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

Philippe Naveau

Laboratoire des Sciences du Climat et l'Environnement (LSCE)
Gif-sur-Yvette, France

25 juin 2015

Applied Math and Earth sciences

"There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only, without knowledge of each other's work, but also will ignore the problems which require mutual assistance".

QUIZ

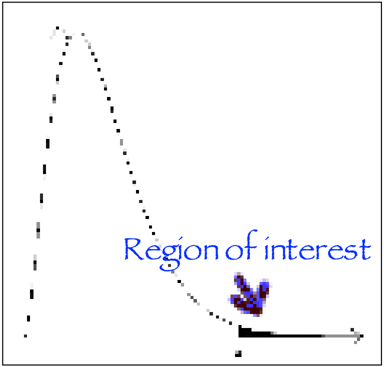
- (A) Gilbert Walker
- (B) Ed Lorenz
- (C) Rol Madden
- (D) Patrick Flandrin

- *Nous avons anticipé dans la mesure du possible mais on ne peut pas prévoir l'imprévisible"* **Xynthia's storm, 25th of Feb, 2010**
- *"Il est impossible que l'improbable n'arrive jamais"* Emil Julius Gumbel (1891-1966)

Extreme quotes

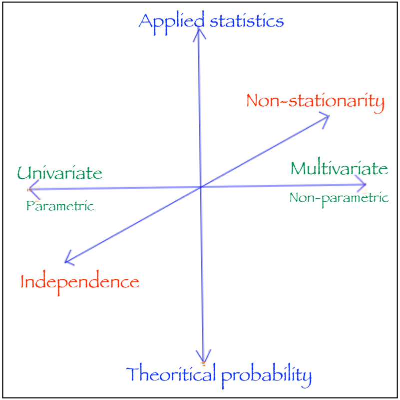
1. *"Man can believe the impossible, but man can never believe the improbable"* Oscar Wilde (Intentions, 1891)

Extreme events ? ... a probabilistic concept linked to the **tail** behavior : low frequency of occurrence, large uncertainty and sometimes strong amplitude.

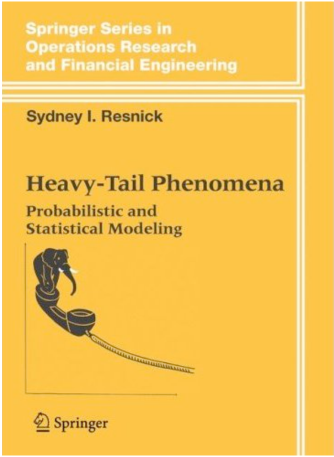


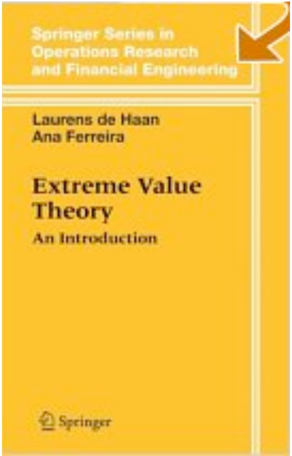
Important issues in Extreme Value Theory

- An asymptotic probabilistic concept
- A statistical approach for extrapolation of quantiles
- A general framework with “weak” assumptions (ie no model for the full data set)
- Assessing uncertainties



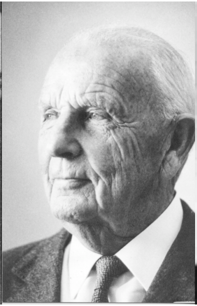
An active statistical and probabilistic field





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. He was a pioneer in the application of extreme value theory, particularly to climate and hydrology.
- 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. He also collected empirical examples of heavy-tailed distributions.

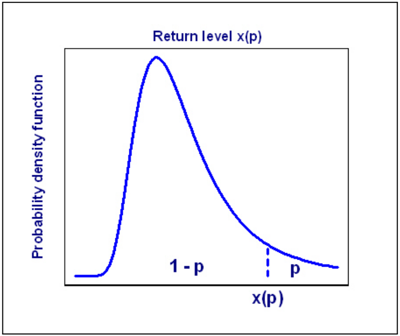
Other important names : Fisher and Tippet (1928), Gnedenko (1943), etc

Return levels and return periods

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.

A Return level interpretation

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



Max-stability

Let $M_n = \max(X_1, \ldots, X_n)$ with X_i iid with distribution F.

Definition : F max-stable if

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} < x\right) = F^n(a_n x + b_n) = F(x)$$

Blackboard

Unit-Frèchet $F(x) = \exp(-1/x)$ for $x > 0$. Then $a_n = n$ & $b_n = 0$

Gumbel $F(x) = \exp(-\exp(-x))$ for all real x. Then $a_n = 1$ & $b_n = \log n$

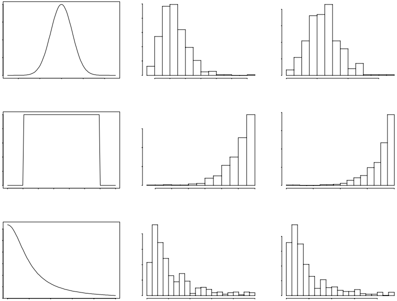
Weibull $F(x) = \exp(-(-x)^\alpha)$ for $x < 0$ (1 otherwise). Then $a_n = n^{-1/\alpha}$, $b_n = 0$

Maxima Distribution

Normal density ⇒ ⇐ Gumbel density

Uniform density ⇒ ⇐ Weibull density

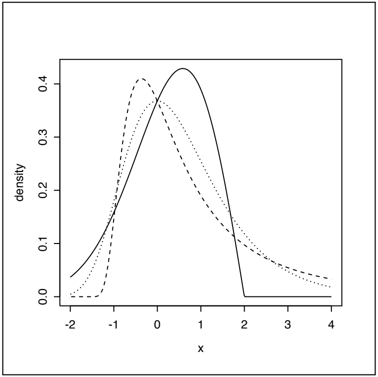
Cauchy density ⇒ ⇐ Fréchet density



$n = 50$ $n = 100$

Generalized Extreme Value (GEV) distribution

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



Home work : show that a GEV is max-stable

GEV and return levels

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

Computing the return level z_p such that $\text{GEV}(z_p) = 1 - p$

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

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

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

Estimating the GEV parameters estimates

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

GEV and return levels estimation

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

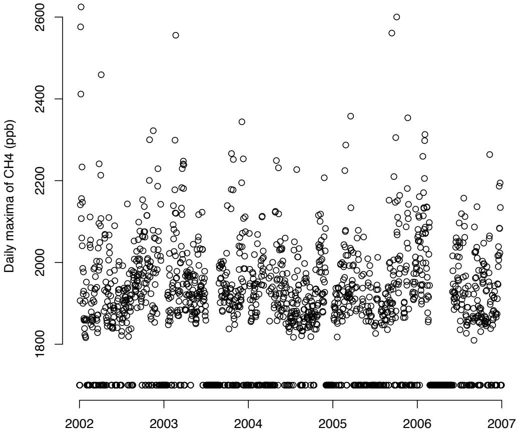
Maximum likelihood estimates of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})^t$

Asymptotically distributed as a multivariate Gaussian vector with mean $\theta = (\hat{\mu}, \hat{\sigma}, \hat{\xi})^t$ and covariance matrix that is the inverse of the **expected information matrix** whose elements are equal

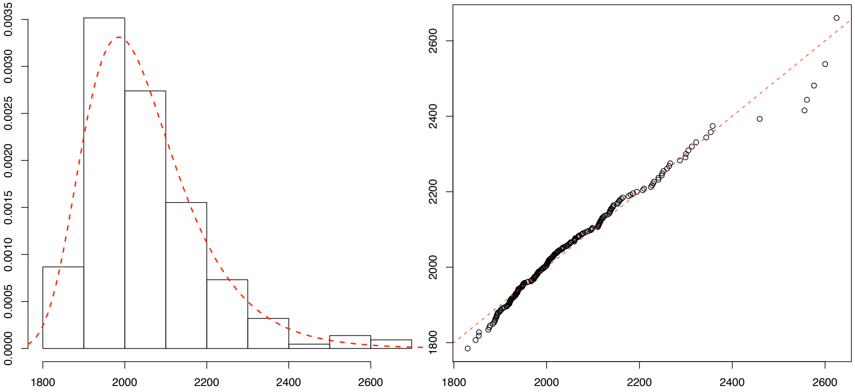
$$\mathbb{E}\left(-\frac{\partial^2 \log l(\theta)}{\partial\theta_i \partial\theta_j} \right)$$

where $l(\theta)$ is the likelihood function of the GEV distributed sample

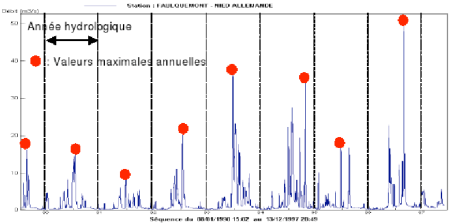
Daily maxima of CH_4 at Gif-sur-Yvette (Toulemonde et al., 2009, Environmetrics)

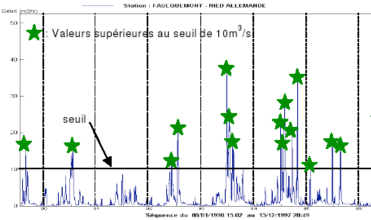


Maxima of CH_4 at Gif-sur-Yvette (Toulemonde et al., 2009, Environmetrics)



Peak over Threshold (POT)

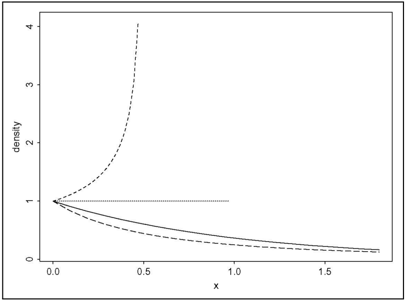




Thresholding : the Generalized Pareto Distribution (GPD)

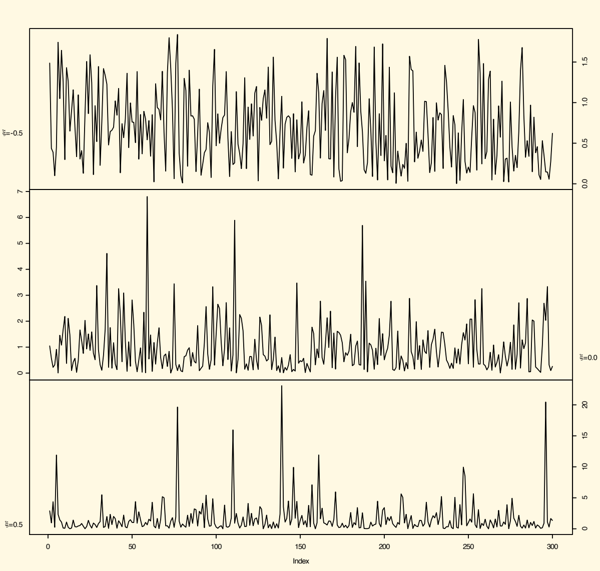
$$\mathbb{P}\{\mathbf{R}-u > y|\mathbf{R} > u\} = \left(1 + \frac{\xi\, y}{\sigma_u}\right)_+^{-1/\xi}$$

Vilfredo Pareto : 1848-1923



Born in France and trained as an engineer in Italy, he turned to the social sciences and ended his career in Switzerland. He formulated the power-law distribution (or ”Pareto’s Law”), as a model for how income or wealth is distributed across society.

GPD : "From Bounded to Heavy tails"



GPD

Two GPD examples

Exponential $F(x) = 1 - \exp(-x)$ for $x > 0$.

Uniform $F(x) = x$ for $0 < x < 1$.

GPD

GPD return level z_p

$$z_p = u + \frac{\sigma_u}{\xi}\left(\left[\frac{p}{\mathbb{P}(R > u)}\right]^{-\xi} - 1\right)$$

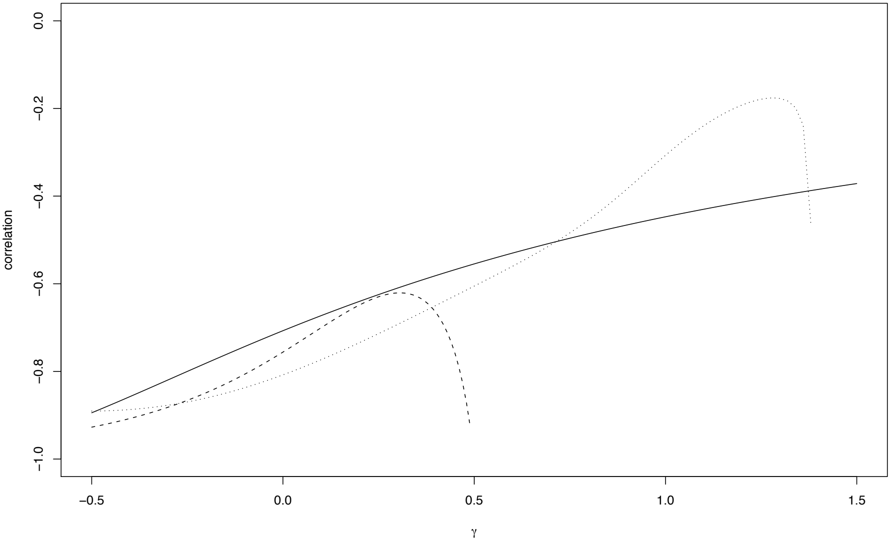
Estimating the return level z_p

$$\hat{z}_p = u + \frac{\hat{\sigma}_u}{\hat{\xi}}\left(\left[\frac{p \times n}{N_u}\right]^{-\hat{\xi}} - 1\right)$$

GPD parameters estimation

- Maximum likelihood estimation (*Smith, 1985*)
- Methods of moments type (PWM and GPWM, *Ribereau et al., 2010*)
- Exhaustive tail-index approaches

GPD estimators



Generalized Pareto Distribution (GPD)

$$\mathbb{P}\{\mathbf{R} - u > y | \mathbf{R} > u\} = \left(1 + \frac{\xi\, y}{\sigma_u}\right)^{-1/\xi}$$

Parameters

- u = predetermined threshold
- σ_u = scale parameter to be estimated
- ξ = shape parameter to be estimated

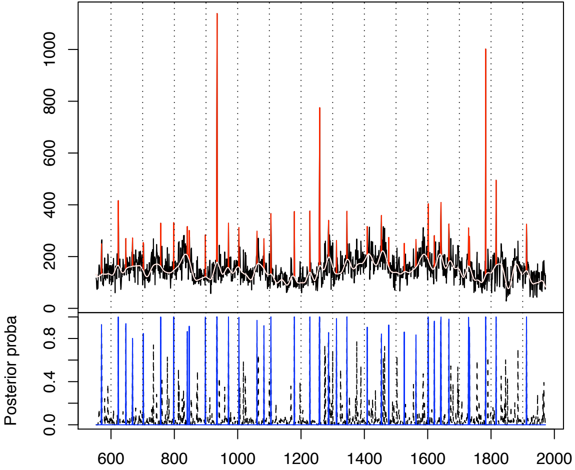
Advantages & Practical issues

- Flexibility to describe three different types of tail behavior
- More data are kept for the statistical inference
- Problem of threshold selection

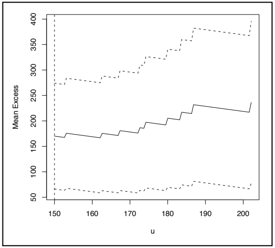
Stability property

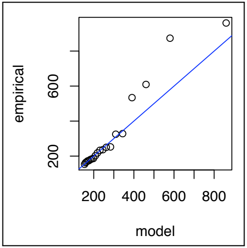
If the exceedance $(\mathbf{R} - u | \mathbf{R} > u)$ follows a GPD(σ_u, ξ) then the higher exceedance $(\mathbf{R} - v | \mathbf{R} > v)$ also follows GPD$(\sigma_u + (v - u)\xi, \xi)$

Crete ice core Greenland (ecm)



GPD diagnostics & models selection for a Crete data





$\hat{\xi} = 0.56\ (0.37)$

Intro summary

Modeling maxima : GEV

Stability for the max operator and $X_0, X_1, \ldots X_n$ idd GEV

$$a \ \max(X_1, \ldots, X_n) + b = X$$

Modeling excedances : GPD

If exceedances $(\mathbf{R} - u | \mathbf{R} > u)$ follows a GPD(σ_u, ξ) then higher exceedances $(\mathbf{R} - v | \mathbf{R} > v)$ also follows GPD$(\sigma_u + (v - u)\xi, \xi)$

A few studies dealing with geophysical extremes

- Casson and Coles (1999) a Bayesian hierarchical model for wind speeds exceedances
- Stephenson and Tawn (2005) Bayesian modeling of sea-level and rainfall extremes
- Chavez and Davison (2005) GAM for extreme temperatures (NAO)
- Wang et al. (2004) Wave heights with covariates
- Turkman et al. (2007), Spatial extremes of wildfire sizes
- Ribatet et al. (2010), Spatial R package for extremes
- Bel, Bacro, Lantujenoul (2010), Spatial extremes
- Extreme snow , Blanchet et al., 2010
- Special issue of the journal Extremes, 2010
- Pratique du calcul bayésien, JJ. Boreux, E. Parent et J. Bernier, 2010
- Biodiversity and extreme temperatures, Sang and Gelfand, 2009
- Lichenometry, Jomelli et al., 2007

Weather station of LYON :
1996-2011
Hourly precipitation
Our goal (joint work with Raphael Huser (Kaust) :
Modeling small, moderate and heavy precipitation

Modeling moderate and heavy precipitation

Desiderata

- a GP tail for extreme rainfall
- Gamma like behavior for moderate precipitation (Katz et al, 2002)
- Very few parameters
- Rapid and efficient estimation schemes
- Easy to obtain return levels
- Simple simulation algorithms
- No threshold selection

Frigessi et al, 2002, Vrac and N. (2007)

$$c\Big[(1 - p_{\mu,\tau}(x)) \text{ light tailed density} + p_{\mu,\tau}(x)\ h_\xi(x/\sigma)/\sigma\Big],$$

where h_ξ a GP density with $\xi > 0$ and the weight function $p_{\mu,\tau}(.)$

$$p_{\mu,\tau}(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\tau}\right).$$

Desiderata

- \+ a GP tail for extreme rainfall
- \+ Gamma like behavior for moderate precipitation
- \- Very few parameters
- – Inference rapid and efficient
- \- Easy to obtain return levels
- \+ Simple simulation schemes
- +/- No threshold selection
- \- One component has to be heavy-tailed

Hydrid-Pareto, Carreau and her co-authors (2006 ; 2009 ; 2011)

Desiderata

+ a GP tail for extreme rainfall
- Gamma like behavior for moderate precipitation
+ Very few parameters (3)
+ MLE Inference
+ Easy to obtain return levels
+ Simple simulation schemes
+ No threshold selection
- Strong dependence among parameters
- Negative values allowed

Going back to Extreme Value Theory (second order condition)

Falk M., Husler J., Reiss R.-D. (2010)

$$\frac{1}{\sigma} h_{\xi}(x/\sigma) \left(1 + O\left(\overline{H}_{\xi}^{\delta}(x/\sigma)\right)\right),$$

where $\kappa > 0$

What's about very low rainfall ?

EVT at work

$$\mathbb{P}(-X > -x \mid -X > -u) \approx \overline{H}_{-1/\kappa}\left(\frac{-x+u}{\nu}\right)$$

where $\kappa > 0$

What's about very low rainfall ?

EVT at work

$$\mathbb{P}(-X > -x | -X > -u) \approx \overline{H}_{-1/\kappa}\left(\frac{-x+u}{\nu}\right)$$

where $\kappa > 0$

$$\mathbb{P}(X \leq x) \approx \text{cst} \times x^{\kappa}, \text{ for any small } x \geq 0.$$

(this condition satisfied by a gamma density $f(x) \propto x^{\kappa-1} e^{-x/\theta}$, $x \geq 0$, $\kappa, \theta > 0$).

Full rainfall range

EVT on both sides of the tails

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - H_{\xi}\left(\frac{x}{\sigma}\right), & \text{for any large } x, \\ \text{cst} \times x^{\kappa}, & \text{for any small } x \text{ near } 0. \end{cases}$$

Full rainfall range

EVT on both sides of the tails

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - H_\xi\left(\frac{x}{\sigma}\right), & \text{for any large } x, \\ \text{cst} \times x^\kappa, & \text{for any small } x \text{ near } 0. \end{cases}$$

Our extended GPD

Random variable definition

$$X = \sigma H_{\xi}^{-1}\{G^{-1}(U)\},$$

where G is a continuous cdf on $[0, 1]$.

Our extended GPD

Random variable definition

$$X = \sigma H_{\xi}^{-1}\{G^{-1}(U)\},$$

where G is a continuous cdf on $[0, 1]$.

Constraints on G

(A)

$$\lim_{u \downarrow 0} \frac{\overline{G}(1-u)}{u} = a, \text{ for some finite } a > 0,$$

(B)

$$\lim_{u \downarrow 0} \frac{G\{u\, v(u)\}}{G(u)} = b, \text{ for some finite } b > 0,$$

where $v(u)$ is any positive function such $v(u) = 1 + o(u)$ as $u \to 0$,

(C)

$$\lim_{u \downarrow 0} \frac{u^{\kappa}}{G(u)} = c, \text{ for some finite } c > 0.$$

Our extended GPD

Random variable definition

$$X = \sigma H_{\xi}^{-1}\{G^{-1}(U)\},$$

where G is a continuous cdf on $[0, 1]$.

Quantiles

$$x_p = F^{-1}(p) = \begin{cases} \frac{\sigma}{\xi}\left[\{1 - G^{-1}(p)\}^{-\xi} - 1\right], & \text{if } \xi > 0, \\ -\frac{\sigma}{\delta}\log\{1 - G^{-1}(p)\}, & \text{if } \xi = 0, \end{cases}$$

CDF

$$F(x) = G\left\{H_{\xi}\left(\frac{x}{\sigma}\right)\right\} \text{ and } \overline{F}(x) = \overline{G}\left\{H_{\xi}\left(\frac{x}{\sigma}\right)\right\}.$$

Examples

Constraints on G

(A)

$$\lim_{u \downarrow 0} \frac{\overline{G}(1-u)}{u} = a, \text{ for some finite } a > 0,$$

(B)

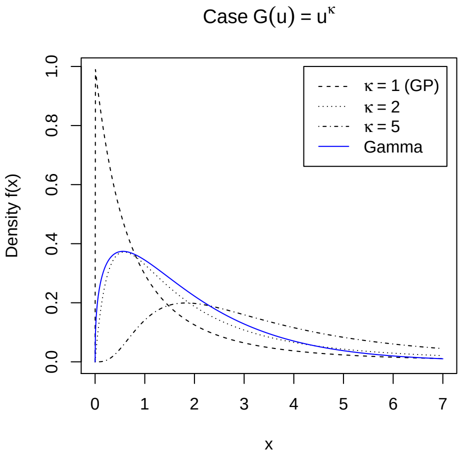
$$\lim_{u \downarrow 0} \frac{G\{u\, v(u)\}}{G(u)} = b, \text{ for some finite } b > 0,$$

where $v(u)$ is any positive function such $v(u) = 1 + o(u)$ as $u \to 0$,

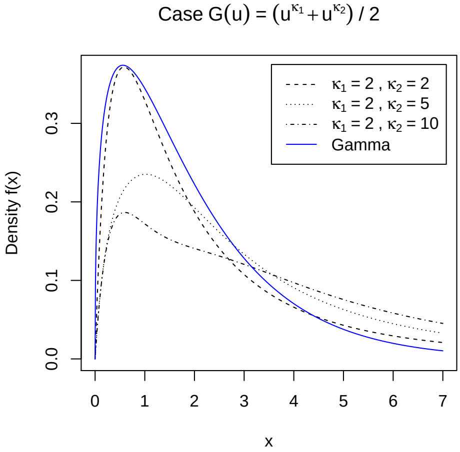
(C)

$$\lim_{u \downarrow 0} \frac{u^{\kappa}}{G(u)} = c, \text{ for some finite } c > 0.$$

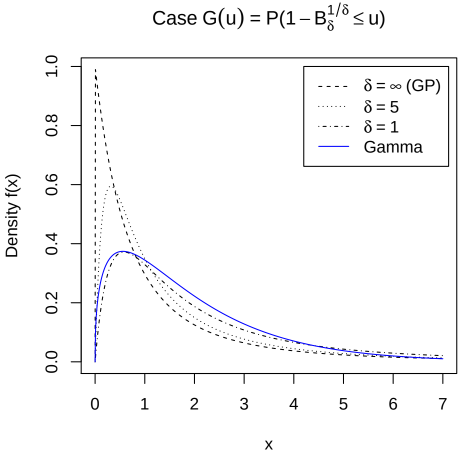
Examples (Papastathopoulos and Tawn, 2013)



Examples


Case G(u) = (u^κ1 + u^κ2) / 2
κ1 = 2 , κ2 = 2
κ1 = 2 , κ2 = 5
κ1 = 2 , κ2 = 10
Gamma
Density f(x)
0.0
0.1
0.2
0.3
0
1
2
3
4
5
6
7
x

Examples (Falk M., Husler J., Reiss R.-D. (2010))



Inference

Probability Weighted Moments

$$\mathbb{E}\left(X\overline{F}^{s}(X)\right) = \frac{\sigma}{\xi}\left(\mathbb{E}\left[\{1 - G^{-1}(U)\}^{-\xi}(1-U)^{s}\right] - \frac{1}{1+s}\right),$$

Maximum Likelihood Estimation

Simulations

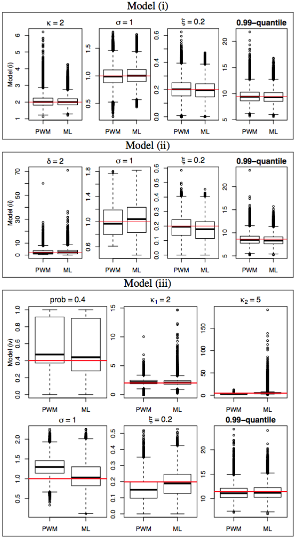


Fig. 4. Boxplots of estimated parameters and 0.99-quantiles using PWMs (left) and ML (right), for model (i) with $\sigma = 1$, $\xi = 0.2$, $\kappa = 2$, model (ii) with $\sigma = 1$, $\xi = 0.2$, $\delta = 2$, and model (iii) with $p = 0.4$, $\kappa_1 = 2$, $\kappa_2 = 5$, $\sigma = 1$ and $\xi = 0.2$. Boxplots are based on 10^5 independent replicates, and true values are represented by horizontal red lines.

Simulations

Table 1. Ratio of root mean squared errors (RMSEs) of PWMs and MLE based on estimates obtained from 10^5 independent datasets of size $n = 300$. Each **cell in bold** represents the ratio of RMSEs for the 99%-**quantile**. Non-bold cells correspond to ratios for each parameter, i.e. $\sigma/\xi/\kappa$ for model (i), and $\sigma/\xi/\delta$ for model (ii).

(i) $G(u) = u^{\kappa}$

κ	ξ = 0.1	0.2	0.3
1	1.06/1.02/1.17 **1.01**	1.06/0.98/1.19 **0.98**	1.11/1.00/1.27 **0.98**
2	1.05/1.02/1.13 **1.01**	1.07/1.00/1.18 **0.99**	1.15/1.05/1.33 **1.02**
5	1.74/1.39/1.29 **1.00**	1.68/1.49/1.31 **1.14**	1.19/1.09/1.34 **1.04**
10	2.40/1.70/1.41 **0.76**	4.21/2.82/1.32 **1.26**	1.39/1.26/0.69 **1.06**

(ii) $G(u) = \overline{V}_{\delta}\{(1-u)^{\delta}\}$

δ	ξ = 0.1	0.2	0.3
0.5	1.00/1.03/0.98 **1.02**	0.91/0.90/0.91 **1.02**	0.92/0.92/0.91 **1.05**
1	0.95/0.99/0.94 **1.03**	0.89/0.90/0.90 **1.02**	0.90/0.92/0.90 **1.05**
2	1.02/1.00/0.95 **1.03**	0.95/0.93/0.91 **1.01**	0.93/0.94/0.92 **1.04**
5	1.06/1.05/0.95 **1.02**	0.99/0.97/0.98 **0.98**	0.98/0.97/1.07 **0.99**

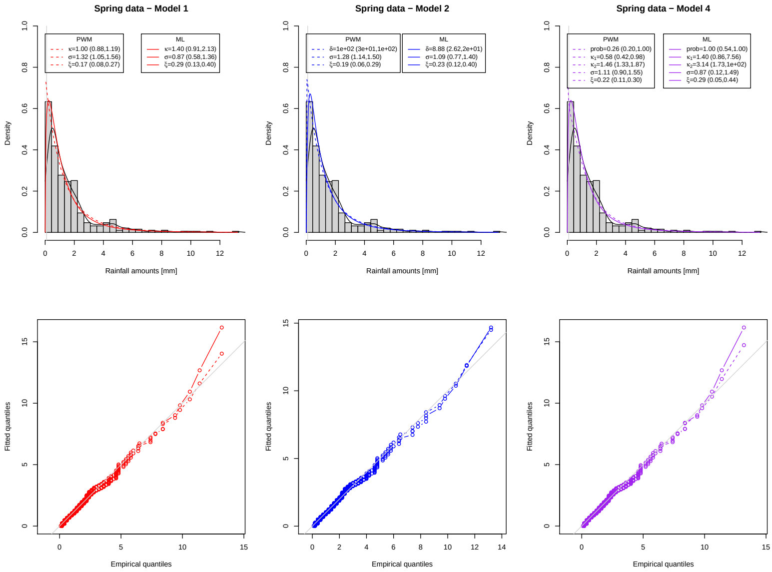
(iii) $G(u) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$

κ_1	κ_2 = 2	5	10	20
1	**0.57**	**0.96**	**0.68**	**0.74**
2	–	**0.95**	**0.98**	**0.71**
5	–	–	**1.29**	**1.04**
10	–	–	–	**1.38**

(iv) $G(u) = [\overline{V}_{\delta}\{(1-u)^{\delta}\}]^{\kappa/2}$

κ	δ = 0.5	1	2	5
1	**1.13**	**1.13**	**1.09**	**1.02**
2	**1.09**	**1.09**	**1.06**	**1.01**
5	**1.10**	**1.09**	**1.07**	**1.04**
10	**1.02**	**1.13**	**1.03**	**1.07**

Back to Lyon

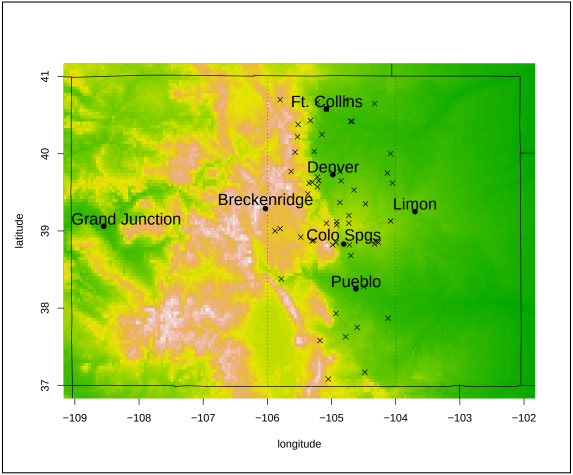


Spring data − Model 1
Spring data − Model 2
Spring data − Model 4
PWM
ML
κ=1.00 (0.88,1.19)
σ=1.32 (1.05,1.56)
ξ=0.17 (0.08,0.27)
κ=1.40 (0.91,2.13)
σ=0.87 (0.58,1.36)
ξ=0.29 (0.13,0.40)
δ=1e+02 (3e+01,1e+02)
σ=1.28 (1.14,1.50)
ξ=0.19 (0.06,0.29)
δ=8.88 (2.62,2e+01)
σ=1.09 (0.77,1.40)
ξ=0.23 (0.12,0.40)
prob=0.26 (0.20,1.00)
κ1=0.58 (0.42,0.98)
κ2=1.46 (1.33,1.87)
σ=1.11 (0.90,1.55)
ξ=0.22 (0.11,0.30)
prob=1.00 (0.54,1.00)
κ1=1.40 (0.86,7.56)
κ2=3.14 (1.73,1e+02)
σ=0.87 (0.12,1.49)
ξ=0.29 (0.05,0.44)
Density
Rainfall amounts [mm]
Fitted quantiles
Empirical quantiles

Summary of our extended GPD

- A distribution in compliance with EVT on both sides
- No threshold needed
- Very few parameters for the full precipitation range
- PWM and MLE inference but Bayesian possible
- Work in progress (bivariate)

Daily precipitation (April-October, 1948-2001, 56 stations)



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.”

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

Bayes' formula = calculating conditional probability

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

Rev. T. Bayes

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

Applied Math and Earth sciences

"There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only, without knowledge of each other's work, but also will ignore the problems which require mutual assistance".

QUIZ

- *(A) Gilbert Walker*
- *(B) Ed Lorenz*
- *(C) Rol Madden*
- *(D) Patrick Flandrin*

FIG. 3. *Photograph of Sir Gilbert T. Walker* (*source*: *Royal Society*; *Taylor*, 1962).

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

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

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** $:= \mathbb{P}(\text{data}|\text{process, parameters}) =$

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

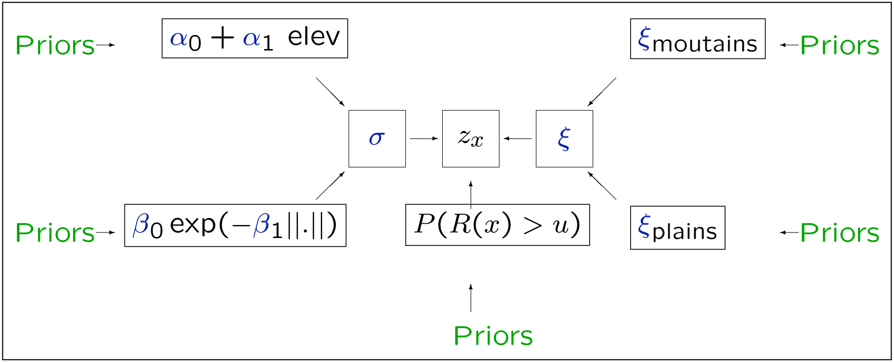
B) **Process layer :=** $\mathbb{P}(\text{process}|\text{parameters})$:

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

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

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

Bayesian hierarchical modeling

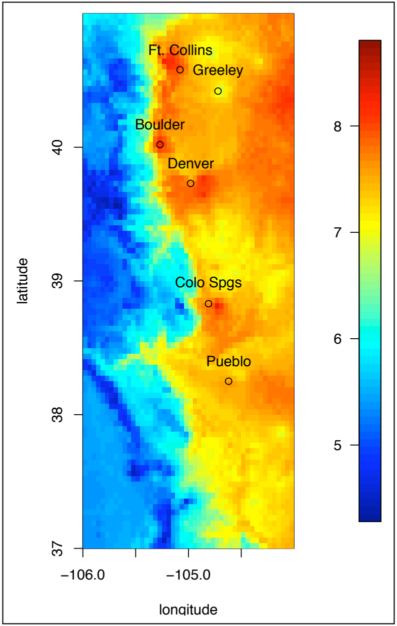


Model selection

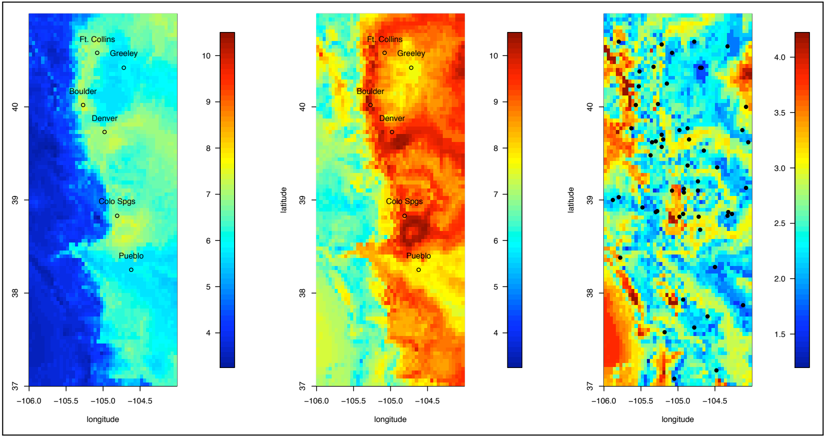
Baseline model		$\bar{D}$	p_D	*DIC*
Model 0:	$\phi = \phi$ $\xi = \xi$	73,595.5	2.0	73,597.2
Models in latitude/longitude space		$\bar{D}$	p_D	*DIC*
Model 1:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,442.0	40.9	73,482.9
Model 2:	$\phi = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	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 $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} \|\mathbf{x}_i - \mathbf{x}_j\|)$.

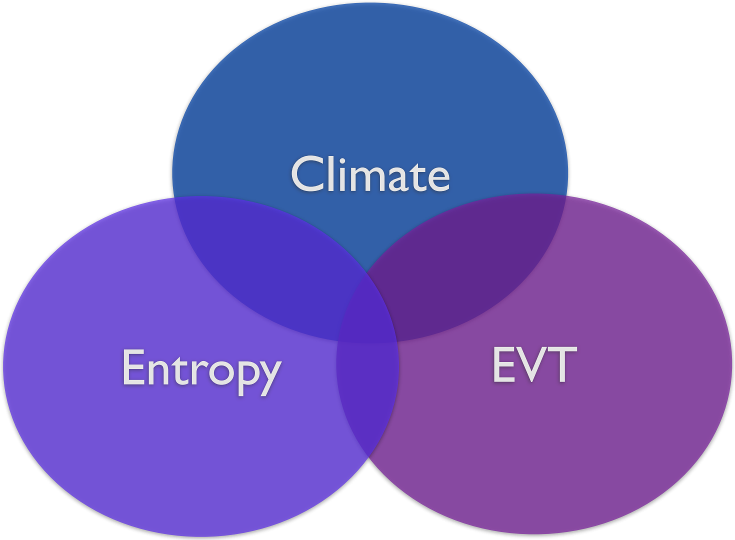
Return levels posterior mean

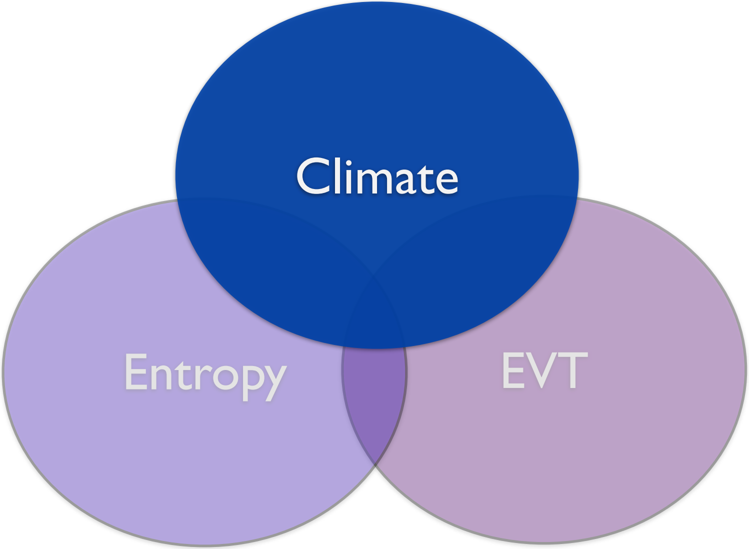


Posterior quantiles of return levels (.025, .975)



A simple and fast tool to deal with non-stationarity in GPD

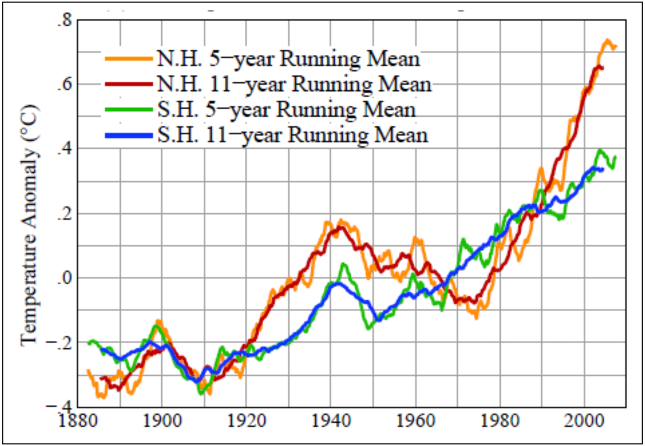



Climate
Entropy
EVT

Objectives

- Detecting changes in time, e.g. are the last 30 year extreme temperatures different from earlier periods ?
- Detecting changes in space, e.g. are extreme temperatures in Paris different from the ones recorded in Trieste ?
- We don't deal with the attribution problem here (see Francis Z.)

Hemispheric mean temperatures (source GISS-NASA 2010)



Temperatures anomalies (1961-1990)

Source : Abarca-Del-Rio and Mestre, GRL. (2006)

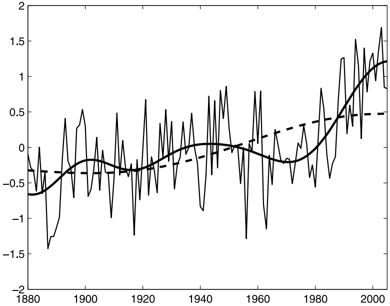
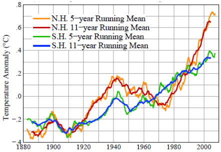


Figure 5. Time series of annual mean temperature anomalies over France (black), (A6 + D5 + D6) reconstruction (solid black), and A6 (dashed).

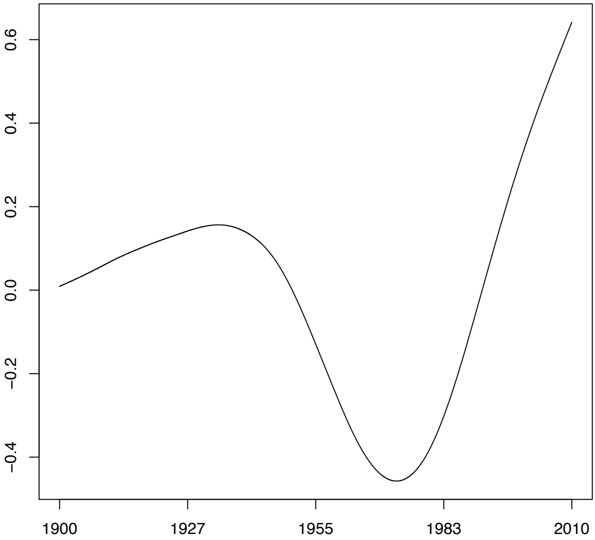


An illustration

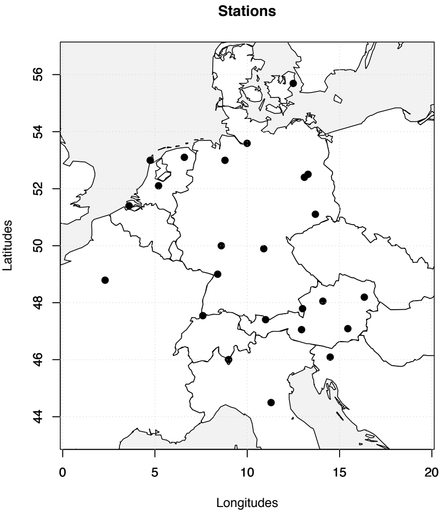
Paris, station Montsouris

- Daily maxima of temperatures
- Years from 1900 to 2010
- 40 515 maxima
- Urban island effect (?)
- Temperature maxima are often of Weibull type (precip light Fréchet)
- Clustering versus declustering
- 30 years = the climatology yard stick

Paris, smooth trend



Beyond Paris



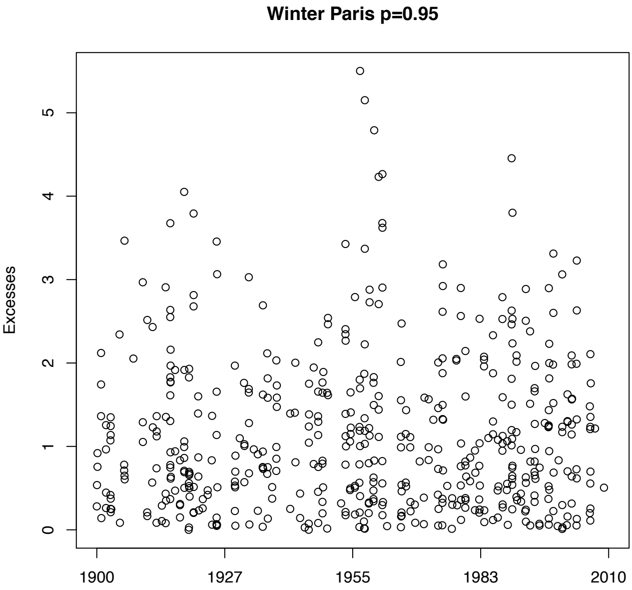
European stations with long time series

Table 1. Ratio of root mean squared errors (RMSEs) of PWMs and MLE based on estimates obtained from 10^5 independent datasets of size $n = 300$. Each **cell in bold** represents the ratio of RMSEs for the 99%**-quantile**. Non-bold cells correspond to ratios for each parameter, i.e. $\sigma/\xi/\kappa$ for model (i), and $\sigma/\xi/\delta$ for model (ii).

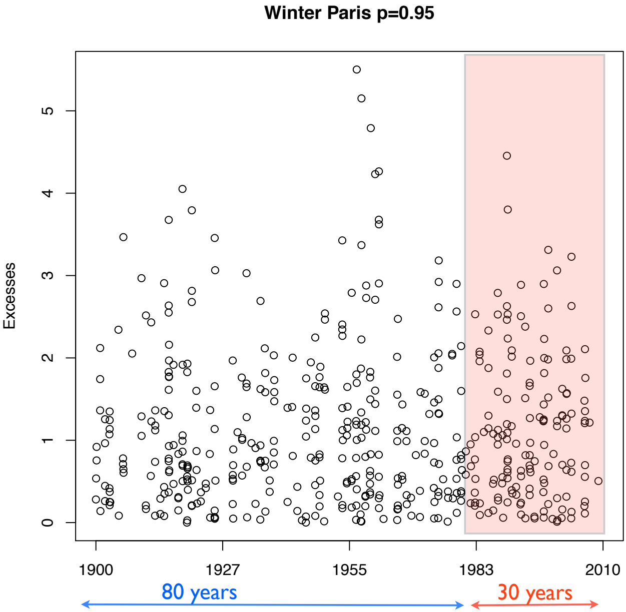
(i) $G(u) = u^\kappa$			
κ	ξ		
	0.1	0.2	0.3
1	1.06/1.02/1.17 **1.01**	1.06/0.98/1.19 **0.98**	1.11/1.00/1.27 **0.98**
2	1.05/1.02/1.13 **1.01**	1.07/1.00/1.18 **0.99**	1.15/1.05/1.33 **1.02**
5	1.74/1.39/1.29 **1.00**	1.68/1.49/1.31 **1.14**	1.19/1.09/1.34 **1.04**
10	2.40/1.70/1.41 **0.76**	4.21/2.82/1.32 **1.26**	1.39/1.26/0.69 **1.06**

(ii) $G(u) = \overline{V}_\delta\{(1-u)^\delta\}$

Paris, analyzing excesses per season



A main question about the last "30 year" extremes



A current approach in the climate community about detecting changes in extremes

- Fitting a GEV or a GPD based model to describe extremes
- Investigating how the GEV or GEV parameters change in space or time.

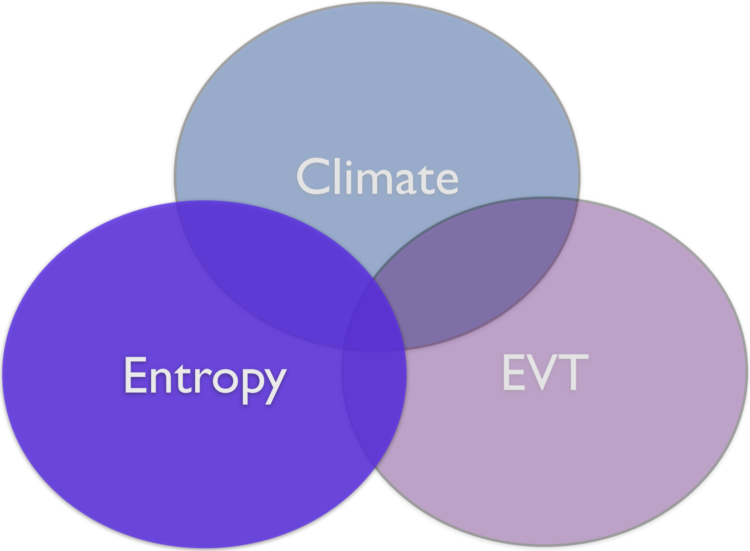
E.g., Jarusková and Rencová (2008), Fowler and Kilsby, (2003), Kharin et al., (2007)

Desiderata of our statistical approach

- Very few assumptions (neither imposing a GPD nor a GEV)
- Fast computations
- Good statistical properties of our estimators
- Interpretability
- Cross discipline tools (statistics and climatology)

Our assumptions

- Trends and annual cycles have been removed
- High detrended excesses can be considered stationary (or even iid)
- The two periods of interest belongs to the same domain of attraction
- The upper end points of both periods are equal
- We do **not** assume that the two periods have necessarily the same shape parameter


Climate
Entropy
EVT

Definitions (Kullback, 1968)

The Kullback-Leibler directed divergence

$$I(f;g) = \mathbb{E}_f \left(\log \left(\frac{f(\mathbf{X})}{g(\mathbf{X}})\right) \right),$$

where f and g pdfs and $\mathbf{X}$ random vector with density f.

The Kullback-Leibler divergence

$$D(f;g) = I(f;g) + I(g;f)$$

a symmetrical measure relative to f and g.

Kullback-Liebler divergence

Advantages

- Interpretability
- A single and simple summary
- Cross discipline tools (statistics and climatology)
- It is not an index
- Links with model selection criteria

Kullback-Liebler divergence

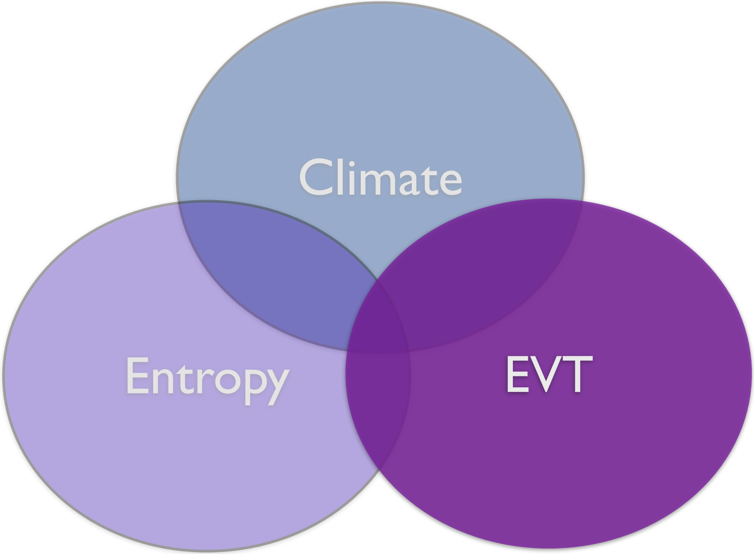
$$I(f;g) = \mathbb{E}_f\left(\log\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}\right)\right), \text{ and } D(f;g) = I(f;g) + I(g;f)$$

How to compute and infer the divergence ?

- Assume f and g have explicit expressions, e.g. GP
- Plug in the mle estimates

Drawbacks

- Strong assumptions : f and g have explicit expressions, e.g. GP
- Difficult to work with the likelihood when the dimension is large (composite likelihoods, etc)


Climate
Entropy
EVT

The “Old” life

Distributions (cdf)

The “Modern” style

Densities (pdf)

$I(f;g) = \mathbb{E}_f\left(\log\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}\right)\right)$

A general comment

The divergence is based on the **probability density functions**, **f** and **g**, but extreme behaviors are better captured by **cumulative distribution functions**, **F** and **G**, or their tails, $1 - \mathbf{F}$ and $1 - \mathbf{G}$.

..., but what is the question for this application ? It is not

- to infer GPD parameters

Entropy for excesses above u

Notations

Let X and Y be two abs. cont. r.v.'s with identical upper end-points.
For any u, we define the random vector $[X|\ X > u]$ by its density

$$f_u(x) = \frac{f(x)}{\overline{F}(u)}, \text{ for all } x_F > x > u.$$

Divergence definition

$$\begin{aligned} I(f_u; g_u) &= \mathbb{E}_{f_u}\left(\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right) \\ &= \frac{1}{\overline{F}(u)} \int_u^{x_F} \log\left(\frac{f_u(x)}{g_u(x)}\right) f(x)dx \end{aligned}$$

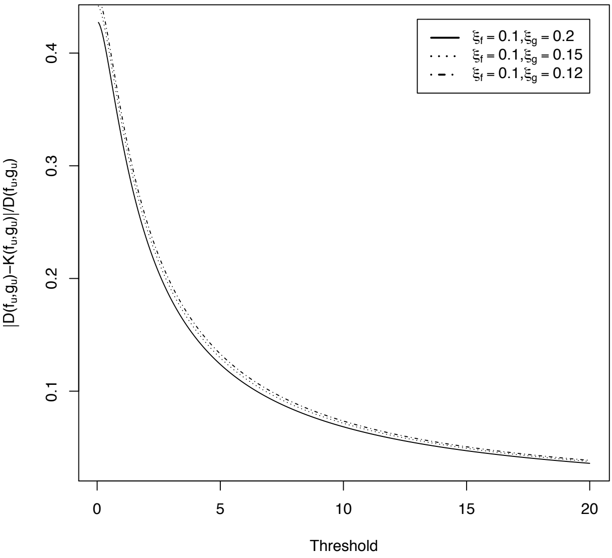
Entropy for excesses above u

An approximation of $I(f_u; g_u) = \mathbb{E}_{f_u}\left(\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right)$

Under mild assumptions, the divergence $D(f_u; g_u) = I(f_u; g_u) + I(g_u; f_u)$ is equivalent to the quantity $K(f_u; g_u) = -L(f_u; g_u) - L(g_u; f_u)$ where

$$L(f_u; g_u) = \mathbb{E}_f\left(\log \frac{\overline{G}(X)}{\overline{G}(u)} \middle| X > u\right) + 1$$

Approximation for the GPD



Necessary condition for applying our divergence approximation

Under mild conditions, the divergence $D(f_u; g_u) = I(f_u; g_u) + I(g_u; f_u)$ is equivalent to the quantity $K(f_u; g_u) = -L(f_u; g_u) - L(g_u; f_u)$ where

$$L(f_u; g_u) = \mathbb{E}_f \left(\log \frac{\overline{G}(X)}{\overline{G}(u)} \middle| X > u \right) + 1$$

Inference

One advantage of $\mathbb{E}_f\left(\log \frac{\overline{G}(X)}{\overline{G}(u)} \middle| X > u\right)$ **over** $\mathbb{E}_{f_u}\left(\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right)$

The Kullback Leibler divergence approximation can be estimated by

$$\widehat{L}(f_u; g_u) = \frac{1}{N_n}\sum_{i=1}^{n} \log \frac{\overline{G_m}(X_i \vee u)}{\overline{G_m}(u)}$$

with G_m is the empirical cdf of the Y_i's and $N_n := \#\{X_i, X_i \geq u\}$

Distribution under the null hypothesis : $\overline{F} = \overline{G}$

Suppose $n = m$

$$\overline{G_n}(x) = \overline{F_n}(x) \text{ in distribution}$$

Distribution under the null hypothesis : $\overline{F} = \overline{G}$

Suppose $n = m$

$$\overline{G_n}(x) = \overline{F_n}(x) \text{ in distribution}$$

The statistic

$$\widehat{L}(f_u; f_u) = \frac{1}{N_n} \sum_{i=1}^{n} \log \frac{\overline{F_n}(X_i \vee u)}{\overline{F_n}(u)}$$

does not depend on the original F and it is a distribution free statistic.

Simulations results

Coming tables

Number of false positive (wrongly rejecting that $f = g$) and negative out of **1000** replicas of two samples of sizes $n = m$ for a 95% level where f and g GP densities with shape parameter ξ_f and ξ_g, respectively.

Bounded tails

Weibull-Weibull case	$\xi_f = -.1$				
n \ ξ_g	$-.2$	$-.15$	$-.1$	$-.08$	$-.05$
50	**107**	**550**	**54**	**816**	**93**
100	**3**	**233**	**58**	**680**	**5**
200	**0**	**35**	**45**	**469**	**0**
500	**0**	**0**	**62**	**94**	**0**
1000	**0**	**0**	**50**	**5**	**0**

Bounded tails : comparing with the Kolmogorov-Smirnov test

Weibull-Weibull case	$\xi_f = -.1$									
n \ ξ_g	$-.2$		$-.15$		$-.1$		$-.08$		$-.05$	
50	262	**107**	691	**550**	26	**54**	889	**816**	241	**93**
100	42	**3**	445	**233**	50	**58**	824	**680**	24	**5**
200	0	**0**	110	**35**	30	**45**	627	**469**	0	**0**
500	0	**0**	0	**0**	52	**62**	189	**94**	0	**0**
1000	0	**0**	0	**0**	42	**50**	17	**5**	0	**0**

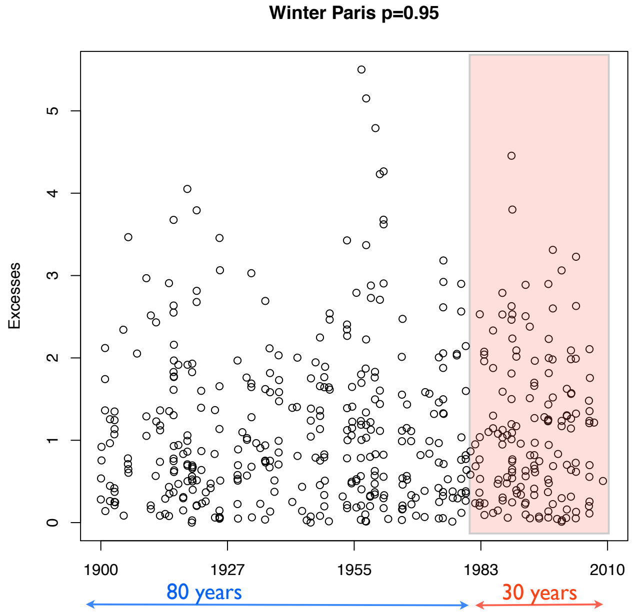
Heavy-tails : comparing with the Kolmogorov-Smirnov test

Fréchet-Fréchet case	$\xi_f = .1$									
n \ ξ_g	.05		.08		.1		.15		.2	
10	986	**946**	991	**939**	15	**49**	986	**954**	963	**950**
100	973	**935**	960	**941**	43	**49**	953	**930**	925	**899**
1000	937	**740**	946	**926**	45	**49**	943	**773**	740	**285**
10000	648	**9**	911	**597**	41	**55**	688	**16**	14	**0**

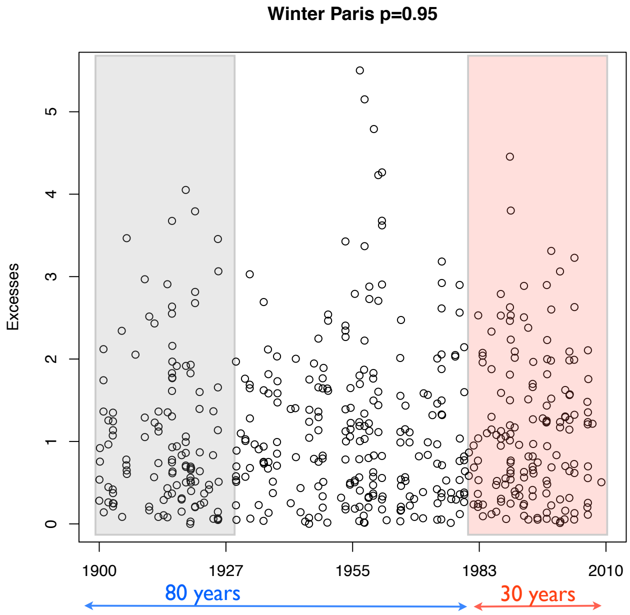
Gumbel case : comparing with the Kolmogorov-Smirnov test

Gumbel-Weibull case	$\xi_f = 0$									
n \ ξ_g	$-.5$		$-.4$		$-.3$		$-.2$		0	
50	804	**155**	868	**377**	921	**628**	949	**799**	28	**48**
100	513	**3**	727	**37**	896	**231**	933	**599**	38	**50**
200	63	**0**	348	**0**	709	**12**	888	**252**	37	**60**
500	0	**0**	1	**0**	134	**0**	616	**4**	58	**53**
1000	0	**0**	0	**0**	0	**0**	229	**0**	36	**54**
Gumbel-Fréchet case	$\xi_f = 0$									
n \ ξ_g	0		$.2$		$.3$		$.4$		$.5$	
50	33	**63**	943	**853**	945	**746**	919	**589**	885	**473**
100	36	**74**	923	**695**	907	**470**	844	**253**	733	**100**
200	30	**57**	902	**412**	807	**120**	606	**18**	392	**1**
500	43	**55**	706	**52**	335	**0**	101	**0**	13	**0**
1000	45	**59**	423	**1**	34	**0**	0	**0**	0	**0**

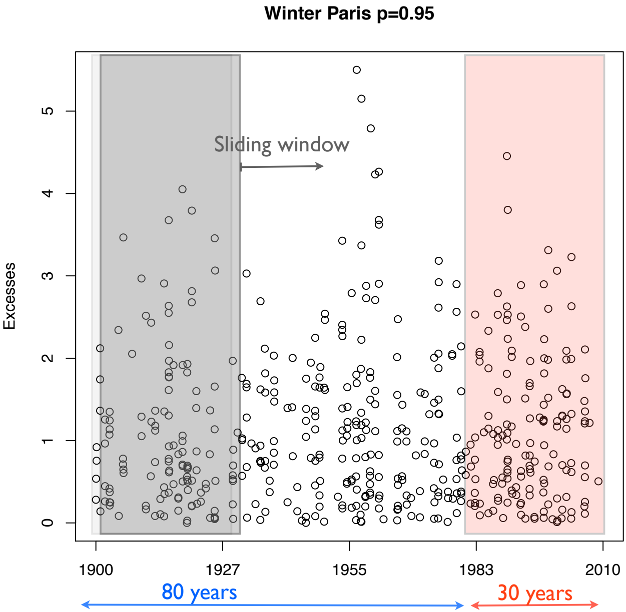
Coming back to our temperatures maxima recorded in Paris



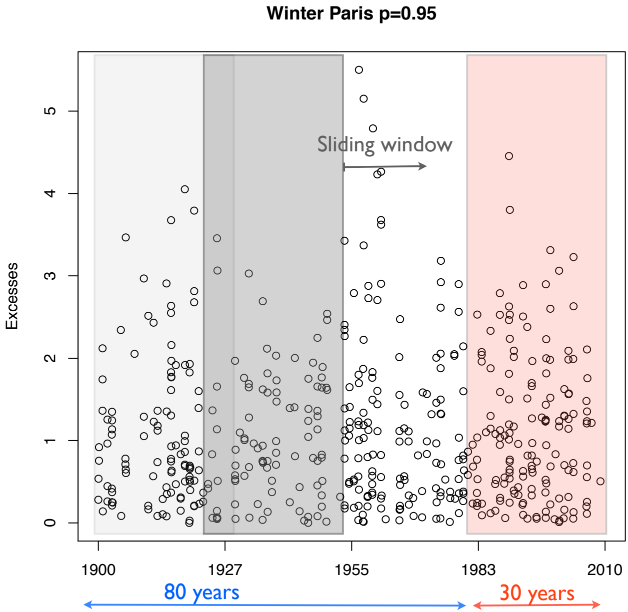
Step A (sliding window of length 30 years)



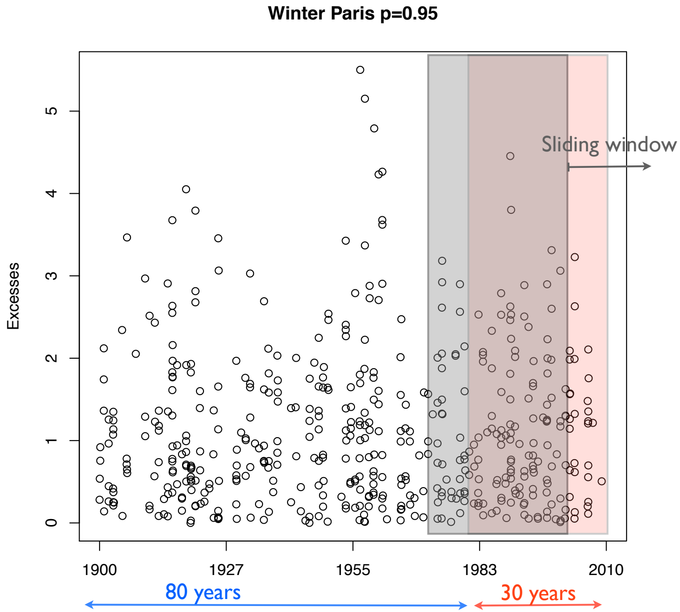
Step A (sliding window of length 30 years)



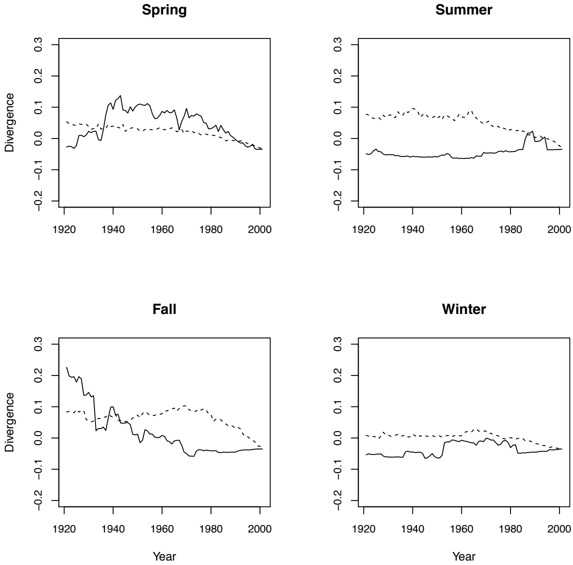
Step A (sliding window of length 30 years)



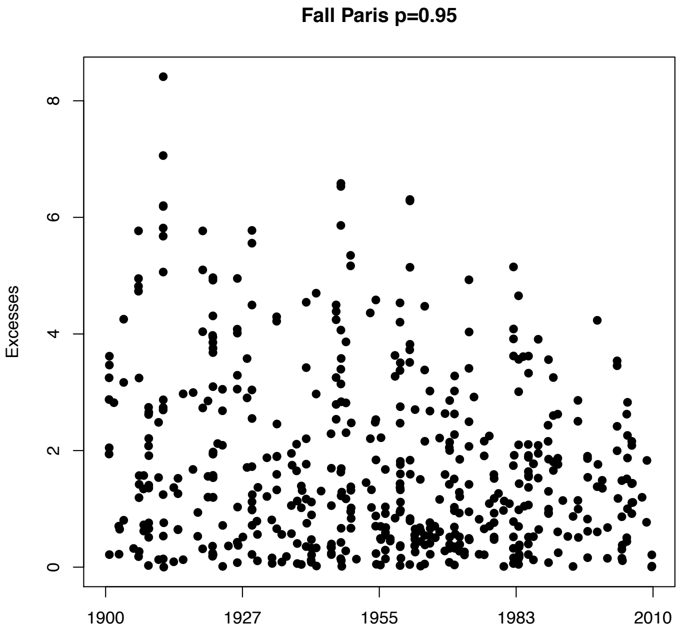
Step A (sliding window of length 30 years)



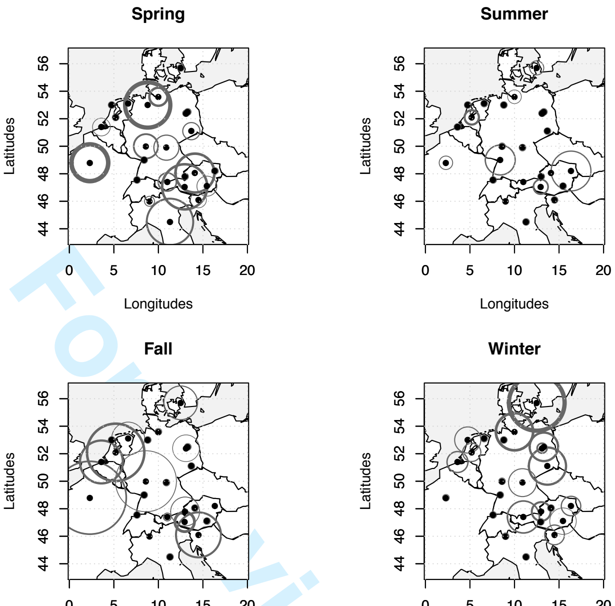
Back to Paris



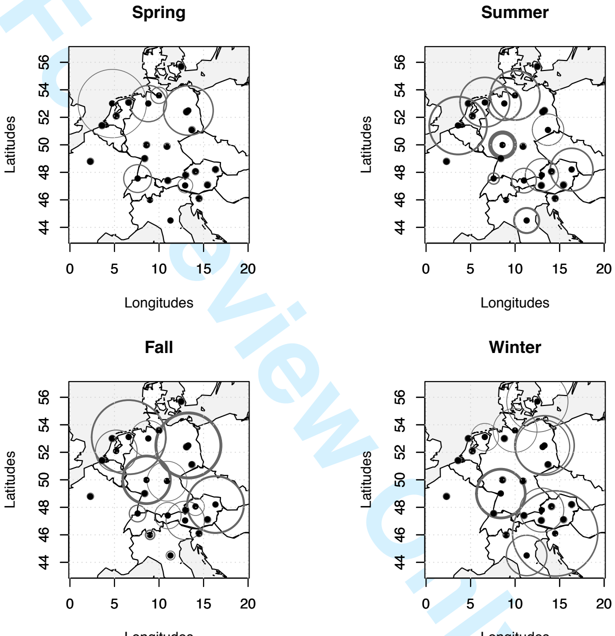
Paris Montsouris



Maxima


Spring
Summer
Fall
Winter
Latitudes
Longitudes
44 46 48 50 52 54 56
0 5 10 15 20

Minima


Spring
Summer
Fall
Winter
Latitudes
Longitudes
44 46 48 50 52 54 56
0 5 10 15 20

Take home messages for detecting with the entropy

- Extremes here mean very rare
- Connections between EVT and Kullback-Leibler divergence
- No need to choose a explicit density
- Fast algorithm
- Limited to one question
- Significant changes over Europe, especially for minima
- More research needed for the multivariate case
- More applications to large datasets

Statistics and Earth sciences

"There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only, without knowledge of each other's work, but also will ignore the problems which require mutual assistance".

QUIZ

- *(A) Sir Gilbert T. Walker (Walker, 1927b, page 321)*

FIG. 3. *Photograph of Sir Gilbert T. Walker (source: Royal Society; Taylor*, 1962).