Applying Multivariate Extreme Value Theory to Environmental Data

Philippe Naveau naveau@lsce.ipsl.fr

Laboratoire des Sciences du Climat et l'Environnement (LSCE)
Gif-sur-Yvette, France
joint work with Dan Cooley and Richard Davis

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

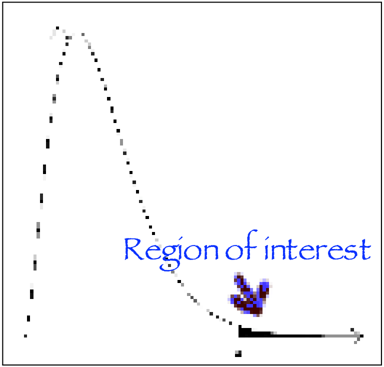
13 décembre 2010

- *"We anticipated as far as possible but one cannot forecast the unforeseeable"*
  Xynthia's storm in France, 25 Feb 2010
- *" It is impossible that the improbable never occurs"*
  Emil Julius Gumbel (1891-1966)

Oscar Wilde perspective

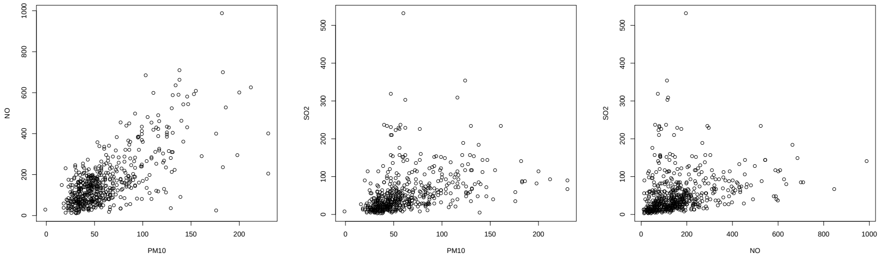
1. *"Man can believe the impossible, but man can never believe the improbable"* Oscar Wilde (Intentions, 1891)

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



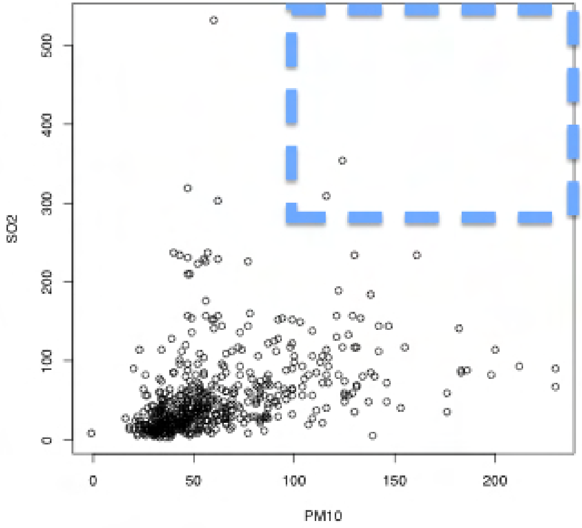
An example

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)



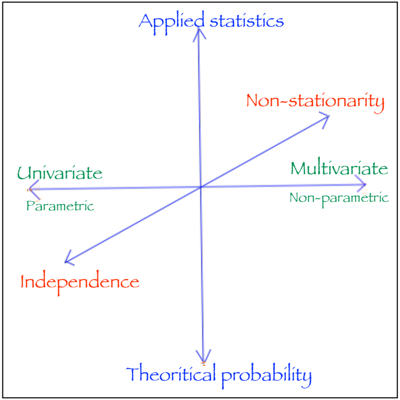
Typical question

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



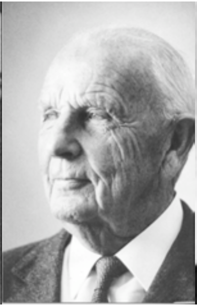
A few facts about Extreme Value Theory

- 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



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

Max-stability

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

Definition : F max-stable if

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

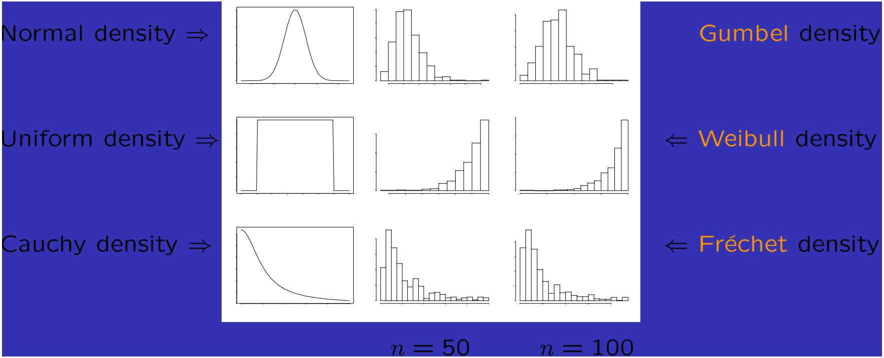
Examples

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 $a_n = 1$ & $b_n = \log n$

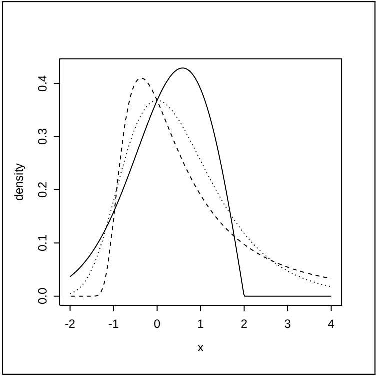
Weibull $F(x) = \exp(-(-x)^\alpha)$ for $x < 0$ (1 otherwise). Then $a_n = n^{-1/\alpha}$, $b_n = 0$

Maxima Distribution

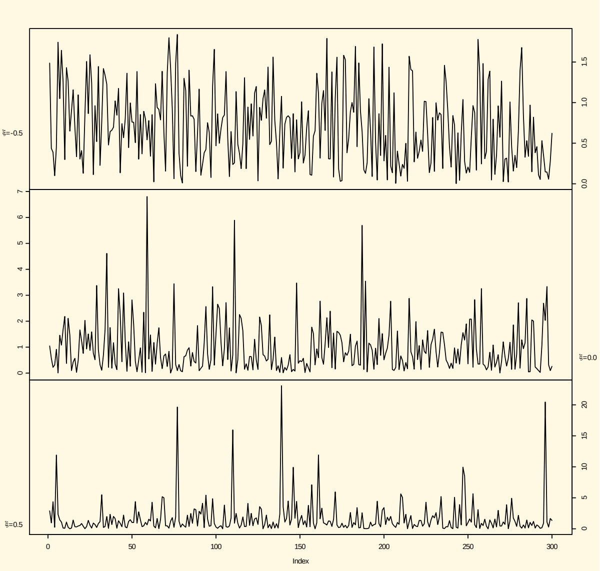


Generalized Extreme Value (GEV) distribution

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



From Bounded to Heavy tails



Intro summary

Modeling maxima : GEV

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

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

Note : Modeling excedances via Generalized Pareto Distribution

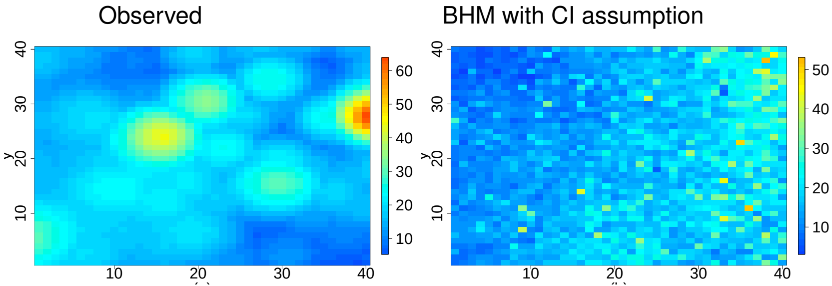
If exceedances $(\mathbf{R} - u | \mathbf{R} > u)$ follows a GPD(σ_u, ξ) then higher exceedances $(\mathbf{R} - v | \mathbf{R} > v)$ also follows GPD$(\sigma_u + (v-u)\xi, \xi)$

A few studies linking EVT with geophysical extremes

- **Special issue of the journal Extremes, 2010**
- 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
- Biodiversity and extreme temperatures, Sang and Gelfand, 2009
- Lichenometry, Jomelli et al., 2007
- Hydrology Katz et al.
- Downscaling Vrac M., Kallache M., Rust H., Friedrichs P., etc
- GCMs and RCMS analysis Smith R., Zwiers F., Maraun D., etc

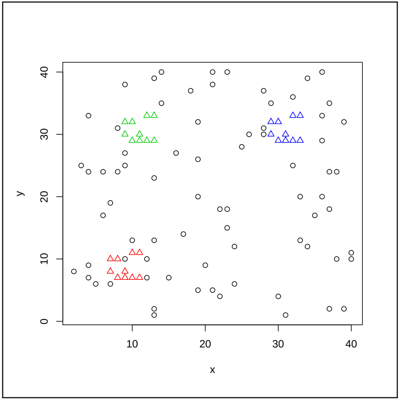
Limits of the univariate approach

Independence or conditional independence assumptions



Ribatet, Cooley and Davison (2010)

Why is Multivariate EVT needed ?



How to perform **spatial interpolation** of extreme events ?

Why is Multivariate EVT needed ?

- Compute confidence intervals
- Calculating probabilities of joint extreme events

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 few geophysical applications of Multivariate EVT

- Sea surges & river flows (Gumbel conditional regression)
  Tawn et al., 2004
- Measuring the spatial dependence among rainfall maxima in Bourgogne (Max-stable processes) :
  Naveau & et al., (2009, Biometrika)
- **Modeling multivariate dependence among pollutants** (spectral EVT measures) :
  Cooley, Davis and Naveau (2009, JMVA)
- Spatial extremes, Bel, Bacro, Lantujenoul (2010)
- Extreme snow, Blanchet et al., 2010

Mutivariate extremes

A few Approaches for modeling multivariate extremes

- **Max-stable processes** : Adapting asymptotic results for multivariate extremes
  Schlather & Tawn (2003), de Haan & Pereira (2005)
- **Complete modeling** : Auto-Regressive spatio-temporal heavy tailed processes, Davis and Mikosch (2007), AR-Gumbel Toulemonde et al. (2009)
- **Copula approach** : uniform marginals with extreme copulas, Genest et al., Charpentier
- Ribatet et al. (2010), Spatial R package for extremes
- Pseudo-likelihood inference Padoan, Ribatet and Sisson

Main question

How to model dependencies among maxima ?

Choice of marginals : unit-Fréchet

$$F(x) = \exp(-1/x), \text{ for } x > 0$$

Fréchet (1878-1973)

Max-stable processes

Max-stability in the univariate case with an unit-Fréchet margin

$$F^t(tx) = F(x), \text{ for } F(x) = \exp(-1/x)$$

Max-stable processes

Max-stability in the univariate case with an unit-Fréchet margin

$$F^t(tx) = F(x), \text{ for } F(x) = \exp(-1/x)$$

Max-stability in the multivariate case with unit-Fréchet margins

$$F^t(tu, tv) = F(u, v)$$

A central question

$$F(u,v) = ?? \text{ such that } F^t(tu, tv) = F(u,v)$$

A central question $F(u,v) = F^t(tu,tv)$

Suppose that $F(u,v) = \exp(-V(u,v))$ and let (X_i, Y_i) iid with distribution $F(u,v)$ and $i = 1,\dots,t$

Link with counting processes

$$
\begin{aligned}
P(\max X_i \leq tu, \max Y_i \leq tv) &= P(\forall i = 1,\dots,t; X_i \leq tu, Y_i \leq tv), \\
&= F^t(tu,tv), \\
&= F(u,v), \\
&= P(\text{Number of points in } [u,\infty) \times [v,\infty)] = 0), \\
&= \frac{(V(u,v))^0 \exp(-V(u,v))}{0!}
\end{aligned}
$$

A central question $F(u, v) = F^t(tu, tv)$

Suppose that $F(u, v) = \exp(-V(u, v))$ and let (X_i, Y_i) iid with distribution $F(u, v)$ and $i = 1, \ldots, t$

Link with counting processes

$$
\begin{aligned}
P(\max X_i \le tu, \max Y_i \le tv) &= P(\forall i = 1, \ldots, t; X_i \le tu, Y_i \le tv), \\
&= F^t(tu, tv), \\
&= F(u, v), \\
&= P(\text{Number of points in } [u, \infty) \times [v, \infty)] = 0), \\
&= \frac{(V(u, v))^0 \exp(-V(u, v))}{0!}
\end{aligned}
$$

Interpretation of $V(u, v)$

It can be viewed as the integrated Poisson intensity of the limit of

$$N_t(u, v) = \sum_{i=1}^{t} I\left[(X_i, Y_i) \notin [0, tu) \times [0, tv)\right]$$

A central question $F(u,v) = F^t(tu,tv)$

Equivalence between F and V

$F(u,v) = F^t(tu,tv)$ is equivalent to $V(u,v) = tV(tu,tv)$

Pseudo-polar coordinates

The special case $r = (u+v)$ and $\omega_1 = \frac{u}{r}, \omega_2 = \frac{v}{r}$ (pseudo-polar coordinates) gives

$$F(u,v) = \exp\left(-\frac{1}{r} \quad V(\omega_1,\omega_2)\right)$$

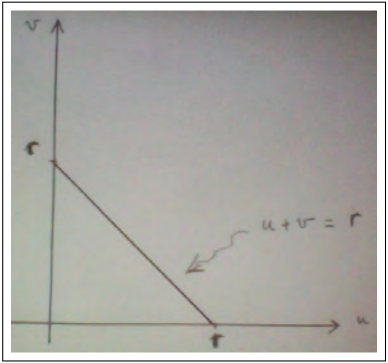
The Poisson intensity can be viewed as a product of two independent components : a radius (strength) and an angular (direction)

Polar coordinates

2D

$r = (u + v)$ and

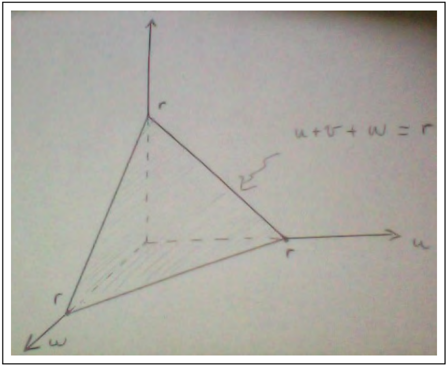
$\omega_1 = \frac{u}{r}, \omega_2 = \frac{v}{r}$



3D

$r = (u + v + w),$

$\omega_1 = \frac{u}{r}, \omega_2 = \frac{v}{r}, \omega_3 = \frac{w}{r}$

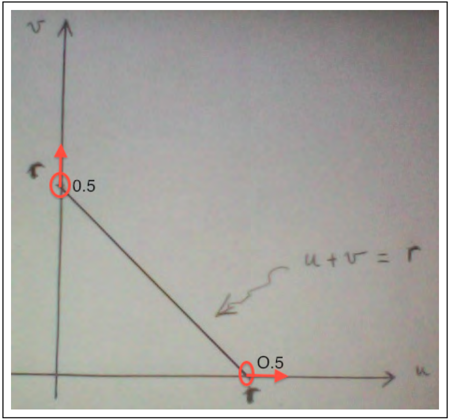


2D Polar coordinates

2D : INDEPENDENT CASE

$r = (u + v)$ and

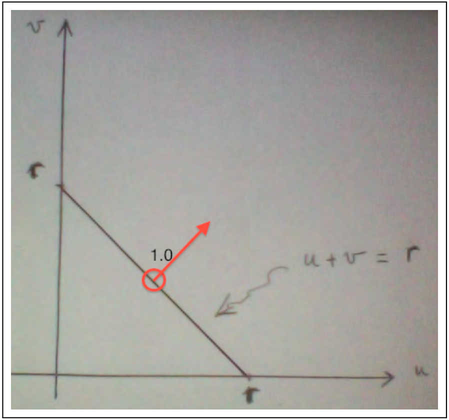
$\omega_1 = \frac{u}{r}, \omega_2 = \frac{v}{r}$



2D : COMPLETE DEPENDENCE

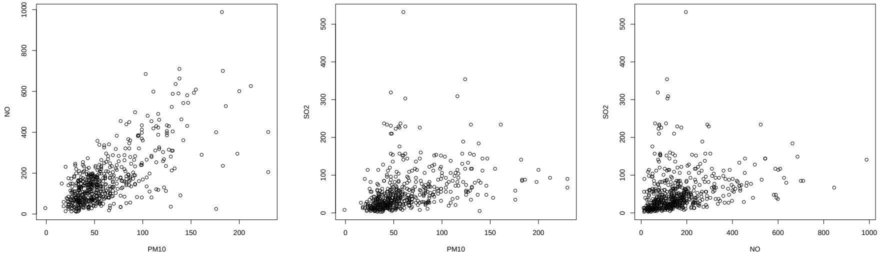
$r = (u + v)$ and

$\omega_1 = \frac{u}{r}, \omega_2 = \frac{v}{r}$



Examples : fitting multivariate maxima

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)



Our strategy

1. Assume observations arise from a max-stable process
2. Find and fit a flexible parametric model for the spectral density
3. Two desiderata : (A) interpretable parameters & (B) going beyond the bivariate case

Multivariate Max-Stable Distributions (de Haan, Resnick)

If $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_p))^T$ has a multivariate max-stable distribution with unit Fréchet margins ($\mathbb{P}(Z(\boldsymbol{x}_i) \leq z) = \exp(-z^{-1})$) then :

$$G(\boldsymbol{z}) = \mathbb{P}(\boldsymbol{Z} \leq \boldsymbol{z}) = \exp[-V(\boldsymbol{z})], \text{ where}$$

$$V(\boldsymbol{z}) = p \int_{S_p} \max_i \left(\frac{w_i}{z_i} \right) dH(\boldsymbol{w}),$$

H is a positive measure on S_p, s.t.

$$\int_{S_p} w_i dH(\boldsymbol{w}) = 1/p,$$

and $S_p = \{\boldsymbol{w} \in \mathbb{R}_+^p | w_1 + \ldots + w_p = 1\}$.

Models for Multivariate MSD's

Exponent measure function $V(\boldsymbol{z})$	Spectral density $h(\boldsymbol{w})$
■ Logistic ■ Asymmetric Logistic (Tawn, 88) ■ Negative Logistic (Joe, 90)	■ Dirichlet (Coles & Tawn, 91) ■ Dirichlet mixture (Boldi & Davison, 2006) ■ **Pairwise Beta** (Cooley, Davis and Naveau)
+ Can obtain $G(\boldsymbol{z})$ – Overparametrized ? – Less flexible ?	+ More flexibility ? – Cannot directly get $G(\boldsymbol{z})$

Dirichlet model (Coles, Tawn, 1991) $\int_{S_p} w_i dH(\boldsymbol{w}) = 1/p$

$$h(\boldsymbol{w};\boldsymbol{\theta}) = \frac{1}{p}(\boldsymbol{m}\cdot\boldsymbol{w})^{-(p+1)}\prod_{j=1}^{p} m_j h^*\left(\frac{m_1 w_1}{\boldsymbol{m}\cdot\boldsymbol{w}},\ldots,\frac{m_p w_p}{\boldsymbol{m}\cdot\boldsymbol{w}};\boldsymbol{\theta}\right)$$

A special case : Dirichlet model

$$h^*(\boldsymbol{w};\boldsymbol{\alpha}) = \frac{\Gamma(\boldsymbol{\alpha}\cdot\boldsymbol{1})}{\prod_{j=1}^{p}\Gamma(\alpha_j)}\prod_{j=1}^{p} w_j^{\alpha_j-1}, \quad \alpha_j > 0, j = 1,\ldots,p.$$

$$h(\boldsymbol{w};\alpha) = \frac{1}{p}\prod_{j=1}^{p}\frac{\alpha_j}{\Gamma(\alpha_j)}\frac{\Gamma(\boldsymbol{\alpha}\cdot\boldsymbol{1}+1)}{(\boldsymbol{\alpha}\cdot\boldsymbol{w})^{p+1}}\prod_{j=1}^{p}\left(\frac{\alpha_j w_j}{\boldsymbol{\alpha}\cdot\boldsymbol{w}}\right)^{\alpha_j-1}$$

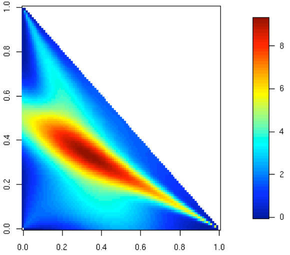
Our Pairwise Beta Model

$$
\begin{aligned}
h_p(\boldsymbol{w}; \alpha, \boldsymbol{\beta}) &= K_p(\alpha) \sum_{i \neq j} h_{i,j}(w_i, w_j; \alpha, \beta_{i,j}), \text{ where} \\
h_{i,j}(w_i, w_j; \alpha, \beta_{i,j}) &= (w_i + w_j)^{(p-1)(\alpha-1)} (1 - (w_i + w_j))^{\alpha - 1} \times \\
& \frac{\Gamma(2\beta_{i,j})}{(\Gamma(\beta_{i,j}))^2} \left(\frac{w_i}{w_i + w_j} \right)^{\beta_{i,j} - 1} \left(\frac{w_j}{w_i + w_j} \right)^{\beta_{i,j} - 1}
\end{aligned}
$$

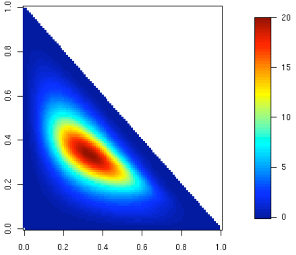
Advantages :

- no adjustment necessary to get center of mass condition $\int w_j dH(\mathbf{w}) = 1/p$
- parameters have some interpretation : α controls overall dependence, $\beta_{i,j}$'s control pairwise dependence
- largely specified by pairwise parameters
- Middle ground between Coles & Tawn (1991) and Boldi & Davison (2007)

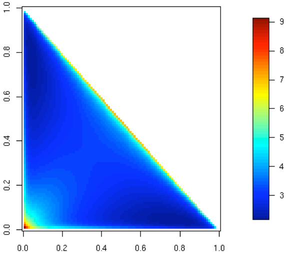
Pairwise Beta Models



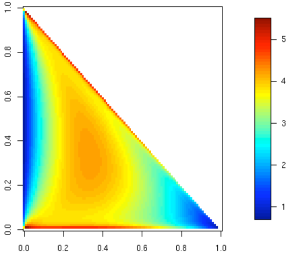
$\alpha = 1, \boldsymbol{\beta} = (2, 4, 15)$



$\alpha = 4, \boldsymbol{\beta} = (2, 4, 15)$



$\alpha = 1, \boldsymbol{\beta} = (2, .5, .5)$



$\alpha = 1, \boldsymbol{\beta} = (2, 2, .5)$

Fitting the spectral density model

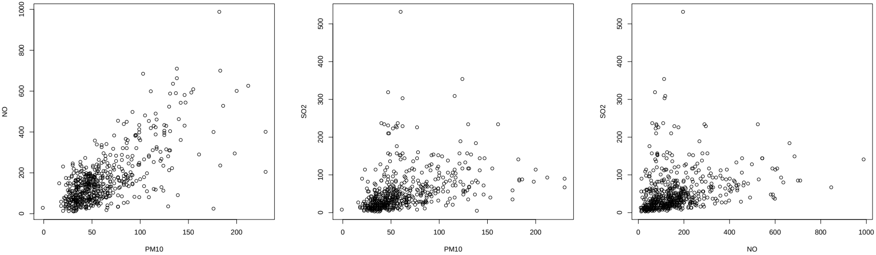
Beirlant at al., (2004), Coles & Tawn (2004), Boldi & Davison (2007)

(a) have common marginals with unit tail index
(b) transform into polar coordinates and select exceedances above t_0
(c) maximize the likelihood

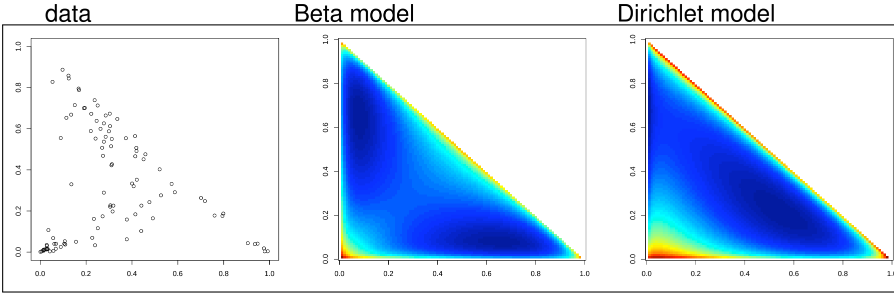
$$L(\boldsymbol{\theta}; (r_{(i)}, \boldsymbol{w}_{(i)}), i = 1, \ldots, N_{t_0}) \approx \exp(-\nu(A)) \prod_{i=1}^{N_{t_0}} d\nu(r_{(i)}, \boldsymbol{w}_{(i)}) = \exp(-t_0^{-1}) \prod_{i=1}^{N_{t_0}} r_{(i)}^{-2} h(\boldsymbol{w}_{(i)}, \boldsymbol{\theta}),$$

Air pollutants example

α	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{4,5}$
—	PM10, NO	PM10, NO2	PM10, O3	PM10, SO2	NO, NO2	NO, O3	NO, SO2	NO2, O3	NO2, SO2	O3, SO2
0.31	4.04	29.69	0.33	0.81	3.51	0.34	0.53	0.61	0.45	0.33
(0.002)	(0.139)	(1.222)	(0.006)	(0.026)	(0.119)	(0.006)	(0.014)	(0.013)	(0.011)	(0.006)

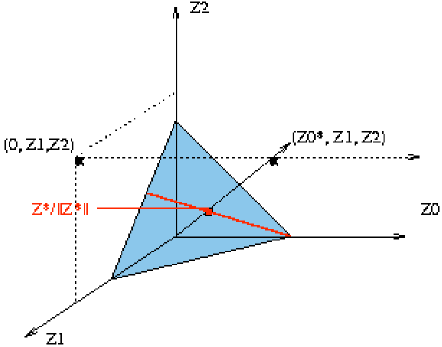


Air pollutants example



100 largest observations.
corners= PM10 (lower right), NO (upper left), SO2 (lower left)

Prediction : Approximating the conditional density ?



If $V(\boldsymbol{z})$ is known and differentiable, then joint density can be obtained exactly. However, we are modeling $h(\boldsymbol{w})$. Assume Z_1, Z_2 are observed and Z_0 is unobserved. Any predictor Z_0^* will yield a point $\boldsymbol{Z}^* = (Z_0^*, Z_1, Z_2)$ which can be mapped back to S_p as $\frac{\boldsymbol{Z}^*}{\|\boldsymbol{Z}^*\|_1}$.

Approximating the conditional density ?

If $V(\boldsymbol{z}) = \mu\{(0, z]^c\}$ is small (i.e. the radius is large), then

$$G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z})) \approx 1 - V(\boldsymbol{z}).$$

Using Coles and Tawn (91) result to estimate the density at $\boldsymbol{z}$:

$$g(\boldsymbol{z}) \approx \frac{\partial}{\partial z_1, \ldots, \partial z_p}[1 - V(\boldsymbol{z})] = \frac{1}{||\boldsymbol{z}||^{-(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right)$$

So conditional density can be approximated by

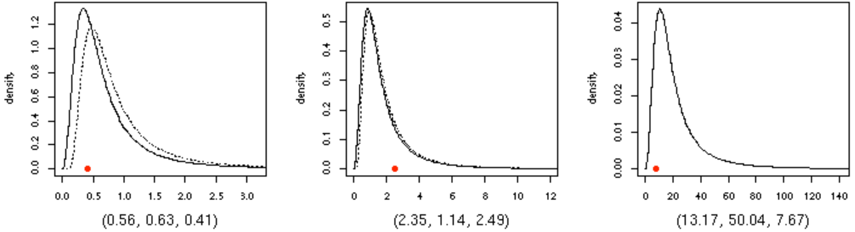
$$g_{Z_p|Z_1,\ldots Z_{p-1}}(z_p|z_1, \ldots, z_{p-1}) \approx \frac{\frac{1}{||\boldsymbol{z}||^{-(p+1)}} h\left(\frac{\boldsymbol{z}}{||\boldsymbol{z}||}\right)}{\int_0^\infty \frac{1}{||\boldsymbol{z}^*||^{-(p+1)}} h\left(\frac{\boldsymbol{z}^*}{||\boldsymbol{z}^*||}\right) d\zeta}$$

where $\boldsymbol{z}^* = (z_1, \ldots, z_{p-1}, \zeta)$.

Approximating the conditional density ?

Three realizations from a trivariate symmetric logistic distribution.

True conditional density (solid line) and approximated conditional density (dotted line)



$||\boldsymbol{z}|| = 1.6$ $||\boldsymbol{z}|| = 6$ $||\boldsymbol{z}|| = 71$

Summary of our spectral approach

- “Simple” and flexible spectral density with interpretable parameters
- Can be used for prediction or interpolation purposes
- Can be generalized (Ballani, Schlather, 2010)
- Can be extended to the asymptotic independent case (Qin, Smith, Ren, 2008)

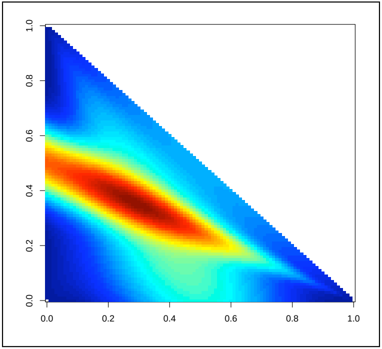
Take home messages

- Multivariate EVT may help characterizing extremes dependencies in space and time
- Physical knowledge should be integrated into the statistical analysis
- Computational issues can be arisen quickly
- Modeling trade off between parametric and non-parametric approaches
- Asymptotic independence can be an issue
- Extremes here means very rare

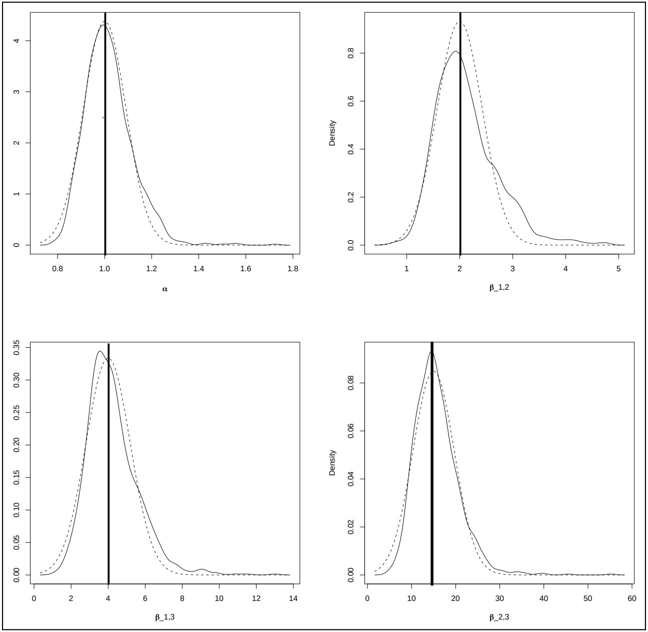
Two advertisements

- Extreme Value Analysis (EVA, Lyon June 27th to July 1st, 2011)
- Environmental Risk and Extreme Events, Workshop, Ascona, July 10-15 2011

An example with $\alpha = 1$, $\beta = (2, 4, 15)$

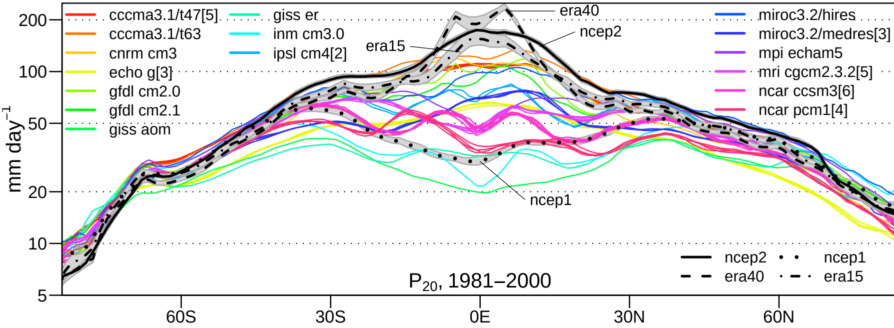


An example with $\alpha = 1$, $\beta = (2, 4, 15)$ ($\ldots$ = asymptotic mle) 200 real * 1000



A main random variable of interest : precipitation

1. Relevant parameter in meteorology and climatology
2. Highly stochastic nature compared to other meteorological parameters



Kharin and Zwiers, Journal of Climate 2007, P_{20} (1981-2000)

Estimating the GPD parameters estimates $(\hat{\sigma}_u, \hat{\xi})$

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

Taking advantages of the stability property

- Mean Excess function

$$\mathbb{E}(\mathbf{R} - u | \mathbf{R} > u) = \frac{\sigma_u + u\xi}{1 - \xi}$$

- the scale parameter varies linearly in the threshold u
- the shape parameter ξ is fixed wrt the threshold u