Hierarchical space-time modeling of threshold exceedances

with an application to hourly space-time precipitation in Southern France

Thomas Opitz, BioSP, INRA Avignon

Workshop *spatiotempmeteo*
Rennes
29/11/2018

Motivation for this talk

- understanding **extreme events** is crucial in many domains
- **extreme events** can extend over **space and time**
  - **duration** of extreme conditions (e.g. droughts)
  - **aggregation** over space and time (e.g. cumulated precipitation over catchment)
- models for **extreme events** should be aware of **extreme value theory**
- we need models that can **incorporate space-time dependence and covariates**
- precipitation in Southern France is a typical example

Why hierarchical models ?

Hierarchical models can flexibly combine **several structured layers** :

- **observation/data process** (measurement errors, preferential sampling)

  is conditioned on

- “smooth” **latent process(es)** capturing trends and dependence

- **parameters** governing observation process and latent process(es) (variance, shape of distribution, dependence over space and time)

Hierarchical models often lack of unconditional closed-form likelihoods but can be **estimated through approximate Bayesian inference**.

However, for **suitably constructed hierarchical models**, we can do **frequentist inference** based on **composite likelihood** even in very high space-time dimension.

Space-time setting

Data observed over space and time :

$$X(s_i, t_j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

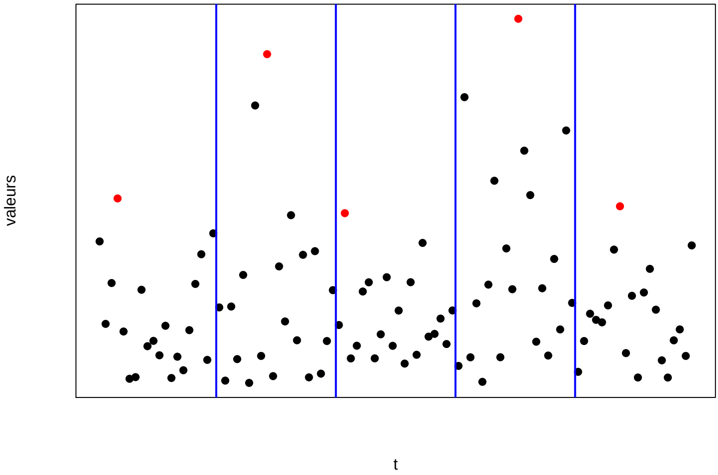
often with a factorial design : observation locations $\times$ observation times

Two popular approaches for tackling extreme value analysis :

- block maxima taken over temporal blocks
- threshold exceedances

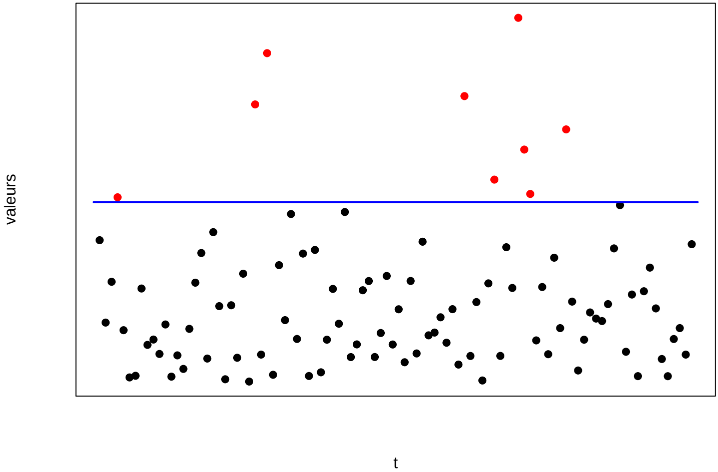
What **asymptotic distributions** can we get for linearly rescaled data when block size or threshold increase ?

Univariate extremes : block maxima



⇝ **generalized extreme value limit distribution** for rescaled maxima

Univariate extremes : threshold exceedances



⇝ **generalized Pareto limit distribution** (GP) for exceedances $Y = (X - u)_+ > 0$ above fixed threshold u

Generalized Pareto distribution

We model the **exceedance** Y with $Y \stackrel{d}{=} (X-u) \mid X > u$ above a **high threshold** u using the **generalized Pareto distribution** :

$$1 - \mathrm{GP}(y;\sigma,\xi) = \mathrm{pr}(Y > y \mid Y > 0) = \begin{cases} \left(1+\xi\frac{y}{\sigma}\right)_+^{-1/\xi} & \xi \neq 0, \\ \exp(-y/\sigma) & \xi = 0, \end{cases}$$

with **shape parameter** ξ and **scale parameter** $\sigma > 0$ [Davison and Smith, 1990, Coles, 2001]

- here we focus on $\xi \geq 0$:
  - $\xi > 0$ power law tail decay
  - $\xi = 0$ exponential tail decay
  - ($\xi < 0$ polynomial tail decay towards finite upper boundary)

We **estimate** 3 **parameters** characterizing the **tail distribution** : σ, ξ, $p = \Pr(X > u)$.

Summaries for bivariate extremal dependence [Coles et al., 1999]

Assume that X_1 and X_2 have both standard exponential marginal distribution.

Tail correlation χ :

$$\chi = \lim_{x\to\infty} \Pr(X_2 > x \mid X_1 > x) = \lim_{x\to\infty} \frac{\Pr(X_1 > x, X_2 > x)}{\Pr(X_1 > x)} \in [0,1]$$

$\Rightarrow$ **asymptotic dependence** if $\chi > 0$

If $\chi = 0$ **(asymptotic independence)**, joint tail decay is faster than marginal tail decay

$\Rightarrow$ characterize fast joint tail decay with new **coefficient $\overline{\chi} \in [-1,1]$** :

$$\lim_{x\to\infty} \Pr\left(\min(X_1,X_2) > y + x \mid (\min(X_1,X_2) > x\right) = \exp\left(\frac{2}{1+\overline{\chi}}y\right), \quad y > 0$$

$\Rightarrow$ **exceedances are (approximately) exponentially distributed** with mean

- 1 for marginal exceedances (by marginal normalization)
- 1 for min-exceedances with asymptotic dependence
- $0 \leq (1+\overline{\chi})/2 \leq 1$ for min-exceedances with asymptotic independence

Asymptotic spatial models

If block maxima or threshold exceedances observed over several spatial sites jointly converge to a limit process, then the limit is

- a max-stable limit process for block maxima,
- a generalized Pareto processes for threshold exceedances.

Spatial dependence in the limit arises only with **asymptotic dependence in data**.
Finite-sample dependence in asymptotically independent data vanishes asymptotically.

⚠ **Asymptotic models are not well adapted to asymptotically independent data.**
They would overestimate joint occurrences of very extreme values.

⚠ In practice, **most weather data look asymptotically independent**
$\Rightarrow$ **need more flexible dependence models for asymptotically independent data**.

Typical structure of a hierarchical exceedance model

Data :

- deterministic space-time threshold $u(s,t)$
- exceedance indicator $e(s,t) = 1(X(s,t) > u(s,t))$
- positive exceedances $y(s,t) = (x(s,t) - u(s,t)) > 0$ with $(s,t) \in \{e(s,t) = 1\}$

Hierarchical model :

$\boldsymbol{\theta} = (\boldsymbol{\theta}_{\boldsymbol{x}_y}, \boldsymbol{\theta}_{\boldsymbol{x}_e}, \theta_{\boldsymbol{p}}, \boldsymbol{\theta}_F, \ldots)$	parameters
$\boldsymbol{x}_y \mid \boldsymbol{\theta}$	latent process(es) for exceedances
$\boldsymbol{x}_e \mid \boldsymbol{\theta}$	latent process(es) for exceedance probability
$e(s,t) \mid \boldsymbol{x}_e, \boldsymbol{\theta} \sim \mathrm{Bernoulli}(p_{s,t,\boldsymbol{x}_e,\boldsymbol{\theta}})$	exceedance indicator data
$y(s,t) \mid \boldsymbol{x}_y, \boldsymbol{\theta} \sim F_{s,t,\boldsymbol{x}_y,\boldsymbol{\theta}}$	positive exceedance data

- **conditional independence** of $e(s,t)$ and $y(s,t)$ with respect to latent processes
- $\boldsymbol{x}_e$ and $\boldsymbol{x}_y$ may be dependent, e.g. to assure that :
  higher exceedance probability $\Rightarrow$ larger exceedances
- **conditional distribution F or unconditional distribution of $y(s,t)$ is GP**

Latent Gaussian models

Idea : embed latent Gaussian processes for the parameters p, σ, ξ

- no closed-form expressions for the "unconditional distribution" of $y(s,t)$
- **approximate Bayesian inference** for parameters and latent variables
  - **Markov-Chain Monte-Carlo**,
    e.g. [Cooley et al., 2007] for spatial modeling of extreme precipitation
  - **Integrated Nested Laplace Approximation** [Opitz et al., 2018]
    (+ penalized complexity prior for the tail index $\xi \geq 0$, constant over space-time)

Example : Space-time tail regression for precipitation

[Opitz et al., 2018] : Daily Dutch precipitation data 1973–1995

Three models with **seasonal effect** (RW) and **spatial effect** (Matérn covariance) :

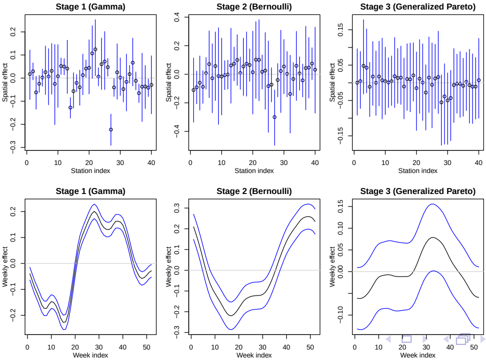
- **gamma model** with mean $\mu(s,t)$ for positive precipitation :

$$\log(\mu(s,t)) = \beta^{\Gamma} + \boldsymbol{x}^{\Gamma}_{\text{space}} + \boldsymbol{x}^{\Gamma}_{\text{season}}$$

$\Rightarrow$ fix space-time threshold $u(s,t)$ as high quantile of gamma fit

- **Bernoulli model** for $e(s,t)$ with $\text{logit}(p(s,t)) = \beta^{p} + \boldsymbol{x}^{p}_{\text{space}} + \boldsymbol{x}^{p}_{\text{season}}$
- **GP model** for exceedances $y(s,t)$ of gamma residuals $x(s,t)/\mu(s,t)$:

$$\log(\alpha\text{-quantile}) = \beta^{\text{GP}} + \boldsymbol{x}^{\text{GP}}_{\text{space}} + \boldsymbol{x}^{\text{GP}}_{\text{season}}$$



Pros and cons of latent Gaussian models for extremes

++Pros++

- many standard Bayesian inference tools (MCMC, INLA)
- spatial and spatio-temporal modeling is difficult without Gaussian processes

−−Cons−−

- Gaussian dependence is not well adapted to extremes, often not strong enough in the tails $\Rightarrow$ latent Gaussian models are useful mainly for **random effect modeling of nonlinear trends** of space/season/time and other covariates
- unconditional distributions of the hierarchical model are not in closed form
- very computer-intensive with high-dimension of datasets and/or latent models

Space-time Gamma-Pareto models [Bacro et al., 2017]

Latent space-time process $\Lambda(s,t)$ with $\Gamma(\mathrm{shape} = 1/\xi_{s,t}, \mathrm{rate} = \sigma_{s,t}/\xi_{s,t})$-margins.

Gamma-Pareto model for exceedances :

$$Y(s,t) \mid \Lambda(s,t) \sim \mathrm{Exp}(\mathrm{rate} = \Lambda(s,t))$$

has **unconditional generalized Pareto margins**

$$\mathrm{pr}(Y(s,t) > y) = (1 - \xi_{s,t} y/\sigma_{s,t})^{-1/\xi}, \quad y, \xi_{s,t}, \sigma_{s,t} > 0.$$

(in practice, we could marginally transform the model, e.g. for $\xi_{s,t} \leq 0$)

For the **exceedance probability** over $u(s,t)$, we set

$$\mathrm{pr}\left(X(s,t) > u(s,t) \mid \Lambda(s,t)\right) = \exp(-\kappa\Lambda(s,t)), \quad \kappa > 0$$

(higher exceedances $\leftrightarrow$ higher exceedance probability)

Lower tail dependence of $\Lambda(s,t)$ at 0 translates into upper tail dependence of $X(s,t)$.

Space-time gamma process

For modeling asymptotically independent threshold exceedances,
need a flexible **space-time gamma process with lower-tail asymptotic independence**.

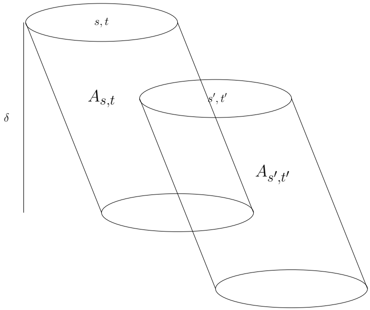
We smooth **space-time gamma white noise** $\Gamma(\mathrm{d}(s,t))$ with a **space-time kernel** :

- gamma white noise process
  - $\Gamma(A) \sim \Gamma(\alpha|A|, \beta)$
    $\Rightarrow$ α controls ruggedness of the space-time surface $\Lambda(s,t)$
  - $\Gamma(A_1) \perp \Gamma(A_2)$ if $A_1 \cap A_2 = \emptyset$
- space-time smoothing $\Lambda(s,t) = \Gamma(A_{s,t})$ with **kernel set** $A_{s,t}$
  $\Rightarrow$ shape and size of $A_{s,t}$ control space-time dependence and margins

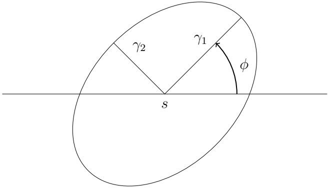
Space-time dependence structure

Use **slated cylinders** (base=space, height=time) for space-time kernel sets $A_{s,t}$
$\Rightarrow$ **space-time dependence is determined by intersection** of A_{s_1,t_1} and A_{s_2,t_2}.

Parameters : velocity (v_x, v_y), ellipse axes (γ_1,γ_2), ellipse orientation ϕ, time range δ



slated space-time cylinders



elliptical spatial base

Bivariate distributions of exceedances $Y(s,t) > 0$

$$Y(s,t) \mid \Lambda(s,t) \sim \mathrm{Exp}(\Lambda(s,t))$$

Consider the decomposition into independent gamma variables :

$$\Lambda(s_1,t_1) = \Lambda_0 + \Lambda_{1\setminus 2}, \qquad \Lambda(s_2,t_2) = \Lambda_0 + \Lambda_{2\setminus 1}$$

with $\Lambda_0 = \Gamma(A_{s_1,t_1} \cap A_{s_2,t_2})$, $\Lambda_{2\setminus 1} = \Gamma(A_{s_2,t_2} \setminus A_{s_1,t_1})$, $\Lambda_{1\setminus 2} = \Gamma(A_{s_1,t_1} \setminus A_{s_2,t_2})$.

$$\Rightarrow \mathrm{pr}(Y_1 > y_1, Y_2 > y_2) = \mathcal{L}_{(\Lambda_1,\Lambda_2)}(y_1,y_2) = \mathcal{L}_{\Lambda_0}(y_1+y_2) \times \mathcal{L}_{\Lambda_{1\setminus 2}}(y_1) \times \mathcal{L}_{\Lambda_{2\setminus 1}}(y_2)$$

with the Laplace transform $\mathcal{L}_\Lambda(y) = (1 - y/\beta_\Lambda)^{-\alpha_\Lambda}$ of $\Lambda \sim \Gamma(\alpha_\Lambda, \beta_\Lambda)$

$\Rightarrow$ **bivariate pdf, cdf and likelihood are in closed form !**

Asymptotic independence : $0 \leq \overline{\chi}(A_1,A_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} < 1$ if $A_1 \neq A_2$

Pairwise likelihood inference for Gamma-Pareto models

A **bivariate censored likelihood** $\ell(\boldsymbol{\theta}; z_1, z_2)$ is readily calculated, and we estimate space-time parameters $\boldsymbol{\theta}$ through the **pairwise likelihood**

$$\mathrm{pl}(\boldsymbol{\theta}) = \sum_{\|s_2 - s_1\| \leq d_{\max}} \sum_{|t_2 - t_1| \leq t_{\max}} \ell(\boldsymbol{\theta}; z(s_1, t_1), z(s_2, t_2)).$$

- maximal spatial distance $d_{\max}$
- maximal temporal lag $t_{\max}$

$\Rightarrow$ keep only observation pairs where dependence matters

We have run a simulation study mimicking the set-up of our data application. Pairwise likelihood estimation was shown to be fast and reliable.

"Épisodes cévenols"

During autumn,
hot and humid air enters from the Mediterranean,
then hits cold temperatures from ocean/mountains ⇒ precipitation.

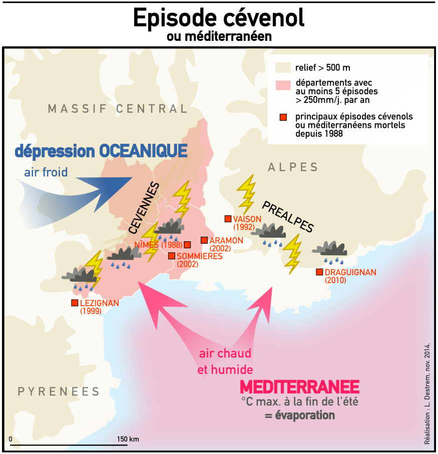
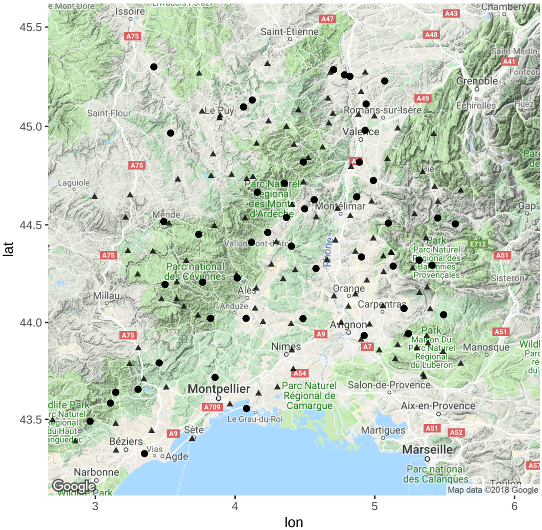


Image credit : Lucas Destrem (Wikipedia)

More general term for extreme precipitation events is **Mediterranean episodes**.

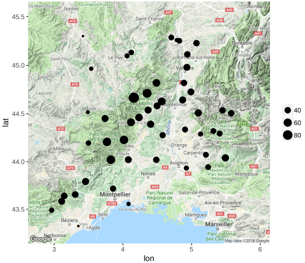
Hourly precipitation data

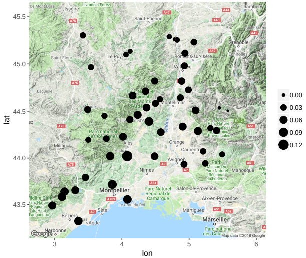
- Southern France, close to Rhone Valley between Mediterranean Sea and Alps
- 213 stations, around $80,000$ hours (1993-2014), but many missing data
- we select 50 stations (black dots) and focus on September to November months

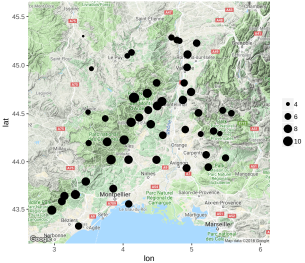


Site-specific generalized Pareto parameters

threshold $u(s)$ (upper left), shape $\xi(s)$ (upper right), scale $\sigma(s)$ (below)



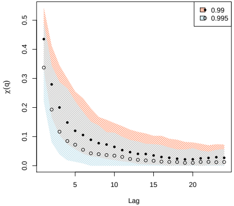




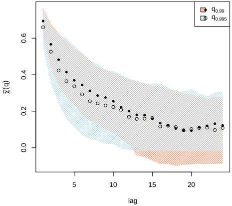
Exploring the extremal dependence in space and time

- tail correlation $\chi(p) = \Pr(X_1 > q_p \mid X_2 > q_p) \in [0, 1]$
- $\chi(p) \to 0$ for $p \uparrow 1$ if asymptotic dependence
- $\overline{\chi}(p) \approx$ ratio of means of joint/marginal exceedances (exponential scale)

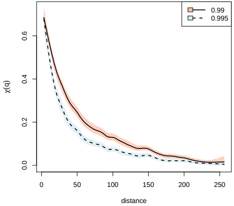
χ(time lag)



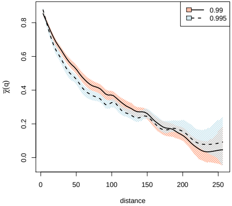
$\overline{\chi}$(time lag)



χ(space lag)



$\overline{\chi}$(space lag)



$\Rightarrow$ **clear evidence of asymptotic independence**

Model fitting by pairwise likelihood

Threshold is 99%-quantile (of dry+wet observations).

Estimated parameters :

- elliptical spatial base :
  - axes $\hat{\gamma}_1 = 170km(20km)$, $\hat{\gamma}_2 = 320km(20km)$
  - rotated by 73 degrees counter-clockwise
- cylinder height $\hat{\delta} = 20h(1h)$ (= maximal time dependence)
- velocity vector is close to $(0, 0)$

$\Rightarrow$ clear **spatial anisotropy**, but no **dominant direction of movement**

Some elements of model comparison

Alternatively, we have also fitted two **threshold-censored Gaussian space-time copula processes** (separable model, frozen field model with velocity).

Based on the **composite likelihood information criterion** (CLIC), our model clearly outperforms the censored Gaussian models.

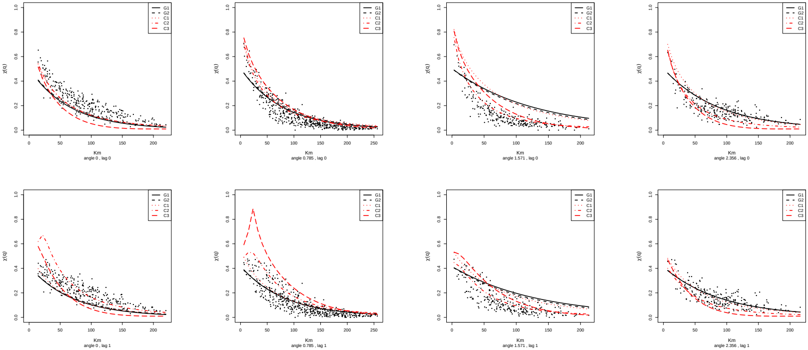
CLIC difference to velocity-free Gamma-Pareto model is very small.

Considering **root mean squared error of conditional exceedance probabilities**, Gamma-Pareto model performs better than Gaussian models at higher thresholds.

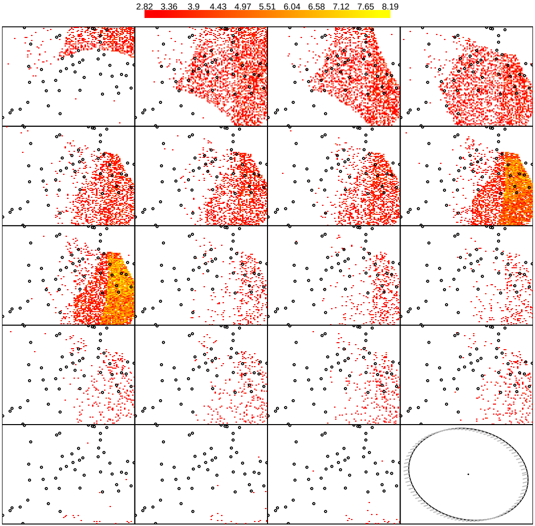
Visual goodness-of-fit diagnostics

Observed vs. fitted conditional exceedance probabilities.

Plots for different angles and lags, with respect to spatial distance.



A 19-hour simulation of our model



Surfaces are “rugged” due to conditional independence assumptions in the model. We could do small-scale smoothing to improve realism of precipitation surface.

Mediterrean precipitation extremes :

- **high-frequency space-time precipitation modeling** is very complex, even more when considering extremes
- **asymptotic independence**
- difficult to reveal velocity patterns
- our model allows for better understanding of the precipitation process and for extreme space-time scenario simulation

Methodology :

- **latent Gaussian hierarchical modeling :** combine tools of extreme value theory and Bayesian spatial modeling
- **latent Gaussian models are good for trends**, but not dependence
- **Gamma-Pareto models with physical interpretation of space-time dependence** ⇒**pairwise likelihood scales very well with high space-time dimension**
- perspectives for **Gamma-Pareto models** :
  - improve joint modeling/estimation of trends and dependence
  - conditional simulation and Bayesian inference of Gamma-Pareto models
  - **"subasymptotic" extensions**, more or less close to the asymptotic GP distribution

— Thank you — Merci — Danke —

Bacro, J.-N., Gaetan, C., Opitz, T., and Toulemonde, G. (2017).
Hierarchical space-time modeling of exceedances with an application to rainfall data.
arXiv :1708.02447.

Coles, S. (2001).
An Introduction to Statistical Modeling of Extreme Values.
Springer, London.

Coles, S., Heffernan, J., and Tawn, J. (1999).
Dependence measures for extreme value analyses.
Extremes, 2(4) :339–365.

Cooley, D., Nychka, D., and Naveau, P. (2007).
Bayesian spatial modeling of extreme precipitation return levels.
Journal of the American Statistical Association, 102(479) :824–840.

Davison, A. C. and Smith, R. L. (1990).
Models for exceedances over high thresholds (with discussion).
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 52(3) :393–442.

Opitz, T., Huser, R., Bakka, H., and Rue, H. (2018).
INLA goes extreme : Bayesian tail regression for the estimation of high spatio-temporal quantiles.
Extremes, 21(3) :441–462.