DATA ANALYSIS
Year 2019–2020

#5 Extreme Value Analysis

Master M1–Marine Physics
Université de Bretagne Occidentale

Jonathan GULA
[gula@univ-brest.fr]

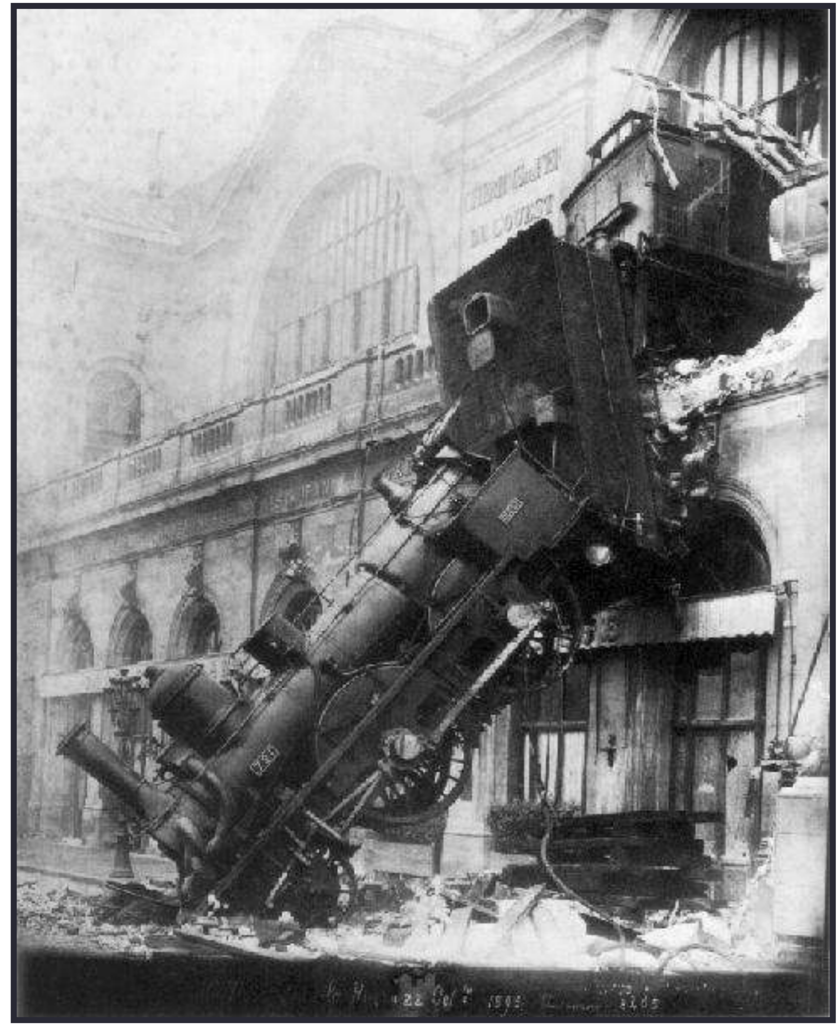
Extreme Events

“Man can believe the impossible. But man can never believe the improbable.”

- Oscar Wilde

“It's impossible that the improbable will never happen.”

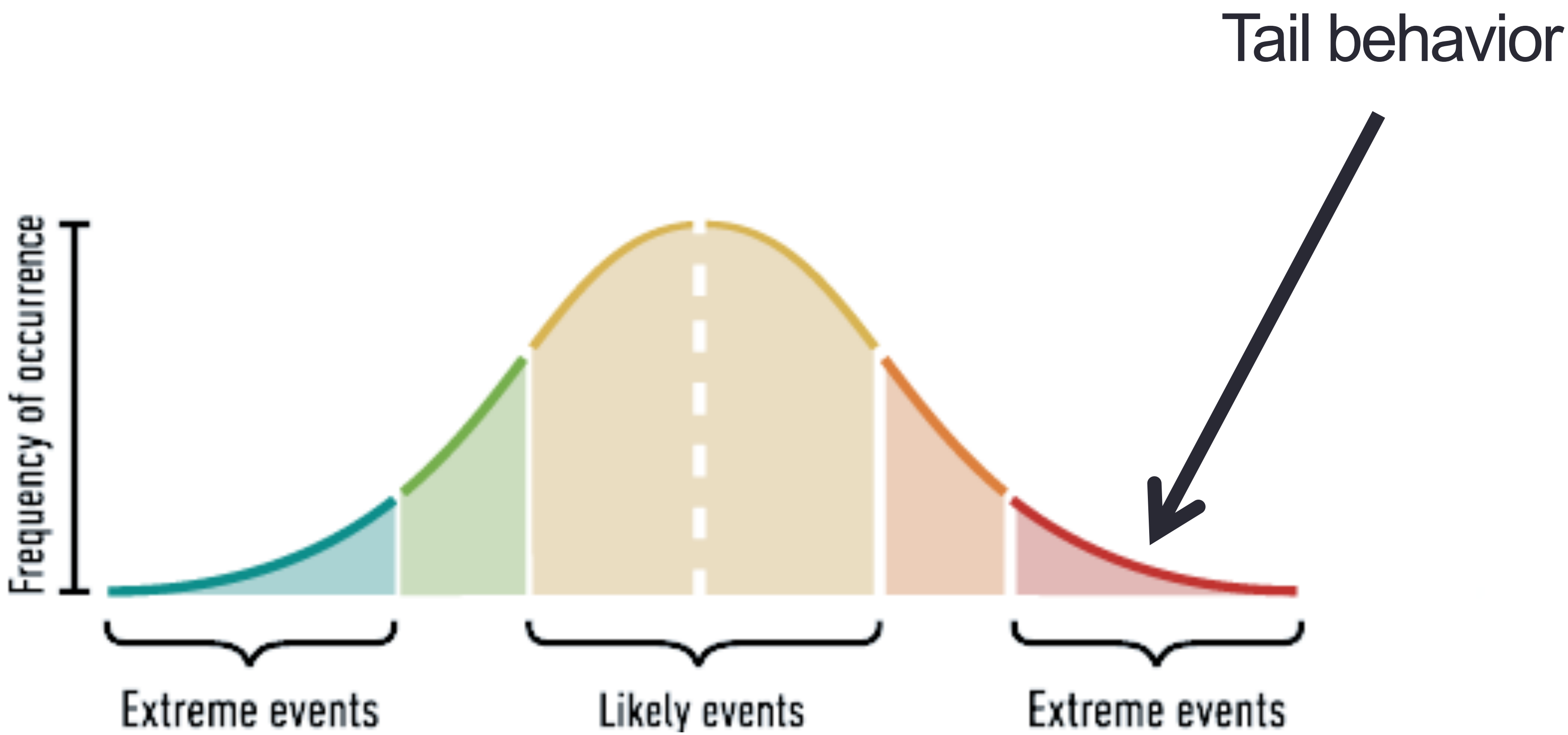
- Emil Gumbel



Extreme Events

- Quantification of extremes is important for environmental disaster planning purposes - flood, wind, mudslides, fire, tornado, temperatures, drought, etc.
- Financial applications: insurance, risk analysis, stock gain/loss, etc.
- Human health: heat waves, ozone/pollutant levels, contamination levels, etc.

Extreme Events



Extreme Events

X(T): **Return Value** X of **Return Period** T

- If T is measured in years: X is the threshold that is exceeded in one year with a probability of 1/T. (One or more exceedances!)
- If T is very large (T >> 1 year) this is equivalent to saying that X is exceeded on average once in T years.

Extreme Events

Example: daily rainfall from Engelberg, Switzerland

- The T=5-year return value rainfall is X=70 mm. <=> A rainfall of 70 mm is exceeded in one year with a probability of 20%.
- The rainfall that occurred on August 21, 1954 (X=89.6 mm) has a return period of T=15 years. <=> Such an event (or larger) is expected on average every 15 years.

Extreme Events

Example: The 100-year flood

Extreme Events

Historic High Water

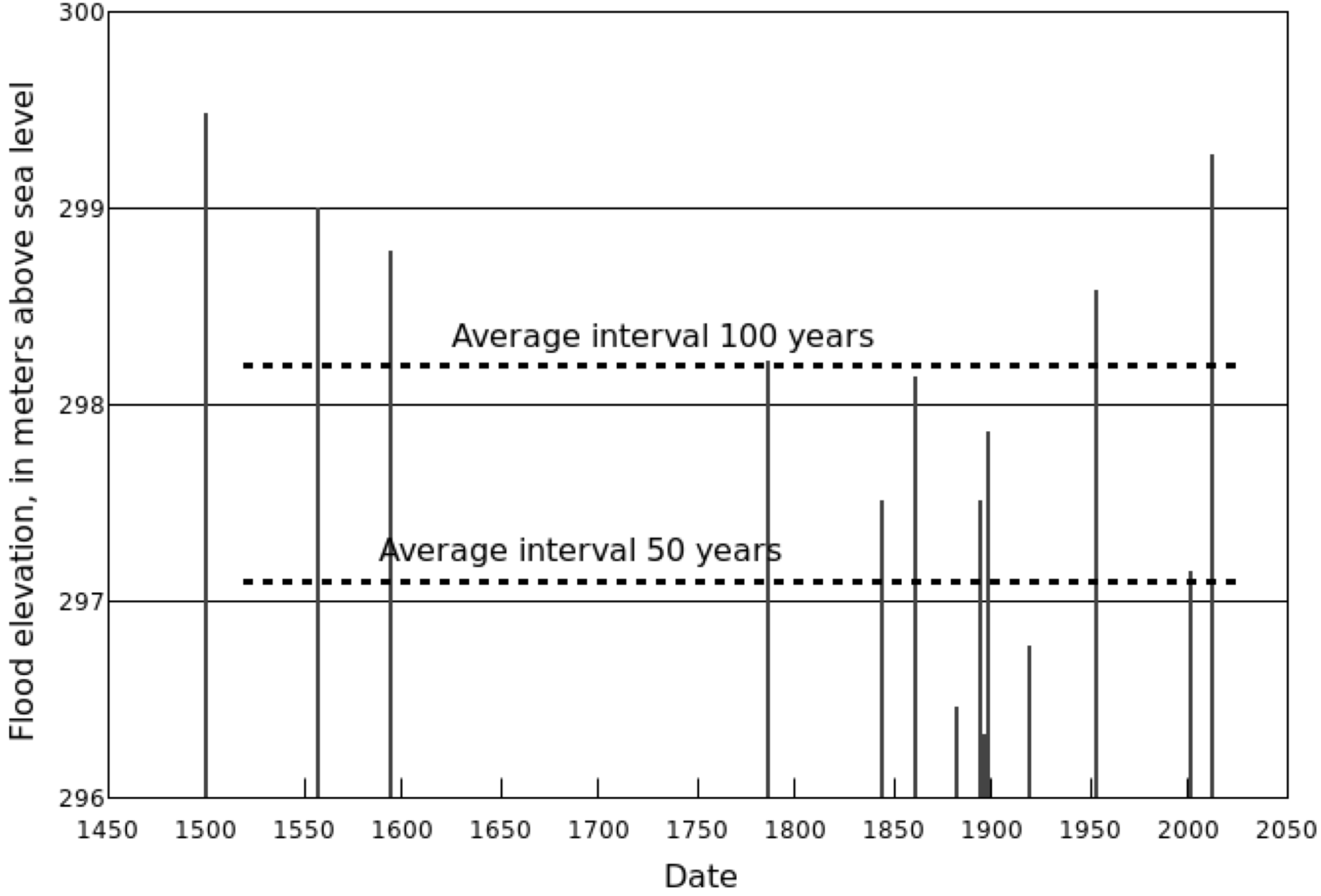
Passau Suffers Worst Flood in 500 Years

Amid the worst flooding seen in half a millennium, the streets of Passau have been transformed into canals, full of soldiers evacuating pensioners in rubber boats. Though the water is receding, the worst is yet to come with the assessment of damages in the Bavarian town.

By Björn Hengst in Passau

Extreme Events

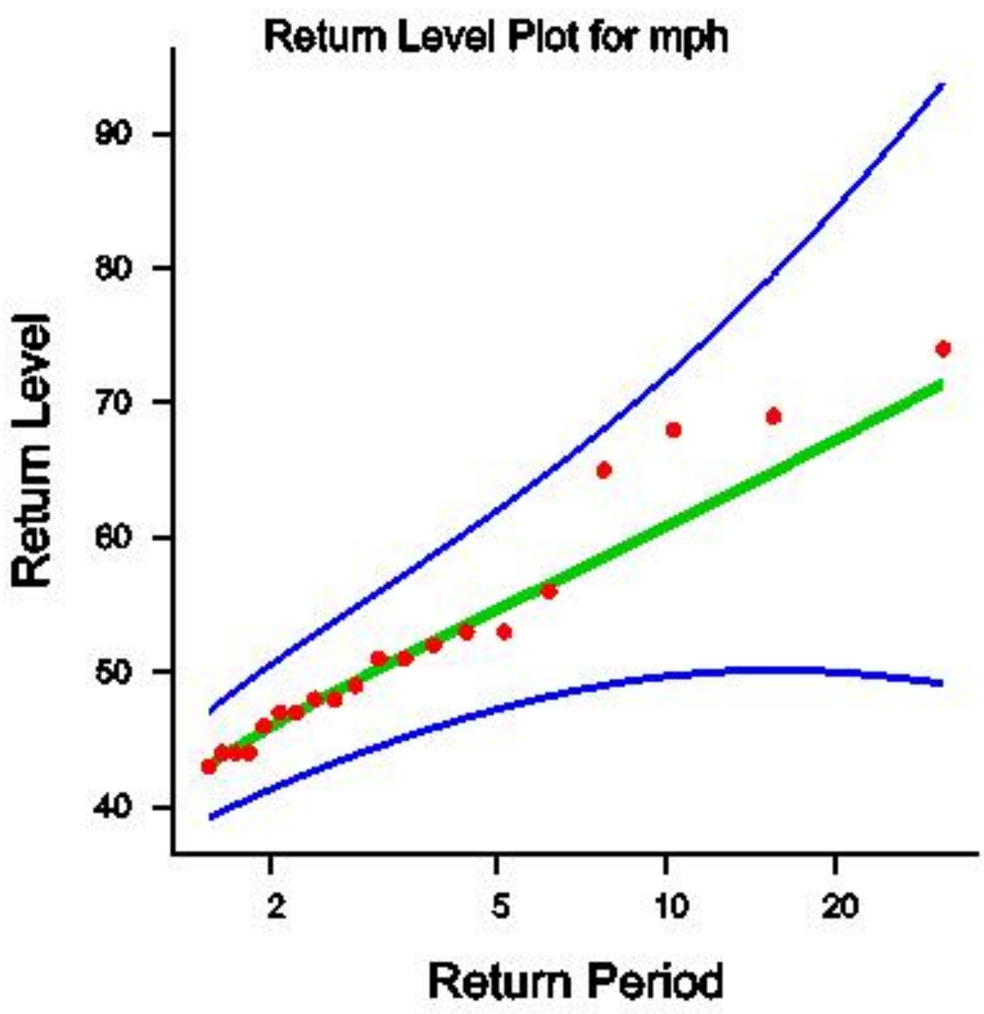
Example: Observed intervals between floods at Passau, 1501-2013



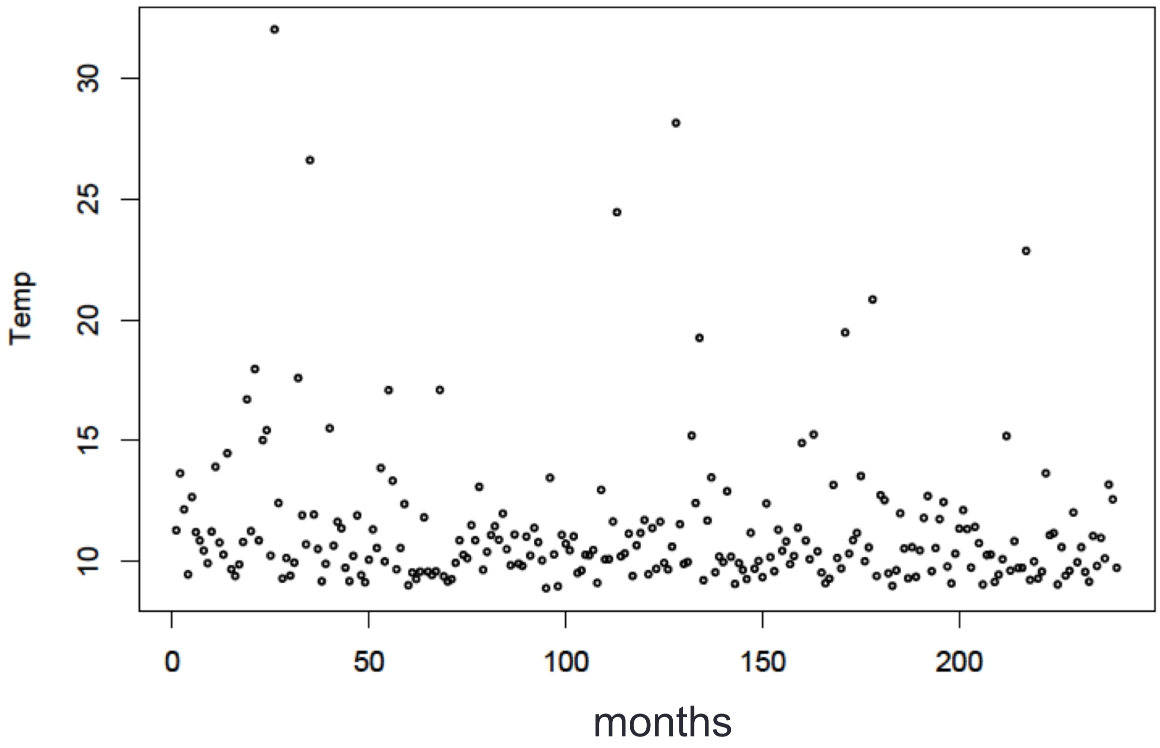
Extreme Events

Purpose of the **Extreme Value Analysis**:

- Find reliable estimates of X(T) for large T (i.e. rare events),
- even for T larger than the period of observation,
- including estimates of the uncertainty of X(T).



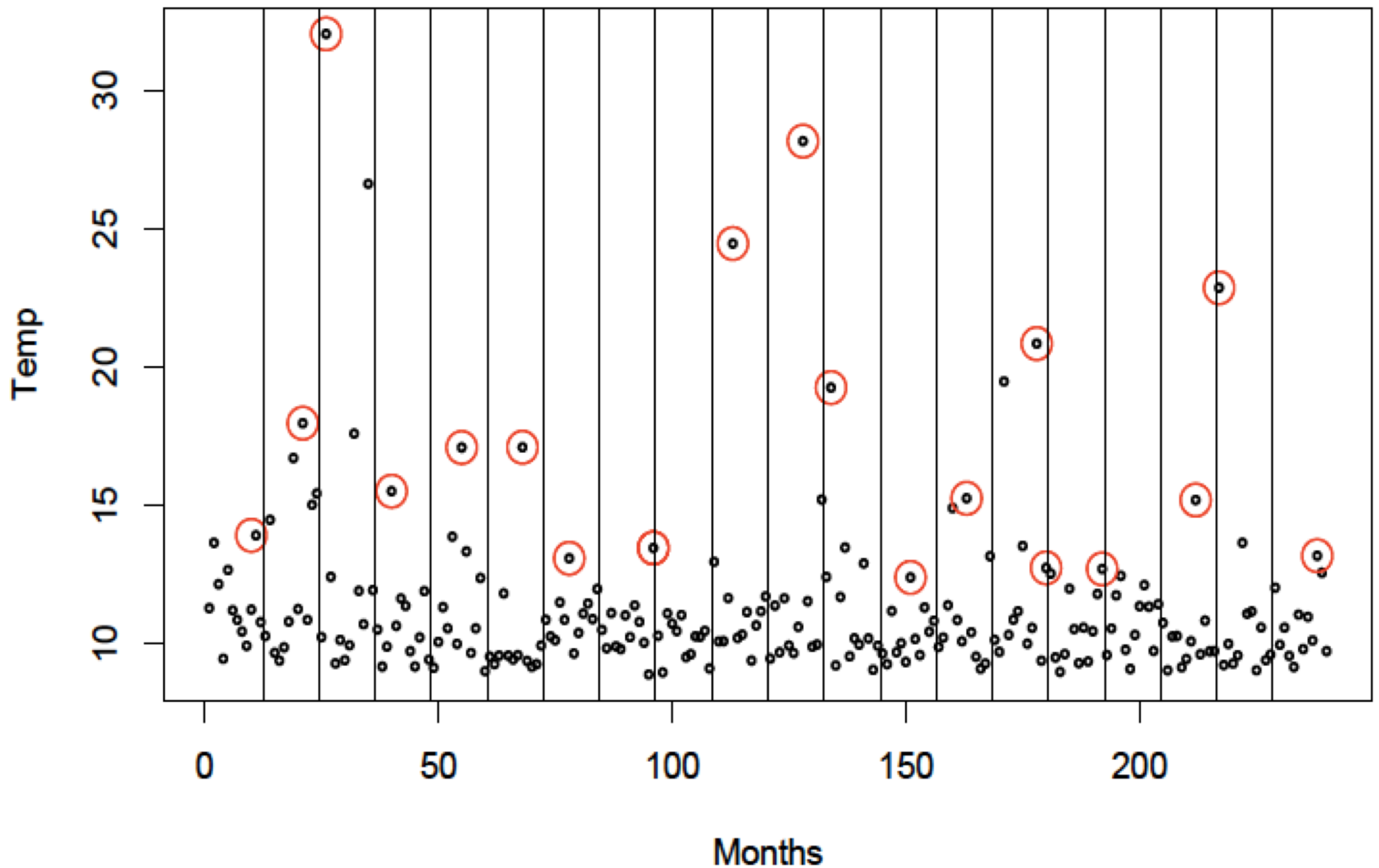
How do we decide what is extreme?



1. The block-maxima approach
2. The peak-over-threshold approach

The Block Maxima Approach:

- Divide full dataset into equal sized chunks of data
- Determine the Max for each block



Asymptotic Distributions Laws

- **The Central Limit Theorem:**

The mean of a large number of independent and identically distributed (iid) random variables is distributed like the Normal Distribution independently of the parent distribution.

- **The Extremal Types Theorem:**

The maximum of a large number of iid random variables is distributed like the Gumbel or Fréchet or Weibull Distributions independently of the parent distribution (if there is convergence at all).

Extremal types distributions

- **The Gumbel distribution (CDF)**

$$G(x) = \exp\left(-\exp(-x)\right)$$

Emil Julius Gumbel
1891-1966

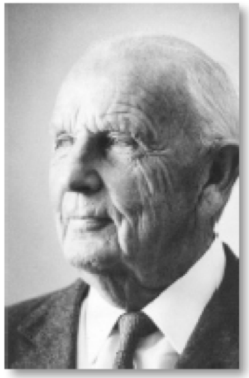
- **The Fréchet distribution (CDF)**

$$G(x) = \begin{cases} 0 & x \le 0 \\ \exp\left(-x^{-\alpha}\right) & x > 0, \alpha > 0 \end{cases}$$

Maurice René Fréchet
1878-1973

- **The Weibull distribution (CDF)**

$$G(x) = \begin{cases} \exp\left(-(-x)^{\alpha}\right) & x < 0, \alpha > 0 \\ 1 & x \ge 0 \end{cases}$$



Ernst Hjalmar Waloddi Weibull
1887-1979

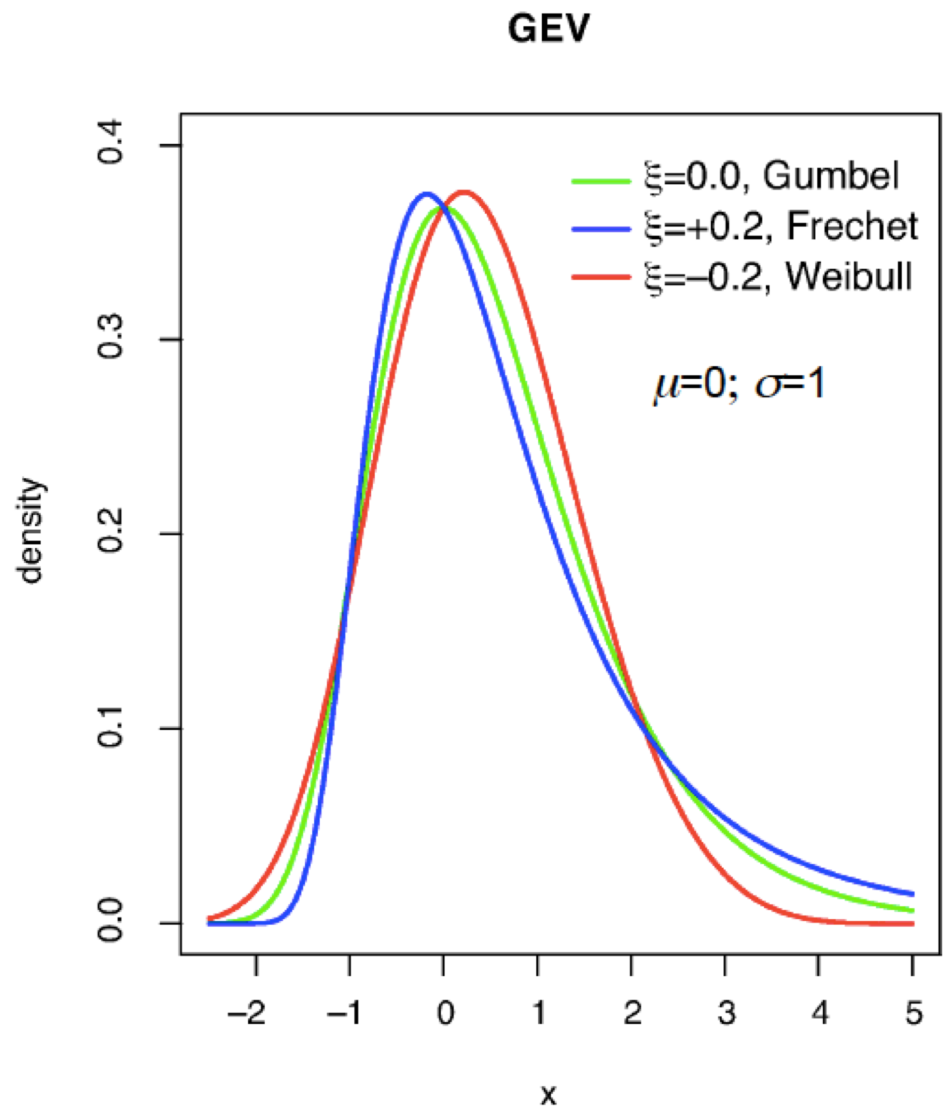
Asymptotic Distributions Laws

- **The GEV Distribution (CDF)**

$$GEV(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

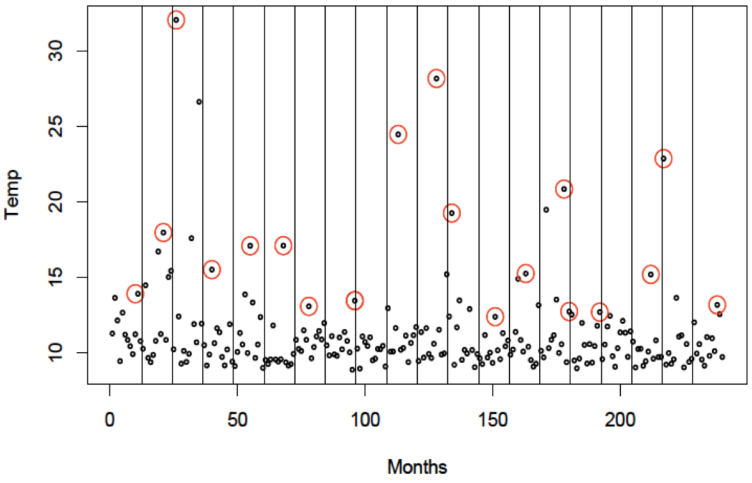
$$\text{where:}\quad 1+\xi\cdot\frac{x-\mu}{\sigma}>0$$

  - A combined parametrisation of all three limit distributions
  - Three parameters: Location (μ), Scale (σ), Shape (ξ)
  - $\xi = 0$: Gumbel, unbounded
    $\xi > 0$: Fréchet, lower bound
    $\xi < 0$: Weibull, upper bound



The Block Maxima Approach:

- Divide full dataset into equal sized chunks of data
- Determine the Max for each block

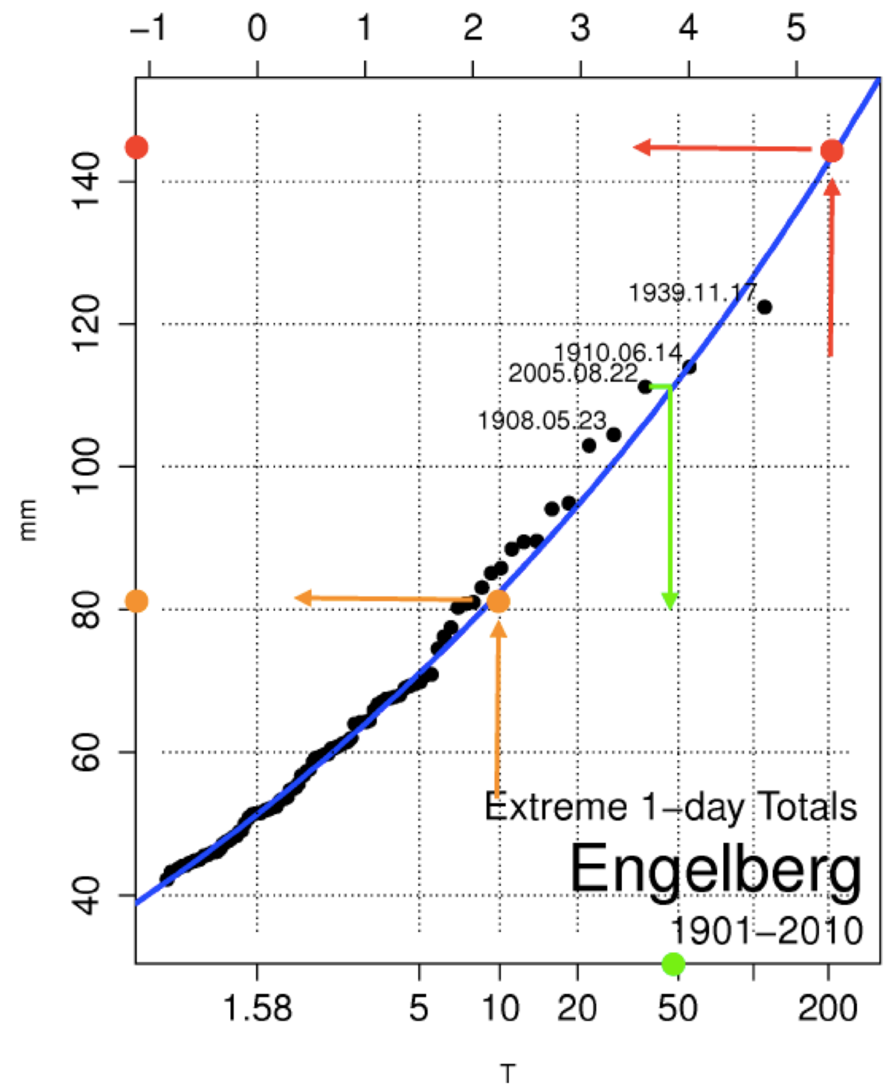


- Fit GEV to the Max and estimate X(T)
  - Estimate parameters of a GEV fitted to the block maxima.
  - Calculate the return value function X(T) and its uncertainty.

The Block Maxima Approach:

Extrapolation:
200-year return value: 145 mm

Return value:
10-year return value: 82 mm

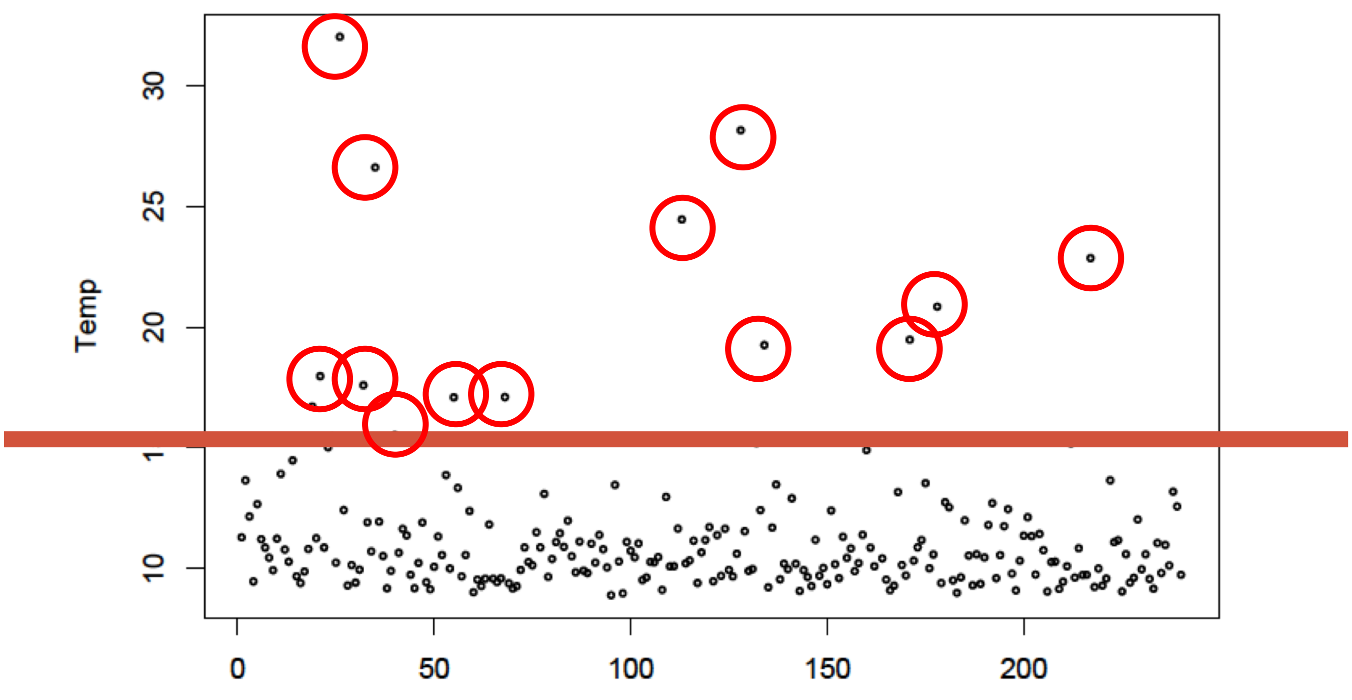


Return period:
Amounts fallen on 2005.08.22 have a return period of 46 years.

The Peak-over-threshold Approach

- Select a large enough threshold
- Extract the exceedances from the dataset.

The differences between the actual values and the threshold value are called exceedances over/under the threshold.



Asymptotic Distributions Laws

- **If Block Maxima of $F(x)$ asymptote to GEV …**

$$\mathrm{prob}\left(M_n < z\right) \approx GEV\left(z;\mu,\sigma,\xi\right) \qquad \text{for } n \to \infty$$

- **… then the distribution of exceedances $E_u(y)$ asymptotes (for large u) to a limit distribution:**

$$E_u\left(y\right) \approx GPD\left(y;\tilde{\sigma},\xi\right) \qquad \text{for } u \to \infty$$

$$GPD\left(y;\tilde{\sigma},\xi\right) = 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-1/\xi} \qquad \text{with} \quad \tilde{\sigma} = \sigma + \xi \cdot \left(u - \mu\right) > 0$$

  - GPD the **Generalized Pareto Distribution**
  - GPD and GEV shape parameters are identical
  - GPD and GEV scale parameters are related.

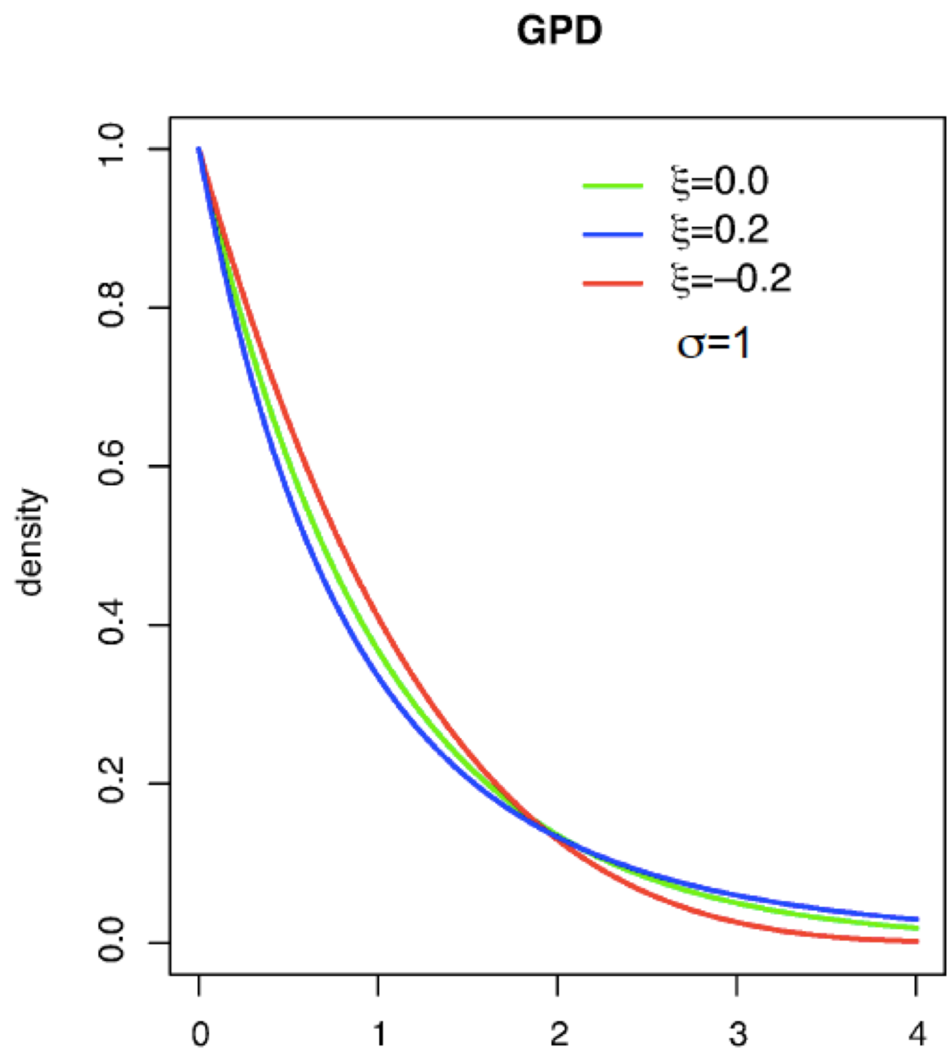
Asymptotic Distributions Laws

- **The GPD (CDF)**

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

$$y \geq 0, \quad 1 + \xi\, y/\sigma \geq 0$$

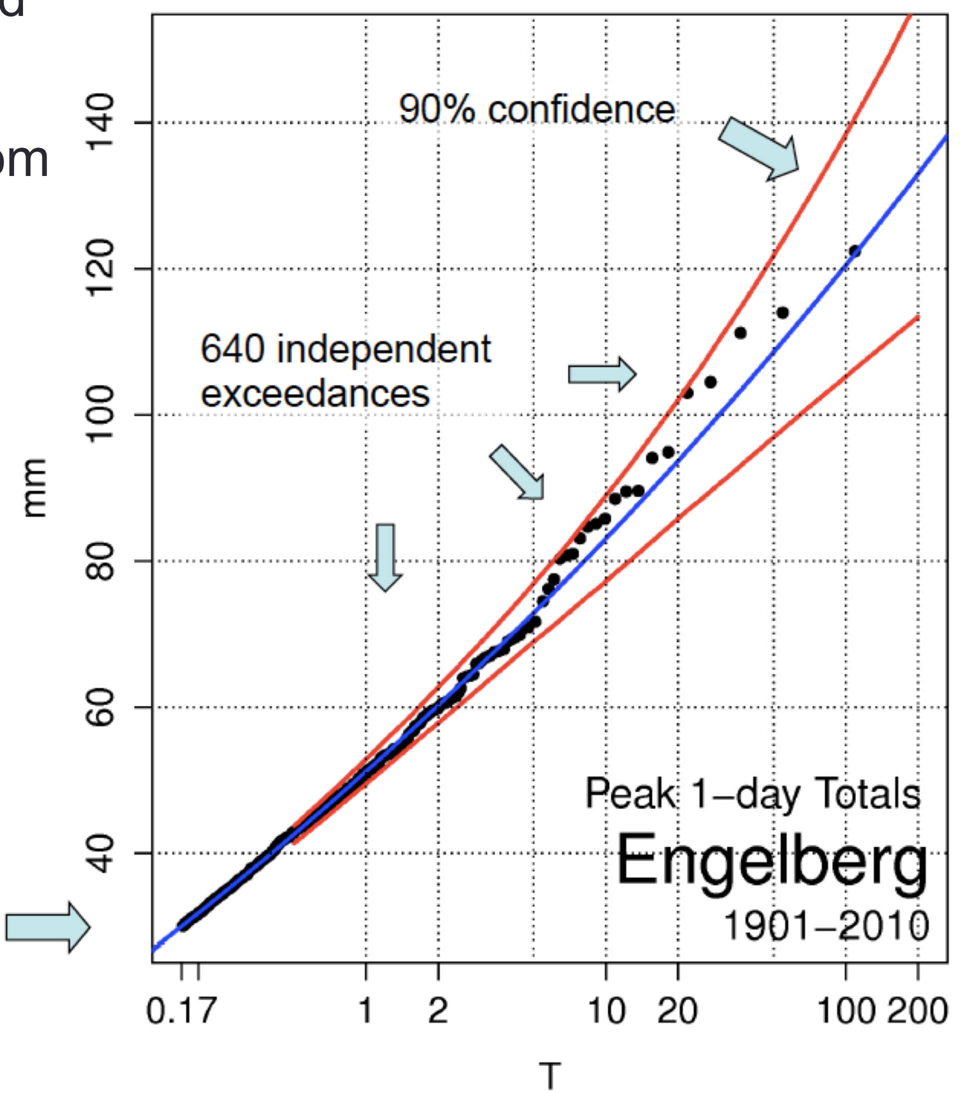
  - Two parameters:
    Scale (σ), Shape (ξ)
  - $\xi = 0$: Exponential Distribution
  - $\xi < 0$: upper bound at $-\sigma/\xi$
  - $\xi > 0$: no upper bound



The Peak-over-threshold Approach

- Select a large enough threshold
- Extract the exceedances from the dataset.
- Fit GPD to exceedances

Threshold:
30 mm



Max-Likelihood
σ = 11.3 mm
ξ = 0.07
$x(T$=200) = 133 mm

Axis in log(T):
Exponential Distribution
(ξ = 0) is a straight line.

Extreme Value Analysis

- See TD5 – Statistical analysis of extreme values

Generalized Maximal Pareto Distributions

- Pickands (1975) demonstrates that when the threshold tends to the upper end of the random variable, the exceedances follow a generalized Pareto distribution, $\mathrm{GPD}_M(\alpha, \kappa)$, with cdf

$$F_M(x) = \begin{cases} 1 - (1 - \kappa x / \alpha)^{1/\kappa}, & \kappa \neq 0, \\ 1 - e^{-x/\alpha}, & \kappa = 0. \end{cases}$$

Generalized Maximal Pareto Distribution

The GPD_M family has two parameters:

- α is a scale parameter ($\alpha > 0$)
- κ is a shape parameter

Note that when $\kappa \le -1/2,\ \mathrm{Var}(X) = \infty.$

The *p*th quantile is ($0 < p < 1$):

$$x(p) = \alpha\left(1 - (1 - p)^{\kappa}\right)/\kappa$$

Special Cases of the Maximal GPD

The GPD_M has three special cases:

1. When $\kappa = 0$, the GPD_M reduces to the Exponential distribution with mean α.
2. When $\kappa = 1$, the GPD_M reduces to the Uniform $U(0, \alpha)$.
3. When $\kappa < 0$, the GPD_M becomes the Pareto distribution.