Extreme Value Analysis
(or how to go beyond the data range)

Philippe Naveau (naveau@lsce.ipsl.fr)
Laboratoire des Sciences du Climat et l'Environnement (LSCE-CNRS)
Gif-sur-Yvette, France

What is an extreme ?

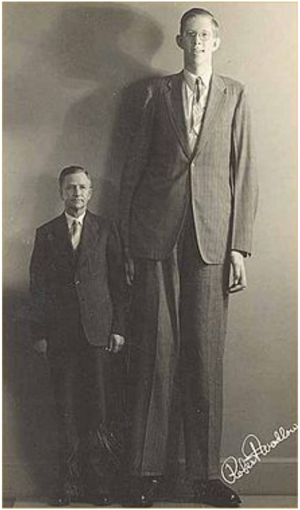
The tallest man living is Sultan Kosen (Turkey, December 1982) who measured 246.5 cm

What is an extreme ?

The tallest man living is Sultan Kosen (Turkey, December 1982) who measured 246.5 cm

"Man can believe the impossible, but man can never believe the improbable"

Oscar Wilde, (Intentions, 1891)



Robert Wadlow (1918-1940, 2.72 m, 8 ft 11.1 in) & his father, H.F. Wadlow (1.82 m, 5 ft 11 1 ?2 in)

NEWS: BOULDER FLOOD

Eight days, 1,000-year rain, 100-year flood

The story of Boulder County's Flood of 2013

By Charlie Brennan and John Aguilar, Camera Staff Writers

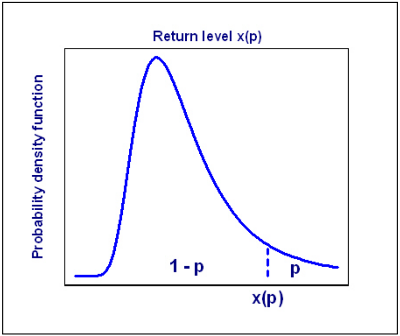
POSTED: 09/21/2013 07:49:24 PM MDT | UPDATED: 3 YEARS AGO

Quantiles and return levels

A **return level** with a **return period** of $T = 1/p$ years is a high threshold z_p whose probability of exceedance is p. E.g., p = 0.01 $\Rightarrow$ T = 100 years.

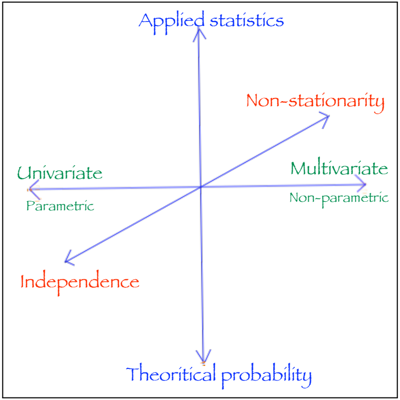
Return levels

- Number of events : Average number of events occurring within a T-year time period is one



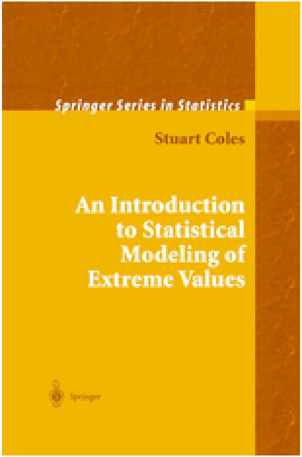
Important facts about Extreme Value Theory

- An asymptotic probabilistic concept
- A statistical approach for extrapolation of quantiles
- Clear assumptions
- Assessment of uncertainties
- Possible couplings with physical numerical models



An active research statistical field

A good intro



A few sessions at EVA in 2019

Extremes and machine learning - organizer Anne Sabourin (*LTCI, Télécom ParisTech)*

- Dan Cooley (*Colorado State University*)
- Laurent Gardes (*IRMA, Université de Strasbourg*)
- Vincent Feuillard (*Airbus Group Innovations*)

Risk analysis in insurance - organizer Liang Peng (*Georgia State University*)

- Claudia Kluppelberg (*Technische Universität München*)
- Jan Beirlant (*KU Leuven*)
- Fan Yang (*University of Waterloo*)

Spatial extremes - organizer Clement Dombry (*Université de Franche-Comté*)

- Ana Ferreira (*Universidade de Lisboa*)
- Marco Oesting (*Universität Siegen*)
- Raphaël de Fondeville (*EPFL*)

Detection and attribution of climate change - organizer Dan Cooley (*Colorado State University*)

- Anna Kiriliouk (*Université de Namur*)
- Richard Smith (*University of North Carolina*)
- Alexis Hannart (*Ouranos*)

Sub-asymptotic spatial extremes - organizer Thomas Opitz (*INRA Avignon*)

- Raphaël Huser (*KAUST*)
- Gwladys Toulemonde (*Montpellier University*)
- Jenny Wadsworth (*Lancaster University*)

Mixtures of dependence types - organizer Jennifer Wadsworth (*Lancaster University*)

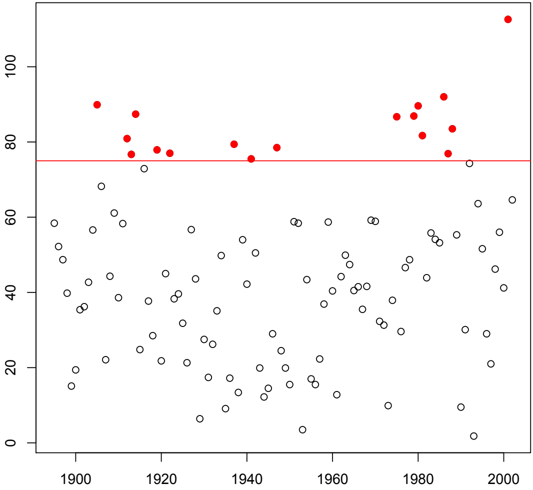
- Anne Sabourin (*Telecom Paris-Tech*)
- Emma Simpson (*Lancaster University*)
- Chen Zhou (*Erasmus University Rotterdam, De Nederlandsche Bank*)

A very short biblio about univariate EVT

- Davison, A. C. and Huser, R. G. (2015) Statistics of extremes. Annual Review of Statistics and Its Application, 2, 203-235.
- Katz, R. W., Parlange, M. B. and Naveau, P. (2002) Statistics of extremes in hydrology. Advances in Water Resources, 25, 1287-1304.
- Stephenson A. Gilleland E. (2006) Software for the analysis of extreme events :The current state and future directions, Extremes,

EVT = Going beyond the data range

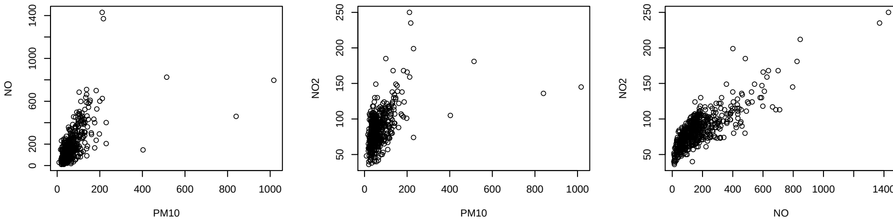
What is the probability of observing data above an high threshold ?



March precipitation amounts recorded at Lille (France) from 1895 to 2002. The 17 black dots corresponds to the number of excesses above the threshold $u_n = 75$ mm. This number can be conceptually viewed as a random sum of Bernoulli (binary) events.

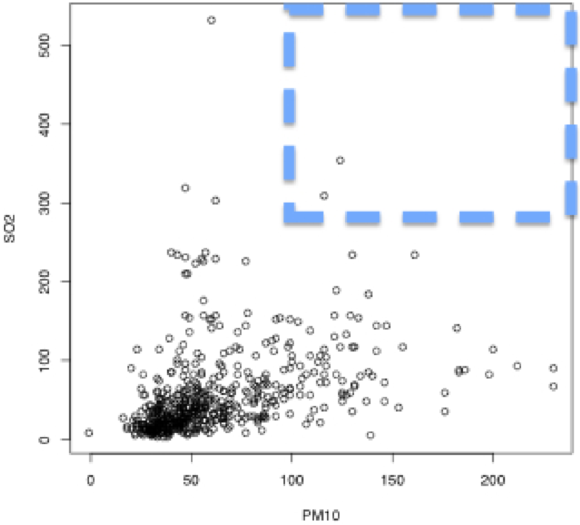
An example in three dimensions

Air pollutants (Leeds, UK, winter 94-98, daily max) NO vs. PM10 (left), SO2 vs. PM10 (center), and SO2 vs. NO (right) (Heffernan& Tawn 2004, Boldi & Davison, 2007)

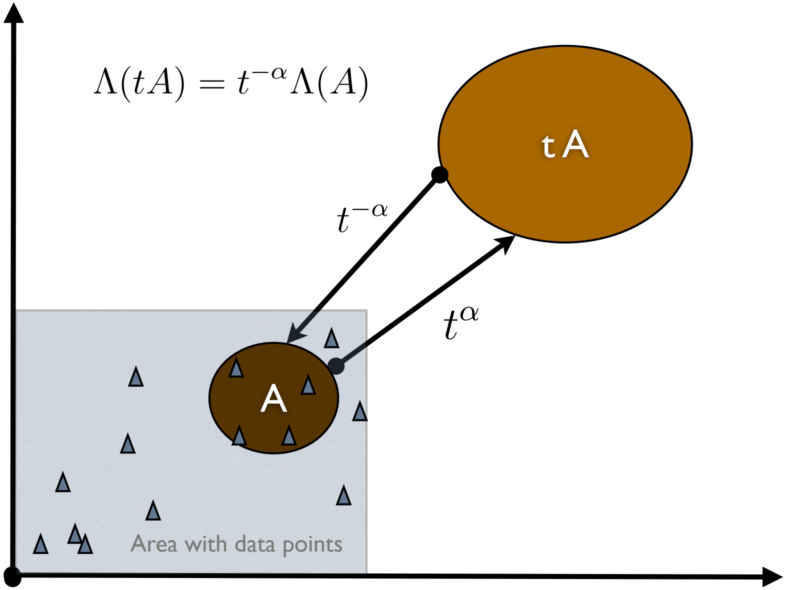


Typical question in multivariate EVT

What is the probability of observing data in the blue box ?



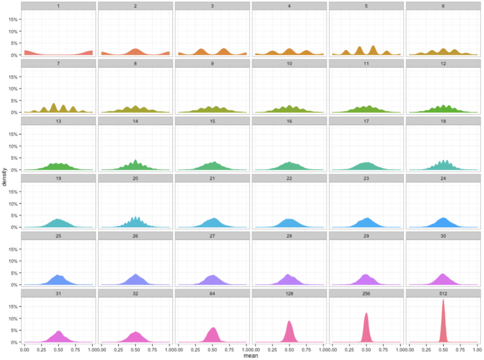
Give me a lift



EVT : It is a kind of Magic

Asymptotic theory

Pierre-Simon Laplace (1749-1827) + De Moivre (1667-1754)

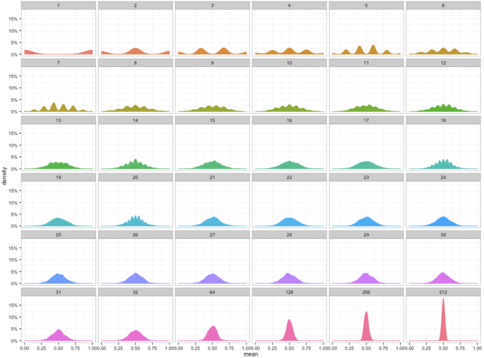


Central limit theorem [1]

The sum of independent random variables, properly normalized, tends toward a normal distribution even if the original variables themselves are not normally distributed (with finite variance).

1. a CLT proof of a result similar to the 1922 Lindeberg CLT was the subject of Alan Turing's 1934 Fellowship Dissertation. Only after submitting the work did Turing learn it had already been proved. Consequently, Turing's dissertation was not published.

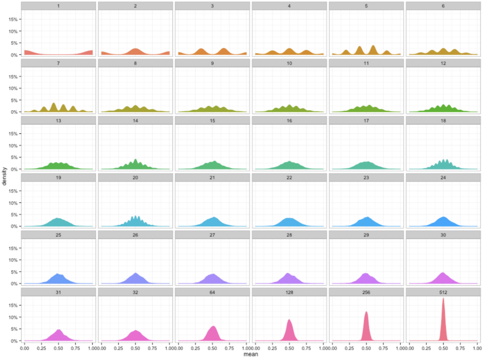
Pierre-Simon Laplace (1749-1827) + De Moivre (1667-1754)



Central limit theorem

$X_1 + \cdots + X_n \sim$ Gaussian with mean $E(X)$ and variance $var(X)/n$, for large n

Pierre-Simon Laplace (1749-1827) + De Moivre (1667-1754)



Central limit theorem

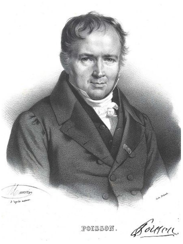
$X_1 + \cdots + X_n \sim$ Gaussian with mean $E(X)$ and variance $var(X)/n$, for large n

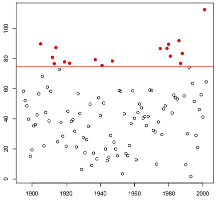
Why a Gaussian limit ?

The normal distribution is **sum-stable** and behaves like an attractor.

“Everyone wants to be **normal**, but no one wants to be **average**”

Simon Denis Poisson (1781-1840)





Counting excesses

As a sum of random binary events, the variable N_n that counts the number of events above the threshold u_n has mean $n\,Pr(X > u_n)$

Poisson's theorem [2] in 1837

If u_n such that

$$\lim_{n\to\infty} n\,Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n follows approximately a **Poisson variable** N.

2. Give HW

Simon Denis Poisson (1781-1840)

Poisson's theorem in 1837

If u_n such that

$$\lim_{n\to\infty} n\, Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n, the number of events above the threshold u_n, follows approximately a **Poisson variable** N.

Simon Denis Poisson (1781-1840)

Poisson's theorem in 1837

If u_n such that

$$\lim_{n \to \infty} n\, Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n, the number of events above the threshold u_n, follows approximately a **Poisson variable** N.

Question 1

Let a_n such that $n\, Pr(X > a_n) = 1$. What is your interpretation of a_n?

Simon Denis Poisson (1781-1840)

Poisson's theorem in 1837

If u_n such that

$$\lim_{n\to\infty} n\,Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n, the number of events above the threshold u_n, follows approximately a **Poisson variable** N.

Question 1

Let a_n such that $n\,Pr(X > a_n) = 1$. What is your interpretation of a_n ?

Question 2

Suppose $u_n > a_n$ such that $n\,Pr(X > u_n) = \lambda \in (0, \infty)$. What is $Pr(X > u_n | X > a_n)$? your interpretation ?

Simon Denis Poisson (1781-1840)

Poisson's theorem in 1837

If u_n such that

$$\lim_{n\to\infty} n\, Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n, the number of events above the threshold u_n, follows approximately a **Poisson variable** N.

Question 1

Let a_n such that $n\, Pr(X > a_n) = 1$. What is your interpretation of a_n ?

Question 2

Suppose $u_n > a_n$ such that $n\, Pr(X > u_n) = \lambda \in (0, \infty)$. What is $Pr(X > u_n | X > a_n)$? your interpretation ?

Question 3 : a special case

Suppose X follows an exponential distribution. Find $u_n(x)$ and a_n such that

$$Pr(X > u_n(x) | X > a_n) = Pr(X > x).$$

How to interpret this result ?

Simon Denis Poisson (1781-1840)

Poisson's theorem in 1837

If u_n such that

$$\lim_{n\to\infty} n\,Pr(X > u_n) = \lambda \in (0, \infty).$$

then N_n, the number of events above the threshold u_n, follows approximately a **Poisson variable** N.

Question 1

Let a_n such that $n\,Pr(X > a_n) = 1$. What is your interpretation of a_n ?

Question 2

Suppose $u_n > a_n$ such that $n\,Pr(X > u_n) = \lambda \in (0, \infty)$. What is $Pr(X > u_n | X > a_n)$? your interpretation ?

Question 3 : a special case

Suppose X follows an exponential distribution. Find $u_n(x)$ and a_n such that

$$Pr(X > u_n(x) | X > a_n) = Pr(X > x).$$

How to interpret this result ?

Poisson and maxima

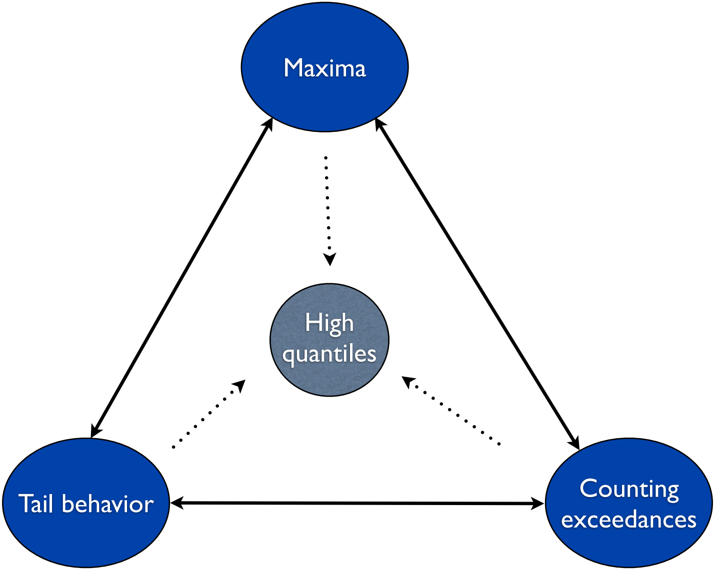
Counting = max

$$Pr(M_n \leq u_n) = Pr(N_n = 0) \text{ with } M_n = \max(X_1, \ldots, X_n)$$

Poisson's at work

$$\lim_{n \to \infty} Pr(M_n \leq u_n) = \lim_{n \to \infty} Pr(N_n = 0) = Pr(N = 0) = \exp(-\lambda)$$

Different modeling options



The fundamental Fisher-Tippet theorem (1928)

Suppose that there exists $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} Pr(M_n \leq a_n x + b_n) \text{ is a non-degenerate distribution}$$

then this limit has to be equal to

$$\text{GEV}(x) = \exp\left\{-\overline{H}(x; \mu, \sigma, \xi)\right\}$$

where

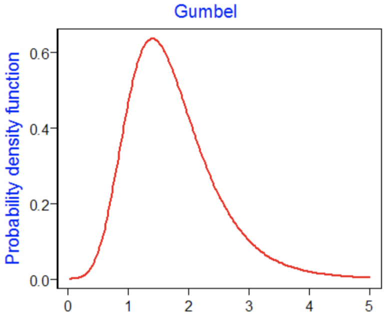
$$\overline{H}(x; \mu, \sigma, \xi) = \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}$$

Gumbel type

(i) $\xi = 0$ (*Gumbel type*, limit as $\xi \to 0$)

"Light" upper tail

"Domain of attraction" for many common distributions (e. g., normal, exponential, gamma)



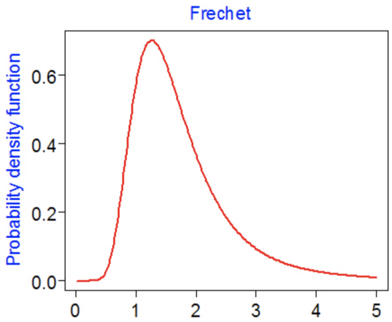
Fréchet (Pareto) type

(ii) $\xi > 0$ (*Fréchet type*)

"Heavy" upper tail with infinite *r*th-order moment if $r \geq 1/\xi$

(e. g., infinite variance if $\xi \geq 1/2$)

Fits precipitation, streamflow, economic damage

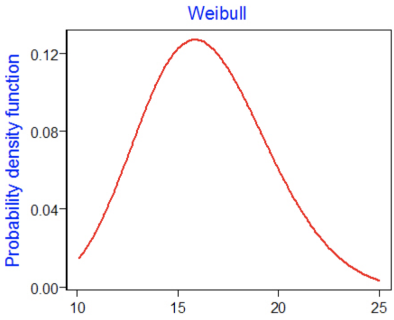


Weibull type

(iii) $\xi < 0$ (*Weibull type*)

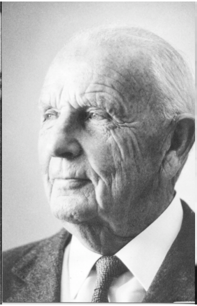
Bounded upper tail [$x < \mu + \sigma / (-\xi)$]

Fits temperature, wind speed, sea level



Extreme Value Theory : Historical perspective

Gumbel (1891-1966)



Weibull (1887-1979)

Fréchet (1878-1973)

- Emil Gumbel was born and trained as a statistician in Germany, forced to move to France and then the U.S. because of his pacifist and socialist views.
- Waloddi Weibull was a Swedish engineer famous for his pioneering work on reliability, providing a statistical treatment of fatigue, strength, and lifetime.
- Maurice Frechet was a French mathematician who made major contributions to pure mathematics as well as probability and statistics.

Other important names : Fisher and Tippet (1928), Gnedenko (1943), see Rick Katz's website

Annual Review of Statistics and Its Application

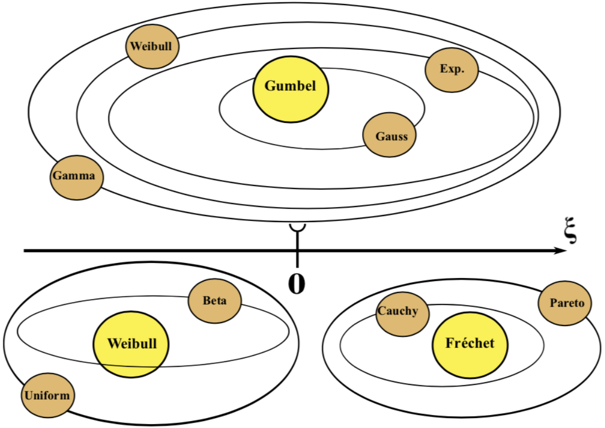
Statistics of Extremes: Animation 1

An animation from the 2015 review by A.C. Davison and R. Huser, "Statistics of Extremes," from the *Annual Review of Statistics and Its Application*.

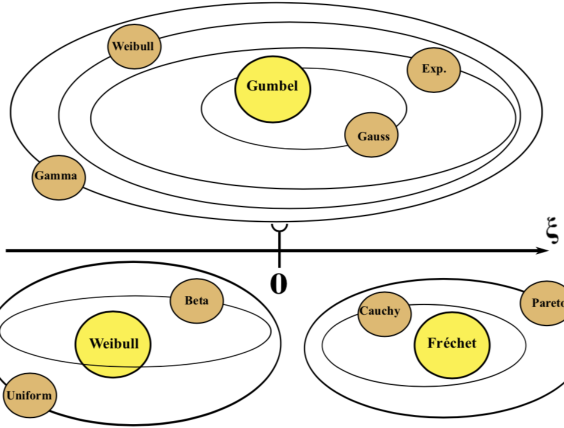
Illustration of the Extremal Types Theorem. For increasing values of n, the left panels display the distribution of the maximum Z_n of n independent uniform (*top row*), standard Gaussian (*second row*), unit exponential (*third row*), and 0.2-Pareto (*bottom row*), i.e., $F(y) = 1 - y^{-0.2}$, $y > 1$, random variables. The right panels display the distribution of $a_n^{-1}(Z_n - b_n)$ for appropriate sequences $a_n > 0$ and b_n. In each case, Z_n is asymptotically degenerate, whereas $a_n^{-1}(Z_n - b_n)$ is not.

View on YouTube | Read Associated Article

A summary of the 1D EVT galaxy



A summary of the 1D EVT galaxy



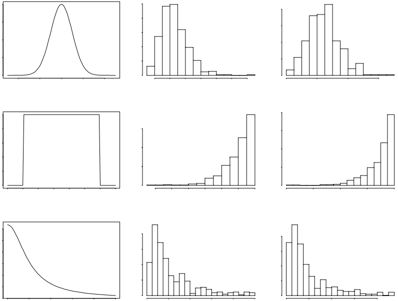
Can you find the misplaced planet ?

Maxima Distribution

Normal density $\Rightarrow$ $\Leftarrow$ Gumbel density

Uniform density $\Rightarrow$ $\Leftarrow$ Weibull density

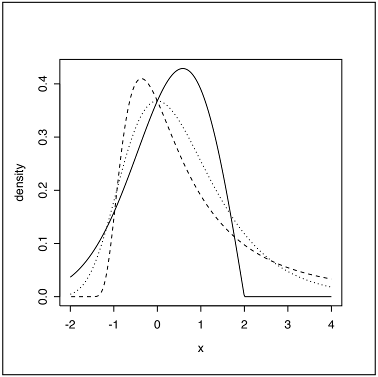
Cauchy density $\Rightarrow$ $\Leftarrow$ Fréchet density



$n = 50$ $n = 100$

Generalized Extreme Value (GEV) distribution

$$\mathrm{pr}\left(\frac{M_n - a_n}{b_n} < x\right) \sim \mathrm{GEV}(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}$$



Home work : show that a GEV is max-stable

A tough example : maxima of normally distributed random variables

Home work simulation

- Generated random sample of length 100 from standard normal distribution and obtain maximum value (Repeated 40,000 times)
- Fit GEV distribution to sample of 40000 maxima
- Check if the estimate of ξ is around -0.1 but not zero

Penultimate approximation

Theory by Fisher and Tippett (1928)

For infinite block sizes, maxima of Gaussian belong to the Gumbel galaxy, but for finite samples sizes the estimated shape parameters belong to the Weibull galaxy (Von Mises conditions)

$$\xi_n = \frac{1}{hazard'(a_n)} \text{ with } hazard(x) = \frac{f(x)}{\overline{F}(x)} \text{ and } \overline{F}(x) = 1 - F(x)$$

Penultimate approximation

Theory by Fisher and Tippett (1928)

For infinite block sizes, maxima of Gaussian belong to the Gumbel galaxy, but for finite samples sizes the estimated shape parameters belong to the Weibull galaxy (Von Mises conditions)

$$\xi_n = \frac{1}{hazard'(a_n)} \text{ with } hazard(x) = \frac{f(x)}{\overline{F}(x)} \text{ and } \overline{F}(x) = 1 - F(x)$$

Back to the Gaussian example

$$hazard(x) \approx x \text{ for large } x \text{ and } a_n \approx \sqrt{2 \log n}$$

Hence,

$$\xi_n \approx \frac{1}{2 \log n} \text{ and } \frac{1}{2 \log 100} \approx -0.109$$

Lessons learned from this Gaussian example

- The convergence towards max-stability can be very very slow
- The sign of the estimated shape parameter from a finite block size should not be over-interpreted
- Theoretical results exist to explain and quantity this phenomenon

Lessons learned from this Gaussian example

- The convergence towards max-stability can be very very slow
- The sign of the estimated shape parameter from a finite block size should not be over-interpreted
- Theoretical results exist to explain and quantity this phenomenon

Practical considerations

- Some light-tailed atmospheric variables like temperatures are often averaged in space, time or both. The central limit theorem makes them very close to Gaussian variables. So, don't be surprised to find negative shape parameters for maxima of averaged values

GEV and return levels

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

Computing the return level z_p such that GEV$(z_p) = 1 - p$

$$z_p = \mathrm{GEV}^{-1}(1-p)$$

Hence, $\boxed{z_p = \mu + \frac{\sigma}{\xi}\left([-\ln(1-p)]^{-\xi} - 1\right)}$

GEV and return levels estimation

$$z_p = \mu + \frac{\sigma}{\xi}\left([-\ln(1-p)]^{-\xi} - 1]\right)$$

Estimating the return level z_p

$$\boxed{\hat{z}_p = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left([-\ln(1-p)]^{-\hat{\xi}} - 1]\right)}$$

Estimating the GEV parameters

- Maximum likelihood estimation
- Methods of moments type (PWM and GPWM, *Ribereau et al., 2010*)
- Exhaustive tail-index approaches

An example from Rick Katz

Data at hand

- **Fort Collins daily precipitation amount**

-- Time series of daily precipitation amount (in), 1900-1999

Annual cycle in precipitation

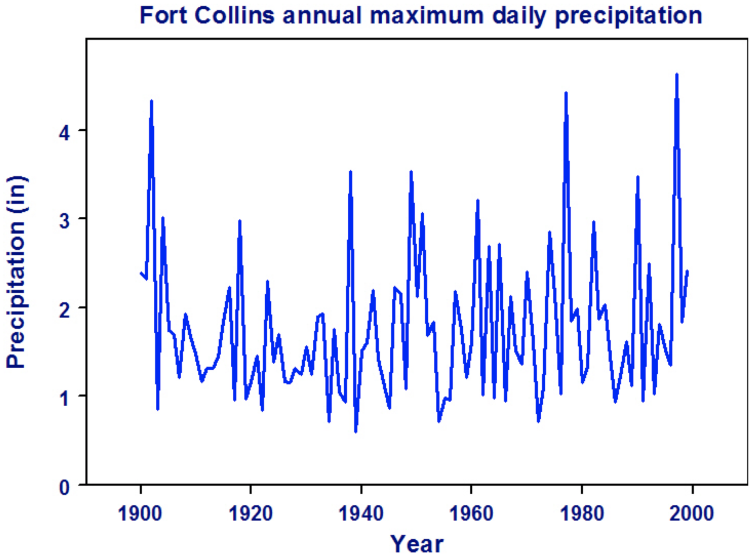
(Peak in mean in May)

Consider annual maxima (block size $n \approx 365$)

No obvious long-term trend in annual maxima ($T = 100$)

Flood on 28 July 1997

(Damaged campus of Colorado State University)

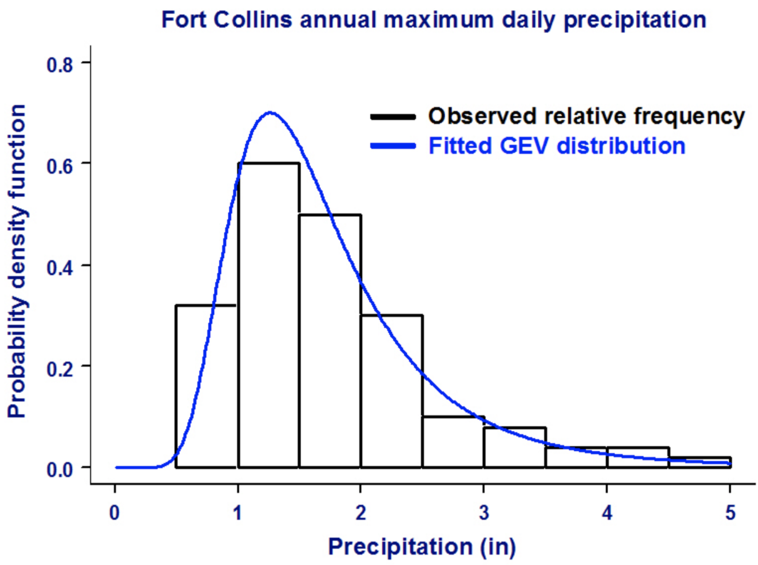


Fort Collins annual maximum daily precipitation
Precipitation (in)
0
1
2
3
4
1900
1920
1940
1960
1980
2000
Year

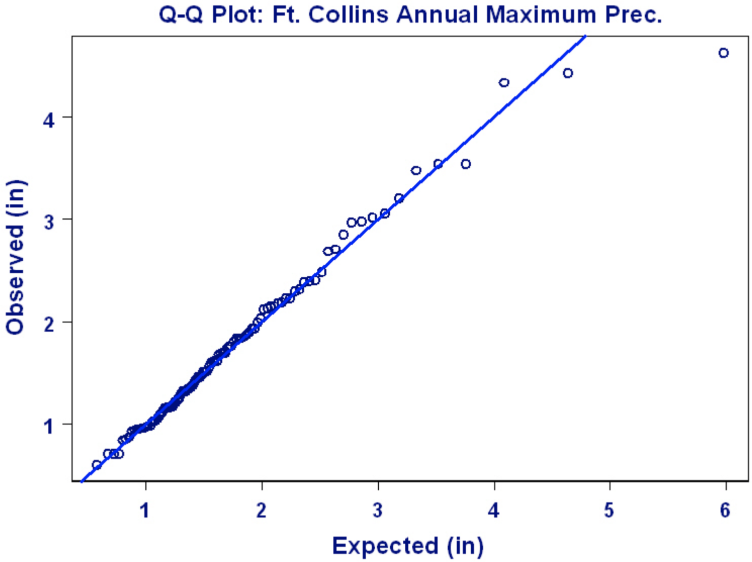
Maximum likelihood estimates

Parameter	Estimate	(Std. Error)
Location μ	1.347 in.	(0.062)
Scale σ	0.533 in.	(0.049)
Shape ξ	0.174	(0.092)

Advice : Always checked the literature on the same type of analysis



Fort Collins annual maximum daily precipitation
Observed relative frequency
Fitted GEV distribution
Probability density function
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
Precipitation (in)


Q-Q Plot: Ft. Collins Annual Maximum Prec.
Observed (in)
Expected (in)
1
2
3
4
5
6

-- **Likelihood ratio test (LRT) for comparing two models**

(e. g., test whether shape parameter $\xi = 0$)

Fit two models:

(i) GEV distribution with $\xi \neq 0$

$-\ln L(x_1, x_2, \ldots, x_T; \mu, \sigma, \xi)$ minimized with respect to μ, σ, ξ

(ii) GEV distribution with $\xi = 0$ (i. e., Gumbel distribution)

$-\ln L(x_1, x_2, \ldots, x_T; \mu, \sigma, \xi = 0)$ minimized with respect to μ, σ

If $\xi = 0$, then 2 [(ii) – (i)] has approximate chi square distribution with 1 degree of freedom (df) for large T

-- Confidence interval

(e. g., for shape parameter ξ)

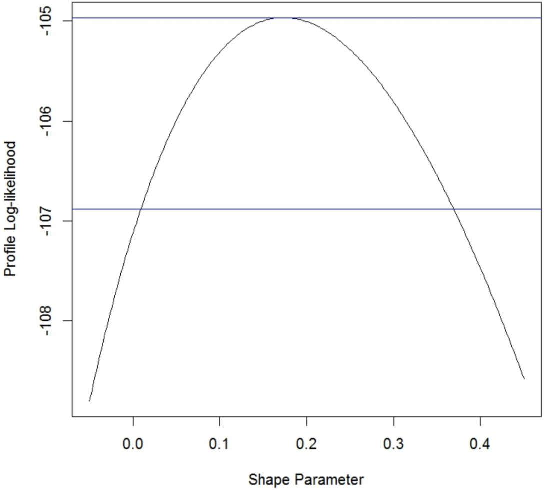
Profile likelihood method:

Minimize $-\ln L(x_1, x_2, \ldots, x_T; \mu, \sigma, \xi)$ with respect to μ, σ as function of ξ

Use chi square distribution with 1 df to determine lower and upper bounds of confidence interval

Alternate technique:

Resampling (parametric or nonparametric bootstrap)



Fort Collins Annual Maximum Daily Precipitation

-- Likelihood ratio test (LRT) for $\xi = 0$

Obtain P-value ≈ 0.038

-- 95% confidence interval for shape parameter ξ

Based on profile likelihood technique, obtain

$$0.009 < \xi < 0.369$$

(Consistent with P-value < 0.05)

Annual block maxima modeling with a GEV

Advantages

- Easy to define an extreme
- Do not need to explicitly model seasonal or diurnal cycles
- Do not need to explicitly model temporal dependence
- Incorporating covariates is simple (e.g., gam for GEV)
- Easy to communicate with stakeholders

Annual block maxima modeling with a GEV

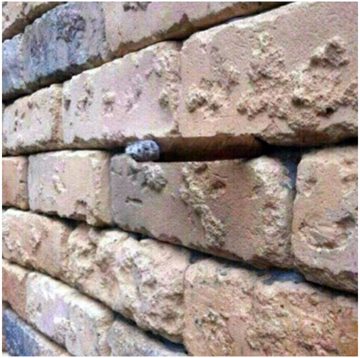
Advantages

- Easy to define an extreme
- Do not need to explicitly model seasonal or diurnal cycles
- Do not need to explicitly model temporal dependence
- Incorporating covariates is simple (e.g., gam for GEV)
- Easy to communicate with stakeholders

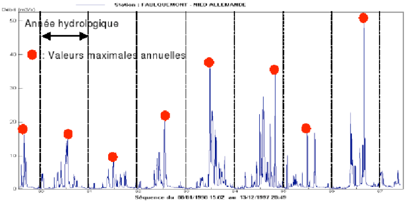
Drawbacks

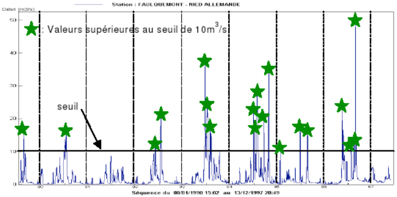
- The block size is fixed
- Only one observation is kept within a block
- Difficult to explain multivariate block maxima

A different point of view : what is the hidden object in this picture ?



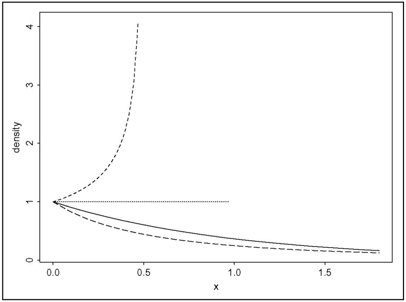
Modeling exceedances





The Generalized Pareto Distribution (GPD)

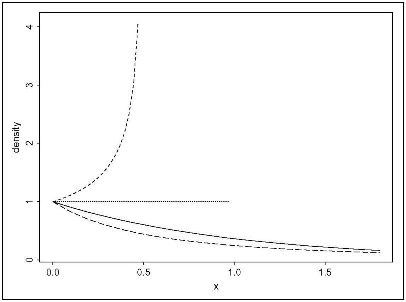
$$\text{pr}\{\mathbf{X} - u > y | \mathbf{X} > u\} = \left(1 + \frac{\xi\, y}{\sigma_u}\right)_+^{-1/\xi}$$



Born in France and trained as an engineer in Italy, Vilfredo Pareto (1848-1923) formulated the power-law distribution (or "Pareto's Law"), as a model for how income or wealth is distributed across society.

The Generalized Pareto Distribution (GPD)

$$\mathrm{pr}\{\mathbf{X} - u > y | \mathbf{X} > u\} = \left(1 + \frac{\xi\, y}{\sigma_u}\right)_+^{-1/\xi}$$



Born in France and trained as an engineer in Italy, Vilfredo Pareto (1848-1923) formulated the power-law distribution (or ”Pareto’s Law”), as a model for how income or wealth is distributed across society.

Home work

Show that the GPD is threshold invariant

Exceedance intensities modelling

The survival

$$P(X > x) = P(X > x | X > u) \times P(X > u)$$

can be approximated by (for u large and $x > u$)

$$\overline{H}_\xi \left(\frac{x - u}{\sigma} \right) \times P(X > u).$$

Exceedance intensities modelling

The survival

$$P(X > x) = P(X > x | X > u) \times P(X > u)$$

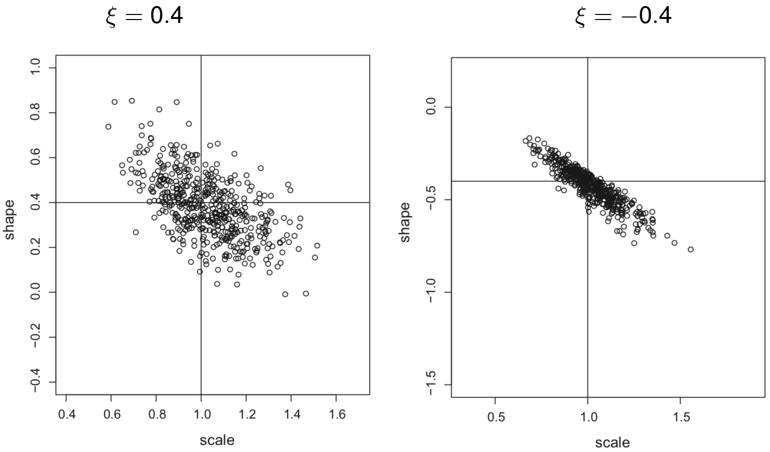
can be approximated by (for u large and $x > u$)

$$\overline{H}_{\xi}\left(\frac{x-u}{\sigma}\right) \times P(X > u).$$

After fitting a GDP to the exceedances above u, $P(X > x)$ is estimated by

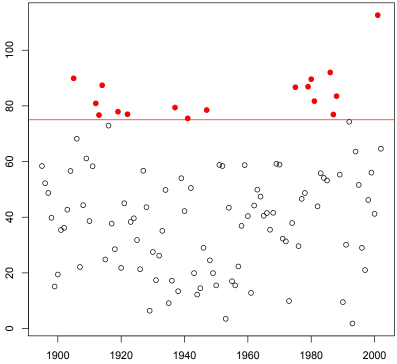
$$\overline{H}_{\hat{\xi}}\left(\frac{x-u}{\hat{\sigma}}\right) \times \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i > u).$$

Over interpreting MLE fits from 500 GP samples of size 100



Can you find the two woman faces ?

Can you find two inferential strategies ?



Animations

Annual Review of Statistics and Its Application

Statistics of Extremes: Animation 2

An animation from the 2015 review by A.C. Davison and R. Huser, "Statistics of Extremes," from the *Annual Review of Statistics and Its Application*.

Illustration of the point process of exceedances and the convergence to the GPD. For increasing values of n, the plots display the point process of rescaled times and rescaled variables, namely $(j/(n+1), (Y_j - b_n)/a_n)$, for data simulated from the uniform (*top left*), standard Gaussian (*top right*), unit exponential (*bottom left*), and 0.2-Pareto distributions. The side plots are histograms of the exceedances over the threshold u (*horizontal blue line*), i.e., $a_n^{-1}(Y_j - b_n)|a_n^{-1}(Y_j - b_n) > u$. The solid red curves are the corresponding asymptotic GPD densities.

View on YouTube | Read Associated Article

Precipitation in Colorado's front range

Data

- 56 weather stations in Colorado (semi-arid and mountainous region)
- Daily precipitation for the months April-October
- Time span = 1948-2001
- Not all stations have the same number of data points
- Precision : 1971 from 1/10th of an inche to 1/100

D. Cooley, D. Nychka and P. Naveau, (2007). Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of The American Statistical Association.*

Pierre Simon Laplace (1749-1827)

“L’analyse des probabilités assigne la probabilité de ces causes, et elle indique les moyens d’accroitre de plus en plus cette probabilité.”
“Essai Philosophiques sur les probabilités” (1774)

Pierre Simon Laplace (1749-1827)

"If an event can be produced by a number of n different causes, then the probabilities of the causes given the event ... are equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of the causes."

$$\mathrm{pr}(\text{cause}_i|\text{event}) = \frac{\mathrm{pr}(\text{event}|\text{cause}_i) \times \mathrm{pr}(\text{cause}_i)}{\sum_{j=1}^{n} \mathrm{pr}(\text{event}|\text{cause}_j) \times \mathrm{pr}(\text{cause}_j)}$$

Bayes' formula = calculating conditional probability

$$\boxed{[x|y] \propto [y|x] \times [x]}$$

REV. T. BAYES

1701(?)- 1761 "An essay towards solving a Problem in the Doctrine of Chances" (1764)

Bayesian approach

$$[x|y] \propto [y|x] \times [x]$$

Advantages

- Integration of expert information via prior $[x]$
- Deals with the full distribution
- Non-Gaussian
- Non-linear

Drawbacks

- Integration of expert information via prior $[x]$
- densities are needed
- Complex algorithmic techniques (MCMC, particle-filtering)
- Can be slow and not adapted for large data sets
- Adequacy with EVT

Hierarchical Bayesian Model with three levels

$$\begin{aligned}\text{pr}(\text{process, parameters}|\text{data}) \quad &\propto \quad \text{pr}(\text{data}|\text{process, parameters})\\ &\quad \times \text{pr}(\text{process}|\text{parameters})\\ &\quad \times \text{pr}(\text{parameters})\end{aligned}$$

Process level : the scale and shape GPD parameters $(\xi(x), \sigma(x))$ are hidden random fields

Our main assumptions

- Process layer : The scale and shape GPD parameters $(\xi(x), \sigma(x))$ are random fields and result from an unobservable latent spatial process
- Conditional independence : precipitation are independent given the GPD parameters

Our main variable change

$$\sigma(x) = \exp(\phi(x))$$

Our three levels

A) **Data layer** := pr(data|process, parameters) =

$$\mathrm{pr}_\theta\{\mathbf{R}(\mathbf{x_i}) - u > y | \mathbf{R}(\mathbf{x_i}) > u\} = \left(1 + \frac{\xi_i\, y}{\exp \phi_i}\right)^{-1/\xi_i}$$

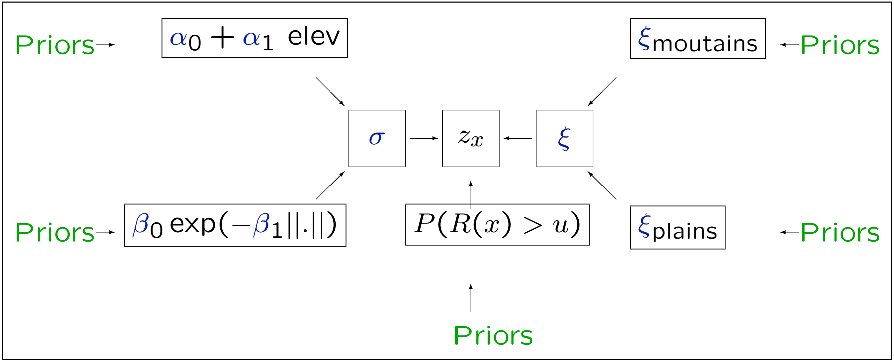
B) **Process layer :=** pr(process|parameters) :

$$\text{e.g. } \phi_i = \alpha_0 + \alpha_1 \times \text{elevation}_i + \text{MVN}\,(0, \beta_0 \exp(-\beta_1 ||x_k - x_j||))$$

$$\text{and } \xi_i \quad = \quad \begin{cases} \xi_{\text{moutains}}, \text{ if } x_i \in \text{mountains} \\ \xi_{\text{plains}}, \text{ if } x_i \in \text{plains} \end{cases}$$

C) **Parameters layer (priors) :=** pr(parameters) :
e.g. $(\xi_{\text{moutains}}, \xi_{\text{plains}})$ Gaussian distribution with non-informative mean and variance

Bayesian hierarchical modeling

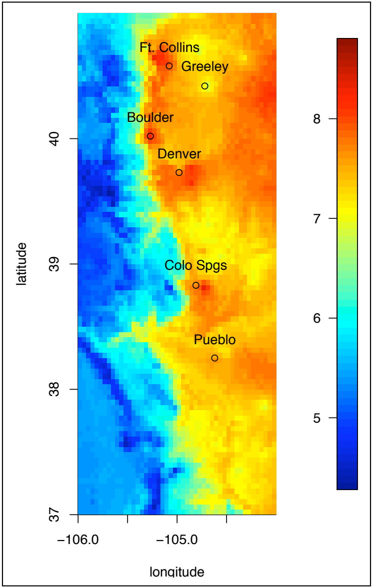


Model selection

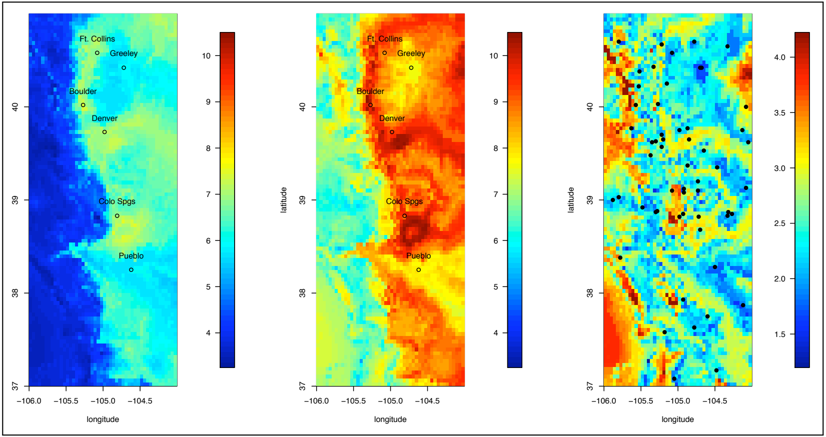
Baseline model		$\bar{D}$	p_D	*DIC*
Model 0:	$\phi = \phi$ $\xi = \xi$	73,595.5	2.0	73,597.2
Models in latitude/longitude space		$\bar{D}$	p_D	*DIC*
Model 1:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,442.0	40.9	73,482.9
Model 2:	$\phi = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,441.6	40.8	73,482.4
Model 3:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,443.0	35.5	73,478.5
Model 4:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \alpha_2(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,443.7	35.0	73,478.6
Models in climate space		$\bar{D}$	p_D	*DIC*
Model 5:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,437.1	30.4	73,467.5
Model 6:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,438.8	28.3	73,467.1
Model 7:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	**73,437.5**	**28.8**	**73,466.3**
Model 8:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	73,436.7	30.3	73,467.0
Model 9:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi + \epsilon_\xi$	73,433.9	54.6	73,488.5

NOTE: Models in the climate space had better scores than models in the longitude/latitude space. $\epsilon_{\cdot} \sim \text{MVN}(0, \Sigma)$, where $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} \|\mathbf{x}_i - \mathbf{x}_j\|)$.

Return levels posterior mean



Posterior quantiles of return levels (.025, .975)



Problem 3 : Means versus maxima

- Simulate a uniform sample of daily values for 1000 years (i.e. m=365,n=1000)
- Compute yearly means and maxima
- Plot the histograms of the sample average and sample maximum
- Repeat this experience for different values of *m* (season, month, etc)
- Repeat this experience with exponential and Gaussian draws

Problem 4 : fitting a GEV distribution

- Simulate three GEV samples of length 100 with $\xi = -0.3$, 0 and 0.3. Can you see a difference ?

Problem 4 : fitting a GEV distribution

- Simulate three GEV samples of length 100 with $\xi = -0.3$, 0 and 0.3. Can you see a difference ?
- Load package ”evd”
- Fit the GEV distribution to Annual Rainfall Maxima at Uccle, Belgium from 1938 to 1972 for four different time scales
  - day = Annual daily rainfall maxima.
  - hour = Annual hourly rainfall maxima.
  - tmin = Annual rainfall maxima over ten minute durations.
  - min = Annual rainfall maxima over one minute durations.