Semiparametric Bayesian multivariate models for extreme exceedances

Manuele Leonelli, Dani Gamerman

Instituto de Matematica, Universidade Federal do Rio de Janeiro

Genova, 12/09/2016

Plan for this talk

- Motivation
- Background:
  - Univariate extreme value theory
  - Multivariate extreme value theory
  - Copulae
- Our proposed models
- Simulations
- An hydrological application

Motivation

Precise knowledge and predicting capabilities for extremes are fundamental in many disciplines:

- Environmental sciences
- Finance and actuarial science
- Engineering and reliability

Evidence of increasing occurrence of extremes and larger insurance and economic losses.

Assessment of extreme dependence is often critical:

- sea level and wave height
- concentration of O_3 and NO_2

Contributions

Standard statistical methods do not guarantee precise extrapolations towards the tail of the distribution where little, if no, data is available $\Longrightarrow$ **extreme value theory**.

For multivariate extremes these rely on highly technical results and are not widely available for use.

We introduce here easily interpretable and flexible multivariate models to investigate both marginal and joint extreme behaviour.

Inference is carried out within the Bayesian paradigm using the MCMC machinery.

Univariate EVT: result 1

Let $X_1, \ldots, X_n$ i.i.d r.v.s and $M_n = \max\{X_1, \ldots, X_m\}$. If there exist sequences $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$, then

$$\lim_{n\to\infty} \mathbb{P}\left((M_n - b_n)/a_n \leq x\right) = H(x),$$

where H is a **generalized extreme value** distribution

$$H(x|\xi, \sigma, u) = \exp\left\{- \left(1 + \xi(x-u)/\sigma\right)^{-1/\xi}\right\}$$

Inference carried over sub-sample maxima.

Univariate EVT: result 2

Let X have d.f. F. Then

$$\lim_{u\to\infty} F(x|u) = \mathbb{P}(X \leq x + u | X > u) = P(x),$$

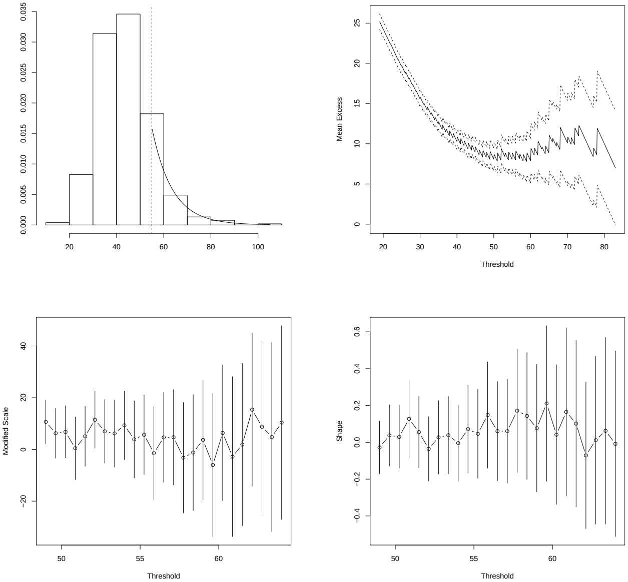
where $P(x)$ is the d.f. of a **generalized Pareto**

$$P(x|\xi,\sigma,u) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma}(x-u))^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-(x-u)/\sigma), & \xi = 0 \end{cases}$$

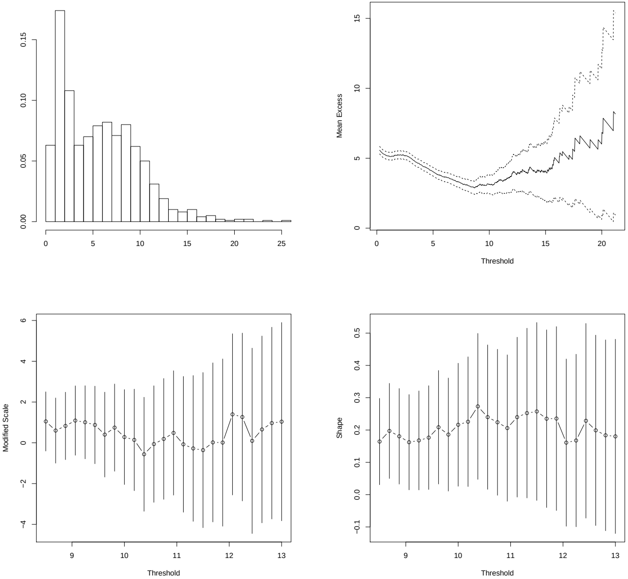
ξ is the *shape*, σ the *scale* and u the *threshold*.

If $\xi \geq 0$, $x \in (u, \infty)$, but if $\xi < 0$, $x \in (u, u - \sigma/\xi)$.

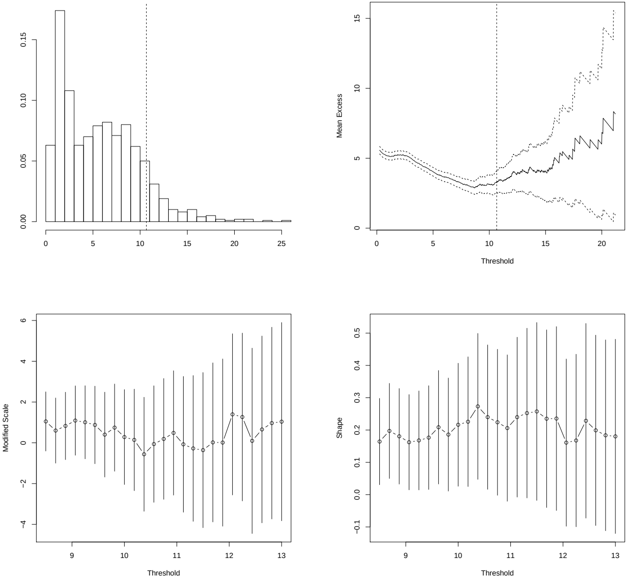
Fitting a GPD: NO_2 data



Fitting a GPD: Simulated data



Fitting a GPD: Simulated data



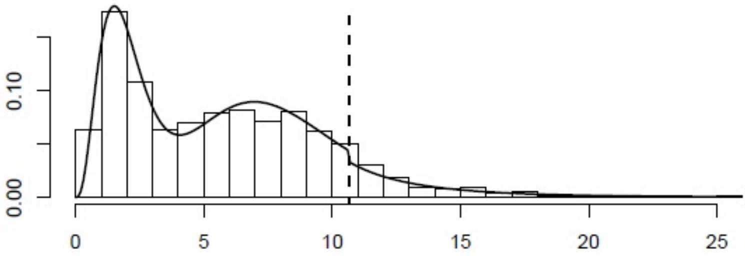
Fitting a GPD: Mixture modelling

Two major drawbacks with standard techniques:

- Most of the data is not formally used for inference
- Arbitrary choice of the threshold

An alternative is a so-called **extreme mixture model**

- GPD parametric model for the tail
- An uncertain threshold
- A non-parametric model for the bulk



Multivariate extremes

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d d-dimensional random vectors with $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$ and

$$\boldsymbol{M}_n = (\max_{1 \leq i \leq n} X_{1,i}, \ldots, \max_{1 \leq i \leq n} X_{d,i})$$

If there are sequences $\boldsymbol{a}_n > \mathbf{0}$ and $\boldsymbol{b}_n \in \mathbb{R}^d$ such that

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x} \right) = G(\boldsymbol{x})$$

then G is the d.f. of a **multivariate extreme value** distribution. The marginals of G are univariate extreme value distributions.

Multivariate extremes

Suppose G_i is unit Fréchet, i.e. $G_i(x) = \exp(-1/x)$. Then

$$G(\boldsymbol{x}) = exp(-V(\boldsymbol{x}))$$

where

$$V(\boldsymbol{x}) = d \int_{\mathcal{S}_d} \max_{i=1,\dots,d} \frac{\omega_i}{x_i} \mathrm{d}H(\boldsymbol{w}),$$

$\mathcal{S}_d$ is the unit simplex and H is a positive finite measure satisfying

$$\int_{\mathcal{S}_d} w_i \mathrm{d}H(\boldsymbol{w}) = \frac{1}{d}, \forall i = 1, \dots, d.$$

V is the **exponent measure** and H is the **spectral measure**.

Fitting multivariate extremes

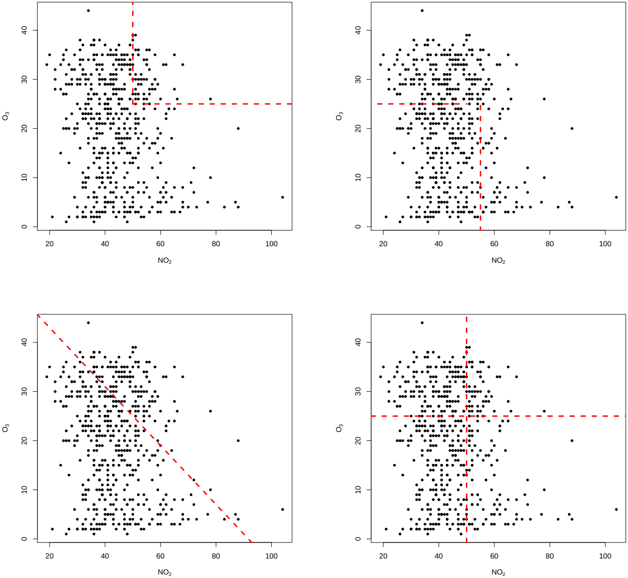
As in the univariate case data above some threshold is supposed extreme and formally used for inference.

Various approaches:

- Parametric:
  - for the exponent measure *(simpler but less flexible) Coles and Tawn 1991, 1994; Jaruskova 2009; Joe 1990;*
  - for the spectral measure *(computationally more intensive) Ballani and Schlather 2011; Boldi and Davison 2007; Cooley et al. 2010;*
- Nonparametric modelling of the spectral measure *(almost exclusively non-Bayesian) Guillotte et al. 2011;*.
- Motivated by different theoretical justifications *Bortot et al. 2000; Heffernan and Tawn 2004; Ramos and Ledford 2009;*.

In all cases some initial non-parametric data transformation is performed (via ECDF)

Multivariate thresholds



Asymptotic independence

Let G be a bivariate EVD for the maxima of (X_1, X_2). Then if

$$G(x_1, x_2) = G(x_1)G(x_2)$$

X_1, X_2 are said to be **asymptotically independent**.
This can be checked by computing

$$\bar{\chi} = \lim_{u \to 1} \mathbb{P}(F_1(X_1) > u | F_2(X_2) > u).$$

If $\bar{\chi} = 0 \implies$ asymptotic independence
If $\bar{\chi} \in (0, 1] \implies$ asymptotic dependence

If $X_1, X_2 \sim \mathcal{N}$, $\text{cor}(X_1, X_2) = \rho \neq 0$, then

$$\lim_{u \to 1} \mathbb{P}(F_1(X_1) > u | F_2(X_2) > u) = 0.$$

Copulae

A copula C is a flexible tool to construct multivariate distributions with given margins. Let $X_1, \ldots, X_d$ be r.v.s with d.f.s $F_1, \ldots, F_d$.
A **copula** C is a function $C : [0,1]^d \to [0,1]$ s.t.

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

- Sklar's theorem guarantees there always exists one such copula;
- If $X_1, \ldots, X_d$ are continuous C is unique;
- C is a d.f. in $[0,1]$ itself;
- separate marginal and dependence modelling.

Copula density

Since C is a d.f. in $[0,1]^d$ it has a density

$$c(u_1, \dots, u_d) = \frac{\mathrm{d}}{\mathrm{d}u_1 \cdots \mathrm{d}u_d} C(u_1, \dots, u_d),$$

and thus

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d).$$

Construction of copulae

Sklar's theorem guarantees that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Calling $u_i = F_i(x_i)$, we have that $x_i = F_i^{-1}(u_i)$ and thus substituting

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)).$$

Often F is chosen to be an elliptical distribution. For example Gaussian

$$C(u_1, \dots, u_d) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

where Φ is the standard normal d.f. and Φ_R is the multivariate d.f. with mean zero and correlation R.
But also Skew-Normal, T, Skew-T: **Elliptical copulae**.

Asymptotic behaviour

Recall, asymptotic dependence can be assessed by

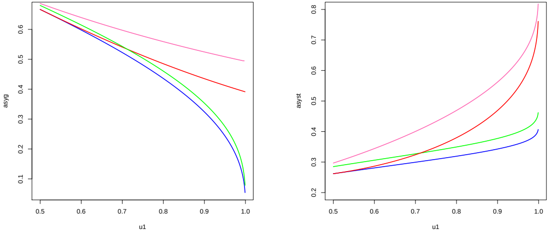
$$\bar{\chi} = \lim_{u \to 1} \mathbb{P}(F_1(X_1) > u | F_2(X_2) > u)$$

Other measure is the sub-asymptotic dependence coefficient

$$\bar{\chi}_{\text{sub}} = \lim_{u \to 1} \frac{2 \log \mathbb{P}(F_1(X_1) > u)}{\log \mathbb{P}(F_1(X_1) > u, F_2(X_2) > u)} - 1 \in (-1, 1]$$

For Normal and skew-Normal $\bar{\chi} = 0$ and $\bar{\chi}_{\text{sub}} \in (-1, 1)$

For T and skew-T $\bar{\chi} \in (0, 1]$ and $\bar{\chi}_{\text{sub}} = 1$.



General idea

We build new models for multivariate extremes that

- marginally utilize flexible extreme mixture models
- exploit the flexibility of copulae to model dependence
- assess extreme dependence from the chosen copula
- formally utilize all data available

Marginal MGPDs

Marginally we use the MGPD model (*Nascimento et al. 2012*)

- mixture of gammas for the bulk (*Wiper et al. 2001*)
- GPD for the tail

$$f(x|\cdot) = \begin{cases} \sum_{i=1}^{k} w_i g_i(x|\mu_i, \eta_i), & x \leq u, \\ \left(1 - \sum_{i=1}^{k} w_i G_i(x|\mu_i, \eta_i)\right) p(x|\xi, \sigma, u), & x > u \end{cases}$$

where p is the density of a GPD, $\sum_{i=1}^{k} w_i = 1$, with $w_i \geq 0$, and G_i is the d.f. of a Gamma with density

$$g_i(x|\mu_i, \eta_i) = \frac{(\eta_i/\mu_i)^{\eta_i}}{\Gamma(\eta_i)} x^{\eta_i - 1} \exp(-(\eta_i/\mu_i)x)$$

Parametrization chosen to address identifiability issues of mixtures.

Joint modelling

The full model is chosen as a mixture of elliptic copulas with MGPD margins.

$$f(\boldsymbol{x}|\cdot) = \sum_{i=1}^{r} \omega_i c_i(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d),$$

where f_i is MGPD, c_i is a copula density and $\sum_{i=1}^{r} \omega_i = 1$, $\omega_i \geq 0$.

So for example if Gaussian

$$f(\boldsymbol{x}|\cdot) = \sum_{i=1}^{r} \omega_i c_i^{\text{gauss}}(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d)$$

where $c_i^{\text{gauss}}(u_1, \ldots, u_d) = |R_i|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathsf{T}}(R_i^{-1} - I_d)\boldsymbol{y}\right)$, with $\boldsymbol{y}^{\mathsf{T}} = (\Phi^{-1}(u_1)), \ldots, \Phi^{-1}(u_d))$.

Some restrictions

- For T copulas each mixture component has the same degrees of freedom ($\in \mathbb{R}^+$)
- For skew-Normal copulas each mixture component has the same skewness parameters
- For skew-T copulas we consider one single copulas with integer degrees of freedom
- one correlation parameter ρ_i for each mixture component. For identifiability $\rho_1 < \rho_2 < \cdots < \rho_r$.

Priors

- μ_{ij}, η_{ij} vague Inverse Gamma and Gamma respectively
- ξ_i, σ_i uninformative prior $\pi(\xi_i, \sigma_i) \propto \sigma_i^{-1}(1+\xi_i)^{-1}(1+2\xi_i)^{-1/2}$ (*Castellanos and Cabras, 2007*)
- u_i Normal distribution with prior mean around a high sample quantile
- ρ_i and δ_{i1}, δ_{i2} (skewness parameters) $\mathcal{U}(-1,1)$
- v (integer) zero-truncated Poisson with mean 25
- v (positive) uninformative prior (*Fonseca et al. 2008*)

$$\pi(v) \propto \left(\frac{v}{v+3}\right)^{1/2} \left(\phi(v/2) - \phi((v+1)/2) - \frac{2(v+3)}{v(v+1)^2}\right)^{1/2},$$

where ϕ is the trigamma function.

Inference

Inference is carried out via MCMC with Metropolis-Hastings steps

- Implementation in OX
- Variances of the proposals are tuned via adaptive M-H
- Proposals:
  - Gamma for parameters in $\mathbb{R}^+$
  - Truncated normal for parameters taking values in continuous spaces
  - a discrete uniform in $\{v-2, v-1, v, v+1, v+2\}$ for integer v.
- 25000 iterations, 5000 burn-in and thinning every 20 iterations (giving an MCMC sample of 1000 observations)
- Number of Gamma mixture components was chosen via investigation of the marginals and then held fixed

Inference

Since inference is via MCMC, we can compute posterior point estimates and credibility intervals for any function of the parameters.

For extremes, interest is on

- high marginal p-quantiles q_i s.t. $\mathbb{P}(X_i > q_i) = p$

$$q_i = u_i + \frac{\sigma_i}{\xi_i}\left[\left(1 - \frac{p - F_i(u_i|\cdot)}{1 - F_i(u_i|\cdot)}\right)^{-\xi_i} - 1\right],$$

  where F_i is the d.f. of the MGPD

- Joint exceedances

$$\mathbb{P}(X_1 > x_1, X_2 > x_2) = 1 - F_1(x_1|\cdot) - F_2(x_2|\cdot) + \sum_{i=1}^{r} \omega_i C_i(F_1(x_1|\cdot), F_2(x_2|\cdot))$$

Predictions

For a parameter vector $\boldsymbol{y}$, the density of a new observation y given a sample $\boldsymbol{x}$ equals

$$f(y|\boldsymbol{x}) = \int f(y, \boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta} = \int f(y|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}|\boldsymbol{x}}(f(y|\boldsymbol{\theta}))$$

This expectation has no closed-form but can be approximated via Monte Carlo

$$\hat{f}(y|\boldsymbol{x}) = \frac{1}{J} \sum_{i=1}^{J} f(y|\boldsymbol{\theta}^{(i)}),$$

where $\boldsymbol{\theta}^{(i)}$ is a value sampled from $\pi(\boldsymbol{\theta}|\boldsymbol{x})$.

Thus we can straightforwardly produce predictions for high quantiles.

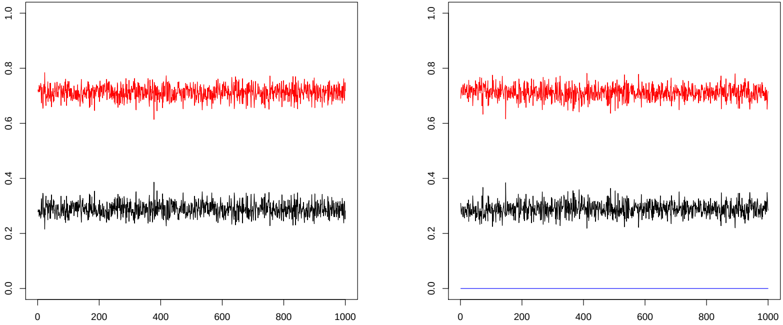
Simulation study

We simulated 1000 observations from 8 models, 4 asymptotically dependent and 4 asymptotically independent:

- Mixture of 2 T-copulae with MGPD margins
- Mixture of 2 Gaussian copulae with MGPD margins
- Skew Normal copula with MGPD margins
- Skew-T copula with MGPD margins
- Morgenstern copula with lognormal-GPD margins
- Asymmetric logistic copula with lognormal-GPD margins
- Cauchy copula with lognormal margins
- Bilogistic copula with lognormal margins

Model selection: copulae weights

- Weights of unnecessary components equal to zero



Model selection: asy. dep. vs indep.

- We are able to identify the right number of components
- These can give an indication of extreme behavior

	Gaussian	Student-T	Skew-Normal
Cauchy	**2**	1	**2**
Asy. log.	1	1	1
Skew T	**2**	1	**2**
2-T	2	2	2
Bilogistic	1	1	1
Morgenstern	1	1	1
Skew-Normal	1	1	1
2-Gauss	2	2	2

Model selection: degrees of freedom

	T1	T2	ST
Cauchy (1)	0.95 (0.82,1.10)		1 (1,1)
Asy. log.	7.32 (4.42,16.07)		9 (4,22)
Skew T (5)	5.63 (3.86,9.36)		6 (4,12)
2-T (6)	2.35 (1.89,3.08)	9.83 (3.60,52.1)	3 (2,3)
Bilogistic	7.11 (4.33,14.6)		18 (6,29)
Morgenstern	38.8 (13.0,**155**)		20 (13,29)
Skew-Normal	28.9 (12.2,**136**)	1	19 (12,29)
2-Gauss	3.22 (2.46,4.50)	16.51 (5.83,**141**)	4 (3,6)

Model selection: BIC/DIC

- 2 Gaussian copulae

	Ind.	G1	G2	T1	T2	SN1	SN2	ST1
BIC	10285	9998	9973	9884	**9668**	10050	9986	9718
DIC	10033	9680	**9604**	9657	9635	9693	9612	9632

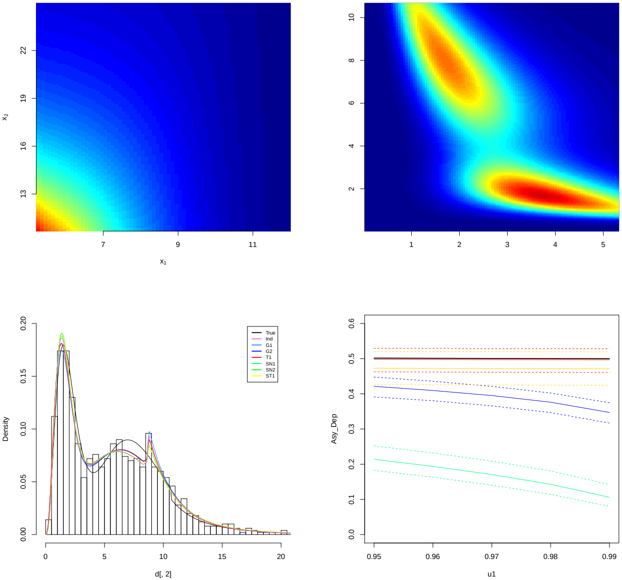
- Skew-T copula

	Ind.	G1	T1	SN1	ST1
BIC	11083	10846	10774	10279	**10278**
DIC	10930	10705	10434	**9865**	9999

- Cauchy copula

	Ind.	G1	G2	T1	SN1	SN2	ST1
BIC	9158	**8923**	8972	8953	8938	8988	8940
DIC	9260	9072	**8928**	9078	9091	8932	8934

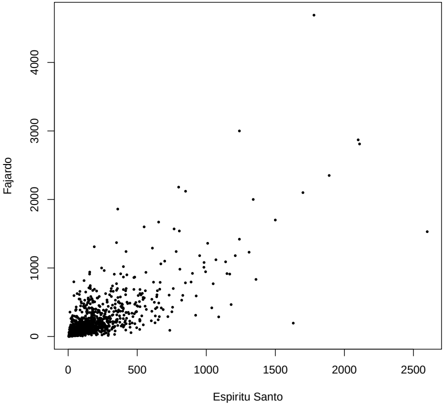
Summaries



x₂
22
19
16
13
7
9
11
x₁
10
8
6
4
2
1
2
3
4
5
True
Ind
G1
G2
T1
SN1
SN2
ST1
Density
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
d[, 2]
Asy_Dep
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.95
0.96
0.97
0.98
0.99
u1

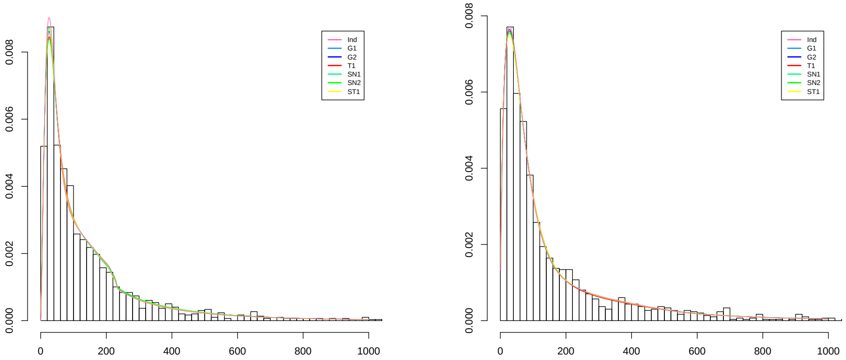
Rivers in Puerto Rico

- Dataset: 2492 weekly maxima of Espiritu Santo and Fajardo rivers
- Randomly selected 1492 observations to fit the models
- 1000 observations to investigate the models' prediction power



Marginals

	Ind.	G2	T1	SN2	ST1
ξ_1	0.26 (0.13,0.43)	0.19 (0.08,0.34)	0.22 (0.11,0.35)	0.18 (0.06,0.30)	0.20 (0.09,0.34)
ξ_2	0.34 (0.15,0.62)	0.27 (0.12,0.48)	0.32 (0.16,0.51)	0.27 (0.11,0.49)	0.28 (0.14,0.51)



Model selection

- T degrees of freedom 5.31 (3.78,7.91), Skew T degrees of freedom 10 (6,22);
- Mixture of 2 Gaussian/ Skew Normal copulae vs 1 T copula
- BIC/DIC

	Ind.	G1	G2	T1	SN1	SN2	ST1
BIC	40753	39518	39496	**39445**	39538	39495	39518
DIC	40765	39747	39618	39494	39896	**39487**	39593

Extremes' prediction

	Prob.	Emp.	Ind.	G2	T1	SN2
	0.90	[402,404]	368 (335,405)	378 (345,413)	373 (343,410)	379 (312,447)
Espiritu	0.95	[570,572]	551 (497,610)	554 (506,608)	554 (509,611)	553 (472,645)
	0.99	[1080,1120]	1128 (971,1354)	1061 (953,1235)	1092 (975,1262)	1047 (913,1249)
	0.999	[2180,2280]	2527 (1850,3971)	2115 (1707,2999)	2277 (1816,3112)	2038 (1645,2846)

	Prob.	Emp.	Ind.	G2	T1	SN2
	0.90	[441,441]	448 (395,492)	448 (402,493)	444 (396,491)	450 (401,496)
Fajardo	0.95	[610,629]	663 (599,747)	659 (602,732)	660 (599,734)	664 (601,739)
	0.99	[1300,1370]	1421 (1207,1732)	1341 (1184,1560)	1388 (1214,1632)	1345 (1167,1592)
	0.999	[2610,8800]	3564 (2536,6458)	2999 (2352,4418)	3379 (2586,4989)	2983 (2322,4627)

	Point	Prob.	Ind.	G2	T1	SN2
	(305,300)	0.10	0.023 (0.019,0.027)	0.093 (0.082,0.11)	0.092 (0.081,0.11)	**0.094 (0.075,0.114)**
Joint	(475,470)	0.05	0.006 (0.005,0.008)	0.042 (0.035,0.051)	0.043 (0.036,0.052)	**0.045 (0.035,0.057)**
	(850,850)	0.01	0.0006 (0.0004,0.0009)	**0.01 (0.008,0.014)**	0.012 (0.009,0.015)	0.0097 (0.0065,0.0137)
	(2000,2500)	0.001	5.0e-6 (1.5e-6,1.4e-5)	0.0004 (0.0002,0.0009)	**0.0008 (0.0004,0.0016)**	0.0005 (0.0002,0.0010)

Discussion

We have introduced novel multivariate models for extremes that

- do not require a pre-specified threshold
- utilize all the data points
- assess extreme dependence and can take into account asymptotic independence
- exploit the Bayesian paradigm to provide estimates and predictions of high quantiles

and explored their performance for estimation and prediction with both synthetic and real datasets.

Extensions

- Extensions into higher dimensions:
  - Identifiability constraint for matrices
  - Use of vine copulae
- Combination of a copula for the bulk and a parametric model of the spectral measure for the "tail"
- Use of covariates, time-dependent copulae, Markov switching models etc...

Extensions

- Extensions into higher dimensions:
  - Identifiability constraint for matrices
  - Use of vine copulae
- Combination of a copula for the bulk and a parametric model of the spectral measure for the "tail"
- Use of covariates, time-dependent copulae, Markov switching models etc...

Thanks for your attention!

References

- Ballani, Schlather, *A construction principle for multivariate extreme value distributions*, Biometrika 83: 715-726 (2011)
- Boldi, Davison, *A mixture model for multivariate extremes*, J. R. Statist. Soc. Ser. B 69:217-229 (2007)
- Bortot, Coles, Tawn, *The multivariate Gaussian tail model: an application to oceanographic data*, J. R. Statist. Soc. Ser. C 49:31-49 (2000)
- Castellanos, Cabras, *A default Bayesian procedure for the generalized Pareto distribution*, J. Statist. Plann. Inference 137: 473-483 (2007)
- Coles, Tawn, *Modelling extreme multivariate events*, J. R. Statist. Soc. Ser. B 53: 377-392 (1991)
- Coles, Tawn, *Statistical methods for multivariate extremes: an application to structural design*, J. R. Statist. Soc. Ser. C 43: 1-48
- Cooley, Davis, Naveau, *The pairwise beta distribution: a flexible parametric multivariate model for extremes*, J. Multivariate Anal. 101:2103-2117 (2010)

References

- Fonseca, Ferreira, Migon, *Objective Bayesian analysis for the Student-T regression model*, Biometrika 95: 325-333 (2008)
- Guillotte, Perron, Segers *Non-parametric Bayesian inference on bivariate extremes*, J. R. Statist. Soc. Ser. B 73:377-406 (2011)
- Heffernan, Tawn, *A conditional approach for multivariate extreme values*, J. R. Statist. Soc. Ser. B 66: 497-546
- Jaruskova, *Modeling multivariate extremes of precipitation series in northern Moravia*, Environmetrics 20: 751-775 (2009)
- Joe, *Families of min-stable multivariate exponential and multivariate extreme value distributions*, Statist. Probab. Letters 9: 7581 (1990)
- Nascimento, Gamerman, Lopes, *A semiparametric Bayesian approach to extreme value estimation*, Stat. Comput. 22: 661-675 (2012)
- Ramos, Ledford, *A new class of models for bivariate joint tails*, J. R. Statist. Soc. Ser. B 71: 219-241 (2009)
- Wiper, Rios Insua, Ruggeri, *Mixtures of gamma distributions with applications*, J. Comput. Graph. Statist. 10: 440-454 (2001)