Section 4:
Extreme Value Analysis –
An Introduction
Motivation

• What can happen in extreme cases?
  o Design of weather exposed constructions (a pylon, wind power plant, …)
  o Plan constructions for the protection against natural disasters.

• Sea dams in the Netherlands are dimensioned for the one in 10’000 year event. (?!)

• Society is risk-averse!
Outline

• Theoretical Background

• Modelling Block Maxima

• Modelling Peaks over Threshold

• Additional Remarks

• Material based on:
  o Coles 2001 (Chap. 1-4)
Section 4: Extreme Value Analysis – An Introduction

Theoretical Background
Return Value and Return Period

- \( X(T) \): 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 (\( T \gg 1 \) year) this is equivalent to saying that \( X \) is exceeded on average once in \( T \) years.
  - \( X(T) \): Amplitude as function of rareness (the quantile function)

- **Example: Engelberg (daily rainfall)**
  - 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.
What if we’d use data only?

Extreme 2-day precipitation in Entlebuch (1901-2004/5)

<table>
<thead>
<tr>
<th>Year</th>
<th>Precipitation (mm)</th>
<th>Empirical return period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1901-2004</td>
<td></td>
<td>2004</td>
</tr>
<tr>
<td>1946.07.05</td>
<td>133</td>
<td>T=52</td>
</tr>
<tr>
<td>1946.08.22</td>
<td>132</td>
<td>T=35</td>
</tr>
<tr>
<td>1984.08.09</td>
<td>150</td>
<td>T=104</td>
</tr>
<tr>
<td>2005.08.21</td>
<td>177</td>
<td>T=?</td>
</tr>
</tbody>
</table>

Empirical return period:
- 2004: T=52, T=35
- 2005: T=105, T=53, T=35, T=27
Extreme Value Analysis

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

• **Procedure**
  o Choose an appropriate parametric distribution function
  o Calibrate it such that it describes available data well
  o Extrapolate distribution function
Distribution of Maxima

- **Independent identically distributed random vars:**
  
  \[ X_1, X_2, X_3, \ldots, X_n \quad X_k \sim F(x) \quad iid \]

  - E.g. representative for daily precipitations in a year \((n=365)\)
  - \(F(x):= \text{prob}(X_k \leq x)\), the *parent distribution* (CDF)

- **The Maximum**

  \[ M_n = \max(X_1, X_2, X_3, \ldots, X_n) \]

  - E.g. the largest 24-hour total in a year
  - A random variable with:
    \[ M_n \sim F^n(x) \]

  - Because:
    \[ \text{prob}(M_n \leq x) = \text{prob}(X_1 \leq x, \ldots, X_n \leq x) = \text{prob}(X_1 \leq x) \cdot \ldots \cdot \text{prob}(X_n \leq x) = F^n(x) \]
Illustration

- Distributions of Maxima ($M_n$) of the exponential distribution
- Distributions of Maxima ($M_n$) of the normal distribution
- For large $n$ the distribution of $M_n$ converges to one shape.
- Asymptotic shapes for the normal and the exponential parent distributions are the same. (The Gumbel Distr.)
- Convergence is fast for exponential but slow for normal parent distribution.
Extremal Types Theorem

• If distribution of $M_n$ converges with large $n$ …
  
  o I.e. if there exist $a_n > 0$, $b_n$, and a non-degenerate distribution $G(x)$ such that:
  
  $$F^n \left( \frac{x - b_n}{a_n} \right) \xrightarrow{n \to \infty} G(x)$$

• … then the limit distribution $G(x)$ is one of …
  
  o The Gumbel distribution
  o The Fréchet distribution
  o The Weibull distribution

• … independent of the parent distribution $F(x)$. 
Extremal Types Distributions

- The Gumbel distribution (CDF)
  \[ G(x) = \exp(-\exp(-x)) \]

- The Fréchet distribution (CDF)
  \[ G(x) = \begin{cases} 
  0 & x \leq 0 \\
  \exp(-x^{-\alpha}) & x > 0, \alpha > 0 
\end{cases} \]

- The Weibull distribution (CDF)
  \[ G(x) = \begin{cases} 
  \exp(-(x)^{\alpha}) & x < 0, \alpha > 0 \\
  1 & x \geq 0 
\end{cases} \]
Asymptotic Distributions Laws

• **The Central Limit Theorem:**
  - The mean of a large number of 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).
Convergence?

- **In theory**
  - No general assurance of convergence but …
  - … convergence is warranted for continuous parent distributions under fairly general regularity assumptions (well-behaviour of the tail).

- **In practice**
  - It is justified to presume that data from natural phenomena have a parent distribution the maxima of which are converging.
  - Much more care is required with the assumption that the asymptotic limit is applicable:
    - slow convergence with some parent distributions
    - threshold processes may influence the tail
Sketch of Proof

- Max - Stability
  - A distribution function $G(x)$ is max-stable if there exist $a_k (>0)$, and $b_k$ so that:

  \[ G^k (x) = G \left( \frac{x - b_k}{a_k} \right), \quad k = 1, 2, 3, \ldots \]

- If $G(x)$ is a limit distribution then it must be max-stable:

  \[
  \begin{align*}
  X_{1,1} & \quad X_{1,2} & \cdots & \quad X_{1,n} & \quad \xrightarrow{n \to \infty} & \quad M_{(n),1} \approx G \\
  X_{2,1} & \quad X_{1,2} & \cdots & \quad X_{2,n} & \quad \xrightarrow{n \to \infty} & \quad M_{(n),2} \approx G \\
  \vdots & & \vdots \\
  X_{k,1} & \quad X_{k,2} & \cdots & \quad X_{k,n} & \quad \xrightarrow{n \to \infty} & \quad M_{(n),k} \approx G \\
  n \to \infty & & & & & \\
  M_{(n,k)} \approx G & \leftrightarrow & \quad M_{(n,k)} \approx G^k
  \end{align*}
  \]
Sketch of Proof

- **Gumbel, Fréchet and Weibull are max-stable.**
  - Can be verified with simple algebra.
  - E.g. Gumbel

\[
G^k(x) = \exp\left(-\exp(-x)\right)^k \\
= \exp(-k \cdot \exp(-x)) \\
= \exp(-\exp(-x + \log(k))) \\
= \exp(-\exp(-(x - \log(k)))) = G\left(x - \log(k)\right)
\]

- **There are no other max-stable distributions than Gumbel, Fréchet and Weibull.**
  - Proof is complex (function theory).
Generalized Extreme Value Distribution

- The GEV Distribution (CDF)

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

where: \( 1 + \xi \cdot \frac{x - \mu}{\sigma} > 0 \)

- A combined parametrisation of all three limit distributions
- Three parameters:
  - Location \( \mu \)
  - Scale \( \sigma \)
  - Shape \( \xi \)
- \( \xi = 0 \): Gumbel, unbounded
- \( \xi > 0 \): Fréchet, lower bound
- \( \xi < 0 \): Weibull, upper bound
Which one is “appropriate”?

• The GEV …
  o … takes a special role in statistics.
  o … is theoretically appropriate to modelling extremes (minima, maxima).

• Independent of the nature of original data
  o Wind gusts,
  o Precipitation,
  o Stock market changes,
  o etc.
History of Extreme Value Statistics

Ronald Aylmer Fisher & L.H.C. Tippett
First statement of extremal types theorem

Boris V. Gnedenko
1912-1995
Unification / extension of Extreme Value Theory

Emil Julius Gumbel
1891-1966
Statistical application of theory to estimate extremes

...
Section 4: Extreme Value Analysis – An Introduction

Modelling of Block Maxima
The Block Maxima Approach

Estimate $X(T)$ (for rare extremes) by parametric modelling of *Maxima* taken from large *blocks* of independent data.

Jenkinson 1955, Gumbel 1958
Procedure

• **Build Blocks**
  - Divide full dataset into equal sized chunks of data
  - E.g. yearly blocks of 365/366 daily precipitation measurements

• **Extract Block Maxima**
  - 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 Gumbel Diagramm \(X(T)\)

\(T\)-axis is transformed such that Gumbel-Distribution is a straight line.

\[
F(x) = GEV(x; \mu, \sigma, \xi) \quad \text{estimated CDF}
\]

\[
Y(x) = -\log(-\log(F(x))) \quad \text{Gumbel Variate}
\]

Horizontal axis is linear in \(Y\).

\[
T(x) = \frac{1}{1 - F(x)} \quad \text{Return period}
\]

\(x_k \quad k = 1, \ldots, N \quad \text{Block Maxima}
\]

\[
\tilde{T}_k = \frac{N + 1}{N + 1 - \text{rank}(x_k)} \quad \text{plotting points of block maxima } x_k
\]
The Gumbel Diagram

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.
Parameter Estimation

• Maximum Likelihood (ML) Estimation
  - Numerically maximize the Likelihood Function
  - Find $\mu$, $\sigma$, $\xi$ such that the probability of drawing $\{x_k, k=1,\ldots, N\}$ as random sample from the GEV would be largest.
  - Pro: Obtain approximate confidence intervals (e.g. Coles 2001).
  - Cons: Robustness for small $N$, numerical procedure

• L-Moments Estimation
  - Set $\mu$, $\sigma$, $\xi$ so that the first 3 sample L-Moments are equal to the first 3 L-Moments of the GEV.
  - Formuli for GEV L-moments from Hosking 1990
  - Pros: Robustness, analytical solution.
  - Cons: Best estimate only. Uncertainties (CIs) to be inferred separately.
Parameter Estimation (Example)

- **L-Moments**
  - $\mu = 53.2$ mm
  - $\sigma = 12.6$ mm
  - $\xi = 0.055$
  - $x(T=200) = 131$ mm

- **Max-Likelihood**
  - $\mu = 53.5$ mm
  - $\sigma = 12.5$ mm
  - $\xi = 0.038$
  - $x(T=200) = 127$ mm
Assumptions

• **Original data are identically distributed**
  - May be violated by seasonality, trends, …

• **Original data are independent**
  - May be violated by serial correlation

• **Asymptotic limit. Maxima from a large number of “originals”**
  - Need large blocks

• **Need enough independent episodes from the season with largest values!**
Assessing Goodness of Fit

• **QQ-plot**
  - Check if GEV can adequately describe the data
  - Too small block size and non-stationarities may manifest in inconsistencies

Engelberg 1-day totals 1901-2010
Uncertainty?

- How accurate are parameter estimates?
- How accurate are return values?
- How far is extrapolation justified?
Let’s „roll“ storms

Yearly maximum wind gusts in Zurich 1981-1999

$X_{100} = 183\, \text{km/h}$

$T(\text{Lothar}) = 16\, \text{a}$
Let’s „roll“ storms

\[ X_{100} = 205 \text{ km/h} \]
\[ X_{100} = 183 \text{ km/h} \]
\[ X_{100} = 165 \text{ km/h} \]
Let’s „roll“ storms

Parametric Resampling Confidence Intervals

$X_{100}$: 145 - 250 km/h

- Large sampling uncertainty.
- Uncertainty of $X(T)$ increases with $T$. 
Confidence Intervals

• **Resampling Confidence Intervals**
  - Simulate random samples (same size) and fit GEV
    - Parametric: random samples from estimated distribution.
    - Non-parametric: Draw samples (with replacement) from the original data.
  - Pros: Generally accurate, applicable to any estimation method
  - Cons: Computationally demanding.

• **Asymptotic Maximum Likelihood Confidence (Delta-Method)**
  - Likelihood-Theory: For large samples the sampling distribution of parameters is multivariate Normal. Variances/covariances are inferred from curvature of Likelihood surface.
  - Pros: fast, analytic
  - Cons: applicable to MLE only, not very accurate for small samples
Confidence Intervals

90% ML confidence

90% resampling confidence

Extreme 1-day Totals
Engelberg 1901–2010

mm
2 10 20 50 100 200

mm
2 10 20 50 100 200

symmetric
asymmetric
What goes wrong here?

Peaks over Threshold

ML-CI (Delta Method)

Index for “Storminess” in Europe.
ERA40, 1958-2002

Vivian, Feb 1990
Confidence Intervals

- **Likelihood-Profile Confidence**
  - Similar to asymptotic ML-CI. Exploits higher moments in the shape of the likelihood function.
  - Converges to asymptotic ML-CI for large samples.
  - Pros: more accurate for small samples, strong shapes, large \( T \) makes better use of information in available data, also accurate for negative shape.
  - Cons: only for ML estimates, computationally demanding.

Coles 2001
ML vs. Likelihood Profile CI

Peaks over Threshold

Vivian, Feb 1990

Index for “Storminess” in Europe.
ERA40, 1958-2002

Likelihood-Profile
ML (Delta Method)
Block Size: What is “large enough”?

- A trade-off between biases (too small block size) and large sampling errors (large blocks, small number of blocks).

- Desirable block size depends on parent distribution
  - Smaller for precipitation (close to exponential distribution)
  - Larger for temperature (close to Normal distribution)

- For the Exponential (Normal) parent distributions GEVs from block sizes >20 (>50) provide a reasonably bias free approximation for return periods between 5 and 100 times the block size.
Other Issues in Practice

• Trends violate iid assumption!
  o Solution: Parametrization of trends. GEV parameters vary smoothly with time (Katz et al. 2002).

• GEV does not make sense for seasonal means (e.g. summer 2003). Not a Maximum of a large block!

• In a network with many stations you will find events with very large “local” return periods every now and then. Citing those results only is misleading. (Don’t fall for sensation journalism!)
Example: August 2005

2005.08.21-22
RR = 205 mm
T > 500a

Extreme 2-Tages Summen
Meiringen
1901 – 2005

Jährlichkeit (Jahre) des 2-Tages Niederschl.: 2005.08.21-22

MeteoSwiss 2006
Section 4: Extreme Value Analysis – An Introduction

Modelling of Peaks over Threshold
The Peak-over-Threshold Approach

Estimate $X(T)$ (for rare extremes) by parametric modeling of independent exceedances above a large threshold.

Note: Hydrologists tend to call this the method of *Partial Duration Series*
Distribution of Exceedances

- Independent identically distributed random variables:
  \[ X_1, X_2, X_3, \ldots, X_N \sim F(x) \text{ iid} \]
  - E.g. daily maximum of wind speed over 30 years of observation
  - \( F(x) := \text{prob}(X_k \leq x) \), the parent distribution

- The Distribution of Exceedances
  - \( u \) a threshold (e.g. \( u=10 \text{ m/s} \), about Beaufort 6 or larger)
  - Exceedances: \( y=x-u \)
  - CDF of exceedances \( E_u(y) \):
    \[ E_u(y) := \text{prob}(X < u + y \mid X > u) \]
    \[ E_u(y) = \frac{F(x = u + y) - F(x = u)}{1 - F(x = u)} \]
Clarification

\[ \text{prob}(X > u + y \mid X > u) = 1 - E_u(y) \]

\[ \text{prob}(X > u + y \mid X > u) = \frac{1 - F(x = u + y)}{1 - F(x = u)} \]

\[ \Rightarrow E_u(y) = \frac{F(x = u + y) - F(x = u)}{1 - F(x = u)} \]
Asymptotic Theorem for Exceedances

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

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

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

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

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

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

See outline of proof in Coles 2001, p.76
Generalized Pareto Distribution

- The GPD (CDF)

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

\( y \geq 0, \quad 1 + \frac{\xi y}{\sigma} \geq 0 \)

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

Vilfredo Federico Damaso Pareto
1848-1923
Procedure

- **Select a threshold** \( u \)
  - Should be large enough to be in asymptotic limit (see later)

- **Extract the exceedances from the dataset**
  - \( n \) values out of the total \( N \) data values
  - Exceedances need to be mutually independent (see later)

- **Fit GPD to exceedances, yields conditional distribution:**
  \[
  \text{prob}(X > x \mid X > u) = 1 - GPD(x - u; \sigma, \xi)
  \]

- **Estimate unconditional distribution and return values**
  \[
  \text{prob}(X > x) = \text{prob}(X > u) \cdot (1 - GPD(x - u; \sigma, \xi))
  \]
  - with \( \text{prob}(X > u) \) estimated as \( n/N \) (the third model parameter)
  - Return values \( X(T) \) from the unconditional distribution
Estimation and CIs

• Exactly as with GEV

• Parameter Estimation
  o Maximum Likelihood Method (find maximum of Likelihood Function numerically)
  o Method of L-Moments (analytical formulæ, see Hosking 1990)

• Confidence Intervals
  o Resampling
  o Asymptotic ML confidence intervals (Delta Method)
  o Likelihood profile

Hosking 1990, Coles 2001
The Exceedance Diagram

Threshold: 30 mm

640 independent exceedances

Max-Likelihood
\( \sigma = 11.3 \text{ mm} \)
\( \xi = 0.07 \)
\( x(T=200) = 133 \text{ mm} \)

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

• **Exceedances are identically distributed**
  - May be violated e.g. by seasonality, by trends
  - May be resolved by stratification, explicit modeling

• **Exceedances are independent**
  - May be violated by serial correlation
  - Much more critical than for block maximum approach
  - In general resolved by *declustering* of original data
  - E.g.: Exceedances should be separated by at least \( m \) days (*runs-declustering*).

• **Sufficiently large threshold (asymptotic limit)**
  - Depends on the parent distribution (fast/slow convergence)
  - Use additional diagnostics for appropriate choice (see later)
Threshold Selection

- Parameter Dependence on Threshold
  - If exceedances of threshold \( u \) are GPD then exceedances of threshold \( v > u \) are also GPD with:

\[
GPD(y, \sigma_v, \xi_v) = \frac{GPD((v - u) + y, \sigma_u, \xi_u)}{GPD((v - u), \sigma_u, \xi_u)} \quad \text{for all } y > 0
\]

- Can only be satisfied if:

\[
\xi_v = \xi_u, \quad \sigma_v - \xi_v v = \sigma_u - \xi_u u
\]

At sufficiently large thresholds \( u \):
- (a) shape \( \xi \) is independent of threshold
- (b) modified scale \( (\sigma - \xi u) \) is independent of threshold
Threshold Selection

Diagnostics for threshold selection

Engelberg, daily precipitation (mm) 1901-2010

Be aware:
Stability within confidence limits does not warrant true stability.

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Threshold Selection

Mean Residual Life Plot:

- If exceedances are GPD

\[
\text{prob}(Y_u < y) = GPD(y, \sigma_u, \xi)
\]

- Then expected value of exceedances is:

\[
E(Y_u) = \frac{\sigma_u - \xi \cdot u}{1 + \xi}
\]

I.e. for sufficiently large thresholds, the mean of the exceedances depends linearly on threshold.
Block-Max vs. Peak-over-Thresh

**Block Maximum:**
GEV, 110 maxima

**Peak-over-Threshold:**
GPD, 640 peaks > 30 mm

---

**Engelberg 1901–2010**

- **Block Maximum:**
  - X(100) = 152 mm
  - 127 mm
  - 101 mm

- **Peak-over-Threshold:**
  - X(100) = 138 mm
  - 120 mm
  - 105 mm
Block-Max vs. Peak-over-Thresh

• **Block Maximum Approach**

  • **Pros**
    - Theoretical assumptions are less critical in practice.
    - Independence of maxima can be achieved by selecting large block size.
    - More easy to apply.

  • **Cons**
    - Estimation uncertainties can be large because sample size is small.

• **Peak-over-Threshold Approach**

  • **Pros**
    - Smaller CIs if a “small” threshold is justified. (More independent exceedances than block maxima.)

  • **Cons**
    - Independence assumption is critical in practice. Need declustering techniques.
    - Less easy to apply in practice.
Section 4: Extreme Value Analysis – An Introduction

Additional Remarks
Additional Remarks

• Precipitation Extremes
  o Dependence of return levels on duration is commonly described by an *Intensity-Duration-Frequency* Curves (also *Depth-Duration-Frequency*) using scaling relationships between durations.

![Log Precipitation Rate vs Log Measurement Interval](image)

Log Precipitation Rate (Intensity) vs Log Measurement Interval (Duration)

Koutsoyannis et al. 1998; Geiger, Zeller, Röthlisberger, 1990
Additional Remarks

- **Hydrological Extremes**
  - Hydrological processes involve threshold processes. Only very few events may have been observed from the extreme tail. Very large block size needed.
  - Statistics alone may be inadequate. Physical modelling of scenarios instead.

Hourly runoff extremes, Langeten

Liechti, 2008
Additional Remarks

- **Spatial Extremes**
  - Utilize concepts of spatial dependence or regional pooling of data to reduce uncertainties in estimation.
  - Combine extreme values analysis and spatial statistics to estimate extremes for unobserved locations.