



FACULTAD DE CIENCIAS

**Statistical Extreme Value Theory.
Application to basins of the Basque Country**

*(Teoría estadística de valores extremos. Aplicación a cuencas
fluviales del País Vasco)*

Pablo Señas Peón

TRABAJO DE FIN DE MÁSTER
PARA ACCEDER AL

**Máster Universitario en Matemáticas y
Computación**

Director: Juan Antonio Cuesta Albertos
Director: Cristina Prieto Sierra

Septiembre - 2020

Abstract

Floods are among the most common and destructive natural disasters, directly and indirectly affecting human life and property. Improving the assessment and management of flood risks is therefore a major priority, and it is a challenge for the scientific and technical communities to address these needs. Within this context, improved forecasting of flood volumes is of great assistance to authorities in taking the most appropriate protection measures.

This report begins with a compilation of the most important concepts and results of the Extreme Value Theory, including proofs of some of the most relevant ones. In the second part, the previous theory will be applied to the expected flood levels in the river Oñati, in the Basque Country. These predictions have been made using several available libraries in the programming language R.

Keywords: Basque Country, Extreme Value Theory, Floods, Rivers,

Resumen

Las inundaciones se encuentran entre las catástrofes naturales más comunes y destructivas, afectando directa e indirectamente a la vida humana y a la propiedad. Por lo tanto, mejorar la evaluación y la gestión de los riesgos de inundación es una prioridad para la sociedad. Dentro de este contexto, la mejora de las predicciones de los volúmenes de crecidas de los ríos es de gran ayuda para la adopción por parte de las autoridades de las medidas de protección más adecuadas.

Este trabajo comienza haciendo una recopilación de los conceptos y resultados más importantes de la Teoría Estadística de los Valores Extremos, incluyendo las demostraciones de algunos de los más relevantes. En la segunda parte se aplicará la teoría anterior a la evaluación de los niveles de crecida esperables en la cuenca fluvial del río Oñati, en el País Vasco. Estas predicciones se han realizado usando librerías previamente disponibles en el lenguaje de programación R.

Palabras clave: Inundaciones, País Vasco, Ríos, Teoría de Valores Extremos

Contents

Introduction	1
1 Statistical Extreme Value Theory	5
1.1 Fisher-Tippett-Gnedenko theorem	5
1.1.1 Proof of the Theorem	5
1.1.2 Domains of attraction and examples	9
1.1.3 Generalized Extreme Value distribution	12
1.2 Estimation of the parameters	13
1.2.1 Block Maxima approach	14
1.2.2 Exceedance approach	20
1.2.3 On the value of ξ	25
1.3 Results under dependence conditions	28
2 River floods in the Basque Country	33
2.1 Motivation	33
2.2 Methodology	36
2.3 Analysis: river Oñati	38
2.3.1 Block Maxima	39
2.3.2 POT approach	44
2.3.3 Conclusion and future lines of work	49
Bibliography	51
Appendices	55
A Additional figures	57
B On the density plot for the GP fitting	61

Introduction

When you have excluded the impossible, whatever remains, however improbable, must be the truth.

Arthur Conan Doyle

Extreme Value Theory may be defined as the branch of Statistics that deals with extreme percentiles of a distribution. Being interested in extreme values of a distribution is as old as statistics itself. In 1709 Nicholas Bernoulli poses a problem on finding the expected lifetime of the last survivor among a group of men with the same age. Later on, in the 19th century, astronomers were interested in finding criteria for assessing whether an extreme value is an outlier or not.

Let X_1, \dots, X_n be independent and identically distributed (iid) random variables (rv's) with distribution F , and suppose we are interested in knowing how the rv $M_n := \max(\{X_1, \dots, X_n\})$ behaves the bigger n gets. Clearly, the cumulative distribution function (cdf) of M_n is, for each n , F^n . However, most of the times F is unknown, and even though it can be estimated from the empirical cdf \tilde{F} , the difference between \tilde{F}^n and F^n is likely to be large. On the other hand, it can be proved ([LLR83, Corollary 1.5.2]) that M_n converges to $x_F := \sup\{x \in \mathbb{R} \cup \{\infty\} : F(x) < 1\}$ almost surely. Unfortunately, this only shows something that we already suspected, but it does not let us extract more information about the behaviour of F . There are several things about F that we may be interested in knowing. For example, the return period of an extreme value, which is the estimated time that we have to wait for that extreme value to occur (in formal terms, if u is the extreme value, then its return period is $(1 - F(u))^{-1}$), or the return level associated to a period, which are *inverse* concepts. Many security protocols are based on return periods/levels. For example, assume $X_1, \dots, X_n \sim F$ describe the mean sea level of a particular place in the year n . Then, the return level for the period of 100 years is the value u such that $(1 - F(u)) = \frac{1}{100}$, that is, the value that we expect to see exceeded once every 100 years. When building a dike, clearly it is not feasible to ensure that it prevents any flood; the best we can do is to prevent very rare ones. That is why its height is sometimes determined because it is the return level of a big period of time. The tools developed by Extreme Value Theory allow us to estimate quantities like these.

Extreme Value Theory was not born until the 1920's. The first results of the decade were only concerned with large values of the normal distribution. For example, in 1925 Leonard Tippett studies the distribution of the largest values of samples coming from the normal distribution for different sample sizes. The first important result is dated from 1927: Maurice Fréchet identifies a possible asymptotic family of distributions for large values, and realizes that extreme values from different distributions can follow the same asymptotic distribution. He is also aware that the limit distributions must meet a certain condition, namely, the *max-stability* condition, which will be introduced in time. However, his work was overshadowed by

the research of Fisher and Tippett, who got the same conclusions a year later, while figuring out two other limit distributions. In 1936 Richard von Mises gives sufficient conditions for a distribution to have extreme values following one of the three limit families, and in 1943 Boris Gnedenko gives necessary and sufficient conditions to this. The result stating that the asymptotic distribution of the normalized maximum, if non-degenerate, is restricted to only three different families, is now known as the Fisher-Tippett-Gnedenko theorem, and such distributions are called the Gumbel, Fréchet and Weibull families. The Fisher-Tippett-Gnedenko theorem is arguably the most important result of Extreme Value Theory, because the way of dealing with extreme values of a distribution is, essentially, dealing with the limit distribution. The first book on Extreme Value Theory, as it known today, and which gave raise to the discipline, is Gumbel's *Statistics of Extremes*, published in 1958 [Gum58]. It contained all the knowledge to the date, and gave several applications on, for example, engineering, climatology or hydrology. These fields still rely on Extreme Value Theory for a multitude of problems.

These are the main results of Classical Extreme Value Theory, which only deals with real valued, independent and identically distributed observations. However, they are also worth to mention: 1) the advance concerning more complex conditions than the ones stated by von Mises and Gnedenko, which have been called second order conditions, and are based on the theory of regularly-varying functions. 2) Pickands-Balkema-de Haan theorem, which approximates the distribution of the exceedances of a variable following a distribution with a non-degenerate limit distribution for its extremes. This theorem may be regarded as the second most important result of Extreme Value Theory.

The second half of the 20th century was focused on studying the behaviour of the extreme values of a distribution when the data in the sample are not independent. Of course, this is motivated by the fact that in practice, it is virtually impossible to have such independent data. The most natural generalization of a sequence of iid rv's is a stationary series: stationarity is considerably more realistic. In general, the Fisher-Tippett-Gnedenko theorem does not hold: consider, for example, the stationary process $X_n = X_1$ for $n \in \mathbb{N}$. However, under rather general dependence conditions, extremes of sequences behave roughly as in the independent case. This makes the first results still very useful and applicable. See, for instance, [LLR83].

Nowadays, Extreme Value Theory is not reduced to the univariate case. Multivariate Extreme Value Theory has also been a topic of interest for a long time. For example, in 1964 Gumbel and Goldstein studied the maximum annual discharge of the Ocmulgee River measured at two different stations [GG64]. Multivariate Extreme Value Theory has been thoroughly studied; however, it is significantly more complicated. Observe that there are problems from the very beginning: for example, what is a multivariate extreme value? In addition, the results reached are not nearly as satisfying as in the univariate case. For instance, now we do not have a finite number of possible limit distributions. A possible reference for this topic is Chapter 8 of [BGT04] or Part II of [HF06]. Extreme values for infinite-dimensional observations have also been studied. For reference, see Part III of [HF06].

In this dissertation, we will discuss the basic results of Univariate Extreme Value Theory, and then we will apply them to a real dataset. In the first part, the Fisher-Tippett-Gnedenko and the Pickands-Balkema-Theorem will be proved, and general results on a distribution to have extreme values belonging to one of the three limit distributions will also be shown. Following Gumbel's motto, *no distribution is stated without an explanation of how the parameters may be estimated*: regrettably, the limit distributions to be dealt with have several

complications when trying to estimate their parameters. The moments of a distribution, for example, may be infinite, and this may have an influence on the estimation method to be used or, in general, on how to deal with the distribution. As a consequence, part of our effort will be devoted to the estimation issue. To this end, there are two approaches: the Block Maxima and the Peak-over-Threshold methods, each of which are motivated by the important aforementioned theorems. We will delve into both. Eventually, Extreme Value Theory deals with the prediction of high quantiles, so we will also talk about that. We will also discuss how some of the results still hold when we allow the sequence to be stationary along with a condition that weakens the dependence of the observations the more separated they are.

The results of Extreme Value Theory refer to both large and small values of a distribution. However, in practice, only one of the situations is discussed, because of the evident relation between the maxima and the minima of observations. As a consequence, results for the maximum (respectively, minimum) can be deduced from results for the minimum (respectively, maximum). We have chosen the maxima approach, which is far more common in the literature.

In the second part of the dissertation we will estimate extreme discharge levels of the river Oñati, in the Basque Country, measured at a gauging station. These data have never been analyzed using these tools before.

It can be claimed that Extreme Value Theory is not an easy discipline, which is expected: the difficulty of predicting extreme values and quantiles come from the fact that it is difficult to obtain samples because they themselves are very difficult to observe. As cited in Gumbel's book: *however big floods get, there will always be a bigger one coming*. The results that we are going to present give us a framework to draw conclusions on higher percentiles than the ones we have observed, and we cannot expect these to be easy.

Given the restrictions on the length of the report, the proofs which are shown in detail are not too long, with the exception of the Fisher-Tippett-Gnedenko theorem, because of its unquestionable importance. We aim to find a proper balance between a self-contained yet rigorous theoretical introduction to Extreme Value Theory and an application which attempts to solve an existent problem. Data used have been made available to me thanks to the *Instituto de Hidráulica Ambiental*, part of the University of Cantabria.

Both parts include tables, images and figures. Some of them have been included in the appendix. Unless stated otherwise, they are my own creation, and have been made using the programming language R and already existent packages. The code is available in https://github.com/senaspablo/TFM_StatisticalExtremeValueTheory_BasinsBasqueCountry.

Chapter 1

Statistical Extreme Value Theory

1.1 Fisher-Tippett-Gnedenko theorem

Let X_1, \dots, X_n be iid with distribution F . As we mentioned earlier, F^n converges to x_F almost surely, so this path gets nowhere if we want to get some insight on how large values of F behave. A possible alternative is to consider linear normalizations, that is, to study the limit of $\mathbb{P}(a_n(M_n - b_n) \leq x) = F^n(a_n^{-1}x + b_n)$ for suitable sequences $a_n > 0$ and $b_n \in \mathbb{R}$. As we will see, this gives us very interesting results.

Example 1. Let $F(x) = \exp(-\exp(-x))$ for $x \in \mathbb{R}$. Then $F^n(x) = \exp(-\exp(-x + \log(n)))$, that is, $F^n(x + \log(n)) = F(x)$, and taking $a_n = 1$ and $b_n = \log(n)$ we get:

$$\mathbb{P}(a_n(M_n - b_n) \leq x) = \mathbb{P}(M_n \leq x + \log(n)) = F^n(x + \log(n)) = F(x)$$

The distribution of the last example is the standard *Gumbel* distribution: its general form is $F_{\mu, \sigma}(x) = \exp(-\exp(-\frac{x-\mu}{\sigma}))$, with $\mu \in \mathbb{R}$ and $\sigma > 0$. This distribution has a main role in this theory, along with the *Fréchet* and the *Weibull* distributions, with cdf's $\exp(-(\frac{x-\mu}{\sigma})^{-\alpha})$ (for $x \geq \mu$) and $\exp(-(-\frac{x-\mu}{\sigma})^\alpha)$ (for $x \leq \mu$), where μ and σ are as before and $\alpha > 0$.

The starting point and keystone to the Extreme Value Theory is the Fisher-Tippett-Gnedenko theorem (see Theorem 1.1.1). The result as we know it today is due to Boris Gnedenko, who proved it in 1943 ([Gne43]). In this section we will expose one of the proofs and some of its implications. The theorem can be informally stated as follows:

Theorem 1.1.1 (Fisher-Tippett-Gnedenko theorem). *Let X_1, \dots, X_n be iid with distribution F and assume there exist sequences $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ such that $\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x) \xrightarrow[n \rightarrow \infty]{w} G(x)$, where $G(x)$ is a non-degenerate cdf. Then $G(x)$ is the cdf of a Gumbel, a Fréchet, or a Weibull distribution.*

Remark. As every distribution of the theorem is continuous in \mathbb{R} , then the convergence is actually pointwise.

Observe that the theorem already assumes that $\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow[n \rightarrow \infty]{w} G(x)$. We will also see necessary and sufficient conditions to determine if this happens. These conditions are, however, not always easy to apply.

1.1.1 Proof of the Theorem

We follow the first chapter of [LLR83]. We will first need to introduce several preliminary notions and results.

If F is an increasing right-continuous function, we define its *generalized inverse* with domain $(\inf\{F(x)\}, \sup\{F(x)\})$ (with one of the endpoints closed if the corresponding infimum/supremum is reached) as $F^{-1}(y) := \inf\{x : F(x) \geq y\}$. Note that if F is a cdf, then F^{-1} is the quantile function of the distribution. These are some basic properties of generalized inverse functions:

Lemma 1.1.2. *Let $F(x)$ be an increasing right-continuous function.*

- 1 *Let $a > 0, b, c \in \mathbb{R}$ and $H(x) = F(ax+b) - c$. Then $H(x)$ is increasing, right-continuous and $H^{-1}(y) = a^{-1}(F^{-1}(y+c) - b)$.*
- 2 *If F^{-1} is continuous, then for all $x \in \mathbb{R}$ $F^{-1}(F(x)) = x$.*
- 3 *If F is a non-degenerate cdf, then there exist $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $-\infty < F^{-1}(y_1) < F^{-1}(y_2) < \infty$.*
- 4 *If F is a non-degenerate cdf and there exist $a, \alpha > 0, b, \beta \in \mathbb{R}$ such that for all $x \in \mathbb{R}$ $F(ax+b) = F(\alpha x + \beta)$, then $a = \alpha$ and $b = \beta$.*

In our way to proving Fisher-Tippett-Gnedenko theorem we need an important result whose proof would deviate us from our topic of interest. Several preliminary results are needed and, as a consequence, we will just formulate it. For a complete proof we refer the reader to [CM90].

Proposition 1.1.3 (Law of convergence of types). *Let $\{F_n\}, G$ be cdf's, where G is non-degenerate, and $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ are sequences such that*

$$F_n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{w} G(x)$$

Then there exists a non-degenerate cdf G^ and sequences $\{\alpha_n > 0\}, \{\beta_n\}$ with*

$$F_n(\alpha_n x + \beta_n) \xrightarrow[n \rightarrow \infty]{w} G^*(x)$$

if and only if the limits $a := \lim_n a_n^{-1} \alpha_n$ and $b := a_n^{-1}(\beta_n - b_n)$ are finite and $a > 0$. Furthermore, when this happens, for all $x \in \mathbb{R}$ $G^(x) = G(ax + b)$.*

Definition 1.1.1. *Let G be a non-degenerate cdf. We say that:*

1. *G is max-stable if for all $n \in \mathbb{N}$ and for all x there exists $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ such that for all $x \in \mathbb{R}$ $G^n(a_n x + b_n) = G(x)$.*
2. *If F is a cdf, then F is in the domain of attraction (for maxima) of G , and it is written $F \in \mathcal{D}(G)$, when there exist sequences $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ such that $F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{w} G(x)$.*
3. *If $G^*(x)$ is another non-degenerate cdf, we say that G and G^* have the same type if for all x there exist $a > 0$ and $b \in \mathbb{R}$ such that for every $x \in \mathbb{R}$ $G^*(ax + b) = G(x)$.*

Remark. The Proposition 1.1.3 is also known as Khintchine's Theorem, and it can be rewritten in terms of the last definition as:

$F_n(a_n x + b_n) \xrightarrow[k \rightarrow \infty]{w} G(x)$ and $F_n(\alpha_n x + \beta_n) \xrightarrow[k \rightarrow \infty]{w} G^*(x)$ for sequences $\{a_n\} \subset \mathbb{R}^+$, $\{\alpha_n > 0\}$, $\{b_n\}$ and $\{\beta_n\}$ and non-degenerate cdf's $G(x)$ and $G^*(x)$ if and only if $G(x)$ and $G^*(x)$ have the same type, $\frac{a_n}{\alpha_n} \rightarrow a > 0$ and $\frac{\beta_n - b_n}{a_n} \rightarrow b \in \mathbb{R}$.

The next result is an immediate consequence of the Convergence of Types theorem. We will express it as a proposition to stress it.

Proposition 1.1.4.

- If $G_1(x)$ and $G_2(x)$ are two non-degenerate cdf's of the same type, then $\mathcal{D}(G_1) = \mathcal{D}(G_2)$.
- If $F_1(x)$ and $F_2(x)$ are two cdf's of the same type and $F_1 \in \mathcal{D}(G)$ for some non-degenerate cdf G , then $F_2 \in \mathcal{D}(G)$.

The converse of the former property holds, as easily seen, but not the other one. Later on, we will see that the exponential and normal families belong to the same domain of attraction and they are not of the same type.

Now, the theorem that we want to prove can be formulated as follows: if G is a non-degenerate cdf such that $\mathcal{D}(G) \neq \emptyset$, then G is the cdf of a distribution Gumbel, Fréchet or Weibull. We saw in the Example 1 that, if $F(x) = \exp(-\exp(-x))$ (i.e. a Gumbel distribution), then $F \in \mathcal{D}(F)$. Doing that, we also proved that F is max-stable. This is not a coincidence: the following result points out that these notions are, in a certain way, the same.

Proposition 1.1.5. *Given G a non-degenerate cdf:*

1. G is max-stable if and only if there exists a sequence $\{F_n\}$ of cdf's and sequences $\{a_n\} \subset \mathbb{R}^+$, $\{b_n\}$ such that for all $k \in \mathbb{N}$

$$F_n(a_{nk}^{-1}x + b_{nk}) \xrightarrow[n \rightarrow \infty]{w} G^{1/k}(x)$$

2. $\mathcal{D}(G) \neq \emptyset$ if and only if G is max-stable. In that case, $G \in \mathcal{D}(G)$.

Proof.

1. Let G be a max-stable cdf, and $\{a_n\} \subset \mathbb{R}^+$, $\{b_n\}$ sequences such that for all $n \in \mathbb{N}$: $G^n(a_n^{-1}x + b_n) = G(x)$. Let $F_n := G^n$. Then, for all $k \in \mathbb{N}$:

$$F_n^k(a_{nk}^{-1}x + b_{nk}) = G^{nk}(a_{nk}^{-1}x + b_{nk}) = G(x)$$

On the other hand, suppose there exist sequences $\{F_n\}$, $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ with

$$F_n(a_{nk}^{-1}x + b_{nk}) \xrightarrow[n \rightarrow \infty]{w} G^{1/k}(x)$$

for all $k \in \mathbb{N}$. Since G is non-degenerate, then so is $G^{1/k}$. Taking $k = 1$ and $k = 2$, we obtain

$$F_n(a_n^{-1}x + b_n) \xrightarrow[n \rightarrow \infty]{w} G(x) \text{ and } F_n(a_{2n}^{-1}x + b_{2n}) \xrightarrow[n \rightarrow \infty]{w} G^{1/2}(x) \text{ for every } x \in \mathbb{R}.$$

From Proposition 1.1.3, there exist $a_2 > 0$ and $b_2 \in \mathbb{R}$ such that $G(a_2x + b_2) = G^{1/2}(x)$. Analogously, for every k there exist $a_k > 0$ and $b_k \in \mathbb{R}$ with $G(a_kx + b_k) = G^{1/k}(x)$ or, equivalently, $G^k(a_kx + b_k) = G(x)$. Therefore, G is max-stable.

2. If G is max-stable, then there exist sequences $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ with $G^n(a_n x + b_n) = G(x)$, so $G \in D(G)$. Now assume $F \in D(G)$, and let $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ be the sequences corresponding to this condition. It is clear that for all $k \in \mathbb{N}$ we have $F^{nk}(a_{nk}x + b_{nk}) \xrightarrow[n \rightarrow \infty]{w} G(x)$, so

$$F^n(a_{nk}x + b_{nk}) \xrightarrow[n \rightarrow \infty]{w} G^{1/k}(x)$$

By the previous result, taking $F_n := F^n$ and $a_{nk}^{-1} := a_{nk}$ we conclude that G is max-stable. □

Corollary 1.1.6. *Let G be a max-stable cdf. Then there exist functions $a(s) > 0$ and $b(s)$ such that for all $x \in \mathbb{R}$, for all $s > 0$, $G^s(a(s)x + b(s)) = G(x)$.*

The Fisher-Tippett-Gnedenko theorem follows immediately from the equivalence between max-stable distributions and distributions with non-empty domain of attraction, along with the following result, which states that a max-stable distributions is either a Gumbel, a Fréchet or a Weibull distribution.

Theorem 1.1.7. *If G is a max-stable cdf, then G is of the same type as one of the following:*

$$\begin{aligned} \text{Type I} \quad G(x) &= \exp(-\exp(-x)) && (\text{Gumbel}) \\ \text{Type II} \quad G(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} && (\text{Fréchet}) \\ \text{Type III} \quad G(x) &= \begin{cases} \exp(-(-x)^\alpha) & x \leq 0 \\ 1 & x > 0 \end{cases} && (\text{Weibull}), \end{aligned}$$

where $\alpha > 0$. The converse is also true.

Proof. Let G be a max-stable cdf, and let $a(s) > 0$ and $b(s)$ functions such that $G^s(a(s)x + b(s)) = G(x)$, whose existence is guaranteed by Corollary 1.1.6. Taking logarithms:

$$-\log(-\log(G(a(s)x + b(s)))) - \log(s) = -\log(-\log(G(x))). \quad (1.1)$$

Let $\Psi(x) := -\log(-\log(G(x)))$, which is increasing and right-continuous, and $U(y) := \Psi^{-1}(y)$. It is a small exercise to prove that, since $G(x)$ is max-stable, then $U(y)$ is defined for every $y \in \mathbb{R}$.

Using Equation 1.1, we obtain $\Psi(a(s)x + b(s)) - \log(s) = \Psi(x)$, and now by the 1 in Lemma 1.1.2, $U(y) = \frac{U(y + \log(s)) - b(s)}{a(s)}$. Therefore $U(y) - U(0) = \frac{U(y + \log(s)) - U(\log(s))}{a(s)}$. If we now consider the change of variables $z := \log(s)$, $\tilde{a}(z) := a(\exp(z))$, $\tilde{U}(y) := U(y) - U(0)$, we get:

$$\tilde{U}(y) = \frac{\tilde{U}(y + z) - \tilde{U}(z)}{\tilde{a}(z)} \quad (1.2)$$

We consider two cases:

- $\tilde{a}(z) \equiv 1$. Then for all y , for all $z \in \mathbb{R}$ we have $\tilde{U}(y + z) = \tilde{U}(y) + \tilde{U}(z)$. It is easily proven (taking y_0 with $\tilde{U}(y_0) \neq 0$ and considering that $\{\frac{n_1}{n_2}y_0 : n_1, n_2 \in \mathbb{Z}\}$ is dense in \mathbb{R}) that $\tilde{U}(y) = \beta y$ for some $\beta > 0$, and therefore $U(y) = \Psi^{-1}(y) = \beta y + U(0)$. Since $\Psi(x)$ is continuous at a dense set, then $x = \Psi^{-1}(\Psi(x)) = \beta \Psi(x) + U(0)$, which allows us to conclude that $-\log(-\log(G(x))) = \Psi(x) = \frac{x - U(0)}{\beta}$ and $G(x) = \exp(-\exp(-\frac{x - U(0)}{\beta}))$. This proves that G is a cdf of Type I.

- $\tilde{a}(z) \neq 1$ for some z . Exchanging y and z in Equation 1.2 and subtracting both expressions, we get

$$\tilde{U}(y)(1 - \tilde{a}(z)) = \tilde{U}(z)(1 - \tilde{a}(y)) \Rightarrow \tilde{U}(y) = \frac{\tilde{U}(z)}{1 - \tilde{a}(z)}(1 - \tilde{a}(y))$$

Let $c = \frac{\tilde{U}(z)}{1 - \tilde{a}(z)}$. We claim that $c \neq 0$, or otherwise $U(y)$ would be constant, which is not possible because G is non-degenerate. Substituting the last expression in Equation 1.2 and simplifying, we obtain $\tilde{a}(y + z) = \tilde{a}(y)\tilde{a}(z)$. With a similar reasoning to the first case, there exists $\beta \neq 0$ such that $\tilde{a}(y) = \exp(\beta y)$. Therefore, $\Phi^{-1}(y) = U(y) = c(1 - \exp(\beta y)) + U(0)$, and with easy manipulations we finally get

$$G(x) = \exp\left(-\left(1 - \frac{x - U(0)}{c}\right)^{-1/p}\right) \quad \left(1 - \frac{x - U(0)}{c} > 0\right)$$

Assume $p > 0$. Since $U(y)$ is increasing, then $c < 0$ (and viceversa). If we change $\alpha := -1/p$, $z := -cx + U(0) + c$, then $G(z) = \exp(-z^\alpha)$, and G is Type II. Analogously, if $p < 0$ then G is Type III.

□

It is clear how the Fisher-Tippett-Gnedenko theorem (Theorem 1.1.1) follows from the previous results. It implies that, for n large enough, if $F \in \mathcal{D}(G)$, then we can approximate the maximum by a distribution G^* of the same type as G :

$$\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x) \approx G(x).$$

Thus $\mathbb{P}(M_n \leq x) \approx G\left(\frac{x - b_n}{a_n}\right) = G^*(x)$.

1.1.2 Domains of attraction and examples

As observed in the beginning, the Fisher-Tippett-Gnedenko theorem only characterizes distributions with a non-empty domain of attraction. A different problem is to determine if a given distribution is in one of the three domains of attraction, and if so, to determine in which one. We will now give necessary and sufficient conditions to guarantee that this happens. We recall the notation $x_F = \sup\{x \in \mathbb{R} \cup \{\infty\} : F(x) < 1\}$.

Theorem 1.1.8 ([LLR83], Theorem 1.6.2). *Let $F(x)$ be a cdf. Then F belongs to the domain of attraction of Type I, II or III if and only if, respectively,*

i There exists a positive function g such that

$$\text{for all } x \in \mathbb{R} \quad \lim_{t \uparrow x_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}.$$

ii $x_F = \infty$ and for some $\alpha > 0$

$$\text{for all } x > 0 \quad \lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}$$

iii $x_F < \infty$ and for some $\alpha > 0$

$$\text{for all } x > 0 \quad \lim_{h \downarrow 0} \frac{1 - F(x_F - xh)}{1 - F(x_F - h)} = x^\alpha$$

Furthermore:

1. if $\gamma_n := F^{-1}(1 - \frac{1}{n})$, the following are possible elections of the constants a_n and b_n :

	Type I	Type II	Type III
a_n	$g(\gamma_n)^{-1}$	γ_n^{-1}	$(x_F - \gamma_n)^{-1}$
b_n	γ_n	0	x_F

2. if F belongs to the domain of attraction of Type I, then $\int_0^\infty (1 - F(u))du < \infty$ and one choice of g is $g(t) = \int_t^{x_F} \frac{1-F(u)}{1-F(t)} du$.

Remark.

1. If X is a non-negative random variable with cdf F , it is well known that $\mathbb{E}[X] = \int_0^{x_F} (1 - F(u))du$. If F belongs to the domain of attraction of Type I, then, as a consequence of Theorem 1.1.8, $\mathbb{E}[X] < \infty$ and $g(t) = \mathbb{E}[X|X \geq t]$, where $g(t)$ is the same as in the theorem.
2. Observe that $\gamma_n = F^{-1}(1 - \frac{1}{n})$ is exceeded with probability $\frac{1}{n}$, that is, γ_n is the return level associated to a period of n units.

Note that, as a consequence of Proposition 1.1.3, the constants are not unique. Theorem 1.1.8 tells us something which could be expected, and it is that the only thing influencing that a distribution belongs to some domain of attraction is the right tail, and the behaviour of the remaining part is irrelevant. Some examples are given below:

Example 2 (Exponential distribution). The exponential distribution belongs to the Type I attraction domain. Let $F(x) = 1 - e^{-x}$ be the cdf of the standard exponential distribution, and $g(t)$ such that $\lim_{t \rightarrow \infty} g(t) = 1$ (for example, $g(t) = 1$). Then:

$$\lim_{t \rightarrow \infty} \frac{1 - F(t + xg(t))}{1 - F(t)} = \lim_{t \rightarrow \infty} \frac{e^{-t-xg(t)}}{e^{-t}} = \lim_{t \rightarrow \infty} e^{-xg(t)} = e^{-x}.$$

Furthermore, $F^{-1}(y) = -\log(1 - y)$, and $\gamma_n = F^{-1}(1 - \frac{1}{n}) = -\log(\frac{1}{n}) = \log(n)$.

Example 3 (Cauchy distribution). It is well known that $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x)$ is the cdf of the standard Cauchy distribution. It can easily be seen that it is in the Type II attraction domain because

$$\lim_{t \rightarrow \infty} \frac{\arctan(tx) - \frac{\pi}{2}}{\arctan(t) - \frac{\pi}{2}} = x^{-1}.$$

And we obtain $\alpha = 1$. Besides, $F^{-1}(y) = \tan(\pi(y - \frac{1}{2}))$, and $\gamma_n = F^{-1}(1 - \frac{1}{n}) = \tan(\pi \frac{n-2}{2n})$. On the other hand, $b_n = 0$, so:

$$\mathbb{P}(\gamma_n^{-1} M_n \leq x) \xrightarrow{n \rightarrow \infty} e^{-x^{-1}} \quad (x > 0)$$

Example 4. Let $K, \alpha > 0, x_F \in \mathbb{R}$ and $F(x) = 1 - K(x_F - x)^\alpha$ defined in $[x_F - K^{-1/\alpha}, x_F]$. It is easy to check that it is in the Type III attraction domain, because:

$$\lim_{h \rightarrow 0} \frac{1 - F(x_F - xh)}{1 - F(x_F - h)} = \frac{K(x_F - x_F + xh)^\alpha}{K(x_F - x_F + h)^\alpha} = x^\alpha.$$

We also have that $F^{-1}(y) = x_F - \left(\frac{1-y}{K}\right)^{1/\alpha}$, and $\gamma_n = F^{-1}(1 - \frac{1}{n}) = x_F - \frac{1}{nk^{1/\alpha}}$, so we have that:

$$a_n = (x_F - \gamma_n)^{-1} = (nk)^{\frac{1}{\alpha}}.$$

Example 5 (Uniform distribution). Let $a < b$ and consider the uniform distribution with parameters a and b . The cdf is $F(x) = \frac{x-a}{b-a}$ in its support. Thus the uniform distribution can be reduced to the previous example: the theorem holds with $\alpha = 1$ and a possible election of the constants is

$$a_n = (x_F - \gamma_n)^{-1} = \frac{n}{a+b} \text{ and } b_n = b$$

Many times the computations involved in Theorem 1.1.8 are complex and it is not easy to compute the corresponding limit. An example of this is the normal distribution, which (as shown in [LLR83, Theorem 1.5.3] by other means) belongs to the Type I domain of attraction. A possible choice of the constants is:

$$a_n = (2 \log(n))^{-\frac{1}{2}} \text{ and } b_n = (2 \log(n))^{\frac{1}{2}} - \frac{1}{2}(2 \log(n))^{-\frac{1}{2}}(\log(\log(n)) + \log(4\pi))$$

There are, however, distributions that do not belong to any domain of attraction. In other words, there are cdf's $F(x)$ where no pair of sequences $\{a_n\} \subset \mathbb{R}^+$, $\{b_n > 0\}$ can be chosen, so that $F^n(a_n x + b_n)$ converges to a non-degenerate distribution. Moreover, some examples of such distributions are very simple, and it is easy to come across them in some problem. The next few results will provide a condition for a distribution not to belong to any domain of attraction.

Proposition 1.1.9 ([LLR83], Theorem 1.5.1). *Let X_1, \dots, X_n be iid with distribution F . Let $\tau \in [0, \infty]$ and assume there exists a sequence $\{u_n\}$ such that*

$$n(1 - F(u_n)) \xrightarrow[n \rightarrow \infty]{} \tau.$$

Then, $\mathbb{P}(M_n \leq u_n) \xrightarrow[n \rightarrow \infty]{} \exp(-\tau)$. The converse also holds (with $\exp(-\infty) := 0$).

Proposition 1.1.10 ([LLR83], Theorem 1.7.13). *Let $F(x)$ be a cdf, $F(x-) := \lim_{t \uparrow x} F(t)$, and X_1, \dots, X_n be iid with distribution F . If $\tau \in (0, \infty)$, then there exists a sequence $\{u_n\}$ such that*

$$\mathbb{P}(M_n \leq u_n) \xrightarrow[n \rightarrow \infty]{} \exp(-\tau)$$

if and only if

$$\lim_{x \uparrow x_F} \frac{F(x) - F(x-)}{1 - F(x-)} = 0 \tag{1.3}$$

In particular, the results just shown ensure that if the above limit is different than 0, then $\mathbb{P}(M_n \leq u_n)$ converges to either 0 or 1. As a consequence F does not belong to any domain of attraction, because the limit in Equation 1.3 does not exist.

Let $X \sim F$ be an integer-valued random variable such that $x_F = \infty$. Then $p(n) = F(n) - F(n-) = \mathbb{P}(X = n)$. If

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}(X = n)}{1 - \mathbb{P}(X \leq n-1)} = \lim_{n \rightarrow \infty} \frac{\mathbb{P}(X = n)}{\mathbb{P}(X \geq n)} \neq 0,$$

then F does not belong to any domain of attraction.

Example 6. The Poisson distribution with parameter $\lambda > 0$ is an example of such distribution.

Recall that the Poisson distribution is defined as $P(n) = e^{-\lambda} \frac{\lambda^n}{n!}$. Therefore, we have

$$\frac{\mathbb{P}(X = n)}{\mathbb{P}(X \geq n)} = \frac{\frac{\lambda^n}{n!}}{\sum_{i=n}^{\infty} \frac{\lambda^i}{i!}} = \frac{1}{1 + \sum_{r=n+1}^{\infty} \frac{n!}{r!} \lambda^{r-n}} \xrightarrow[n \rightarrow \infty]{} 1.$$

Note that the series goes to 0 because

$$\sum_{r=n+1}^{\infty} \frac{n!}{r!} \lambda^{r-n} \leq \sum_{r=1}^{\infty} \left(\frac{\lambda}{n}\right)^r \xrightarrow{n \rightarrow \infty} 0.$$

Example 7. The geometric distribution with parameter $p \in (0, 1)$ does not belong to any domain of attraction.

$$\frac{\mathbb{P}(X = n)}{\mathbb{P}(X \geq n)} = \frac{(1-p)^{n-1}p}{\sum_{k=n}^{\infty} (1-p)^{k-1}p} = \frac{(1-p)^{n-1}}{\frac{(1-p)^{n-1}}{p}} = p.$$

1.1.3 Generalized Extreme Value distribution

Gumbel, Fréchet and Weibull distributions, also known as extreme value distributions, can all be grouped within a single family. This has been called Generalized Extreme Value Distribution (GEV distribution), and its cdf is

$$G_{\xi, \mu, \sigma}(x) = \begin{cases} \exp(-e^{-\frac{x-\mu}{\sigma}}) & \xi = 0 \\ \exp\left(-\left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) & \xi \neq 0, 1 + \xi \frac{x-\mu}{\sigma} > 0 \\ 0 & \xi > 0, \frac{x-\mu}{\sigma} \leq -\frac{1}{\xi} \\ 1 & \xi < 0, \frac{x-\mu}{\sigma} \geq -\frac{1}{\xi} \end{cases} \quad (1.4)$$

The quantile and density functions are, respectively:

$$G^{-1}(y) = \begin{cases} \mu + \frac{\sigma}{\xi} ((-\log(y))^{-\xi} - 1) & \xi \neq 0 \\ \mu - \sigma \log(-\log(y)) & \xi = 0 \end{cases} \quad (1.5)$$

$$g_{\xi, \mu, \sigma}(x) = \begin{cases} \exp(-e^{-\frac{x-\mu}{\sigma}}) e^{-\frac{x-\mu}{\sigma}} \frac{1}{\sigma} & \xi = 0 \\ \exp\left(-\left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) & \xi \neq 0 \end{cases} \quad (1.6)$$

whenever the above expressions make sense. Figure 1.1 shows the pdf and the cdf of distributions of such family.

It is clear that if $\xi = 0$, $G_{0, \mu, \sigma}$ belongs to the Gumbel family. Furthermore, if $\xi > 0$ or $\xi < 0$, then $G_{\xi, \mu, \sigma}$ belongs to the Fréchet or Weibull family, respectively. There is a very clear advantage in merging these distributions under just one: in a practical problem, where we do not know the distribution our data came from. The first step would be choosing one of the three types, using some rule or technique.

With the GEV family, the type is determined by ξ parameter. If the estimator for ξ allows us, for example, to test hypotheses about its value, or to compute confidence intervals, then we can use these tools to assess which of the families best fits our data. In the next section we will discuss estimators for ξ , μ and θ .

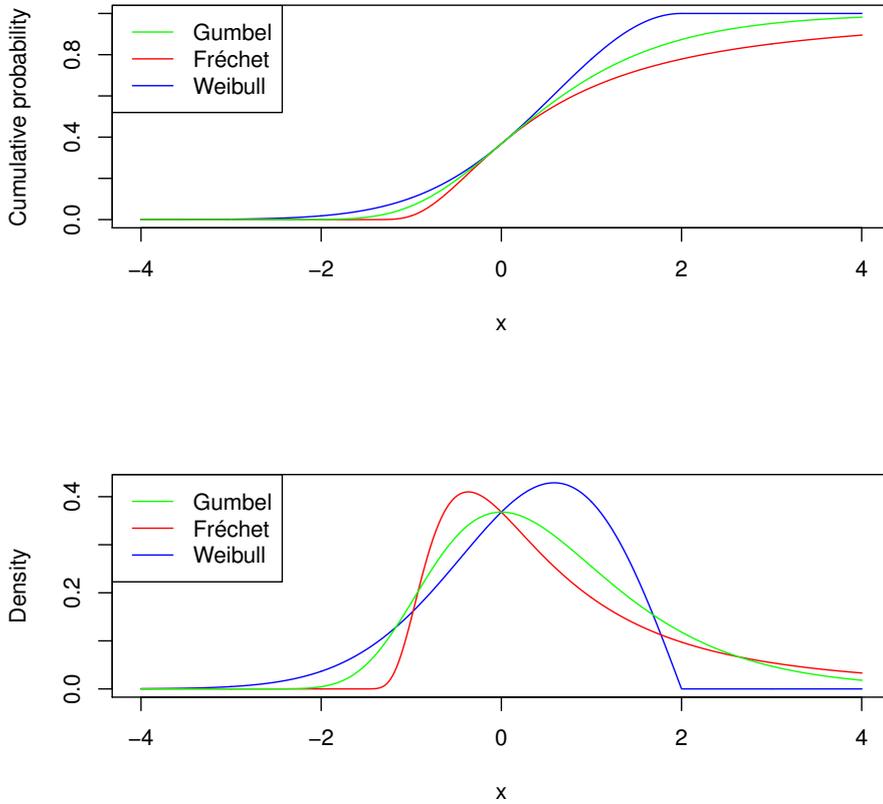


Figure 1.1: cdf and density function of the GEV distribution with parameters $\xi \in \{0, \frac{1}{2}, -\frac{1}{2}\}$, $\mu = 0$, $\sigma = 1$.

Remark. The definition of the GEV distribution can also be found in the literature exchanging ξ for $-\xi$. Consequently, all the expressions and formulas derived from this are slightly different, so special caution should be taken.

1.2 Estimation of the parameters

We can talk about two kinds of methods:

- *The block maxima approach.* Let us assume we have a sample X_1, \dots, X_n be iid with distribution F , where F is unknown but it is assumed that it belongs to a domain of attraction. The most rudimentary option would be to divide this collection into m blocks of the same size that do not overlap, and consider a new collection $\{Z_i\}_{i=1}^m$ where each Z_i is the maximum of the i -th block. Then we would get a new collection of iid rv's Z_1, \dots, Z_m and we assume that their common distribution is $G_{\xi, \mu, \sigma}(x)$ for some ξ , $\mu \in \mathbb{R}$ and $\sigma > 0$. That is, the approximation of the Fisher-Tippett-Glivenko Theorem is considered exact. From here, (ξ, μ, σ) can be estimated with any of the available methods.
- *The exceedance approach.* It may seem that potentially relevant information is lost if only the maximum is used. Let $u \in \mathbb{R}$ so that every $X_i > u$ could be considered an

extreme value. A second option is to study the distribution of the exceedances, $X_i - u$, when $X_i > u$. We will see that if F is in some domain of attraction, then we have an approximation of the distribution of the excesses that is very manageable.

In the end, our purpose is a problem of estimating quantiles: given a small p , we want to compute x_p so that $x_p = F^{-1}(1 - p)$, or the other way around. In the case of the GEV and the Generalized Pareto distributions (which will be introduced in Subsection 1.2.2), the quantile function can easily be expressed in terms of the parameters of the corresponding distributions. On the other hand, we want the estimators we obtain to have desirable properties; for example, to be asymptotically normal. This not only allows us to obtain confidence intervals for the estimators, but also for the quantile itself. The justification for this is found in the Delta method, which, roughly speaking, follows from Taylor expansion.

Theorem 1.2.1 (Delta method). ([Vaa98, Theorem 3.1]) *Let $\Phi : D \subset \mathbb{R}^k \rightarrow \mathbb{R}^d$ be a differentiable mapping at $\theta \in D$. If $\{T_n\} \subset D$, $\{r_n\} \subset \mathbb{R}$ are sequences such that $r_n(T_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} T$, where $r_n \rightarrow \infty$, then $r_n(\Phi(T_n) - \Phi(\theta)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \Phi'(\theta)T$.*

We also recall the definition of a consistent estimator:

Definition 1.2.1. $\hat{\theta}_n$ is a strongly (resp. weakly) consistent estimator of θ if $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{} \theta$ almost surely (resp. in probability).

1.2.1 Block Maxima approach

This method was the first, historically speaking, and continues to be widely used. Sometimes their use is mandatory: it may be that the data available are precisely maxima of blocks of maxima of some process. Moreover, in many cases, the nature of the problem leads to this: blocks may correspond to natural periods of time of space. However, when this is not the case, the choice of block size can be problematic: a very small block causes a large bias, because the approximation by $G_{\xi, \mu, \sigma}$ requires a sufficiently large sample size. On the other hand, a large block provides little data to make relevant estimates, bringing very high variances.

In the following lines we will describe the most popular procedures based on Block Maxima. From now on, assume $\{X_i\}_{i=1}^n$ is iid with distribution $G_{\xi, \mu, \gamma}$, for unknown (ξ, μ, σ) .

Maximum Likelihood Estimator

Let $\theta := (\xi, \mu, \sigma)$ and let g_θ be the density function of the GEV distribution (Equation 1.6). Recall that $\hat{\theta}_n^{MLE}$ is a maximum likelihood estimator (MLE) if it maximizes $\prod_{i=1}^n g_\theta(X_i)$ or, equivalently, $\sum_{i=1}^n \log(g_\theta(X_i))$. In the case of the GEV distribution, the last expression (the logarithm of the likelihood function, or the log-likelihood function) is:

$$l_{(\xi, \mu, \sigma)}(X_1, \dots, X_n) = \begin{cases} -n \log(\sigma) - \frac{\xi+1}{\xi} \sum_{i=1}^n \left[\log \left(1 + \frac{\xi}{\sigma} (X_i - \mu) \right) - \left(1 + \frac{\xi}{\sigma} (X_i - \mu) \right)^{-1/\xi} \right] & \xi \neq 0 \\ -n \log(\theta) - \sum_{i=1}^n \left[\exp \left(-\frac{X_i - \mu}{\sigma} \right) - \frac{X_i - \mu}{\sigma} \right] & \xi = 0 \end{cases} \quad (1.7)$$

It is well known that under some smoothness conditions on the variation of the log-likelihood with respect to (ξ, μ, σ) (usually known as 'regularity conditions'), the MLE estimator is strongly consistent and asymptotically normal:

$$\hat{\theta}_n^{MLE} \xrightarrow[n \rightarrow \infty]{c.s.} \theta, \quad \sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, I_{\theta_0}^{-1}), \quad \text{where } I_{\theta_0} = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \log(g_\theta(x)) \mid \theta = \theta_0 \right].$$

These regularity conditions do not hold in the GEV distribution, and the situation becomes problematic. For example, it is always the case that $\hat{\xi}_n^{MLE} > -1$, so there is no consistent estimator when $\hat{\xi} < -1$ (see [Dom15, Remark 4]). The classical regularity conditions, however, can be relaxed, and the same results can be obtained for some subfamilies of the GEV distribution. The first results can be found in [Smi85]. A full and detailed discussion is in [BS17].

Theorem 1.2.2. ([BS17, Proposition 3.3]) *The MLE is strongly consistent and asymptotically normal whenever $\xi > -\frac{1}{2}$.*

Furthermore, if $\Gamma(x)$ denotes the Gamma function, $\Psi(x) := \log(\Gamma(x))'$, $\gamma := -\Gamma'(1)$ is the Euler-Mascheroni constant, $p := (1 + \xi)^2\Gamma(1 + 2\xi)$ and $q := \Gamma(2 + \xi) \left(\Psi(1 + \xi) + \frac{1+\xi}{\xi} \right)$, the entries of I_θ are ([BGT04, Appendix 5.9.1]):

$$\begin{aligned} I_\theta(1, 1) &= \frac{1}{\sigma^2\xi^2}(1 - 2\Gamma(2 + \xi) + p) \\ I_\theta(1, 2) &= -\frac{1}{\sigma\xi^2} \left(1 - \gamma - q + \frac{1 - \Gamma(2 + \xi)}{\xi} + \frac{p}{\xi} \right) \\ I_\theta(1, 3) &= -\frac{1}{\sigma^2\xi}(p - \Gamma(2 + \xi)) \\ I_\theta(2, 2) &= \frac{1}{\xi^2} \left[\frac{\pi^2}{6} + \left(1 - \gamma + \frac{1}{\xi} \right) - \frac{2q}{\xi} + \frac{p}{\xi^2} \right] \\ I_\theta(2, 3) &= -\frac{1}{\sigma\xi} \left(q - \frac{p}{\xi} \right) \\ I_\theta(3, 3) &= \frac{p}{\sigma^2} \end{aligned}$$

Therefore, in order to get the approximated distribution of the MLE, we would replace the inverse of I_θ for its estimated using the plug-in method. With this we can, for example, compute confidence intervals or carry out test hypothesis.

The properties of the MLE (asymptotically unbiased, asymptotically efficient or lowest variace) makes it the preferred method for some people whenever the size sample is large enough. When this is not the case, some authors prefer to use other methods.¹ However, the method that we will show next is one of the most popular estimation alternatives, especially in hydrology.

Method of L -moments

If X is a rv, then we recall that the r -th moment, the r -th centered moment and the r -th standardized moment are, respectively:

$$\mathbb{E}(X^r), \quad \mu_r := \mathbb{E}((X - \mathbb{E}(X))^r), \quad \frac{\mu_r}{\mu_2^{r/2}}.$$

¹There are modifications of the MLE that supposedly solve the problems reported for the MLE when the sample size is not large, but we will not delve into it. See Coles and Dixon: *Likelihood-Based Inference for Extreme Value Models* (1990) <https://doi.org/10.1023/A:1009905222644>, where they propose a penalizing factor to the log-likelihood function.

Moments are a way of describing a distribution. The first moment (the mean) gives an idea of the location of the distribution; the second centered moment (the variance) measures the dispersion of the distribution with respect to the mean; the third standardized moment (skewness) measures the asymmetry of the distribution, and the fourth standardized moment (kurtosis) measures how concentrated the data are in the tails. These are the moments that are commonly used when describing a distribution, although there are also interpretations for moments of larger orders.

Historically, the first parameter estimation method is *the method of moments*: if we have k parameters to estimate, then the first k moments are matched to the population ones, obtaining a system of equations according to the parameters. Solving this system we obtain an estimate of the parameters. One of the advantages of this method is that it is simple to calculate: the only complication is in solving a system of equations, which is done with numerical techniques in most cases. However, it has several disadvantages: for example, sometimes relevant moments do not exist, or the estimators obtained by this method may not belong to the parametric space, or it may be very biased.

Fortunately, moments are not the only descriptions that exist of a distribution. We will introduce the L -moments, and then the estimation with them, which has the same spirit as the method of moments. The L -moments are linear combinations of expectations of the order statistics.

Definition 1.2.2. *We define the r -th L -moment as follows:*

$$\lambda_r := \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathbb{E}(X_{r-j:r}),$$

where $X_{r-j:r}$ is the $(r-j)$ -th order statistic from a sample of size r . We also define the r -th standardized L -moment as $\tau_r := \frac{\lambda_r}{\lambda_2}$.

Remark. The L -moments can also be seen as linear combinations of Probability Weighted Moments (P.W.M.), which are extensions of moments. Define the quantities:

$$M_{p,r,s} = \mathbb{E}(X^p F(X)^r (1 - F(X))^s), \quad \alpha_k = M_{1,0,k} \text{ and } \beta_k = M_{1,k,0},$$

where $X \sim F$. Note that $M_{p,0,0}$ is the usual p -th moment. It can be checked that:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r p_{r,k}^* \alpha_k = \sum_{k=0}^r p_{r,k}^* \beta_k, \text{ where } p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}. \quad (1.8)$$

We will use the second equality to estimate the L -moments, and we will compute the parameters of the GEV distribution from the estimation of the L -moments. This was done for the first time in [HWW85].

The use of L -moments instead of usual moments is not justified if it is not demonstrated that they have some advantage over them. Furthermore, we want to be able to interpret them (at least) as well as conventional moments. The following result solves the first problem:

Theorem 1.2.3 ([Hos90], Theorem 1). *If the mean of a distribution is finite, then every L -moment exists and it is finite. Furthermore, if this happens, the distribution is characterized by its L -moments.*

It is well known that this property is very far from happening to traditional moments. For example, the distributions defined by the following density functions:

$$f_0(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{\log^2(x)}{2}\right) \text{ and } f_a(x) = f_0(x)(1 + a \sin(2\pi \log(x))) \quad (-1 \leq a \leq 1)$$

have the same moments ([Dur19, Section 3.3.5]).

On the other hand, the first L -moments are not hard to interpret. Clearly, λ_1 is the mean (or L -location). Moreover:

- $\lambda_2 = \frac{1}{2}(\mathbb{E}(X_{2:2}) - \mathbb{E}(X_{1:2}))$, or L -scale: the more disperse the distribution is, the bigger this difference will be. Therefore, this gives an idea of the dispersion of the distribution. Besides, note that λ_2 is always positive (whenever the distribution is non-degenerate). This is not evident from the definition via P.W.M., which is how they are sometimes defined.
- $\lambda_3 = \frac{1}{3}(\mathbb{E}(X_{3:3}) - 2\mathbb{E}(X_{2:3}) + \mathbb{E}(X_{1:3})) = \frac{1}{3} \cdot \mathbb{E}[(X_{3:3} - X_{2:3}) + (X_{1:3} - X_{2:3})]$: if the left tail of the distribution is heavier than the right tail, then $X_{1:3}$ will generally be further from $X_{2:3}$ than $X_{2:3}$ from $X_{3:3}$, and therefore $|\mathbb{E}(X_{1:3} - X_{2:3})| > |\mathbb{E}(X_{2:3} - X_{3:3})|$, which means $\lambda_3 < 0$. On the contrary, if the right tail is heavier, then $\lambda_3 > 0$; if the distribution is symmetric, then $\lambda_3 = 0$. As a consequence, we can think of λ_3 as a measure of the asymmetry of the distribution, and so is τ_3 (or L -skewness).
- $\lambda_4 = \frac{1}{4}(\mathbb{E}(X_{4:4}) - 3\mathbb{E}(X_{3:4}) + 3\mathbb{E}(X_{2:4}) - \mathbb{E}(X_{1:4})) = \frac{1}{4} \cdot \mathbb{E}[X_{4:4} - X_{1:4} - 3(X_{3:4} - X_{2:4})]$: this compares the separation of the extreme order statistics with respect to the central order statistics. This interpretation is somewhat hard to do. For example, if X is a uniform random variable, $\mathbb{E}(X_{i:4} - X_{i-1:4}) = \mathbb{E}(X_{j:4} - X_{j-1:4})$ whenever this expression make sense for i and j , and thus $\lambda_4 = 0$. This is the limit case, in which the distribution is 'flat' (its density function, in its support, is constant). The further the distribution is from being flat, the further from 0 this quantity is. τ_4 is known as the L -kurtosis.

We have just seen how the interpretation of L -moments is not worse than that of the usual moments, excepting the difficulty of reading L -kurtosis.² In fact, it could be argued that for this purpose the L -kurtosis is better: for example, for all $r \geq 3$ τ_r is bounded in absolute value by 1 (see [Hos90, Theorem 2]). This allows a better interpretation of the L -skewness compared to the skewness, which is not bounded.

As previously mentioned, the estimators of λ_r will be derived from the estimators of β_k , applying the plug-in method to Equation 1.8:

$$\hat{\lambda}_{r+1} = \sum_{k=0}^r p_{r,k} \hat{\beta}_k. \quad (1.9)$$

An unbiased estimator for β_k is

$$\hat{\beta}_k := \frac{1}{n} \sum_{j=k+1}^n \left(\prod_{s=1}^k \frac{(j-s)}{(n-s)} \right) x_{j:n}. \quad (1.10)$$

The following estimator, asymptotically biased, is also reported to perform well (see [HWW85]):

$$\hat{\beta}_k[p_{j,n}] = \frac{1}{n} \sum_{j=1}^n p_{j,n} x_{j:n}, \quad \text{where } p_{j,n} = \frac{j-\alpha}{n+\beta} \quad (0 < \alpha < 1).$$

²However, the interpretation of the usual kurtosis is not very clear either.

It has been found that whenever estimating the parameters of the GEV distribution and then estimating high quantiles if the size sample is not large, estimation using P.W.M. (or L -moments: they enjoy the same properties) performs better than MLE (see, for example, [HWW85]). It has been reported that the best results were obtained when taking $\alpha = 0.35$, $\beta = 0$. This result was obtained through simulations (again, see [HWW85]).

However, L -moments do not exist whenever $\xi > 1$, and they should not be used if there is evidence that this is the case (in the same way that MLE estimation should not be used if it may be the case than $\xi < -1$).

We conclude this section with the equations of the first L -moments for the GEV family. They have been extracted from [HW97, Appendix A6]. If $\xi \neq 0$:

$$\lambda_1 = \mu - \frac{\sigma}{\xi}(1 - \Gamma(1 - \xi)) \quad (1.11)$$

$$\lambda_2 = -\frac{\sigma}{\xi}(1 - 2^\xi)\Gamma(1 - \xi) \quad (1.12)$$

$$\tau_3 = 2 \cdot \frac{1 - 3^\xi}{1 - 2^\xi} - 3 \quad (1.13)$$

$$\tau_4 = \frac{5(1 - 4^\xi) - 10(1 - 3^\xi) + 6(1 - 2^\xi)}{1 - 2^\xi}. \quad (1.14)$$

From the estimator of τ_3 , $\hat{\xi}$ can be computed, and with the relations

$$\sigma = -\frac{\lambda_2 \hat{\xi}}{(1 - 2^{\hat{\xi}})\Gamma(1 - \hat{\xi})} \text{ and } \mu = \lambda_1 + \frac{\sigma}{\hat{\xi}}(1 - \Gamma(1 - \hat{\xi}))$$

the rest of the estimators can be obtained. In [HWW85] the authors proposed the approximation

$$\xi \approx -7.8590c - 2.9554c^2, \quad \text{where } c = \frac{2}{3 + \tau_3} - \frac{\log(2)}{\log(3)},$$

which has an absolute error with an upper bound $9 \cdot 10^{-4}$ whenever $\xi \in [-\frac{1}{2}, \frac{1}{2}]$. This approximation avoids solving the system using numerical methods, and as a consequence, computing the estimators is very fast. On the other hand, when $\xi = 0$, the expressions above should be understood as the limit when $\xi \rightarrow 0$. The result is:

$$\lambda_1 = \mu + \sigma\gamma \quad (1.15)$$

$$\lambda_2 = \sigma \log(2) \quad (1.16)$$

$$\tau_3 = \frac{\log(9) - \log(8)}{\log(2)} \quad (1.17)$$

$$\tau_4 = \frac{16 \log(2) - 10 \log(3)}{\log(2)}. \quad (1.18)$$

The estimation via L -moments is asymptotically normal if $\xi < \frac{1}{2}$ (see [HWW85]). Again, observe the resemblance with the MLE, which is asymptotically normal for $\xi > -\frac{1}{2}$. The expression of the covariance matrix of the estimators via L -moments (denoted by $(\hat{\mu}^L, \hat{\sigma}^L, \hat{\xi}^L)$) is very difficult to describe. It is derived from the covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$, which can be found in [HWW85, Appendix C], and then using the Delta method. It has the form

$$\frac{1}{n} \begin{bmatrix} \sigma^2 w_{11} & \sigma^2 w_{12} & \sigma w_{13} \\ \sigma^2 w_{12} & \sigma^2 w_{22} & \sigma w_{23} \\ \sigma w_{13} & \sigma w_{23} & w_{33} \end{bmatrix},$$

where w_{ij} depend only on ξ . Some values are shown below ([HWW85, Table 1]):

ξ	w_{11}	w_{12}	w_{13}	w_{22}	w_{23}	w_{33}
0.4	1.6637	1.3355	1.1405	1.8461	1.1628	2.9092
0.3	1.4153	0.8912	0.5640	1.2574	0.4442	1.4090
0.2	1.3322	0.6727	0.3926	1.0013	0.2697	0.9139
0.1	1.2915	0.5104	0.3245	0.8440	0.2240	0.6815
0	1.2686	0.3704	0.2992	0.7390	0.2247	0.5633
-0.1	1.2551	0.2411	0.2966	0.6708	0.2447	0.5103
-0.2	1.2474	0.1177	0.3081	0.6330	0.2728	0.5021
-0.3	1.2438	-0.0023	0.3297	0.6223	0.3033	0.5294
-0.4	1.2433	-0.1205	0.3592	0.6368	0.3329	0.5880

Elemental Percentile Method

The Elemental Percentile Method (EPM) was first introduced for the GEV family in [CH94], although it can be used for other distributions, as we will see. It can be described as follows:

1. For every combination of indices $i < j < k \in \{1, \dots, n\}$:
 - (a) Match empirical and theoretical quantiles for i, j, k , that is, consider the system $\{G_{\xi, \mu, \sigma}(X_{l:n}) = p_{l,n} : l = i, j, k\}$, where $p_{l,n} = \frac{l-\alpha}{n+\beta}$, and $0 < \alpha < 1$ and $\beta \geq 0$.
 - (b) Compute solutions $\hat{\xi}_{i,j,k}$, $\hat{\mu}_{i,j,k}$ and $\hat{\sigma}_{i,j,k}$ of the system (most of the times, via numerical methods).
2. Apply a centralization and robust function to the collection of estimators and obtain a final estimator $\hat{\theta}^{EPM} := \{\hat{\xi}^{EPM}, \hat{\sigma}^{EPM} \hat{\mu}^{EPM}\}$.

Given $i < j < k \leq n$, the system to be solved (if $\xi \neq 0$) for $\xi_{i,j,k}$, $\mu_{i,j,k}$ and $\sigma_{i,j,k}$ is:

$$\begin{cases} p_{i:n} = \exp\left(-\left(1 + \frac{\xi_{i,j,k}}{\sigma_{i,j,k}}(x_{i:n} - \mu_{i,j,k})\right)^{-1/\xi_{i,j,k}}\right) \\ p_{j:n} = \exp\left(-\left(1 + \frac{\xi_{i,j,k}}{\sigma_{i,j,k}}(x_{j:n} - \mu_{i,j,k})\right)^{-1/\xi_{i,j,k}}\right) \\ p_{k:n} = \exp\left(-\left(1 + \frac{\xi_{i,j,k}}{\sigma_{i,j,k}}(x_{k:n} - \mu_{i,j,k})\right)^{-1/\xi_{i,j,k}}\right). \end{cases}$$

Eliminating $\mu_{i,j,k}$ and $\sigma_{i,j,k}$, we obtain:

$$\frac{x_j - x_k}{x_i - x_k} = \frac{(-\log(p_{j:n}))^{-\xi} - (-\log(p_{k:n}))^{-\xi}}{(-\log(p_{i:n}))^{-\xi} - (-\log(p_{k:n}))^{-\xi}} = \frac{1 - A_{jk}^{-\xi}}{1 - A_{ik}^{-\xi}},$$

where $A_{ik} = \frac{\log(p_{i:n})}{\log(p_{k:n})}$. The rest of the parameters can be computed using the relations:

$$\sigma_{i,j,k} = \xi_{i,j,k} \cdot \frac{x_{i:n} - x_{j:n}}{(-\log(p_{i:n}))^{-\xi} - (-\log(p_{j:n}))^{-\xi}} \text{ and } \mu_{i,j,k} = x_{i:n} + \frac{\sigma_{i,j,k}}{\xi_{i,j,k}}(1 - (-\log(p_{i:n}))^{-\xi}).$$

On the other hand, if there is evidence that $\xi = 0$, then it is enough to consider two indices $i < j$ and the system is:

$$\begin{cases} p_{i:n} = \exp\left(-e^{-\frac{x_{i:n} - \mu_{i,j}}{\sigma_{i,j}}}\right) \\ p_{j:n} = \exp\left(-e^{-\frac{x_{j:n} - \mu_{i,j}}{\sigma_{i,j}}}\right). \end{cases}$$

In this case, exact solutions can be found for $\mu_{i,j}$ and $\sigma_{i,j}$:

$$\sigma_{i,j} = \frac{x_{j:n} - x_{i:n}}{\log\left(\frac{\log(p_{i:n})}{\log(p_{j:n})}\right)} \text{ and } \mu_{i,j} = x_{i:n} + \sigma_{i,j}(\log(-\log(p_{i:n}))).$$

We now do some remarks on the *EPM*:

- The main advantage of this method over the MLE and the estimation via L -moments is that it can be applied regardless of the possible value of ξ (recall MLE nor L -moments should be applied when $\xi < -1$ or $\xi > 1$ respectively). In fact, according to [CH94], outside of a moderate range of ξ , EPM performs better than the L -moments.
- However, the number of combinations of possible indices $i < j < k$ grows speedily with n . A possible alternative is fixing $k = n$: this has the advantage of avoiding inconsistent estimators with the data whenever the right tail is finite ($\xi < 0$), which is a reported problem of the L -moments ([ZBK10]). To further accelerate the process, $i = 1$ can also be fixed.
- The estimators we get are weakly consistent, but we lack results concerning asymptotic normality. A possible alternative to this, as suggested in [CH94], is to use confidence intervals based on parametric bootstrap.

1.2.2 Exceedance approach

The methods that we will introduce in this section are commonly known as Peak-over-Threshold (POT) methods. Its use requires to handle the so-called Generalized Pareto distribution.

Definition 1.2.3. *We define the Generalized Pareto (GP) distribution as the one with cdf:*

$$H_{\xi,\sigma}(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \xi = 0, x > 0 \\ 1 - (1 + \xi\frac{x}{\sigma})^{-1/\xi} & \xi \neq 0, 1 + \xi\frac{x}{\sigma} > 0 \\ 1 & \xi < 0, x \geq \frac{\sigma}{|\xi|} \end{cases}$$

We show the cdf and the density functions of GP distributions in Figure 1.2.

Remark.

- As particular cases of the GP distribution, we find the exponential distribution when $\xi = 0$, the uniform distribution when $\xi = -1$ or the Pareto distribution if $\xi < 0$.
- It is very common to find the GP distribution defined with $-\xi$ instead of ξ , similarly to what happened to the GEV distribution. It can also be found defined with a localization parameter, exchanging x for $x - \mu$ (and the support of the distribution changing accordingly). However, in Extreme Value Theory, the GP distribution is used because of Theorem 1.2.4, which makes clear why this location parameter is not necessary.

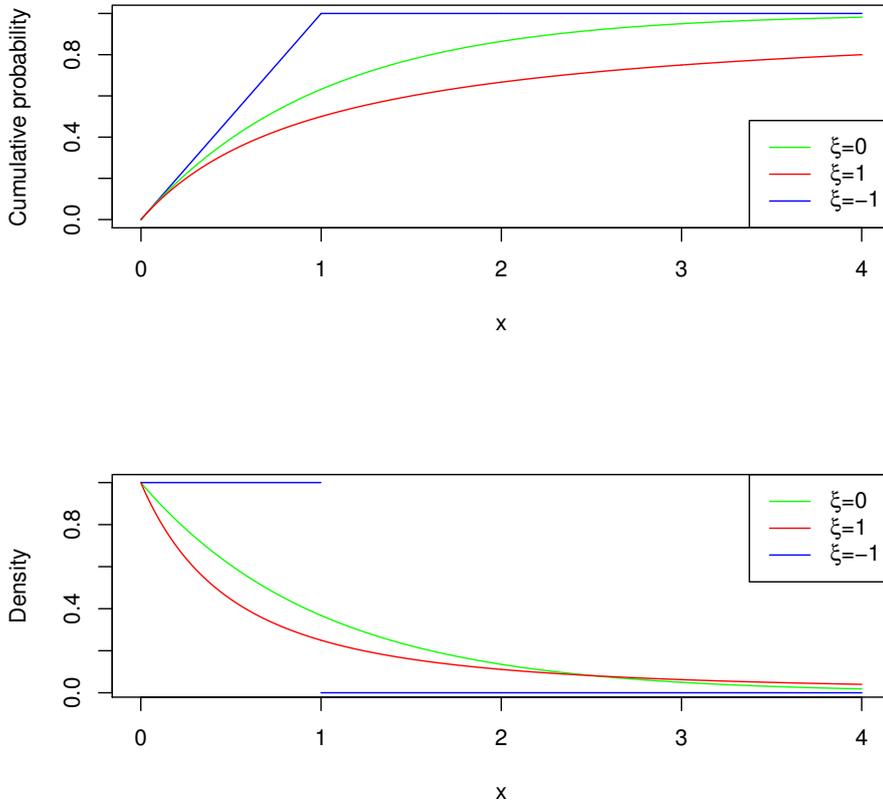


Figure 1.2: Cumulative distribution and density functions of the GP distribution with parameters $\xi \in \{0, 1, -1\}$, $\sigma = 1$.

As mentioned in the introduction, if X is a rv with distribution belonging to some domain of attraction, and u is large enough (in terms which we have not specified yet) then the distribution of the exceedances $X - u$ can be approximated by a GP distribution. Furthermore, the parameters are very related to the ones from the GEV distribution.

Theorem 1.2.4 (Pickands–Balkema–de Haan theorem). *Let X_1, \dots, X_n be iid with distribution F such that $F \in \mathcal{D}(G)$ for some G . Then for large enough u there exist ξ and $\tilde{\sigma}$ such that*

$$\mathbb{P}(X - u \leq y | X > u) \approx H_{\xi, \tilde{\sigma}}(y).$$

Sketch of proof. Let us compute the value of

$$\mathbb{P}(X - u > y | X - u > 0) = \frac{1 - F(u + y)}{1 - F(u)}.$$

As already mentioned, for large enough n we can accept the approximation $F^n(x) \approx G_{\xi, \mu, \sigma}(x)$ for all $x \in \mathbb{R}$ and some $\xi, \mu \in \mathbb{R}$, $\sigma > 0$. Therefore,

$$\log(F(x)) \approx \frac{1}{n} \log(G_{\xi, \mu, \sigma}(x))$$

whenever $G_{\xi,\mu,\sigma}(x) \neq 0$. If $u \in \mathbb{R}$ is such that $F(u) \approx 1$, then $\log(F(u)) \approx -(1 - F(u))$, which implies

$$1 - F(u) \approx \frac{1}{n} \log(G_{\xi,\mu,\sigma}(u)) = \frac{1}{n} \left(1 + \frac{\xi}{\sigma}(u - \mu) \right)^{-1/\xi}.$$

Analogously, if $y \in (0, x_F - u)$:

$$1 - F(u + y) \approx \frac{1}{n} \log(G_{\xi,\mu,\sigma}(x)) = \frac{1}{n} \left(1 + \frac{\xi}{\sigma}(u + y - \mu) \right)^{-1/\xi}.$$

Finally, dividing, we obtain:

$$\frac{1 - F(u + y)}{1 - F(u)} = \left(1 + \frac{\xi}{\tilde{\sigma}}y \right)^{-1/\xi} \quad \tilde{\sigma} = \sigma + \xi(u - \mu)$$

and we conclude that $\mathbb{P}(X - u \leq y | X - u > 0) \approx H_{\xi,\tilde{\sigma}}(y)$. \square

Remark. A detailed proof can be found in [Pic75].

Compared to parameter estimation methods for the GEV distribution, there is many literature studying the estimation of parameters for the Generalized Pareto distribution, specially in this context.³ An exhaustive comparison of many estimation methods within the POT approach can be found in [ZBK10]. We will now briefly discuss the estimators introduced in the block maxima approach. Assume $\xi \neq 0$, or otherwise the GP distribution reduces to the exponential distribution, which can be estimated by any means (for example, the MLE and the estimator via L -moments is the same).

- MLE. Its theoretical behaviour is very similar to the MLE for the GEV distribution: it exists whenever $\xi > -1$ and it is asymptotically normal when $\xi > -\frac{1}{2}$ ([Dom15]). According to [ZBK10], MLE it is the best method for estimating the parameters when the sample size is large enough and ξ belongs to $(-\frac{1}{2}, \frac{1}{2})$. The log-likelihood function is:

$$l(x_1, \dots, x_n; \xi, \sigma)(X_1, \dots, X_n) = \begin{cases} -n \log(\sigma) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log\left(1 + \frac{\xi}{\sigma} X_i\right) & \xi \neq 0 \\ -n \log(\sigma) - \frac{1}{\sigma} \sum_{i=1}^n X_i & \xi = 0 \end{cases} \quad (1.19)$$

And the asymptotic covariance matrix is ([BGT04, p. 162]) is

$$I_\theta = \frac{1}{2\sigma + 1} \begin{bmatrix} \frac{2}{\xi+1} & \frac{1}{(\xi+1)\sigma} \\ \frac{1}{(\xi+1)\sigma} & \frac{1}{\sigma^2} \end{bmatrix} \quad I_\theta^{-1} = (1 + \xi) \begin{bmatrix} 1 + \xi & -\sigma \\ -\sigma & 2\sigma^2 \end{bmatrix}$$

- L -moments: as we already mentioned, they enjoy the same properties as the PWM. For the latter, it has been reported (see [ZBK10]) that they perform better than the rest of the estimators if ξ is in the range $(0, \frac{1}{2})$, regardless of the sample size. However,

³This can be due to the majority opinion that the exceedance approach is better than the block maxima approach, mainly because they use more data. However, this has been questioned: for a detailed discussion, see Bücher, Zhou: *A horse racing between the block maxima method and the peak-over-threshold approach* (2020). Statistical Science. To appear.

L -moments are only defined for $\xi < 1$. Below we can find the first four L -moments for the GP distribution, as reported in [HW97, Appendix A5]:

$$\lambda_1 = \frac{\sigma}{1 - \xi} \quad (1.20)$$

$$\lambda_2 = \frac{\sigma}{(1 - \xi)(2 - \xi)} \quad (1.21)$$

$$\tau_3 = \frac{1 + \xi}{3 - \xi} \quad (1.22)$$

$$\tau_4 = \frac{(1 + \xi)(2 + \xi)}{(3 - \xi)(4 - \xi)}, \quad (1.23)$$

The calculation of the estimators $\hat{\xi}^L$ and $\hat{\sigma}^L$ from the estimators of the L -moments (shown in Equations 1.9 and 1.10 or in any of the variants following these equations) is simply: a closed expression of $\hat{\xi}^L$ can be found from Equation 1.22, and then use 1.20 or 1.21 to compute $\hat{\sigma}^L$. The estimators are asymptotically normal when $\xi < \frac{1}{2}$, and the covariance matrix of (ξ^L, σ^L) is [BGT04, Section 5.6]:

$$\frac{1-2\xi}{3-2\xi} \begin{bmatrix} (1 - \xi)(2 - \xi)^2(1 - \xi + 2\xi^2) & -\sigma(2 - \xi)(2 - 6\xi + 7\xi^2 - 2\xi^3) \\ -\sigma(2 - \xi)(2 - 6\xi + 7\xi^2 - 2\xi^3) & \sigma^2(7 - 18\xi + 11\xi^2 - 2\xi^3) \end{bmatrix} \quad (1.24)$$

- Elemental Percentile Method: as explained, this method is able to perform where the MLE and the estimator using L -moments cannot be used, as it is defined for every ξ . Since we only have two parameters, then we only need to pick two parameters.

In this setting, if $\delta := \frac{\sigma}{\xi}$, given $i < j \leq n$ the system to be solved is:

$$\begin{cases} p_{i:n} = 1 - \left(1 + \frac{x_{i:n}}{\delta_{i,j}}\right)^{-1/\xi_{i,j}} \\ p_{j:n} = 1 - \left(1 + \frac{x_{j:n}}{\delta_{i,j}}\right)^{-1/\xi_{i,j}} \end{cases}$$

Eliminating $\xi_{i,j}$ we get the following equation for $\delta_{i,j}$:

$$\log(1 - p_{i:n}) \log\left(1 + \frac{x_{j:n}}{\delta_{i,j}}\right) = \log(1 - p_{j:n}) \log\left(1 + \frac{x_{i:n}}{\delta_{i,j}}\right),$$

and, using the relations

$$\xi_{i,j} = \frac{\log\left(1 - \frac{x_{i:n}}{\delta_{i,j}}\right)}{\log(1 - p_{i:n})} \quad \text{and} \quad \sigma_{i,j} = \delta_{i,j} \xi_{i,j}.$$

we can compute the estimators $\hat{\xi}_{i,j}$ and $\hat{\sigma}_{i,j}$.

Once again, to compute the variance estimators of $\hat{\xi}^{EPM}$ and $\hat{\sigma}^{EPM}$ and to obtain confidence intervals, parametric bootstrap is suggested [CH97]. It is also advisable to fix $j = n$ to reduce the computational cost of the estimation, and to ensure that the estimator is consistent with the observed data, which may be a problem if $\xi < 0$.

If the block maxima methods entail a tradeoff regarding the block size, the *POT* approach has an analogous problem with the choice of the threshold u : if u is too small, then the approximation $\log(F(u)) \approx -(1 - F(u))$ is poor, and the corresponding approximation of the exceedances by a GP distribution is very weak. On the contrary, if u is very large, then the approximation is valid but we may end up with too few data to make the pertinent estimations. To help us make the choice, there are several methods available, as we show now.

Choosing the threshold⁴

There are several heuristic formulas that are used in practice, even though they lack, to a greater or lesser extent, formal arguments supporting them. One example are percentile rules: when doing this, generally the 0.9 quantile is chosen as a threshold. Upper order statistics are also considered: and usual elections $k = \sqrt{n}$ or $k = \frac{n^{\frac{2}{3}}}{\log(\log(n))}$. A detailed discussion on different ways to choose the threshold can be found in [SM12].

We will use graphical methods, which allow us to assess in every situation which the better option is. A drawback of this approach is that sometimes plots are not easy to read. We first note that if $Y \sim G_{\xi, \sigma}$, then $\mathbb{E}[Y] = \frac{\sigma}{1-\xi}$ if $\xi < 1$ (and ∞ otherwise). Now, if u_0 is a valid threshold (in the sense of being large enough), and $Y_{u_0} = X - u_0$ is an exceedance of u_0 , where $X \sim G_{\xi, \mu, \sigma}$, then $\mathbb{E}[Y_{u_0}] = \frac{\sigma_0}{1-\xi}$, where $\sigma_0 = \sigma + \xi(u_0 - \mu)$. Besides, since u_0 is valid, then every $u > u_0$ is a valid threshold as well. If Y_u is an exceedance of u , with the conditions just imposed, we have the relation:

$$\mathbb{E}[Y_u] = \frac{\sigma + \xi(u - \mu)}{1 - \xi} = \frac{\sigma_0 + \xi(u - u_0)}{1 - \xi},$$

and this expression is linear in u . As a consequence, if we plot the points

$$\left\{ \left(u, \frac{1}{\#\{i : X_i > u\}} \sum_{i: X_i > u} (X_i - u) \right) : u < M_n \right\},$$

then for some u the plot should start to look like a straight line, and any of these values of u could be an adequate election of a threshold. This plot is called the *mean residual life plot* (see, for reference, [GR10]). Moreover, since the parameters of the line are related to the parameters of the GP distribution, one could think of doing linear regression on the part of the plot that could be identified as a good option as a way to estimate them. This has been done in [ZW07], where they find good results. It has also been used in [CH97] to measure the goodness of fit of the *EPM* estimator. It is usual to plot the mean residual life plot along with confidence intervals based on asymptotic normality of the sample mean. Both things are easy to compute since, for a given threshold, it only involves computing the mean and the standard deviation of the exceedances over that threshold.

Another option to choose the threshold is based on the idea that the quantities ξ and $\sigma^* := \sigma_u - \xi u$, where σ_u is the second parameter of the GP distribution with threshold u , do not depend on u when it is a valid threshold. As a consequence, if we plot these values against u , the plots should begin being constant at some point. Confidence intervals can be computed as well: confidence intervals for $\hat{\xi}$ depend on the estimation method that has been used, whereas confidence intervals for $\hat{\sigma}^*$ require the Delta method. These plots are sometimes referred to as *parameter stability plots*.

The last option is considerably more expensive, because it requires fitting the model several times and compute the corresponding confidence intervals. However, the mean life residual plot may not be easy to read for extreme values of u , because the lack of points above those thresholds may increase the variance of the sample mean. As a consequence, this option may be helpful if the first interpretation gets difficult. Note that these methods can also be useful for evaluating the goodness of the fit and not only for choosing the threshold.

⁴For this discussion, we have followed Section 4.3 of [Col01].

1.2.3 On the value of ξ

The importance of ξ is evident: it is the parameter that determines the shape of the GP and the GEV distributions and, especially, it has a great influence on the right tail. When $\xi < 0$ the right tail is bounded; when $\xi = 0$ it decreases exponentially, and if $\xi > 0$ it decreases polynomially, that is, the distribution is a heavy-tailed distribution.

We have also seen how, depending on the value of ξ , some estimation methods are more appropriate than others. As a consequence, it is very important to have a preliminary idea of its value. There are several ways to achieve this:

Depending on the the problem, some values are more expected than others. For example, according to [HWW85], in 1975, the fitting of 32 annual flood series belonging to the *Natural Environment Research Council* database found ξ to be in the range $(-0.48, 0.32)$, and in more general settings it belongs to $(-0.5, 0.5)$: that is why the MLE or the estimation via L -moments are considered most of the times. However, the estimation of the extreme values of sea waves measured in Bilbao in 1997 is far from that range: ξ goes from -0.682 to -1.271 (see [CH97]).

A more rigorous alternative, if possible, would consist in having a preliminary estimation of ξ using an easy-to-compute estimator to have a preview of the shape of the distribution, and ideally to do hypothesis testing. This would allow us to best decide which methods can be used. Fortunately, there exist estimators as described. They are based on more complex conditions for a distribution to belong to a domain of attraction than the ones we have seen. We will settle for just describing the estimators.

Pickands and moment estimator

The Pickands estimator is one of the most simple estimators available. A complete discussion with the results that we are going to mention can be found in [DH89]. It is defined as:

$$\hat{\xi}_n^{P,k} := (\log(2))^{-1} \log \left(\frac{X_{n-k+1:n} - X_{n-2k+1:n}}{X_{n-2k+1:n} - X_{n-4k+1:n}} \right),$$

where $k = k(n)$. An explanation of the origin of this estimator can be found in [BGT04]. It can be proved that the Pickands estimator is weakly consistent if $k(n) \rightarrow \infty$ and $\frac{k(n)}{n} \rightarrow 0$, and strongly consistent if, in addition, $\frac{k(n)}{\log(\log(n))} \rightarrow \infty$. Furthermore, under considerably more complicated conditions, stated in Theorem 2.3 and Theorem 2.5 of the aforementioned paper, $\hat{\xi}_n^{P,k}$ is unbiased and asymptotically normal with variance

$$\frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1)\log(2))^2 k(n)}.$$

One drawback of the Pickands estimator is the fact that only 4 values are used, which undoubtedly seems like a waste of information. This may explain why the variance is so high. However, it is very valued for its simplicity. An alternative which has been observed to perform better than the Pickands estimator, while having similar properties, is the moment estimator.

Consider first $M_n^{(1)} := \frac{1}{k} \sum_{i=0}^{k-1} \log\left(\frac{X_{n-i:n}}{X_{n-k:n}}\right)$. This is known as the *Hill estimator*, which historically speaking is the first estimator based on the underlying conditions used by the

estimators that we are mentioning. It is easy to observe that it only makes sense when $\xi > 0$. It enjoys very similar properties to the Pickands estimator, while the variance when it is asymptotically normal is $\frac{\xi^2}{k(n)}$, which is considerably smaller. Now, define the following quantities:

$$M_n^{(2)} := \frac{1}{k} \sum_{i=0}^k \log^2 \left(\frac{X_{n-i:n}}{X_{n-k:n}} \right) \text{ and } \hat{\xi}_n^{M,k} := M_n^{(1)} + 1 - \frac{1}{2} \left[1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right]^{-1}.$$

The statistic $\hat{\xi}_n^{M,k}$ is known as the *moment estimator*. It has the following properties:

- Weakly consistent if $x_F > 0$, $\frac{k(n)}{n} \rightarrow 0$ and $k(n) \rightarrow \infty$.
- Strongly consistent if the previous condition hold and $\frac{k(n)}{\log(n)^\delta} \rightarrow \infty$ for some $\delta > 0$.
- Under the same conditions as the Pickands estimator, asymptotically normal with variance

$$\begin{cases} \frac{1+\xi^2}{k(n)} & \text{if } \xi \geq 0 \\ \frac{(1-\xi)^2(1-2\xi)}{k(n)} \left(4 - \frac{8(1-2\xi)}{1-3\xi} + \frac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)} \right) & \text{if } \xi < 0. \end{cases}$$

The previous results can be found in [DEH89]. Computationally speaking, it is not significantly more involved than the Pickands estimator and, as mentioned, it is found to perform better. A possible inconvenience is the requirement that $x_F > 0$. However, it can be easily solved by doing a proper shift of the data.

To sum up, any of these estimators are appropriate enough for our purpose, which is to do a preliminary study on the shape of the distribution. Finally, we will introduce graphical methods on the right tail of the distribution, to try and get more additional information on this topic.

Graphical methods: P-P Plot

The P-P plot, along with the Q-Q plot, is one of the most used graphical methods to evaluate how feasible it is for a sample to come from a given parametric family of distributions $\{F_\theta, \theta \in \Theta\}$. The idea consists of looking for transformations g of $(0, 1)$ and h of the sample space such that

$$g(F_\theta(x)) = ah(x) + b,$$

where a and b may depend on θ . As a consequence, if we have a sample X_1, \dots, X_n identically distributed and \tilde{F} is its empirical cdf, then the points

$$\left\{ (h(X_i), g(\tilde{F}(X_i))) : i \in \{1, \dots, n\} \right\}$$

should resemble a straight line when $X_i \sim F_\theta$. We can use this procedure to evaluate which one of the domains of attraction best fits our data, and have a preliminary idea on what we can expect, and what methods we can use.

Recall that the cdf of the Gumbel distribution (the subfamily of the GEV distribution corresponding to $\xi = 0$) is $F_{\mu,\sigma}(x) = \exp(-\exp(-\frac{x-\mu}{\sigma}))$. Taking logarithms, we obtain:

$$-\log(-\log(F(x))) = \frac{1}{\sigma}x - \frac{\mu}{\sigma},$$

so if $g(F(x)) = -\log(-\log(F(x)))$ and $h(x) = x$ we have the desired transformations.

On the other hand, if $G_{\xi,\mu,\sigma}(x)$ is the cdf of the GEV distribution (Equation 1.4), derivating $g(G(x))$ twice we get:

$$g''(G(x)) = \begin{cases} \frac{-\xi}{(\sigma+\xi(x-\mu))^2} & \xi \neq 0 \\ 0 & \xi = 0 \end{cases}$$

If we have a sample from a distribution F and we consider the Gumbel P-P plot, we can argue that:

- If the right tail looks like a straight line ($g'' = 0$), we have a hint that F belongs to the Type I domain of attraction.
- If the right tail looks convex ($g'' > 0$), we have a hint that F belongs to the Type II domain of attraction.
- If the right tail looks concave ($g'' < 0$), we have a hint that F belongs to the Type III domain of attraction.

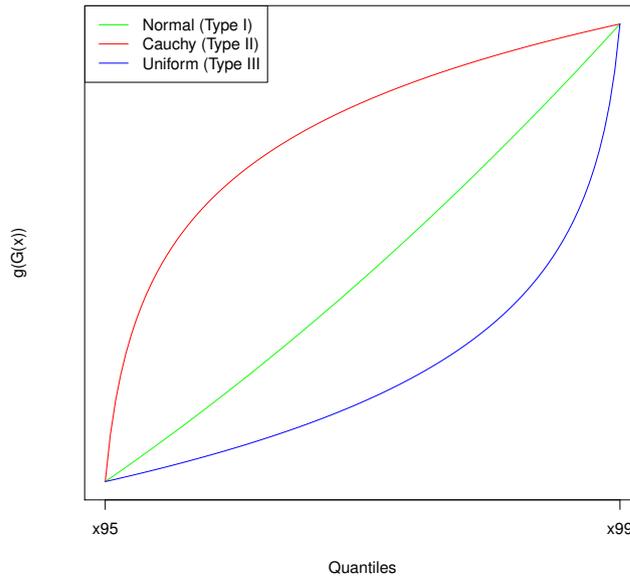


Figure 1.3: Gumbel P-P Plot where the right tail of the Normal ([LLR83, Theorem 1.5.3]), Cauchy (Example 3) and Uniform (Example 5) distributions have been plotted. The parameters of the distributions have been adjusted so that the quantiles (0.95, 0.999) are the same.

1.3 Results under dependence conditions

The theoretical results that we have seen so far rely on the observations being iid. However, in the vast majority of practical situations this is not the case, and therefore it is of great interest to know what happens if we relax this condition in some way. For instance, the observations being identically distributed may be a reasonable assumption, but independence is very rarely the case. For example, in real-life problems extreme values tend to come along with others (we say that they are *clustered*). In this section we will first prove a result giving conditions under which the possible limit distributions for the normalized maxima are the same, that is, GEV distributions. Later on, we will see that, in terms yet to be defined, we can use the same tools as in the iid case.

The technique to prove the more refined version of the Fisher-Tippett-Gnedenko theorem will be different from the iid case. We want to prove that if $\{X_n : n \in \mathbb{N}\}$ is a sequence of rv's with certain dependence conditions, and there exist sequences $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ such that if $\mathbb{P}(a_n(M_n - b_n) \leq x)$ converges to a non-degenerate cdf, then that distribution is max-stable (or, equivalently, GEV).

From the characterization of max-stable distributions in Proposition 1.1.5, it can be deduced that G is max-stable if and only if there exist sequences $\{a_n\} \subset \mathbb{R}^+$ and $\{b_n\}$ such that for every $k \in \mathbb{N}$:

$$\mathbb{P}(a_{nk}(M_n - b_{nk}) \leq x) \xrightarrow[n \rightarrow \infty]{w} G^{1/k}(x).$$

Now, assume $\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow[n \rightarrow \infty]{w} G(x)$. Note that, for any fixed $k \in \mathbb{N}$, the convergence still holds changing n for nk . It is clear to see that G is max-stable if, under the previous assumption, for all $k \geq 2$:

$$\mathbb{P}(a_{nk}(M_{nk} - b_{nk}) \leq x) - \mathbb{P}^k(a_{nk}(M_n - b_{nk}) \leq x) \xrightarrow[n \rightarrow \infty]{} 0, \quad (1.25)$$

and this is what we will prove. We will now settle the terminology we will use in this section.

Definition 1.3.1.

- If i_1, \dots, i_r is a set of indices, $F_{i_1, \dots, i_r}(x_1, \dots, x_r)$ will denote the joint cdf of the random vector $(X_{i_1}, \dots, X_{i_r})$. If $x_1 = \dots = x_r = x$, we will write $F_{i_1, \dots, i_r}(x)$.
- If $j_1 \leq j_2$, then $[j_1, j_2]$ will denote the set $\{j \in \mathbb{N} : j_1 \leq j \leq j_2\}$. Its length is $j_2 - j_1 + 1$. If $[j_1, j_2]$ are $[j_3, j_4]$ intervals, we say they are separated by m if $j_3 - j_2 = m$.
- If $E = \{i_1, \dots, i_r\} \subset \mathbb{N}$ is a set of indices and $\{X_n : n \in \mathbb{N}\}$ a sequence of rv's, $M(E) := \max\{X_{i_j} : j \in [1, r]\}$.

The first condition that we impose to the process is stationarity: recall that the sequence $\{X_n : n \in \mathbb{N}\}$ is stationary if for every combination $r, s \in \mathbb{N}$, $i_1, \dots, i_r \in \mathbb{N}$, $F_{i_1, \dots, i_r} = F_{i_1+s, \dots, i_r+s}$. From now on, every process will be stationary, unless specified.

Definition 1.3.2. Let $\{u_n\}$ be a real sequence. We say that the condition $D(u_n)$ holds for the process $\{X_n : n \in \mathbb{N}\}$ if for every chain $i_1 < \dots < i_p$ and $j_1 < \dots < j_{p'} \leq n$ such that $j_1 - i_p \geq s$,

$$|F_{i_1, \dots, i_p, j_1, \dots, j_{p'}}(u_n) - F_{i_1, \dots, i_p}(u_n)F_{j_1, \dots, j_{p'}}(u_n)| \leq \alpha_{n,s},$$

where $\alpha_{n,s}$ decreases in s and for all $\lambda > 0$, $\lim_n \alpha_{n, [n\lambda]} = 0$.

In the last expression, the left hand side is 0 if the rv's $\{X_n : n \in \mathbb{N}\}$ are independent. If $D(u_n)$ holds, then the dependence for events of the form $X_i \leq u_n$ decreases the more separated the corresponding index collections are. In real problems, this translates into the dependence becoming weaker the more separated the measures are, for instance, space or time-wise. The condition $D(u_n)$ is, as a consequence, a formalization of a dependence that turns out to be very natural and feasible in lots of situations. We will show that under this condition, the Fisher-Tippett-Gnedenko theorem still holds.

Lemma 1.3.1. *Let $\{u_n\}$ be a sequence such that $D(u_n)$ holds for a sequence $\{X_n : n \in \mathbb{N}\}$ of rv's, and let $n, r, k \in \mathbb{N}$ and E_1, \dots, E_r be subintervals of $[1, n]$ pairwise separated by, at least, k . Then*

$$\left| \mathbb{P}(\cap_{j=1}^r \{M(E_j) \leq u_n\}) - \prod_{j=1}^r \mathbb{P}(\{M(E_j) \leq u_n\}) \right| \leq (r-1)\alpha_{n,k}$$

Proof. We apply induction on r . Let $A_j := \{M(E_j) \leq u_n\}$ for $j = 1, \dots, r$.

- $r = 2$: $|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| = |F_{E_1 E_2}(u_n) - F_{E_1}(u_n)F_{E_2}(u_n)| \leq \alpha_{n,k}$.
- First, observe that

$$\left| \mathbb{P}(\cap_{j=1}^r A_j) - \prod_{j=1}^r \mathbb{P}(A_j) \right| \leq \left| \mathbb{P}(\cap_{j=1}^r A_j) - \mathbb{P}(\cap_{j=1}^{r-1} A_j)\mathbb{P}(A_r) \right| + \mathbb{P}(A_r) \left| \mathbb{P}(\cap_{j=1}^{r-1} A_j) - \prod_{j=1}^{r-1} \mathbb{P}(A_j) \right|$$

By induction assumption, the second term of the right hand side can be bounded by $(r-2)\alpha_{n,k}$. On the other hand, if $\bar{E}_r = \cup_{j=1}^{r-1} E_j$ and $\bar{A}_r = \{M(\bar{E}_r) \leq u_n\}$, then the first term can be written as $|\mathbb{P}(\bar{A}_r \cap A_r) - \mathbb{P}(\bar{A}_r)\mathbb{P}(A_r)|$, which can be bounded by $\alpha_{n,k}$, because $D(u_n)$ holds. □

For large n , let $k < n$ and $n'_k := \lfloor \frac{n}{k} \rfloor$. Assume that there exists $m \in \mathbb{N}$ such that $k < m < n'_k$, which is the case if n is large enough. Define:

$$\begin{aligned} I_j &:= [n'_k(j-1) + 1, jn'_k - m], & I_j^* &:= [jn'_k - m + 1, jn'_k] \quad (j \in [1, k]) \\ I_{k+1} &:= [(k-1)n'_k + m + 1, kn'_k], & I_{k+1}^* &:= [kn'_k + 1, kn'_k + m]. \end{aligned}$$

Note that the only interval which is not contained in $[1, n]$ is I_{k+1}^* . Besides, the intervals $I_1, I_1^*, \dots, I_k, I_k^*$ are consecutive, and their lengths are $n'_k - m$ and m respectively.

Lemma 1.3.2. *Under the previous previous notation, let $\{u_n\}$ be a sequence such that $D(u_n)$ holds for the sequence of rv's $\{X_n\}$. Then*

1. $0 \leq \mathbb{P}(\cap_{j=1}^r \{M(I_j) \leq u_n\}) - \mathbb{P}(M_n \leq u_n) \leq (k+1)\mathbb{P}(M(I_1) \leq u_n < M(I_1^*))$.
2. $\left| \mathbb{P}(\cap_{j=1}^r \{M(I_j) \leq u_n\}) - \mathbb{P}^k(M(I_1) \leq u_n) \right| \leq (k-1)\alpha_{n,m}$.
3. $\left| \mathbb{P}^k(M(I_1) \leq u_n) - \mathbb{P}^k(M_{n'_k} \leq u_n) \right| \leq k\mathbb{P}(M(I_1) \leq u_n < M(I_1^*))$.

Proof. We note that the probabilities $\mathbb{P}(M(I_j) \leq u_n)$ and $\mathbb{P}(M(I_j) \leq u_n \leq M(I_j^*))$ do not depend on the value of j because of the stationarity of the process. The first two results follow easily from this observation. The latter also follows taking into account that

$$\mathbb{P}(M(I_1) \leq u_n) - \mathbb{P}(M_{n'_k} \leq u_n) = \mathbb{P}(M(I_1) \leq u_n < M(I_1^*)),$$

and also that $y^k - x^k \leq k(y-x)$ whenever $0 \leq x \leq y \leq 1$. □

From the previous results we can deduce the inequality

$$\left| \mathbb{P}(M_n \leq u_n) - \mathbb{P}^k(M_{n'_k} \leq u_n) \right| \leq (2k+1)\mathbb{P}(M(I_1) \leq u_n < M(I_1^*)) + (k-1)\alpha_{n,m}, \quad (1.26)$$

adding and subtracting $\mathbb{P}(\cap_{j=1}^k \{M(I_j) \leq u_n < M(I_j^*)\})$ and $\mathbb{P}^k(M(I_1) \leq u_n)$ and doing the proper reorderings. Observe that if $u_n = a_n^{-1}x + b_n$ for some $x \in \mathbb{R}$, changing n for nk and n'_k for n , the expression that we are bounding in absolute value is the left hand side of (1.25). Since $\alpha_{n,m}$ goes to 0, it is enough to conveniently bound $\mathbb{P}(M(I_1) \leq u_n < M(I_1^*))$ to finish our task:

Proposition 1.3.3. *Under the previous previous notation, if $D(u_n)$ holds for some sequence $\{u_n\}$ and r, m meet $n \geq (2r+1)mk$, then*

$$\mathbb{P}(M(I_1) \leq u_n < M(I_1^*)) \leq \frac{1}{r+1} + 2r\alpha_{n,m}.$$

Proof. Define $E_j := [2(j-1)m+1, (2j-1)m]$ for $j \in [1, r]$. Since $n'_k \geq (2r+1)m$, then E_1, \dots, E_r is a collection of subintervals of I_1 pairwise separated by m and separated by at least m from I_1^* . Note that for every $j \in [1, r]$ $\mathbb{P}(M(E_j) \leq u_n) = \mathbb{P}(M(I_1^*) \leq u_n) =: p$. On the other hand:

$$\begin{aligned} \mathbb{P}(M(I_1) \leq u_n < M(I_1^*)) &\leq \mathbb{P}(\cap_{j=1}^r \{M(E_j) \leq u_n\}, \{M(I_1^*) > u_n\}) \\ &= \mathbb{P}(\cap_{j=1}^r \{M(E_j) \leq u_n\}) - \mathbb{P}(\cap_{j=1}^r \{M(E_j) \leq u_n\}, \{M(I_1^*) \leq u_n\}) \end{aligned}$$

The last expression can be bounded in absolute value by $p^r - p^{r+1} - 2r\alpha_{n,m}$ adding and subtracting $p^r - p^{r+1}$ and using the Lemma 1.3.1. The result follows now observing that $p^r - p^{r+1} \leq \frac{1}{r+1}$ (which can be proved from the geometric series expression of $(1-p)^{-1}$). \square

It is now easy to put the pieces together to prove the Fisher-Tippett-Gnedenko theorem for stationary sequences with the $D(u_n)$ condition for some u_n .

Theorem 1.3.4. *Let $\{X_n : n \in \mathbb{N}\}$ be a stationary sequence $\{X_n\}$ of rv's and $\{a_n\} \subset \mathbb{R}^+$, $\{b_n\}$, sequences such that $\mathbb{P}(a_n(M_n - b_n) \leq x)$ converges to a non-degenerate cdf $G(x)$. If $D(u_n)$ holds for every sequence of the form $u_n = a_n^{-1}x + b_n$, then $G(x)$ is the cdf of a GEV distribution.*

Proof. As already discussed, it suffices to show that for all $k \geq 2$

$$\left| \mathbb{P}(M_n \leq u_n) - \mathbb{P}^k(M_{n'_k} \leq u_n) \right| \xrightarrow{n \rightarrow \infty} 0.$$

Combining (1.26) and Proposition 1.3.3 we have:

$$\left| \mathbb{P}(M_n \leq u_n) - \mathbb{P}^k(M_{n'_k} \leq u_n) \right| \leq \frac{2k+1}{r+1} + (2r(2k+1) + k-1)\alpha_{n,m},$$

and this goes to 0 letting first n and then r go to ∞ . \square

Remark. The condition $D(u_n)$ feels like an unnecessary restriction, considering that in most real-life processes the dependence decreases the more separated the variables are, regardless of its actual values. As a consequence, the following definition captures more suitably the notion of dependence we want:

Definition 1.3.3. We say that the condition D holds for the process $\{X_n : n \in \mathbb{N}\}$ if for every chain $i_1 < \dots < i_p$ and $j_1 < \dots < j_{p'}$ such that $j_1 - i_p \geq l$ and every $u \in \mathbb{R}$:

$$|F_{i_1, \dots, i_p, j_1, \dots, j_{p'}}(u) - F_{i_1, \dots, i_p}(u)F_{j_1, \dots, j_{p'}}(u)| \leq g(l),$$

where $g(l) \xrightarrow[l \rightarrow \infty]{} 0$.

Once we have found out conditions under which the non-empty domains of attraction are the same, it is interesting to know the influence of these conditions on the limit distribution, compared to the process being iid.

For a stationary process $\{X_n : n \in \mathbb{N}\}$ it is now useful to define the iid process $\{\tilde{X}_n : n \in \mathbb{N}\}$ with the same marginal distributions as $\{X_n : n \in \mathbb{N}\}$. The following theorem relates the convergence of the normalized maximum of $\{\tilde{X}_1, \dots, \tilde{X}_n\}$, denoted as \tilde{M}_n , which we already know, with the convergence of the normalized M_n .

Theorem 1.3.5 ([BGT04], Theorem 10.4). Suppose $\mathbb{P}(a_n^{-1}(\tilde{M}_n - b_n) \leq x) \xrightarrow[n \rightarrow \infty]{w} \tilde{G}(x)$ for sequences $\{a_n\} \subset \mathbb{R}^+$, $\{b_n\}$ and a non-degenerate cdf $\tilde{G}(x)$. If $D(u_n)$ holds for every sequence of the form $u_n = a_n x + b_n$ such that $\tilde{G}(x) > 0$ and $\mathbb{P}(a_n^{-1}(M_n - b_n) \leq x)$ converges for some x , then there exists $\theta \in [0, 1]$ such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow[n \rightarrow \infty]{} (\tilde{G}(x))^\theta.$$

The constant θ is called the *extremal index* of the process $\{X_n : n \in \mathbb{N}\}$. One of the possible interpretations of the extremal index is the following: it is the inverse of the expected number of observations over a threshold in a cluster containing at least another exceedance. Therefore, the bigger the extremal process, the less clusterized the extremal values are. Clearly, if the process is independent then $\theta = 1$, although the contrapositive proposition is not true: see, for example, [LLR83, Theorem 3.5.2]. On the other hand, under very relaxed conditions ([BGT04, Condition 10.8]) the extremal index is strictly positive.

Estimation of the parameters

If we were to use the techniques described in the previous section, in which we are treating our sample as iid, then we would be estimating $\tilde{G}(x)$ (actually, $\tilde{G}(a_n x + b_n)$), which belongs to the same family as $\tilde{G}(x)$, but we are interested in estimating $G(x)$. To do this, it is necessary that we estimate θ . If we do not do it, we are overestimating the quantiles we are interested in (or, equivalently, we are underestimating the probabilities of events), because:

$$G^{-1}(q) = \tilde{G}^{-1}(q^{\frac{1}{\theta}}) \leq \tilde{G}^{-1}(q).$$

Its estimation is itself a wide topic, and we do not aim to delve into it, but to have a way of estimating it. The following is one of the possible estimators, which is based on the interpretation of θ we have seen.⁵ If u is a threshold and $r < n$, we define

$$\bar{\theta}_n^R(u; r) = \frac{n \sum_{i=1}^{n-r} \mathbf{1}(X_i > u, M([i+1, i+r]) \leq u)}{(n-r) \sum_{i=1}^n \mathbf{1}(X_i > u)}$$

As it can be seen from the last paragraphs, the Block Maxima approach is unaffected by the conditions we have imposed to the process for two reasons: first, taking blocks it is very

⁵Which is not the only one. See, for example, (10.10) in [BGT04].

likely that we are getting rid of the clusters by taking just one value per block; second, we can assume that the block maxima are independent. This assumption is reasonable if the length of the blocks is large enough, as a consequence of the long-range dependence being weak (as formalized by the condition D , which most of the times holds). This implies that we can use the described tools to estimate \tilde{G} . As a consequence, to estimate the joint cdf G , the only thing left to do is estimating the extremal index. It is worth noting that depending on the nature of the problem, a cluster may belong to different blocks, but this issue is easily solvable. For example, in hydrology. when studying floods, the water year (starting in October 1 and finishing in September 30) is taken instead of the usual year; this solves the potential problem of a flood occurring in late December and early January.

However, the Peak-over-Threshold approach must be reconsidered. This is because, in order for the exceedances over a valid threshold to follow a GP distribution, these exceedances had to be independent. As we have already mentioned, most of the times they are clustered. As a consequence, this approach requires a way of declusterizing, that is, localice the clusters and pick the maxima of each cluster. From here, the GP distribution can be fit. Independence of the cluster maxima is also a reasonable assumption, as proposed by [Col01].

Having a methodical way of declustering is, as estimating the extremal index, a deep issue, and we do not intend to focus here. A simple way of achieving this, used by [Haa90], is the following: given $i \in \mathbb{N}$, select only the values larger than both the i preceding and the i following observations over a threshold.

Chapter 2

River floods in the Basque Country

We know that records must be broken in the future, so if a flood design is based on the worst case of the past then we are not really prepared against floods. Materials will fail due to fatigue, so if the body of an aircraft looks fine to the naked eye, it might still suddenly fail if the aircraft has been in operation over an extended period of time. Our theory has by now penetrated the social sciences, the medical profession, economics and even astronomy. We believe that our field has come of age. ([GLS94])

Introduction

Extreme Value Theory is greatly used in different disciplines of knowledge. For example, if we look at recent publications of the journal *Extremes* (electronic ISSN 1572-915X, see <https://www.springer.com/journal/10687> for information), we not only find theoretical research on topics we have briefly mentioned; we can also see applications of this framework to engineering, risk theory, environmental or social sciences.¹

The first application of Extreme Value Theory dates from 1938, when Gumbel studied radioactive emissions. However, he would soon direct his interest in applying these tools to hydrology, namely, to the study of floods, to which he was devoted in the 1940s. This would be a recurrent topic in his career: recall that he was the first in applying multivariate extreme value techniques, and he did it in 1964 with the level of water measured at two different points of the same river. Since then, hydrology has benefited much from using these methods.

In the following pages we will use Extreme Value Theory tools to study and get insight on the floods of the river Oñati, in the Basque Country. As already mentioned, I have had access to the data thanks to *Instituto de Hidráulica Ambiental* of the University of Cantabria. As far as I am concerned, this is the first time that Extreme Value Theory is used to handle this data.

2.1 Motivation

The Basque Country is a region where the risks of floods is inherently high. This is because of the occupation of alluvial plains (the surroundings of rivers or, more general, streams, which are very likely to be affected if the volume flow of the stream increases) to economic and

¹For the latter, see the very interesting [RZ17].

social activities. For example, the city center of Vitoria-Gasteiz is built on the alluvial plain of the river Zadorra. However, the most illustrative example is Bilbao, the cornerstone of the Basque economy (see [MM20]). Bilbao was built in the place of confluence of the Cantabrian Sea with the river Ibaizabal², and some parts of the Estuary of Bilbao, where the Ibaizabal and the river Nervion meet, is navigable by big ships. This was one of the keys of the big economical development that Bilbao still has. Of course, this is not free of risks: in 1983 an event of 'cold drop' occurred in the Basque Country, Cantabria and eastern Asturias, but had its strongest effect in Bilbao, where the river Ibaizabal overflowed (in some points, up to 5m). The aftermath: 34 deaths and material damages worth 200.000 million pesetas (Spanish old coin; 1200 million euros, not taking inflation into account) and the industry of the region, main economic force of the Basque Country, completely shut down.

In a setting like this, great and damaging floods are not a possibility, but a certainty. It is important to be prepared in two ways. It is needed:

- An emergency plan to know how to proceed when a catastrophe happens along with a prediction system to foresee potentially harmful events. For example, in the disaster of 1983 autonomous communities in Spain were just born (the Statute of Autonomy of the Basque Country dates from 1979), so the organisms which should take care of these policies were not fully prepared. As a consequence, their emergency plan was untested and improvable. Their current emergency plan can be found in [URA15].
- Some knowledge on the behaviour of these disasters and their frequency to happen. This is commonly expressed in terms of return levels and return periods, which are good ways of describing extreme-value distributions.

Both things are connected, and in particular the first one is dependent on having some insight on the frequency of events whose consequences we want to reduce as much as possible. There is more than one way to proceed on this. At first, hydrological models for floods were built by simple interpolation from observed values. As stated by Gumbel in [Gum41]: “these formulas are sometimes constructed *ad hoc* [...] and have no general meaning. Most of them are rather complicated.”

Later on, proper statistical distributions appeared with the purpose of trying to fit extreme values. For instance, a widely used distribution is the log-Pearson type III (that is, a random variable X follows a log-Pearson type III if $\log(X)$ follows a Pearson type III), which is recommended by the United States Geological Survey in their *Bulletin 17*, or *Guidelines for Determining Flood Flow Frequency* ([ECF19]); another distribution is the 5-parameter Wakeby distribution. These are specific approaches. We will use the more general framework provided by Extreme Value Theory. Next, we will provide some information of the context of floods in the Basque Country.

Floods in the Basque Country

Floods are fundamentally caused by precipitations, although the morphology of the river is also a factor to take into account. We can find two types of rivers in the Basque Country, depending whether the river merges with the Cantabrian Sea or with the Mediterranean Sea. These types of rivers are very different:

²In Basque, *wide river*.

- Rivers merging with the Cantabrian Sea are generally short with big slopes, which increases the speed of the water. This makes the flood appear very quickly. Furthermore, these basins are prone to carry solid sediments, so the danger of these floods is greater.
- Rivers merging with the Mediterranean Sea³ usually receive less precipitations in their basins, which are longer and not very steep.

As a consequence, the first rivers are more likely to overflow, but they are not the only ones: some Mediterranean rivers lack of enough drainage. This issue causes flows at certain places for two reasons: precipitations occurring in that place and the rise of rivers passing over there. An example of this is the river Zadorra.

From a meteorological point of view, these are the main sources of floods in the Basque Country:

1. Very intense and short duration storms. The pluviometer in storms like these can measure more than 10 l/m^2 every 10 minutes. Fortunately these storms usually last 20 or 30 minutes at most. They can cause important problems in urban areas and roads due to the lack of drainage.
2. Stationary storms and cold drops.⁴ Stationary storms are due to a front between cold and warm air masses, when neither of them can replace the other and they do not move or move very slowly (stationary front). The storm generated is very intense and usually lasts for several hours. Big basins are also affected by these precipitations and the floods can be devastating. They generally happen in summer.
3. Stationary and occluded fronts. An occluded front appears when a cold air mass replace the warm mass. The passing of stationary and/or occluded fronts in Winter and Autumn can leave moderate but extended precipitations, and can last several days.

Floods can cause severe dangers to the population, goods and services. To minimize this risk, an emergency plan dictating systems of prediction, vigilance and action is currently in motion. There is a progression of four tracking status which indicates the danger of an adverse meteorological event:

- **Green:** there is no risk.
- **Yellow:** there is no risk for the population but some particular activities may be affected in some way.
- **Orange:** in some cases the damage may be important and life of people may be in danger. This is when the alert is declared.
- **Red:** the risk of a meteorological adverse situation is extreme and the damages could be great, both for people and material elements. In this case, an alarm is declared.

There is a continuous monitoring of meteorological phenomena that allows the Basque government to declare the emergency state in real time. The criteria for changing the status depends on both return periods associated to the phenomenon observed and how endangered and disrupted the population may be. For example, yellow code is expected to be declared

³Actually, the only river merging with the Mediterranean Sea is the river Ebro; the rest of the rivers are tributary of the Ebro.

⁴*Cold drop* is Spanish terminology. A proper translation could not be found.

many times over a year, whereas orange code should happen few times. On the other hand, the return period associated to a red code phenomenon is several years, so we should not expect that to happen often.

This progression is specialized in different ways for different phenomena, such as raining, wave heights or discharge of rivers. In the last case, which is the one we will focus on, the orange code is related to the river overflowing in few and separated regions, without much harm involved. In the red code, the flooded areas are important, and great damages to people, infrastructure and industry may occur. Obviously, for different rivers the particular criteria are different.

2.2 Methodology

When have encountered several problems, some of which we have already mentioned. Some of them are easily solvable, whereas others require more thinking and inevitably sacrificing the extent of the conclusions we can reach.

- In the Block Maxima approach, since the blocks will coincide with years, we will have as many data as years available, which implies that the quality of the estimation of the GEV distribution may not be great. This is a very common problem in Extreme Value Theory, because the ability to systematically take (reliable) measurements, and the interest in doing so, is new. On the other hand, we already know that some methods work better than others when the sample size is small, so we will take this into account. There is also a 'minor' problem with the assumption that the annual block maxima are iid, which is generally accepted. A way of enforcing this hypothesis is using water years instead of the usual years, which capture better the behaviour of precipitations and discharges if we want to compare them from year to year. Apart from doing this, we will not pay any more attention to this issue.
- For the river Oñati, we have data measured every 10 minutes, which is a double-edged sword. On the one hand it is always desirable to have as many data as possible; even though in this particular case having a large sampling rate may seem unnecessary, it can serve useful purposes (for example, identifying invalid measurements). On the other hand, when using the Peak-over-Threshold approach, we first need to decluster our data.⁵ As a consequence, we need to find a way of doing this. An easy way was described at the end of the previous chapter, but more options may be available.
- The mentioned issues are the least of our worries if the data cannot be assumed to be stationary. While it is true that non-stationary extreme value theory exists (the idea can be summarized saying that the parameters of the GEV distribution are functions of the time), there are some reasons why we will not use them.⁶ This is an important issue, since we face non-stationarity because of two different reasons:
 1. Clearly, in the course of a year, our measurements cannot be considered stationary, as the precipitations behave differently in winter than in summer, for example. Our solution will be to keep only the winter values. This partially solves the problem, because while it is reasonable enough to claim that winter precipitations,

⁵Recall that this is to be done in such a way that the exceedances over a good threshold can be decently assumed to be independent.

⁶The main reason is the fact that these tools would have required a huge prior theoretical treatment, which is not feasible in this context.

and as a consequence (as long as nothing happens to a basin) discharges, are stationary, the biggest floods do not always occur during winter. In fact, recall the Bilbao catastrophe happened in August of 1983. Stationary storms and DANAs, as mentioned above, are more likely to happen in summer. However, it can be argued that using only winter months is good enough for our purpose: only 3 out of the 15 biggest floods from 2011 to 2018 did not happen in winter, as reported by [URA18]. As a result, we will not be studying all kinds of floods, just the winter ones.

2. Another issue is the potential effect of climate change in precipitations and floods in the Basque Country. Two studies have been carried out and they can be found summarized in [URA18]. The studies differ in some things, but they agree that one of the potential consequences of climate change may be the increase in the 100 year-return level. Although one of them also claims changes in return levels associated to periods of less than 100 years, the other do not. Both studies have been done taking into account two levels of the Representative Concentration Pathway (RCP), which aims to describe different possible scenarios. RCP 4.5 (intermediate) and RCP 8.5 (the worst scenario) were considered, and while no relevant outcome would come out under the former, under RCP 8.5 the 100 year-return level for precipitations is expected to increase in the range 33 – 47%. This problem is not really something we can tackle, we just have to be aware that this is another source of uncertainty in the predictions.

We will now summarize the steps we will follow in our procedure, taking into account the aforementioned problems and other issues that appeared along the way:

1. Preprocess the original data and remove invalid data, obvious outliers, etc.
2. For the Block Maxima approach:
 - (a) Extract annual maxima with water year instead of the usual year.
 - (b) Fit a GEV distribution to the annual maxima, with a suitable method for estimation of parameters.
 - (c) Analyze the goodness of the fit by graphical methods.
 - (d) Compute the levels for 2-year, 50-year and 100-year periods and return periods for the three discharges that dictate the declarations of yellow, orange and red code.
3. For the Peak-over-Threshold approach:
 - (a) Find a good threshold, taking into account the tradeoff between being too small and, thence, a bad approximation to the GP distribution, or being too big and not having enough exceedances. We will use the techniques described in Subsection 1.2.2.
 - (b) Find a way of declustering the data. As our data are split along winters and they are sufficiently separated, we can assume that these winter blocks are independent and decluster each of the blocks on their own.
 - (c) We will fit the exceedances over the chosen threshold with the GP distribution and the estimation methods we have shown in Subsection 1.2.2. With the best fitting, we will compute return periods and return levels as in the Block Maxima approach and compare its results.

If this process go well (i.e., graphical methods validate what the pertinent estimators say, Block Maxima and POT approaches allow us to draw similar conclusions, etc) we will also have validated the theory shown in the previous chapter using a real dataset. This is itself valuable.

We have written the code in the programming language R for this task, making use of the the packages:

- *extRemes: Extreme Value Analysis* [GK16] provides most of the functions we will use.
- *EnvStats: Package for Environmental Statistics* [Mil13] provides the EPM method for fitting the GEV distribution.
- *Dowd* [Ach16] for the Pickands estimator.
- For very specific tasks not covered in any of these packages, functions have been written. They can be found in the script *Functions.R*.

2.3 Analysis: river Oñati

In the following pages we will study floods of the river Oñati, both with the Block Maxima and the Peak-over-Threshold approaches. We have chosen it because it is among the list of the rivers that the Basque Country marks as relevant, according to mixed criteria on the amount of people living close to them and the potential economic loss of there is a flood.

The river Oñati is born in the Basque town of Oñati, belonging to the province of Gipuzkoa. It is a tributary of the river Deba, which flows into the Cantabrian Sea, and it has been historically exploited for industrial purposes. The river Deba has always been affected by floods, as can be checked in the registry of floods of the Basque Government ([URA18I]), which is dangerous considering the wide occupation of its alluvial planes. The river Oñati, one of its main tributaries, has also been victim of several floods, and it is important because there exist hydroelectric power stations using water from its basin; it also has rich biodiversity and an appealing landscape, which attracts people from outside the region.⁷

We will study the discharge of the river Oñati measured at a gauging station which have been taking measures once every 10 minutes since August of 1989, with the exception of the time between October of 2002 and July of 2003. In practice, we have 30 years to draw conclusions. The discharges corresponding to the declaration of yellow, orange and red code, as decided by the Basque Government, are:

	Discharge (m^3/s)
Yellow code (Warning)	80.48
Orange code (Alert)	99.48
Red code (Alarm)	120.02

⁷For example, look at the trail PR-GI 3003: Camino del agua (<https://www.turismodebagoiena.eus/es/que-hacer/uraren-bidea-pr-gi-3003/>) or the Arritzutz cave (<http://www.onatiturismo.eus/es/listings/arrikutz-onatiko-kobak/>). Websites last visited on 19/09/2020.

2.3.1 Block Maxima

As discussed, in practice the Block Maxima approach is highly convenient because in most cases independence of the maxima can be assumed with no concerns, and it may be one of the reasons it is still widely used. We have removed some invalid data from the dataset and have taken annual maxima according to the water year, considering only winter months. In this context, winter usually has a broader meaning than late December to mid-March. According to [Cou99]: “winter will in general refer to the entire cool portion of the year, not just December-February as in much meteorological literature”. We have followed this idea and we have taken the period from October 1st to March 31th. This is the block maxima we have obtained for the river Oñati:

Year	Discharge (m^3/s)	Year	Discharge (m^3/s)
1989	20.28	2004	42.84
1990	53.22	2005	79.55
1991	53.47	2006	63.02
1992	100.10	2007	55.85
1993	79.25	2008	83.50
1994	62.66	2009	73.95
1995	39.63	2010	57.45
1996	65.88	2011	112.13
1997	69.25	2012	77.21
1998	66.04	2013	50.08
1999	35.42	2014	101.09
2000	64.34	2015	79.71
2001	25.48	2016	92.08
2002	39.45	2017	49.52
2003	45.02	2018	85.51

Table 2.1: Annual maxima of discharges of the river Oñati.

Note that according to our data, the red code has never been declared. This is only partially true: there was a flood in June of 1993 with a discharge of $223 m^3/s$ which was catastrophic. A picture can be found in the following url: <https://www.diariovasco.com/alto-deba/onati/crecida-onati-desbordamiento-20180413001247-ntvo.html> (last access on 09/19/2020).

It would be useful to have a preliminary idea on the value of ξ . The Gumbel P-P plot has been computed (Figure 2.1), with the plotting position $\widehat{F}(x) = \frac{x}{n+1}$. It seems to have a straight trend, that is, we have a clue of the limit distribution being Gumbel ($\xi = 0$).

Another option is computing the inexpensive Pickands estimator. The *Dowd* package has functions named *PickandsEstimator* and *PickandsPlot*, although judging by the succinct documentation, it seems to compute the moment estimator. Since this estimator serves a similar purpose, we have used it. Unfortunately, the results are disappointing (see Figure 2.2): the estimation of ξ under different values of k does not get stable, even though at the end it could be argued that this happens. However, it is not possible to extend the plot (the value of k must be less than a quarter of the total sample size) and those values are very different from the estimations of the GEV distribution we will show next.

This may be due to the fact that these estimators require an iid sample, which is not the case. Trying to find a getaway from this seems like it does not make up for the gains, so it

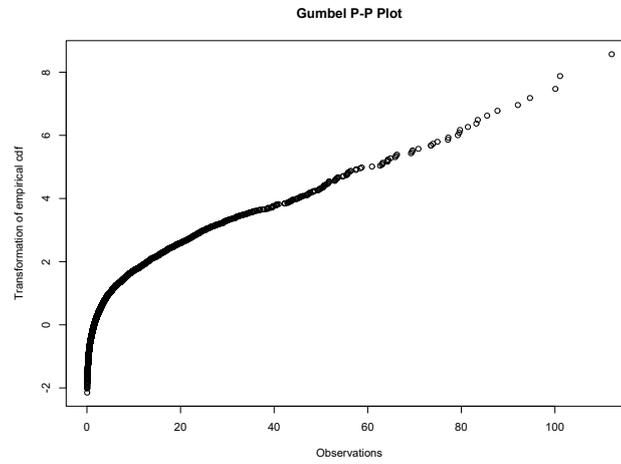


Figure 2.1: Gumbel P-P plot for discharges of the river Oñati.

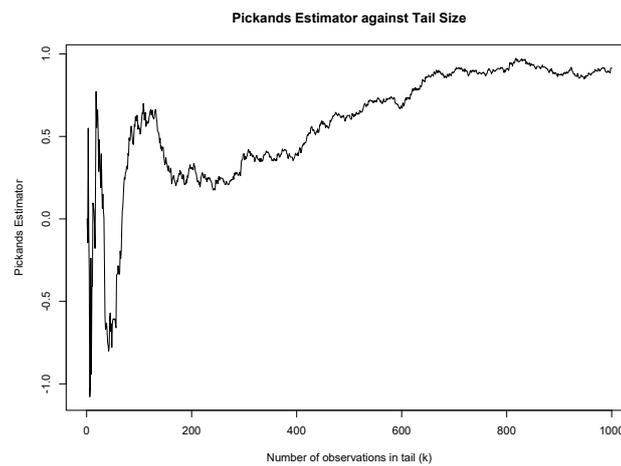


Figure 2.2: Shape estimator for discharges of extremes of the river Oñati considering different tail sizes.

Parameter	Method	95% lower CI	Estimation	95% upper CI
μ	MLE	47.457	56.092	64.726
	L -moments	47.352	55.518	65.200
	EPM		55.248	
σ	MLE	15.384	21.569	27.753
	L -moments	15.526	22.281	29.437
	EPM		21.701	
ξ	MLE	-0.523	-0.260	0.003
	L -moments	-0.562	-0.234	0.048
	EPM		-0.263	

Table 2.2: Estimation of parameters of the GEV distribution, using MLE and L -moments, and 95% confidence intervals, for annual block maxima of discharges of the river Oñati.

seems that it is not possible to get a preliminary idea of the value of ξ , and we will go straight to fitting the GEV distribution.

In order to fit the GEV distribution to the discharges, we have used the function *fevd* from the *extRemes* package for MLE and L -moments and *egevd* from *EnvStats* for EPM. The latter, unfortunately, does not include confidence intervals, and we have not been able to find a library that does. The results can be found in Table 2.2. For the MLE and the estimation via L -moments, a 95% confidence interval for each of the parameters is included. The computation of the covariance matrix of the parameters for the estimation via L -moments is done through parametric bootstrap, whereas the MLE covariances are computed by default using the asymptotic normality of Theorem 1.2.2 and the formulas below such theorem. Confidence intervals have been obtained through the function *ci* in the package *extRemes*.

We will disregard two of the methods and work with the remaining. The EPM would be the choice when ξ is large in absolute value, which does not seem to be the case here. As a consequence, we are not using it anymore. To decide between the MLE or the estimation via L -moments, we will look at diagnostic plots generated by the *fevd* function. By default, it generates 4 plots: a Q-Q plot with the fitted distribution, another Q-Q plot with simulated data, a superposition of the density function corresponding to the estimated parameters and a non-parametric estimation of the density producing the data obtained with the kernel-smoothing method, and the return level plotted against the return period. The last plot has been transformed in such a way that when it is convex, it is a possible evidence that the right tail follows a Weibull distribution (note that this is the opposite of the Gumbel P-P Plot).

We see that both sets of diagnostic plots, displayed in Figures 2.3 and A.1 (the latter can be found in the appendix), are fairly similar. As we know, the Q-Q plot gives an idea of the goodness of the fit depending on how close the points are to a straight line (in this case, the line $y = x$) plotted on the same graph. In this case, the qualitative differences are nonexistent: both models give quite a good fit although they slightly overestimate *small values* (return levels of periods corresponding to few years) while underestimating *large values*. This can also be seen in the last plot, where the L -moments plot seems to be slightly better adjusted than the MLE; however, this difference is almost imperceptible, and the points lie within the 95% confidence interval in both cases. On the other hand, it seems like the empirical quantiles are better adjusted to the quantiles generated from the fitted distribution in the MLE rather than in the estimation via L -moments. It is necessary to remark that this plot changes with every call to the *plot* function because it is based on simulated data, and thus

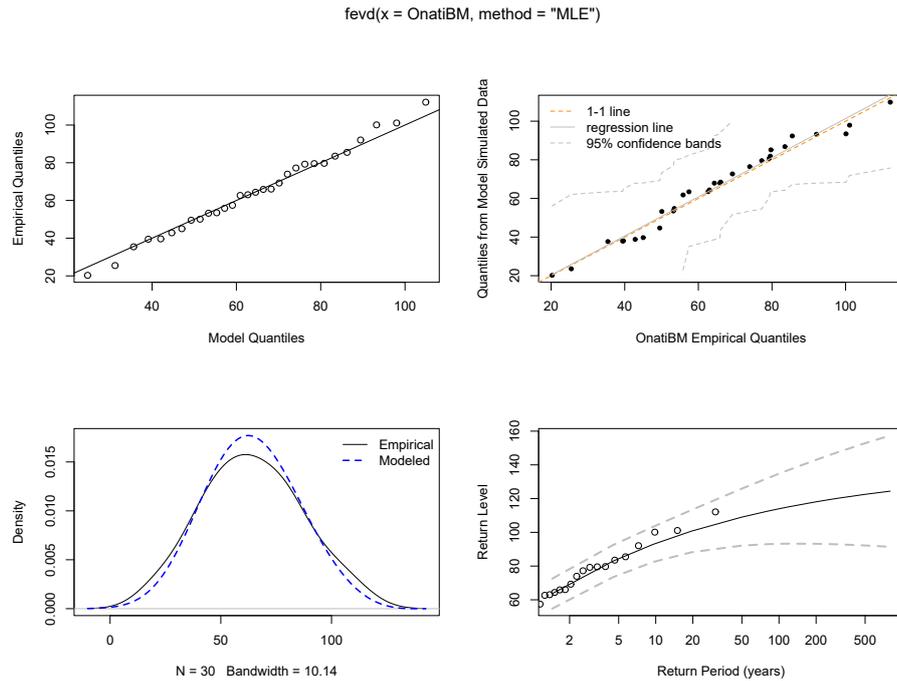


Figure 2.3: Diagnostic plots of the GEV fitting via MLE.

this observation is done after several runs to see the general behaviour. Arbitrary plots have been chosen in both cases. Lastly, it is not clear whether the density plot is better adjusted at its peak by the MLE plot rather than the L -moments plot, because the latter is not fully shown. However, at the sides L -moments performs better. In conclusion, it seems that there is no clear winner between both estimations. Therefore, the selection of one of them will be done later.

We mentioned in the first chapter that a useful preliminary step is testing the hypothesis $\xi = 0$. As we see, in both cases, we cannot reject the hypothesis $\xi = 0$ with a 0.95 significance level, even though we are close. In fact, the ML model rejects it with a significance level of 0.948. However, flood risks are huge problems in so many ways that it is not advisable to be liberal when assessing these risks. This, along with the fact that the qualitative difference of $\xi < 0$ or $\xi = 0$ is huge (in essence, having an upper bound or not on the floods) makes us take the case $\xi = 0$ into account. The estimation of the parameters of the Gumbel distribution using MLE (*fevd* does not allow to fit a Gumbel distribution using L -moments), along with a 95% confidence interval, can be found in Table 2.4. The corresponding diagnostic plot can be found in Figure A.2. As we can see, the parameter estimation is very similar but, at the same time, more accurate, since Gumbel confidence intervals are narrower. In addition, the parameters μ and σ when $\xi < 0$ are greater than those from the case $\xi = 0$, which makes sense: since the right tail of the Gumbel is heavier than when $\xi < 0$, then in order for the latter to fit large values, μ and σ must increase.

On the other hand, it seems from the diagnostic plots that the Gumbel model performs worse than both MLE and L -moments, since each of the plots is worse fitted than its analogue in the other models.

These models are not usually just analyzed by its parameters, since they do not have an

	95% lower CI	Point estimation	95% upper CI
μ	45.427	53.206	60.906
σ	15.044	20.519	25.996

Table 2.4: Estimation of parameters of the Gumbel distribution, using MLE, along with a 95% confidence interval, for annual block maxima of discharges of the river Oñati.

easy interpretation due to the complex expression of the GEV distribution:⁸. That is, in part, why return levels (in the end, quantiles) are commonly used to describe these functions. This is a sensible choice, considering that sometimes return levels are precisely what we want to know from a given phenomenon. In Table 2.3 we compute return levels associated to return periods of 2, 50 and 100 years, using the function *return.level*, found in the library *extRemes*.

Period	Method	95% lower CI	Return level	95% upper CI
2-year	GEV (MLE)	54.740	63.631	72.523
	GEV (<i>L</i> -moments)	54.546	63.633	73.107
	Gumbel	52.074	60.727	69.380
50-year	GEV (MLE)	92.107	108.950	125.794
	GEV (<i>L</i> -moments)	93.243	112.506	133.449
	Gumbel	108.233	133.274	158.314
100-year	GEV (MLE)	93.179	113.939	134.699
	GEV (<i>L</i> -moments)	95.243	118.261	146.617
	Gumbel	118.882	147.601	176.319

Table 2.3: Confidence intervals of 2, 50 and 100 year return-levels as computed with the GEV (fitted via MLE and *L*-moments) and Gumbel distribution.

The first observation we make is that MLE and *L*-moments estimations are pretty similar in all three periods. It is worth noting that for MLE and *L*-moments, the 2-year return level has been exceeded exactly 15 times over the last 30 years, which is an exciting result for its accuracy. The 50-year and 100-year return levels have not been exceeded, which is not too surprising taking into account that we only have data for 30 years. Furthermore, even though their differences increase with the return period, this is not very noticeable (3.556 years in the 50-year period, and 4.322 years in the 100-year period). MLE estimations seem more conclusive than estimation via *L*-moments, since the corresponding confidence interval is narrower. On the other hand, Gumbel estimations for the return level quickly go off, even though their estimators are still technically compatible with the other ones if we look at confidence intervals. This was already expected from a theoretical point of view, as mentioned earlier. Besides, it could also be seen from the last of the diagnostic plots (Figures 2.3, A.1 and A.2) that Gumbel estimation of return levels would eventually surpass those from a bounded GEV distribution. Finally, note that Gumbel confidence intervals are wider than the other two when the periods are 50 and 100 years.

To end this section, we will compute the return periods of the discharges that mark the declaration of yellow, orange and red code. These are $80.48m^3/s$, $99.48m^3/s$ and $120.02m^3/s$ respectively. This is not directly implemented in the *extRemes* package, but can be easily

⁸In spite of μ and σ being location and dispersion measurements of the distribution, they do not coincide with the mean nor with the standard deviation. Similarly, the precise meaning of σ , even though its sign determines the boundedness of the right tail, is unknown.

	Emergency levels	Return periods		
		MLE	L -moments	Gumbel
Yellow	80.48	4.33	4.19	4.29
Orange	99.48	17.77	14.63	10.04
Red	120.02	290.25	126.79	26.45

Table 2.5: Return period (years) of the discharge levels (m^3/s) chosen by the Basque Government to declare an emergency situation, using GEV distributions.

computed using *pevd* function from that library, which computes the *cdf* of a GEV or GP distribution. It is shown in Table 2.5.

It was expected that Gumbel estimations were going to be lesser than the other two, since it gives bigger results for the return levels. However, which we did not expect is the difference in the estimation of the return period for the red code with MLE and L -moments. Note that, amongst all the estimations we have shown, this is the first in which they greatly differ. This illustrates how sensitive extreme value estimation is and the inherent uncertainty of dealing with extreme values. Recall that red code has never been declared in Winter in the river Oñati, and it comes as no surprise that we struggle when drawing conclusions from something we have never seen. Furthermore, the small sample size hinders our task even more.

2.3.2 POT approach

Selection of the threshold

We now proceed to study the statistical behaviour of exceedances over a valid threshold. Recall that, by a valid threshold, in theory, we mean a value u such that, if F is our target cdf, then $\log(F(u)) \approx -(1 - F(u))$. In practice, however, we need to have values larger than the threshold to fit the GP distribution, so u cannot be very large. There are several ways to choose this threshold: for example, the threshold of the exceedances series stored in the National River Flow Archive, dependent on the UK Centre for Ecology & Hydrology are taken in such a way that there are, on average, 5 exceedances per year.

We will go ahead with the methods described in the previous chapter. The first one, the mean life residual plot, relies on the linearity on the set of valid thresholds of the expectance of the GP distribution.

It should be noted that, theoretically, before being able to use the mean life residual plot properly, the declustering of the exceedances must be done. Obviously, the problem is that we do not know what those exceedances are before choosing the threshold, so we will be cautious when drawing conclusions from the plot. According to Figure 2.4, even though we cannot say anything conclusive, it does seem like valid thresholds are bigger than 20. We will now draw the parameter stability plot for ξ : if a fitted GP distribution with threshold u has parameters σ_u and ξ_u , if we plot ξ_u against u , then for valid thresholds the graph should be constant. In order to fit the GP distribution, we first need to find clusters of exceedances and pick cluster maxima, which are assumed to be independent. An easy way to do this has already been described. We will use the *decluster* function from the package *extRemes*, which computes an estimate of the number of clusters, their maxima and an estimation of the extremal index. It also allows us to separate measures of different years, which are assumed to be independent and, therefore, that they belong to different clusters. In Figure 2.5 we

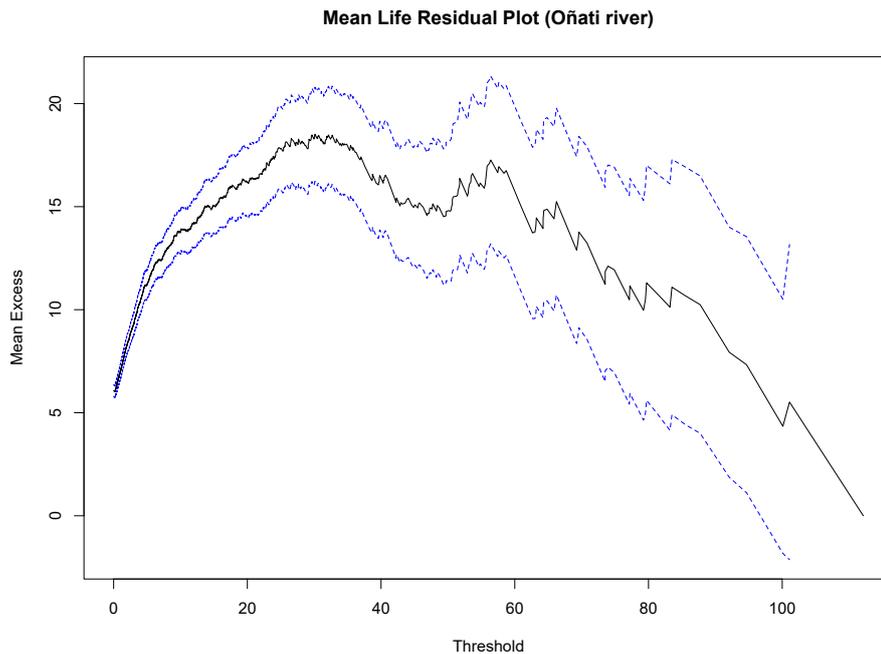


Figure 2.4: Mean Life Residual Plot with 95% confidence interval.

see an example of a declustering. The narrow year corresponds to the period 2002-2003 with almost no measurements in Winter.

From Figure 2.6 and Figure 2.7, which contain parameter stability plots for ξ using *MLE* and *L*-moments, we get very similar results to Figure 2.4, and we see more clearly that appropriate values of thresholds start more or less halfway 20 and 40, as the value of $\hat{\xi}_u$ is almost constant. Furthermore, in both plots the 95% confidence interval contains the fitted value of ξ from the Block Maxima approach, which gives us more evidence. Considering the mean life residual plot and both parameter stability plots, we have decided to pick $u = 30$. This gives us 102 different clusters, which amounts to an average of 3.4 exceedances per year. In comparison to the aforementioned criterion of having at least an average of 5 exceedances per year, this is very poor. However, to get this value, the maximum possible threshold would be close to 22.2, which does not seem like a good choice if we look at the plots.

Fitting the GP distribution

We now turn to estimating the *GP* distribution of exceedances of discharges over $30m^3/s$. This will allow us to answer the same questions than with the previous approach and, hopefully, with similar answers.⁹ We will only use *MLE* and *L*-moments, because we are already confident on where ξ lies, and there is no need to use *EPM*.

⁹Recall that both approaches are expected to generate similar results. The Block Maxima is used because of its convenience, while the advantages of using POT is that the sample size is usually higher. A big drawback of the latter is that the preprocess of the data is a delicate matter, and the conclusions extracted can be very sensitive to both the way of finding clusters and the election of the threshold.

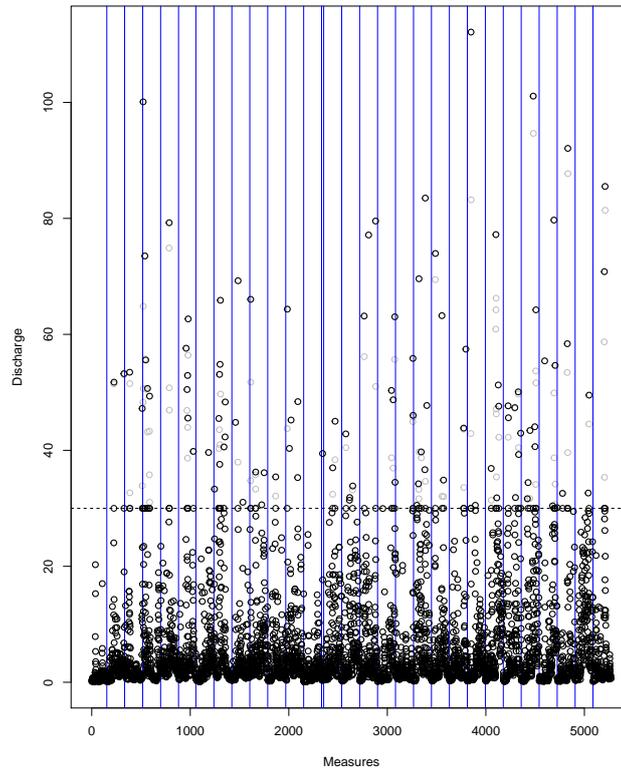


Figure 2.5: Example of output of *decluster* with the threshold $u = 25$. Blue vertical lines determine years.

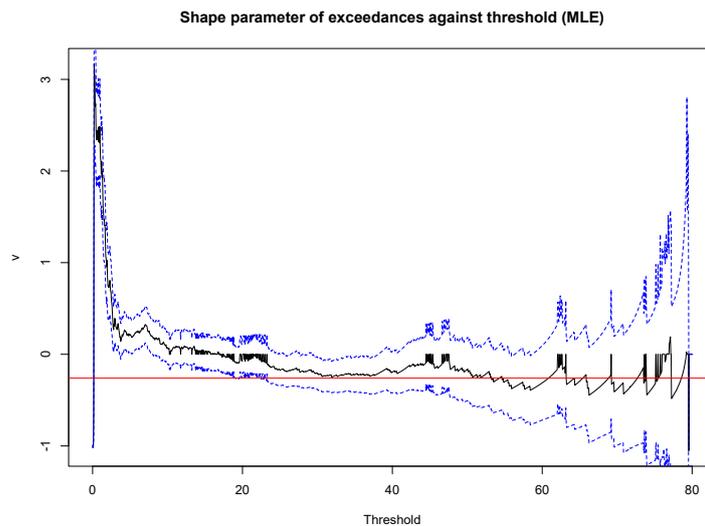


Figure 2.6: Parameter ξ of the fitted GP distribution (using MLE) for different thresholds with 95% confidence interval (computed through asymptotic normality). The red line denotes -0.26 , the fitted value for ξ in the Block Maxima section.

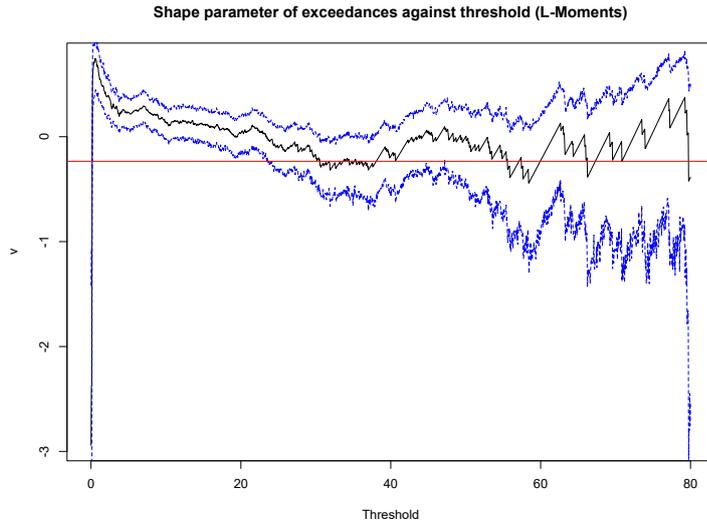


Figure 2.7: Parameter ξ of the fitted GP distribution (using L -moments estimation) for different thresholds with 95% confidence interval (computed through parametric bootstrap). The red line denotes -0.234 , the fitted value for ξ in the Block Maxima section.

Parameter	Method	95% lower CI	Point estimation	95% upper CI
$\tilde{\sigma}$	MLE	18.529	25.009	31.489
	L -moments	18.443	24.564	33.071
ξ	MLE	-0.385	-0.210	-0.034
	L -moments	-0.452	-0.189	0.008

Table 2.6: 95% confidence level for the parameters of the GP distribution ($u = 30$) using MLE and L -moments.

The results appear on Table 2.6. We can make the following remarks:

- The estimation of ξ is perfectly compatible with its estimation when fitting the GEV distribution. This was already known, since it was an auxiliary reason when choosing the thresholds. However, we could argue that this estimation is better, since the confidence intervals are narrower both for MLE and L -moments estimation. Again, comparing lengths of the confidence intervals, the MLE seems more precise. This is good, because this estimator is closer to the GEV estimators.
- If we look at the diagnostic plots (Figures A.3 and A.4) of the GP fittings, we find even less differences than with the previous approach. The implementation in the library *extRemes* of the densities plot for the GP fitting had a major flaw that we have fixed. The issue and our proposed solution is described in Appendix B.
- Recall, from Pickands-Balkema-de Haan Theorem that, if μ , σ and ξ are the parameters of the GEV distribution and $\tilde{\sigma}$ is the scale parameter of the GP distribution fitting exceedances over a threshold u , then $\tilde{\sigma} = \sigma + \xi(u - \mu)$. If we compute a confidence interval for $\tilde{\sigma}$ using this formula, the parameters from the MLE of the GEV distribution, and Delta-Method, the 95% confidence interval we get is $(15.950, 40.766)$ with a point estimation of 28.318. It is clearly worse than the other two estimators, since the confidence interval is considerably wider. This is understandable because two different

factors are adding uncertainty to the estimation worse: the error from the GEV distribution and the choice of the threshold. Nevertheless, it is interesting to note that both approaches are compatible.

- Similar conclusions are drawn when testing for $\xi = 0$ against $\xi \neq 0$, and we seem to be in the border of both possibilities, even though this time with MLE we would reject $\xi = 0$ with a significance level of 0.95. If we fit the exponential distribution (recall, GP with $\xi = 0$) for these exceedances, we get an estimation of $\tilde{\sigma}$ of 20.654 with a 95% confidence interval of (16.626, 24.682). These results are partially compatible with the estimations of Table 2.6, since MLE and L -moments confidence intervals contain the exponential estimator, but the latter does not contain the MLE, and hardly contains the one from L -moments. Note that in the Block Maxima approach all three estimations are fully compatible. This may be due to the sensitivity of the threshold selection, the way of clustering data, the small sample size, or perhaps due to the case that extreme models with $\xi = 0$ do not fit these data properly. The last one may be the case, since diagnostic plots for the exponential distribution (Figure A.5) seem to fit worse than the other two sets of plots (Figures A.3 and A.4).

Just like the GEV distribution, exceedances models are also compared and described by means of return levels. Table 2.7 shows 2-year, 50-year and 100-year return levels along with 95% confidence intervals. Comparing this table with the analogous for the GEV distribution (Table 2.3) we find interesting results. While the 2-year level differs from the GEV estimates, the 50-year and the 100-year levels are more similar. It is surprising, since in case of expect differences, we would expect them when talking about large return levels, just as happened in the Block Maxima approach. However, this has not been the case. This is not the only *strange* thing going on with the estimation of the 2-year return level under this approach, since the confidence interval coming from MLE is considerably wider than the L -moments one. This is the first time we have observed this phenomenon, which does not occur in the rest of the return levels.

Period	Method	95% lower CI	Return level	95% upper CI
2-year	GP (MLE)	60.115	69.727	79.338
	GP (L -moments)	63.550	69.740	75.322
	Gumbel	62.115	69.896	77.676
50-year	GP (MLE)	86.387	108.786	131.186
	GP (L -moments)	89.385	110.815	139.846
	Gumbel	115.632	136.378	157.124
100-year	GP (MLE)	87.601	114.269	140.938
	GP (L -moments)	90.758	116.834	155.523
	Exponential	127.156	150.694	174.2325

Table 2.7: Confidence intervals of 2, 50 and 100 year return-levels as computed with the GP (fitted via MLE and L -moments) and exponential distribution.

Finally, in Table 2.8 we can find the estimates for the return periods of the yellow, orange and red code, just like we did in Table 2.5 within the Block Maxima approach. Once again, we observe similar qualitative results in both frameworks. The return periods of the exponential distribution give lower estimates. Note that this is an expected result on extreme quantities, since floods within this model are not bounded. It is very interesting to note, on the other hand, that MLE and L -moments estimation have a similar behaviour in both approaches, since they have been almost equal up to computing return periods of large discharges.

	Emergency levels	Return periods		
		MLE	L -moments	Exponential
Yellow	80.48	4.06	3.97	3.38
Orange	99.48	18.93	16.88	8.50
Red	120.02	238.53	152.53	22.98

Table 2.8: Return period (years) of the discharge levels (m^3/s) chosen by the Basque Government to declare an emergency situation, using GP distributions.

2.3.3 Conclusion and future lines of work

To conclude, we will now expose several interesting conclusions we have learnt from the study of the floods of the river Oñati, and possible task that could be done to further refine the analysis.

First, thorough the Peak-over-Threshold approach we have seen, as a general rule, how the results coincide with the results extracted from the Block Maxima approach. This is good news for two reasons: it validates the theory we have developed in the first chapter, and the elections we have made, such as choosing the clustering technique or the threshold that would be used to fit the GP distribution. As we have seen, small differences (for example, in the parameters) become huge when trying to get insight from something extreme, such as large return levels or periods. However, our results have been consistently compatible using different methods. This cannot be but a consequence of elections done right.

In this line, it is surprising how reliable graphical methods, such as the mean life residual plot, or the parameter stability plot, have been. Clearly, they are bound to work from a theoretical point of view, but we have used them with a real application, and they have performed meritoriously. Described rules of thumb such as the 0.9 quantile of the dataset (15.79) or the \sqrt{n} -order statistic (0.165), where n is the sample size, would have clearly been wrong.

There are two things to be discussed. First of all: Block Maxima or Peak-over-Threshold? If we look at the conciseness of the results, then POT has shorter confidence intervals, and we can interpret that as having a more conclusive result. This is expected, since we are fitting a distribution with less parameters and more data at the same time, so our certainty increases. On the other hand, it is clearly so much harder, due to the already mentioned elections that have to be done before fitting the GP distribution, and that can end up distorting our problem. Note, in addition, that we have assessed the quality of the POT methods using Block Maxima, but otherwise we would not have had any clue on whether we were on the right path or not. Taking into account how easy the block maxima methods are to set up, it seems a good idea to use it to validate exceedances thresholds, but use the latter to draw conclusions.

The other thing to discuss is whether to accept $\xi = 0$ or not. As we have seen, this is very relevant when estimating return periods and return levels. While it is true that diagnostic plots have shown a slightly better performance for *full* models, it may be better to adopt this model. The main reason, as we already mentioned, is the fact that $\xi = 0$ implies that floods are not bounded. If $\xi < 0$, then modelled floods are bounded, and this upper bound is a function of the parameters estimated. This is not only realistically wrong, but also imprudent if we want to get insight on such a destructive and dangerous event.¹⁰ For

¹⁰Recall Gumbel's quote: *however big floods get, there will always be a bigger one coming.*

example, even though we can confidently assume that human life is not longer than 200 years, at this moment we cannot bound winter floods in the river Oñati with the same confidence, especially considering the flood of $223m^3/s$ in June of 1993. In principle, this could also be possible in winter.

There are certain things that could extend the work we have done to try and get more conclusive results:

1. Recall that the variances of L -moments estimations are computed through parametric bootstrap, while MLE variances are computed through results of asymptotic normality. However, these results exist for L -moments, although the formulas for Block Maxima reported in the literature are extremely complicated. It would be interesting to see the comparison between confidence intervals of MLE and L -moments using the same procedure, and check if MLE estimators are still more conclusive than L -moments.¹¹
2. Recall that we are only modelling winter floods, since our model required stationarity of the data. A possible option would be to introduce a non-stationary model. As we briefly mentioned, these models treat the parameters of the GEV/GP distributions as functions of the time. The first task would be to see what kind of functional relations are more adequate and then proceed to fit the model; in this case, seasonality should also be considered for the election of the function. However, if the small sample size already seems like a problem, it would be even worse with more parameters to estimate. In general, note that having a larger sample size would solve most of our problems. For example, we would have more information on whether $\xi = 0$ or not, and in general our estimations would be more precise.

With all the limitations, both theoretical (e.g., need of stationarity) and practical (e.g., incomplete data, small sample size,...) we have obtained valuable and consistent results. This shows the power of Extreme Value Theory.

¹¹For small sample sizes. If the sample size is big, then we already know that the MLE is the way to go.

Bibliography

- [Ach16] D. Acharya. “Dowd: Functions Ported from ‘MMR2’ Toolbox Offered in Kevin Dowd’s Book Measuring Market Risk”. In: (2016). R package version 0.12. URL: <https://CRAN.R-project.org/package=Dowd> (visited on 09/19/2020).
- [BGT04] J. Beirlant et al. *Statistics of Extremes*. John Wiley & Sons, Ltd New York, 2004. DOI: 10.1002/0470012382.
- [BS17] A. Bücher and J. Segers. “On the maximum likelihood estimator for the Generalized Extreme-Value distribution”. In: *Extremes* 20.4 (2017), pp. 839–872. DOI: 10.1007/s10687-017-0292-6.
- [CH94] E. Castillo and A.S. Hadi. “Parameter and quantile estimation for the generalized extreme-value distribution”. In: *Environmetrics* 5.4 (1994), pp. 417–432. DOI: 10.1002/env.3170050405.
- [CH97] E. Castillo and A.S. Hadi. “Fitting the Generalized Pareto Distribution to Data”. In: *Journal of the American Statistical Association* 92.440 (1997), pp. 1609–1620. DOI: 10.1080/01621459.1997.10473683.
- [CM90] J.A. Cuesta and C. Matrán. “A Short Proof of the Law of Convergence of Types”. In: *Sankhyā: The Indian Journal of Statistics* 52.2 (1990), pp. 259–260.
- [Col01] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer London, 2001. DOI: 10.1007/978-1-4471-3675-0.
- [Cou99] National Research Council. *Improving American River Flood Frequency Analyses*. National Academies Press Washington, DC, 1999. DOI: 10.17226/6483.
- [DEH89] A. L. M. Dekkers, J. H. J. Einmahl, and L. De Haan. “A Moment Estimator for the Index of an Extreme-Value Distribution”. In: *The Annals of Statistics* 17.4 (1989), pp. 1833–1855. DOI: 10.1214/aos/1176347397.
- [DH89] A. L. M. Dekkers and L. de Haan. “On the Estimation of the Extreme-Value Index and Large Quantile Estimation”. In: *The Annals of Statistics* 17.4 (1989), pp. 1795–1832. DOI: 10.1214/aos/1176347396.
- [Dom15] C. Dombry. “Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework”. In: *Bernoulli* 21.1 (2015), pp. 420–436. DOI: 10.3150/13-bej573.
- [Dur19] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 2019. DOI: 10.1017/9781108591034.
- [ECF19] J.F. England et al. *Guidelines for determining flood flow frequency—Bulletin 17C*. US Geological Survey, 2019. DOI: 10.3133/tm4b5.
- [GG64] E. J. Gumbel and Neil Goldstein. “Analysis of Empirical Bivariate Extremal Distributions”. In: *Journal of the American Statistical Association* 59.307 (1964), pp. 794–816. DOI: 10.1080/01621459.1964.10480728.

- [GK16] E. Gilleland and R.W. Katz. “extRemes 2.0: An Extreme Value Analysis Package in R”. In: *Journal of Statistical Software* 72.8 (2016), pp. 1–39. DOI: 10.18637/jss.v072.i08.
- [GLS94] J. Galambos, J. Lechner, and E. Simiu, eds. *Extreme Value Theory and Applications*. Springer New York, 1994. DOI: 10.1007/978-1-4613-3638-9.
- [Gne43] B. Gnedenko. “Sur La Distribution Limite Du Terme Maximum D’Une Serie Aleatoire”. In: *The Annals of Mathematics* 44.3 (1943), p. 423. DOI: 10.2307/1968974.
- [GR10] S. Ghosh and S. Resnick. “A discussion on mean excess plots”. In: *Stochastic Processes and their Applications* 120.8 (2010), pp. 1492–1517. DOI: 10.1016/j.spa.2010.04.002.
- [Gum41] E. J. Gumbel. “The Return Period of Flood Flows”. In: *The Annals of Mathematical Statistics* 12.2 (1941), pp. 163–190. DOI: 10.1214/aoms/1177731747.
- [Gum58] E. J. Gumbel. *Statistics of Extremes*. Columbia University Press New York, 1958. DOI: 10.7312/gumb92958.
- [Haa90] L. de Haan. “Fighting the arch-enemy with mathematics”. In: *Statistica Neerlandica* 44.2 (1990), pp. 45–68. DOI: 10.1111/j.1467-9574.1990.tb01526.x.
- [HF06] L. de Haan and A. Ferreira. *Extreme Value Theory*. Springer New York, 2006. DOI: 10.1007/0-387-34471-3.
- [Hos90] J. R. M. Hosking. “L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.1 (1990), pp. 105–124. DOI: 10.1111/j.2517-6161.1990.tb01775.x.
- [HW97] J. R. M. Hosking and J.R. Wallis. *Regional Frequency Analysis*. Cambridge University Press, 1997. DOI: 10.1017/cbo9780511529443.
- [HWW85] J. R. M. Hosking, J. R. Wallis, and E. F. Wood. “Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments”. In: *Technometrics* 27.3 (1985), pp. 251–261. DOI: 10.1080/00401706.1985.10488049.
- [LLR83] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer New York, 1983. DOI: 10.1007/978-1-4612-5449-2.
- [Mil13] S.P. Millard. *EnvStats: An R Package for Environmental Statistics*. New York: Springer, 2013. DOI: 10.1007/978-1-4614-8456-1.
- [MM20] J.J. Morales and C. Martínez. “El País Vasco funciona al ritmo que marca la Ría”. In: *El País* (2020). URL: <https://elpais.com/espana/2020-07-09/el-pais-vasco-funciona-al-ritmo-que-marca-la-ria.html> (visited on 09/19/2020).
- [Pic75] J. Pickands III. “Statistical Inference Using Extreme Order Statistics”. In: *The Annals of Statistics* 3.1 (1975), pp. 119–131. DOI: 10.1214/aos/1176343003.
- [RZ17] H. Rootzén and D. Zholud. “Human life is unlimited - but short”. In: *Extremes* 20.4 (2017), pp. 713–728. DOI: 10.1007/s10687-017-0305-5.
- [SM12] C. Scarrott and A. MacDonald. “A review of extreme value threshold estimation and uncertainty quantification”. In: *REVSTAT - Statistical Journal* 10.1 (2012), pp. 33–60. URL: <https://www.ine.pt/revstat/pdf/rs120102.pdf> (visited on 09/19/2020).

- [Smi85] R. L. Smith. “Maximum likelihood estimation in a class of nonregular cases”. In: *Biometrika* 72.1 (1985), pp. 67–90. DOI: 10.1093/biomet/72.1.67.
- [URA15] URA (Agencia vasca del agua). *Plan especial de emergencias ante el riesgo de inundaciones de la comunidad autónoma del País Vasco*. 2015. URL: https://www.euskadi.eus/contenidos/informacion/planes_inundaciones/es_doc/adjuntos/PEE%20inundaciones-es.pdf (visited on 09/19/2020).
- [URA18] URA (Agencia vasca del agua). *Revisión y actualización de la evaluación preliminar del riesgo de inundación. 2º ciclo*. 2018. URL: https://www.uragentzia.euskadi.eus/contenidos/informacion/2011_epri/es_doc/adjuntos/EPRI_2CICLO_DHC_ORI_Memoria_cas.pdf (visited on 09/19/2020).
- [URA18I] URA (Agencia vasca del agua). *Revisión y actualización de la evaluación del riesgo de inundación (EPRI 2º ciclo). Anexo 1: Registro de eventos de inundación*. 2018. URL: https://www.uragentzia.euskadi.eus/contenidos/informacion/2011_epri/es_doc/EPRI_2CICLO_DHC_ORI_Anexo%201_Registro_inundaciones_cas.pdf (visited on 09/19/2020).
- [Vaa98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. DOI: 10.1017/cbo9780511802256.
- [YSH15] T.W. Yee, J. Stoklosa, and R.M. Huggins. “The VGAM Package for Capture-Recapture Data Using the Conditional Likelihood”. In: *Journal of Statistical Software* 65.5 (2015), pp. 1–33. DOI: 10.18637/jss.v065.i05.
- [ZBK10] P. de Zea Bermudez and S. Kotz. “Parameter estimation of the generalized Pareto distribution—Part I”. In: *Journal of Statistical Planning and Inference* 140.6 (2010), pp. 1353–1373. DOI: 10.1016/j.jspi.2008.11.019.
- [ZW07] M. Zagorski and M. Wnek. “Analysis of the turbine steady-state data by means of Generalized Pareto Distribution”. In: *Mechanical Systems and Signal Processing* 21.6 (2007), pp. 2546–2559. DOI: 10.1016/j.ymsp.2007.01.002.

Appendices

Appendix A

Additional figures

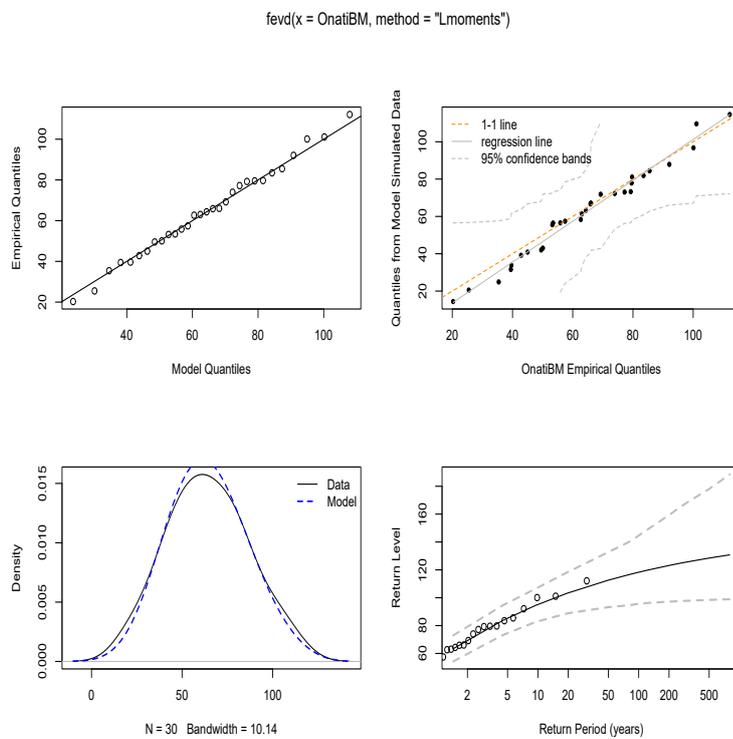


Figure A.1: Diagnostic plots of the GEV fitting via L -moments.

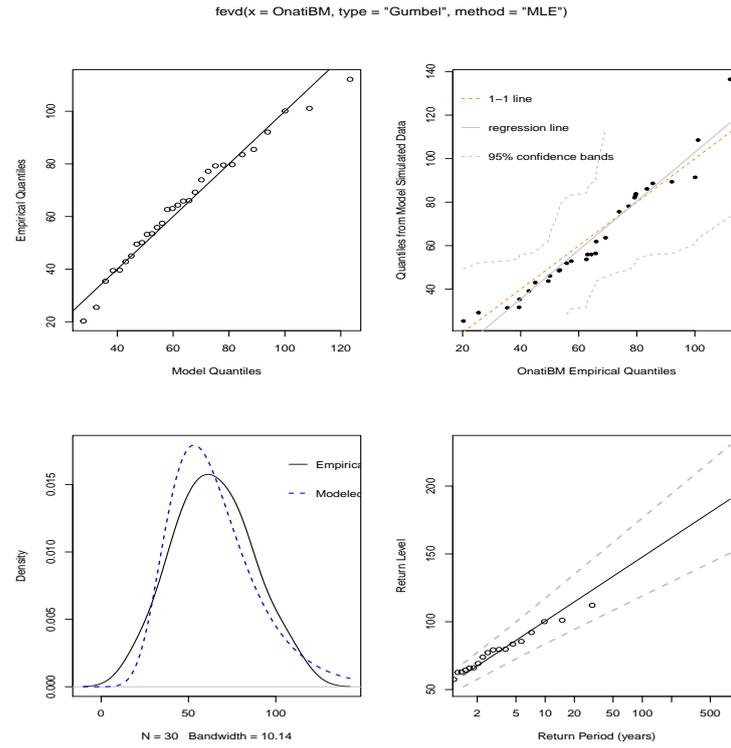


Figure A.2: Diagnostic plots of the Gumbel fitting via MLE.

Diagnostic plots for GP fitting with threshold 30 (MLE)

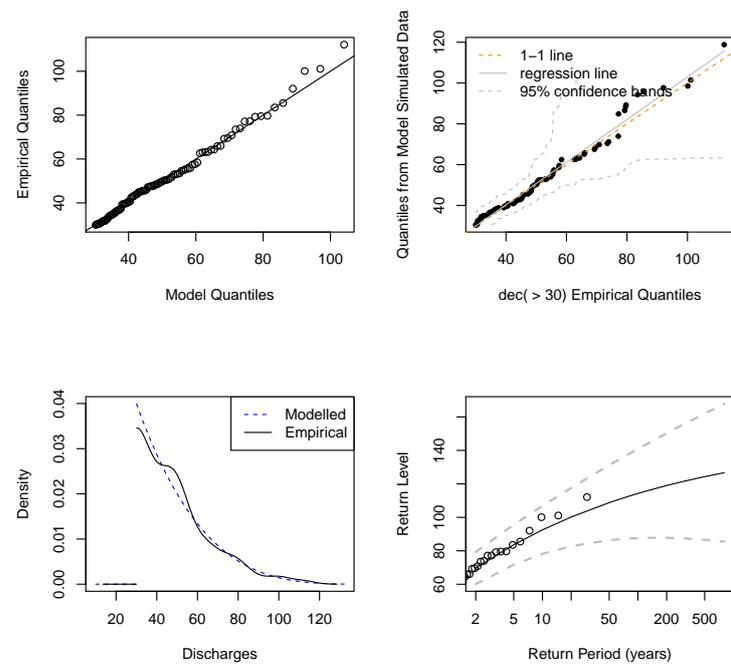


Figure A.3: Diagnostic plots of the GP fitting via MLE.

Diagnostic plots for GP fitting with threshold 30 (L-moments)

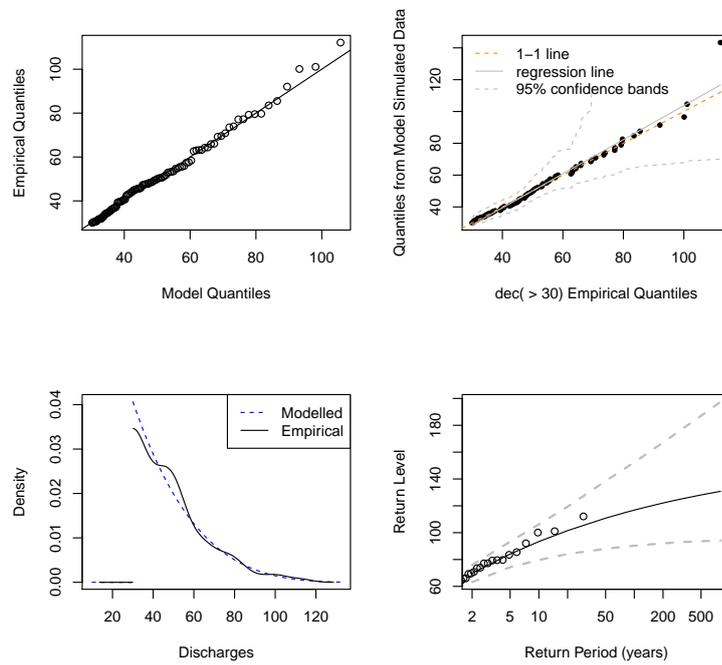


Figure A.4: Diagnostic plots of the GP fitting via L -moments.

Diagnostic plots for GP fitting with threshold 30 (Exponential)

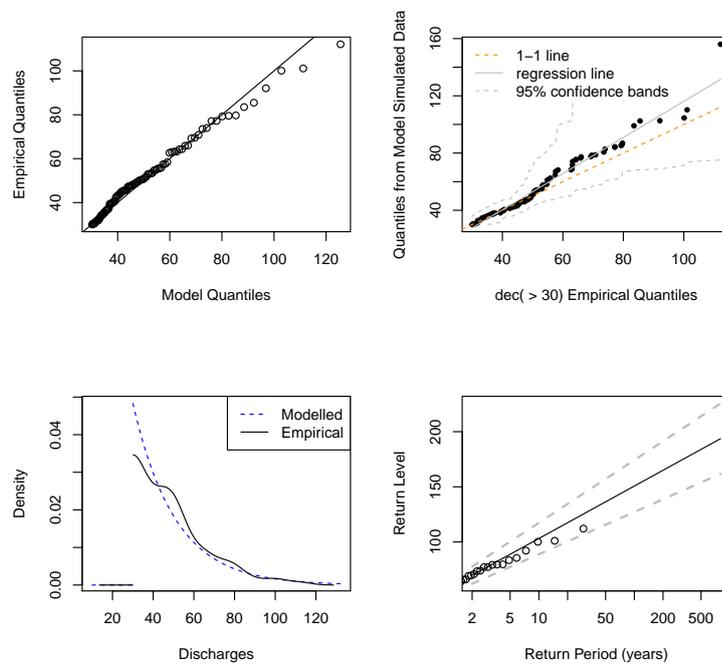


Figure A.5: Diagnostic plots of the GP (Exponential) fitting via MLE.

Appendix B

On the density plot for the GP fitting

Recall that one of the diagnostic plots available in the GP fitting via the *fevd* function in the library *extRemes* is a superposition of two estimations of the density of the distribution generating the data. The so-called empirical one makes no assumption on this distribution, while the model one is based on the assumption that it belongs to the model. According to the documentation of the library, the empirical density is computed via the *density* function found in the base libraries of R. The result of those computations are shown in Figure B.1. This figure shows that there are some serious issues with that plot. This is for two reasons:

- The empirical density should be 0 below the threshold (30 in this case), precisely because we are dealing with exceedances above such threshold. Actually, these begin being 0; however, since *density* works through kernel estimation, the values above the threshold are taken into consideration when computing estimations of the density to the left, what makes this estimation begin strictly positive even for values below 20.
- For the same reason, the first values at the right of the threshold are shrunk as a result of the kernel estimation and the fact that there are no observations at all at the left of the threshold T . In fact, the maximum of the density function should be reached at its threshold, but it is not the case in the plot. This is because it is influenced by the values at its left.

To solve this, we have decided to do the following. Assume x are the values chosen to estimate the density function and $f(x)$ their kernel estimations.

1. Consider the vector $X^- = (x_1^-, \dots, x_n^-)$ whose components are the values of x at the left of the threshold in reversed order, i.e., $x_1^- > x_2^- > \dots > x_n^-$.
2. Let $X^+ = (x_1^+, \dots, x_n^+)$ the ordered vector composed by the first n values of x at the right of the threshold and define

$$\tilde{f}(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \leq T \\ f(x_i^+) + f(x_i^-) & \text{for } i = 1, \dots, n \\ f(x) & \text{if } x > T \text{ and } x \notin X^+. \end{array} \right\}$$

The fact that the function density chooses the values x equispaced, makes the new function \tilde{f} to be a new density function. Moreover, the characteristics of the initial sample, jointly with the properties of the kernel estimators, makes the function \tilde{f} to achieve its maximum at the threshold.

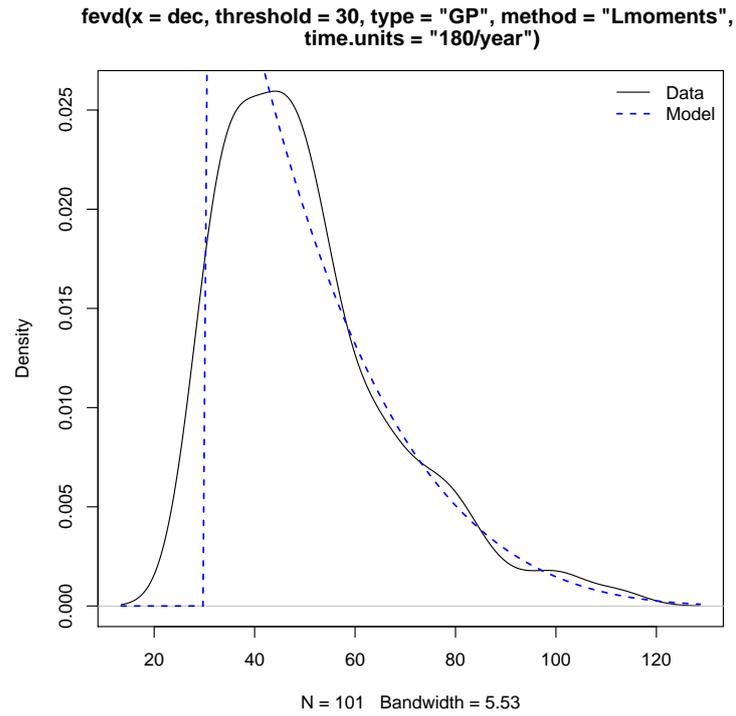


Figure B.1: Density diagnostic plot with the default implementation.

The code of our implementation can be found in the function `plotGPDensity` of the file `functions.R`. Figure B.2 includes a plot of these new points, joined via linear interpolation, and maintaining the density of the fitted GP distribution.

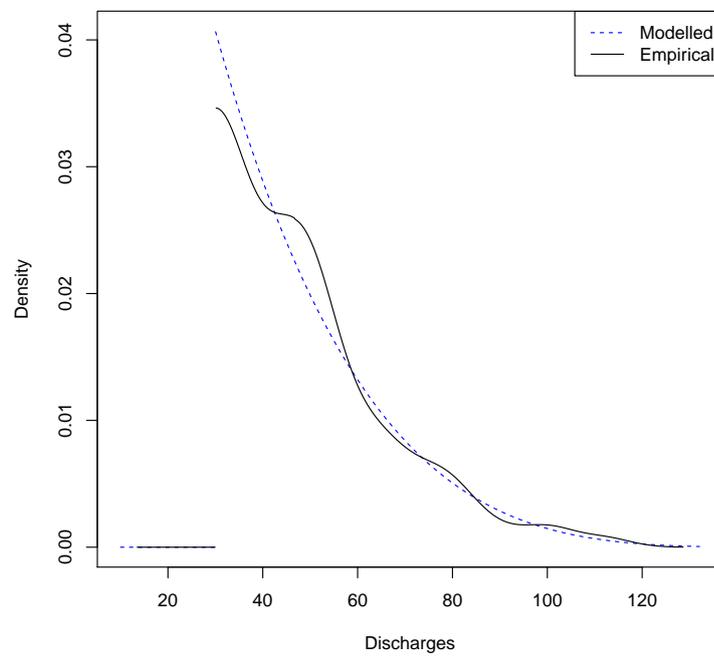


Figure B.2: Our density diagnostic plot.