Univariate (and multivariate) extreme value theory

Mathieu Ribatet (mathieu.ribatet@ec-nantes.fr)

Ecole Centrale de Nantes

Course details

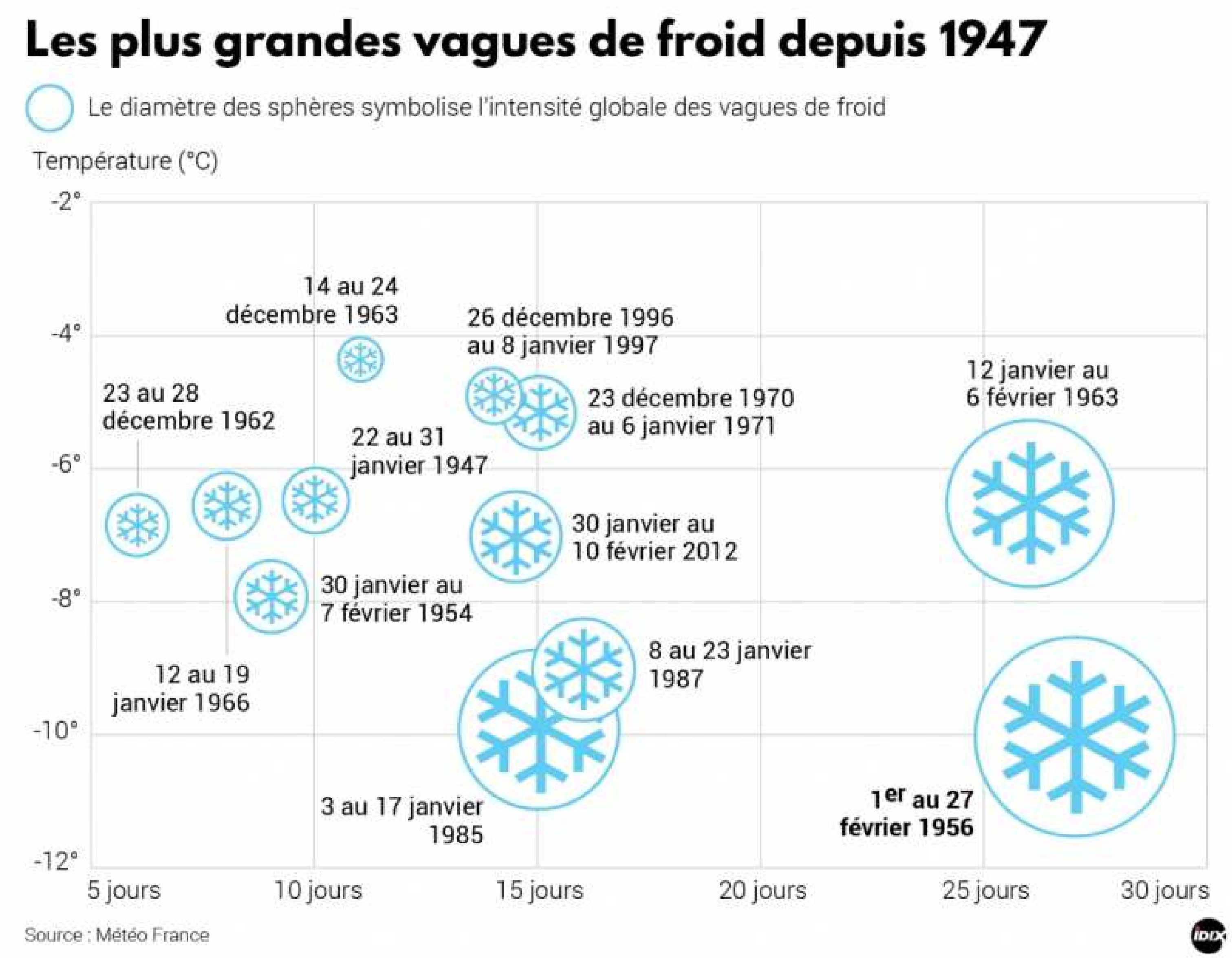
- Supplementary material (if any) can be downloaded from

  `http://mribatet.perso.math.cnrs.fr/teaching.html`

- Grading: Written exam

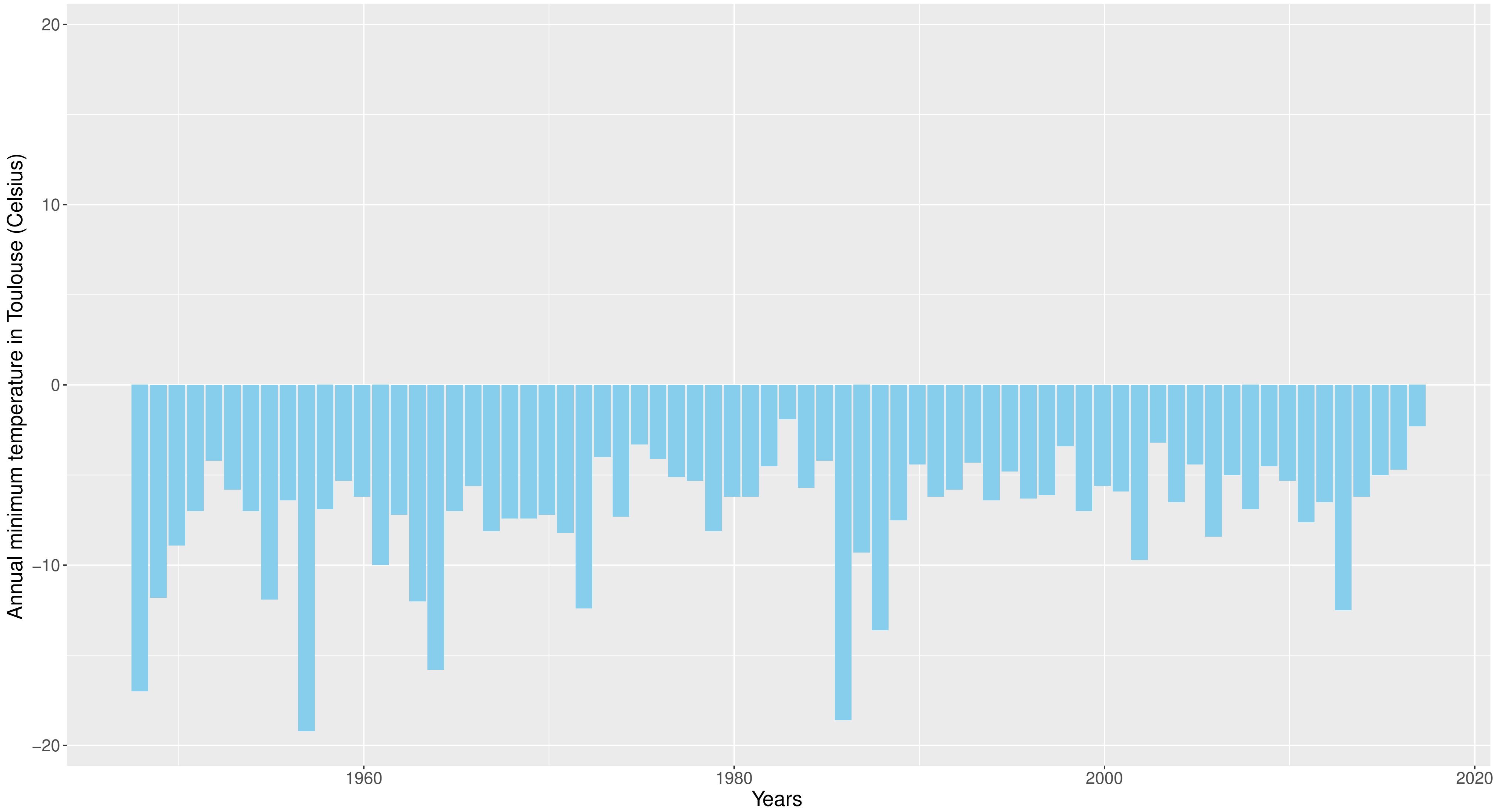
Bibliography

- Coles (2001) An Introduction to Statistical Modeling of Extreme Values, Springer
- de Haan and Ferreira (2006) Extreme Value Theory: An Introduction, Springer
- Resnick (1987) Extreme values, Regular variation and Point processes, Springer
- Embrechts, Klüppelberg and Mikosch (1997) Modelling Extreme Events for Insurance and Finance, Springer


Les plus grandes vagues de froid depuis 1947
Le diamètre des sphères symbolise l'intensité globale des vagues de froid
Température (°C)
-2°
-4°
-6°
-8°
-10°
-12°
14 au 24 décembre 1963
26 décembre 1996 au 8 janvier 1997
23 au 28 décembre 1962
23 décembre 1970 au 6 janvier 1971
12 janvier au 6 février 1963
22 au 31 janvier 1947
30 janvier au 10 février 2012
30 janvier au 7 février 1954
8 au 23 janvier 1987
12 au 19 janvier 1966
3 au 17 janvier 1985
1er au 27 février 1956
5 jours
10 jours
15 jours
20 jours
25 jours
30 jours
Source : Météo France

Environmental extremes

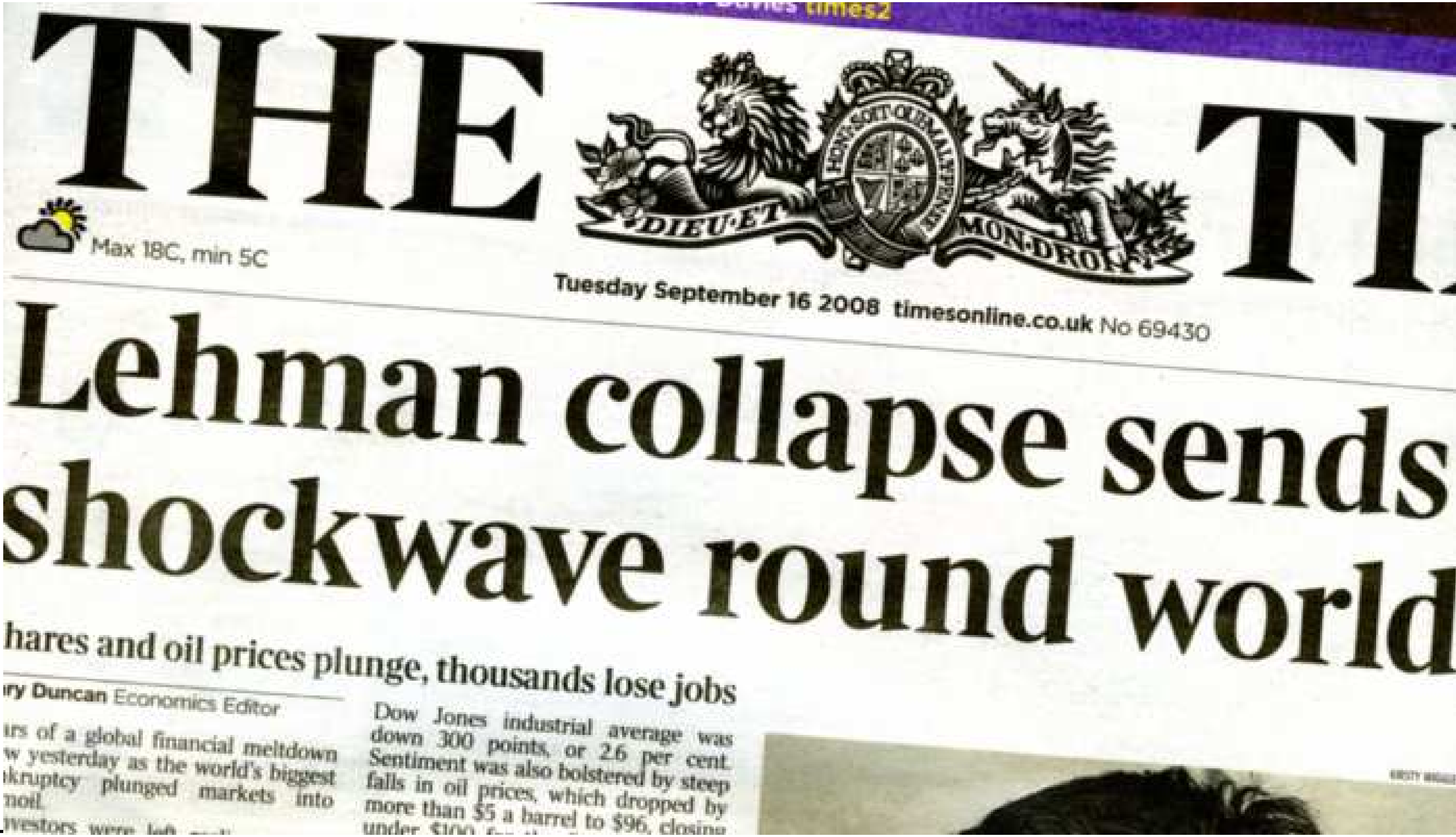
Environmental extremes



Environmental extremes (2)

Knowledge of the distribution of environmental extremes might be useful for

- economic reasons, prevent any severe dammage due to a storm, extremely cold temperatures, ... yielding to economic losses;
- policy management to characterize the potential human losses if some extreme weather events occur—France 2003.



THE TI

Max 18C, min 5C

Tuesday September 16 2008 timesonline.co.uk No 69430

Lehman collapse sends shockwave round world

hares and oil prices plunge, thousands lose jobs

ry Duncan Economics Editor

ars of a global financial meltdown w yesterday as the world's biggest kruptcy plunged markets into noil.

Dow Jones industrial average was down 300 points, or 2.6 per cent. Sentiment was also bolstered by steep falls in oil prices, which dropped by more than $5 a barrel to $96, closing under $100


CHRISTIE'S
LEHMAN BROTHERS
CHRISTIE'S

Financial extremes

Financial extremes (2)

Knowledge of the distribution of financial extremes might be useful to

- be in agreement with the Basel committe, e.g., characterize the value at risk;
- assess to which extent a given company is “at risk”;
- derive optimal portfolio management such as extension of the Markowitz framework.

History

- 1930: Foundations of asymptotic arguments from Fisher and Tippett
- 1940s: Unification and extension of the asymptotic theory by Gnedenko and von Mises
- 1950s: First statistical modelling from asymptotic distribution by Gumbel and Jenkinson
- 1960s: Multivariate maxima
- 1970s: Threshold exceedances
- 1980s: Extremes for stationary processes, point processes approaches
- 1990s: Multivariate modelling strategies, Bayesian approaches
- 2000s: Softwares
- 2010s: Spatial extremes

1. Block maxima

Set up

- Let $X_1, \ldots, X_m \overset{\text{iid}}{\sim} F$ and define the (block) maximum $M_m = \max\{X_1, \ldots, X_m\}$. Clearly we have

$$\begin{aligned}\Pr(M_m \leq x) &= \Pr(X_1 \leq x, \ldots, X_m \leq x)\\ &= \Pr(X_1 \leq x) \times \cdots \times \Pr(X_m \leq x)\\ &= F(x)^m.\end{aligned}$$

- F is unknown so approximate F^m with some relevant distribution.

- As $m \to \infty$ we have

$$F(x)^m \longrightarrow \begin{cases} 0, & F(x) < 1,\\ 1, & F(x) = 1,\end{cases}$$

so $M_m \xrightarrow{\text{D}} x_+$ where $x_+ = \sup\{x \in \mathbb{R} \colon F(x) < 1\}$. We say that the limiting distribution is degenerate.

How to avoid degenerate limiting distribution?

- You already met degenerate distribution, e.g., provided $\mathbb{E}(|X|) < \infty$,

$$\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i \xrightarrow{\text{D}} \mathbb{E}(X), \qquad m \to \infty.$$

- Question: How would you get a non degenerate distribution?

How to avoid degenerate limiting distribution?

- You already met degenerate distribution, e.g., provided $\mathbb{E}(|X|) < \infty$,

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^{m} X_i \xrightarrow{\text{D}} \mathbb{E}(X), \qquad m \to \infty.$$

- Question: How would you get a non degenerate distribution?

- We will just do the same with M_m!

Example 1. Find suitable (normalizing) sequences such that maxima of independent random variables from the

i) Exponential(1)
ii) (unit) Fréchet, i.e., $\Pr(X \leq x) = \exp(-1/x)$, $x > 0$
iii) Uniform(0,1)

distributions have non–degenerate limiting distributions.

Numerical illustration

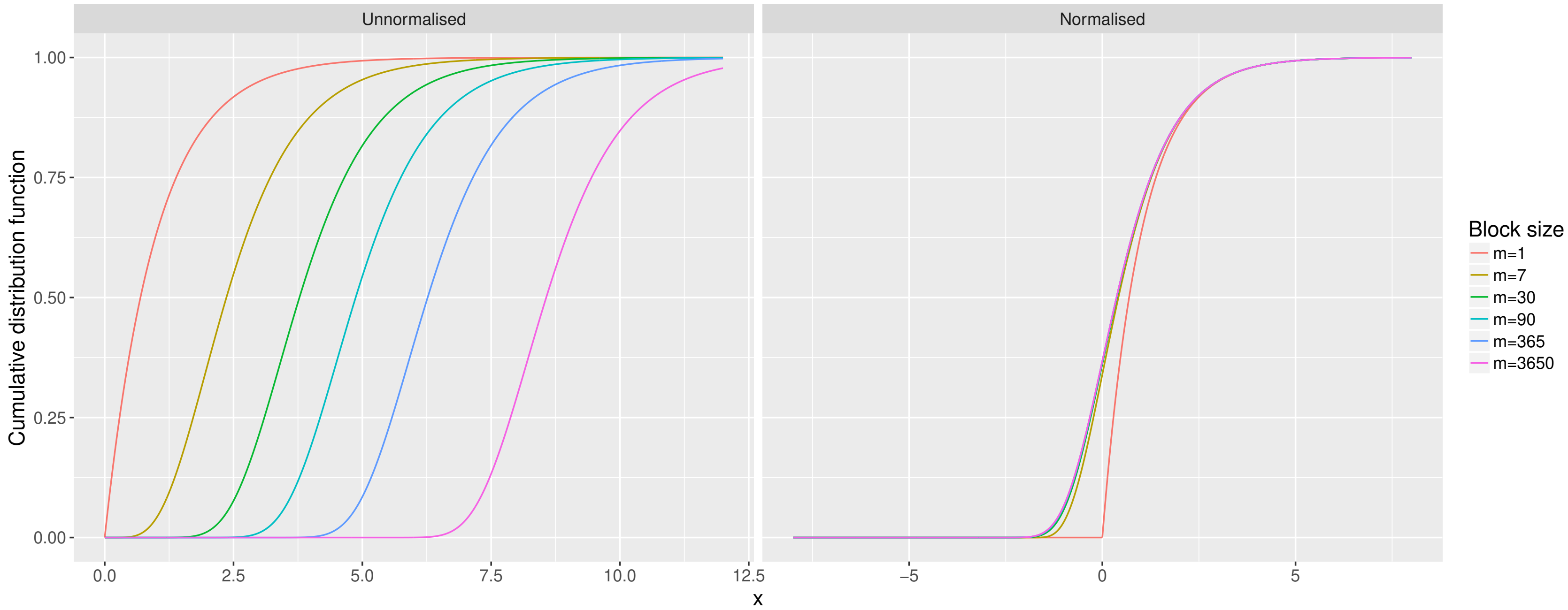


Figure 1: *Distribution of maxima (left) and normalized maxima (right) with* $m = 1, 7, 30, 90, 365, 3650$ *standard Exponential random variables.*

Definition 1. A distribution H is said to be max-stable if for any $k \in \mathbb{N}_*$

$$H^k(x) = H(a_k x + b_k),$$

for some constants a_k and b_k.

Definition 2. Two distributions F and G are of the same type if there are constants $a > 0$ and $b \in \mathbb{R}$ such that $G(ax + b) = F(x)$ for all $x \in \mathbb{R}$.

Clearly if a limiting distribution H for normalized maxima exists, it must be max-stable since as $m \to \infty$

$$\Pr\left(\frac{M_{mk} - b_{mk}}{a_{mk}} \le x\right) \longrightarrow H(x),$$

$$\Pr\left(\frac{M_m - b_m}{a_m} \times \frac{a_m}{a_{mk}} + \frac{b_m - b_{mk}}{a_{mk}} \le x\right)^k \longrightarrow H\left\{\frac{x - \beta(k)}{\alpha(k)}\right\}^k,$$

as the convergence to types theorem states that

$$\frac{a_m}{a_{mk}} \longrightarrow \alpha(k) > 0, \qquad \frac{b_k - b_{mk}}{a_{mk}} \longrightarrow \beta(k), \qquad m \to \infty.$$

Theorem (Extremal types theorem)**.** *If there exist sequences of constants* $\{a_m > 0 : m \geq 1\}$ *and* $\{b_m \in \mathbb{R} : m \geq 1\}$ *such that, as* $m \to \infty$,

$$\Pr\left(\frac{M_m - b_m}{a_m} \leq x\right) \longrightarrow H(x),$$

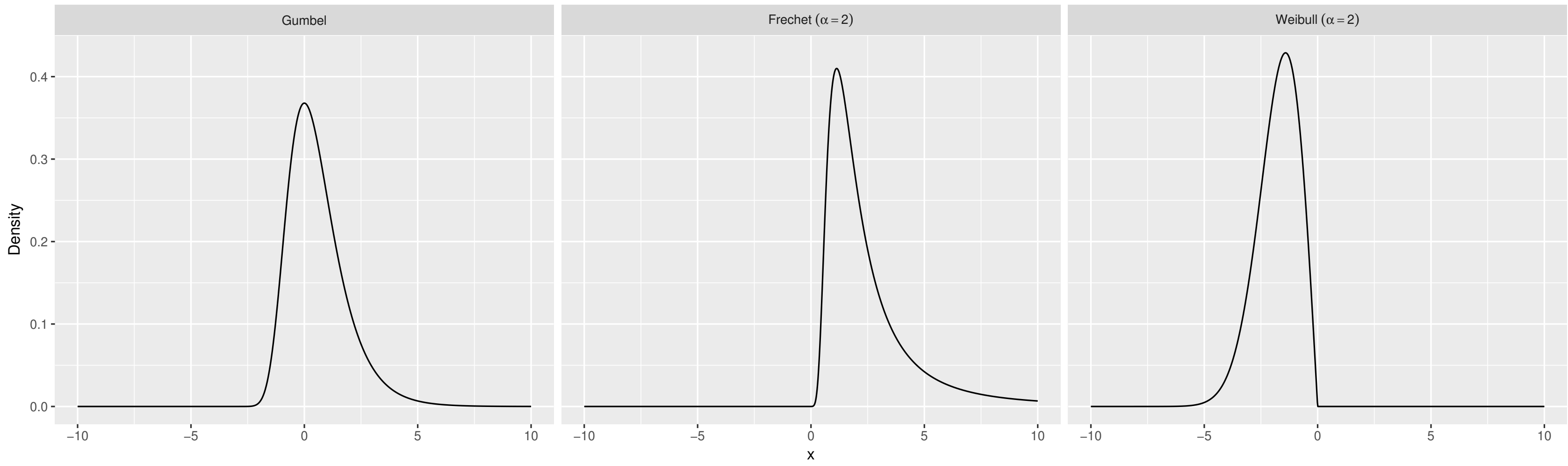
for some non–degenerate distribution H*, then* H *has the same type as one of the following distributions:*

I: $H(x) = \exp\{-\exp(-x)\}$, $x \in \mathbb{R}$;

II: $$H(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0, \alpha > 0; \end{cases}$$

III: $$H(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}, & x < 0, \alpha > 0, \\ 1, & x \geq 0. \end{cases}$$

The three limiting families



- The three limiting distribution are known respectively as the Gumbel, Fréchet and Weibull distributions.
- Note that Fréchet is lower bounded, Weibull is upper bounded.

Statistical application

- From a statistical perspective, we assume that for some (unknown) $a > 0$ and $b \in \mathbb{R}$,

$$\Pr\left(\frac{M_m - b}{a} \leq x\right) \approx H(x),$$

or in other words,

$$\Pr(M_m \leq x) \approx H\left(\frac{x-b}{a}\right) = H_2(x),$$

where H_2 is of the same type as H.
- We thus fit one of the three family to a series of observations of M_m.

- From a statistical perspective, we assume that for some (unknown) $a > 0$ and $b \in \mathbb{R}$,

$$\Pr\left(\frac{M_m - b}{a} \leq x\right) \approx H(x),$$

or in other words,

$$\Pr(M_m \leq x) \approx H\left(\frac{x-b}{a}\right) = H_2(x),$$

where H_2 is of the same type as H.
- We thus fit one of the three family to a series of observations of M_m.
- It is a bit unfortunate that we need to consider three different families. . .

Definition 3. A random variable X has a Generalized Extreme Value distribution GEV(μ, σ, ξ) if its c.d.f. is

$$H(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, \qquad 1 + \xi\frac{x-\mu}{\sigma} > 0.$$

The GEV distribution has three parameters: a location $\mu \in \mathbb{R}$, a scale $\sigma > 0$ and a shape $\xi \in \mathbb{R}$.
The case $\xi = 0$ is derived by a continuity extension, i.e.,

$$H(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \qquad x \in \mathbb{R}.$$

- The shape parameter controls the tail, i.e.,
  - $\xi > 0$ corresponds to the heavy–tailed (Fréchet) case;
  - $\xi = 0$ corresponds to the light–tailed (Gumbel) case;
  - $\xi < 0$ corresponds to the short–tailed (Weibull) case.

Theorem. *If there exist sequences of constants* $\{a_m > 0 : m \geq 1\}$ *and* $\{b_m \in \mathbb{R} : m \geq 1\}$ *such that, as* $m \to \infty$,

$$\Pr\left(\frac{M_m - b_m}{a_m} \leq x\right) \longrightarrow H(x),$$

for some non–degenerate distribution H, *then*

$$H(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, \qquad 1 + \xi\frac{x-\mu}{\sigma} > 0,$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $\xi \in \mathbb{R}$.

Theorem. *If there exist sequences of constants* $\{a_m > 0 : m \geq 1\}$ *and* $\{b_m \in \mathbb{R} : m \geq 1\}$ *such that, as* $m \to \infty$,

$$\Pr\left(\frac{M_m - b_m}{a_m} \leq x\right) \longrightarrow H(x),$$

for some non–degenerate distribution H, *then*

$$H(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, \qquad 1+\xi\frac{x-\mu}{\sigma} > 0,$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $\xi \in \mathbb{R}$.

Remark. Watch out! The theorem above states that if the limit exists it has to be GEV. In general there is no guarantee that such a limit exists, e.g., Poisson distribution.

Definition 4. A distribution F is said to belong to the (max) domain of attraction of the $\begin{array}{l}\text{Gumbel}\\\text{Fréchet}\\\text{Weibull}\end{array}$ distribution if the limiting distribution of $\frac{M_m - b_m}{a_m}$ is $\begin{array}{r}\text{Gumbel}\\\text{Fréchet}\\\text{Weibull}\end{array}$

Example 2. The (max) domain of attraction of the Normal distribution is Gumbel.

- In practice the notion of domain of attraction is of little interest since typically F is unknown—and so is the domain of attraction!

- How one can determine sequences $\{a_m : m \geq 1\}$ and $\{b_m : m \geq 1\}$?
- The von Mises conditions give sufficient (but not necessary) simple conditions, i.e., for a (smooth enough) distribution F the Mills ratio is

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

Then with

$$b_m = F^{-1}\left(1 - \frac{1}{m}\right), \qquad a_m = r(b_m), \qquad \xi = \lim_{x \to x_+} r'(x),$$

the limit distribution of $(M_m - b_m)/a_m$ is GEV with shape ξ.

Example 3. Use the von Mises conditions to check the limiting distribution of maxima from the uniform, exponential, Fréchet and Gaussian distribution.

- Convergence to the limiting distribution may be slow.
- Taking $\xi_m = r'(b_m)$ may give better approximation to the distribution of $(M_m - b_m)/a_m$ for finite m than does using the limiting approximation.

Example 4. For the $N(0,1)$ case we have

$$\xi_7 \approx -0.324, \quad \xi_{30} \approx -0.176, \quad \xi_{90} \approx -0.13, \quad \xi_{365} \approx -0.097, \quad \xi_{3650} \approx -0.068,$$

so the distribution of M_m is short–tailed compared to the Gumbel limit—even when m is very large!

- Convergence to the limiting distribution may be slow.
- Taking $\xi_m = r'(b_m)$ may give better approximation to the distribution of $(M_m - b_m)/a_m$ for finite m than does using the limiting approximation.

Example 4. For the $N(0,1)$ case we have

$$\xi_7 \approx -0.324, \quad \xi_{30} \approx -0.176, \quad \xi_{90} \approx -0.13, \quad \xi_{365} \approx -0.097, \quad \xi_{3650} \approx -0.068,$$

so the distribution of M_m is short–tailed compared to the Gumbel limit—even when m is very large!

- As a consequence, even if we were sure about the Gumbel limit, one should prefer fitting a GEV with an arbitrary shape parameter ξ.

Illustration of the penultimate approximation

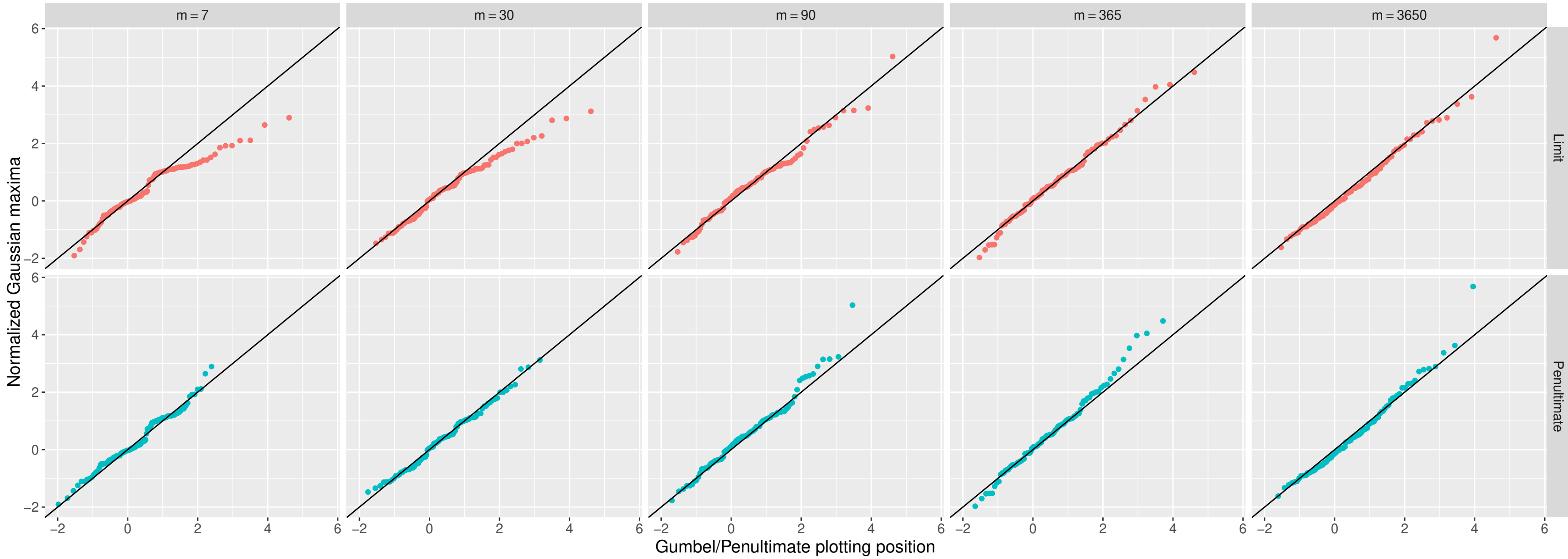


Figure 2: *Illustration of the penultimate approximation with 100 replicated of renormalized N(0,1) maxima with* $m = 7, 30, 90, 365, 3650$.

Recall the extremal types theorem

Theorem. *If there exist sequences of constants* $\{a_m > 0 : m \geq 1\}$ *and* $\{b_m \in \mathbb{R} : m \geq 1\}$ *such that, as* $m \to \infty$,

$$\Pr\left(\frac{M_m - b_m}{a_m} \leq x\right) \longrightarrow H(x),$$

for some non–degenerate distribution H*, then*

$$H(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, \qquad 1 + \xi\frac{x-\mu}{\sigma} > 0,$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $\xi \in \mathbb{R}$.

Statistical interpretation

- We observe a time series of, say, daily values $X_1, X_2, \ldots$ supposed to be independent and identically distributed from F
- We compute the maxima $M_m = \max(X_1, \ldots, X_m)$ of blocks of the original time series
  - Environmental applications: annual maxima with $m = 365$, monthly maxima $m = 30$
  - Finance: annual maxima with $m = 250$, monthly maxima $m = 20$
- We suppose that this new time series of block maxima follows the GEV distribution with unknown parameters μ, σ and ξ.
- We then estimate the parameters and use our fitted GEV for estimations.

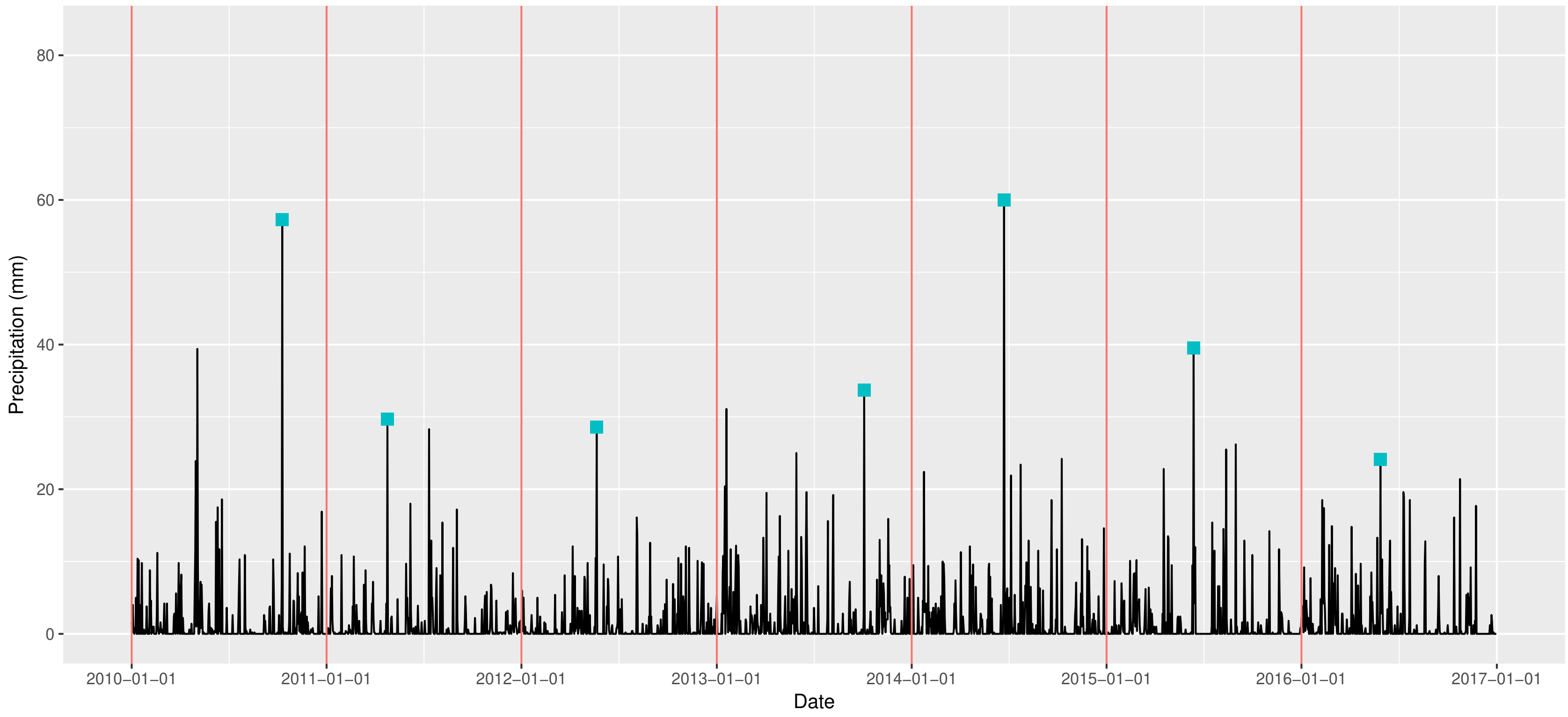


Figure 3: *Annual maxima for precipitation (mm) recorded at Toulouse–Blagnac.*

- Watch out for seasonality, starting/ending of blocks, e.g., hydrological years.

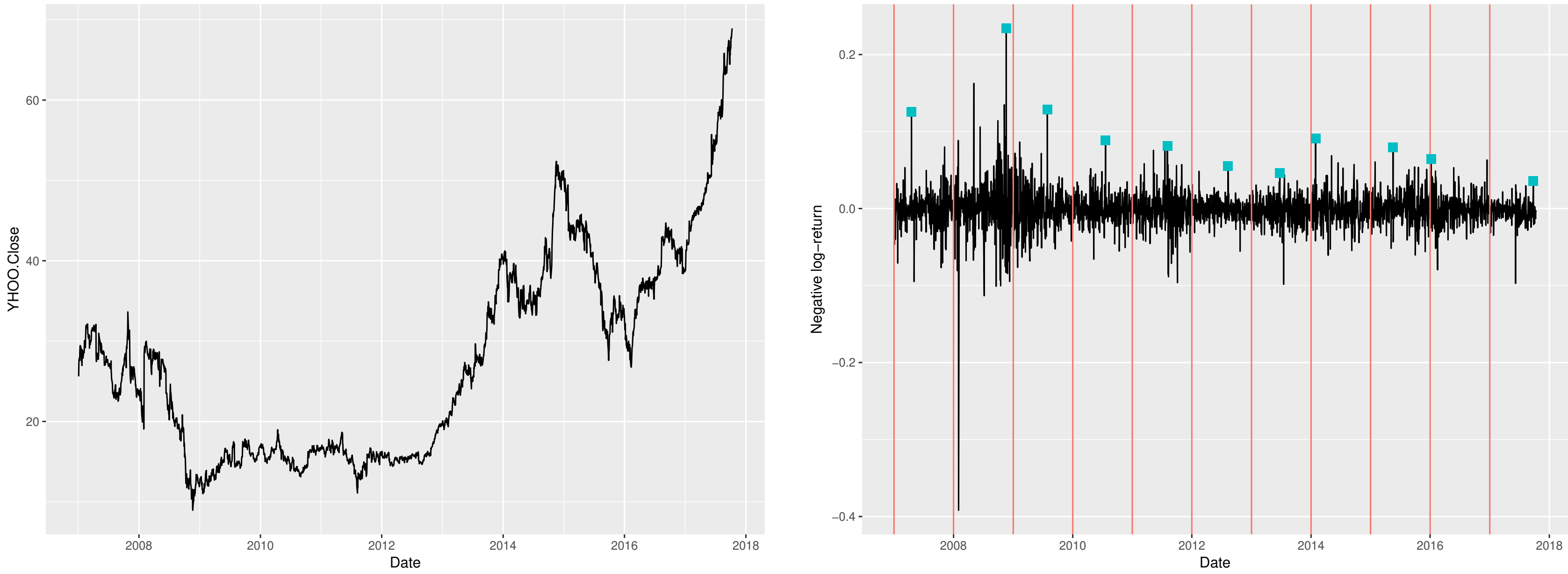


Figure 4: *Illustration about the use of (negative) log-returns, i.e.,* $Y_t = -\log(X_t/X_{t-1})$—*Yahoo closing prices.*

- It is common practice to work on the (negative) log-return to cancel out trends and mitigate the volatility.

Quantiles for the GEV

- Let $p \in (0, 1)$, the quantile y_p of a GEV(μ, σ, ξ) with exceedance probability p, i.e., $F(y_p) = 1 - p$, is

$$y_p = \mu - \sigma \frac{1 - \{-\log(1-p)\}^{-\xi}}{\xi}.$$

- In environmental application we say that y_p is the return level associated with the return period $1/p$.
- In finance we say that y_p is the Value at Risk (VaR).
- In both cases, y_p is a quantile.

Why the phrasing "return period"?

- Let $T = 1/p$ with $p \in (0,1)$.
- Let $Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} Y$ and consider the random variable

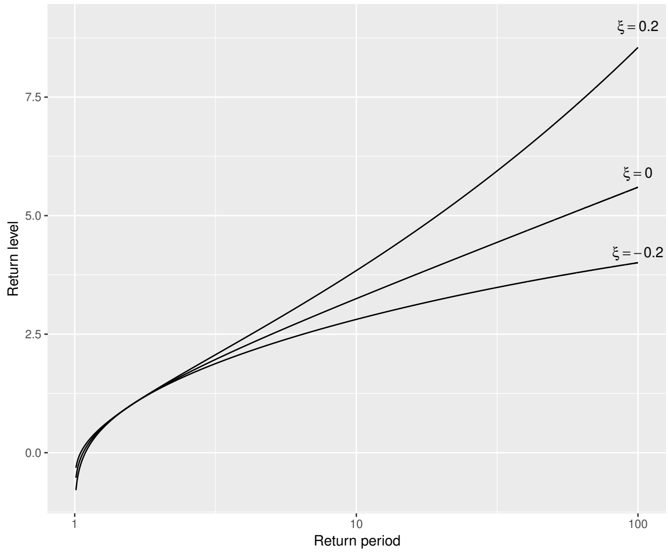
$$I = \arg\min \{i \geq 1 : Y_i \geq y_p\}, \qquad \Pr(Y \geq y_p) = p.$$

- Clearly $I \sim \text{Geom}(p)$ as Pr("success") = $\Pr(Y \geq y_p) = p$.
- Hence $\mathbb{E}(I) = 1/p = T$, that is, y_p is expected to be exceeded once every $T = 1/p$ observations.
- If the Y_i's are block maxima, it is expected to be exceeded once every $T = 1/p$ blocks, e.g., years.

Remark. Don't be fooled! It doesn't refer to any kind of periodicity…

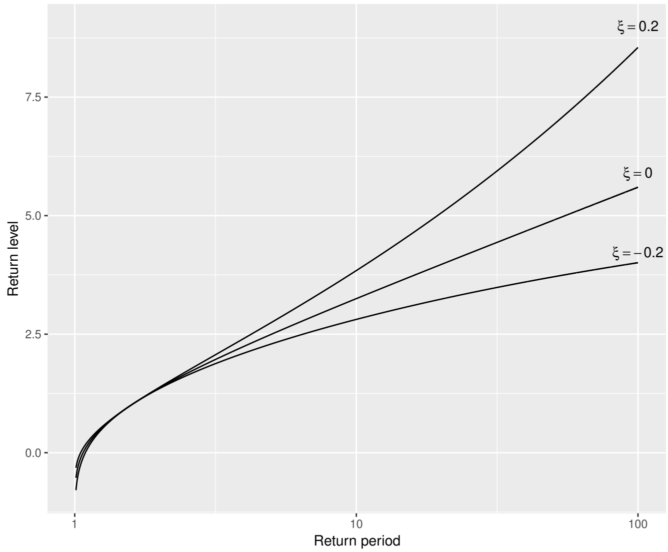
It is common practice to show results using a return level plot, i.e., plotting on a log–scale for the x–axis the function

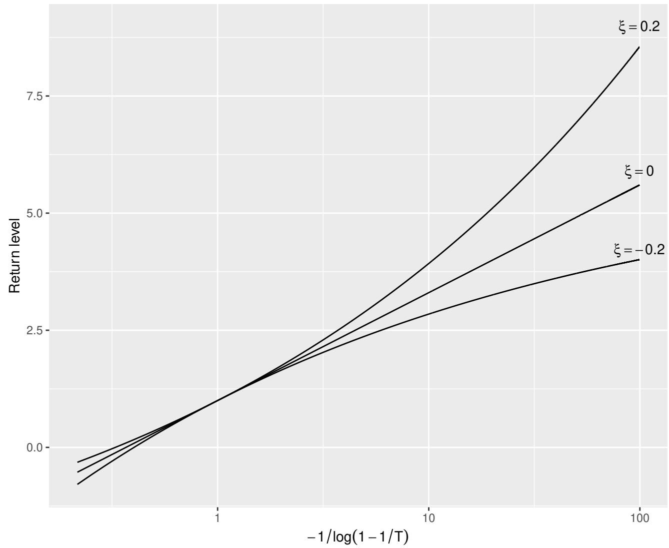
$$f : T \longmapsto y_p = \mu - \sigma \frac{1 - \{-\log(1-p)\}^{-\xi}}{\xi}, \qquad p = \frac{1}{T}.$$



It is common practice to show results using a return level plot, i.e., plotting on a log–scale for the x–axis the function

$$f : T \longmapsto y_p = \mu - \sigma \frac{1 - \{-\log(1-p)\}^{-\xi}}{\xi}, \qquad p = \frac{1}{T}.$$





Remark. Sometimes the x–axis is not the return period T but rather $-1/\log(1-1/T)$. Since $-1/\log(1-1/T) \sim T$ as $T \sim \infty$, both plots are roughly the same.

Given observed block maxima $Y_1, \ldots, Y_n$, we want to estimate the GEV parameters (μ, σ, ξ). One could use

- moment based estimators—usually not relevant as moments might not exist with extremes.
- probability weighted moments—good small sample performance but not very flexible.
- likelihood based approaches (by far the most used approach)
  - flexible and usually efficient;
  - model selection is easy (AIC, BIC, Likelihood ratio, ...)
  - can be embedded, if necessary, into a Bayesian framework.

Example: Precipitation extremes at Toulouse–Blagnac


```
> library(evd)##R package for EVT (other alternatives exist)
> head(data)##data is an R matrix giving the *raw* data
  Years Precip
1  1947    0.2
2  1947    0.2
3  1947    0.0
4  1947    0.0
5  1947    6.0
6  1947    0.4
> block.max <- aggregate(Precip~Years, FUN = max, data = data)
> (fitted <- fgev(block.max[,"Precip"]))

Call: fgev(x = block.max[, "Precip"])
Deviance: 543.5806

Estimates
     loc     scale     shape
35.12564   9.90881   0.01489

Standard Errors
    loc    scale    shape
1.33591  0.97678  0.09087
...
```

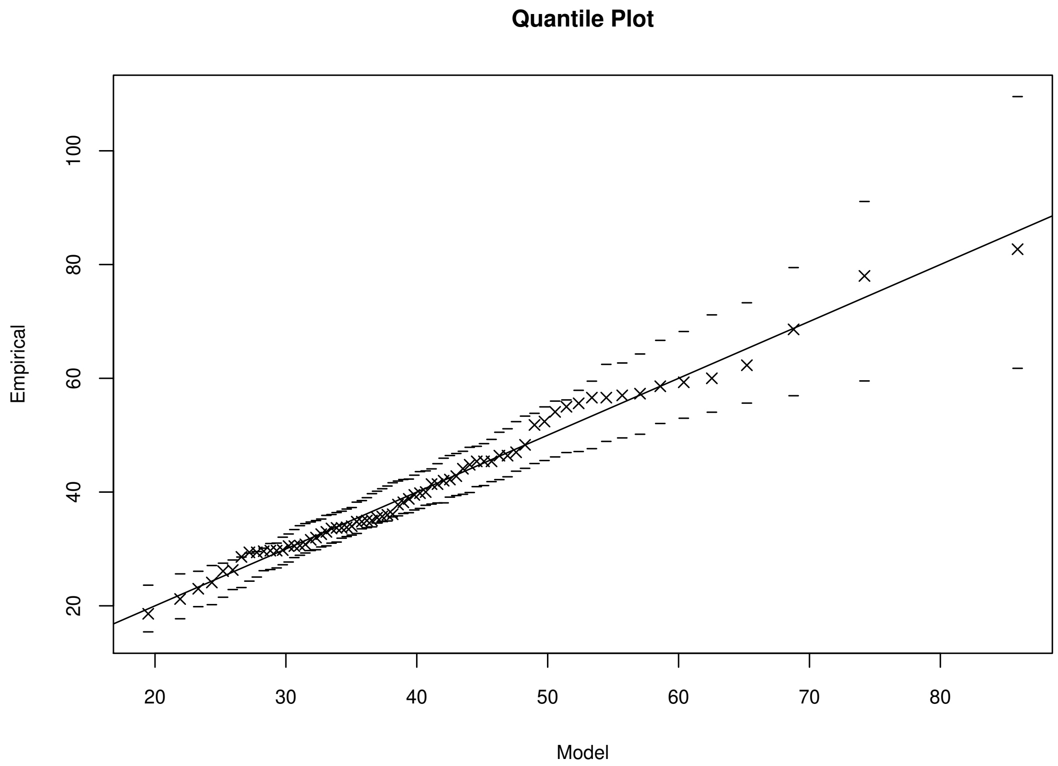
- QQ–plots are useful for model checking, identifying possible outliers
- Given a sample $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} F$ we plot the order statistics $Y_{(1)} < \cdots < Y_{(n)}$ against the plotting positions of F, e.g., the fitted GEV,

$$F^{-1}\left(\frac{1}{n+1}\right) < \cdots < F^{-1}\left(\frac{n}{n+1}\right).$$

- If F is a sensible statistical model, one should get a points lying close to a straight line of unit slope through the origin.

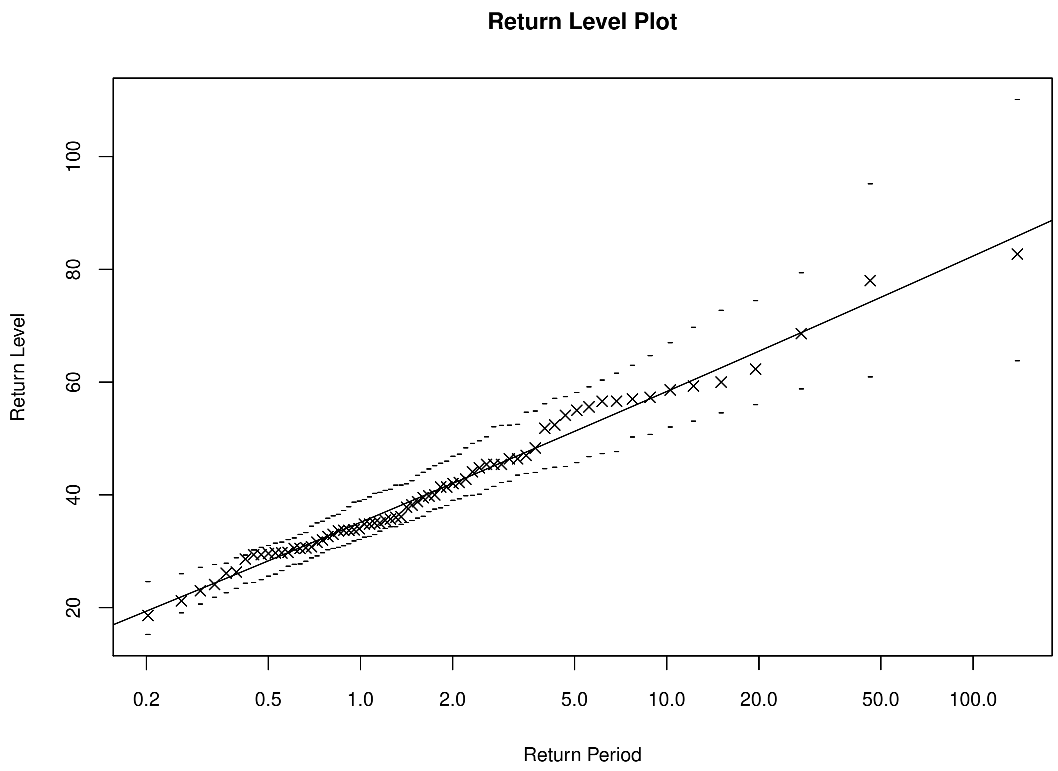
QQ–plot: Toulouse–Blagnac


```
> qq(fitted)
```



Remark. The plot shows 95% pointwise confidence intervals obtained by parametric bootsrap.

```
> rl(fitted)
```



Remark. The plot shows empirical points, i.e., $\{(n+1)/(n+1-i), Y_{(i)}\}$ and pointwise confidence intervals as before.

- The output of `fgev` gives the standard errors for $\hat{\mu}, \hat{\sigma}$ and $\hat{\xi}$ from which one easily get symmetric confidence intervals, e.g.,

$$\hat{\mu} \pm z_{1-(1-\alpha)/2} \times \text{std.err}(\hat{\mu}), \qquad \Phi(z_{1-(1-\alpha)/2}) = 1 - (1-\alpha)/2.$$

- Symmetry is not always a good thing so one could use confidence intervals based on the profile likelihood.
- As the likelihood ratio statistics satisfies

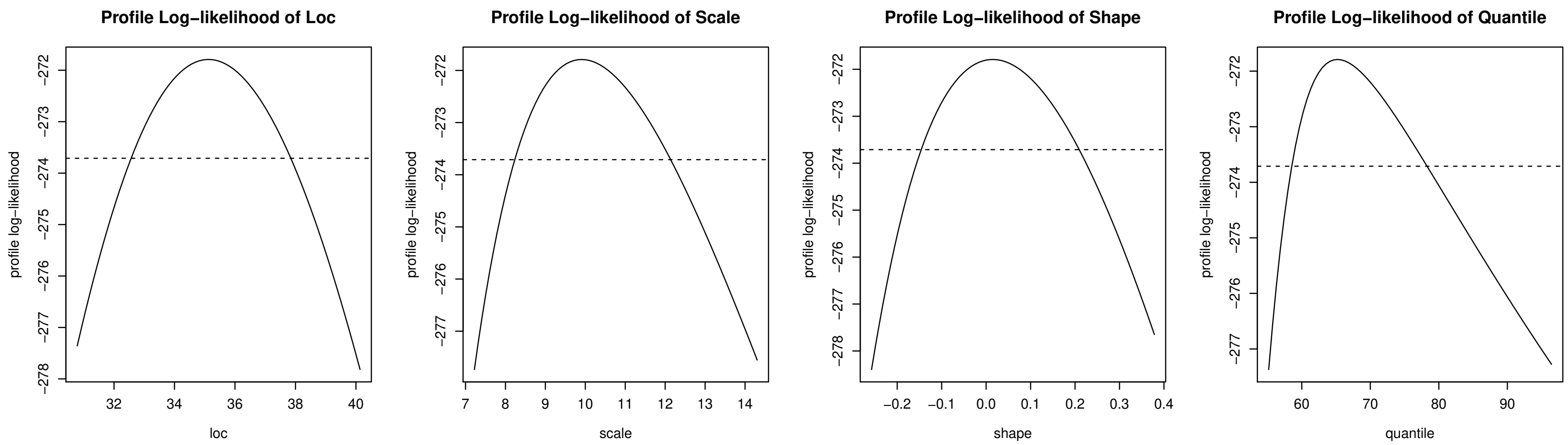
$$W(\xi_0) := 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_{\xi=\xi_0})\} \xrightarrow{D} \chi_1^2, \qquad n \to \infty,$$

we can find $I = [\xi_-, \xi_+]$ such that for all $\xi_0 \in I$

$$\Pr\left\{\chi_1^2 > W(\xi_0)\right\} > \alpha.$$

Profile likelihood: Toulouse–Blagnac


```
> plot(profile(fitted))
> plot(profile(fgev(block.max[,"precip"], prob = 0.05), "qua
```



- What does mean this (optional) argument `prob = p`?
- It is just a reparametrization of the GEV with parameters (y_p, σ, ξ), i.e., we substitute μ in the density by

$$\mu = y_p + \sigma \frac{1 - \{-\log(1-p)\}^{-\xi}}{\xi}$$

- And fit the GEV as usual.
- We can then profile the likelihood w.r.t. y_p as any other parameter !

2. Threshold exceedances

Another representation for extremes

- Previously we characterize extremes using block maxima

$$M_m = \max_{j=1,\dots,m} X_j.$$

- Another approach consists in considering threshold exceedances

$$\{X_j - u : X_j - u > 0\},$$

for some (high enough) threshold u.

Theorem. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with distribution F and sequences $\{a_m > 0 : m \geq 1\}$, $\{b_m \in \mathbb{R} : m \geq 1\}$ such that*

$$\Pr\left(\frac{M_m - b_m}{a_m} \leq x\right) \longrightarrow H(x), \qquad m \to \infty,$$

where H is a (non degenerate) GEV. Then

$$\Pr\{X > u_m(u + x) \mid X > u_m(u)\} \longrightarrow 1 - \tilde{H}(x), \qquad m \to \infty,$$

with $u_m(x) = a_m x + b_m$ for all $x \in (0, \infty)$ and

$$\tilde{H}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\tau}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\tau}\right), & \xi = 0, \end{cases}$$

where $\tau = \sigma + \xi(u - \mu)$. The limiting distribution is the Generalized Pareto Distribution GPD(τ, ξ).

- We observe a time series of, say, daily values $X_1, X_2, \ldots$ supposed (for now) to be independent and identically distributed from F.
- We choose, and not estimate, a large enough threshold u—common practice is to take $u = F^{-1}(0.95)$ but see later.
- Compute the exceedances

$$\{X_i - u \colon X_i > u\}.$$

- And fit a GPD to these exceedances.

Environmental applications

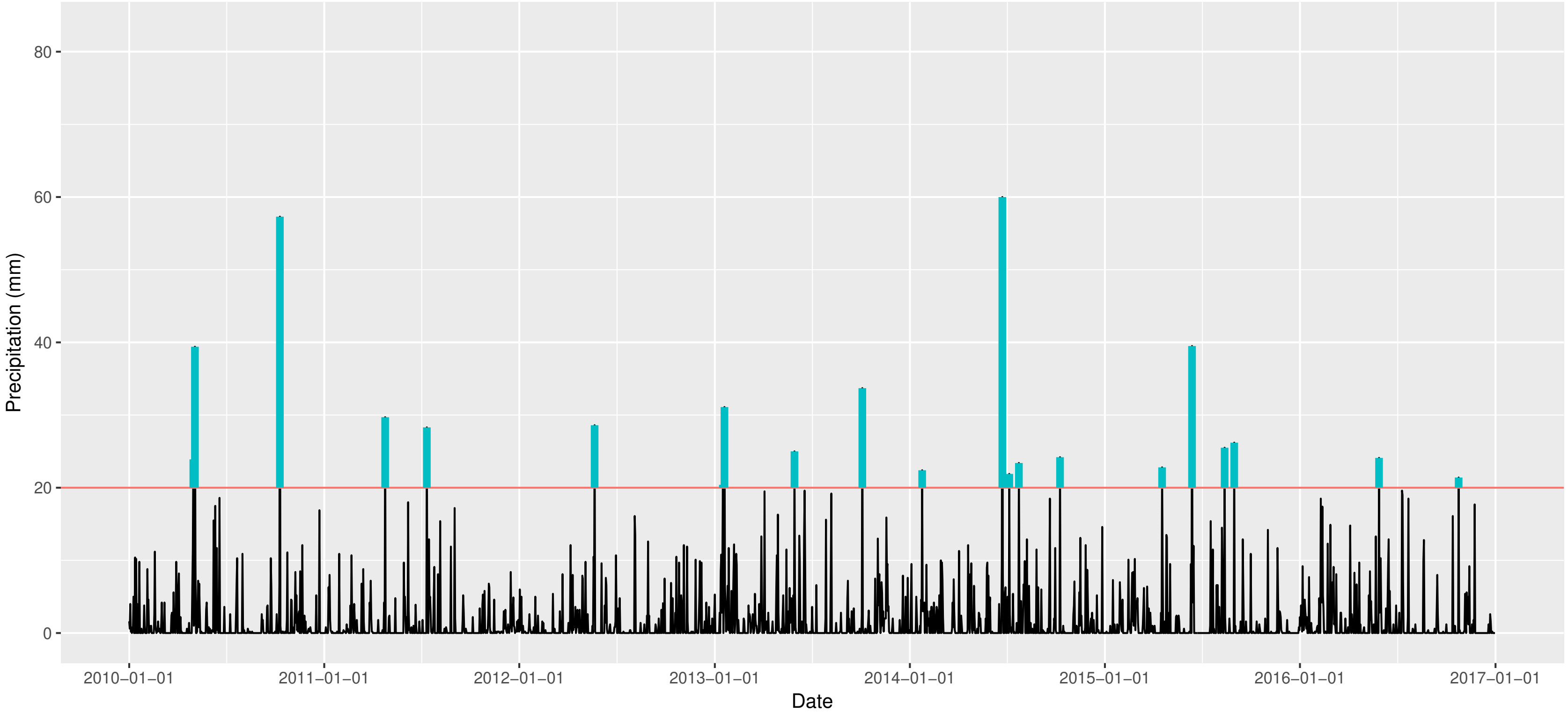


Figure 5: *Exceedances above* $u = 20$ (mm) *recorded at Toulouse–Blagnac.*

- ☐ Watch out for seasonality, temporal dependences

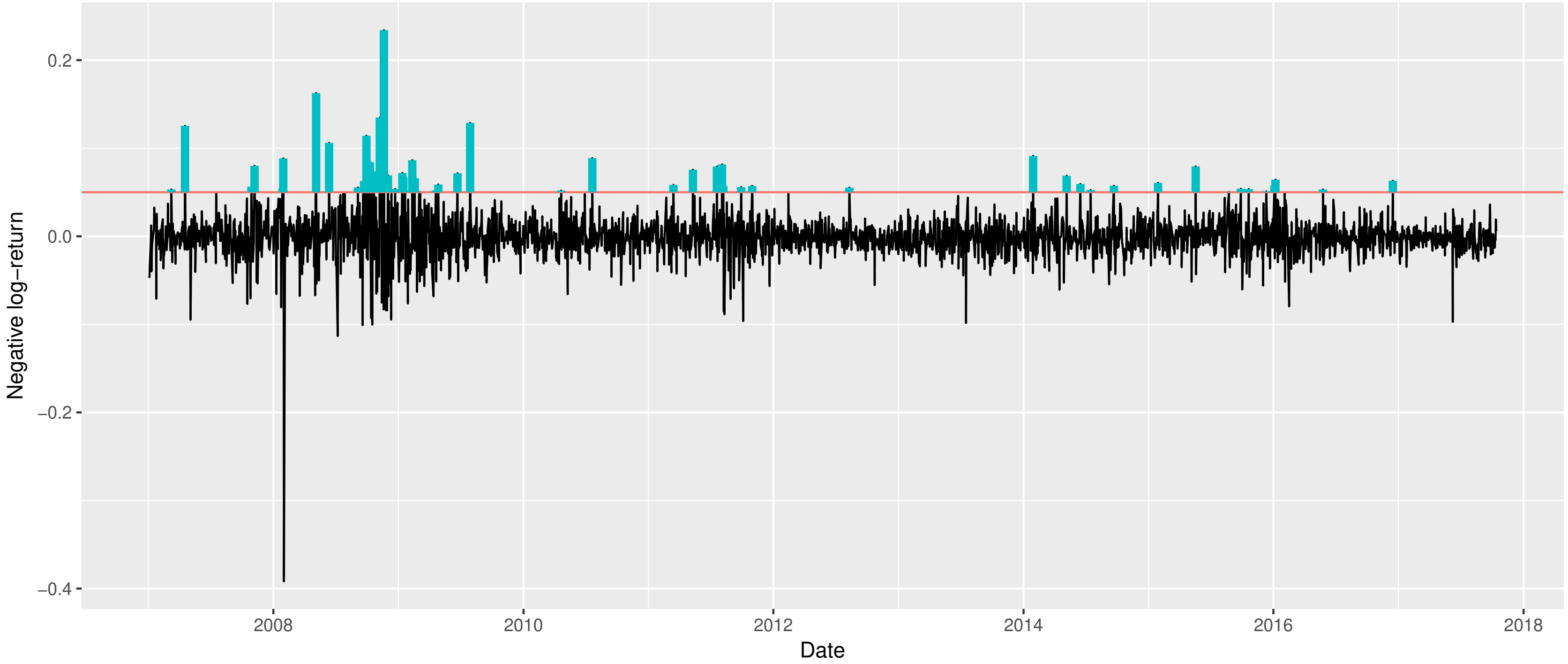


Figure 6: *(Negative) Log-returns exceedances with threshold $u = 0.1$—Yahoo closing prices.*

Quantiles for the GPD

- Let $p \in (0,1)$, the quantile y_p of a GPD(τ, ξ) with exceedance probability p, i.e., $F(y_p) = 1 - p$, is

$$y_p = \tau \frac{p^{-\xi} - 1}{\xi}.$$

- Or depending on the situtation we can write

$$y_p = u + \tau \frac{p^{-\xi} - 1}{\xi},$$

if we work on the orginal scale.

Return levels for the GPD

- Recall that the GPD is an asymptotic model for conditional exceedances.
- For some threshold u, the return level $y_p > u$ of with exceedance probability p satisfies

$$\Pr(X > y_p) = \Pr(X > y_p \mid X > u)\Pr(X > u) = p.$$

- Hence we get

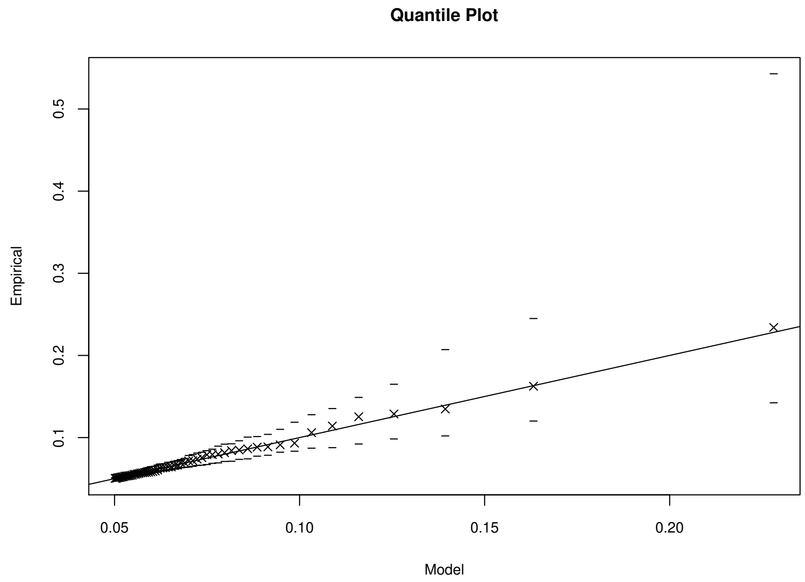
$$y_p = u + \tau \frac{(p/p(u))^{-\xi} - 1}{\xi}, \qquad p(u) = \Pr(X > u),$$

and y_p is expected to be exceeded once every $1/p$ observations.
- It is often more convenient to work on an annual scale so if we have n_y observations per year, y_p is expected to be exceeded once every $1/(pn_y)$ years.

Example: Yahoo negative log-returns


```
> library(evd)## For EVT
> library(quantmod)## To get the Yahoo data
> getSymbols("YHOO", src = "google")
> head(YHOO)## YHOO is a xts object giving the *raw* data
> nlogreturn <- -diff(log(YHOO$YHOO.Close))
> (fitted <- fpot(nlogreturn, 0.05, npp = 250))

Call: fpot(x = nlogreturn, threshold = 0.05, npp = 250)
Deviance: -320.2334

Threshold: 0.05
Number Above: 58
Proportion Above: 0.0214

Estimates
  scale    shape
0.01743  0.28939

Standard Errors
   scale     shape
0.003793  0.179451
...
```

```
> qq(fitted)
```



Return level plot: Yahoo


```
> rl(fitted)
```

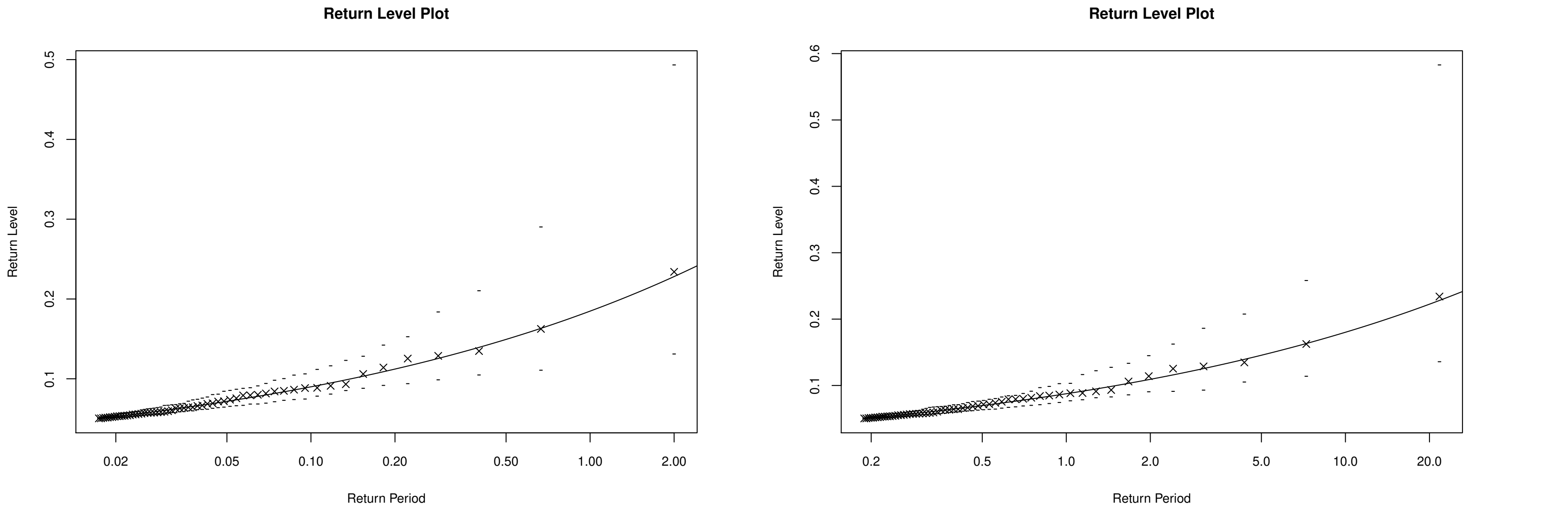
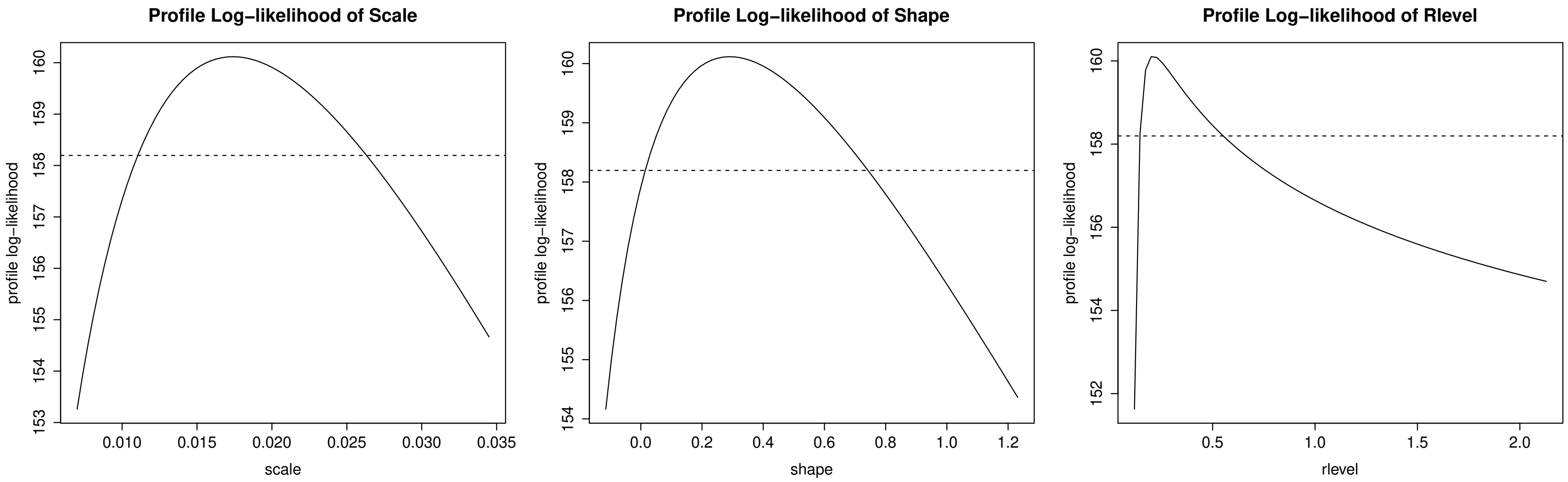


Figure 7: *Return level plot for the Yahoo data set. Left: without specifying the* `npp` *argument. Right: With* `npp = 250`*.*

Profile likelihood: Yahoo


```
> plot(profile(fitted))
> plot(profile(fpot(nlogreturn, 0.05, mper = 20, npp = 250)),
                    "rlevel")
```



Threshold selection

- Remember that threshold u is not a parameter of the GPD. We should fix it. But how?
- Intuitively one should expect a bias/variance tradeoff:
  - if u is too low: far from the asymptotic regime → bias
  - if u is too high: only few exceedances → large variance
- The basic idea is to check whether some properties of the GPD are met for a sequence of increasing thresholds $\{u_m : m \geq 1\}$.

Threshold stability of the GPD

Proposition 1. *If* $X - u_0 \mid X > u_0 \sim GPD(\tau, \xi)$ *then for all* $u \geq u_0$,

$$X - u \mid X > u \sim GPD(\tilde{\tau}, \xi), \qquad \tilde{\tau} = \tau + \xi(u - u_0).$$

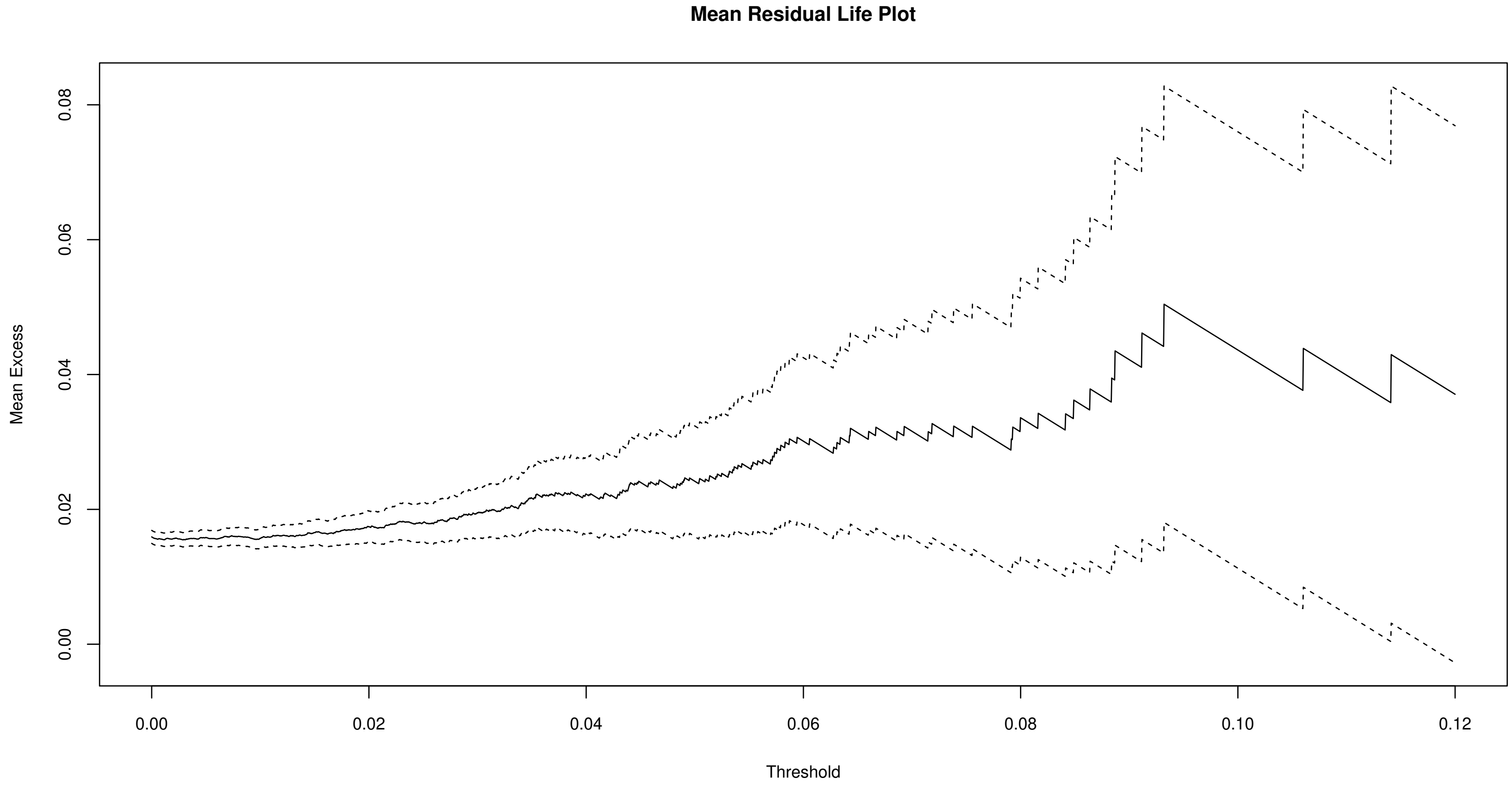
Proposition 2. *If $X - u_0 \mid X > u_0 \sim GPD(\tau, \xi)$, $\xi < 1$, then for all $u \geq u_0$*

$$MRL(u) = \mathbb{E}\,(X - u \mid X > u) = \frac{\tau(u_0) + \xi u}{1 - \xi}$$

- Hence if the GPD assumption is sensible for some threshold u_0, then the function $u \mapsto MRL(u)$ should be linear in u, $u \geq u_0$.
- We then define a sequence of increasing threshold $\{u_m : m \geq 1\}$, compute the empirical version of MRL(u_m) and check for linearity.

Mean residual life plot: Yahoo


```
> mrlplot(nlogreturn, c(0, 0.12))
```



Parameters stability

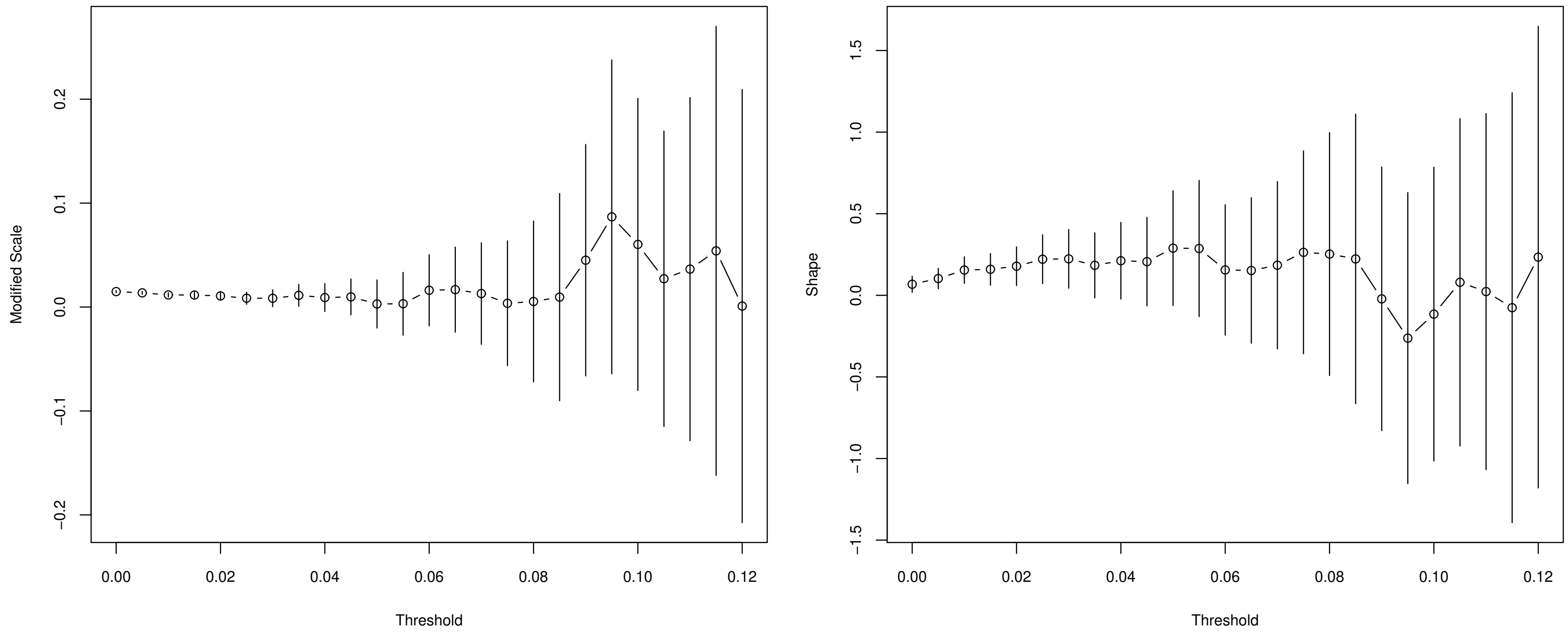
- Let $X - u_0 \mid \{X > u_0\} \sim \text{GPD}(\tau, \xi)$ then we know that for all $u \geq u_0$

$$X - u \mid \{X > u\} \sim \text{GPD}(\tilde{\tau}, \xi), \qquad \tilde{\tau} = \tau + \xi(u - u_0).$$

- Hence the function $\tau_* : u \mapsto \tilde{\tau} - \xi u$ should be constant and the shape parameter should be the same.
- It suggests to define a sequence of increasing threshold $\{u_m : m \geq 1\}$, fit a GPD to exceedances above threshold u_m, and check for stability of τ_* and ξ.

Parameters stability: Yahoo


```
> tcplot(nlogreturn, c(0, 0.12))
```



- A threshold around $u = 0.04$ seems appropriate here.

3. Point process

Point processes

Definition 5 (Informal)**.** A point process $\{X_i : i \in I\}$ is a stochastic process whose realization is a collection of points "falling" in a space $\mathscr{X}$. These points are often called atoms.

- The distribution of a point process is characterized through its counting measure

$$N(A) = \sum_{i \in I} \delta_{X_i}(A),$$

$A \subset \mathscr{X}$ Borel set and δ the Dirac function.

- Its intensity measure is defined by

$$\Lambda : A \longmapsto \mathbb{E}\{N(A)\}.$$

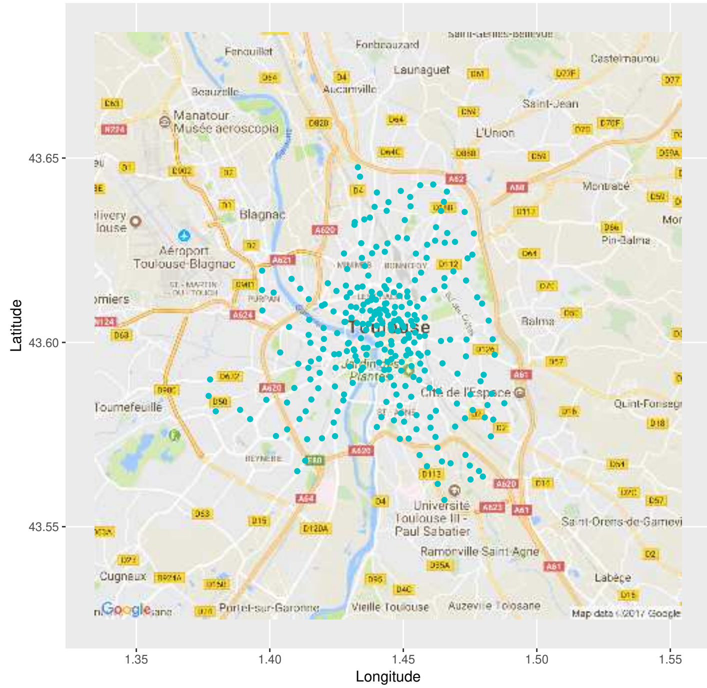


Figure 8: *Locations of the bike share program in Toulouse. Can be seen as a point process on* $\mathscr{X}$ = Toulouse.

Definition 6. A point process with intensity measure Λ is a Poisson point process if for all $k \geq 1$ and disjoint Borel sets $A, A_1, \ldots, A_k \subset \mathscr{X}$,

i) $N(A) \sim \text{Poisson}\{\Lambda(A)\}$;
ii) $N(A_1), \ldots, N(A_k)$ are independent random variables.

Remark. The intensity measure Λ is not necessarily finite. We only require it to be σ–finite, i.e., one may have $\Lambda(\mathscr{X}) = \infty$ but we can find a partition $\cup_{i \in I} A_i = \mathscr{X}$ such that $\Lambda(A_i) < \infty$, $i \in I$, where I is at most countable.

Definition 7. A Poisson point process on $\mathscr{X}$ with intensity measure Λ is regular if for all Borel set $A \subset \mathscr{X}$

$$\Lambda(A) = \int_A \lambda(s) \mathrm{d}s.$$

The function λ is non-negative and is called the intensity function.

Proposition 3. *Let $\{X_1, \ldots, X_n\}$ be a realization of a Poisson point process on $\mathscr{X}$ with intensity measure Λ. The likelihood is*

$$\exp\{-\Lambda(\mathscr{X})\} \prod_{i=1}^{n} \lambda(X_i).$$

Theorem. *Under the framework of convergence of M_m to the GEV, the sequence of point processes living in $\mathscr{X} = [0,1] \times \mathbb{R}$*

$$\{\mathscr{P}_m\}_{m\geq 1} = \left\{ \left(\frac{i}{m+1}, \frac{X_i - b_m}{a_m} \right) : i = 1, \ldots, m \right\}_{m \geq 1}$$

converges to a Poisson point process *(PPP) on $[0,1] \times C$ with intensity measure*

$$\Lambda\{[a,b] \times (z,\infty)\} = (b-a)\left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-1/\xi},$$

where $C = \{x \in \mathbb{R} : 1 + \xi(x-\mu)/\sigma > 0\}$.

Illustration convergence to a PPP

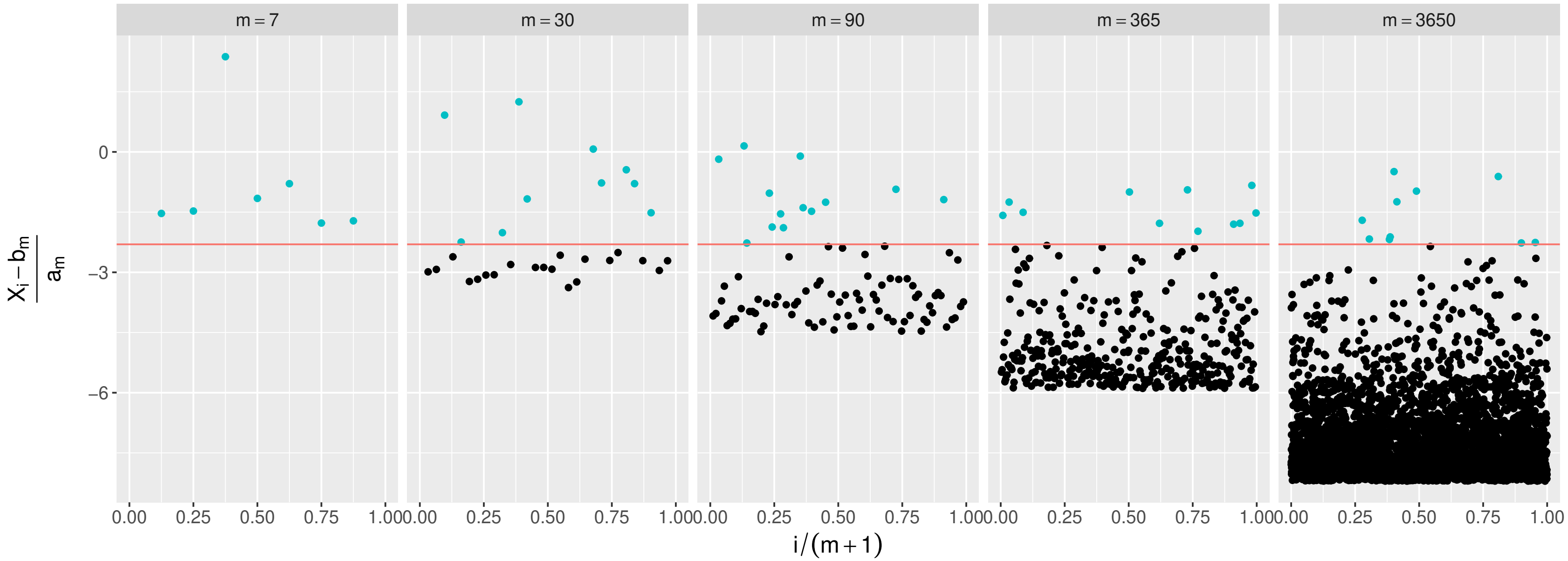


Figure 9: *Illustration of the convergence to a PPP with* $m = 7, 30, 90, 365, 3650$ *for standard Exponential random variables—*$a_m = 1$ *and* $B_m = \log m$*. The threshold is* $u = -\log 10$*.*

Statistical interpretation

- For a threshold u large enough, we fit a PPP to the exceedances with intensity measure

$$\Lambda\{(a,b)\times(x,\infty)\} = (b-a)\times\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}, \qquad x>u, \quad (a,b)\subset[0,1].$$

- In practice it is more convenient to scale the parameter to an annual scale, i.e.,

$$\Lambda\{(a,b)\times(x,\infty)\} = n_{\text{year}}(b-a)\times\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}, \qquad x>u, \quad (a,b)\subset[0,1],$$

where n_{year} is the number of years of data.

- For all $x > u$, the expected number of exceedances above x in a year is

$$\mathbb{E}\left[N\left\{(0, n_{\text{year}}^{-1}) \times (x, \infty)\right\}\right] = \Lambda\left\{(0, n_{\text{year}}^{-1}) \times (x, \infty)\right\} = \left(1 + \xi\frac{x - \mu}{\sigma}\right)^{-1/\xi}.$$

- Hence the T–year return level y_p, $p = 1/T$, satisfies

$$T\left(1 + \xi\frac{y_p - \mu}{\sigma}\right)^{-1/\xi} = 1 \Longleftrightarrow y_p = \mu + \sigma\frac{p^{-\xi} - 1}{\xi}.$$

Remark. It is a the return level derived from a GPD(σ, ξ) with threshold μ restricted to the set $\{x > u\}$.

Example: Yahoo negative log-returns


```
> (fitted <- fpot(nlogreturn, 0.05, model = "pp", npp = 250))

Call: fpot(x = nlogreturn, threshold = 0.05, model = "pp", npp = 250)
Deviance: -398.3905

Threshold: 0.05
Number Above: 58
Proportion Above: 0.0213

Estimates
    loc    scale    shape
0.08803  0.02907  0.30824

Standard Errors
     loc     scale     shape
0.007316  0.006808  0.190213

Optimization Information
  Convergence: successful
  Function Evaluations: 89
  Gradient Evaluations: 13
```

QQ–plot: Yahoo


```
> qq(fitted)
```

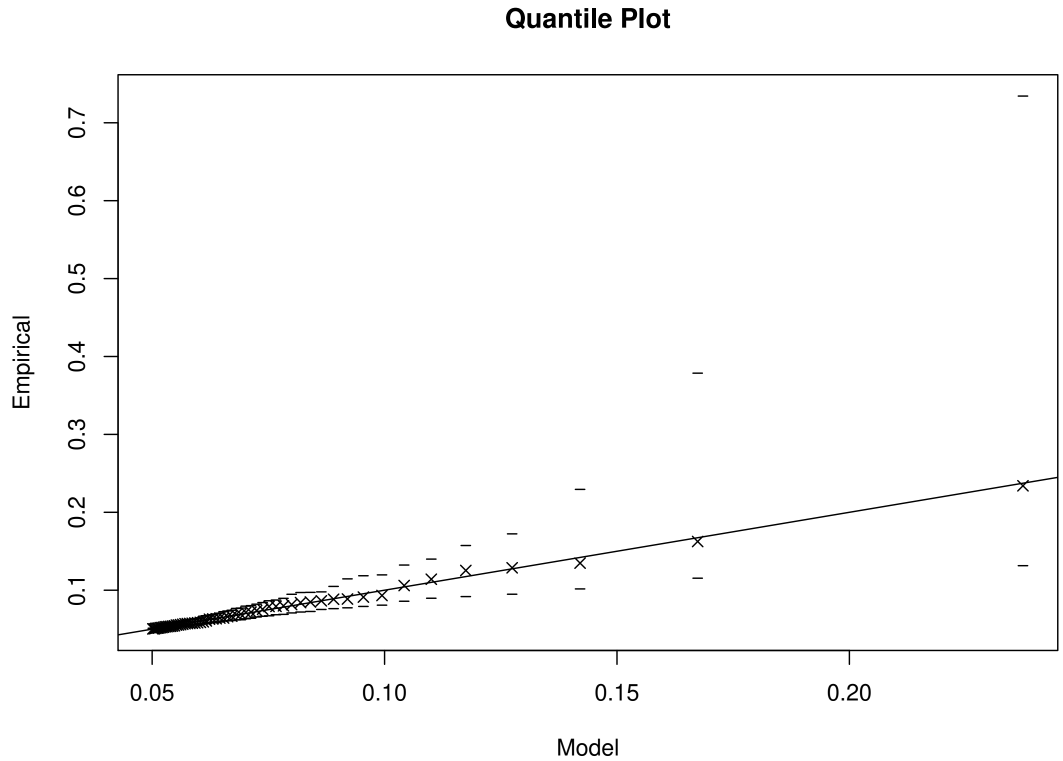


Figure 10: *QQ-plots for the Yahoo data with the PPP approach.*

```
> rl(fitted)
```

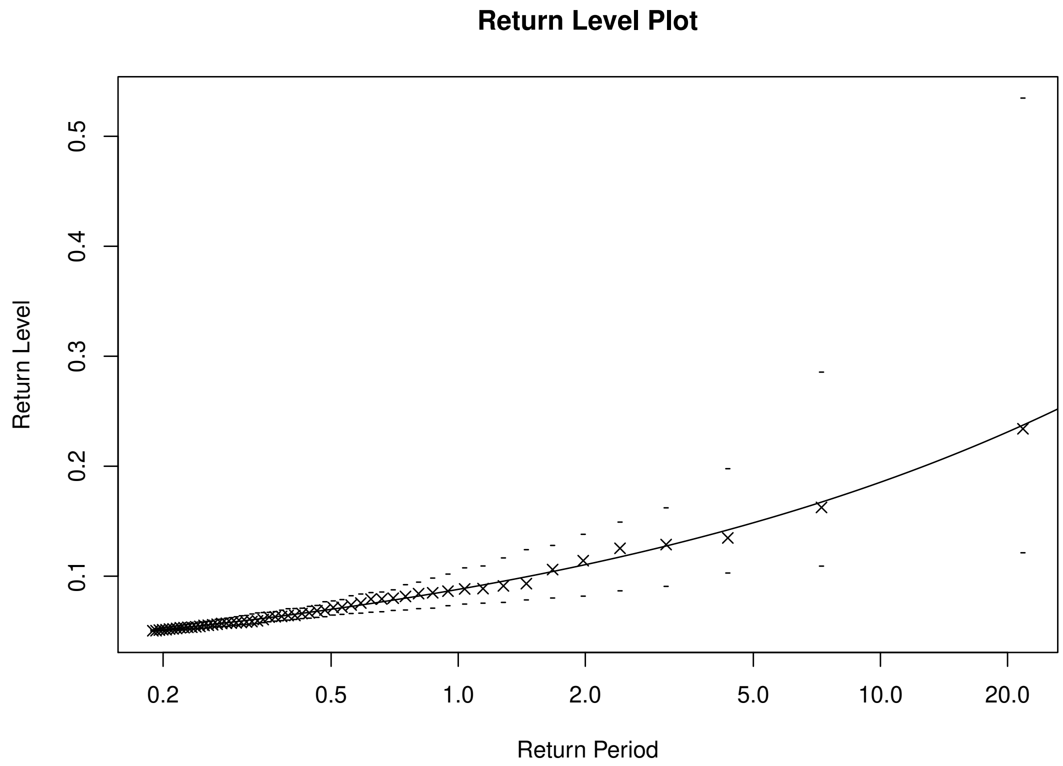


Figure 11: *Return level plot for the Yahoo data set with the PPP approach.*

Profile likelihood: PPP

- Unfortunately the `evd` package appears to be broken when we try to profile the PPP likelihood.
- Hence we will try to do it as a homework or during the lab session.

Yahoo: GEV / GPD / PPP approaches

```
> getSymbols("YHOO", src = "google")
> head(YHOO)## YHOO is a xts object giving the *raw* data
> nlogreturn <- -diff(log(YHOO$YHOO.Close))
> nlogreturn[1] <- 0

> quarter.max <- aggregate(YHOO.Close ~ quarters(index(YHOO)):years(index(YHOO)),
>                           FUN = max, data = nlogreturn)

> prob <- 1 / 10##10 years return level
> gev <- fgev(quarter.max$YHOO.Close)
> gpd <- fpot(nlogreturn, 0.05, npp = 250, mper = 1 / prob)
> ppp <- fpot(nlogreturn, 0.05, model = "pp", npp = 250)

> qgev(1 - prob/4, gev$param["loc"], gev$param["scale"], gev$param["shape"])
0.1691627
> gpd$param["rlevel"]
0.180289
> qgpd(1 - prob, ppp$param["loc"], ppp$param["scale"], ppp$param["shape"])
0.1854948
```

4. Non-stationary sequences

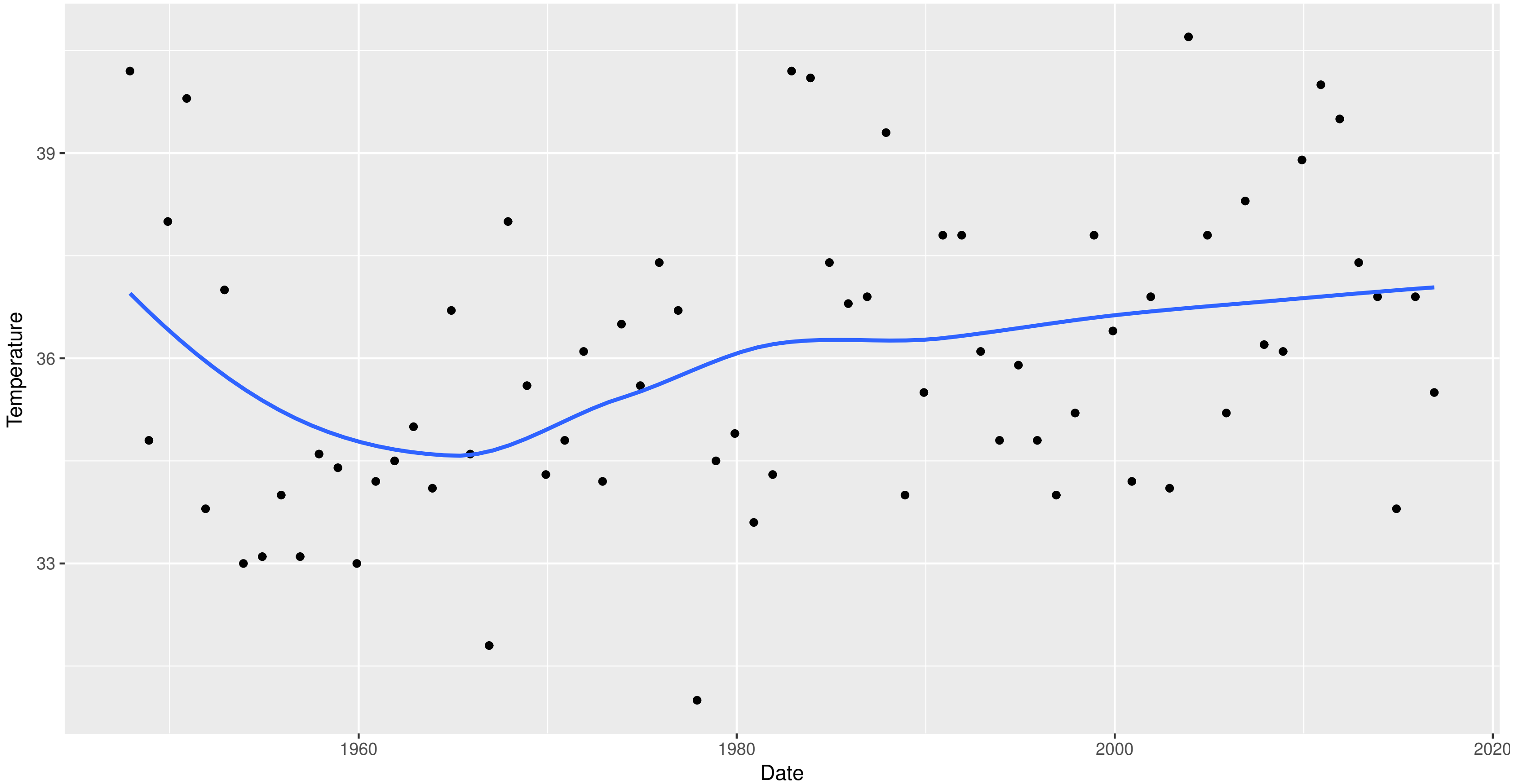
- In many situtation the i.i.d. assumption is not appropriate.
- In this section we will focus on situations in case of failure of the i.d. assumption;
- It is often the case with environmental processes which typically involve:

  1. seasonality, i.e., spring, summer, fall, winter;
  2. trends, e.g., global warming.

Two strategies

- Two (simple) strategies are possible:
  1. you restrict your analysis to specific seasons, i.e., modelling seasonal extremes;
  2. you emmbed the seasonal pattern into the parameters of the GEV/GPD/PPP.
- The first approach is straightforward as it is just a classical EVT analysis applied to a subset of our data.
- The second is (very) slightly more elaborate.

An example: Toulouse summer maxima temperatures



Assume assume for the GEV that $\mu(t) = \beta_0 + \beta_1 t/100$—increase of β_1 °C in a century.

```
> covar <- data.frame(year = scale(1:nrow(summer.max), scale = FALSE)) /
          100
> (fit <- fgev(summer.max$Temperature, nsloc = covar))

Call: fgev(x = summer.max$Temperature, nsloc = covar)
Deviance: 295.6636

Estimates
    loc  locyear    scale    shape
35.1082   3.3582   1.8451  -0.1394

Standard Errors
    loc  locyear    scale    shape
0.24441  1.16260  0.16944  0.07912

Optimization Information
  Convergence: successful
  Function Evaluations: 20
  Gradient Evaluations: 10
```

Model selection

- Is this trend really necessary, i.e.,

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0?$$

- How would you do this?

Do you understand those lines?


```
> z <- abs(fit$par["locyear"] / fit$std.err["locyear"])
> 2 * pnorm(z, lower.tail=FALSE)
   locyear
0.003870264
```

☞ Conclusion?

Do you understand those lines?


```
> z <- abs(fit$par["locyear"] / fit$std.err["locyear"])
> 2 * pnorm(z, lower.tail=FALSE)
    locyear
0.003870264
```

☞ Conclusion?

```
> fit0 <- fgev(summer.max$Temperature)
> W <- 2 * (logLik(fit) - logLik(fit0))
> pchisq(W, df = 1, lower.tail=FALSE)
'log Lik.' 0.005746295 (df=4)
```

☞ Conclusion?

- From the asymptotic normality of the MLE we know that

$$\sqrt{n}(\hat{\beta}_1 - \beta_{1,*}) \xrightarrow{\text{d}} N(0, \sigma^2), \qquad n \to \infty.$$

- Hence under H_0, i.e., $\beta_{1,*} = 0$, we have

$$\sqrt{n}\frac{\hat{\beta}_1}{\sigma} \xrightarrow{\text{d}} N(0,1), \qquad n \to \infty.$$

This is known as the Wald test.

Theoretical versions (2)

- Using Taylor expansion of the log-likelihood $\ell(\theta_*)$ around $\hat{\theta}$ we have

$$\begin{aligned}
\ell(\theta_*) &\dot{\sim} \ell(\hat{\theta}) + (\theta_* - \hat{\theta})^\top \nabla \ell(\hat{\theta}) + \frac{1}{2}(\theta_* - \hat{\theta})^\top \nabla^2 \ell(\hat{\theta}(\theta_* - \hat{\theta})^\top \\
&\dot{\sim} \ell(\hat{\theta}) + \frac{1}{2}(\theta_* - \hat{\theta})^\top \nabla^2 \ell(\hat{\theta})(\theta_* - \hat{\theta})
\end{aligned}$$

- Hence we conclude that as $n \to \infty$

$$\begin{aligned}
W = 2\{\ell(\hat{\theta}) - \ell(\theta_*)\} &= -\sqrt{n}(\theta_* - \hat{\theta})^\top \frac{1}{n} \nabla^2 \ell(\hat{\theta}) \sqrt{n}(\theta_* - \hat{\theta}) \\
&\xrightarrow{\text{d}} \chi^2_p, \qquad p = |\theta_*|.
\end{aligned}$$

- This is known as the likelihood ratio test.

5. Stationary sequences

- Sor far we analyzed the asymptotic behaviour of i.i.d. random variable.
- In many situations, e.g., Yahoo time series, this assumption is unrealistic !
- What happens there is some serial dependance ?

Definition 8. A stationary sequence $\{X_i : i \geq 1\}$ is said to satisfy the $D(u_n)$ condition, if for all $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > \ell_n$, we have

$$|\Pr(X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n) -$$
$$\Pr(X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n)\Pr(X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n)| \leq \alpha(n, \ell),$$

where $\alpha(n, \ell_n) \to 0$ for some sequences $\ell_n = o(n)$ as $n \to \infty$.

- Roughly speaking the $D(u_n)$ condition imposes that the two blocks X_i's and X_j's are close to being independent as long as they are sufficiently "far apart".
- One way to avoid long–range dependence.

Theorem. *Let $X_1, X_2, \ldots$ be a stationary sequence and define $M_n = \max(X_1, \ldots, X_n)$. If there exists 2 sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G(z), \qquad n \to \infty,$$

where G is a non degenerate distribution and the $D(u_n)$ condition is met with $u_n = a_n z + b_n$ for all $z \in \mathbb{R}$ such that $G(z) > 0$, then necessarily G is of the GEV form.

Theorem. *Let $X_1, X_2, \ldots$ be a stationary sequence and define $M_n = \max(X_1, \ldots, X_n)$. If there exists 2 sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G(z), \qquad n \to \infty,$$

where G is a non degenerate distribution and the $D(u_n)$ condition is met with $u_n = a_n z + b_n$ for all $z \in \mathbb{R}$ such that $G(z) > 0$, then necessarily G is of the GEV form.

Remark. The GEV parameters for the stationary sequence will not be the same as the ones for an i.i.d. sequence.

Theorem. *Let $X_1, X_2, \ldots$ be a stationary sequence and $X_1^*, X_2^*, \ldots$ an i.i.d. sequence with the same marginal distribution as the X_i's. Define $M_n = \max(X_1, \ldots, X_n)$ and $M_n^* = \max(X_1^*, \ldots, X_n^*)$. Under the hypothesis of the previous theorem, we have*

$$\Pr\left(\frac{M_n^* - b_n}{a_n} \leq z\right) \longrightarrow G_*(z), \qquad n \to \infty,$$

if and only if

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \longrightarrow G(z), \qquad n \to \infty,$$

where

$$G(z) = G_*^{\theta}(z), \qquad 0 < \theta \leq 1.$$

☞ θ is called the extremal index.

Impact of the extremal index

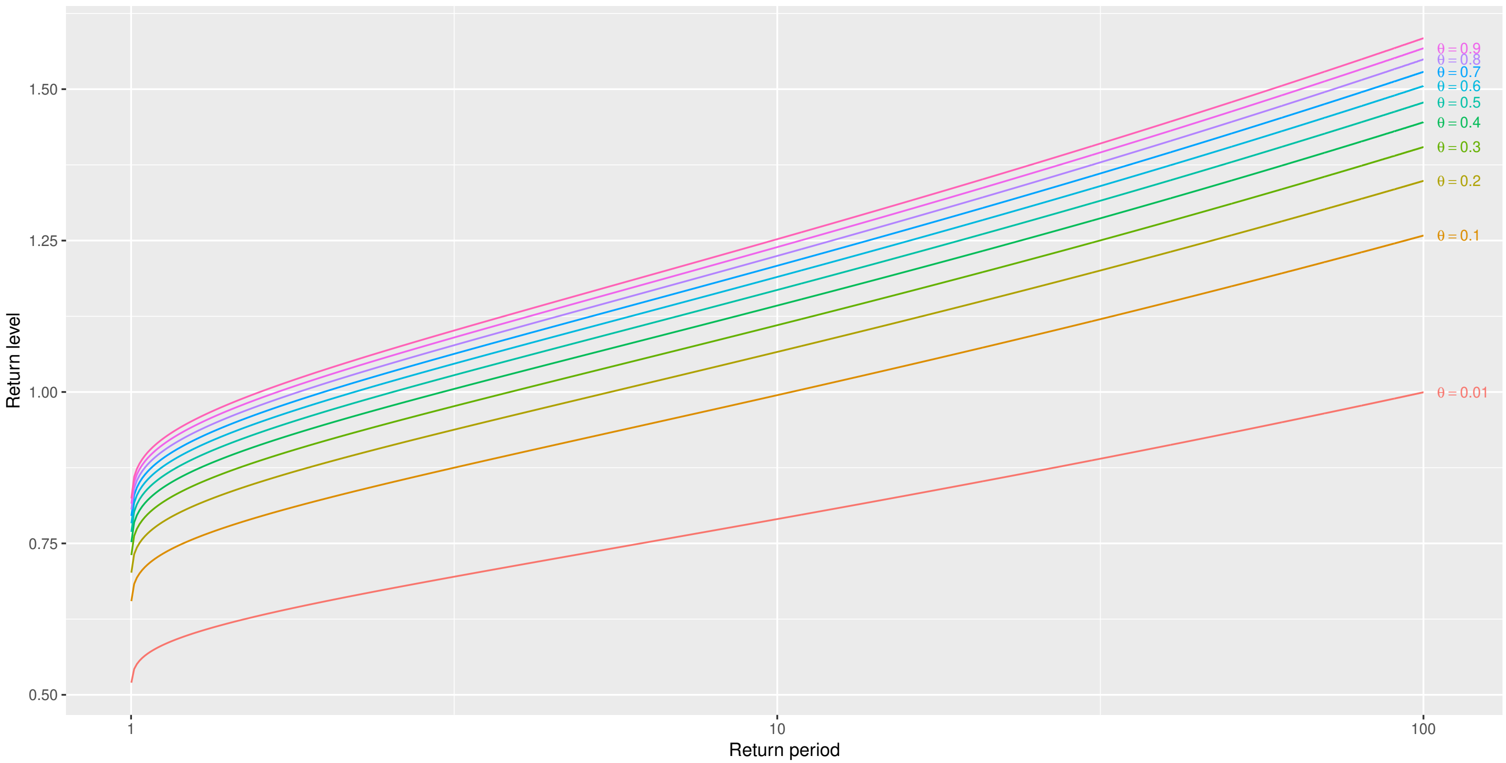


Figure 12: *Impact of the extremal index θ on return levels.*

Impact of the extremal index

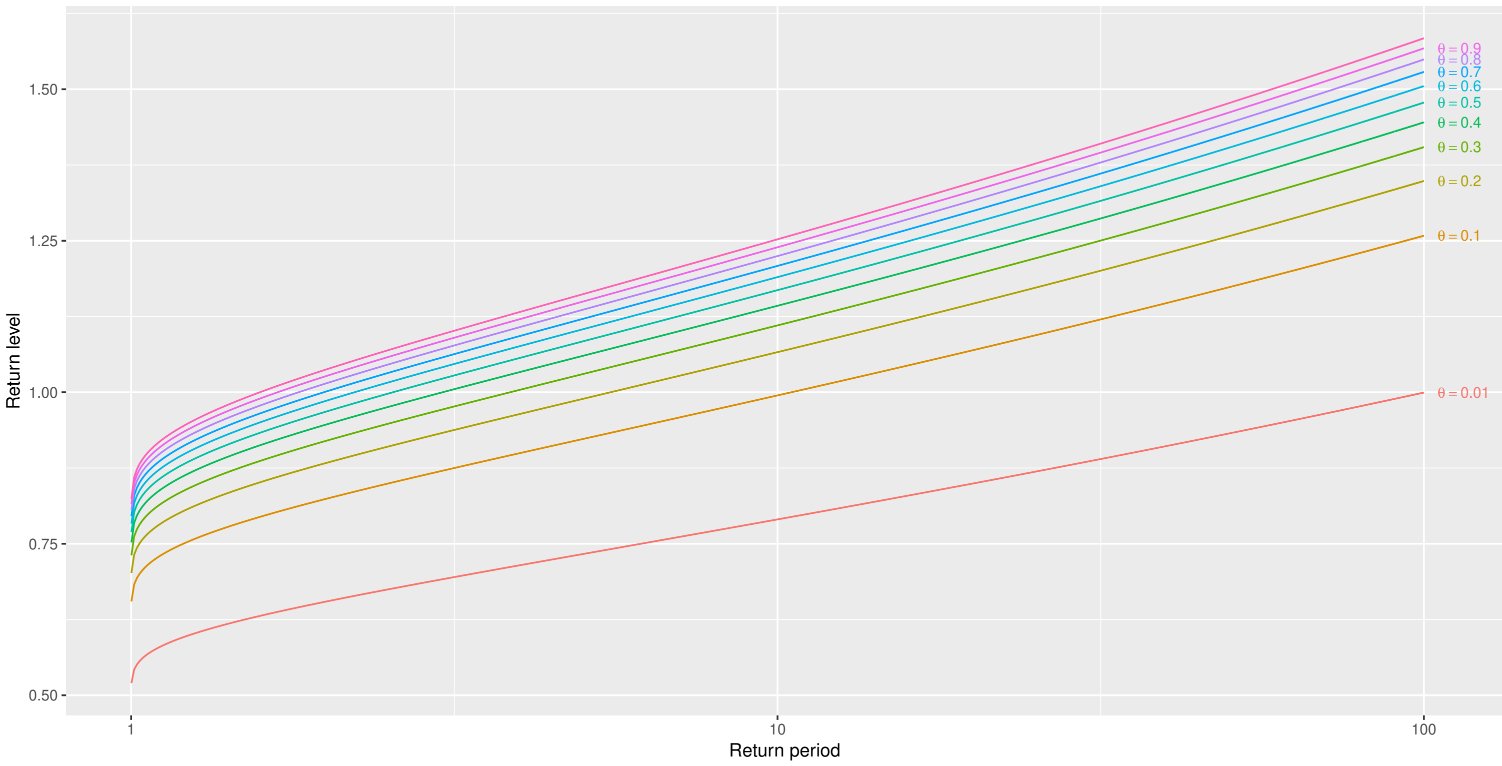


Figure 12: *Impact of the extremal index θ on return levels.*

- As expected, M_n is stochastically smaller than M_n^*—as maxima taken over dependent r.v. is likely to be smaller than for independent r.v.

Connection between G and G_*^θ

- Note that we have

$$G_*^\theta(z) = \exp\left\{-\theta\left(1+\xi\frac{z-\mu}{\sigma}\right)^{-1/\xi}\right\}$$
$$= \exp\left\{-\left(1+\xi\frac{z-\mu_*}{\sigma_*}\right)^{-1/\xi}\right\},$$

where $\mu_* = \mu - \frac{\sigma}{\xi}\left(1-\theta^\xi\right)$, $\sigma_* = \sigma\theta^\xi$.

☞ No change for block maxima of stationary sequence, since you will estimate μ_*, σ_* and ξ directly.

- ☐ Wait!!! There is dependence so the likelihood is not

$$L(\mu, \sigma, \xi; m_1, \ldots, m_{\tilde{n}}) = \prod_{i=1}^{\tilde{n}} f_{GEV}(m_i; \mu, \sigma; \xi) \ldots$$

☞ No change for block maxima of stationary sequence, since you will estimate μ_*, σ_* and ξ directly.

- ☐ Wait!!! There is dependence so the likelihood is not

$$L(\mu,\sigma,\xi;m_1,\ldots,m_{\tilde{n}}) = \prod_{i=1}^{\tilde{n}} f_{GEV}(m_i;\mu,\sigma;\xi)\ldots$$

- ☐ Yes but no! If beginnings / endings of blocks are well defined[1], assumption of mutual independence between block maxima makes sense ⇒ Likelihood still valid.

☞ This will be however a bit different for the GPD // PPP approaches.

[1] Why should it always be “calendar year” blocks?

- The extremal index θ has two alternative definitions:
  - As the reciprocal of the limiting expected cluster size

$$\theta^{-1} = \lim_{n\to\infty} \mathbb{E}\left(\sum_{i=1}^{p_n} 1_{\{X_i > u_n\}} \mid M_{p_n} > u_n \right),$$

for sequences such that $n\{1 - F(u_n)\} \to \lambda \in (0, \infty)$ and $p_n = o(n)$.

  - As the limiting probability that an exceedance over u_n is the last one

$$\theta = \lim_{n\to\infty} \Pr\left\{ \max(X_2, \ldots, X_{p_n}) \le u_n \mid X_1 \ge u_n \right\}.$$

Watch out!

- Consider the 10–years return event $A = \{X > z_{10}\}$.
- This event is expected to occur 10 times in a century.

Watch out!

- Consider the 10–years return event $A = \{X > z_{10}\}$.
- This event is expected to occur 10 times in a century.
- However we have

$$\Pr(A \text{ not seen in the next 10 years}) = \begin{cases} \left(1 - \frac{1}{10}\right)^{10} \approx 0.35, & \theta = 1 \\ \left(1 - \frac{1}{10}\right)^{10\theta} = 0.90, & \theta = 0.1. \end{cases}$$

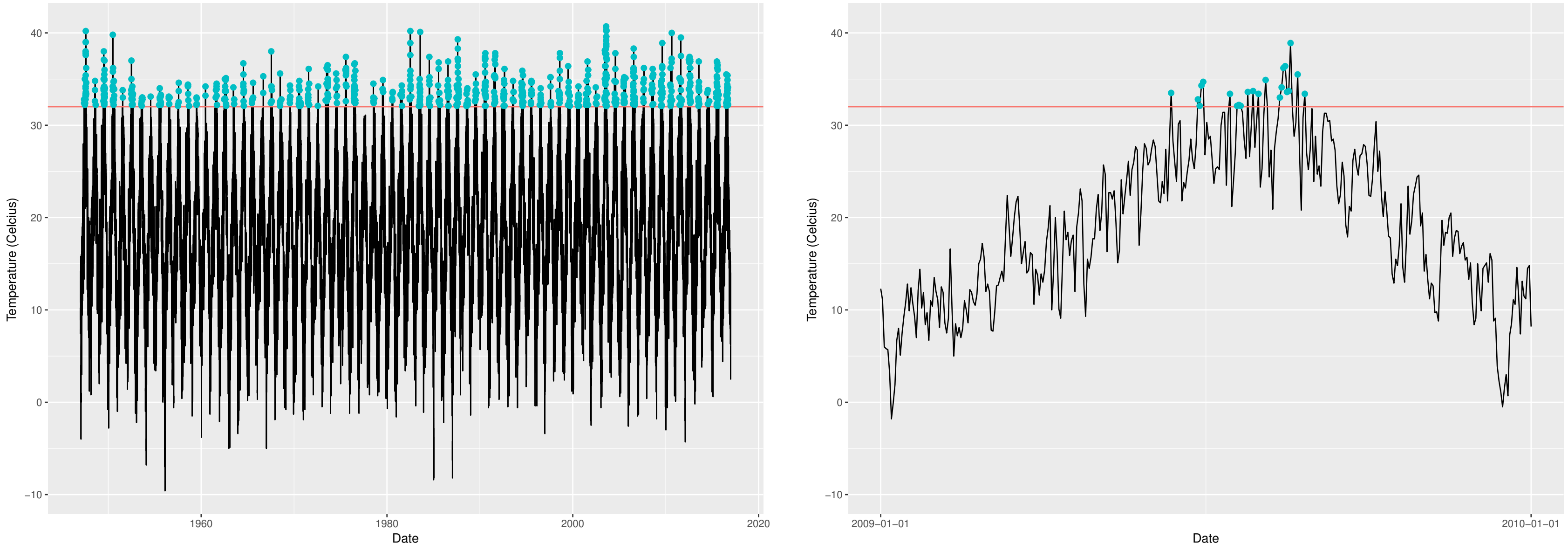
Watch out!

- Consider the 10–years return event $A = \{X > z_{10}\}$.
- This event is expected to occur 10 times in a century.
- However we have

$$\Pr(A \text{ not seen in the next 10 years}) = \begin{cases} \left(1 - \frac{1}{10}\right)^{10} \approx 0.35, & \theta = 1 \\ \left(1 - \frac{1}{10}\right)^{10\theta} = 0.90, & \theta = 0.1. \end{cases}$$

☞ When $\theta = 0.1$, these "expected 10 extremes" will tend to occur simultaneously leading to a higher probability of seeing none of them within the next 10 years.

Exceedances for stationary sequences



- Two approaches are possible:
  1. either you discard some observations to be closer to the i.i.d. assumption;
  2. or you take into account for such a serial dependence, e.g., assume a Markovian structure… Not discussed here!

- Because we usually use the MLE to fit the GPD // PPP, we have to use cluster maxima only for inference.
- Still the expected annual number of exceedances above z is

$$\Lambda\{(0, n_{\text{year}}^{-1}) \times (z, \infty)\} = \theta \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-1/\xi},$$

and the T–year return level y_p, $p = 1/T$, satisfies

$$T\theta \left(1 + \hat{\xi} \frac{y_p - \mu}{\sigma}\right)^{1/\xi} = 1 \iff y_p = \mu + \frac{\sigma}{\xi}\left\{(\theta T)^{\xi} - 1\right\},$$

where the extremal index θ can be estimated separately by

$$\hat{\theta} = \frac{n_c}{n_u}, \qquad n_c = \#\ \text{clusters}, \quad n_u = \#\ \text{exceedances above } u.$$

- Because we usually use the MLE to fit the GPD // PPP, we have to use cluster maxima only for inference.
- Still the expected annual number of exceedances above z is

$$\Lambda\{(0, n_{\text{year}}^{-1}) \times (z, \infty)\} = \theta \left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi},$$

and the T–year return level y_p, $p = 1/T$, satisfies

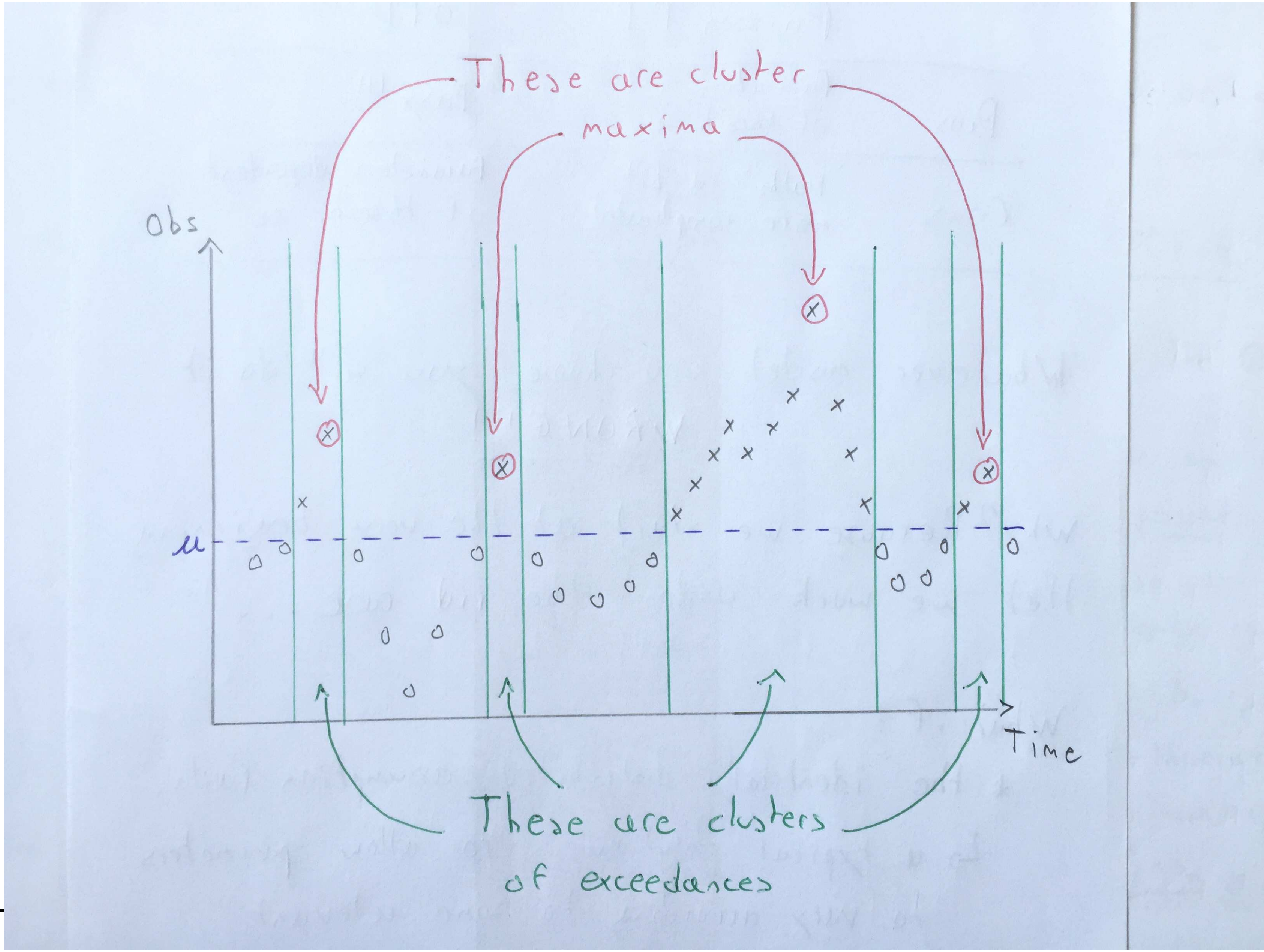
$$T\theta \left(1 + \hat{\xi} \frac{y_p - \mu}{\sigma}\right)^{1/\xi} = 1 \iff y_p = \mu + \frac{\sigma}{\xi} \left\{ (\theta T)^{\xi} - 1 \right\},$$

where the extremal index θ can be estimated separately by

$$\hat{\theta} = \frac{n_c}{n_u}, \qquad n_c = \#\text{ clusters}, \quad n_u = \#\text{ exceedances above } u.$$

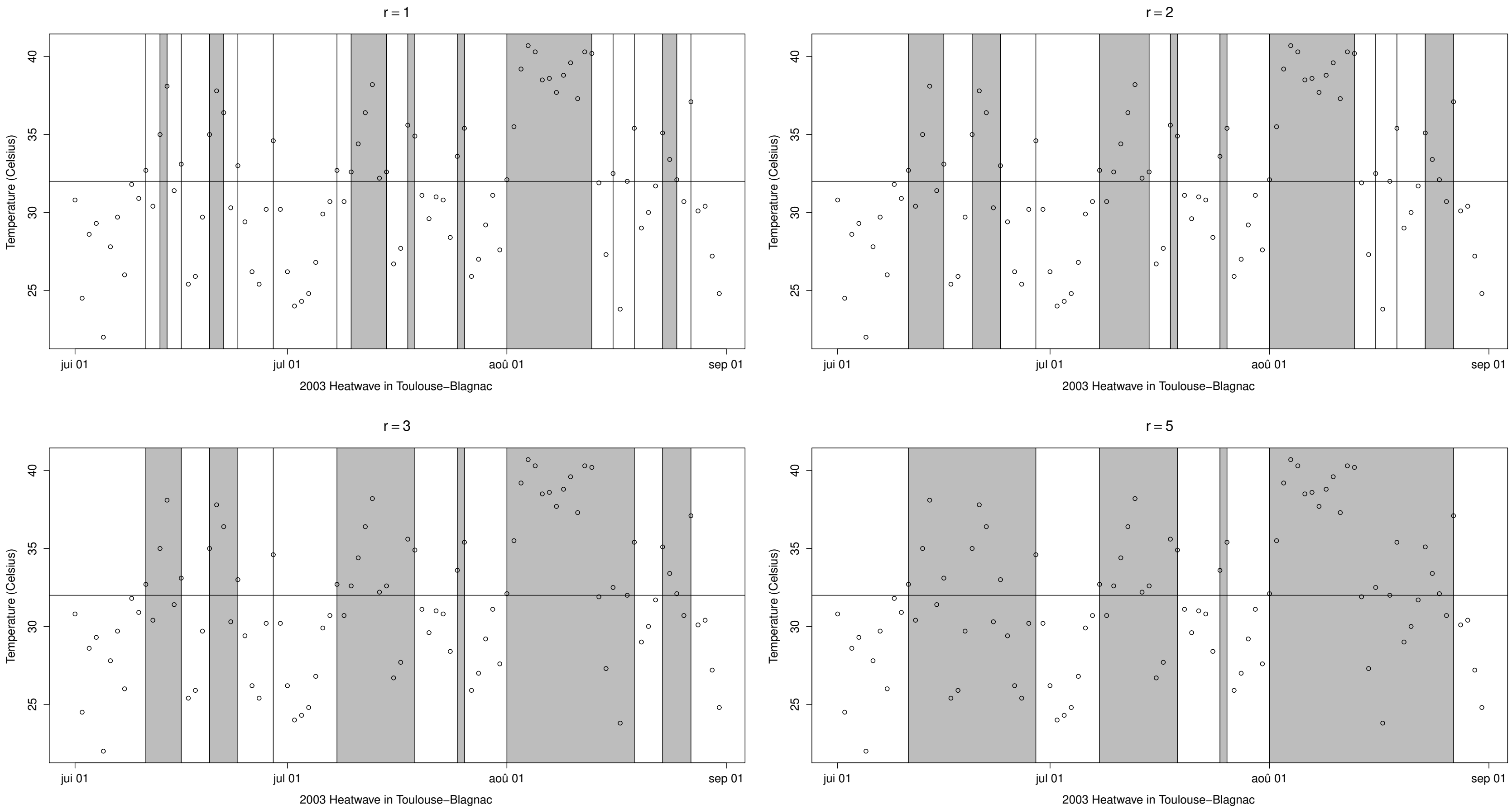
☞ We need to define cluster of exceedances!

Declustering: runs method



Declustering at Toulouse–Blagnac


```
> clusters(data, thresh, r, plot = TRUE)
```



```
> fpot(df$Temperature, 32, "pp", npp = 365.25, cmax = TRUE, r = 3)

Call: fpot(x = df$Temperature, threshold = 32, model = "pp", npp = 365.25,
           cmax = TRUE, r = 3)
Deviance: 807.4698

Threshold: 32
Number Above: 779
Proportion Above: 0.0305

Clustering Interval: 3
Number of Clusters: 330
Extremal Index: 0.4236

Estimates
    loc    scale    shape
35.5823   1.8864  -0.2533

Standard Errors
    loc    scale    shape
0.19606  0.08125  0.04456
```

- ☐ Here the extremal index estimate is $\hat{\theta} \approx 0.42$, i.e., clusters tends to be of size 2.5.
- ☞ It is typical, I think, for temperatures data but might be very different for, say, rainfall where $\theta \approx 1$.

	GEV	PPP	PPP ($r = 1$)	PPP ($r = 3$)	PPP ($r = 10$)
μ	35.2 (0.3)	36.6 (0.1)	35.8 (0.2)	35.6 (0.2)	34.6 (0.3)
σ	1.9 (0.2)	1.51 (0.07)	1.72 (0.07)	1.88 (0.08)	2.65 (0.16)
ξ	-0.16 (0.09)	-0.19 (0.03)	-0.23 (0.04)	-0.25 (0.04)	-0.35 (0.07)
y_{100}	41.3 (0.94)	41.1 (1.2)	40.3 (1.1)	40.1 (1.2)	39.6 (2.8)
θ	—	—	0.54	0.42	0.21