An Introduction to Extreme Value Theory

Petra Friederichs

Meteorological Institute
University of Bonn

COPS Summer School, July/August, 2007

Applications of EVT

Finance

· distribution of income has so called fat tails

· value-at-risk: maximal daily lost

· re-assurance

Hydrology

· protection against flood

· Q100: maximal flow that is expected once every 100 years

Meteorology

· extreme winds

· risk assessment (e.g. ICE, power plants)

· heavy precipitation events

· heat waves, hurricanes, droughts

· extremes in a changing climate

Application

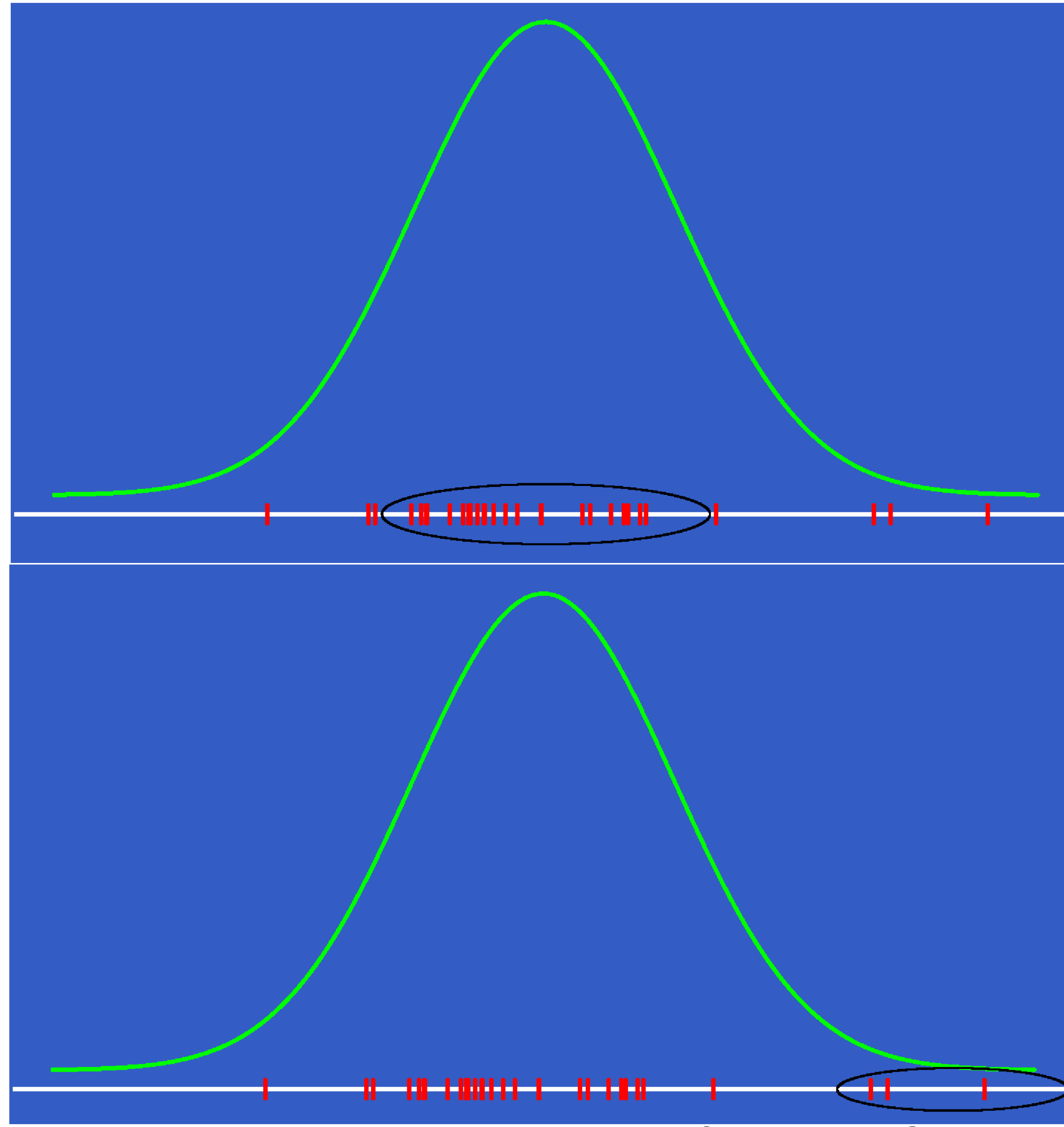
In classical statistics:
focus on AVERAGE behavior of stochastic process

central limit theorem

In extreme value theory:
focus on extreme and rare events

Fisher-Tippett theorem

What is an extreme?



from Ulrike Schneider

Extreme Value Theory

Block Maximum

$$M_n = max\{X_1, \ldots, X_n\}$$

for $n \rightarrow \infty$

M_n follows a **Generalized Extreme Value (GEV)** distribution

Peak over Threshold (POT)

$$\{X_i - u | X_i > u\}$$

very large threshold u

follow a **Generalized Pareto Distribution (GPD)**

Poisson-Point GPD Process

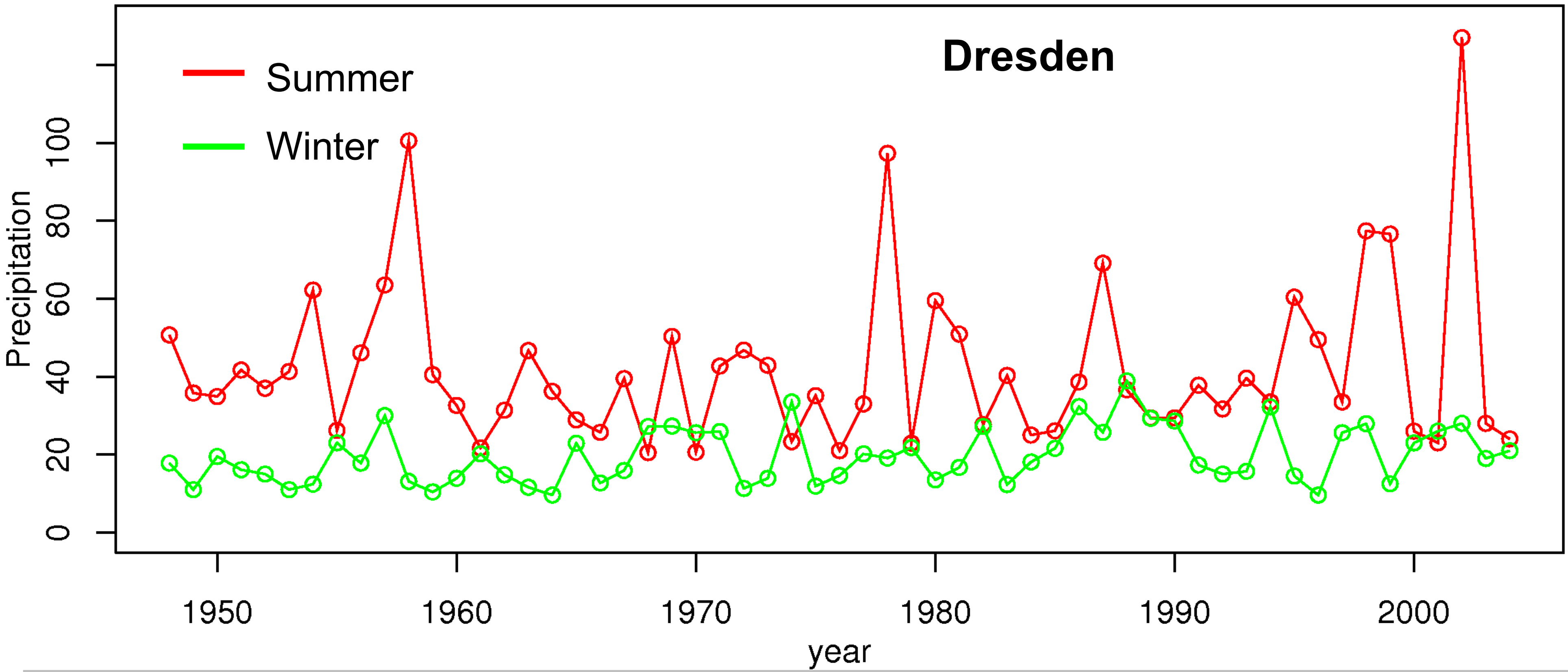
combines POT with Poisson point process

Extreme Value Theory

Block Maximum $M_n = max\{X_1, \ldots, X_n\}$

Example**:** station precipitation (DWD) at Dresden

1948 – 2004: 57 seasonal (Nov-March and May-Sept)
(maxima over approximately 150 days)



GEV – Fisher-Tippett Theorem

The distribution of $M_n = max\{X_1, \ldots, X_n\}$

converges to ($n \rightarrow \infty$)

$$\xi \neq 0 \qquad G(y) = \exp(-[1 + \xi(\frac{y-\mu}{\sigma})]^{-1/\xi})$$

$$\xi = 0 \qquad G(y) = \exp(-\exp(-\frac{y-\mu}{\sigma}))$$

which is called the **Generalized Extreme Value (GEV)** distribution.

It has three parameters

- μ location parameter
- σ scale parameter
- ξ shape parameter

GEV – Types of Distributions

GEV has 3 types depending on shape parameter ξ

$$x = \frac{y - \mu}{\sigma}$$

Gumbel $\xi = 0$

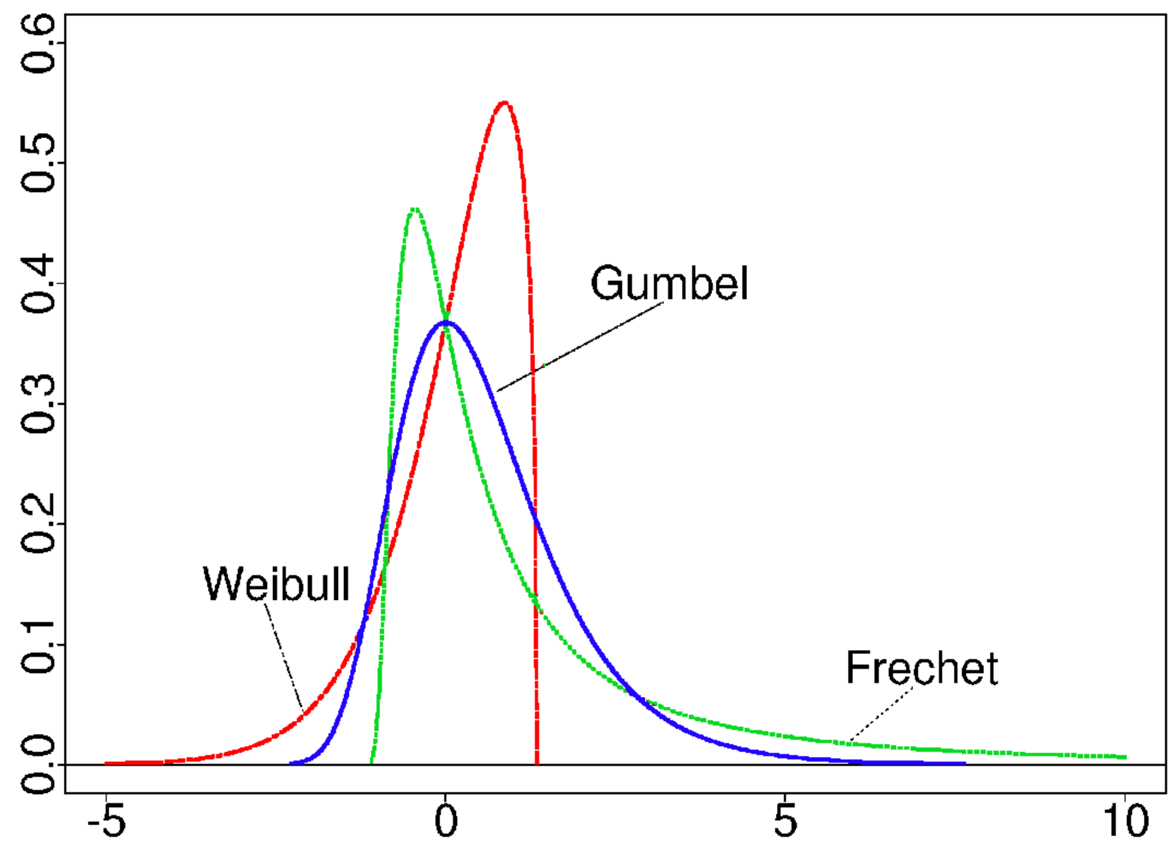
$$G(x) = \exp(-\exp[-x])$$

Fréchet $\xi = 1/\alpha > 0$

$$G(x) = \exp(-[1 + \frac{x}{\alpha}]^{-\alpha})$$

Weibull $\xi = -1/\alpha < 0$

$$G(x) = \exp(-[1 - \frac{x}{\alpha}]^{\alpha})$$



GEV – Types of Distributions

GEV has 3 types depending on shape parameter ξ

$$x = \frac{y - \mu}{\sigma}$$

Gumbel $\xi = 0$

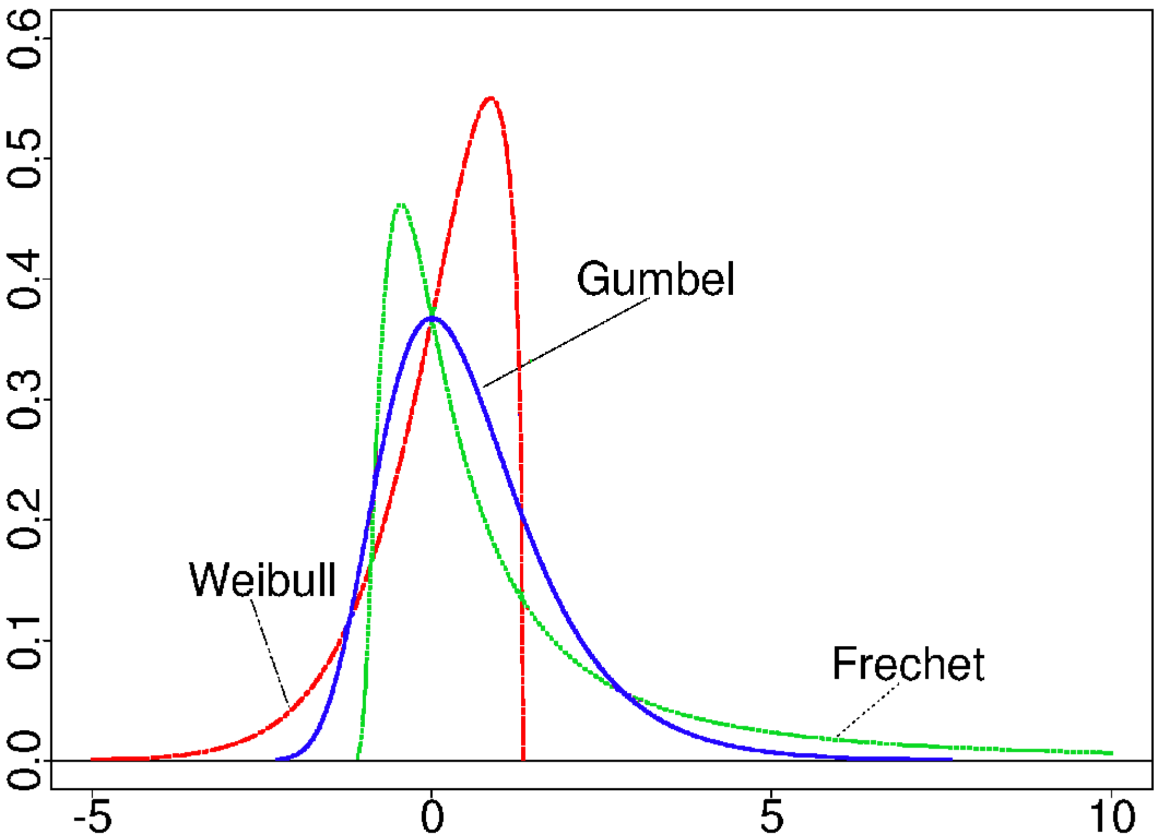
exponential tail

Fréchet $\xi = 1/\alpha > 0$

so called „fat tail“

Weibull $\xi = -1/\alpha < 0$

upper finite endpoint



GEV – Types of Distribution

Conditions to the sample $\{X_1, \ldots, X_n\}$ from which the maxima are drawn X_i must be **independently identically distributed** (i.i.d.)

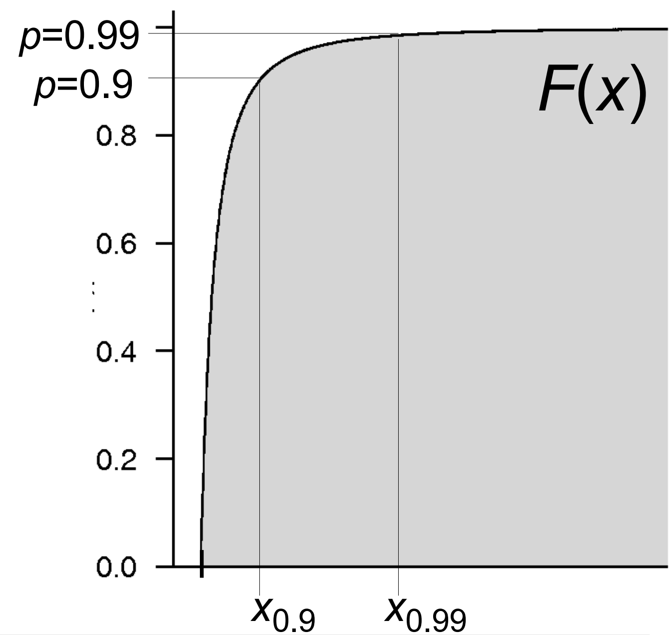
Let $F(x)$ be the distribution of X_i

$F(x)$ is in the domain of attraction of a **Gumbel** type GEV iff

$$\lim_{x \to \infty} \frac{1 - F(x + tb(x))}{1 - F(x)} = e^{-t}$$

for all $t > 0$

Exponential decay in the tail of $F(x)$



GEV – Types of Distribution

Conditions to the sample $\{X_1, \ldots, X_n\}$ from which the maxima are drawn

X_i must be **independently identically distributed** (i.i.d.)

Let $F(x)$ be the distribution of X_i

$F(x)$ is in the domain of attraction of a **Frechet** type GEV iff

$$\lim_{x \to \infty} \frac{1-F(\lambda x)}{1-F(x)} = \lambda^{-1/\xi}$$

for all $\lambda > 0$

polynomial decay in the tail of $F(x)$

GEV – Types of Distribution

Conditions to the sample $\{X_1, \ldots, X_n\}$ from which the maxima are drawn

X_i must be **independently identically distributed** (i.i.d.)

Let $F(x)$ be the distribution of X_i

$F(x)$ is in the domain of attraction of a **Weibull** type GEV iff

there exists ω_F with $F(\omega_F)=1$

and

$$\lim_{x\to\infty}\left(1-F\left(\omega_F-\frac{1}{\lambda x}\right)\right)\left(1-F\left(\omega_F-\frac{1}{x}\right)\right)^{-1}=\lambda^{1/\xi}$$

for all $\lambda>0$

$F(x)$ has a finite upper end point ω_F

GEV – return level

Of interest often is **return level** z_m

value expected every *m* observation (block maxima)

$$Prob(y > z_m) = 1 - G(y \leq z_m) = \frac{1}{m}$$

calculated using invers distribution function (**quantile function**)

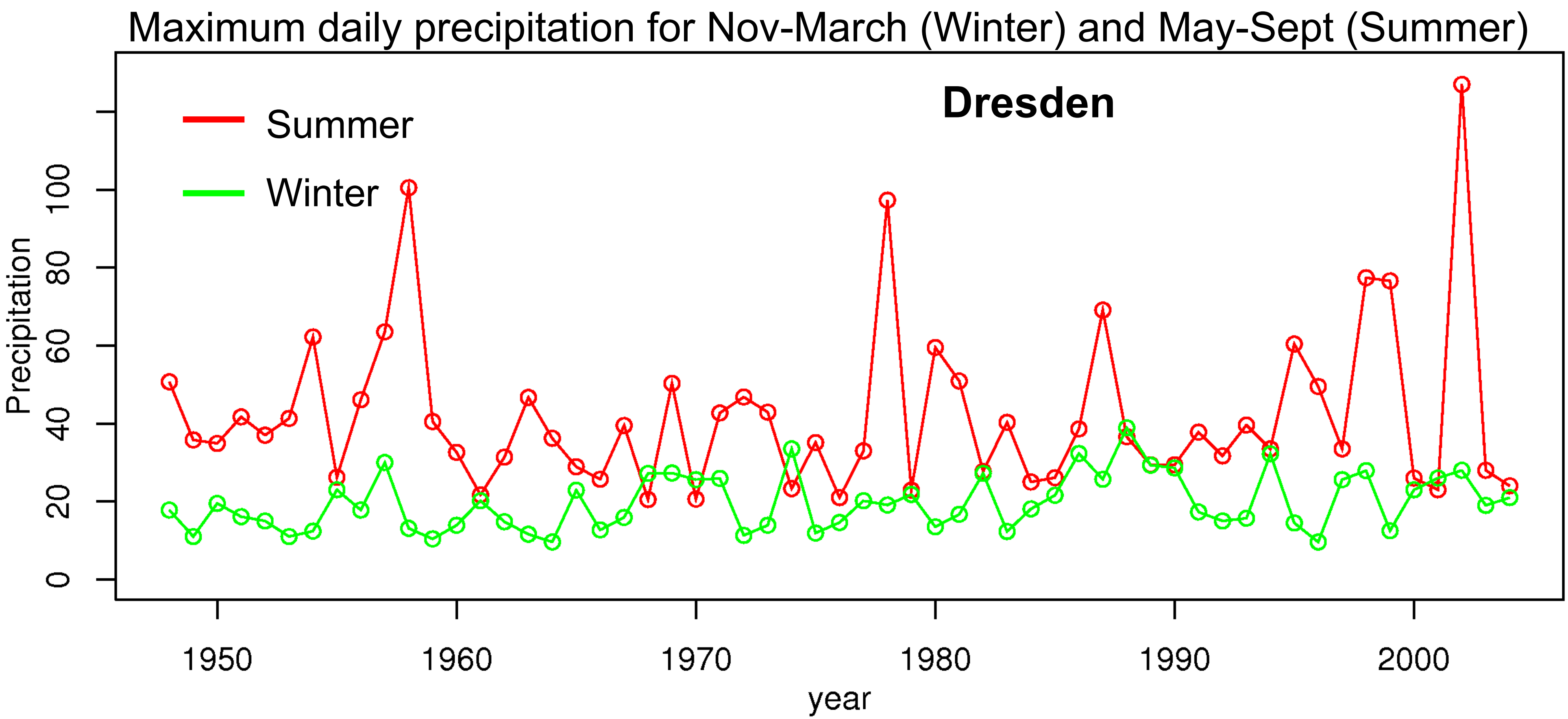
$$z_m = G^{-1}(1 - \frac{1}{m}) = \mu - \frac{\sigma}{\xi}(1 - (-\log(1 - 1/m))^{-\xi})$$

$$z_m = \mu - \sigma \log(-\log(1 - 1/m))$$

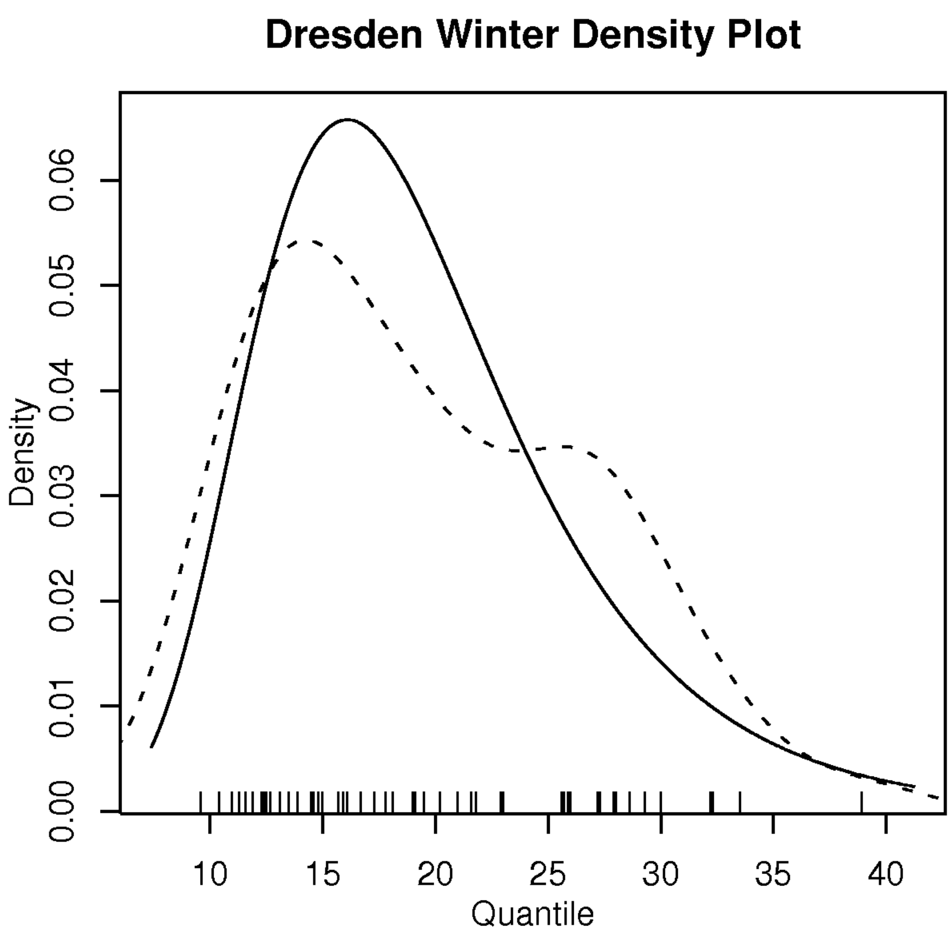
can be estimated empirically as the

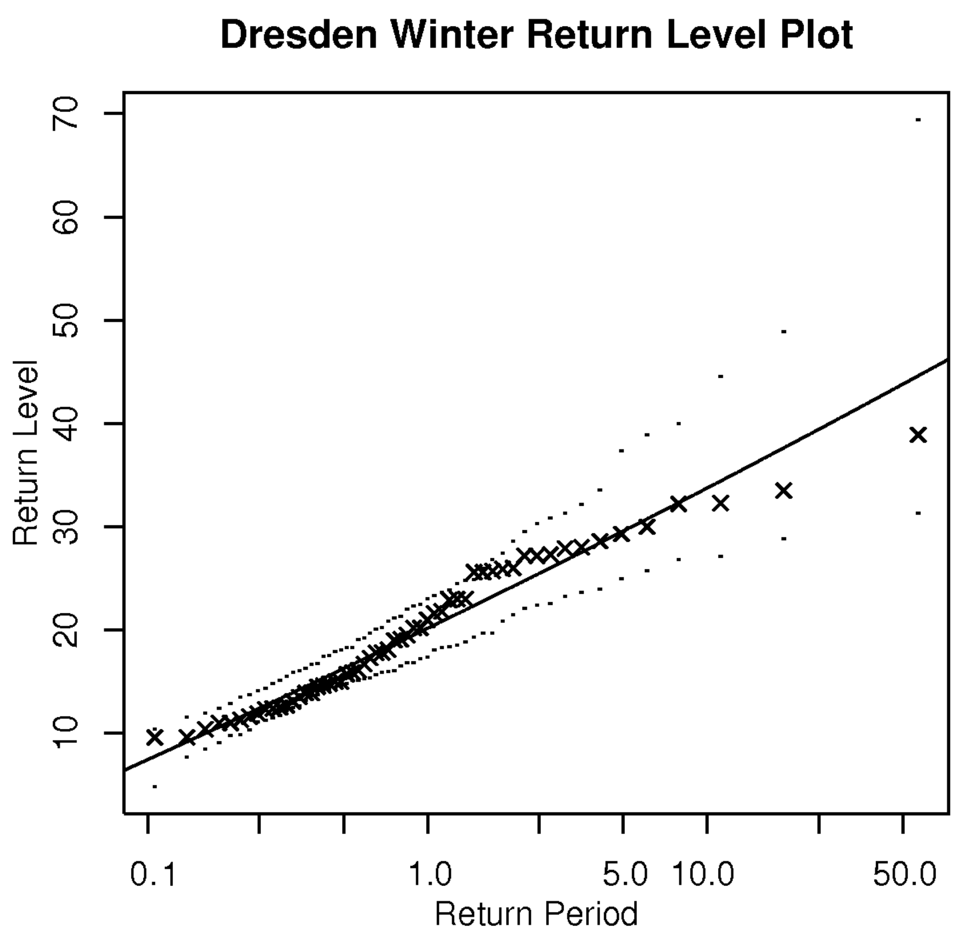
$$\hat{z}_m = \hat{G}^{-1}(1 - \frac{1}{m}) = inf(y \mid F(y) \geq \frac{1}{m})$$

Block Maxima

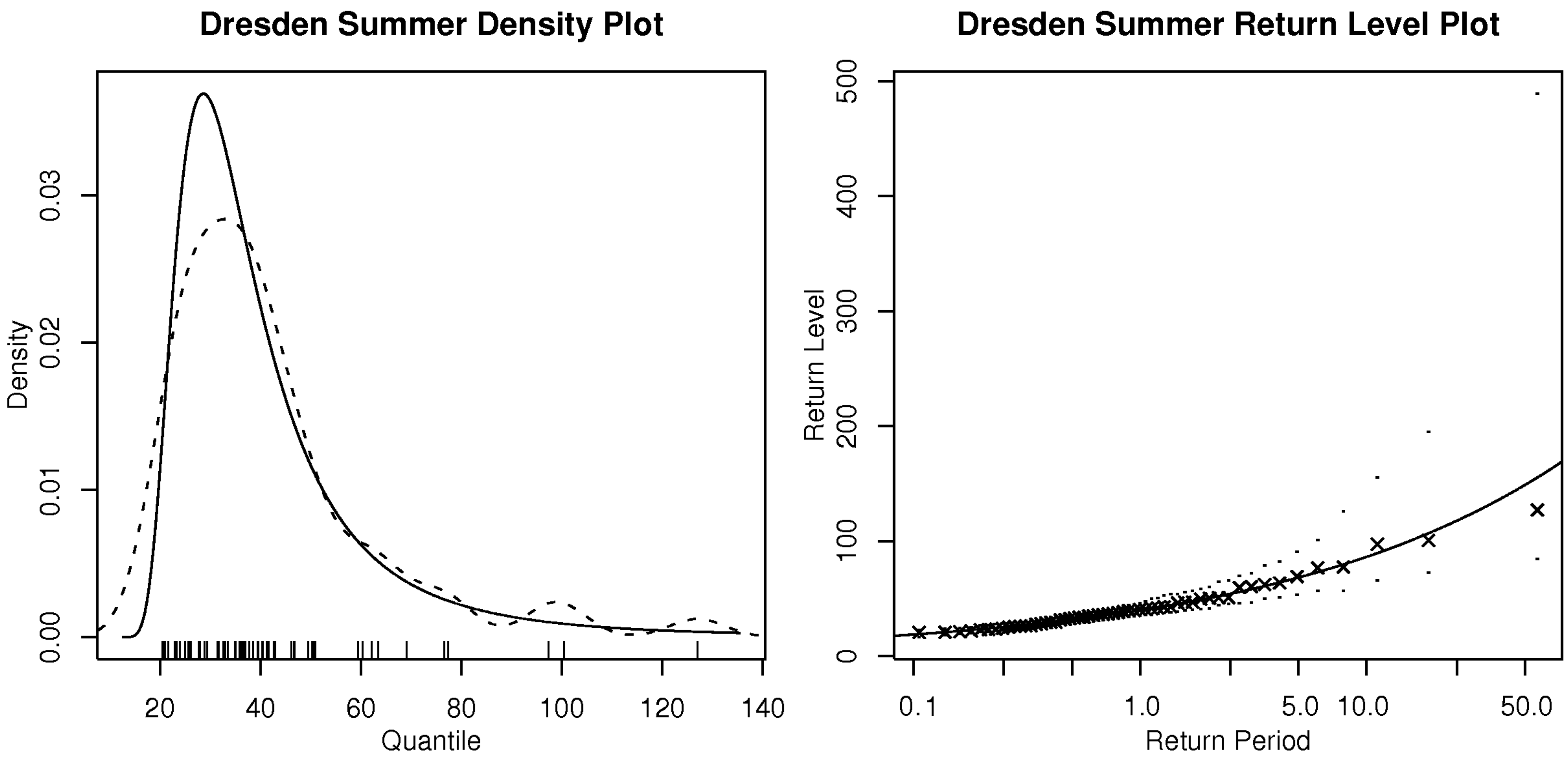


Block Maxima





Block Maxima



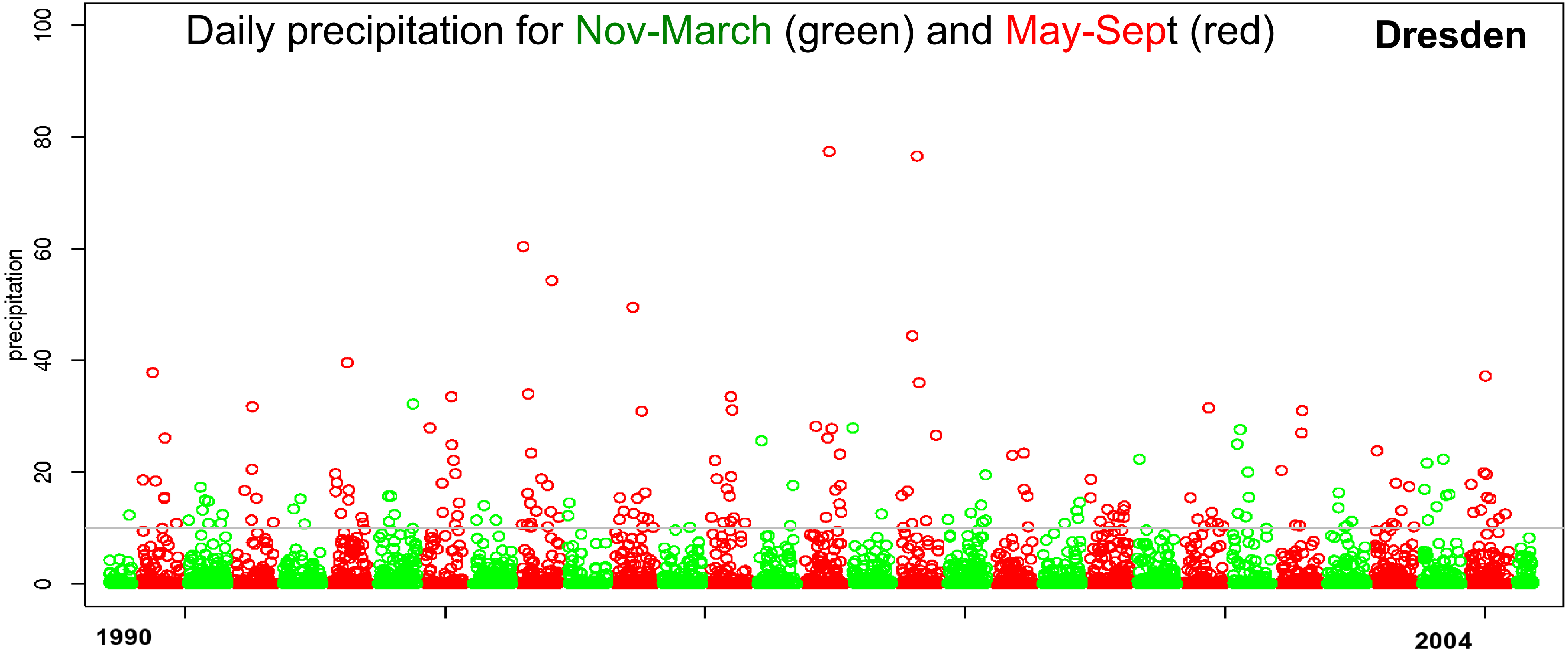
Peak over Threshold

Peak over Threshold (POT)

$$\{X_i - u | X_i > u\}$$

very large threshold u

follow a **Generalized Pareto Distribution (GPD)**



Peak over Threshold

Peak over Threshold (POT)

The distribution of $Y_i := X_i - u | X_i > u$

exceedances over large threshold u

are asymptotically distributed following a

Generalized Pareto Distribution (GPD)

$$H(y|X_i > u) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-1/\xi}$$

two parameters

- σ scale parameter
- ξ shape parameter

advantage: more efficient use of data

disadvantage: how to choose threshold not evident

POT – Types of Distribution

GDP has same 3 types as GEV depending on shape parameter ξ

Gumbel $\xi = 0$

$$H(y) = 1 - \exp(-\frac{y}{\sigma_u})$$

exponential tail

Pareto (Fréchet) $\xi > 0$

$$1 - H(y) \sim c\, y^{-1/\xi}$$

polynomial tail behavior

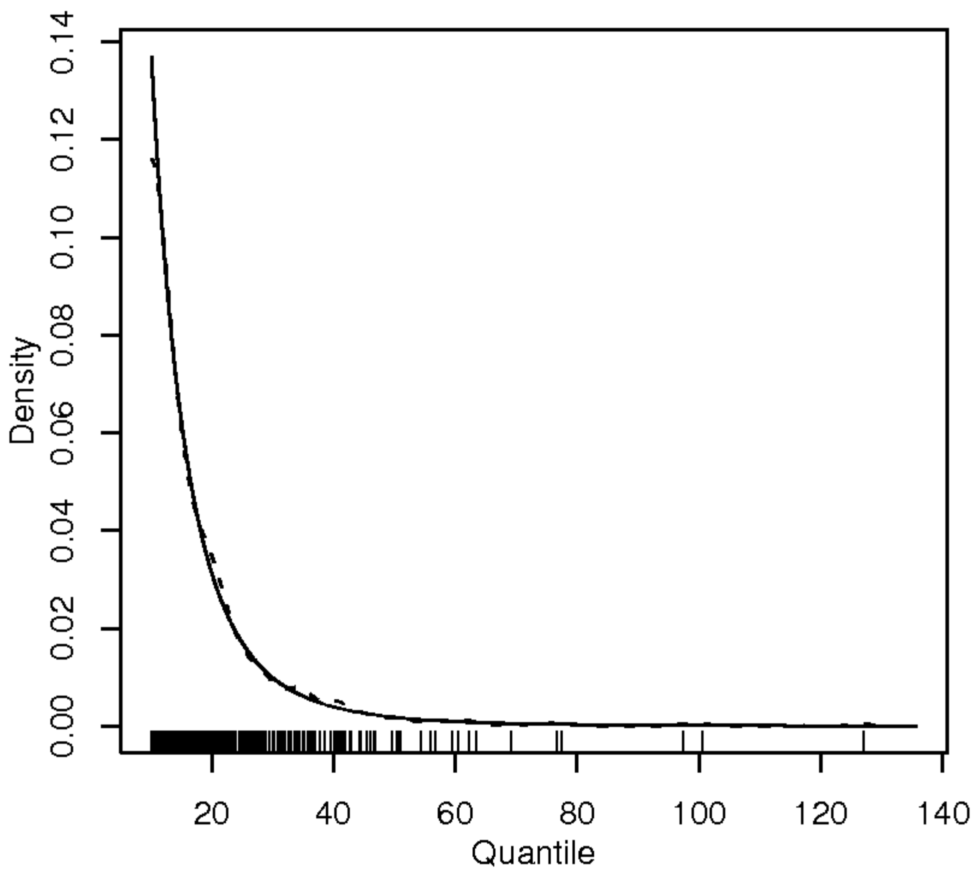
Weibull $\xi < 0$

has upper end point $\omega_F = \dfrac{\sigma_u}{|\xi|}$

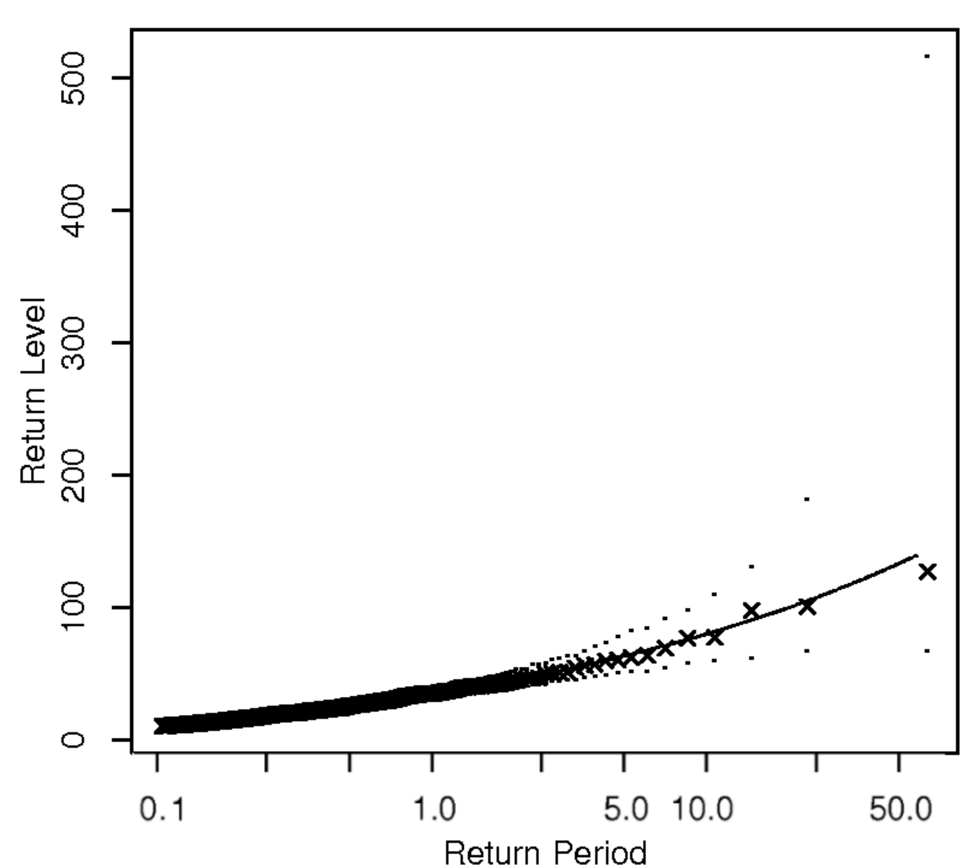
Peak over Threshold

POT: threshold u = 15*mm*

Dresden Summer Density Plot



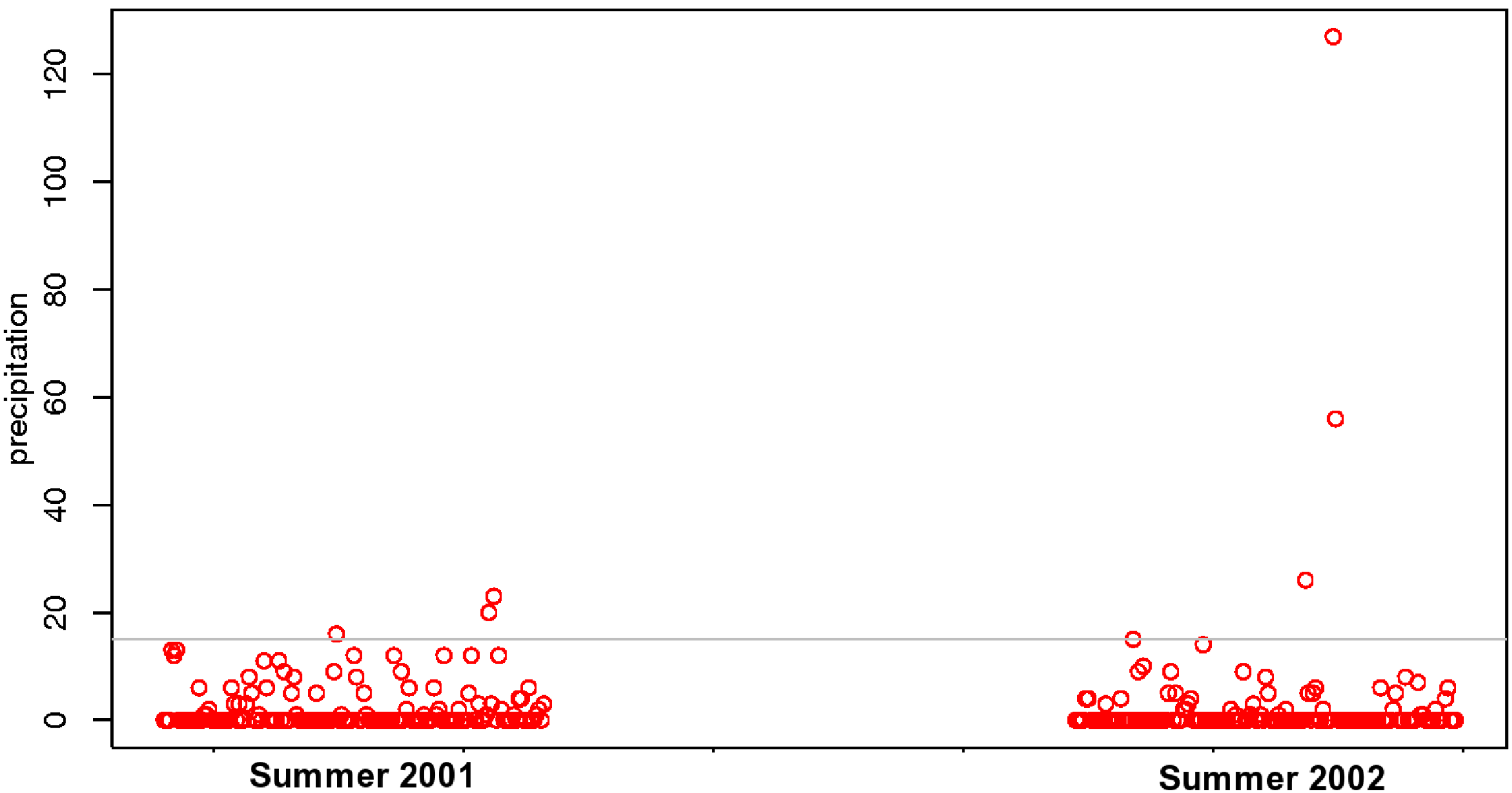
Dresden Summer Return Level Plot



Poisson Point – GPD Process

Poisson point – GPD process with intensity

$$\Lambda(A) = (t_2 - t_1)\left[1 + \xi\left(\frac{y-u}{\sigma_u}\right)\right]^{-1/\xi} \quad \text{on} \quad A = (t_1, t_2) \times (y, \infty)$$



Estimation and Uncertainty

General concept of estimating parameters from a sample

$$\{y_i\},\ i=1,\ldots,n$$

Maximum Likelihood (ML) Method

Assume $\{y_i\}$ are draw from a GEV (GPD,...) with unknown parameters

$$y_i \sim F(y|\mu,\sigma,\xi)$$

and PDF

$$f(y|\mu,\sigma,\xi)=F'(y|\mu,\sigma,\xi)$$

Maximum Likelihood Method

The likelihood L of the sample is then

$$L(\mu,\sigma,\xi)=\prod_{i=1}^{n} f(y_i|\mu,\sigma,\xi)$$

It is easier to minimize the negative logarithm of the likelihood

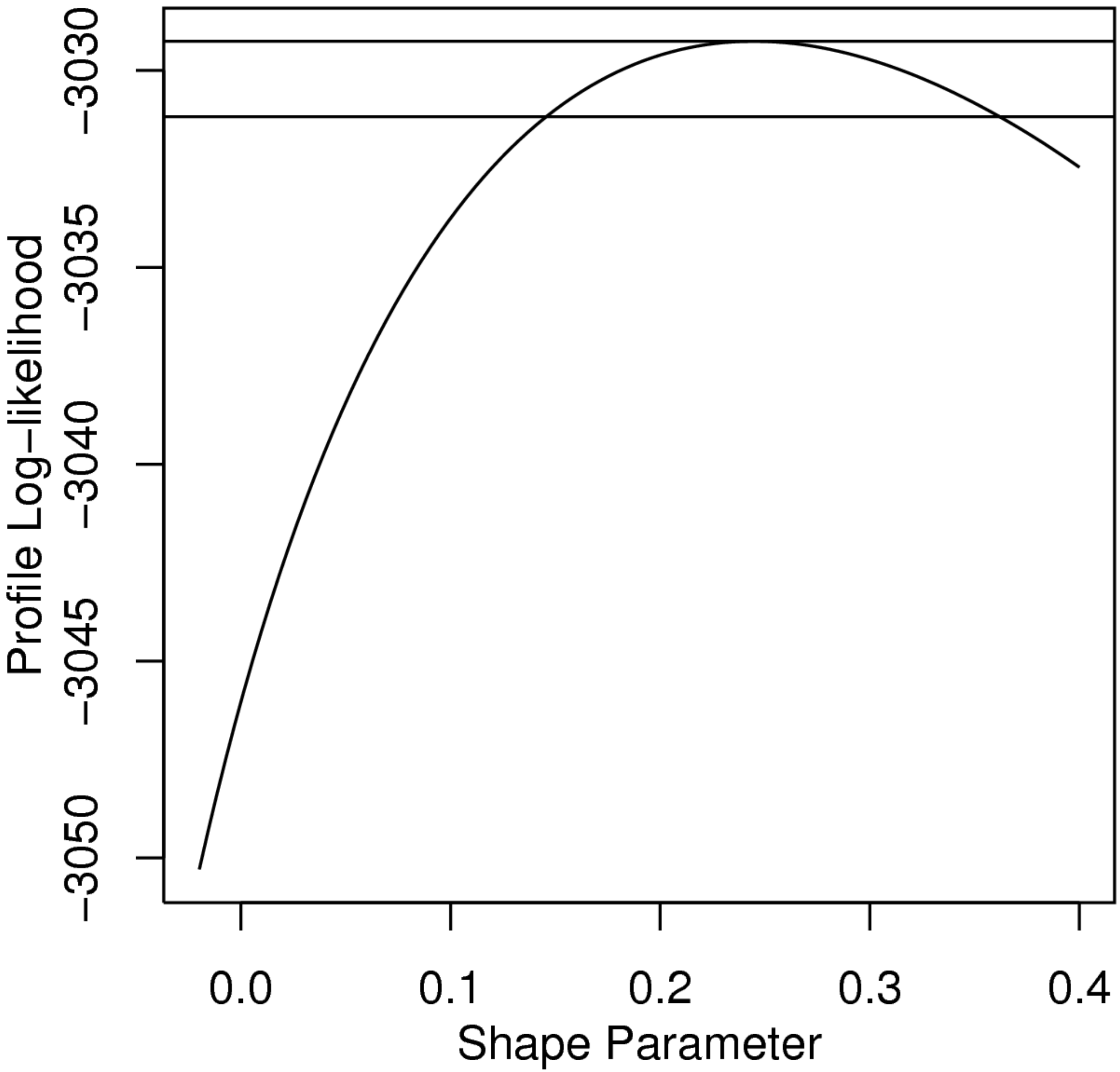
$$l(\mu,\sigma,\xi)=-\sum_{i=1}^{n} \log f(y_i|\mu,\sigma,\xi)$$

in general there is no analytical solution for the minimum with respect to the parameters

Minimize using numerical algorithms.

The estimates $\hat{\mu},\hat{\sigma},\hat{\xi}$ maximize the likelihood of the data.

Profile log Likelihood



To Take Home

There exists a well elaborated **statistical theory for extreme values**.

It applies to (almost) all (univariate) extremal problems.

EVT: extremes from a very large domain of stochastic processes follow one of the three types: **Gumbel**, **Frechet/Pareto**, or **Weibull**

Only those three types characterize the behavior of extremes!

Note: Data need to be in the asymptotic limit of a EVD!

References

Coles, S (2001): An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. Springer Verlag London. 208p

Beirlant, J; Y. Goegebeur; J. Segers; J. Teugels (2005): Statistics of Extremes. Theory and Applications. John Wiley & Sons Ltd. 490p

Embrechts, Küppelberg, Mikosch (1997): Modelling Extremal Events for Insurance and Finance. Springer Verlag Heidelberg.648p

Gumbel, E.J. (1958): Statistics of Extremes. (Dover Publication, New York 2004)

R Development Core Team (2003): R: A language and environment for statistical computing, available at http://www.R-project.org

The **`evd`** and **`ismev`** Packages by Alec Stephenson and Stuard Coles