Space-time modelling and simulation of extreme rainfall

Gwladys Toulemonde[1,2]

[1] *IMAG, Université Montpellier, CNRS, France*

[2] *INRIA, Lemon, Montpellier, France*

with the members of the Cerise and Fraise lefe projects !

MASCOT 2021 Meeting

Proposition of a hierarchical space-time model for extreme precipitation data

s = "**Montpellier**", t= "30 Sept. 2014"

s = "**Venice**", t= "8 Nov. 2014"

Montpellier

JN Bacro

Avignon

T Opitz

Venice

C Gaetan

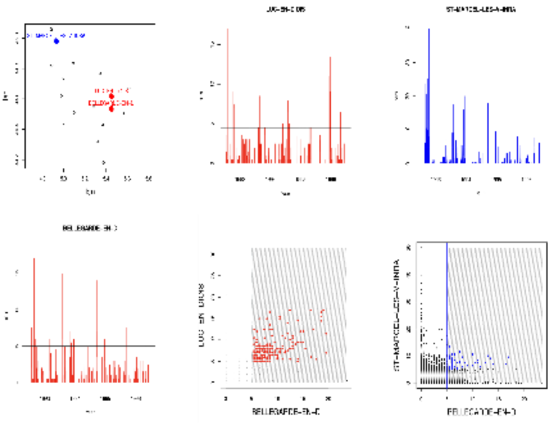
Cerise and Fraise LEFE projects

lefe

Rainfall data

- Hourly observations at 50 rainfall stations for the years 1993 to 2014 from September to November (54542 hours)
- Moderately large dataset (50×54542 observations)

Three stations in France



- clusters of strong values over space and time,
- strong variations at very small spatial and temporal scales

Bivariate max-stable distributions

Let $(X_i, Y_i) \sim F$ be independent random vectors with w.l.g. unit Fréchet margins $K(x) = \exp(-1/x)$, $x > 0$. If a non-degenerate limit distribution for $(M_{x,n}, M_{y,n}) = (\max_{i=1,\dots,n} X_i, \max_{i=1,\dots,n} Y_i)$ exists $(F \in D(G))$,

$$\lim_{n\to\infty} \mathbb{P}\left(M_{x,n} \le nx, M_{y,n} \le ny\right) = G(x,y)$$

then G is max-stable: $G^k(kx_1, kx_2) = G(x,y)$

- If

$$G(x,y) = K(x)\,K(y) = \exp\left(-\frac{1}{x}\right)\exp\left(-\frac{1}{y}\right)$$

$\hookrightarrow$ *ultimately, normalized maxima of X and Y are independent.*

(X,Y) are said to be Asymptotically Independent (AI).

Otherwise, (X,Y) are Asymptotically Dependent (AD).

Dependence measures χ and $\overline{\chi}$

Let $(X,Y) \sim F \in D(G)$, with F_X and F_Y margins.

The χ parameter

$$\chi = \lim_{u\to 1} \mathbb{P}\left(F_Y(Y) > u | F_X(X) > u\right)$$
$$= \lim_{u\to 1} 2 - \frac{\log \mathbb{P}(F_X(X)\le u, F_Y(Y)\le u)}{\log \mathbb{P}(F_X(X)\le u)}$$
$$\equiv \lim_{u\to 1} \chi(u)$$

- $\chi > 0 \Rightarrow X$ and Y are **AD**;

$\hookrightarrow$ χ quantifies the strength of the extremal dependence.

- $\chi = 0 \Rightarrow X$ and Y are **AI**.

$\hookrightarrow$ χ unable to provide dependence information for AI case !

The $\overline{\chi}$ parameter

$$\overline{\chi} = \lim_{u\to 1} \frac{2\log \mathbb{P}(F_X(X)>u)}{\log \mathbb{P}(F_X(X)>u, F_Y(Y)>u)} - 1$$
$$\equiv \lim_{u\to 1} \overline{\chi}(u)$$

- $\overline{\chi} = 1 \Rightarrow X$ and Y are **AD**.

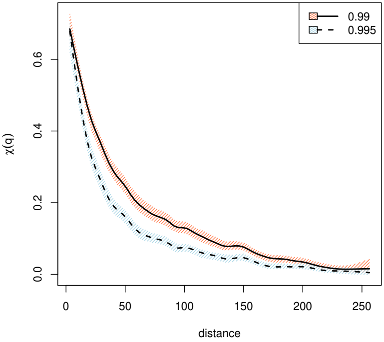
- $-1 \le \overline{\chi} < 1 \Rightarrow X$ and Y are **AI**; moreover $\overline{\chi}$ provides a measure that increases with dependence strength.

Example 1: Gaussian vectors with correlation parameter $\rho \neq 1$: $\chi = 0$, $\overline{\chi} = \rho$.

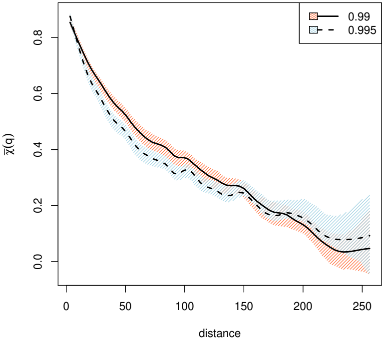
Example 2: For **max-stable distribution**, $\chi(u) = \chi$ (*same dependence structure* $\forall u$)

Our rainfall data: extremal dependence measure I

Spatial lag: $x = (s,t), x' = (s + h_s, t)$



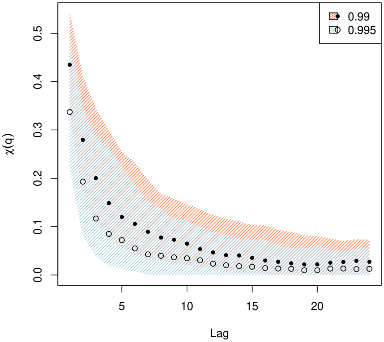
$\chi_{x,x'}(u)$



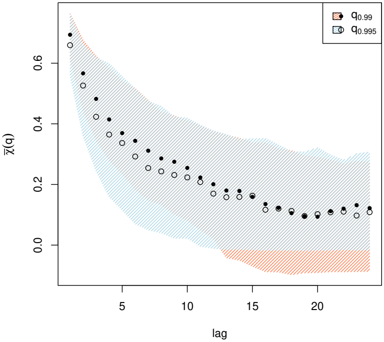
$\overline{\chi}_{x,x'}(u)$

Our rainfall data: extremal dependence measure II

Temporal lag: $x = (s,t), x' = (s, t + h_t)$



$\chi_{x,x'}(u)$



$\overline{\chi}_{x,x'}(u)$

Space-time setup

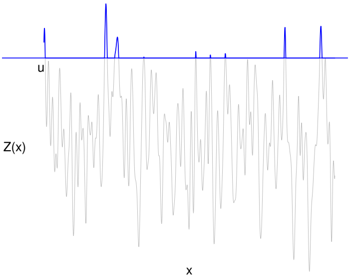
Bacro, J.N., Gaetan, C., Opitz, T., Toulemonde, G. (2020). "Hierarchical space-time modeling of exceedances with an application to rainfall data", *JASA*.

- $\{Z(x), x \in D\}$, **space-time** process where $x = (s, t)$, $D \subset \mathbb{R}^2 \times \mathbb{R}_+$
  - s space coordinate
  - t time coordinate
- Types of concern when dealing with extreme values of the processes:
  - accurate inference for marginal distributions
  - assessment of the space-time dependence of the extreme values

    Possibly **asymptotically independent**
- What *Extreme* means for a process? no unique definition

Exceedances

- Model for tail behaviour of $Z(x)$ by fixing a "high" threshold u and focusing only on the (left-censored) values above u (**exceedances**)



$\hookrightarrow$ We explicitly model the original event data

Marginal modelling: Generalized Pareto (GP) distribution

- Distribution for (censored) exceedances : the GP distribution

$\rightsquigarrow$ asymptotic justification for $u \to \tau_F$ (upper endpoint)

$$\begin{aligned} \Pr(Z(x) - u \leq y | Z(x) > u) &\simeq 1 - \left(1 + \xi \frac{y}{\psi}\right)_+^{-1/\xi} \\ &:= H(y; \xi, \psi), \qquad y \geq 0 \end{aligned}$$

- A different look at the GP distribution (when $\xi > 0$): GP distribution **as a Gamma mixture for the rate of the exponential distribution**:

$$V \mid G \sim Exp(G),\ G \sim \text{Gamma}(1/\xi, \psi/\xi) \ \Rightarrow V \sim GP(\cdot; \xi; \psi).$$

Hierarchical space-time process with GP marginals

Hierarchical formulation for exceedances (following an idea of Bortot and Gaetan, 2014)

$$Y(x) := (Z(x) - u)\,\mathbb{1}\{Z(x) > u\}$$

- latent space-time process *with Gamma marginals*

$$\mathbf{G(x)} \sim \mathsf{Gamma}(\alpha, \beta)$$

$\hookrightarrow$

$$Y(x) | [\mathbf{G(x)}, Y(x) > 0] \sim \mathsf{Exp}(\mathbf{G(x)})$$

$$P(Y(x) > 0 \,|\, \mathbf{G(x)}) = \exp(-\kappa \mathbf{G(x)})$$

where $\kappa > 0$ is a parameter controlling the rate of upcrossings of the threshold.

$\rightsquigarrow$ joint space-time structure of the zero part and the positive part in the distribution of $Y(x)$

Multivariate distribution over the threshold

Exploiting a direct connection between probabilities for $Y(\cdot)$ and $\mathscr{L}_G(\cdot)$, we obtain:

$$\boxed{\Pr(Z(x) > u) = \mathrm{E}[\Pr(Z(x) > u \mid G(x))] = \mathscr{L}_{G(x)}(\kappa)}$$

⇝ Data $\boldsymbol{z} = (z_1, \ldots, z_n)'$; for $\boldsymbol{z} \geq \boldsymbol{u}$,

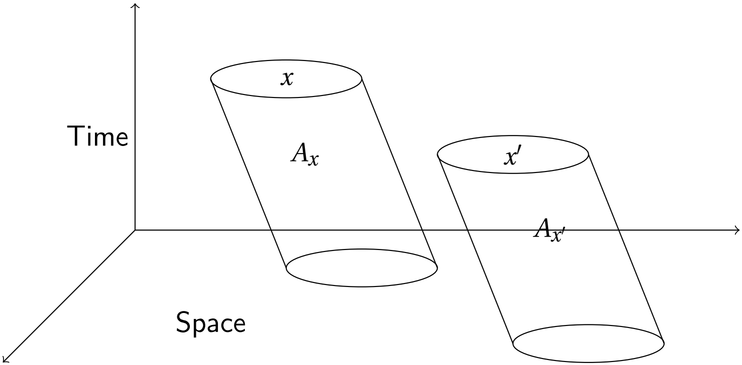
$$\boxed{\Pr(Z(x_1) > z_1, \ldots, Z(x_n) > z_n) = \mathscr{L}_{\boldsymbol{G}}(\boldsymbol{z} - (\boldsymbol{u} - \boldsymbol{\kappa}))}$$

↪ Multivariate densities can be evaluated as soon as $\mathscr{L}_{\boldsymbol{G}}$ is known.

- Notation for bivariate distribution with $z_1 > u$ and $z_2 > u$: $\Pr(Z(x_i) \leq z_i, Z(x_j) \leq z_j) = H(z_i, z_j)$

Which space-time process $G(\cdot)$ with Gamma marginals?

Based on **Slated elliptical cylinder + Gamma random field**
(Huser and Davison, 2014) (Wolpert and Ickstadt, 1998)



- The **ellipse** describes the spatial influence zone of a storm
- The ellipse (storm) moves through space with a **velocity** ω
- The **duration** of a storm is $\delta > 0$

Which space-time process $G(\cdot)$ with Gamma marginals?

We propose to model the space-time process $\{G(x), x \in \mathscr{D}\}$ as a **convolution** of a **Gamma random field** $\Gamma(\cdot)$ (Wolpert and Ickstadt, 1998)

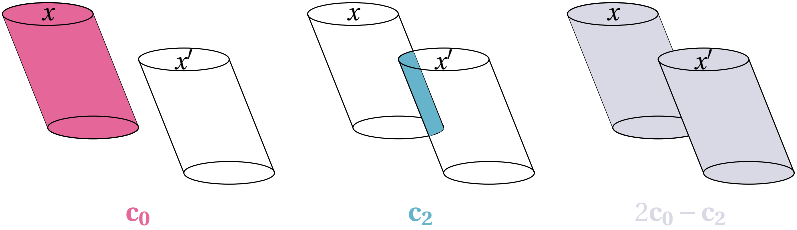
$$G(x) = \int_{A_x} \Gamma(dx') = \Gamma(A_x).$$

with $\Gamma(.)$ a Gamma RF defined on the space-time domain
$\mathscr{D} = \mathbb{R}^2 \times \mathbb{R}_+$
such that

- for any set A, $\Gamma(A) := \int_A \Gamma(dx) \sim \mathrm{Gamma}(\alpha(A), \beta)$;
- for any sets A_1, A_2 such that $A_1 \cap A_2 = \emptyset$, $\Gamma(A_1)$ and $\Gamma(A_2)$ are **independent** random variables.

Extremal dependence of $Z(\cdot)$: **Asymptotic Independence**

$$\chi_{\mathbf{x},\mathbf{x}'} = \mathbf{0} \text{ and } \overline{\chi}_{x,x'} = \frac{\mathbf{c_2}}{2\mathbf{c_0} - \mathbf{c_2}} \geq 0$$

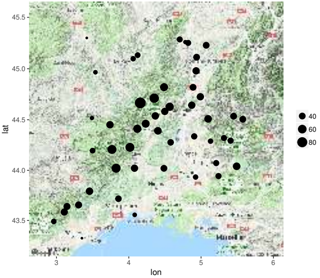


Application to rainfall data

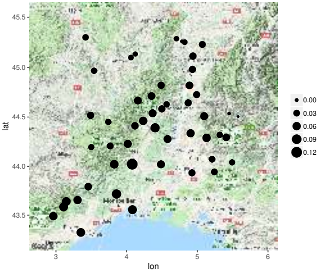
- Hourly observations at 50 rainfall stations for the years 1993 to 2014 from September to November (54542 hours)
- Moderately large dataset (50×54542 observations)

Application to rainfall data

- Marginal distributions are not stationary in space



99% quantile

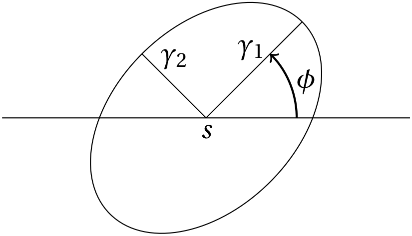


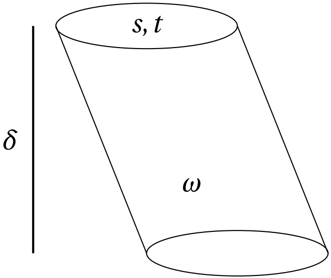
ξ (tail parameter)

- Fit a GP distribution separately to each site s with thresholds chosen as the empirical quantiles $q_{0.99}(s)$
- Transform the exceedances to the same GP distribution

Space-time dependence parameters

$\theta = (\phi, \gamma_1, \gamma_2, \delta, \omega)'$

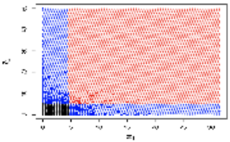




Inferential issues: composite likelihood approach

Let u be a sufficiently high threshold

- Different (censored) likelihood contribution $L(z_1, z_2; \theta)$ of $Z(x_1) = z_1$, $Z(x_2) = z_2$



- Weighted composite likelihood (Lindsay, 1988, Bevilacqua et al., 2012)

$$PL(\theta) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} L(z_i, z_j; \theta) w_{ij}$$

w_{ij} positive weights that depend on the space-time distance.

↪ Then we maximise pairwise weighted censored log-likelihood to obtain parameter estimations.

Application to rainfall data

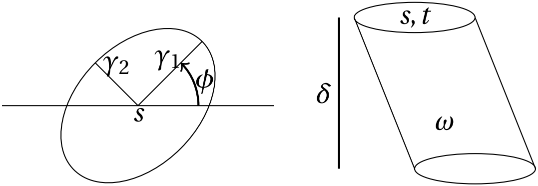
Two models for space-time dependence

G1 Gamma-Pareto process

G2 model G1 without velocity ($\omega = 0$)

Estimates, standard errors (in italic) values of the Gamma-Pareto fitted models.

$\theta = (\phi, \gamma_1, \gamma_2, \delta, \omega)'$



Model	Parameters					
	γ_1	γ_2	ϕ	δ	ω_1	ω_2
G1	165.062	318.823	0.085	20.184	0.723	0.446
	23.459	*19.811*	*0.026*	*0.948*	*0.195*	*0.009*
G2	175.817	294.323	0.041	20.036	0	0
	11.879	*25.291*	*0.064*	*1.039*	-	-

Parameter units are kms for γ_1 and γ_2, radians for ϕ, hours for δ and kms per hour for ω_1 and ω_2.

Comparison with other AI processes

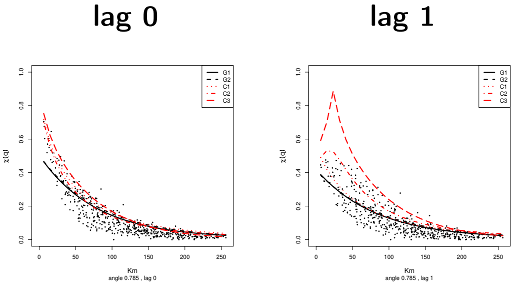
Comparison with three variants of a censored anisotropic Gaussian space-time copula.

- **C1** Space-time separable model
- **C2** Non-separable model (frozen field, Christakos, 2017)
- **C3** Non-separable model with spherical correlation function

Comparison according to

- **CLIC** (minimum for our Gamma-Pareto process G1)
- **Bivariate conditional probabilities** $\Pr(Z(s,t) > q | Z(s', t-h_t) > q)$
- **RMSE based on multivariate conditional probability** $\chi^*_{s_i;h_t}(q) := \Pr(Z(s_j,t) > q, s_j \in \partial s_i | Z(s_i, t-h_t) > q)$

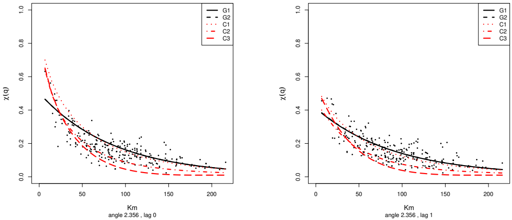
Angle $\pi/4$



Estimated probabilities $\Pr(Z(s,t) > q | Z(s', t-h_t) > q)$ *along different directions and at different temporal lags* h_t*. Dotted points correspond to empirical estimates. The value* q *is fixed to the empirical* 99% *quantile.*

Angle $3\pi/4$

lag 0 **lag 1**



Estimated probabilities $\Pr(Z(s,t) > q | Z(s', t-h_t) > q)$ *along different directions and at different temporal lags* h_t*. Dotted points correspond to empirical estimates. The value* q *is fixed to the empirical* 99% *quantile.*

RMSE

- Compute
  - empirical estimates $\widehat{p}_i(h_t)$ of the multivariate conditional probability

$$\chi^*_{s_i;h_t}(q) := \Pr(Z(s_j,t) > q, s_j \in \partial s_i | Z(s_i, t-h_t) > q)$$

    where ∂s_i is the set of the four nearest neighbors of site s_i, $i = 1, \ldots, 50$.
  - Monte-Carlo estimates $\widetilde{p}_i^{(j)}(h_t)$, $j = 1, \ldots, 200$.
- Calculate site-specific root mean squared errors (RMSE)

$$\mathrm{RMSE}_i(h_t) = \left\{ \frac{\sum_{j=1}^{200} (\widetilde{p}_i^{(j)}(h_t) - \widehat{p}_i(h_t))^2}{200} \right\}^{1/2},$$

and the resulting total $\mathrm{RMSE}(\mathbf{h_t}) = \sum_{\mathbf{i=1}}^{\mathbf{50}} \mathrm{RMSE}_{\mathbf{i}}(\mathbf{h_t})$.

RMSE

	RMSE(0)		RMSE(1)		RMSE(2)	
	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$
G1	2.614	**2.096**	**1.901**	**1.643**	**1.475**	**1.496**
G2	2.605	**2.072**	**1.907**	**1.626**	**1.477**	**1.480**
C1	**2.240**	2.455	2.053	2.428	1.779	1.928

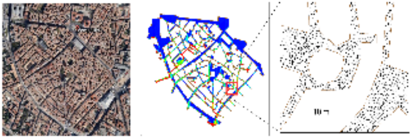
Table 1: Total root mean squared errors for the estimates of $\chi^*_{s_i;h_t}(q)$ with $h_t = 0, 1, 2$ hours.

Conclusions on this part

- A **space-time** model **for threshold exceedances** of data with **asymptotically vanishing dependence strength** with **physical interpretation**.
- Extensions to asymptotic dependence are possible.
- Simulations of exceedances

Why simulate extreme rainfall ?

Example of simulated water level and speed in Montpellier with a urban flood model.



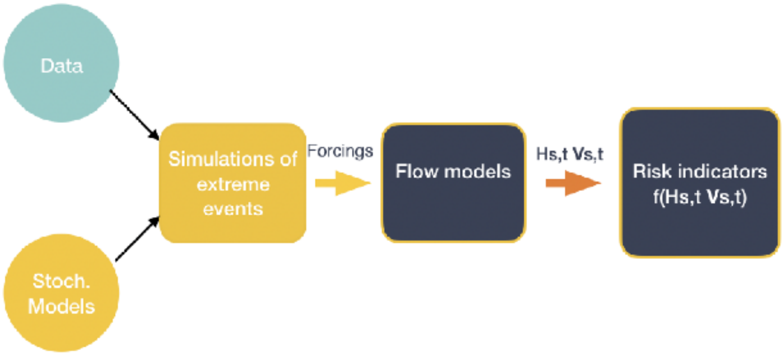
Left : study area (600m x 600m). Center : simulated water depths. Right: detail view of the mesh. The lowest depths in blue and the largest depths (5 cm) in red

↪ Input for urban flood models: **rainfall forcing**.

- Exploration of not already observed scenarios from limited observations
- Stochastic inputs for impact studies

Reconstructing extreme space-time rainfall forcing scenarios as close to reality as possible is therefore a crucial issue.

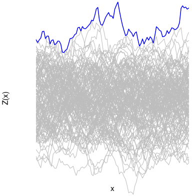
Urban flood risk study



Toulemonde, G., Carreau, J., Guinot, V. (2020). "Space-time simulations of extreme rainfall : why and how ?" in S. Manou-Abi, S. Dabo-Niang, J. Salone (eds), Mathematical Modeling of Random and Deterministic Phenomena, Wiley.

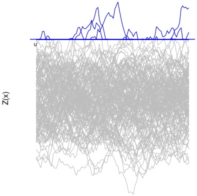
Which extremal behaviour of $Z = \{Z(x), x \in D\}$?

- what does it mean rainfall extreme we would like to simulate ?
- Events satisfying an exceedance condition



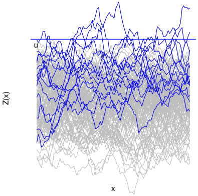
$\{\max_i Z_i(x)\}$

Max-stable



$\{\max(Z(x), u)\}$

Gamma-Pareto processes



$\{Z(x) | \sup_{x \in D} Z(x) > u\}$

Pareto processes

ℓ-Pareto process

$\{Z(x), x \in D | \ell(Z(x)) > u\}$

Semi-parametric simulation method

(Chailan, R., Toulemonde, G., and Bacro, J. (2017); Palacios-Rodriguez F., Toulemonde G., Carreau J., Opitz T. (2020))

Construction of standard space-time ℓ-Pareto processes

(Based on Ferreira and de Haan, 2014; Dombry and Ribatet, 2015)

$$Z(s,t) := R\mathbf{Y}(\mathbf{s},\mathbf{t})$$

with $R \sim \text{Pareto}(1, \gamma_R)$ independent of $Y(s,t) \geq 0$, $\ell(Y(s,t)) = 1$ with ℓ a cost functional (a continuous non negative function that is homogeneous).

Semi-parametric simulation method

- **Standardisation** $\{Z^*(s,t), s \in \mathscr{S}, t \in \mathscr{T}\}$ the Pareto standardised process.
- **Extraction**
  - Defining extreme episodes → Cost functional ℓ+ threshold u
  - Select the m most extreme episodes $\{Z^*_{[i]}(s,t), s \in \mathscr{S}_i \subset \mathscr{S}, t \in \mathscr{T}_i \subset \mathscr{T}\}$, $i \in \{1,\ldots,m\}$

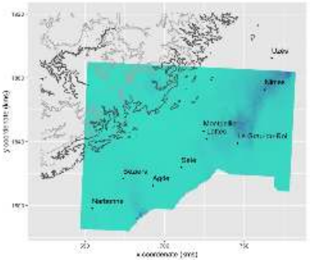
For each $i \in \{1,\ldots,m\}$,

- **Lifting procedure**
  - Non-parametric approach for the dependence structure
  - Sample R_i according to a Pareto r.v. with shape 1 and scale $\alpha > 0$ and generate

$$V_i(s,t) = R_i \frac{Z^*_{[i]}(s,t)}{\ell(Z^*_{[i]}(s,t))} = R_i Y_i(s,t).$$

- **Back-transformation to original scale**

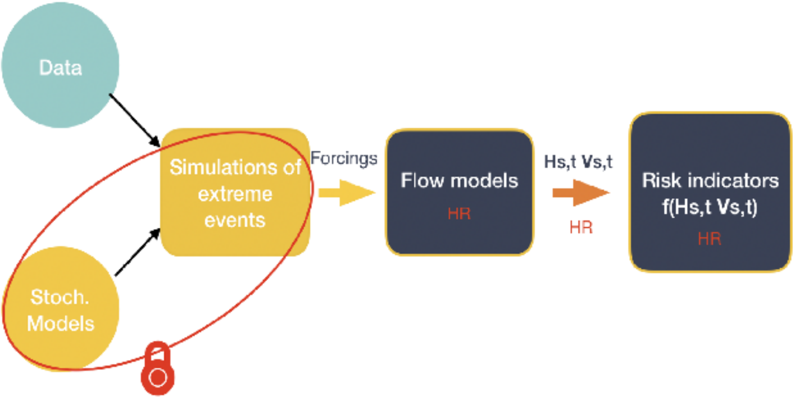
Application to precipitation in Mediterranean France



- Reanalysis data-set
- Hourly rainfall totals (mm).
- $133 kms \times 104 kms$ grid with $1 km$ spatial resolution.
- Years: from 1997 to 2007. $N = 87642$ hours time steps.

- ℓ: Space-time neighborhoods(15 kms, $24h$)
- $u = 0.99$-quantile.

Some perspectives about urban flood risk study



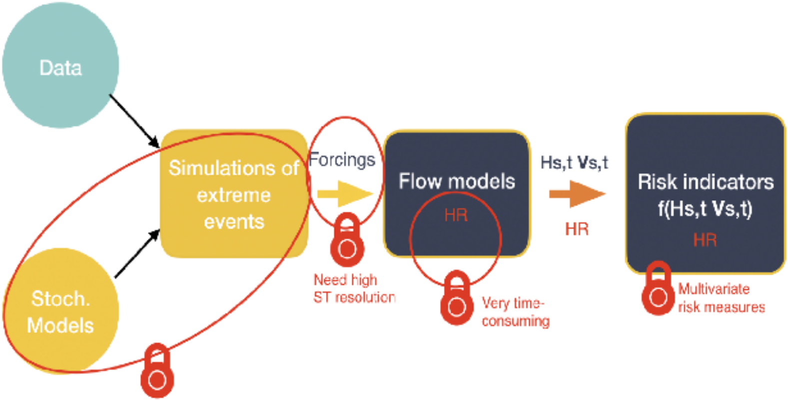
Statistical modelling of extreme events

- Framework:
  - **multivariate**,
  - **temporal**,
  - **spatial**

  → taking into account associated complex dependence.

- Three main issues **(I1)**, **(I2)** and **(I3)**
  - **(I1) Asymptotic independence** (hybrid according components)
  - **(I2) Spatial and/or temporal non-stationarity of the dependence structure** (Carreau J., Toulemonde G., 2020)
  - **(I3) Combination of extreme and non-extreme events**.

Some perspectives about urban flood risk study



Some references

- Bacro, J.N., Gaetan, C., Opitz, T., Toulemonde, G. (2020). Hierarchical space-time modeling of exceedances with an application to rainfall data, *JASA*, 115(530), 555-569.
- Carreau J., Toulemonde G. (2020). Spatial dependence structure for flood-risk rainfall. *Spatial Statistics*, 40.
- Chailan, R., Toulemonde, G., and Bacro, J. (2017). A semiparametric method to simulate bivariate space-time extremes. *The Annals of Applied Statistics*, 11(3):1403-1428.
- Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717-1737.
- Huser, R., Davison, A. C. (2014). Space-time modelling of extreme events, *JRSS B*, 76, 439-461.
- Palacios-Rodriguez F., Toulemonde G., Carreau J., Opitz T. (2020). Generalized Pareto processes for exploring and simulating space-time extremes : application to rainfall data. *SERRA*, 34, 2033-2052.
- Toulemonde, G., Carreau, J., Guinot, V. (2020). Space-time simulations of extreme rainfall : why and how ? in *Mathematical Modeling of Random and Deterministic Phenomena*, Wiley.
- Wolpert, R. L., Ickstadt, K. (1998). Poisson/Gamma random fields for spatial statistics, *Biometrika*, 85, 251-267.

THANK YOU !