^z, \qquad \text{if} \quad z \le 0 \text{ and } n > |z|. \end{aligned}$$

When choosing $a_n = R$ and $b_n = (R-L)/n$, we find the unit Reversed Weibull distribution $We(1,1)$ in the limit as expected.

However, for inferential purpose, this is not of particular interest, because from the expression in 2.5, we know that

$$a_n^{-1}(X_{(n)} - b_n) \xrightarrow{d} G_{\xi,\mu,\sigma}(z), \qquad \text{as } n \to \infty. \tag{2.20}$$

After some algebra, this leads to

$$X_{(n)} \xrightarrow{d} G_{\xi,\mu^*,\sigma^*}(z), \qquad \text{as } n \to \infty, \tag{2.21}$$

with the sequences a_n and b_n being absorbed into the new location and scale parameters μ^* and σ^*. We can then ignore the normalizing constants in practical applications and fit directly the GEV in our set of maxima $X_{(n),k}$. The pertaining estimated parameters will implicitly take the normalization into account, i.e. it will estimate μ^* and σ^*. As also stated in.., the shape parameter is invariant.

But what about the fact that $X_{(n)}$ non-normalized is degenerate (see intro) ?

Maximum domain of attraction

The preceding results can be more easily summarized and obtained when considering *maximum domain of attraction* (MDA). The term "*maximum*" is typically used to distinguish from *sum-stable* distribution. As we study here only the maxima, there are no confusion possible in our work. We will then preferably write only *domain of attraction* in the following for convenience, and thus consider these two names as synonyms.

Definition 2.5 (Domain of attraction)**.** *We say that a distribution F is in the (***maximum***)* **domain of attraction** *of an extreme value family G_k (see eq.2.11-2.13), denoted by $F \in D(G_k)$, if there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that the distribution of $a_n^{-1}(X_{(n)} - b_n)$ converges weakly (see* **??***) to G_k where $X_{(n)}$ is as defined earlier with distribution F.*

The definition is well-defined in the sense that $F \in D(G_i)$ and $F \in D(G_j)$ implies $\xi_i = \xi_j$, writing by ξ_k the extreme value index pertaining to the extreme value distribution G_k.

Before going further with the characterization of the three domains of attraction of our purpose, we think important to introduce a new theorem from Gnedenko [**?**]

Theorem 2.5 (Convergence to Types Theorem)**.** *Let F_n be a sequence of random variables converging weakly (see appendix A.1.1) to F. Let $a_n > 0$ and $b_n \in \mathbb{R}$ such that $a_nF_n + b_n \Rightarrow F'$, where both F and F' are non-degenerate (see* **??** *). Then,*

$$a_n \to a \quad \textit{and} \quad b_n \to b, \qquad a > 0 \quad \textit{and} \quad b \in \mathbb{R}.$$

Equivalently, if G_n, G, G' are distribution functions with G, G' being non-degenerate, and there exists $a_n, a'_n > 0$ and $b_n, b'_n \in \mathbb{R}$ such that

$$G_n(a_nx + b_n) \xrightarrow{d} G(x) \quad \textit{and} \quad G_n(a'_nx + b'_n) \xrightarrow{d} G'(x),$$

at all continuity points of F, respectively F', then there exists constants $A > 0$ and $B \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{a_n}{a'_n} \to A, \qquad \lim_{n\to\infty} \frac{(b_n - b'_n)}{a'_n} \to B,$$

and $G'(Ax + B) = G(x)$ $\forall x \in \mathbb{R}$.

We have now all the necessary tools to the pertaining domains of attractions. But, before proceeding, we would like to point out that the fact that the characterization of the first domain of attraction (Gumbel class) is much more complex than the two following (Fréchet and Weibull class) and requires much more technicalities going beyond the scope of this thesis. Moreover, despite this class is important in theory, it is less relevant for our purpose of modelling extremes. It often requires other generalizations, for instance with additional parameters to surpass the issues of fitting empirical data. [PF15] In the last paragraph, we will present the unified framework, the domain of attraction pertaining to the GEV distributions, which is a kind of summary for the three first domains of attraction presented.
In each of the characterization of the domains of attractions, we will present some of their most useful, necessary and sufficient conditions ... together with their *von Mises conditions*, initially from [] but revisited in [**?**]. These conditions are very important in practice and sometimes more intuitive because they make use of the *hazard function*, defined by, for sufficiently smooth distributions :

$$r(x) = \frac{f(x)}{\bar{F}(x)} = \frac{f(x)}{1 - F(x)}. \tag{2.22}$$

This involves the density function $f(x) = \frac{dF(x)}{d(x)}$ in the numerator.[**?**] We can conversely define the *reciprocal hazard function* simply by $\tilde{r}(x) = 1/r(x)$.

I. **Domain of attraction for Gumbel distribution G_1** We derive here two ways of formulating necessary and sufficient condition for a distribution function F to be in the domain of attraction of G_1, namely $F \in D(G_1)$.

- From (mettre vrmt??) [dHF06, pp.20], for finite or infinite right endpoint x_* with $\int_x^{x_*} \int_t^{x_*} \bar{F}(s) ds dt < \infty$, the function
$$h(x) := \frac{(\bar{F}(x)) \int_x^{x_*} \int_t^{x_*} (\bar{F}(s)) \; ds \; dt}{\left(\int_x^{x_*} (\bar{F}(s)) ds \right)^2},$$
must satisfy $\lim_{t \uparrow x_*} h(t) = 1$. [Reminder: $\lim_{t \uparrow y}(\,\cdot\,)$ means that t is approaching y from below, i.e. from values smaller than y in a increasing manner, and vice-versa for $\lim_{t \downarrow y}(\,\cdot\,)$].

- From [BGST06, pp.72], for some auxiliary function b, for every $v > 0$, the condition
$$\frac{\bar{F}(x + b(x)v)}{\bar{F}(x)} \to e^{-v}, \tag{2.23}$$
must hold as $x \to x_*$. Then,
$$\frac{b(x + vb(x))}{b(x)} \to 1.$$

voir lien avec la GPD!! ecrire en hazard rate?

A lot of more precise characterizations and conditions together with proofs can be found, for example in [dHF06, pp.20-33]. We can also mention a condition that is based on the von Mises function[].

However, we present his ***von Mises criterion*** as in [BGST06, pp.73]:

If the *hazard function* $r(x)$ (2.22) is ultimately positive in the neighbourhood of x_*, is differentiable there and satisfies

$$\lim_{x \uparrow x_*} \frac{dr(x)}{dx} = 0, \tag{2.24}$$

then $F \in D(G_1)$. (compare hazar convergence rates of the three types !!!!!)

Examples of distributions in $D(G_1)$ Intuitively, we can remark that all distributions which are exponentially decaying will have this propensity to be in the Gumbel domain of attraction. For instance, the *Exponential*, the *Gamma*, the $Weibull$, the *logistic*, etc. To see that, by a Taylor expansion, we have that

$$\bar{G}_1(x) = 1 - \exp(-e^{-x}) \sim e^{-x}, \qquad \text{as } x \to \infty.$$

Hence, the Gumbel domain of attractions G_1 decays exponentially (as tend their pertaining distributions).

$\boxed{\textbf{II.}}$ **Domain of attraction for Fréchet distribution $G_{2,\alpha}$** Let $\alpha := \xi^{-1} > 0$ be the *index* of the Fréchet distribution $G_{2,\alpha}$ (see (2.12)). Then, $F \in G_{2,\alpha}$ if and only if

PAS F !!!!!!!!

$$\bar{F}(x) = x^{-\alpha} L(x), \tag{2.25}$$

for some slowly varying function L. See for example theorem **2.4** (?). In this case and with $b_n = 0$,

$$F^n(a_n x) \to G_2(x), \qquad \forall x \in \mathbb{R},$$

with

$$a_n := F^{\leftarrow}(1 - \frac{1}{n}) = \Big(\frac{1}{1-F}\Big)^{\leftarrow}(n),$$

where we define the quantity $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$ for $t < 0 < 1$ as the *generalized inverse of* F with which we can retrieve $x_t = F^{\leftarrow}(t)$, the t-quantile of F. Even if we deal in this text only with continuous and strictly increasing distribution functions (?), we think it is more reliable(?) to consider generalized inverse instead of the ordinary inverse, for sake of generalization.

This previous theorem informs us that all distribution functions $F \in D(G_{2,\alpha})$ have necessarily an infinite right endpoint, that is $x_* = \sup\{x : F(x) < 1\} = \infty$. These distributions are all with regularly varying right-tail with index $-\alpha$. In short,

$$F \in D(G_{2,\alpha}) \Longleftrightarrow \bar{F} \in R_{-\alpha}.$$

Finally, we must also present the (revisited) ***Von Mises condition*** for this domain of attraction which state the following in [FM93] : if F is absolutely continuous with density f and right endpoint $x_* = \infty$, such that

$$\lim_{x \uparrow \infty} x\ r(x) = \alpha > 0,$$

where $r(x)$ is the *hazard function* defined in eq.(2.22), then $F \in D(G_{2,\alpha})$. In words, it means that... We illustrate this with the standard Pareto distribution case (as previsouly in), that is

$$F(x) = \left(1 - (\frac{x_m}{x})^{\alpha}\right)1_{x \geq x_m}, \qquad \alpha > 0 \text{ and } x_m > 0.$$

Clearly, we can see that by setting $K = x_m^{\alpha}$, we have

$$\bar{F}(x) = Kx^{-\alpha}.$$

Therefore, we have that $a_n = (Kn)^{\alpha^{-1}}$ and $b_n = 0$.

Examples of distributions in $\boldsymbol{D(G_{2,\alpha})}$ These distributions are typically very-fat tailed (and hence, heavy-tailed, see) distributions, such that $E(X_+)^{\delta} = \infty$ for $\delta > \alpha$. This class of distributions is appropriate for phenomena with extremely large maxima (like...). [**?**] Common distributions include Pareto, Cauchy, Burr, stable distributions with $\alpha < 2$, etc. An example to see that, is again by Taylor expansion at the tail of $G_{2,\alpha}$ with $\alpha > 0$

$$\bar{G}_{2,\alpha}(x) = 1 - \exp(-x^{-\alpha}) \sim x^{-\alpha}, \qquad x \to \infty, \tag{2.26}$$

showing that $G_{2,\alpha}$ tends to decrease as a *power law.* See for example eq.

$\boxed{\textbf{III.}}$ **Domain of attraction for Weibull distribution $\boldsymbol{G_{3,\alpha}}$** We say that $F \in G_{3,\alpha}$ (2.13) with index $\alpha > 0$ if and only if there exists finite right endpoint $x_* \in \mathbb{R}$ such that

$$\bar{F}(x_* - x^{-1}) = x^{-\alpha}L(x), \tag{2.27}$$

where $L(\,\cdot\,)$ is a slowly varying function (see).

For $F \in D(G_{3,\alpha})$, we have also

$$a_n = x_* - F^{\leftarrow}(1 - n^{-1}), \qquad b_n = x_*.$$

Hence

$$a_n^{-1}(X_{(n)} - b_n) \xrightarrow{d} G_{3,\alpha}.$$

[see references [domain of attraction course]]

Finally, we still present the ***Von Mises condition*** from [FM93] related to the $G_{3,\alpha}$) domain of attraction. It states that for F having positive derivative on some $[x_0, x_*)$, with finite right endpoint $x_* < \infty$, then $F \in D(G_{3,\alpha})$ if

$$\lim_{x \uparrow x_*}(x_* - x)r(x) = \alpha > 0, \qquad \int_{-\infty}^{x_*} \bar{F}(u)du < \infty, \tag{2.28}$$

where $r(x)$ is again the *hazard function* defined in (2.22).

Examples of distributions in $\boldsymbol{D(G_{3,\alpha})}$ Weibull's domain of attraction thus includes all the distribution functions that are bounded to the right ($x_* < \infty$). As most phenomena are typically bounded, we will think as the Weibull for the most attractive and flexible class for modelling extremes. But, in practice, the Fréchet one is often more preferable in an extreme analysis context because allowing for arbitrarily large values.

[put general case pp.73-75 beirlant] ?

Closeness under tail equivalence property An interesting property of all these three types of domain of attraction $D(G_{k,\alpha})$ is that those are *closed under tail-equivalence* **??**. It this sense,

- For the **Gumbel** domain of attraction, let $F \in D(G_{1,\alpha})$. If H is another distribution function such that, for some $b > 0$,

$$\lim_{x \uparrow x_*} \frac{\bar{F}(x)}{\bar{H}(x)} = e^b, \tag{2.29}$$

  then H is also in the domain of attraction of the Gumbel distribution, $H \in D(G_{1,\alpha})$.

- For the **Fréchet** domain of attraction, let $F \in D(G_{2,\alpha})$. If H is another distribution function such that, for some $c > 0$,

$$\lim_{x \to \infty} \frac{\bar{F}(x)}{\bar{H}(x)} = c^\alpha, \tag{2.30}$$

  then H is also in the domain of attraction of the Fréchet distribution, $H \in D(G_{2,\alpha})$.

- For the **Weibull** domain of attraction, let $F \in D(G_{3,\alpha})$. If H is another distribution function such that, for some $c > 0$,

$$\lim_{x \uparrow x_*} \frac{\bar{F}(x)}{\bar{H}(x)} = c^{-\alpha}, \tag{2.31}$$

  then H is also in the domain of attraction of the Weibull distribution, $H \in D(G_{3,\alpha})$

For more informations about the characterizations of the, one can refer to ....

Domain of attraction of the GEV [Coles slides 30] (and see stat extremes beirlant!!) The conditions that have been stated the three preceding domains of attraction can be restated under this "unified" framework for the GEV distribution defined in (2.6) For a given distribution function F, by letting the sequences b_n, a_n, and the shape parameter such that[2]

$$b_n = F^{\leftarrow}(1 - 1/n), \qquad a_n = r(b_n) \qquad \text{and} \qquad \xi = \lim_{n \to \infty} \tilde{r}(x),$$

with $r(\cdot)$ the *hazard* function defined (in 2.22). Hence, $a_n^{-1}(X_{(n)} - b_n)$ has limiting distribution

$$\begin{cases} \exp\Big\{-\big[1 + \xi\sigma^{-1}(x-\mu)\big]_+^{-\xi^{-1}}\Big\}, & \xi \neq 0, \\ \exp\big\{-e^{-x}\big\}, & \xi = 0, \end{cases}$$

which is the GEV (see eq2.11-2.14) [see if put location an scale parameters]

Among a lot of characterizations available, we present the most []:

Theorem 2.6. *If there exist a positive, measurable function $u(\cdot)$, then for $-\infty < \xi < \infty$, $F \in D(GEV)$ if and only if :*

$$\lim_{v \uparrow x_*} Pr\Big\{\frac{X - v}{u(v)} > x \mid X > v\Big\} := \lim_{v \uparrow x_*} \frac{\bar{F}(v + xu(v))}{\bar{F}(v)} = \begin{cases} (1+\xi x)^{-\xi^{-1}}, & \xi \neq 0 \\ e^{-x}, & \xi = 0 \end{cases} \tag{2.32}$$

for $(1 + \xi x) > 0$.

[2] We think important to recall again the reader the difference of parametrization $\xi = \alpha^{-1}$

The Skewed Generalized Extreme Value Distribution

[?]

Return Levels and Return Periods

Return levels play a major role in environmental analysis.

Assuming for this introductory example our time unit reference is in year -as usually assumed in meteorological analysis-, let us consider the *m-year return level* r_m and define it, (at first sight), as the high quantile for which the probability that the annual maximum exceeds this quantile is 1/m, which is called the *return period.* Let $\{X_{(n),y}\}$ denote the iid sequence of random variables representing the annual maximum for yeay y. From (2.6), we then have (check the implication)

$$F(r_m) = \Pr\{X_{(n),y} \leq r_m\} = 1 - 1/m \tag{2.33}$$

$$\Leftrightarrow \quad \left[1 + \xi\left(\frac{r_m - \mu}{\sigma}\right)\right]^{-\xi^{-1}} = \frac{1}{m}. \tag{2.34}$$

Hence, by inverting this relation, and letting $y_m = -\log(1 - m^{-1})$, we can get the quantile of the GEV that we name as the *return level* r_m

$$r_m = \begin{cases} \mu + \sigma\xi^{-1}(y_m^{\xi} - 1) & \text{if } \xi \neq 0, \\ \mu + \sigma\log(y_m) & \text{if } \xi = 0. \end{cases} \tag{2.35}$$

(See for eq. index ())

Hence, we can directly retrieve it from our estimation of the three GEV parameters.

However, we recall that the definition of return period is easily misinterpreted and the given above is thus not universally accepted. To evaporate (vanish) this issue, it is important to distinguish stationary from non-stationary sequences.

Explore why the return leels go beyond the right endpoint of the distribution (when $\xi < 0$ as here), for which return period, etC...

Return Level Plot

Standard errors of the estimates As usual, the standard errors of these estimates are important to compute, for example to construct confidence intervals (but they can be quite misleading!), and hence the return level plot. We naturally expect these to increase with the return period. As r_m is a function of the GEV parameters, we can use the *delta method* (see..)to approximate the variance of $\hat{r}_m$. Specifically,

$$\text{Var}(\hat{r}_m) \approx \nabla r_m^{'} V \nabla r_m, \tag{2.36}$$

with V the variance-covariance matrix of the estimated parameters $(\hat{\mu}, \hat{\sigma}, \hat{\xi})'$ and

$$\begin{aligned} \nabla r_m^{'} =& \left[\frac{\partial r_m}{\partial \mu}, \frac{\partial r_m}{\partial \sigma}, \frac{\partial r_m}{\partial \xi}\right] \\ =& \Big[1,\ \xi^{-1}(y_m^{-\xi} - 1),\ \sigma\xi^{-2}(1 - y_m^{-\xi}) - \sigma\xi^{-1}y_m^{-\xi}\log y_m\Big], \end{aligned} \tag{2.37}$$

with $y_m = -\log(1-m^{-1})$, with the gradient evaluated at the estimates $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. But a problem arise for the so-computed standard errors when considering long-range return levels. (GRAPH??) They can increase so drastically with the return period that the confidence intervals of the *return level plot* can become difficult to work with. To try to get rid of this issue with in section **3..** by constructing intervals on the basis of the *profile* log-likelihood.

Stationarity Under an assumption of a stationary sequence, the return level is the same for all years, and this gives rise to the notion of the return period (or m-year event). Hence, the return period of a particular event is the inverse of the probability that the event will be exceeded in any given year. The m-year return level is associated with a return period of m years. However, there are two main interpretations in this context for return periods.

[AEH+13a, pp.100]

Denoting $X_{(n),y}$ the annual maximum for year y. Omitting the notational dependence on block size n, we assume $\{X_{(n),y}\} \overset{iid}{\sim} F$.

1. The first interpretation of the m-year event is **the expected waiting time until an exceedance occurs**. To see that, letting T be the year of the first exceedance, we we can write

$$\begin{aligned}
\Pr\{T=t\} &= \Pr\{X_{(n),1} \le r_m, \ldots, X_{(n),t-1} \le r_m, X_{(n),t} > r_m\} \\
&= \Pr\{X_{(n),1} \le r_m\} \ldots \Pr\{X_{(n),t-1} \le r_m\} \Pr\{X_{(n),t} > r_m\} && \text{[from iid assumption]} \\
&= \Pr\{X_{(n),1} \le r_m\}^{t-1} \Pr\{X_{(n),1} > r_m\} && \text{[from stationarity]} \\
&= F^{t-1}(r_m)(1-F(r_m)) \\
&= (1-1/m)^{t-1}(1/m).
\end{aligned} \tag{2.38}$$

   We easily recognize that T has geometric density with parameter $1/m$. From simple properties of geometric distributions, we found its expected values is $1/(1/m)$, that is the expected waiting time for an m-year event is m years.

2. The second interpretation of the m-year event is that **the expected number of events in a period of m years is exactly 1**. To see that, we define

$$N = \sum_{y=1}^{m} I(X_{(n),y} > r_m)$$

   as the random variable representing the number of exceedances in m years (where I is indicator function). We can view each year as a "trial", and from the fact that we have assumed $\{X_{(n),y}\}$ are iid, we can compute the probability that the number of exceedances in m-years is k

$$\Pr\{N=k\} = \binom{m}{k}(1/m)^k(1-1/m)^{m-k}$$

   from which we recognize a well-know distribution, that is $N \sim$ Binomial$(m, 1/m)$. Again from properties of this distribution, we easily find that N has an expected value of 1.

Non-stationarity From the definition on non-stationary process, the modelling of return period will change over time. Hence, we introduce the notation of the distribution function F_y of a particular $X_{(n),y}$. We must study

$p(y) = \Pr(X_{(n),y} > r) = 1 - F_y(r)$. If we estimate F_y, we can retrieve easily p(y). $F_y(r_p(y)) = 1 - p$ with the exceedance level $r_p(y)$ changing with year. It shows the changing nature of "risk".

Return period as expected waiting time

Return period as expected number of events

Peaks-Over-Threshold Method

Seuils meteo : 0C (gel permanent), 25C et 30C pour les Tx 0C (gel) et 20C pour les Tn
—> Use this for thresholds ?

Preliminaries: some intuition

The threshold models relying on the *Peaks-Over-Threshold* (POT) method are useful to propose a better (?) alternative than the blocking method in **2.1**. With this new method, we consider a more natural way of determining whether an observation is extreme or not, by focusing only on all observations that are greater than a pre-specified *threshold*. As we saw, estimates of the GEV parameters are sensitive to the size of block chosen to identify extremes (see) while we will investigate that the estimates of the GPD parameters are more stable in this sense. Henceforth POT avoids the problem that can arise by considering the maximum of blocks only (), but this method also brings its own problems (). Be aware that this method brings lots of problems with the independence condition... And, especially for temperature data, where for example during heat or cold waves...

Let's consider a sequence $\{X_j\}$ of n iid random variables having marginal distribution function F. We are then regarding for observations that exceed a well-chosen (see) threshold u, which must obviously be smaller than the right endpoint $x_* = \sup\{x : F(x) < 1\}$ of F. The aim here is to here is to find a "child" probability distribution function (fig.? -video youtube), say H, from the underlying (parent) distribution F, that will allow us to model the exceedance $Y = X - u$, and with H then expressed as $\boxed{H(y) = \Pr\{X - u \leq y | X > u\}}$. Typically, threshold models can therefore be regarded as the conditional survival function of the exceedances Y, knowing that the threshold u is exceeded [BGST06, pp.147] :

$$\Pr\{Y > y \mid Y > 0\} = \Pr\{X - u > y \mid X > u\} = \frac{\bar{F}(u+y)}{\bar{F}(u)}. \tag{2.39}$$

or in terms of the exceedance distribution function $F^{[u]}(x) = \Pr\{X \leq u + x | X > u\}$ [RT07, pp.12], [**?**] and [Ros15] :

$$F^{[u]}(x) = \frac{\Pr\{X - u \leq x, X > u\}}{\Pr\{X > u\}} = \frac{F(x+u) - F(u)}{\bar{F}(u)} \tag{2.40}$$

making use of the well-known conditional probability law. One can remark that (2.39) is actually the survivor of the exceedance distribution function, that is $\bar{F}^{[u]}$.

These intuitive characterizations we have given above about the modelling of the threshold exceedances in term of probability distribution function can be useful to understand the following.

However, if the parent distribution F were known, we would be able to compute the distribution of the threshold exceedances in (2.39). [Col01, pp.74] But as for the GEV in the method of block-maxima (section **2.1**), the distribution F is not known in practice, as we will see also in (..). Hence, and as usual in statistics[3], we must again rely on approximations. This time, we will try to approximate (2.40)

Characterization of the Generalized Pareto Distribution

Anageously to the *Fisher-Tippett* theorem in section **2.1** which applies for the block maxima, we have now to define a new theorem which applies for values above a predefined threshold. From this result 2.6(?), these two theorems form together the basis of Extreme Value Theory.

Theorem 2.7 (POT-stability). *[RT07, pp.25] The max-stability theorem in* **??** *can be applied and are formulated here by the fact that the GP distribution functions H are the only continuous one such that, for certain choice of constants a_u and b_u,*

$$F^{[u]}(a_u x + b_u) = F(x).$$

This will be useful for modelling the exceedances in the following theorem (?). And for the examples (see ex. p.25)

Theorem 2.8 (Pickands–Balkema–de Haan). *discovered by [BdH74] and [Iii75] which showed that the distribution of a threshold u of normalized excesses $F^{[u]}(x)(b_u + a_u x)$, as the threshold approaches the right endpoint x_* of F, is the Generalized Pareto Distribution (**GPD**) $H_{\xi,\sigma_u}(y)$. That is, if X is a random variable for which (2.5) holds, and for the approximating GP distribution function possessing the same left endpoint u as the exceedance distribution function $F^{[u]}$, we have [RT07, pp.27]:*

$$|F^{[u]}(x) - H_{\xi,\sigma_u}(x)| \longrightarrow 0, \qquad u \to x_*,$$

or, in an other, maybe more intuitive formulation [Col01] :

$$\Pr\{X \le y \mid X > u\} \longrightarrow H_{\xi,\sigma_u}(y), \qquad u \to x_*, \tag{2.41}$$

where the **GPD** is defined as :

$$H_{\xi,\sigma_u}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)_+^{-\xi^{-1}}, & \xi \neq 0, \\ 1 - \exp\left\{-\frac{y}{\sigma_u}\right\}_+, & \xi = 0, \end{cases} \tag{2.42}$$

where we recall again that $y = x - u > 0$, and where the scale parameter is denoted σ_u to emphasize its dependency with the chosen threshold u :

$$\sigma_u = \sigma + \xi(u - \mu), \tag{2.43}$$

where one can also remark that the location parameter μ does not appear anymore in (2.42) as it does appear in 2.47.

[3]We would like to quote here the well-known phrase in statistics "*All models are wrong, but some are useful*" from George Box & Draper (1987), *Empirical model-building and response surfaces*, Wiley, p.424

Outline proof of the GPD and justification from GEV As we did for block-maxima approach in section **2.1.1** (2.1.1-2.10), we think it is interesting to have a formal and comprehensive, and still not too technical, intuitive view of where are the GPD from. We remind that we aim here at retrieving the GPD $H_{\xi,\sigma_u}(y)$ (2.41-2.42) from probability distributions as expressed in (2.39-2.40).

$\boxed{Proof}$ We start with X having distribution function F. From the GEV theorem in section **2.1.** (see 2.5-2.6), we have for the largest order statistic, for large enough n,

$$F_{X_{(n)}}(z) = F^n(z) \approx \exp\left\{ -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\xi^{-1}} \right\}, \tag{2.44}$$

with $\mu, \sigma > 0$ and ξ the GEV parameters. hence, by simply taking logarithm on both sides, we have

$$n \ln F(z) \approx -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\xi^{-1}}. \tag{2.45}$$

- We also have that, from Taylor expansion, $\ln F(z) \approx -\big[1-F(z)\big]$ as both sides go to zero when $z \to \infty$. Therefore, substituting into (2.45), we get the following for large u :

$$1-F(u) \approx n^{-1}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\xi^{-1}}.$$

Or, specially expressed for our purpose of retrieving something in the form of (2.39-2.40), with $y > 0$,

$$1-F(u+y) \approx n^{-1}\left[1+\xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-\xi^{-1}}.$$

- Whence we get for 2.39, with some mathematical manipulations, as $u \to x_*$:

$$\begin{aligned}
\Pr\{X > u+y \mid X > u\} = \frac{\bar{F}(u+y)}{\bar{F}(u)} &\approx \frac{n^{-1}\big[1+\xi\sigma^{-1}(u+y-\mu)\big]^{-\xi^{-1}}}{n^{-1}\big[1+\xi\sigma^{-1}(u-\mu)\big]^{-\xi^{-1}}} \\
&= \left[1+\frac{\xi\sigma^{-1}(u+y-\mu)}{1+\xi\sigma^{-1}(u-\mu)}\right]^{-\xi^{-1}} \\
&= \left[1+\frac{\xi y}{\sigma_u}\right]^{-\xi^{-1}},
\end{aligned}$$

where σ_u is still linear in the threshold u (2.47), that is $\sigma_u = \sigma + \xi(u-\mu)$. By simply reverting the probability as in (2.40), we have then

$$\begin{aligned}
\Pr\{X-u \le y \mid X > u\} &= 1-\Pr\{X > u+y \mid X > u\} \\
&= 1-\left(1+\frac{\xi y}{\sigma_u}\right)^{-\xi^{-1}}
\end{aligned} \tag{2.46}$$

which is $GPD(\xi, \tilde{\sigma})$ as required and σ_u is as defined in (2.47)

□

More comprehension can come from [RT07, pp.27-28] or if one wants to compute rates of convergence.

Dependence of the scale parameter σ We chose to express the scale parameter as σ_u to emphasize its dependency with the threshold u. If we increase the threshold, say to $u' > u$, then the scale parameter will be adjusted following :

$$\sigma_{u'} = \sigma_u + \xi(u' - u), \tag{2.47}$$

and in particular, this adjusted parameter $\sigma_{u'}$ will increase if $\xi > 0$ and decrease if $\xi < 0$. If $\xi = 0$, there would be no change in the scale parameter[4]. We think important to point out the fact that, similarly as mentioned for the GEV models in (2.6), the scale parameter σ_u for GPD models is not the usual standard deviation, but does govern the "size" of the excesses. [AEH$^+$13b, pp.20]

We will later discuss the threshold choice in section **3.**

Three different types of GPD and duality with GEV One will remark the similarity with the GEV distributions as the parameters of the GPD of the threshold excesses are uniquely determined by the corresponding GEV distribution parameters of block-maxima (see outline proof in the above to convince yourself). Hence, the shape parameter ξ of the GPD is equal to that of the corresponding GEV and, most of all, it is invariant[5] while the computation of σ_u will not be affected by changes of the corresponding μ or σ in the GEV, from the self-compensation arising in (2.47). [Col01, pp.76]

Hence, as for the block-maxima approach, there are also three possible families of the GPD depending on the value of the shape parameter ξ which determines the qualitative behaviour of the GPD. [HW87], [SG95]

- The **first** type, call it $H_{0,\sigma_u}(y)$, comes by letting the shape parameter $\xi \to 0$ in 2.42, giving :

$$H_{0,\sigma_u}(y) = 1 - \exp\Big(-\frac{y}{\sigma_u}\Big), \qquad y > 0, \tag{2.48}$$

  where we recognize that it corresponds to an **exponential distribution**, and hence light-tailed, with parameter $1/\sigma_u$, namely $Y \sim \exp(\sigma_u^{-1})$.

- The **second** and the **third** types, that is when $\xi < 0$ and $\xi > 0$ (resp.), differ only by their support :

$$H_{\xi,\sigma_u}(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-\xi^{-1}}, \text{ for } \begin{cases} y > 0, & \xi > 0, \\ 0 < y < u + \sigma_u/|\xi|, & \xi < 0. \end{cases} \tag{2.49}$$

  Therefore, if $\xi > 0$ the corresponding GPD is of **Pareto**-type, hence is heavy-tailed, and has no upper limit while if $\xi < 0$, the associated GPD has an upper bound in $y_* = u + \sigma_u/|\xi|$

[4]This is consistent with the *memoryless property* of the exponential distribution H_{0,σ_u} (**??**), for which we give more details in

[5]For instance, choosing different block size in the GEV modelling would shift its (estimated) parameters while GPD (estimated) parameters are *stable*.

and is then **Beta**-type distribution. A special case arise when $\xi = -1$ where the pertaining distribution becomes Uniform$(0, \sigma)$. [**?** , pp.186]

Some plots ?

After looking at the behaviour of the density of these functions, we will procure a more comprehensive view by defining some examples of how to retrieve these different types of Generalized Pareto Distributions.

Density functions of the GPD

$$h_{\xi,\sigma_u}(y) = \frac{\xi}{\sigma_u}\left(1 + \xi\frac{y}{\sigma_u}\right)^{-\xi^{-1}-1} \tag{2.50}$$

Examples of the GPD as limiting distribution for exceedances We have seen in the previous paragraph that if we can have an approximate distribution G for block-maxima, then threshold excess will have a corresponding distribution given by a member of the Generalized Pareto family. Whence the shape parameter ξ, as for GEV distributions, is determinant for controlling the behaviour of the GPD, and thus leads to the three different types in ((2.48)-(2.49)).

1. The first type

The choice of a threshold will be discussed in section **3.5.1**.
From [BGST06, p.147-],

Point Process Approach

Following mainly [Col01], with some further concepts taken from ? ,

As for the two preceding methods, the point process approach aims at modelling some sequences which are initially assumed to be independent (??)

[coles, pp.124] Here, Point Process could be seen as a kind of summary of the two previous methods, leading to nothing new. However, this approach if often preferred :

1. Its interpretation **unifies** the **models** considered so far.
2. Its likelihood enables a more natural formulation of non-stationarity in excess models from the Generalized Pareto model, see section **2.2**.

Furthermore, we will see that the parametrization of the point process model is invariant to threshold choice so that this variation would only affect the well-known (already mentioned) bias-variance trade-off in the inference. Interesting if seasonal modelling.

Non-homogeneous Poisson process

$$N(A) \sim \mathrm{Poi}(\Lambda(A)), \qquad \Lambda(A) = \int_A \lambda(x)dx. \tag{2.51}$$

Method for Minima

transposer ! enoncé modeles (1pg) pas tout

Methods of Inference

We decide to present in this section the two mains methods of inference for GEV distributions. First, the likelihood-based methods for their wide applicability, and easy interpretability. Then, we will broadly present the bayesian methods for their increasing supports in this domain, and easy adjustability. Finally, we will present some other well-known methods that are also widely used to estimate GEV parameters like the Hill or the moment estimator.

As we already discussed in section **2.1.1.** (see (2.20)-(2.21)), a great advantage for the modelling is that we do not have to find the normalizing sequences

Likelihood-based Methods

A potential difficulty with the use of likelihood methods for the GEV concerns the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. Such conditions are not satisfied by the GEV model because the end points of the GEV distribution are functions of the parameter values, $\mu - \sigma/\xi$ is an upper end-point of the distribution when $\xi < 0$, and a lower end-point (?) when $\xi > 0$. [**?**]

, from decreasing order of [Col01, pp.55]

1. $\boldsymbol{\xi < -1}$: MLE's are unlikely to be obtainable. This is due to
2. $\boldsymbol{\xi \in (-1, -0.5)}$: MLE's are generally obtainable but their standard asymptotic properties do not hold.
3. $\boldsymbol{\xi > -0.5}$: MLE's are regular, in the sense of having the usual asymptotic properties.

But fortunately, in practice, the problematic cases in the two first situations ($\xi \leq 0.5$) are rarely encountered for environmental problems. This situation corresponds to distributions (in the Weibull or in the Beta? family) with very short bounded upper tail, see for example figure **??**. And if it is the case, Bayesian inference, which do not depend on these regularity conditions, may be preferable.

Problems of this simple method arise when the approximate normality of the MLE cannot hold. Hence, the underlying inferences are not sustainable. This is the reason why another method is usually more preferable, the *profile likelihood*.

Profile Likelihood

The standard likelihood is not the most accurate method for inference. In section **2.5**, the problem was that confidence intervals computed in the usual method, with standard errors computed by the Delta method in 2.37, was not reliable for inference on return levels. This is due to rejection of the normal approximation (see...) because of the severe asymmetries that are often observed in the likelihood surface for return levels, especially for large quantiles. [**?**]

This is why it is sometimes useful to consider other approach, for example the *profile likelihood* which is often more convenient when a single parameter is of interest, for instance the return level r_m. Let's denote it θ_j. Now let's consider a parameter vector $\boldsymbol{\theta} = (\theta_j, \boldsymbol{\theta_{-j}})$ where $\boldsymbol{\theta_{-j}}$ corresponds to all components of $\boldsymbol{\theta}$ except θ_j. $\boldsymbol{\theta_{-j}}$ can be seen as a vector of nuisance parameters. The profile log-likelihood for θ_j is defined by

$$\ell_p(\theta_j) = \arg\max_{\boldsymbol{\theta_{-j}}} \ell(\theta_j, \boldsymbol{\theta_{-j}}). \tag{3.1}$$

Henceforth for each value of θ_j, the profile log-likelihood is the maximised log-likelihood with respect to $\boldsymbol{\theta_{-j}}$, i.e. to all other components of $\boldsymbol{\theta}$. Generalization where $\boldsymbol{\theta_j}$ is of dimension higher than one is possible.

Return levels Here, we are now specifically interested in computing the profile log-likelihood for the estimation of the return level $\theta_j = r_m$. To do that, we present a method which consists of three main steps :

1. To include r_m as a parameter of the model, by **??** we can rewrite μ as a function of ξ, σ and r_m :
$$\mu = r_m - \sigma\xi^{-1}\Big[\Big(-\log\{1-m^{-1}\}\Big)^{-\xi} - 1\Big].$$
By plugging it in the log-likelihood in 3.2-3.3, we obtain the new GEV log-likelihood $\ell(\xi, \sigma, r_m)$ as a function of r_m.
2. We maximise this new likelihood $\ell(\xi, \sigma, r_m = r_m^-)$ at some fixed low value of $r_m = r_m^- \leq r_m^+$ with respect to the "nuisance" parameters (ξ, σ) to obtain the profile log-likelihood
$$\ell_p(r_m = r_m^-) = \operatorname*{arg\,max}_{(\xi,\sigma)} \ \ell\Big(r_m = r_m^- \ , \ (\xi, \sigma)\Big).$$
We choose arbitrarily large value of the upper range r_m^+ and conversely for starting point of r_m^-.
3. Repeat previous step for a range of values of r_m such that $r_m^- \leq r_m \leq r_m^+$ and then choose r_m which attain the maximum value of $\ell_p(r_m)$.

From this, we easily obtain the *profile log-likelihood plot*

Generalized Extreme Value distribution

(return level) [extremesin climate change p.106] "Approximate confidence intervals for the return level can be obtained by the delta method (Casella and Berger, 2002, Sect. 5.5.4) which relies on the asymptotic normality of maximum-likelihood estimators and produces a symmetric confidence interval. Alternatively, profile likelihood methods (Coles 2001, Sect. 2.6.6) provide asymmetric confidence intervals, which better capture the skewness generally associated with return level estimates." stationary case

We are now considering a sequence $\{Z_i\}_{i=1}^n$ of independent random variables sharing each the same GEV distribution. From the densities of the GEV distribution $g_\xi(z)$ defined in (2.15-2.16), we derive the log-likelihood $\log\big[L(z;\mu,\sigma,\xi)\big]$, for the two different cases $\xi \neq 0$ or $\xi = 0$ respectively:

1.
$$\ell(\mathbf{z};\mu,\sigma,\xi \neq 0) = -m\log\sigma - (1+\xi^{-1})\sum_{i=1}^{n}\log\left[1+\xi\Big(\frac{z_i-\mu}{\sigma}\Big)\right]_+ - \sum_{i=1}^{n}\left[1+\xi\Big(\frac{z_i-\mu}{\sigma}\Big)\right]_+^{-\xi^{-1}}. \tag{3.2}$$

2.
$$\ell(\mathbf{z};\mu,\sigma) = -m\log\sigma - \sum_{i=1}^{n}\Big(\frac{z_i-\mu}{\sigma}\Big) - \sum_{i=1}^{n}\exp\Big\{-\Big(\frac{z_i-\mu}{\sigma}\Big)\Big\}. \tag{3.3}$$
using the Gumbel limit $\xi \to 0$ of the GEV, see 2.16.

Generalized Pareto Distribution

As we have seen, excess-over-threshold models rely on .... From 2.50, we can write the *log-likelihood* of the GPD :

$$\ell(\mathbf{z};\xi,\sigma_u) = -n\ln\sigma_u - (1+\xi^{-1})\sum_i^n \ln(1+\xi\sigma_u^{-1}z_i), \qquad (1+\xi\sigma_u^{-1}z_i) > 0. \tag{3.4}$$

Bayesian Methods

see evdbayes pdf package r

Bayesian inference usually provides a viable alternative in cases when MLE (for example) breaks down. And actually, we are not so fo far from the problematic situations depicted in section **3.1**

For an asymmetric distribution, the HPD interval can be a more reasonable summary than the central probability interval (see illustration ...). For symmetric densities, HPD and central interals are the same while HPD is shorter for asymmetric densities. See [**?**]....

As the dependence becomes stronger, the run length n must be larger in order to achieve the same precision. Dependence exists both within the output for a single parameter (autocorrelations) and across parameters (cross-correlations).

Other Methods

[BGST06, pp.140]

The Hill estimator

[only alid for pareto type $\xi > 0$ tails ?]

The probability-view [BGST06, pp.103] which is the simplest and most intuitive considering the peaks-over-threshold case.

http://freakonometrics.hypotheses.org/12226

Problem : see [pp.105]

Pickands estimator

The Moments estimator

The Probability-Weighted-Moments Estimator

[**?**]

The L-Moments Estimator

[**?**]

Estimators based on generalized quantile

Bootstrap Methods

Monte-Carlo based methods, same as Bayesian.

Study and comparisons on the performance (coverage,..) of the methods used for the CI (boot, bayesian, likelihood, asymptotics,...)

Moving Block Bootstrap

[Bootstrap and other resampling in .... pp.13]

Model Selection

Threshold choice for the excess models

Single threshold selection involves a **bias-variance trade-off**. That is, (raccourcir)

- **Lower threshold** will induce **higher the bias** due to model misspecification. In other words, the threshold must be sufficiently high to ensure that the asymptotics underlying the GPD approximation are reliable.
- **Higher threshold** will induce higher estimation uncertainty, i.e. **higher variance** of the parameter estimate as the sample size is reduced for high threshold.

(Following [LLR83], this is practically equivalent to estimation of the k^{th} upper order statistic $X_{(n-k+1)}$ called the "tail fraction" below. To ensure tail convergence, as $n \to \infty$, $k \to \infty$ but at a reduced rate such that $k/n \to 0$, i.e. the quantile level of the threshold increases at a faster rate as the sample size n grows.

)

Mean residual life function or *mean excess function* , following again [BGST06, pp.14-19], [Col01, pp.78-80],

$$\begin{aligned} mrl(u_0) &:= E(X - u_0 \mid X > u_0) \\ &= \frac{\int_{u_0}^{x_*} \bar{F}(u)du}{\bar{F}(u_0)}, \end{aligned} \tag{3.5}$$

for X having survival function $\bar{F}(u_0)$ computed at u_0, with again $x_* = sup\{x : F(x) < 1\}$ denoting the right endpoint of the support of F. It denotes, in an actuarial context, the expected remaining quantity or amount to be paid out when a level u_0 has been chosen. However, even if it is mainly applied in an actuarial context or in survival analysis in the literature (see [**?**] for a well-known example), there are also interesting and reliable applications in our more environmental purposes as we will see in the following. Moreover, this function has interesting properties about the tail of the underlying distribution of X [BGST06, pp.16]. In fact, we expect the following :

- If $mrl(u_0)$ is constant, then X has exponential distribution.

- If $mrl(u_0)$ ultimately increases, then X has a heavier tail than the exponential distribution.
- If $mrl(u_0)$ ultimately decreases, then X has a lighter tail than the exponential one.

(and vice-ersa, goes it in the two sens??)

This can be particularly interesting for our purpose when considering threshold models. For this case, we can suppose the excesses of a threshold generated by the sequence $\{X_i\}$ follow a generalized Pareto distribution (see 2.2). Knowing the theoretical mean of this distribution, we retrieve, provided the shape parameter $\xi < 1$ and denoting σ_u the scale parameter corresponding to excess of a threshold $u > u_0$,

$$\begin{aligned} mrl(u) := E(X-u \mid X>u) &= \frac{\sigma_u}{1-\xi} \\ &= \frac{\sigma_{u_0}+\xi u}{1-\xi}, \end{aligned} \tag{3.6}$$

from the threshold u dependence with the scale parameter σ (see 2.47). Hence, we remark that $mrl(u)$ is linearly increasing in u, with gradient $\xi(1-\xi)^{-1}$ and intercept $\sigma_{u_0}(1-\xi)^{-1}$. Furthermore, we can estimate empirically this function intuitively by

$$\widehat{mrl}(u) = \frac{1}{n_u}\sum_{i=1}^{n_u}(x_{[i]}-u), \tag{3.7}$$

where we let the $x_{[i]}$ denoting the (i-th out of the) n_u observations that exceed u.

Mean residual life plot This leads to an interesting tool for our purpose, the *mean residual life plot*. It comes from combining the linearity detected between $mrl(u)$ and u in (3.6) with (3.7). Therefore, a reliable information can be retrieved from the point of the points

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_{[i]}-u)\right) : u < x_{max}\right\}. \tag{3.8}$$

Even if its interpretation is not easy, this graphical procedure will give insights for the choice of a suitable threshold u_0 to model extremes via general Pareto distribution, that is the threshold u_0 above which we can detect linearity in the plot. Relying on this well-chosen threshold u_0, the generalized Pareto distribution should be a good approximation. Remind however that its interpretation is often subjective. Furthermore, information in the far right-hand-side of this plot is unreliable. Variability is high due to the limited amount of data (exceedances) above very high thresholds. This can be seen for example on larger confidence intervals.

From [Col01, pp.83-84]

"Substantial subjectivity in interpreting these diagnostic plots, and the resulting uncertainty. Similar challenges are seen with the River Nidd data, shown in Tancredi et al. (2006), and many other exam- ples in the literature. These examples suggests that a more ‘objective’ thresh- old estimation approach is needed and that uncertainty must be accounted for."

see mixture pdf]

Varying threshold

Mixture Models

[?]

The threshold is either implicitly or explicitly defined as a parameter to be automatically estimated, and in most cases the un- certainty associated with the threshold choice can be accounted for naturally in the inferences.

Cross-validation [?] Besides all these methods that are very subjective,...

or see gelman bayesian book pp.169

Dispersion index plot

Model Diagnostics : Goodness-of-Fit

After having fitted a statistical model to data, it is important to assess its accuracy in order to infer reliable conclusions from this model.

Ideally, we aim to check that our model fits well the whole population, that is the whole distribution of maxima. all the past and future temperature maxima that will arise " As this cannot be achieved in practice, it is common to assess a model with the data that were used to estimate this model. We will talk a bit about the problem that could arise from this methodology in the next section,and the problem of overfitting(?).

Diagnostic Plots : Quantile and Probability Plots

From [BTV96, pp.18-36], together with the nice view of [Col01, pp.36-37], we present two major diagnostic tools which aims at assessing the fitting of a particular model (or distribution) against the real distribution coming from the data used to construct the model. These are called the *quantile-quantile* plot (or *qq*-plot) and the *probability* plot.

These diagnostics are popular by their easy interpretation and by the fact that they can both have graphical (i.e. subjective, qualitative, quick) view but also a more precise (i.e. objective, quantitative, rigorous) analysis can be derived, for example from the theory of linear regression. For these two diagnostic tools, we use the order statistics introduced in (1.1) but now we rather consider an **ordered sample** of independent observations :

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)} \tag{3.9}$$

coming from a population from which we fit the estimated model (distribution) $\hat{F}$ and where $x_{(1)}$ (resp. $x_{(n)}$) is thus the minimum (resp. maximum) observation in the sample. These tools will thus help us to know if the fitted model model $\hat{F}$ is reasonable for the data.

Quantile plot Given the sample in (3.9), a *qq-plot* consists of the locus of points

$$\left\{ \left(\hat{F}^{\leftarrow}\Big(\frac{i}{n+1}\Big) \ , \ x_{(i)} \right) \ : \ i = 1, ..., n \right\}. \tag{3.10}$$

This graphic compares the ordered quantiles $\hat{F}^{\leftarrow}\left(\frac{i}{n+1}\right)$ of the fitted model $\hat{F}$ against the ordered observed quantiles, i.e. the ordered sample from (3.9).

We used the continuity correction $\frac{i}{n+1}$.....

Probability plot Given the same sample in (3.9), a *probability plot* consists of the points

$$\left\{\left(\hat{F}(x_{(i)})\ ,\ \frac{i}{n+1}\right)\ :\ i=1,...,n\right\}. \tag{3.11}$$

This graph compares the estimated probability of the ordered values $x_{(i)}$ from the fitted model $\hat{F}$ against the probability...

From these two graphical diagnostic tools, the interpretation is the same and we will consider that $\hat{F}$ fits well the data if the plot looks linear, i.e. the points of the plots lie close to the unit diagonal.

Besides the fact that the probability and the quantile plots contain the same information, they are expressed in a different scale. That is, after changing the scale to probabilities or quantiles (with probability or quantile transforms), one can gain a better perception and both visualizations can sometimes lead contradictory conclusions, especially in the graphical inspection. Using both is thus preferable to make our model's diagnostic more robust.

The disadvantage of Q–Q plots is that the shape of the selected parametric distribution is no longer visible [BGST06][pp.62?]

Overfitting problem [**?**]? A problem of these diagnostics could arise when we focus on prediction accuracy and as we mentioned the fact that the model is fitted on the data. This well-known problem is called *overfitting*, and can be easily explained by the following :

- We are looking for a model which fits the data at best, i.e. for points which are the nearest possible of the diagonal line.
- But, the so-constructed model from which we put the diagnostic is fitted from these original data against which we make the comparison.
- Hence, there could be a incentive to fit a model which fits the most perfectly the available data, that is which points on the diagnostic plots is the nearest possible of the diagonal line. The model is then the best to fit the data at hand
- But, this is a catastrophy when we are seeking at making good predictions from the fitted model, that is making a guess on new, unseen, unavailable data. The model has then lost flexibility, it is not regularized and cannot generalize. (unless the feature space, hear the initial data space, has been completely explored (—>infinite data ?))

See the link with the trade-off bias-variance for threshold selection.

Gumbel plots

Z and W statistic plots

Stationary time Series

In practical environmental applications, the independence assumption is rarely completely fulfilled []. This sounds obvious in extreme values analysis of temperature data, since we expect the temperature .... Whereas it was not really problematic for block-maxima, it was much more painful for POT as we have seen both in section 2. However, here we can see in our case that it does not really happens...(verif.. and why??)

See [BGST06, pp.375]

From now, we considered $X_{(n)} = \max_{1\leq i\leq n} X_i$ where we have assumed $X_1, \ldots, X_n$ are independent random variables. For sake of simplicity, we abandon this notation. In the sequel, this will be denoted by $\tilde{X}_{(n)} = \max_{1\leq i\leq n} \tilde{X}_i$ where $\tilde{X}_1, \ldots, \tilde{X}_n$ will typically denote a sequence of independent random variables, so that the maximum $\tilde{X}_{(n)}$ is composed of (plays with) independent random variables only. We are now interested by modelling $X_{(n)} = \max_{1\leq i\leq n} X_i$ where $\{X_i\}$ will now denote a *stationary* sequence of random variables sharing the same marginal distribution as $\{\tilde{X}_i\}$, F.

Definition 4.1 (Stationarity)**.** *We say that the sequence* $\{X_i\}$ *is* **stationary** *if*

More generally, for $h \geq 0$ *and* $n \geq 1$*, the distribution of the lagged random vector* $(X_{1+h}, \ldots, X_{n+h})$ *does not depend on* h *when the sequence is said to be (strongly) stationary.*

Note that we will only focus on weak(?) stationarity.

Let us denote $F_{i_1,\ldots,i_p}(u_1, \ldots, u_p) := \Pr\{X_{i_1} \leq u_1, \ldots, X_{i_p} \leq u_p\}$ as the joint distribution function of $(X_{i_1}, \ldots, X_{i_p})$ for any arbitrary positie integers $(i_1, \ldots, i_p)$.

Definition 4.2 ($D(u_n)$ dependence condition)**.** *From [**?**] and following [BGST06, Col01, pp.373-374, pp.93] Let* $\{u_n\}$ *be a sequence of real numbers. The* $\boldsymbol{D(u_n)}$ **condition** *holds if for any set of integers* $i_1 < \cdots < i_p$ *and* $j_1 < \cdots < j_q$ *such that* $j_1 - i_p > \ell$*, we have that*

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_q}(u_n, \ldots, u_n; u_n, \ldots u_n) - F_{i_1,\ldots,i_p}(u_n, \ldots, u_n)F_{j_1,\ldots,j_q}(u_n, \ldots u_n)| \leq \beta_{n,\ell} \ , \tag{4.1}$$

where $\beta_{n,\ell}$ *is nondecreasing and* $\lim_{n\to\infty} \beta_{n,\ell_n} = 0$*, for some sequence* $\ell_n = o(n)$*, as* $n \to \infty$*.*

This condition ensures that, when the sets of variables are separated by a relatively short distance, typically $s_n = o(n)$, the long-range dependence between such events is limited, in a sense that is sufficiently close to zero to have no effect on the limit extremal laws.

From this result, we can retrieve the *extreme-value theorem*

Result is remarkable in the sense that, provided a series has limited long-range dependence at extreme levels (D(u_n)) condition makes precise), maxima of stationary series follow the same distributional limit laws as those of independent series. [S.Coles 2001 p.94]

For specific sequence of theresholds u_n that increase with n.

Theorem 4.1 (Limit distribution of maxima under $D(u_n)$)**.** *From [**?**]. Let* $\{X_n\}$ *be a stationary sequence with* $X_{(n)} = \max(X_1, X_2, \ldots X_n)$*. If there exists sequences* $\{a_n > 0\}$ *and* $\{b_n\}$ *such that* $D(u_n)$ *condition holds, then*

$$Pr\{X_{(n)} \leq u_n\} \longrightarrow H(x) \qquad n \to \infty, \tag{4.2}$$

where H *is non-degenerate as defined... and* $D(u_n)$ *is satisfied with* $u_n = a_n x + b_n$ *for every real* x*.*

[bootstrap and other....]

Theorem 4.2 (Leadbetter 1983)**.** *From [Col01, pp.] Let* $\{X_n^*\}$ *be a stationary sequence and let* $\{X_n\}$ *be a sequence of iid random variables. By defining* $X_{(n)}^* = \max\{X_n^*\}$ *and* $X_{(n)} = \max\{X_n\}$*, we have under regularity conditions,*

$$Pr\{a_n^{-1}(X_{(n)} - b_n) \leq x\} \longrightarrow G(x), \qquad n \to \infty,$$

for normalizing sequences $\{a_n > 0\}$ *and* $\{b_n\}$*, where* G *is non-degenrate, if and only if*

$$Pr\{a_n^{-1}(X_{(n)}^* - b_n \leq x\} \longrightarrow G^*(x), \qquad n \to \infty.$$

G^* *is the limit distribution coming from a stationary process, defined by*

$$G^*(x) = G^\theta(x), \tag{4.3}$$

for some constant $\theta \in (0, 1]$ *which is called the* **extremal index***.*

The extremal index

The *extremal index* is an important indicator quantifying the extent of extremal dependence, or equivalently the degree at which the assumption of independence is violated. From eq.(4.3), it is clear that if $\theta = 1$, then the process is independent, but the converse does not hold while the case $\theta = 0$ will not be considered as it is too "far" from independence (check with data?) and brings problems, see for example [BGST06, pp.379-380]. Moreover, the results of Theorem 4.2. would not hold true.

.. However, the maximum has a tendency to decrease as .... [Col01, pp.96]

Formally, it can be defined as

$$\theta = \lim_{n\to\infty} \Pr\{\max(X_2, \ldots, X_{p_n}) \leq u_n \mid X_1 \geq u_n\}, \tag{4.4}$$

where $p_n = o(n)$ and the sequence u_n is such that $\Pr\{X_{(n)} \leq u_n\}$ converges. [Col01][slides]

Hence, θ can be thought as the probability that an exceedance over a high threshold is the final element in a *cluster* of exceedances.

Cluster of exceedance From eq.(4.4), we can now state that extremes have the tendency to occur in cluster, whose *mean cluster size* is θ^{-1} at the limit. Equivalently(?), θ^{-1} is the factor with which the mean distance between cluster is increased.

Identifying clusters and declustering as the distribution of a cluster maximum is the same as the marginal distribution of an exceedance. + slide 82-83(?)

However, [pp.178 Coles], information in discarded when one considers *declustering*. And this information could be substantially important in meteorological applications, for instance to determine heat or cold waves.

New parameters When $\theta > 0$, we have from Theorem 4.2 that G^* is an EV distribution but with different scale and location parameters than G. If we note by (μ^*, σ^*, ξ^*) the parameters pertaining to G^* and those from G kept in the usual way, we have the following relationships when $\xi \neq 0$

$$\mu^* = \mu - \sigma\xi^{-1}(1-\theta^{\xi}), \qquad \sigma^* = \sigma\theta^{\xi}. \tag{4.5}$$

In the Gumbel case ($\xi = 0$), we have $\sigma^* = \sigma$ and $\mu^* = \mu + \log\theta$. The fact that $\xi^* = \xi$ is

Return levels From that (see clusters), one can see that the proba ility of an exceedance is variable (see coles, pp.103 or slide 82) (...)

$$r_m = u + \sigma\xi^{-1}\Big[(m\zeta_u\theta)^{\xi} - 1\Big] \tag{4.6}$$

It is important to take that into account as ignorance of this "dependence" can lead to overestimation of the return level.

Tail dependence

From [RT07, section 2.6], [Col01, section 8.4] or [BTV96, section 9.4.1,10.3.4] + see tail dependence function `atdf()` in `R`.

The auto-tail dependence function using $\chi(u)$ and/or $\bar{\chi}(u)$ employs X against itself at different lags.

a possible estimator (this used by `atdf()`) can come from the sample version

$$\rho_n(u,h) = \frac{1}{n(1-u)}\sum_{i\le n} I(\min(x_i, x_{i+h}) > x_{[nu]:n}) \tag{4.7}$$

(compare with beirlant notations!!!!)$x_{[nu]:n}$

Modelling

Block-Maxima

The modelling with the techniques provided by the GEV distributions (see section **2.1.**)as seen in the beginning of the text can be used in the similar way as we have seen from eq.4.3) or in section 4.1 that the shape parameter remains invariant. The difference is that the effective number of maxima $n =$ will be reduced and hence the convergence will slower.

However, a problem is still unsolved [Col01][pp.98] and is related to the approximations in the limit. Indeed, as the effective number of observations is reduced from n to $n\theta$, the approximation is expected to be poorer, and this "problem" will be exacerbated with increased levels of dependence in the series.

Thresholds models

Practically speaking, one might expect a threshold based analysis to result in estimates of return levels with much reduced standard errors as "all" the extremes are included in the analysis, i.e. those who excess a threshold u. The example in section .... illustrated this

However, the fact that this method deals with "all" the extremes brings also some problems, and especially the issue of *temporal dependence* (see plot of acf or pacf wrt u) which is illustrated by the fact that the extremes have a tendency to *cluster*. Inferences based on the likelihood found in eq.(3.4) which relies on the independence assumption are now invalid.

Several methods can be used :

- Filtering out an (approximate) independent sequence of threshold exceedances.
- Declustering

Non-Stationary Extreme Time Series

Whereas we have considered and relaxed during the previous section the first "i" of the "iid" assumption made during the whole chapter **2**, we will now tackle the last part "**id**", i.e. the strong (?) assumption that the observations are **i**dentically **d**istributed.

The stationarity assumption is very poor to hold for climatologic data [**?**]. It is also the case for temperature data.

OUR AIM HERE IS TO MODEL THE ***EVOLUTION*** OF As we are dealing with time varying sequences, we can

- Positive trend
- Seasonality

The aim of our modelling will more focus on a different parametrization for the mean, thus in allowing the location parameter to vary through time/seasons.

- Variation in time through t accounting for the season : $\mu(t) = \beta_0 + \mathbb{1}_i(t)$ where i=1,2,3,4 represent the seasons.

Block-Maxima

As we continue to consider modelling as yearly blocks, we do only face nonstationary concerns for the trend which is (probably) imputed to the Global Warming. The evidence of seasonality arising when we decrease the length of the blocks is not an issue for yearly modelling. However, we loose information, or comparatively, we do not use all the information as at least one half

Diagnostics

Gumbel plot (slide 94) coles

Part II

Experimental Framework : Global....

In this part, we will focus on the application of the methods seen during the theoretical part.

Analysis of maximum temperatures in Uccle

PUT the examples right in the place where it is mentioned in the theory.! "As we have seen in section 2.1.1.... and in section 2.2.2......"

We must choose a block-length which is large enough for the limiting arguments supporting the GEV approximation (see (2.5)) to be valid, either a large bias in the estimates could occur. For example, if this is too short, the maxima may be too close of each other to assume independence. But a large block-length implies less data to work with, and thus a large variance of the estimates.. So we must find a compromise between bias and variance.

TABLE with nested models (gumbel, GEV, + linear trend, etc etc)

Conclusion

"A key issue in applications is that inferences may be required well beyond the observed tail of the data, and so an assumption of stability is required:" [**?**]

"Another approach would be to use something other than time as the covariate in the model. For instance, one could imagine linking temperature data directly to CO2 level rather than time. However, linking to a climatological covariate makes extrapolation into the future more difficult, as one would need to extrapolate the covariate as well. No obvious climatological covariate comes to mind for the Red River application. "

Timescale-uncertainty effects on extreme value analyses seem not to have been studied yet. For stationary models (Sect. 6.2), we anticipate sizable effects on block extremes–GEV estimates only when the uncertainties distort strongly the blocking procedure. For nonstationary models (Sect. 6.3), one may augment confidence band construction by inserting a timescale simulation step (after Step 4 in Algorithm 6.1) [Mud14, pp.262]

(!!!!! delete unuseful equation numbering ?!!!!)

Replace "distribution functions" by "cdf".

voir notation vectors (en gras ou avec bar en bas?)

attention aux notation homogenes (ex : partie stationnary,...) –> pour dénoter les maximums,....

Appendices

Statistical concepts

Convergence concepts

Weakly convergence We say that a sequence of random variables X_n *converges weakly* to

Varying functions

Figures

do not fortget [**?**]§
Beirlant: 2004 !!

Bibliography

[AEH⁺13a] Amir AghaKouchak, David Easterling, Kuolin Hsu, Siegfried Schubert, and Soroosh Sorooshian, editors. *Extremes in a Changing Climate*, volume 65 of *Water Science and Technology Library*. Springer Netherlands, Dordrecht, 2013.

[AEH⁺13b] Amir AghaKouchak, David Easterling, Kuolin Hsu, Siegfried Schubert, and Soroosh Sorooshian, editors. *Extremes in a Changing Climate*, volume 65 of *Water Science and Technology Library*. Springer Netherlands, Dordrecht, 2013.

[BdH74] A. A. Balkema and L. de Haan. Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792–804, October 1974.

[BGST06] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, March 2006. Google-Books-ID: jqmRwfG6aloC.

[BTV96] Jan Beirlant, Jozef L. Teugels, and Petra Vynckier. *Practical Analysis of Extreme Values*. Leuven University Press, 1996. Google-Books-ID: ylR3QgAACAAJ.

[Col01] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, London, 2001.

[dHF06] L. de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer series in operations research. Springer, New York ; London, 2006. OCLC: ocm70173287.

[FM93] Michael Falk and Frank Marohn. Von Mises Conditions Revisited. *The Annals of Probability*, 21(3):1310–1328, July 1993.

[FT28] R. A. Fisher and Leonard H. C. Tippett. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *ResearchGate*, 24(02):180–190, January 1928.

[Gne43] B. Gnedenko. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

[HW87] J. R. M. Hosking and J. R. Wallis. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):339, August 1987.

[Iii75] James Pickands Iii. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131, January 1975.

[KTWK⁺02] A. M. G. Klein Tank, J. B. Wijngaard, G. P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A. F. V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L. V. Alexander, and P. Petrovic. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, 22(12):1441–1453, October 2002.

[LLR83] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer New York, New York, NY, 1983.

[Mud14] Manfred Mudelsee. *Climate Time Series Analysis*, volume 51 of *Atmospheric and Oceanographic Sciences Library*. Springer International Publishing, Cham, 2014.

[PF15] E. C. Pinheiro and S. L. P. Ferrari. A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data. *arXiv:1502.02708 [stat]*, February 2015.

[Res87] Sidney I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY, 1987.

[Ros15] Gianluca Rosso. Extreme Value Theory for Time Series using Peak-Over-Threshold method. *arXiv preprint arXiv:1509.01051*, 2015.

[RT07] Rolf-Dieter Reiss and Michael Thomas. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields ; [Includes CD-ROM]*. Birkhäuser, Basel, 3. ed edition, 2007. OCLC: 180885018.

[SG95] V. P. Singh and H. Guo. Parameter estimation for 3-parameter generalized pareto distribution by the principle of maximum entropy (POME). *Hydrological Sciences Journal*, 40(2):165–181, April 1995.