Dirichlet mixture model for multivariate extremes re-parametrization and inference with censored data.

Anne Sabourin

Télécom ParisTech

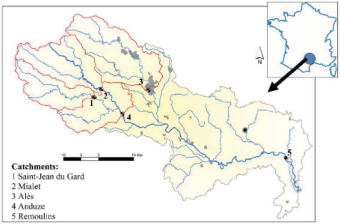
Joint work with Philippe Naveau (LSCE, Saclay), Anne-Laure Fougères (ICJ, Lyon 1), Benjamin Renard (IRSTEA, Lyon).

November 19th, 2013
MISTIS workshop : 'copulas and extremes', Grenoble

Censored Multivariate extremes: floods in the 'Gardons'

joint work with Benjamin Renard

- Daily streamflow data at 4 neighbouring sites :
  St Jean du Gard, Mialet, Anduze, Alès.
- **Joint distributions of extremes ?**
  $\rightarrow$ probability of simultaneous floods.
- Recent, 'clean' series very short
- Historical data from archives, depending on 'perception thresholds' for floods (Earliest: 1604). $\rightarrow$ censored data



Gard river   Neppel *et al.* (2010)

How to use all different kinds of data ?

Wishes

- Flexible model for the dependence structure of large excesses (non parametric) in moderate dimension
- Uncertainty assessment (Bayesian framework)
- Use the dependence structure to improve marginal estimation at poorly gauged sites
  (joint estimation margins + dependence)

Outline

Multivariate extremes and model uncertainty

Dirichlet mixture model: a reparameterization

Historical, censored data in the Dirichlet model

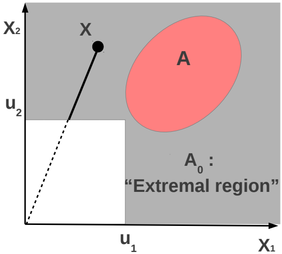
Multivariate extremes

- Random vectors $\mathbf{Y} = (Y_1, \dots, Y_{d,})$; $\quad Y_j \geq 0$
- Margins: $Y_j \sim F_j$, $1 \leq j \leq d$
  (Generalized Pareto above large thresholds)
- **Standardization** ($\to$ unit Fréchet margins)

$$X_j = -1/\log[F_j(Y_j)] \; ; \quad P(X_j \leq x) = e^{-1/x}, \quad 1 \leq j \leq d$$

- Joint behaviour of extremes: distribution of $\mathbf{X}$ above large thresholds ?

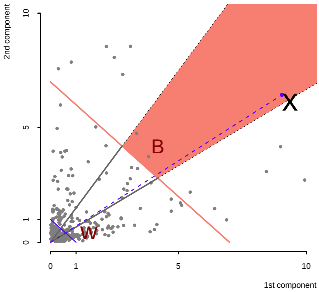
$P(\mathbf{X} \in A | X \in A_0)$? $\qquad (A \subset A_0, \mathbf{0} \notin A_0)$, A_0 'far from the origin'.

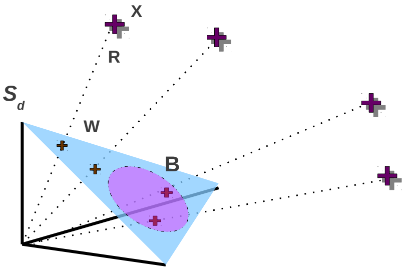


Polar decomposition and angular measure

- Polar coordinates: $R = \sum_{j=1}^{d} X_j$ (L_1 norm) ; $\mathbf{W} = \frac{\mathbf{X}}{R}$.
- $\mathbf{W} \in$ simplex $\mathbf{S}_d = \{\mathbf{w} : \quad w_j \geq 0, \sum_j w_j = 1\}$.
- **Angular probability measure**:

$$H(B) = P(\mathbf{W} \in B) \quad (B \subset \mathbf{S}_d).$$





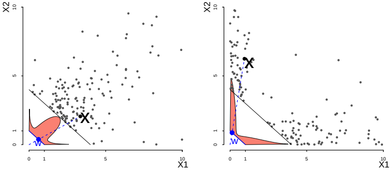
Fundamental Result

de Haan, Resnick, 70's, 80's

- Radial homogeneity (regular variation)

$$P(R > r, \mathbf{W} \in B | R \geq r_0) \underset{r_0 \to \infty}{\sim} \frac{r_0}{r} H(B) \quad (r = c\, r_0, c > 1)$$

- Above large radial thresholds, R is independent from W
- H (+ margins) entirely determines the joint distribution



- One condition only for genuine H: **moments constraint**

$$\int \mathbf{w}\ \mathrm{d}H(\mathbf{w}) = (\frac{1}{d}, \ldots, \frac{1}{d}).$$

  Center of mass at the center of the simplex.
- Few constraints: **non parametric** family !

Estimating the angular measure: non parametric problem

- **Non parametric estimation** (empirical likelihood, Einmahl *et al.*, 2001, Einmahl, Segers, 2009, Guillotte *et al*, 2011.) No explicit expression for asymptotic variance, Bayesian inference with $d = 2$ only.

- Compromise: **Mixture** of countably many parametric models $\rightarrow$ Infinite-dimensional model + easier Bayesian inference (handling parameters).

  Dirichlet mixture model

  (Boldi, Davison, 2007 ; Sabourin, Naveau, 2013)

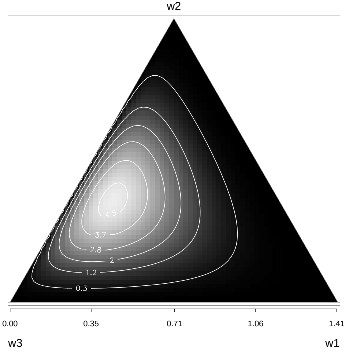
- Can Dirichlet mixtures be used with **censored data** ?

Outline

Dirichlet distribution

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

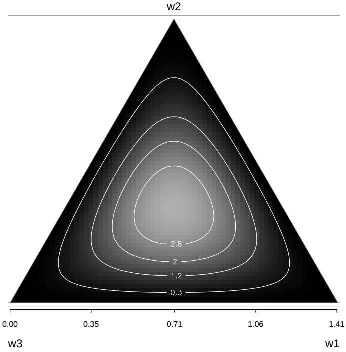
- $\boldsymbol{\mu} \in \mathring{\mathbf{S}}_d$: location parameter (point on the simplex): 'center';
- $\nu > 0$: concentration parameter.



Dirichlet distribution

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

- $\boldsymbol{\mu} \in \mathring{\mathbf{S}}_d$: location parameter (point on the simplex): 'center';
- $\nu > 0$: concentration parameter.



Dirichlet mixture model

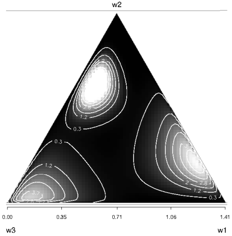
Boldi, Davison, 2007

- $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot\,,1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$, $\qquad \psi = (\boldsymbol{\mu}, \mathbf{p}, \boldsymbol{\nu})$,

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

- Moments constraint $\to$ on $(\boldsymbol{\mu}, p)$:

$$\sum_{m=1}^{k} p_m \,\boldsymbol{\mu}_{.,m} = (\frac{1}{d}, \dots, \frac{1}{d})\,.$$



Weakly dense family ($k \in \mathbb{N}$) in the space of admissible angular measures

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$. Boldi, Davison, 2007
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin ; Sabourin, Renard, *in preparation*

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$. Boldi, Davison, 2007

  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.

  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin ; Sabourin, Renard, *in preparation*

Re-parametrization

- How to build a prior on $(p\,,\,\boldsymbol{\mu})$?
- Constraint on **center of mass**: $\sum_j p_j\,\boldsymbol{\mu}_{\cdot,j}$
- Sequential construction : Use **associativity** properties of barycenter.
- Intermediate variables: **partial centers of mass** ; determined by **eccentricity parameters** $(e_1,\ldots,e_{k-1}) \in (0,1)^{k-1}$.
- Deduce last $\boldsymbol{\mu}_{\cdot,k}$ from first ones: **no more constraints** !

Bayesian model

- New parameter : $\theta_k = (\boldsymbol{\mu}_{\cdot,1:k-1}, e_{1:k-1}, \nu_{1:k})$
- Unconstrained parameter space : union of product spaces ('rectangles')

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

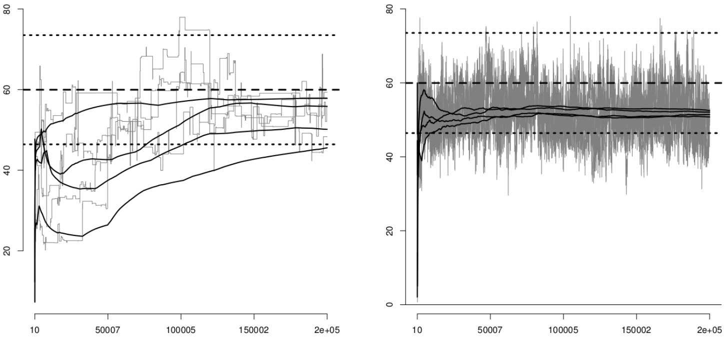
  - Inference: Gibbs + Reversible-jumps.

- Restriction (numerical convenience) : $k \leq 15$, $\nu < \nu_{\text{max}}$, *etc* ...

- 'Reasonable' prior $\simeq$ 'flat' and rotation invariant.
  Balanced weight and uniformly scattered centers.

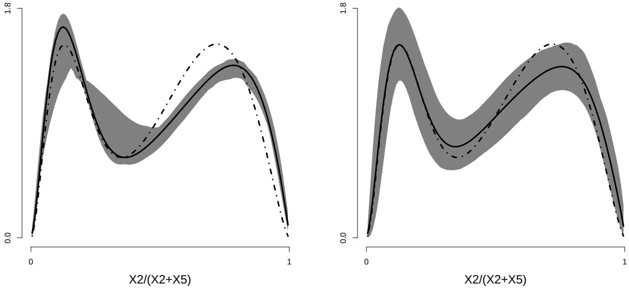
Resuts in the re-parametrized version

- asymptotics:
  - Posterior consistency : $\forall U$ weakly open in Θ, containing θ_0, $\pi_n(U) = \pi(U|\text{data}_{1:n}) \xrightarrow[n\to\infty]{} 1$.
  - Markov chain's ergodicity: $\sum_{t=1}^{T} g(\theta_t) \xrightarrow[T\to\infty]{} \mathbb{E}_{\pi_n}(g)$
- empirical convergence checks:
  Better mixing :



Resuts in the re-parametrized version

- asymptotics:
  - Posterior consistency : $\forall U$ weakly open in Θ, containing θ_0, $\pi_n(U) = \pi(U|\text{data}_{1:n}) \xrightarrow[n\to\infty]{} 1$.
  - Markov chain's ergodicity: $\sum_{t=1}^{T} g(\theta_t) \xrightarrow[T\to\infty]{} \mathbb{E}_{\pi_n}(g)$
- empirical convergence checks:
  Better coverage of credible sets (d=5, bivariate margins, simulated data)



Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$. Boldi, Davison, 2007
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin ; Sabourin, Renard, *in preparation*

Bayesian inference and censored data

- Two issues : (i) parameters constraints (ii) censorship

(i) Bayesian framework: MCMC methods to sample the posterior distribution.
Constraints $\Rightarrow$ Sampling issues for $d > 2$. Boldi, Davison, 2007
  - Re-parametrization: No more constraint, fitting is manageable for $d = 5$: Sabourin, Naveau, 2013

(ii) Censoring: data$\neq$ points but segments or boxes in $\mathbf{R}^d$.
  - Intervals overlapping threshold: extreme or not ?
  - Likelihood: density $\frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w})$ integrated over boxes.
  - Sabourin ; Sabourin, Renard, *in preparation*

Outline

Multivariate extremes and model uncertainty

Dirichlet mixture model: a reparameterization

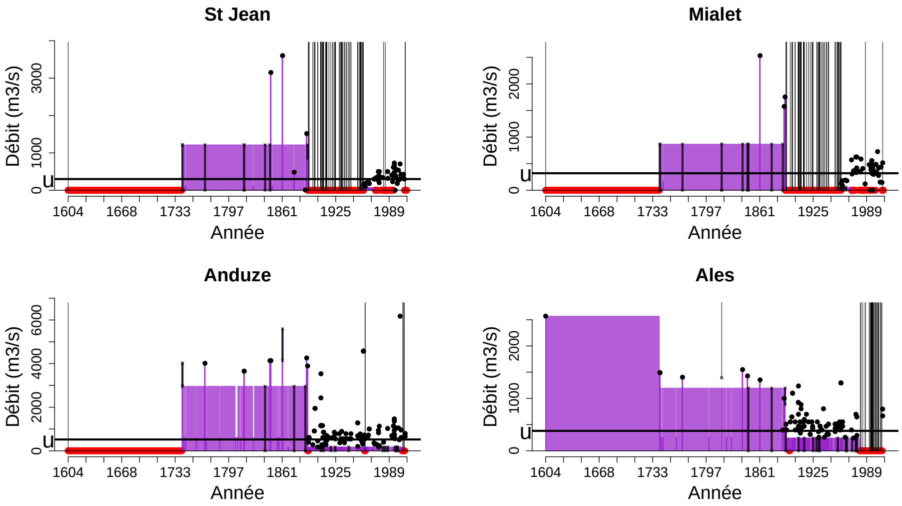
Historical, censored data in the Dirichlet model

Multivariate extremes for regional analysis in hydrology

- Many sites, many parameters for marginal distributions, short observation period.
- 'Regional analysis': replace time with space.
  Assume some parameters constant over the region and use extreme data from all sites.
- Independence between extremes at neighbouring sites ?
  Dependence structure ?
  - Idea: use multivariate extreme value models

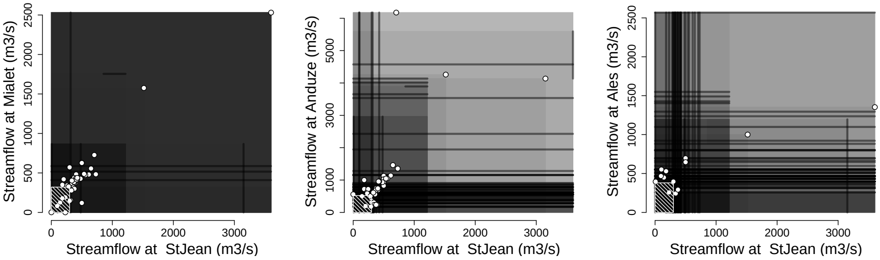
Censored data: univariate and pairwise plots

Univariate time series:

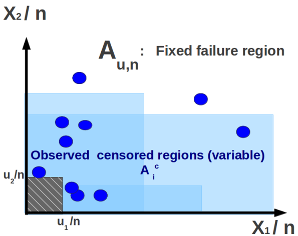


Censored data: univariate and pairwise plots

Bivariate plots:



Data overlapping threshold and Poisson model



How to include the rectangles overlapping threshold in the likelihood ?

$$\left\{ \left(\frac{t}{n}, \frac{\mathbf{X}_t}{n} \right), \ 1 \le t \le n \ \right\} \sim \text{Poisson Process } (\mathrm{Leb} \times \lambda) \ \text{ on } [0,1] \times A_{u,n}$$

λ: ' exponent measure', with Dirichlet Mixture angular component

$$\frac{\mathrm{d}\lambda}{\mathrm{d}r \times \mathrm{d}\mathbf{w}}(r, \mathbf{w}) = \frac{d}{r^2} h(\mathbf{w}) \,.$$

Overlapping events appear in Poisson likelihood as

$$\mathbf{P}\left[N \left\{ (\frac{t_2}{n} - \frac{t_1}{n}) \times \frac{1}{n} A_i \right\} = 0 \right] = \exp\left[-(t_2 - t_1) \lambda(A_i) \right]$$

‘Censored’ likelihood: density integrated over boxes

- Ledford & Tawn, 1996: partially extreme data censored at threshold,
  - GEV models
  - Explicit expression for censored likelihood.

- Here: *idem* + natural censoring
  - Poisson model (Threshold excesses)
  - No closed form expression for integrated likelihood.

- Two terms without closed form:
  - Censored regions A_i overlapping threshold:

$$\exp\{-(t_2 - t_1)\lambda(A_i)\}$$

  - Classical censoring above threshold

$$\int_{\text{censored region}} \frac{d\lambda}{d\mathbf{x}}.$$

Data augmentation

One more Gibbs step, no more numerical integration.

- Objective: sample $[\theta|Obs] \propto$ likelihood (censored obs)
- Additional variables (replace missing data component): $\mathcal{Z}$.
  Full conditionals $[\theta|\mathbf{Z}, Obs], [Z_i|Z_{j\neq j}, \theta, Obs], \ldots$ explicit (Thanks Dirichlet): $\rightarrow$ Gibbs sampling.
- Consistency condition:

$$\int [z, \theta|Obs]_+ \, \mathrm{d}z = [\theta|Obs]$$

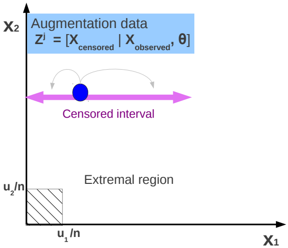
- Sample $[z, \theta|Obs]_+$ (augmented distribution) on $\Theta \times \mathcal{Z}$.

Censored regions above threshold

$$\int_{\text{Censored region}} \frac{d\lambda}{dx} \, dx_{j_1:j_r} \quad :$$

Generate missing components under univariate conditional distributions

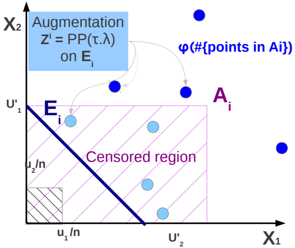
$$\mathbf{Z}^j_{1:r} \sim [X_{\text{missing}} | X_{\text{obs}}, \theta]$$



Dirichlet $\Rightarrow$ Explicit univariate conditionals
Exact sampling of censored data on censored interval

Censored regions overlapping threshold

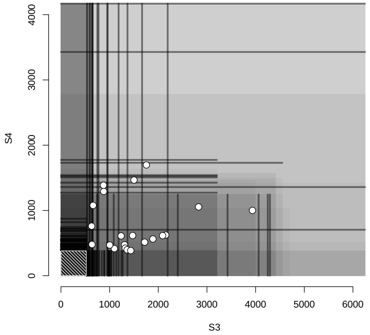
$$e^{-(t_{2,i}-t_{1,i})\lambda(A_i)} \Leftrightarrow \begin{cases} \text{augmentation Poisson process } N_i \text{ on } E_i \supset A_i. \\ + \\ \text{Functional } \varphi(N_i) \end{cases}$$

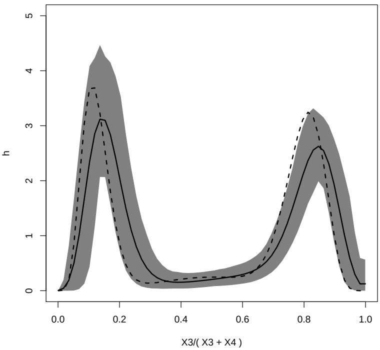


$$[z,\theta|Obs] \propto \underbrace{\ldots}_{\text{density terms, prior, augmented missing components}} [N_i]\varphi(N_i)$$

Simulated data (Dirichlet, $d = 4$, $k = 3$ components), same censoring as real data

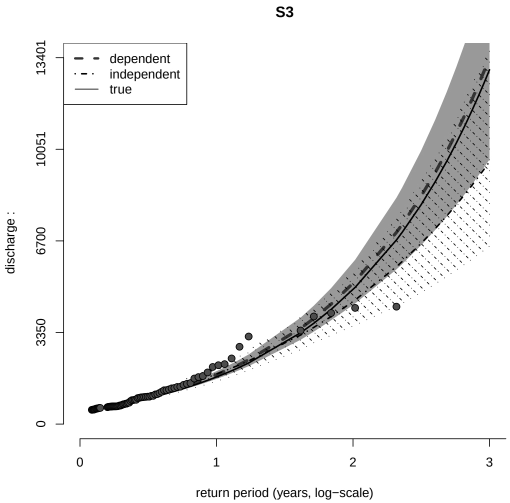
Pairwise plot and angular measure density
(true/ posterior predictive)



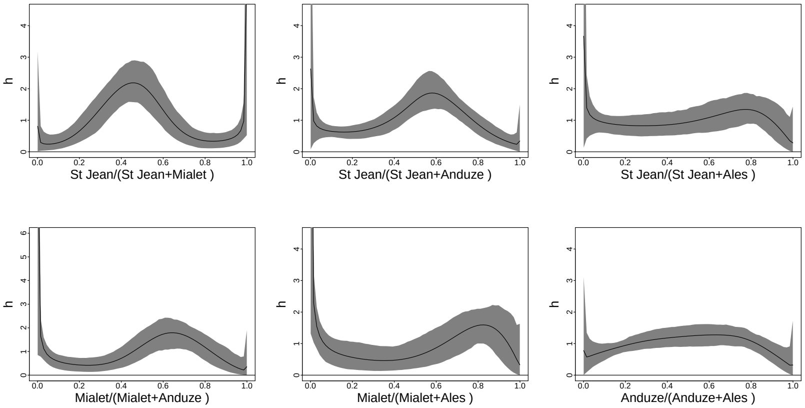


Simulated data (Dirichlet, $d = 4$, $k = 3$ components), same censoring as real data

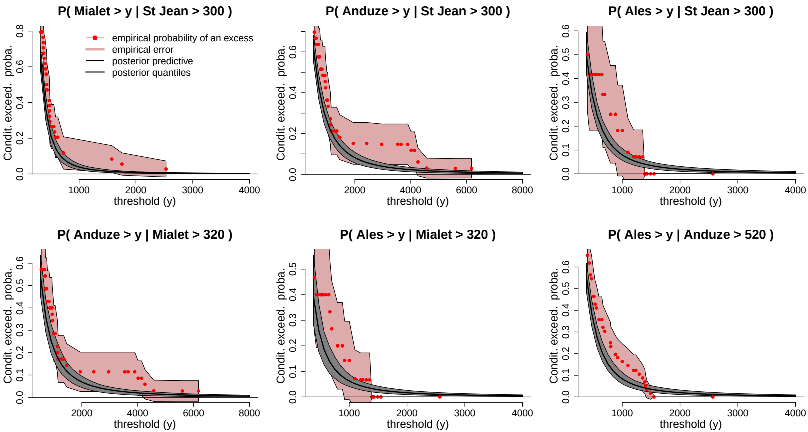
Marginal quantile curves: better in joint model.



Angular predictive density for Gardons data



Conditional exceedance probability



Conclusion

- Building Bayesian multivariate models for excesses:
  - Dirichlet mixture family: ‘non’ parametric, Bayesian inference possible up to re-parametrization
  - Censoring $\rightarrow$ data augmenting (Dirichlet conditioning properies)
  - Two packages `R`:
    - `DiriXtremes`, MCMC algorithm for Dirichlet mixtures,
    - `DiriCens`, implementation with censored data.
- High dimensional sample space (GCM grid, spatial fields) ?
  - Impose reasonable structure (sparse) on Dirichlet parameters
  - Dirichlet Process ? Challenges :
    Discrete random measure $\neq$ continuous framework

Bibliographie I

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

Coles, SG and Tawn, JA
Modeling extreme multivariate events
JR Statist. Soc. B, 53:377–392, 1991

Gómez, G., Calle, M. L., and Oller, R.
Frequentist and bayesian approaches for interval-censored data.
Statistical Papers, 45(2):139–173, 2004.

Hosking, J.R.M. and Wallis, J.R..
Regional frequency analysis: an approach based on L-moments
Cambridge University Press, 2005.

Ledford, A. and Tawn, J. (1996).
Statistics for near independence in multivariate extreme values.
Biometrika, 83(1):169–187.

Neppel, L., Renard, B., Lang, M., Ayral, P., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., and Vinet, F. (2010).
Flood frequency analysis using historical data: accounting for random and systematic errors.
Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55(2):192–208.

Resnick, S. (1987).
Extreme values, regular variation, and point processes, volume 4 of Applied Probability. A Series of the Applied Probability Trust.
Springer-Verlag, New York.

Bibliographie II

Sabourin, A., Naveau, P. and Fougères, A-L. (2013)
Bayesian model averaging for multivariate extremes.
Extremes, 16(3) 325–350

Sabourin, A., Naveau, P. (2013)
Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization.
Computation. Stat and Data Analysis

Schnedler, W. (2005).
Likelihood estimation for censored random vectors.
Econometric Reviews, 24(2):195–217.

Tanner, M. and Wong, W. (1987).
The calculation of posterior distributions by data augmentation.
Journal of the American Statistical Association, 82(398):528–540.

Van Dyk, D. and Meng, X. (2001).
The art of data augmentation.
Journal of Computational and Graphical Statistics, 10(1):1–50.