Methods in climatology

II. Extreme Value Analysis

Motivation

The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe

David Barriopedro,[1]* Erich M. Fischer,[2] Jürg Luterbacher,[3] Ricardo M. Trigo,[1] Ricardo García-Herrera[4]

The summer of 2010 was exceptionally warm in eastern Europe and large parts of Russia. We provide evidence that the anomalous 2010 warmth that caused adverse impacts exceeded the amplitude and spatial extent of the previous hottest summer of 2003. "Mega-heatwaves" such as the 2003 and 2010 events likely broke the 500-year-long seasonal temperature records over approximately 50% of Europe. According to regional multi-model experiments, the probability of a summer experiencing mega-heatwaves will increase by a factor of 5 to 10 within the next 40 years. However, the magnitude of the 2010 event was so extreme that despite this increase, the likelihood of an analog over the same region remains fairly low until the second half of the 21st century.

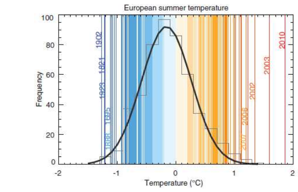
Increasing greenhouse gas concentrations are expected to amplify the variability of summer temperatures in Europe (1–5). Along with mean warming, enhanced variability results in more frequent, persistent, and intense heatwaves (6–10). Consistent with these expectations, Europe has experienced devastating heatwaves in recent years. The exceptional summer of 2003 (1, 11–13) caused around 70,000 heat-related deaths, mainly in western and central Europe (14). In summer 2010, many cities in eastern Europe recorded extremely high values of daytime (for example, Moscow reached 38.2°C), nighttime (Kiev reached 25°C), and daily mean (Helsinki reached 26.1°C) temperatures (fig. S1). Preliminary estimates for Russia referred a death toll of 55,000, an annual crop failure of ~25%, more than 1 million ha of burned areas, and ~US$15 billion (~1% gross domestic product) of total economic loss (15). During the same period, parts of eastern Asia also experienced extremely warm temperatures, and Pakistan was hit by devastating monsoon floods.

In order to characterize the magnitude and spatio-temporal evolution of the 2010 event in a historical context, we used daily mean data sets

[1]Instituto Dom Luiz, University of Lisbon, 1749-016 Lisbon, Portugal. [2]Institute for Atmospheric and Climate Science, Eidgenössische Technische Hochschule (ETH) Zurich, 8092 Zurich, Switzerland. [3]Department of Geography, Justus-Liebig-University of Giessen, D-35390 Giessen, Germany. [4]Agencia Estatal de Meteorología (AEMET), 28071 Madrid, Spain.

*To whom correspondence should be addressed. E-mail: dbarriopedro@fc.ul.pt

220 8 APRIL 2011 VOL 332 SCIENCE www.sciencemag.org

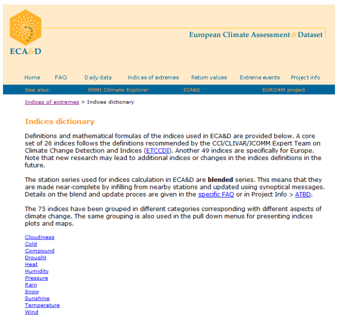


The flood in Moravia and Silesia in July 1997

52 victims - material damage 63 billions of Czech crowns

Terminology

1) Descriptive Extremity Indices http://www.ecad.eu/indicesextremes/



Examples:

R20mm

- *Very heavy precipitation days (precipitation ≥ 20 mm) (days)*

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the no of days where:

$RR_{ij} \geq 20\ mm$

R95p

- *Days with RR > 95th percentile of daily amounts (very wet days) (days)*

Let RR_{wj} be the daily precipitation amount at wet day w ($RR \geq 1.0$ mm) of period j and let $RR_{wn}95$ be the 95th percentile of precipitation at wet days in the 1961-1990 period. Then counted is the no of days where:

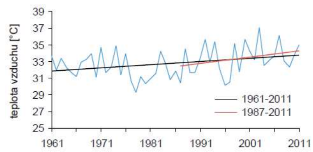
$RR_{wj} > RR_{wn}95$

Terminology

2) Extreme Value Analysis (EVA)

Analysis of frequency occurrence and intensity of rare events – „extremes" that occur with low probability

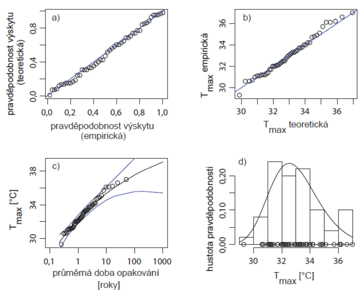
Analysis of annual maximum precipitation, air temperatures exceeding very high threshold, …



Variability of annual absolute maximum air temperatures at Brno, Tuřany (1961-2011), linear trends (19612-2011 and 1987-2011)

- **Extreme Value Theory (EVT)**
- **Extreme Value Distribution (EVD)**

EVA example



N [years]	2	5	10	20	50	100
T_{max} [°C]	32.8	34.4	35.3	36.1	36.9	37.5

Estimates of mean return periods N [years] of annual absolute maximum air temperatures at Brno, Tuřany (1961-2011),

---

Purpose

- 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 $X(T)$

$min\{x_1, … , x_n\} = - max\{x_1, … , x_n\}$

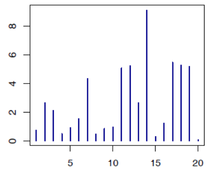
Main steps

- Choose an *appropriate* parametric distribution function
- Calibrate it such that it describes available data well
- Extrapolate distribution function

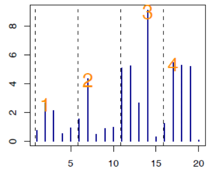
Which is „appropriate" distribution?

Extreme Value Theory (EVT)

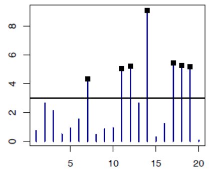
Input data



BM (Block Maxima)



POT (Peak Over Threshold)



Extreme Value Distributions (EVD)

Generalized Extreme Value distribution (GEV) - zobecněné rozdělení extrémních hodnot

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.

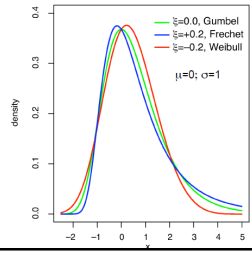
3 parametric distribution: location (μ), scale (σ), shape (ξ)

GEV cumulative density function

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

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

ξ =0: *Gumbel*, unbounded
ξ >0: *Fréchet*, lower bound
ξ <0: *Weibull*, upper bound



Modelling Block Maxima (BM)

- **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 fitting to the block maxima.
  - **Maximum Likelihood (ML) Estimation** – is prefered when i) samples are sufficiently large; ii) climate is not statinary. In this case LME may include „covariates"
  - **L-Moments Estimation** – when sampleas are small
  - **Method of moments** - undrestimate long-period return values
- **Calculate the return value function *X(T)* and its uncertainty**
  (confidence intervals)

Modelling Peak Over Threshold (POT)

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

For large *u* exceedances $E_u(y)$ asymptotes to a limit distribution:

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

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

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

(GPD = zobecněné Paretovo rozdělení)

Generalized Pareto distribution (GPD)

(zobecněné Paretovo rozdělení extrémních hodnot)

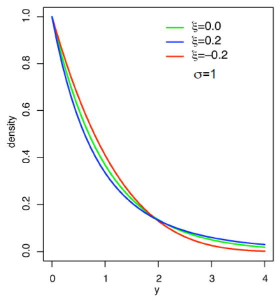
GPD has two parameters: scale (σ), shape (ξ)

- ξ =0: *Exponential Distribution*
- ξ < 0: upper bound at −σ/ξ
- ξ > 0: no upper bound

GPD cumulative density function:

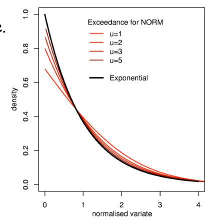
$$GPD(y;\sigma,\xi) = 1-\left(1+\xi\frac{y}{\sigma}\right)^{-1/\xi}$$

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

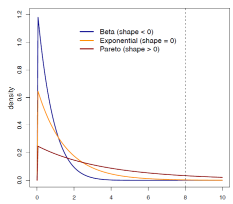


Generalized Pareto distribution (GPD)

- As *u* increases *Eu* converges to one shape.
- The limit distribution is the exponential distribution.



Distributions of Exceedance (density of *Eu*) for the standard *normal distribution* with different thresholds *u*.



Example of generalized Pareto (GP) probability density functions for each of the three types of tail behavior. Scale parameter varies to magnify differences in the tail behavior. Dashed vertical line shows upper bound of beta distribution.

Modelling Peak Over Threshold (POT)

- **Select a threshold** ***u***
  - should be large enough to be in asymptotic limit
- **Extract the exceedances from the dataset**
  - *n* values out of the total *N* data values
  - exceedances need to be mutually independent
- **Fit GPD to exceedances, yields conditional distr.:**
  - $\mathrm{prob}(X > x \mid X > u) = 1 - GPD(x - u; \sigma, \xi)$
- **Estimate uncond. distribution and return values**
  - $\mathrm{prob}(X > x) = \mathrm{prob}(X > u) \cdot (1 - GPD(x - u; \sigma, \xi))$
  - with prob(*X>u*) estimated as *n/N* (the third model parameter)
  - Return values *X*(*T*) from the unconditional distribution

Modelling Peak Over Threshold (POT)

Exceedances are identically distributed, may be violated e.g. by seasonality, by trends

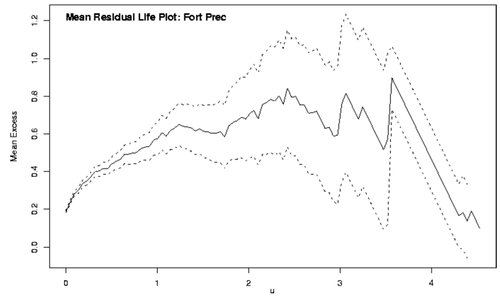
Exceedances are independent

- may be violated by serial correlation
- much more critical than for block maximum approach
- in general solved by *declustering* of original data
- e.g. exceedances should be separated by at least x days.

Modelling Peak Over Threshold (POT)

Threshold Selection

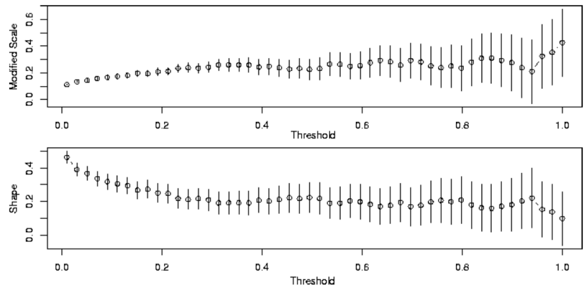
- **mean residual life plot** - the idea is to find the lowest threshold where the plot is nearly linear; taking into account the 95% confidence bounds.



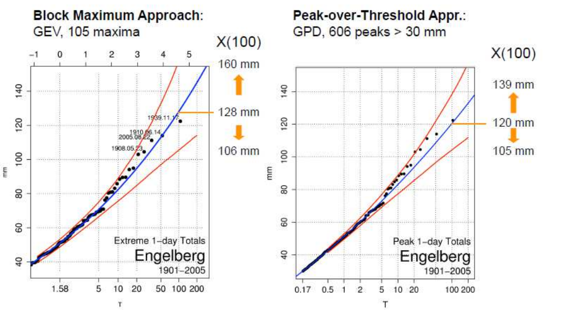
Modelling Peak Over Threshold (POT)

Threshold Selection

Fitting data to a GPD Over a Range of Thresholds and stability of the parameter estimates is checked



BM versus POT



Source: Analysis of Climate and Weather Data, Extreme Value Analysis - An Introduction, christoph.frei [at] meteoswiss.ch
ftp://ftp.pmodwrc.ch/pub/people/anna.shapiro/analisys%20of%20climate/Xstat%5B1%5D.pdf

BM versus POT

BM

- Theoretical assumptions are less critical in practice.
- Independence of maxima can be achieved by selecting large block size.
- More easy to apply
- Estimation uncertainties can be large because small sample size.

POT

- More efficient if a "small" threshold is justified. (More independent exceedances than block maxima.)
- Independence assumption is critical in practice. Need declustering techniques.
- Needs diagnostics for threshold selection. Choice somewhat ambiguous in practice.
- Less easy to apply in practice.

BM versus POT

- The POT approach typically utilizes more of the available data than the block maxima approach.
- However, it can be common for threshold excesses to cluster above a high threshold; especially with atmospheric data – consequently confidence intervals too narrow
- The block maxima approach may include points that are not very extreme
- In some cases it might miss extreme values simply because a larger value occurred somewhere else in the block (e.g., the second, or third, point that exceeds the threshold).
- The block maxima approach typically satisfies the independence assumption to a good approximation, and is easily interpretable in terms of return values.

General comments

- When analyzing extremes of atmospheric phenomena, one often encounters **non-stationarity** in the data (the df is not constant over time)
- The df for the extremes may have a gradual trend or shift over time; even abrupt changes
- The usual method for analyzing such data is to fit an EVD with parameters that vary as a function of a **covariate** (e.g., time is often used)

Validation and assesment of uncertainty

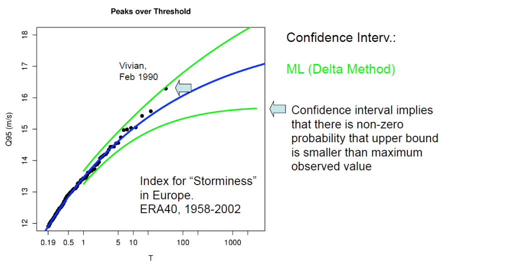
- Tests of goodness-of-fit
- Standard errors and confidence intervals
- Q-Q plot – comparison of observed and estimated quantiles
- All observed extremes must be feasible to occur under the fitted distribution
- Otherwise shape parameter should be adjusted
- Return level estimates for large T are prone to large sampling errors
- Confidence decreases rapidly when the period is more than about two times the length of the original data

General comments

Quality control and dealing with „outliers“

Fitted distribution may be very sensitive to the inclusion/exclusion of the outlier

- Inclusion – quality of the fit is reduced
- Exclusion – return periods are underestimated – not recommended approach



Source: Analysis of Climate and Weather Data, Extreme Value Analysis - An Introduction, christoph.frei [at] meteoswiss.ch ftp://ftp.pmodwrc.ch/pub/people/anna.shapiro/analisys%20of%20climate/Xstat%5B1%5D.pdf

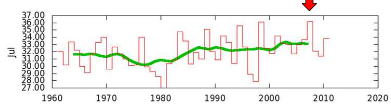
EVA tools

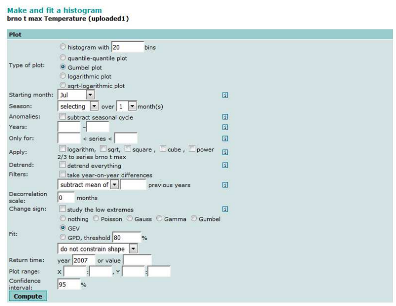
Climate Explorer

https://climexp.knmi.nl

Maximum July air temperatures, Brno, Tuřany, 1961 – 2010

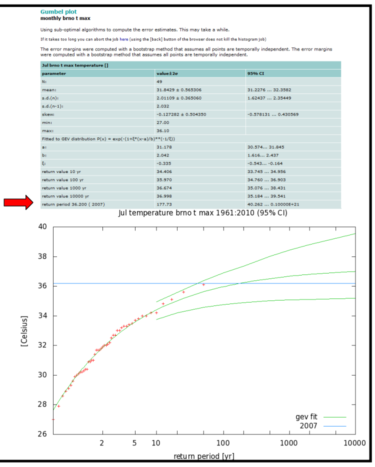
Abs. Max – 36.2°C (2007)





EVA tools

Climate Explorer



Gumbel plot
monthly brno t max

Using sub-optimal algorithms to compute the error estimates. This may take a while.

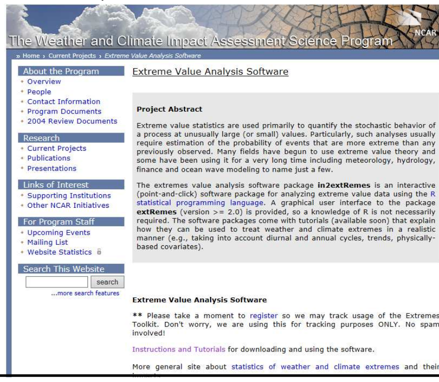
If it takes too long you can abort the job here (using the [back] button of the browser does not kill the histogram job)

The error margins were computed with a bootstrap method that assumes all points are temporally independent. The error margins were computed with a bootstrap method that assumes all points are temporally independent.

Jul brno t max temperature []		
parameter	**value±2σ**	**95% CI**
N:	49	
mean:	31.8429 ± 0.565306	31.2276 ... 32.3582
s.d.(n):	2.01109 ± 0.365060	1.62437 ... 2.35449
s.d.(n-1):	2.032	
skew:	-0.127282 ± 0.504350	-0.578131 ... 0.430569
min:	27.00	
max:	36.10	
Fitted to GEV distribution P(x) = exp(-(1+ξ*(x-a)/b)**(-1/ξ))		
a:	31.178	30.574... 31.845
b:	2.042	1.616... 2.437
ξ:	-0.335	-0.543... -0.164
return value 10 yr	34.406	33.745 ... 34.956
return value 100 yr	35.970	34.760 ... 36.903
return value 1000 yr	36.674	35.076 ... 38.431
return value 10000 yr	36.998	35.184 ... 39.541
return period 36.200 (2007)	177.73	40.262 ... 0.10000E+21

EVA tools

in2extRemes **http://www.assessment.ucar.edu/toolkit/**



The Weather and Climate Impact Assessment Science Program

NCAR

» Home › Current Projects › *Extreme Value Analysis Software*

About the Program
- Overview
- People
- Contact Information
- Program Documents
- 2004 Review Documents

Research
- Current Projects
- Publications
- Presentations

Links of Interest
- Supporting Institutions
- Other NCAR Initiatives

For Program Staff
- Upcoming Events
- Mailing List
- Website Statistics

Search This Website

search

...more search features

Extreme Value Analysis Software

Project Abstract

Extreme value statistics are used primarily to quantify the stochastic behavior of a process at unusually large (or small) values. Particularly, such analyses usually require estimation of the probability of events that are more extreme than any previously observed. Many fields have begun to use extreme value theory and some have been using it for a very long time including meteorology, hydrology, finance and ocean wave modeling to name just a few.

The extremes value analysis software package **in2extRemes** is an interactive (point-and-click) software package for analyzing extreme value data using the R statistical programming language. A graphical user interface to the package **extRemes** (version >= 2.0) is provided, so a knowledge of R is not necessarily required. The software packages come with tutorials (available soon) that explain how they can be used to treat weather and climate extremes in a realistic manner (e.g., taking into account diurnal and annual cycles, trends, physically-based covariates).

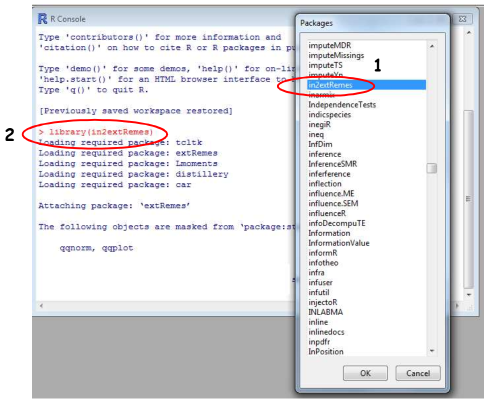
Extreme Value Analysis Software

** Please take a moment to register so we may track usage of the Extremes Toolkit. Don't worry, we are using this for tracking purposes ONLY. No spam involved!

Instructions and Tutorials for downloading and using the software.

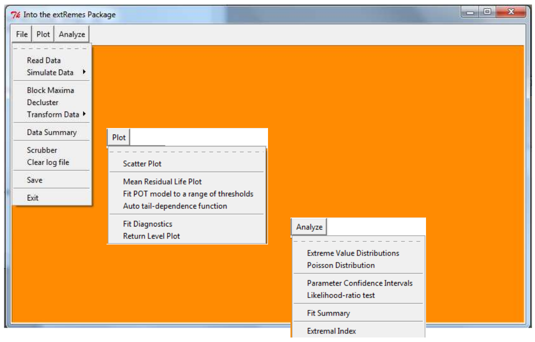
More general site about statistics of weather and climate extremes and their

EVA tools in2extRemes

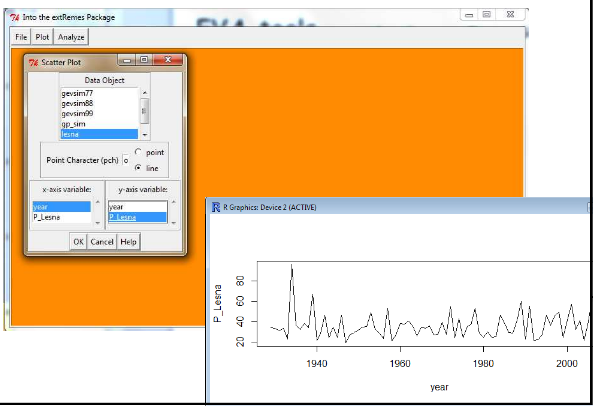


EVA tools in2extRemes – GUI

> in2extRemes()

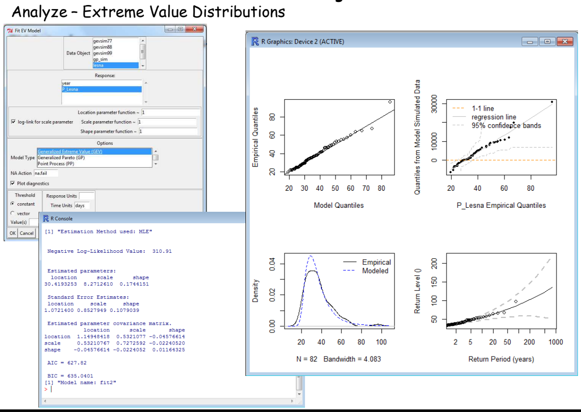


EVA tools in2extRemes – plot data



EVA tools in2extRemes – fitting GEV to data

Analyze – Extreme Value Distributions



EVA tools in2extRemes – estimate N return values (N=100)

Analyze – Parameter Confidence Intervals