Going beyond the Hill: An introduction to Multivariate Extreme Value Theory

Philippe Naveau naveau@lsce.ipsl.fr

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

Books: Coles (1987), Embrechts et al. (1997), Resnick (2006)

FP7-ACQWA, GIS-PEPER, MIRACLE & ANR-McSim, MOPERA

22 juillet 2014

Statistics and Earth sciences

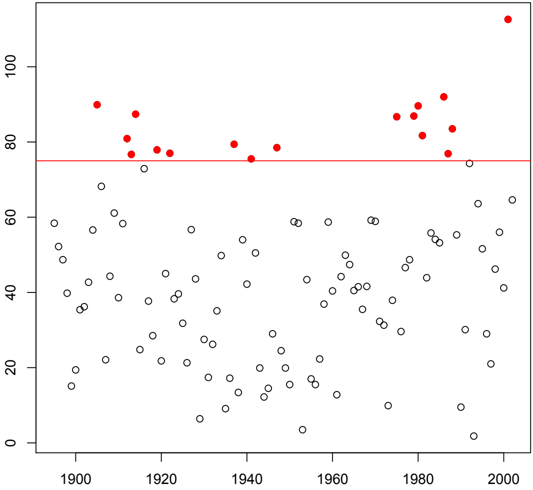
"There is, today, always a risk that specialists in two subjects, using languages full of words that are unintelligible without study, will grow up not only, without knowledge of each other's work, but also will ignore the problems which require mutual assistance".

QUIZ

- (A) Gilbert Walker
- (B) Ed Lorenz
- (C) Rol Madden
- (D) Francis Zwiers

EVT = Going beyond the data range

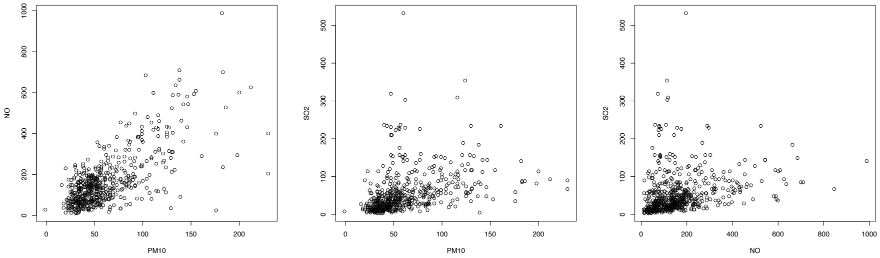
What is the probability of observing data above an high threshold ?



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

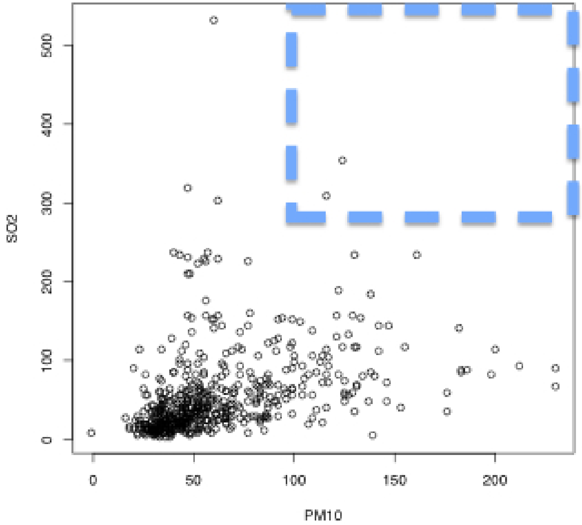
An example in three dimensions

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)

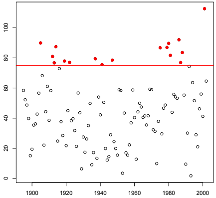


Typical question in multivariate EVT

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



Siméon 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 [1] 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.

---

1. Give HW

Poisson and maxima

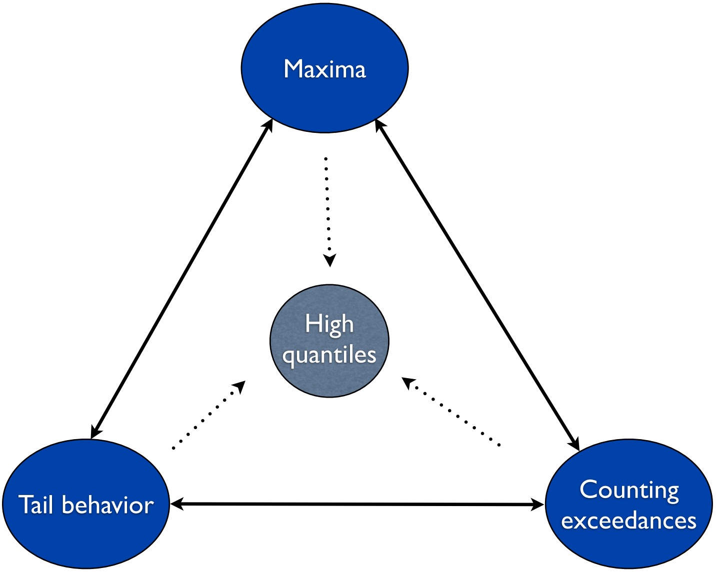
Counting = max

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

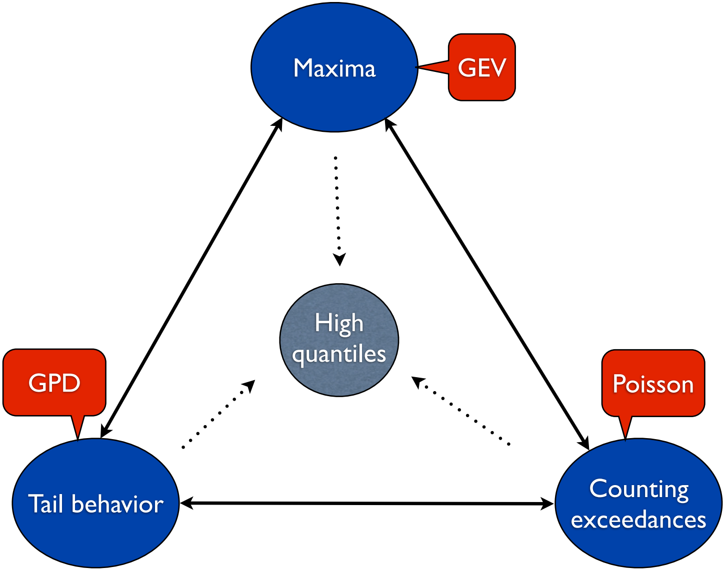
Poisson's at work

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

Equivalences



An univariate summary

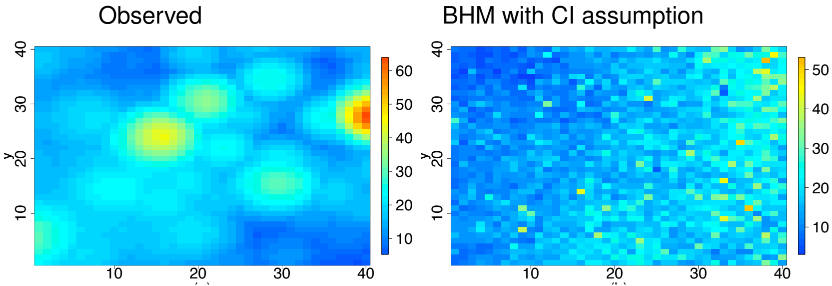


A few studies linking EVT with geophysical extremes

- Casson and Coles (1999) a Bayesian hierarchical model for wind speeds exceedances
- Stephenson and Tawn (2005) Bayesian modeling of sea-level and rainfall extremes
- Cooley et al. (2007) a Bayesian hierarchical GPD model that pooled precipitation data from different locations
- 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
- Lichenometry, Jomelli et al., 2007
- Hydrology Katz et al.
- Downscaling Vrac M., Kallache M., Rust H., Friedrichs P., etc
- GCMs and RCMS analysis Zwiers F., Maraun D., etc
- Attribution Smith R.

Limits of the univariate approach

Independence or conditional independence assumptions



Ribatet, Cooley and Davison (2010)

Why is Multivariate EVT needed ?

- Compute confidence intervals
- Calculating probabilities of joint extreme events
- Clustering of regions
- Extrapolation of extremes
- Downscaling of extremes
- Trading time for space (for small data sets)
- etc

A fundamental question[2] for iid bivariate vector (X_i, Y_i)

Suppose that we have unit Fréchet margins at the limit

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x) = \lim_{n\to\infty} P(\max(Y_1,\ldots,Y_n)/a_n \le x) = \exp(-x^{-1})$$

with a_n such that

$$P(X > a_n) = 1/n$$

2. L. de Hann, S. Resnick

A fundamental question[2] for iid bivariate vector (X_i, Y_i)

Suppose that we have unit Fréchet margins at the limit

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x) = \lim_{n\to\infty} P(\max(Y_1,\ldots,Y_n)/a_n \le x) = \exp(-x^{-1})$$

with a_n such that

$$P(X > a_n) = 1/n$$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = ??$$

2. L. de Hann, S. Resnick

Why is the solution so ugly ?

If

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = G(x,y)$$

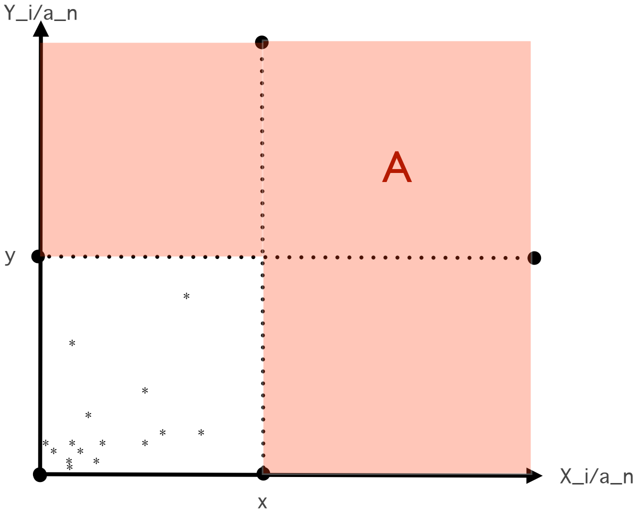
then

$$G(x,y) = \exp\left(-\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w)\right)$$

where $H(.)$ such that $\int_0^1 w\, dH(w) = 1$

Still counting

$$P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = P(N_n(A) = 0)$$



Still counting

$$P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = P(N_n(A) = 0)$$

Poisson again

If

$$\lim_{n\to\infty} E(N_n(A)) = \Lambda(A),$$

then

$$\lim_{n\to\infty} P(N_n(A) = 0) = P(N(A) = 0) = \exp(-\Lambda(A))$$

Still counting

$$P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = P(N_n(A) = 0)$$

Poisson again

If

$$\lim_{n\to\infty} E(N_n(A)) = \Lambda(A),$$

then

$$\lim_{n\to\infty} P(N_n(A) = 0) = P(N(A) = 0) = \exp(-\Lambda(A))$$

One of the main question

- What are the properties of $\Lambda(A)$?

Back to univariate case : Fréchet margins

Poisson condition

$$\lim_{n \to \infty} nP(X/a_n \in A_x) = \Lambda_x(A_x)$$

with

$$\Lambda_x(A_x) = x^{-1}, \text{ for } A_x = [x, \infty)$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) =$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-x^{-1} - y^{-1})$$

Hence

$$x^{-1} + y^{-1} = \Lambda_x(A_x) + \Lambda_y(A_y) = \Lambda(A)$$

Special cases

The independent case

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-x^{-1} - y^{-1})$$

Hence

$$x^{-1} + y^{-1} = \Lambda_x(A_x) + \Lambda_y(A_y) = \Lambda(A)$$

The general case

$$\Lambda(A) \le \Lambda_x(A_x) + \Lambda_y(A_y)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) =$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\max(1/x, 1/y))$$

Hence,

$$\max(1/x, 1/y) = \max(\Lambda_x(A_x),\ \Lambda_x(A_y)) = \ \Lambda(A)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\max(1/x, 1/y))$$

Hence,

$$\max(1/x, 1/y) = \max(\Lambda_x(A_x),\ \Lambda_x(A_y)) = \Lambda(A)$$

The general case

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \le \Lambda(A)$$

Special cases

The dependent case $X_i = Y_i$

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\max(1/x, 1/y))$$

Hence,

$$\max(1/x, 1/y) = \max(\Lambda_x(A_x),\ \Lambda_x(A_y)) = \Lambda(A)$$

The general case

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \le \Lambda(A)$$

$$\max(\Lambda_x(A_x),\ \Lambda_x(A_y)) \le \Lambda(A) \le \Lambda_x(A_x) + \Lambda_y(A_y)$$

Scaling property

Univariate case with $\Lambda_x(A_x) = x^{-1}$

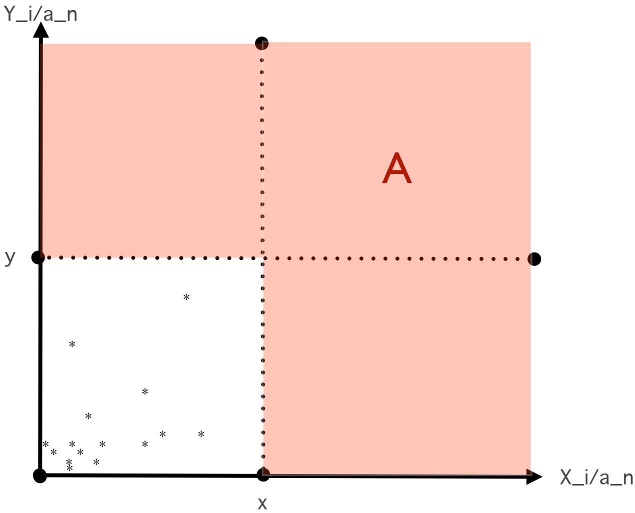
$$\Lambda_x(tA_x) = t^{-1}\Lambda_x(A_x)$$

Bivariate case

$$\Lambda(tA) = t^{-1}\Lambda(A)?$$

Going back to maxima

$$
\begin{aligned}
\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) &= \exp(-\Lambda(A)) \\
&= P(M_X \le x, M_Y \le y)
\end{aligned}
$$



Going back to maxima

$$P(M_X \leq x, M_Y \leq y) = \exp(-\Lambda(A))$$

Scaling

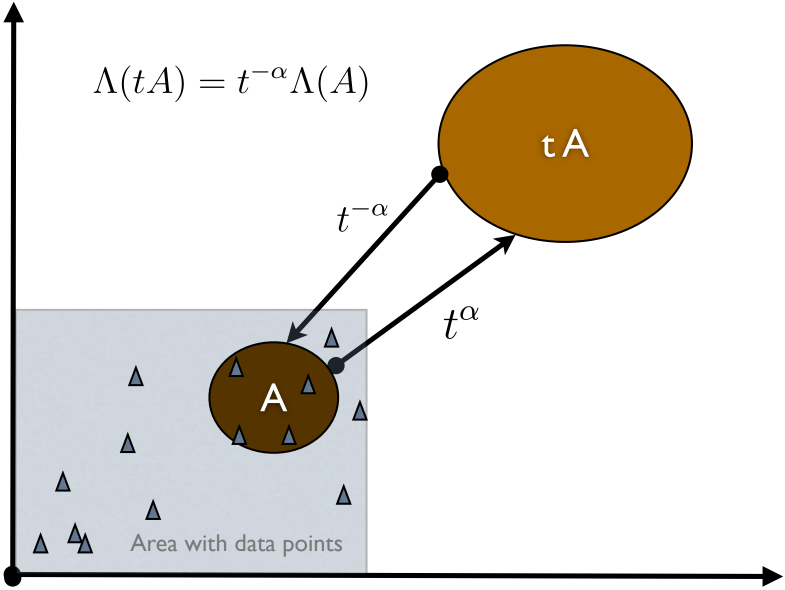
$$\Lambda(tA) = t^{-1}\Lambda(A)$$

is equivalent to

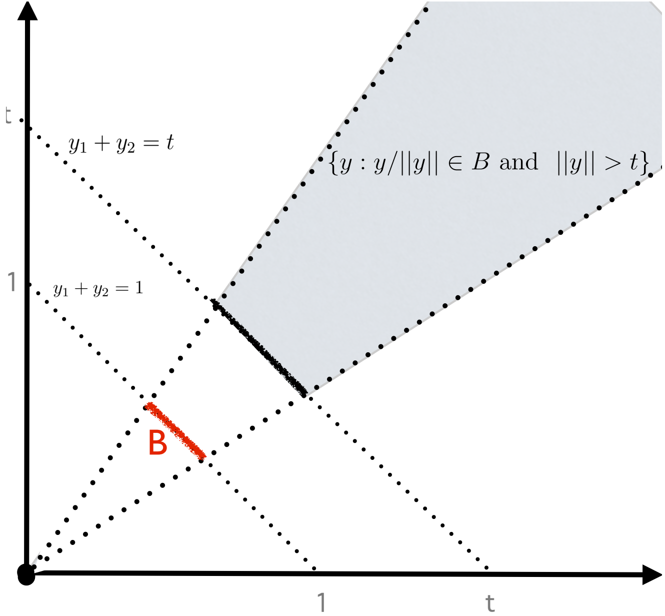
Max-stability

$$\begin{aligned} P^t(M_X \leq t\,x, M_Y \leq t\,y) &= (\exp(-\Lambda(tA)))^t = \exp(-t\Lambda(tA)) \\ &= \exp(-\Lambda(A)) \\ &= P(M_X \leq x, M_Y \leq y) \end{aligned}$$

Scaling property : an essential property of inference



Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$ with $||\mathbf{y}|| = y_1 + y_2$



Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$

A special case

$$A = \{\mathbf{z} = (x, y) : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}$$

where $||\mathbf{z}|| = x + y$ and B any set belonging to the unit sphere

A surprising property

$$\begin{aligned} tA &= \{t\mathbf{z} : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}, \\ &= \{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}, \text{ with } \mathbf{u} = t\mathbf{z}. \end{aligned}$$

This implies

$$\Lambda\left(\{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}\right) = t^{-1} H(B)$$

where $H(.)$ is the mean measure restricted to the unit sphere and often called the spectral measure.

Interpreting the scaling property $\Lambda(tA) = t^{-1}\Lambda(A)$

A special case

$$A = \{\mathbf{z} = (x, y) : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}$$

where $||\mathbf{z}|| = x + y$ and B any set belonging to the unit sphere

A surprising property

$$\begin{aligned} tA &= \{t\mathbf{z} : \mathbf{z}/||\mathbf{z}|| \in B \text{ and } ||\mathbf{z}|| > 1\}, \\ &= \{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}, \text{ with } \mathbf{u} = t\mathbf{z}. \end{aligned}$$

This implies

$$\Lambda\left(\{\mathbf{u} : \mathbf{u}/||\mathbf{u}|| \in B \text{ and } ||\mathbf{u}|| > t\}\right) = t^{-1}H(B)$$

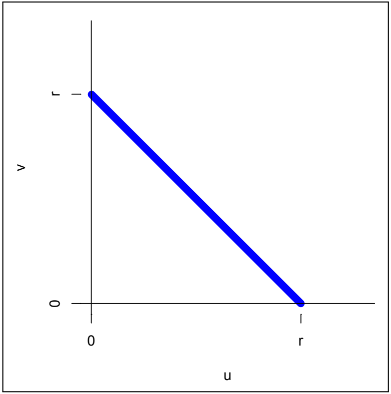
where $H(.)$ is the mean measure restricted to the unit sphere and often called the spectral measure.

Independence between the strength of event $||\mathbf{z}|| = x + y$ and the location on the unit simplex

Polar coordinates

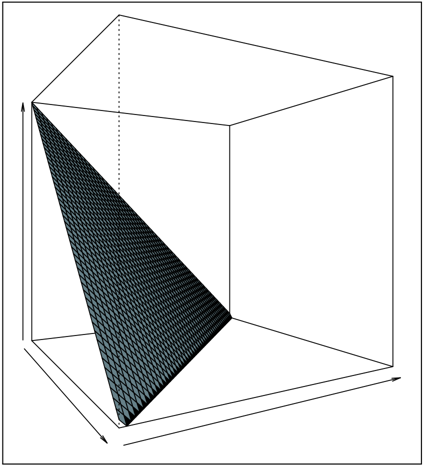
2D

$r = (u + v)$ and
$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



3D

$r = (u + v + w)$,
$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}, \theta_3 = \frac{w}{r}$

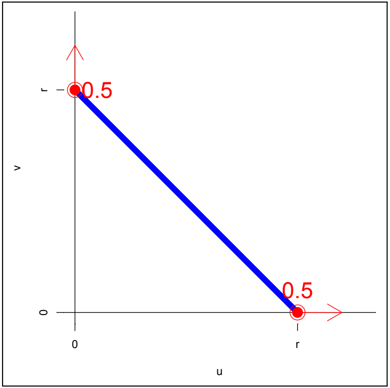


2D Polar coordinates

2D : INDEPENDENT CASE

$r = (u + v)$ and

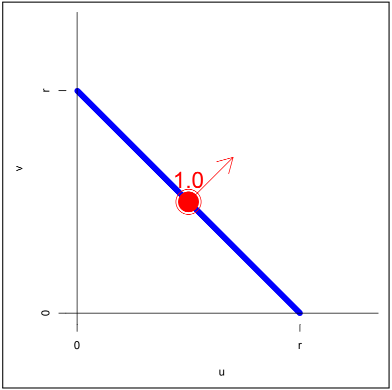
$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



2D : COMPLETE DEPENDENCE

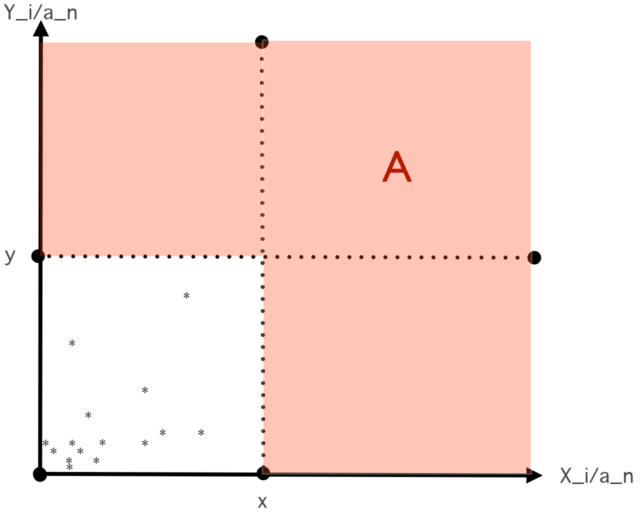
$r = (u + v)$ and

$\theta_1 = \frac{u}{r}, \theta_2 = \frac{v}{r}$



Again, back to maxima

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\Lambda(A))$$



Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = \exp(-\Lambda(A))$$

Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\dots,X_n)/a_n \leq x, \max(Y_1,\dots,Y_n)/a_n \leq y) = \exp(-\Lambda(A))$$

Changing coordinates : $r = u + v$ **and** $w = u/(u+v)$

$$\begin{aligned}(u,v)\notin A \quad &\Leftrightarrow \quad u < x \text{ and } \ v < y,\\ &\Leftrightarrow \quad r < x/w \text{ and } \ r < y/(1-w),\\ &\Leftrightarrow \quad r < \min(x/w, y/(1-w))\end{aligned}$$

Back to maxima

How to express A in

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \leq x, \max(Y_1,\ldots,Y_n)/a_n \leq y) = \exp(-\Lambda(A))$$

Changing coordinates : $r = u + v$ and $w = u/(u+v)$

$$\begin{aligned}
(u,v) \notin A &\Leftrightarrow u < x \text{ and } v < y, \\
&\Leftrightarrow r < x/w \text{ and } r < y/(1-w), \\
&\Leftrightarrow r < \min(x/w, y/(1-w))
\end{aligned}$$

Computing $\Lambda(A)$

$$\begin{aligned}
\Lambda(A) &= \int_{w\in[0,1]} \int_{r>\min(x/w,y/(1-w))} r^{-2} dH(w) \\
&= \int_{w\in[0,1]} \max(w/x, (1-w)/y) dH(w)
\end{aligned}$$

Rewriting the counting rate in function of $H(dw)$

$$\Lambda(A) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) H(dw)$$

Scaling property checked

$$\Lambda(tA) = t^{-1}\Lambda(tA)$$

Max-stable vector

If

$$\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)/a_n \le x, \max(Y_1,\ldots,Y_n)/a_n \le y) = G(x,y)$$

then

$$-\log G(x,y) = \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\, dH(w)$$

where $H(.)$ such that $\int_0^1 w\, dH(w) = 1$

Max-stable vector properties

$$G(x, y) = \exp\left[-\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w)\right]$$

and $H(.)$ such that $\int_0^1 w \, dH(w) = 1$

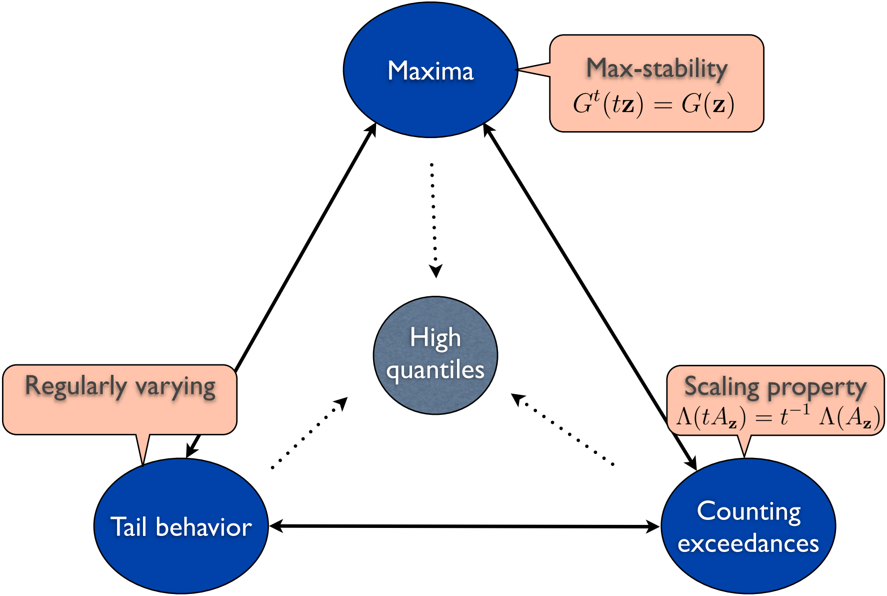
Max-stability

$$G^t(tx, ty) = G(x, y), \text{ for any } t > 0$$

Marginals : unit-Fréchet

$$G(x, \infty) = G(\infty, x) = \exp(-1/x)$$

A multivariate summary



Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of Wired.*

A quick summary about the basics

Learned lessons

- Multivariate maxima can be handled with Poisson counting processes
- "Polar coordinates" allows to see the independence between the strength of the event and the dependence structure that lives on the simplex
- The dependence structure has not explicit expressions (in contrast to the margins and to the Gaussian case)
- Max-stable property = scaling property for the Poisson intensity
- Conceptually easy to go from the bivariate to the multivariate case

Remaining questions

- How to make the inference of the dependence structure ?
- How can we use this dependence structure ?
- No easy regression scheme (how to do D&A, see Francis' talk) ?

Inference

Strategies for either the marginal behavior or the dependence

- Parametric : (+) Reduce dimensionality & easy to deal with covariates (-) impose a parametric form, model selection needed
- Non-parametric : (+) General without strong assumptions, (-) no practical for large dimension (curse of dimensionality), difficult to insert covariates

Inference

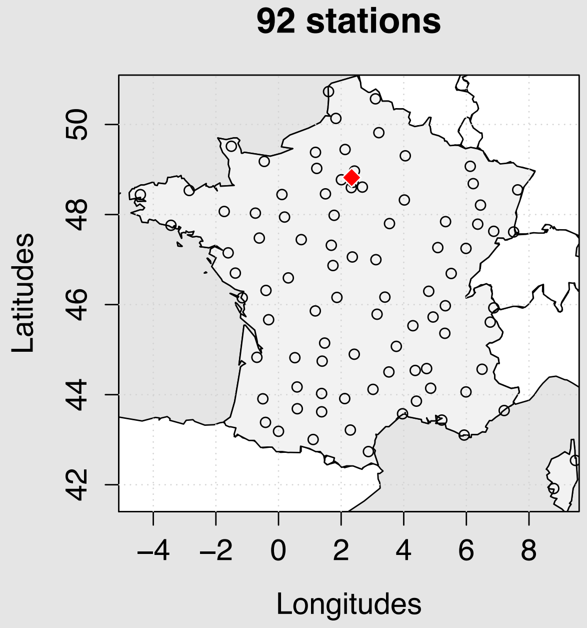
Strategies for either the marginal behavior or the dependence

- Parametric : (+) Reduce dimensionality & easy to deal with covariates (-) impose a parametric form, model selection needed
- Non-parametric : (+) General without strong assumptions, (-) no practical for large dimension (curse of dimensionality), difficult to insert covariates

Techniques

- Maximizing the likelihood : (+) easy to integrate covariates (-) impose a parametric form, no straightforward for large dimension
- Bayesian inference : (+) easy to insert expert knowledge, (-) no straightforward for large dimension (slow)
- Methods of moments : (+) fast and simple to understand, can be non-parametric (-) no straightforward to have covariates

Hourly precipitation in France, 1992-2011 (Olivier Mestre)



Our game plan to handle extremes from this big rainfall dataset

	Spatial scale	
	Large (country)	Local (region)
Problem	Dimension reduction	Spectral density in moderate dimension
Data	Weekly maxima of hourly precipitation	Heavy hourly rainfall excesses
Method	Clustering algorithms for maxima	Mixture of Dirichlet

Without imposing a given parametric structure

Clustering of maxima (joint work with E. Bernard, M. Vrac and O. Mestre)

Task 1

Clustering 92 grid points into around 10-20 climatologically homogeneous groups wrt spatial dependence

Clusterings

Challenges

- Comparing apples and oranges
- An average of maxima (centroid of a cluster) is not a maximum
- variances have to be finite
- Difficult interpretation of clusters

Questions

- How to find an appropriate metric for maxima ?
- How to create cluster centroids that are maxima ?

A central question (assuming that $\mathbb{P}[M(x) < v] = \mathbb{P}[M(y) < u] = \exp(-1/u)$)

$$\mathbb{P}[M(x) < u, M(y) < v] = \exp\left[-\int_0^1 \max\left(\frac{w}{u}, \frac{1-w}{v}\right) \, dH(w)\right]$$

θ = Extremal coefficient

$$\mathbb{P}\left[M(x) < u, M(y) < u\right] = \left(\mathbb{P}\left[M(x) < u\right]\right)^{\theta}$$

Interpretation

- Independence $\Rightarrow \theta = 2$
- $M(x) = M(y) \Rightarrow \theta = 1$
- Similar to correlation coefficients for Gaussian but ...
- No characterization of the **full** bivariate dependence

A L1 marginal free distance (Cooley, Poncet and N., 2005, N. and al., 2007)

$$d(x,y) = \frac{1}{2}\mathbb{E}\left|F_y(M(y)) - F_x(M(x))\right|$$

A L1 marginal free distance (Cooley, Poncet and N., 2005, N. and al., 2007)

$$d(x,y) = \frac{1}{2}\mathbb{E}\,|F_y(M(y)) - F_x(M(x))|$$

If $M(x)$ and $M(y)$ bivariate GEV, then

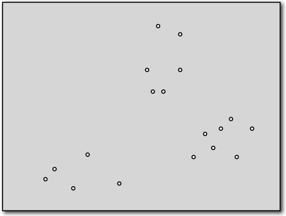
$$\text{extremal coefficient} = \frac{1+2d(x,y)}{1-2d(x,y)}$$

Clusterings

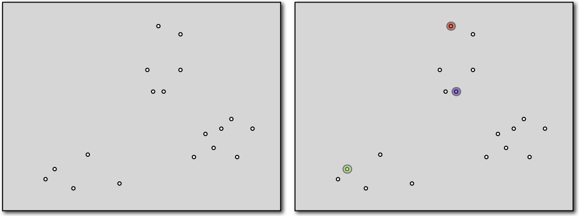
Questions

- How to find an appropriate metric for maxima ?
- **How to create cluster centroids that are maxima** ?

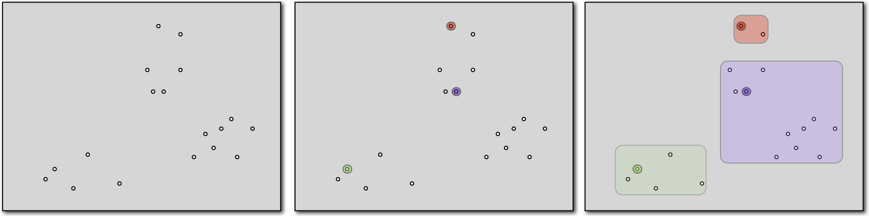
Partitioning Around Medoids (PAM) (Kaufman, L. and Rousseeuw, P.J. (1987))



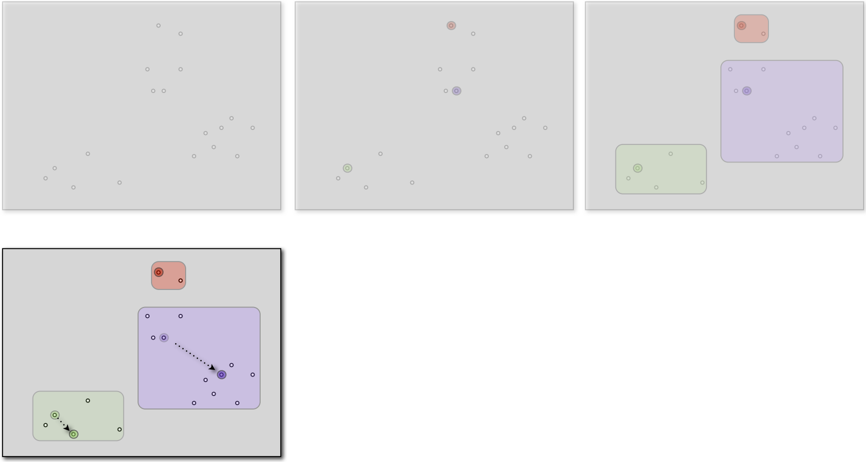
PAM : Choose K initial mediods



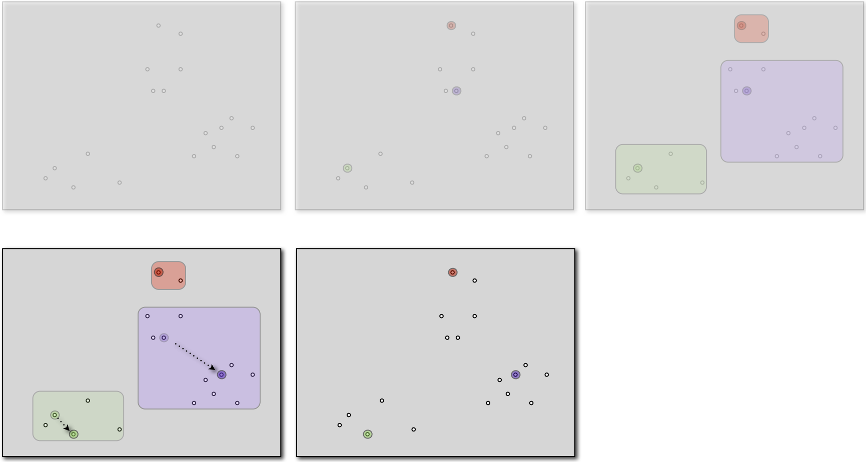
PAM : Assign each point to each closest mediod



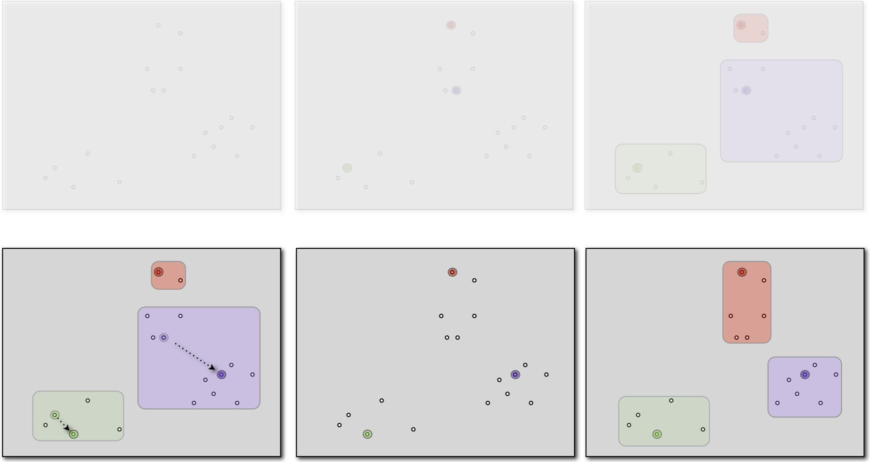
PAM : Recompute each mediod as the gravity center of each cluster



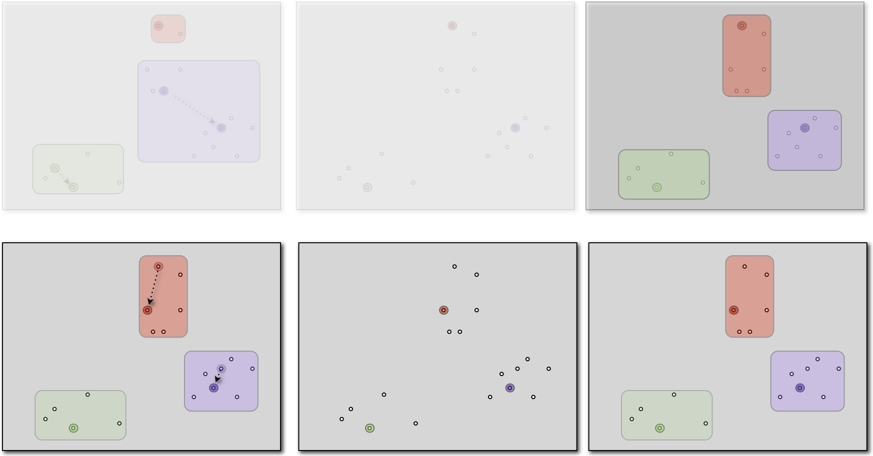
PAM : continue if a mediod has been moved



PAM : Assign each point to each closest mediod

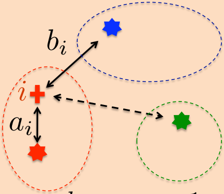


PAM : Recompute each mediod as the gravity center of each cluster



- **Clustering validation**

SILHOUETTE COEFFICIENT

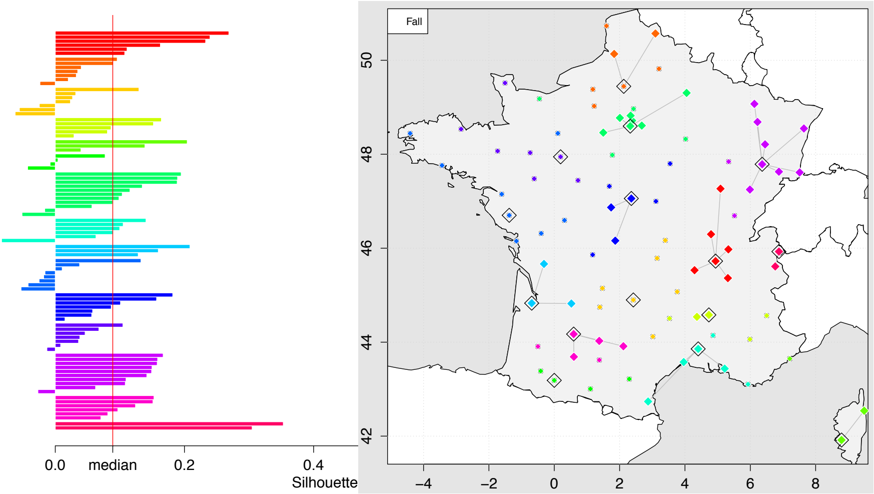


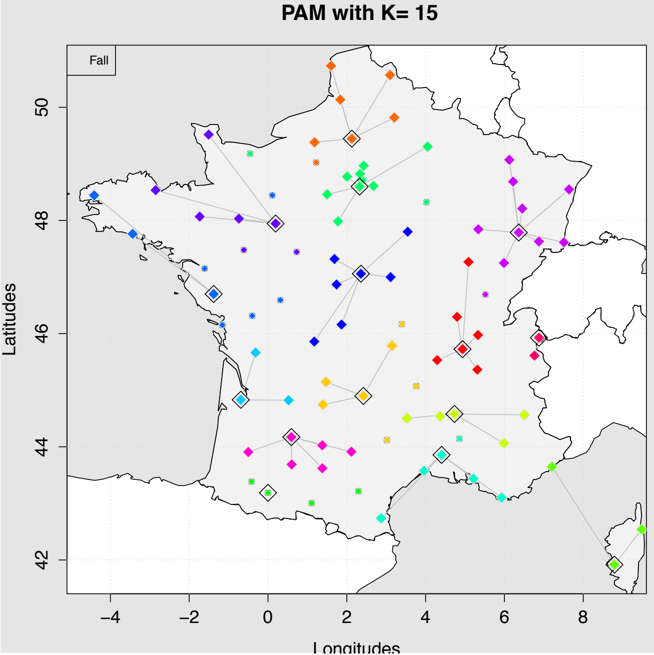
$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

$a_i \ll b_i, \quad s_i \approx 1 \quad \rightarrow$ Well classified

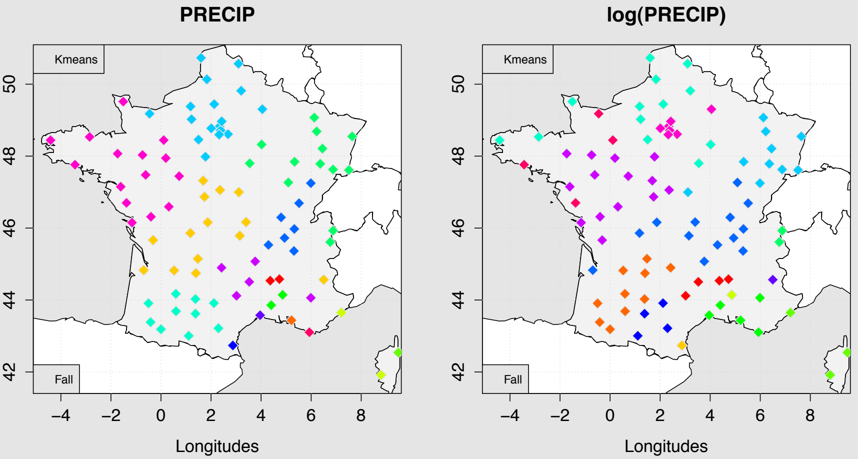
$a_i \sim b_i, \quad s_i \approx 0 \quad \rightarrow$ Neutral

$a_i \gg b_i, \quad s_i \approx -1 \quad \rightarrow$ Badly classified


0.0
median
0.2
0.4
Silhouette
Fall
50
48
46
44
42
−4
−2
0
2
4
6
8


PAM with K= 15
Fall
Latitudes
Longitudes
50
48
46
44
42
−4
−2
0
2
4
6
8

Applying the kmeans algorithm to maxima (15 clusters)



Summary on clustering of maxima

- Classical clustering algorithms (kmeans) are not in compliance with EVT
- Madogram provides a convenient distance that is marginal free and very fast to compute
- PAM applied with mado preserves maxima and gives interpretable results
- R package available on my web site

Project : Dimension reduction (via clustering ?)

- Are clusters of maxima change over time, say pre-industrial, today, future ?
- How robust are clusters of maxima in climate models (is it model sensitive) ?
- Are clusters of maxima different from classical patterns (EOF) ?
- PAM applied with mado preserves maxima and gives interpretable results
- Can we compute the FAR within a given cluster ?
- What about the marginal behavior (the intensity) ?

Data = field of temperature yearly maxima or precipitation (per season ?)

Methods of moments in a non-stationary spatial case [3]

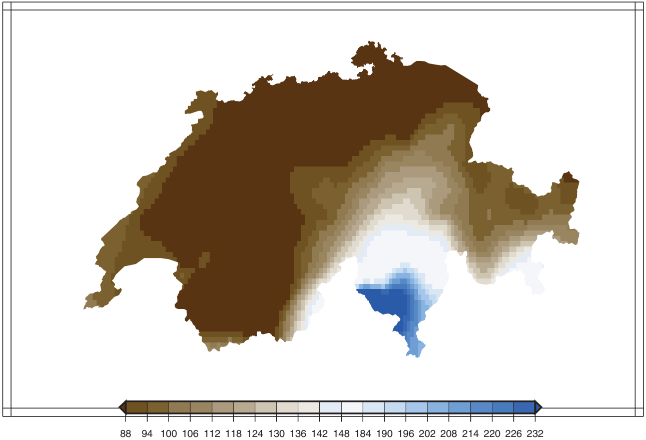


Figure 4. Inferred 50 year return levels in mm for heavy precipitation in Switzerland, see Figure 3.

3. Naveau, Toreti, Smith, Xoplaki, WRR, 2014.Daily precipitation recorded (220 stations) in Switzerland from 2001 to 2010 in autumn. Excesses over the 90^{th} percentile by using a 2-dimensional spatial kernel. To estimate threshold values, universal kriging applied to the station-based thresholds by using elevation as external drift.

Methods of moments in a non-stationary spatial case [4]

Probability Weighted Moments (PWM), see Hoskings and colleagues)

$$\mu_r = \mathbb{E}\left[Z\overline{G}^r(Z)\right]$$

4. Naveau, Toreti, Smith, Xoplaki, WRR, 2014

Methods of moments in a non-stationary spatial case [4]

Probability Weighted Moments (PWM), see Hoskings and colleagues)

$$\mu_r = \mathbb{E}\left[Z\overline{G}^r(Z)\right]$$

PWM for the GPD in the IID case

$$\mu_r = \frac{\sigma}{(1+r)(1+r-\xi)},$$

4. Naveau, Toreti, Smith, Xoplaki, WRR, 2014

Methods of moments in a non-stationary spatial case [4]

Probability Weighted Moments (PWM), see Hoskings and colleagues)

$$\mu_r = \mathbb{E}\left[Z\overline{G}^r(Z)\right]$$

PWM for the GPD in the IID case

$$\mu_r = \frac{\sigma}{(1+r)(1+r-\xi)},$$

PWM and GPD parameters for $\xi < 1.5$

$$\sigma = \frac{2.5\mu_{1.5}\mu_1}{2\mu_1 - 2.5\mu_{1.5}} \text{ and } \xi = \frac{4\mu_1 - (2.5)^2\mu_{1.5}}{2\mu_1 - 2.5\mu_{1.5}}.$$

An estimation of μ_r can be obtained by noticing that $\overline{G}_{\sigma,\xi}(Z)$ follows a uniform distribution on [0, 1].

4. Naveau, Toreti, Smith, Xoplaki, WRR, 2014

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\boldsymbol{x})$ followed a GP($\sigma(\boldsymbol{x}), \xi$)

Now $\sigma(\boldsymbol{x})$ can vary according to a covariate $\boldsymbol{x}$,

$$\mu_r(\boldsymbol{x}) = \mathbb{E}[Y(\boldsymbol{x})\overline{G}^r_{\sigma(\boldsymbol{x}),\xi}(Y(\boldsymbol{x}))],$$

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\boldsymbol{x})$ followed a GP($\sigma(\boldsymbol{x}), \xi$)

Now $\sigma(\boldsymbol{x})$ can vary according to a covariate $\boldsymbol{x}$,

$$\mu_r(\boldsymbol{x}) = \mathbb{E}[Y(\boldsymbol{x})\overline{G}^r_{\sigma(\boldsymbol{x}),\xi}(Y(\boldsymbol{x}))],$$

A simple rewriting

$$\mu_r(\boldsymbol{x}) = \sigma(\boldsymbol{x})\frac{1}{(1+r)(1+r-\xi)} = \sigma(\boldsymbol{x})\mathbb{E}[Z\overline{G}^r_{1,\xi}(Z)],$$

where Z follows GP$(1, \xi)$ distribution.

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\boldsymbol{x})$ followed a GP($\sigma(\boldsymbol{x}), \xi$)

Now $\sigma(\boldsymbol{x})$ can vary according to a covariate $\boldsymbol{x}$,

$$\mu_r(\boldsymbol{x}) = \mathbb{E}[Y(\boldsymbol{x})\overline{G}^r_{\sigma(\boldsymbol{x}),\xi}(Y(\boldsymbol{x}))],$$

A simple rewriting

$$\mu_r(\boldsymbol{x}) = \sigma(\boldsymbol{x})\frac{1}{(1+r)(1+r-\xi)} = \sigma(\boldsymbol{x})\mathbb{E}[Z\overline{G}^r_{1,\xi}(Z)],$$

where Z follows GP$(1, \xi)$ distribution.

A new system

$$\xi = \frac{(1+s)^2 - (1+r)^2\alpha_{rs}}{(1+s) - (1+r)\alpha_{rs}} \text{ and } \sigma(\boldsymbol{x}) = \mu_0(\boldsymbol{x})(1-\xi),$$

with

$$\alpha_{rs} = \frac{\mathbb{E}[Z\overline{G}^r_{1,\xi}(Z)]}{\mathbb{E}[Z\overline{G}^s_{1,\xi}(Z)]}.$$

The only variables depending on $\boldsymbol{x}$ are $\sigma(\boldsymbol{x})$ and $\mu_0(\boldsymbol{x})$.

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\mathbf{x})$ followed a GP($\sigma(\mathbf{x}), \xi$)

Suppose that $\widehat{\mu}_0(\mathbf{x})$ and $\hat{\alpha}$ represent any estimators for $\mu_0(\mathbf{x})$ and α_{rs},

$$\widehat{\xi} = \frac{9 - 4\hat{\alpha}}{3 - 2\hat{\alpha}} \quad \text{and} \quad \widehat{\sigma}(\mathbf{x}) = \widehat{\mu}_0(\mathbf{x})(1 - \widehat{\xi})$$

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\boldsymbol{x})$ followed a GP($\sigma(\boldsymbol{x}), \xi$)

Suppose that $\widehat{\mu}_0(\boldsymbol{x})$ and $\hat{\alpha}$ represent any estimators for $\mu_0(\boldsymbol{x})$ and α_{rs},

$$\widehat{\xi} = \frac{9 - 4\hat{\alpha}}{3 - 2\hat{\alpha}} \quad \text{and} \quad \widehat{\sigma}(\boldsymbol{x}) = \widehat{\mu}_0(\boldsymbol{x})(1 - \widehat{\xi})$$

A kernel regression approach for $\widehat{\mu}_0(\boldsymbol{x})$

Let K be a weighting Kernel, e.g. a standard Gaussian pdf, we set

$$\widehat{\mu}_0(\boldsymbol{x}) = \frac{1}{\sum_i K(\boldsymbol{x} - \boldsymbol{x}_i)} \sum_{i=1}^{n} Y(\boldsymbol{x}_i)\, K(\boldsymbol{x} - \boldsymbol{x}_i).$$

Methods of moments in a non-stationary spatial case

Non-stationary case with $Y(\boldsymbol{x})$ followed a GP($\sigma(\boldsymbol{x}), \xi$)

Suppose that $\widehat{\mu}_0(\boldsymbol{x})$ and $\hat{\alpha}$ represent any estimators for $\mu_0(\boldsymbol{x})$ and α_{rs},

$$\widehat{\xi} = \frac{9 - 4\hat{\alpha}}{3 - 2\hat{\alpha}} \quad \text{and} \quad \widehat{\sigma}(\boldsymbol{x}) = \widehat{\mu}_0(\boldsymbol{x})(1 - \widehat{\xi})$$

A kernel regression approach for $\widehat{\mu}_0(\boldsymbol{x})$

Let K be a weighting Kernel, e.g. a standard Gaussian pdf, we set

$$\widehat{\mu}_0(\boldsymbol{x}) = \frac{1}{\sum_i K(\boldsymbol{x} - \boldsymbol{x}_i)} \sum_{i=1}^{n} Y(\boldsymbol{x}_i)\, K(\boldsymbol{x} - \boldsymbol{x}_i).$$

Estimation of α_{rs}

Replace the unobserved Z_i's by their estimated renormalized version $Z_i' = Y(\boldsymbol{x}_i)/\widehat{\mu}_0(\boldsymbol{x}_i)$. Then, simply use your favorite inference PWM methods to estimate $\mathbb{E}[Z'\overline{G'}^{r}_{1,\xi}(Z')]$ for $r = 1, 2$.

Simulations

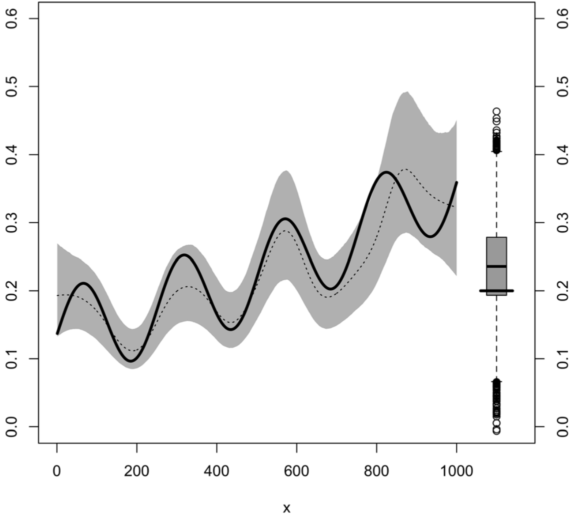


Figure 1. For a $GPD(\sigma(x), \xi)$, the solid black line represents the true scale parameter $\sigma(x)$ in function of *x* (*x* axis). The shape parameter is constant and equals to 0.2 (right axis). From one realization, the boxplot and the gray 90% confidence intervals represent the estimated shape and scale (left axis) obtained by resampling, respectively.

Simulations

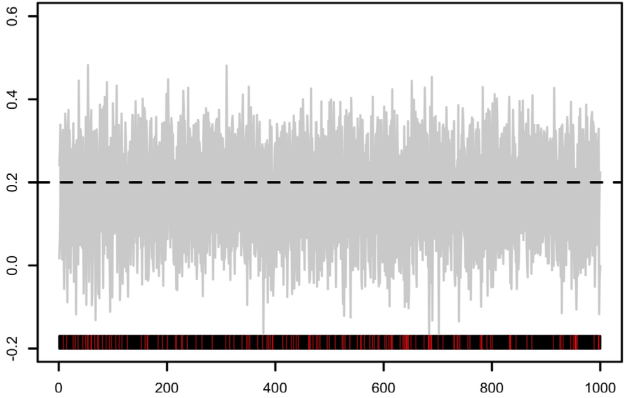


Figure 2. Estimated shape parameter (*y* axis) from 1000 replicas (*x* axis) based on the setup described in Figure 1. The vertical red lines correspond to the samples outside of the estimated 90% coverage probability. As expected for 1000 replicas, around 100 false positive (red lines) occurrences are detected.

Daily precipitation recorded in Switzerland 2001-2010 Autumn ($u = 90^{th}$)

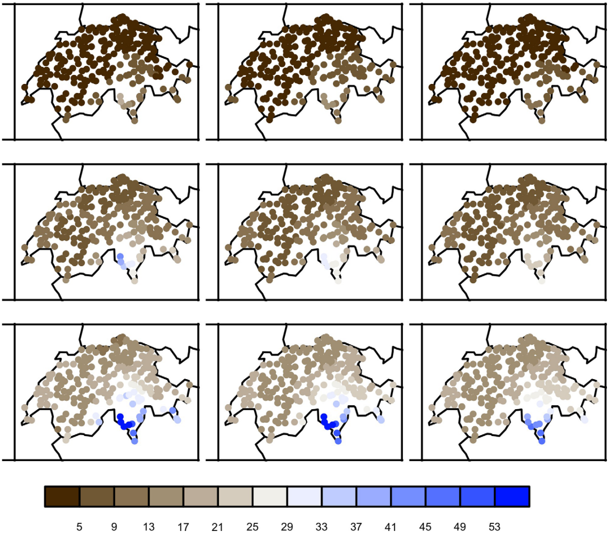


Figure 3. Inferred scale parameter obtained from heavy precipitation (i.e., threshold at the 90% quantile of wet days) recorded at 220 stations in Switzerland from 2001 to 2010 in autumn. The top, middle, and bottom rows correspond to the 5%, median, and 95% values, respectively. The columns from the left represent three different bandwidths, 0.3, 0.5, and 0.7, respectively.

Heavy rainfall in Switzerland

Pros and cons about the inference

- Parametric structure with a GPD : (+) Reduce dimensionality & easy to deal with covariates (-) impose a parametric form, model selection needed
- Non-parametric for the scale parameter
- (+) Fast and conceptually easy (method of moments)
- (-) Independent assumption
- (-/+) Constant shape parameter

Bayesian inference with hidden structures

Notations

- Model = statistical model
- Data $y = (y_1, \ldots, y_n)$
- Hidden signal $x = (x_1, \ldots, x_n)$

Problems at hand

- Model $[y|x]$, the likelihood distribution
- Choose $[x]$ the prior
- Model $[x_t|x_{t-1}]$, the dynamical part of the unobserved system
- **Find $[x|y]$ the inverse probability (posterior)**

A classical and old problem

The problem

- **Find $[x|y]$ the inverse probability (posterior)**

Different names

- Statistical data assimilation
- Statistical inverse problem
- Latent variables
- Filtering methods (Kalman, particles, etc)
- State-space modeling
- Bayesian hierarchical model
- Mixed models

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)

Bayesian vs frequentist statistics

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

Frequentist statistics

- Trust your data and your model
- Find estimators of $[x|y]$ by maximizing the likelihood $[y|x]$ (if necessary, penalize it with prior $[x]$)

Bayesian statistics

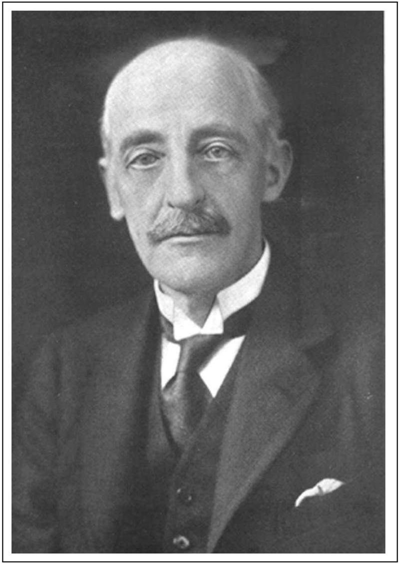
- **Find and trust expert information (independent of our data) through prior** $[x]$
- Trust your data and your model
- Update your expert information via the data, i.e. find posterior $[x|y]$ by using $[x|y] \propto [y|x][x]$

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) Gilbert Walker*
- *(B) Ed Lorenz*
- *(C) Rol Madden*
- *(D) Francis Zwiers*

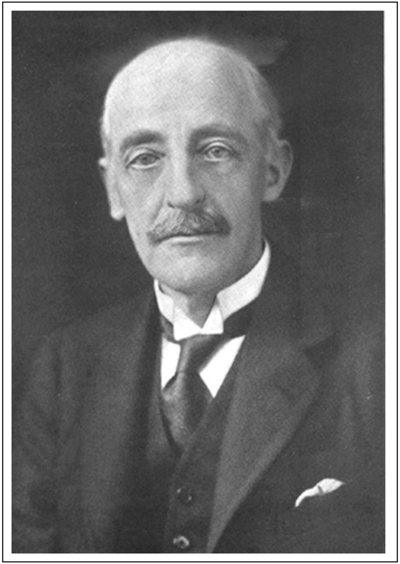


Statistics and Earth sciences

"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) Gilbert Walker*
- *(B) Ed Lorenz*
- *(C) Rol Madden*
- *(D) Francis Zwiers*



Bayesian approach

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

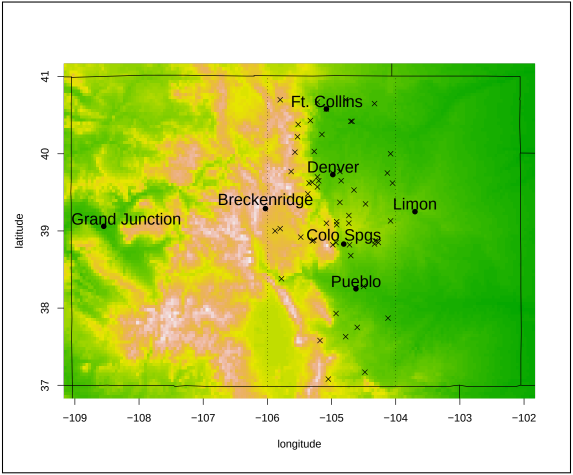
Advantages

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

Drawbacks

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

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



Precipitation in Colorado's front range

Data

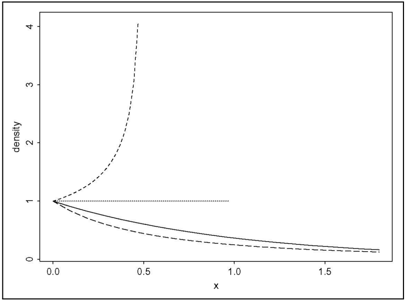
- 56 weather stations in Colorado (semi-arid and mountainous region)
- Daily precipitation for the months April-October
- Time span = 1948-2001
- Not all stations have the same number of data points
- Precision : 1971 from 1/10th of an inche to 1/100

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

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.

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

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

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 three levels

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

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

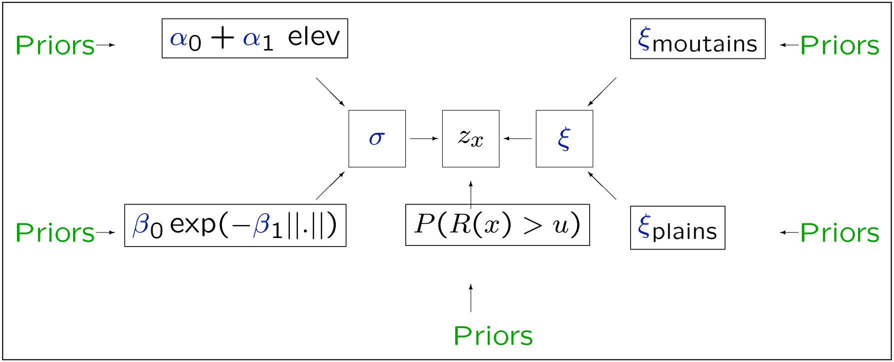
B) **Process layer :=** $\mathbb{P}(\text{process}|\text{parameters})$:

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

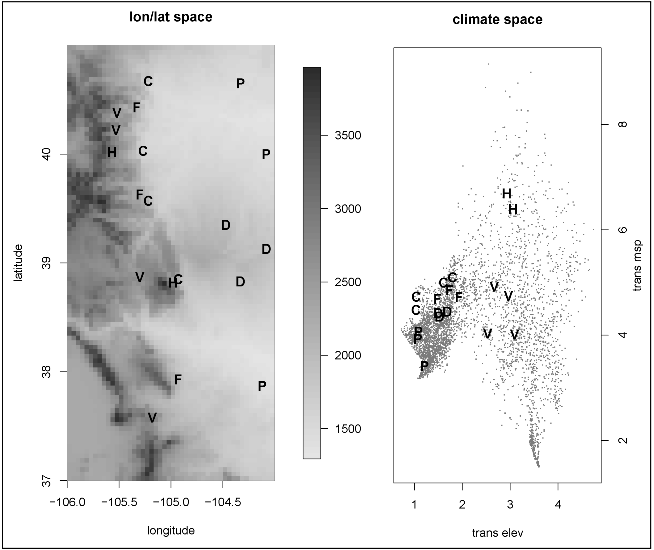
$$\text{and } \xi_i \quad = \quad \begin{cases} \xi_{\text{moutains}}, \text{ if } x_i \in \text{mountains} \\ \xi_{\text{plains}}, \text{ if } x_i \in \text{plains} \end{cases}$$

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

Bayesian hierarchical modeling

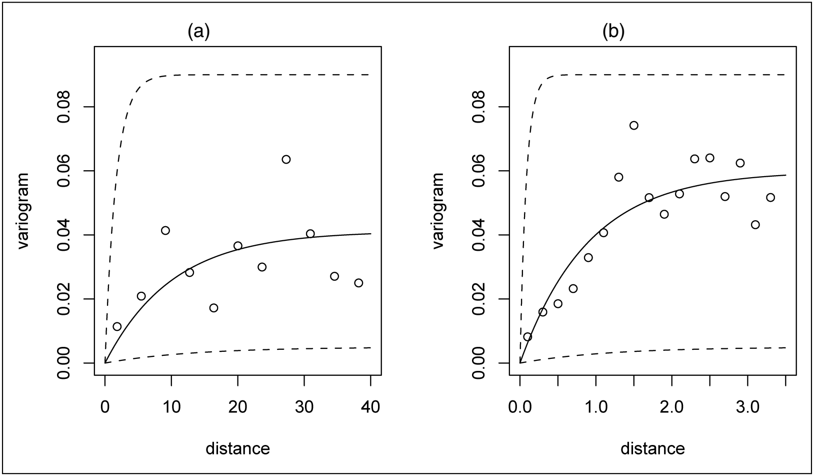


Climate space



foothill cities(C),plains(P),Palmer Divide(D),Front Range(F),mountain valley(V),and high elevation(H)

Priors for the spatial compoment



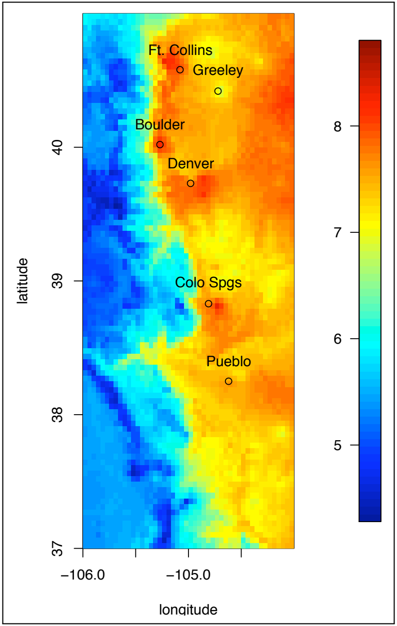
Traditional Space (a) & Climate Space (b). The dashed lines denote the envelope of possible variograms given the sill and range priors

Model selection

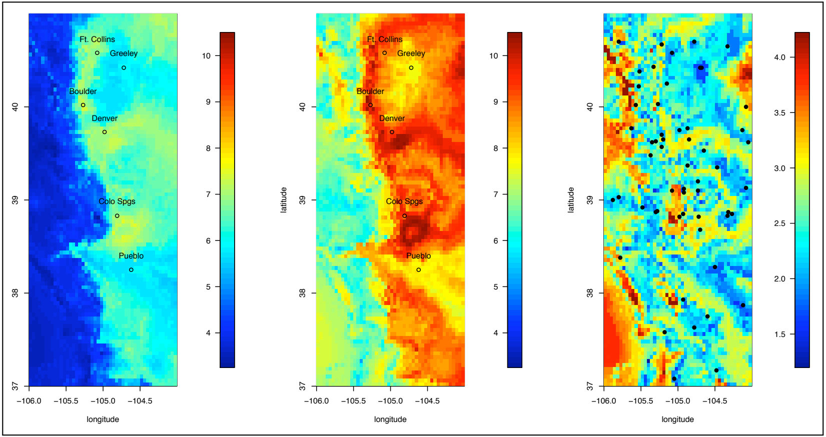
Baseline model		$\bar{D}$	p_D	*DIC*
Model 0:	$\phi = \phi$ $\xi = \xi$	73,595.5	2.0	73,597.2
Models in latitude/longitude space		$\bar{D}$	p_D	*DIC*
Model 1:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,442.0	40.9	73,482.9
Model 2:	$\phi = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	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)



Take-home messages for this rainfall application

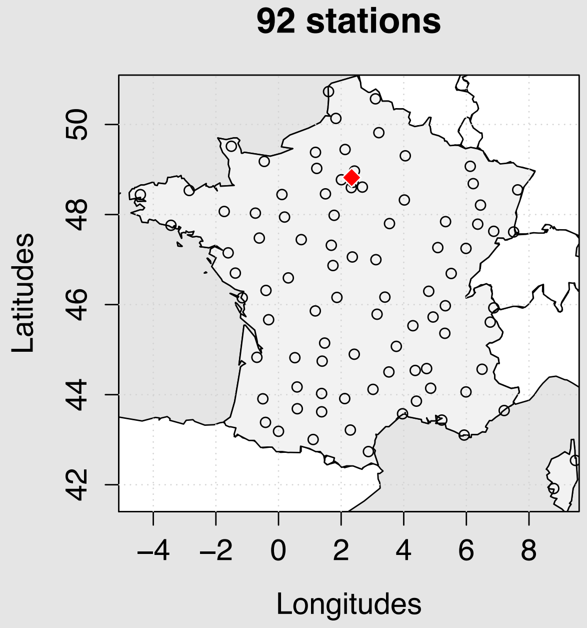
Positive points

- Take advantage of Extreme Value Theory
- Spatial dependencies are captured within the process layer
- The hierarchical Bayesian framework provides a rich and flexible family for modeling complex data sets

Drawbacks

- Computer-intensive implementation (MCMC)
- Difficulty to set the “spatial” priors
- Conditional independence of the observations

Hourly precipitation in France, 1992-2011 (Olivier Mestre)



Our game plan to handle extremes from this big rainfall dataset

	Spatial scale	
	Large (country)	Local (region)
Problem	Dimension reduction	Spectral density in moderate dimension
Data	Weekly maxima of hourly precipitation	Heavy hourly rainfall excesses
Method	Clustering algorithms for maxima	Mixture of Dirichlet

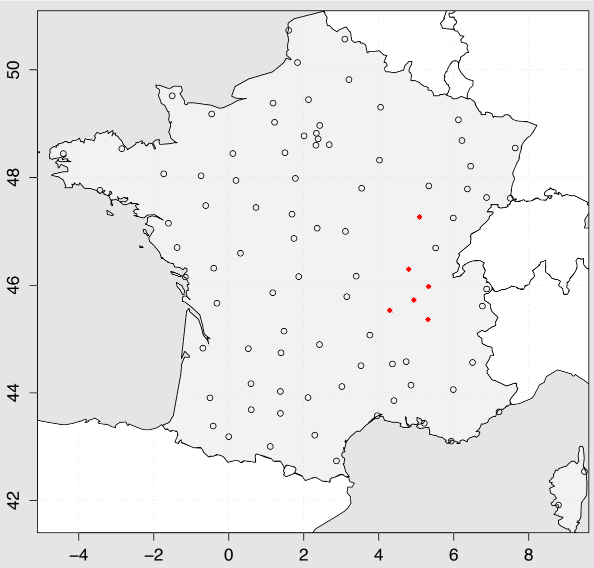
Without imposing a given parametric structure

Our game plan to handle extremes from this rainfall dataset

Spatial scale

	Large (country)	Local (region)
Problem	Dimension reduction	Spectral density in moderate dimension
Data	Weekly maxima of hourly precipitation	Heavy hourly rainfall excesses
Method	Clustering algorithms for maxima	Mixture of Dirichlet

Back to the cluster



Bayesian Dirichlet mixture model for multivariate excesses (joint work with A. Sabourin)

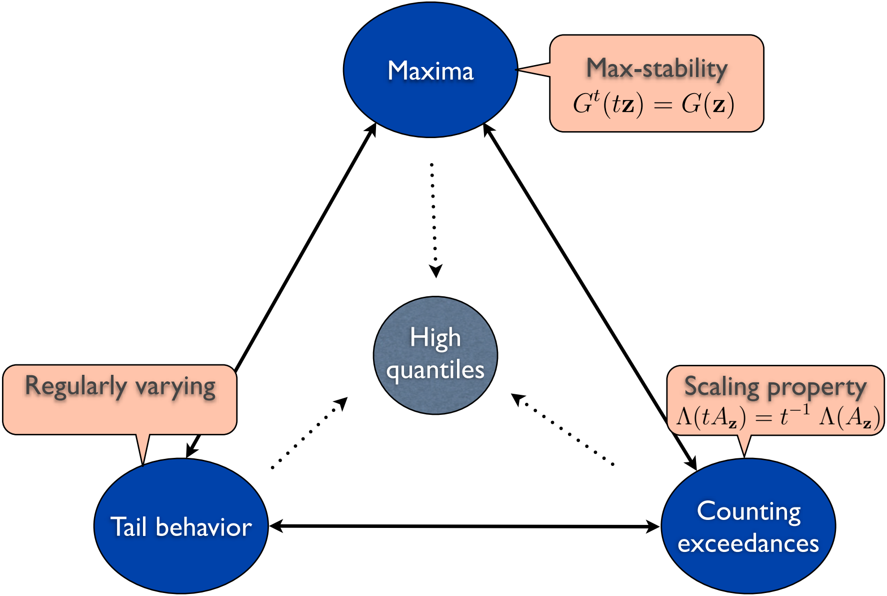
Meteo-France data

Wet hourly events at the regional scale (temporally declustered) of moderate dimensions (from 2 to 8)

Task 2

Assessing the dependence among rainfall excesses

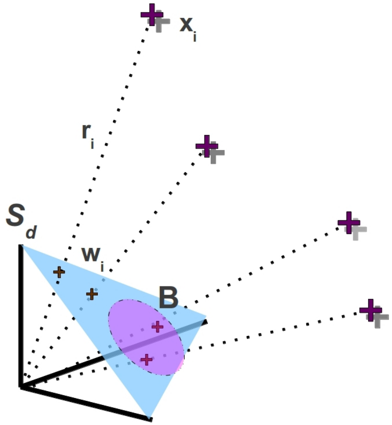
Multivariate Extreme Value Theory (de Haan, Resnick and others)



Defining radius and angular points

Example with $d = 3$ and $\mathbf{X} = (X_1, X_2, X_3)$ such that $\mathbf{P}(X_i < x) = e^{\frac{-1}{x}}$

$$\text{Simplex } \mathbf{S}_3 = \left\{ \mathbf{w} = (w_1, w_2, w_3) : \sum_{i=1}^{3} w_i = 1,\ w_i \geq 0 \right\}.$$



Mathematical constraints on the distribution of the angular points H

$$\mathbf{P}\left(\mathbf{W} \in B, R > r\right) \underset{r \to \infty}{\sim} \frac{1}{r} H(B)$$

Features of H

- H can be non-parametric
- The gravity center of H has to be centered on the simplex

$$\forall i \in \{1, \dots, d\}, \; \int_{\mathbf{S}_d} w_i \, \mathrm{d}H(\mathbf{w}) = \frac{1}{d}$$

A few references on Bayesian non-parametric and semi-parametric spectral inference

M.-O. Boldi and A. C. Davison.
A mixture model for multivariate extremes.
JRSS : Series B (Statistical Methodology), 69(2) :217–229, 2007.

S. Guillotte, F. Perron, and J. Segers.
Non-parametric bayesian inference on bivariate extremes.
JRSS : Series B (Statistical Methodology), 2011.

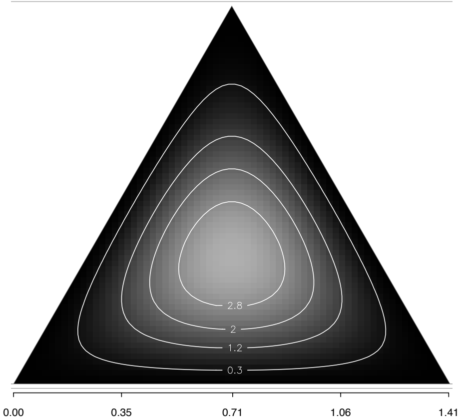
A. Sabourin and P. Naveau.
Bayesian Drichlet mixture model for multivariate extremes.
CSDA, 2013, in press.

P.J. Green.
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.
Biometrika, 82(4) :711, 1995.

Roberts, G.O. and Rosenthal, J.S.
Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains
The Annals of Applied Probability,16,4,2123 :2139, 2006.

Dirichlet distribution

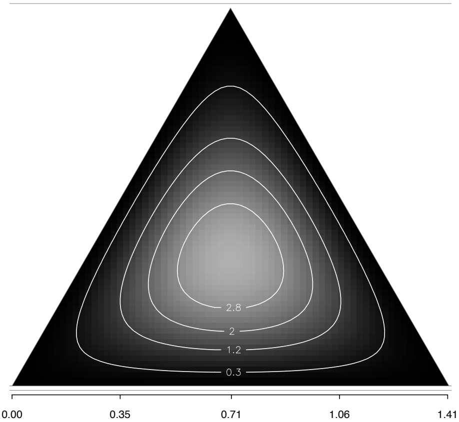
$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d,\ \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^d \Gamma(\nu\mu_i)} \prod_{i=1}^d w_i^{\nu\mu_i - 1}\,.$$



$\boldsymbol{\mu} = (1/3, 1/3, 1/3)$ and $\nu = 9$

Dirichlet distribution

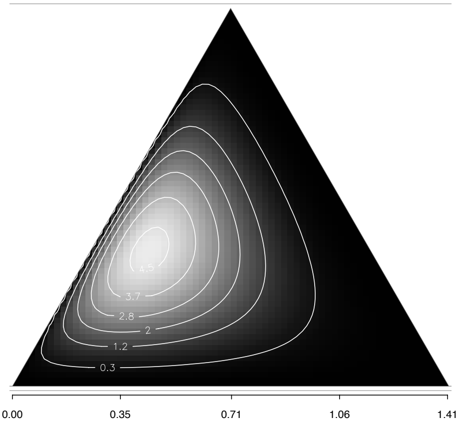
$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d,\ \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^d \Gamma(\nu\mu_i)} \prod_{i=1}^d w_i^{\nu\mu_i - 1}\,.$$



$\boldsymbol{\mu} = (1/3, 1/3, 1/3)$ and $\nu = 9$

Dirichlet distribution

$$\forall \mathbf{w} \in \mathring{\mathbf{S}}_d,\ \text{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^{d} \Gamma(\nu\mu_i)} \prod_{i=1}^{d} w_i^{\nu\mu_i - 1}.$$



$\boldsymbol{\mu} = (.15, .35, .05)$ and $\nu = 9$

But this one is not centered ! !

Mixture of Dirichlet distribution

Boldi and Davision, 2007

$$h_{(\boldsymbol{\mu},\mathbf{p},\boldsymbol{\nu})}(\mathbf{w}) = \sum_{m=1}^{k} p_m \, \mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}_{\cdot\,,\,m}, \nu_m)$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot\,,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$

Mixture of Dirichlet distribution

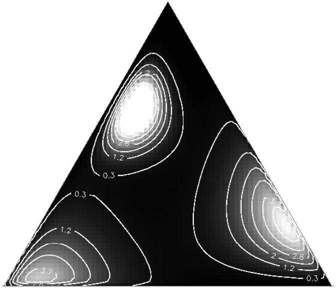
Boldi and Davision, 2007

$$h_{(\boldsymbol{\mu},\mathbf{p},\boldsymbol{\nu})}(\mathbf{w}) = \sum_{m=1}^{k} p_m \,\mathrm{diri}(\mathbf{w} \mid \boldsymbol{\mu}_{\cdot\,,\,m}, \nu_m)$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot\,,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$

Constraint on (μ, p)

$$p_1\,\boldsymbol{\mu}_{.,1} + \cdots + p_k\,\boldsymbol{\mu}_{.,k} = \left(\tfrac{1}{d}, \ldots, \tfrac{1}{d}\right)$$



Inference of Dirichlet density mixtures

Boldi and Davison (2007)

Prior of $[\mu | \mathbf{p}]$ defined on the set

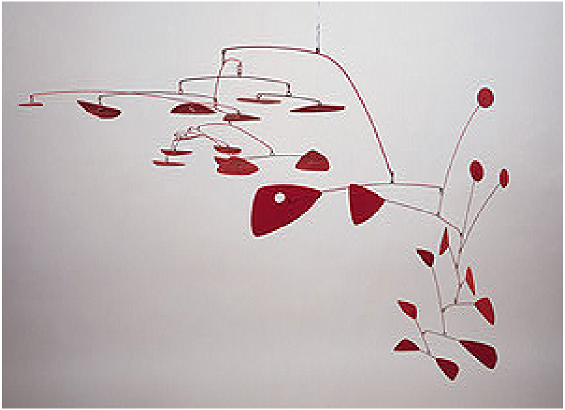
$$p_1 \boldsymbol{\mu}_{.,1} + \cdots + p_k \boldsymbol{\mu}_{.,k} = (\tfrac{1}{d}, \ldots, \tfrac{1}{d})$$

- Sequential inference : first **p**, then $\boldsymbol{\mu}$ one coordinate after the other
- skewed, not interpretable, slow sampling
- Difficult inference in dimension > 3

Inference of Dirichlet density mixtures

How to build priors for $(\mathbf{p}, \mu)$ such that

$$p_1 \boldsymbol{\mu}_{.,1} + \cdots + p_k \boldsymbol{\mu}_{.,k} = \left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$$



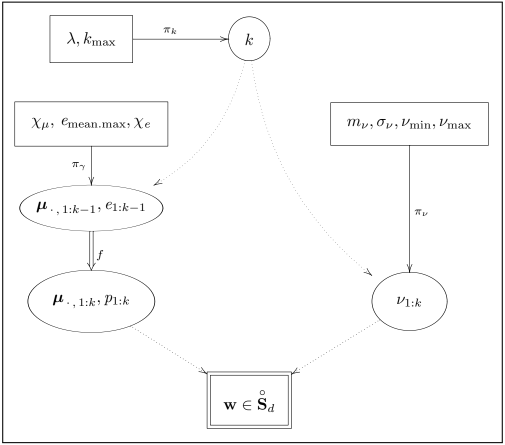
Unconstrained Bayesian modeling for $\Theta = \coprod_{k=1}^{\infty} \Theta_k; \quad \Theta_k = \left\{ (\mathbf{S}_d)^{k-1} \times [0,1)^{k-1} \times (0, \infty]^{k-1} \right\}$

Prior

$$k \sim \text{Truncated geometric}$$
$$\boldsymbol{\mu}_{.,m} | (\boldsymbol{\mu}_{.,1:m-1}, e_{1:m-1}) \sim \text{Dirichlet}$$
$$e_m | (\boldsymbol{\mu}_{.,1:m}, e_{1:m-1}) \sim \text{Beta}$$
$$\nu_m \sim \text{logN}$$

Posterior sampling : MCMC reversible jumps

Summary of the Bayesian scheme



Summary of the Bayesian schemes

Boldi and Davison (2012) Our approach

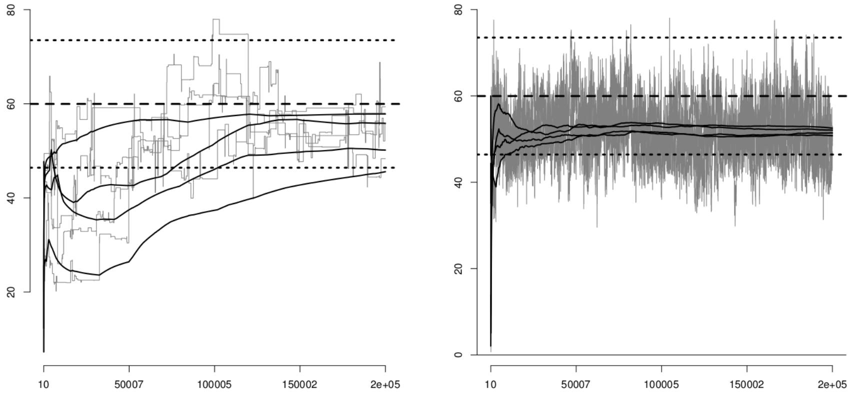
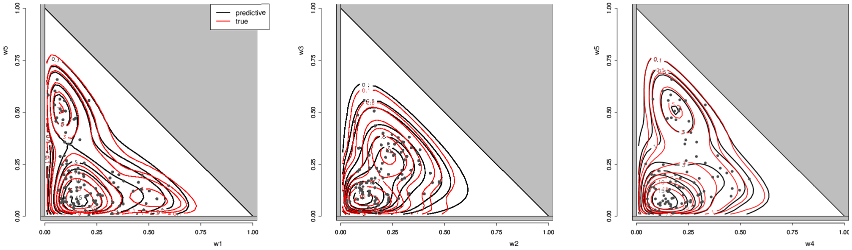


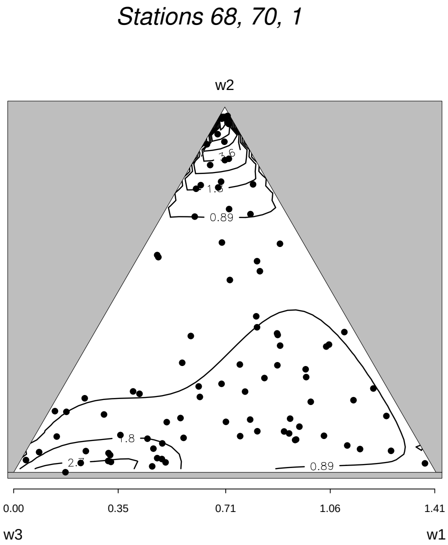
Figure 5: Convergence monitoring with five-dimensional data in the original DM model (left panel) and in the re-parametrized v
with four parallel chains in each model. Grey lines: Evolution of $\langle g, h_{\theta_t(j)} \rangle$. Black, solid lines: cumulative mean. Dashed line

Simulation example with $d = 5$ and $k = 3$



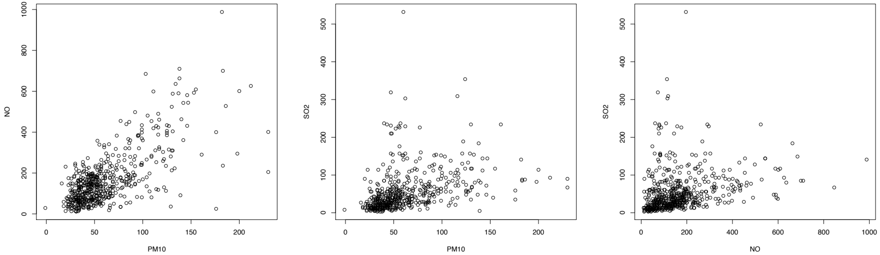
$$T_2 = 150\,10^3, \; T_1 = 50\,10^3.$$

Back to our excesses of the "Lyon" cluster



Coming back to Leeds

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)



Coming back to Leeds

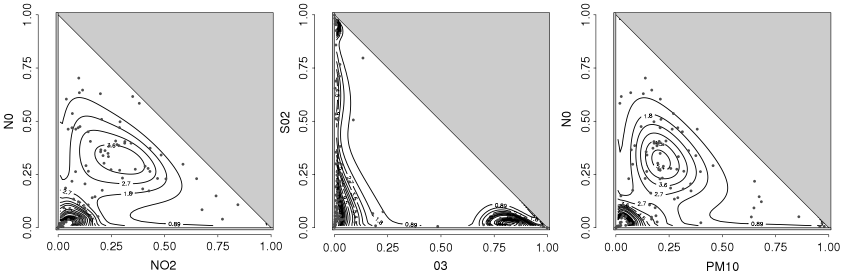


Fig. 6. Five dimensional Leeds data set: posterior predictive density. Black lines: projections of the predictive angular density defined on the four-dimensional simplex $\mathbf{S}_5$ onto the two-dimensional faces. Gray dots: projections of the 100 points with greatest L^1 norm.

Take home messages

Conclusions

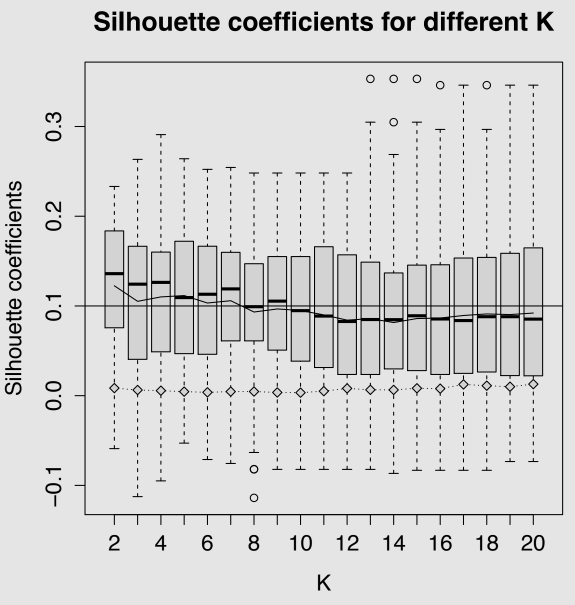
- Clustering of weekly maxima with PAM is fast and gives spatially coherent structures
- Bayesian semi-parametric mixture can handle moderate dimensions and provide credibility intervals

Going further

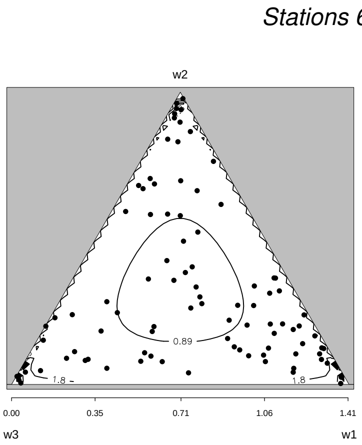
- Anne Sabourin = a Bayesian semi-parametric mixture for censored data with an application to paleo-flood data

References

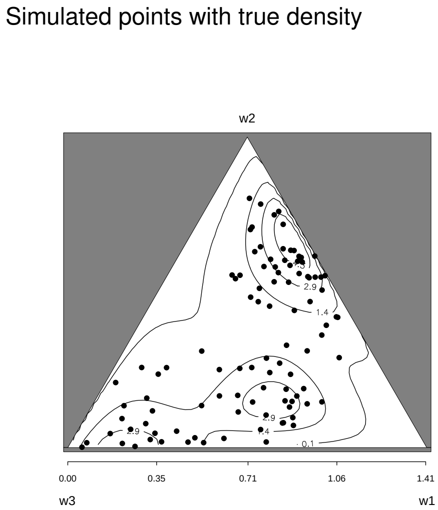
- Bernard, E., et al.. Clustering of maxima : Spatial dependencies among heavy rainfall in france. Journal of Climate, 2013, [**R package**].
- Sabourin, A. , Naveau, P. Dirichlet Mixture model for multivariate extremes. To appear in Computational Statistics and Data Analysis. [**R package**].
- Naveau P. et al., Modeling Pairwise Dependence of Maxima in Space. Biometrika, (2009)

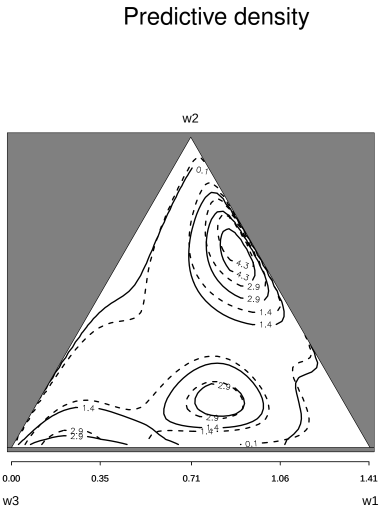

Silhouette coefficients for different K
Silhouette coefficients
0.3
0.2
0.1
0.0
−0.1
2
4
6
8
10
12
14
16
18
20
K

Different results from different Monte Carlo chains ?

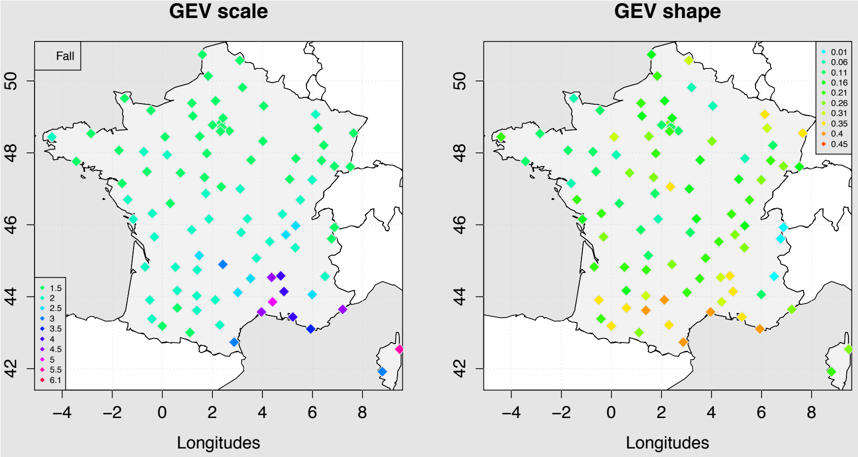


Simulation example with $d = 3$ and $k = 3$





The scale and shape GEV parameters

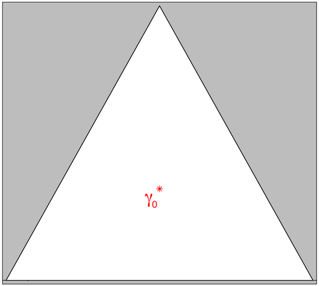


Take home messages from part I

- Extremes here means very rare
- It is possible to estimate the dependence between bivariate extremes
- Multivariate EVT may help characterizing extremes dependencies in space or time
- Modeling trade off between parametric and non-parametric approaches
- Challenges to go beyond the bivariate case and to have flexible parametric models

New parametrisation

Ex : $k = 4$ and $d = 3$

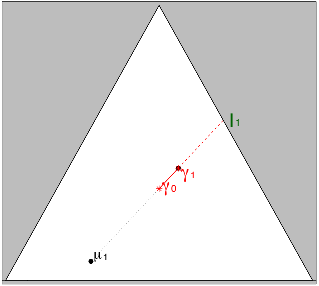


γ_m : "Equilibrium" centers built from $\boldsymbol{\mu}_{\cdot,m+1}, \ldots, \boldsymbol{\mu}_{\cdot,k}$.

$$\gamma_m = \sum_{j=m+1}^{k} \frac{p_j}{p_{m+1} + \cdots + p_k} \boldsymbol{\mu}_{\cdot,j}$$

New parametrisation

Ex : $k = 4$ and $d = 3$

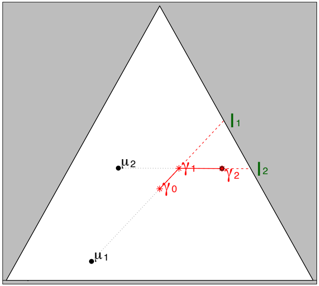


$$\boldsymbol{\mu}_{.,1},\ e_1 \quad \Rightarrow \gamma_1 : \frac{\overline{\gamma_0\,\gamma_1}}{\overline{\gamma_0\, I_1}} = e_1\ ;$$

$$\Rightarrow p_1$$

New parametrisation

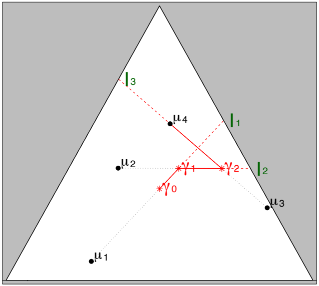
Ex : $k = 4$ and $d = 3$



$$\mu_{.,2},\ e_2 \quad \Rightarrow \gamma_2 : \frac{\overline{\gamma_1\,\gamma_2}}{\overline{\gamma_1\,l_2}} = e_2\ ;$$
$$\Rightarrow p_2$$

New parametrisation

Ex : $k = 4$ and $d = 3$

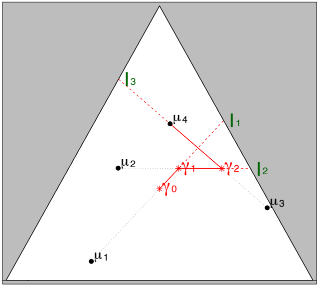


$$\boldsymbol{\mu}_{.,3},\ e_3 \quad \Rightarrow \gamma_3 : \frac{\overline{\gamma_2\,\gamma_3}}{\overline{\gamma_2\,I_3}} = e_3\ ; \quad \boldsymbol{\mu}_{.,4} = \gamma_3.$$

$$\Rightarrow p_3, p_4$$

New parametrisation

***Ex*: $k = 4$ and $d = 3$**



Parametrisation of h with $\theta = (\boldsymbol{\mu}_{.,1:k-1}, \mathbf{e}_{1:k-1}, \nu_{1:k})$

$(\boldsymbol{\mu}_{.,1:k-1}, e_{1:k-1})$ gives $(\boldsymbol{\mu}_{.,1:k}, p_{1:k})$