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

\[ s = \text{"Montpellier"}, \quad t = \text{"30 Sept. 2014"} \]

\[ s = \text{"Venice"}, \quad t = \text{"8 Nov. 2014"} \]
Rainfall data

- Hourly observations at 50 rainfall stations for the years 1993 to 2014 from September to November (54542 hours)
- Moderately large dataset ($50 \times 54542$ 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 \((M_{x,n}, M_{y,n}) = (\max_{i=1,...,n} X_i, \max_{i=1,...,n} Y_i)\) exists \((F \in D(G))\),

\[
\lim_{n \to \infty} \mathbb{P}(M_{x,n} \leq nx, M_{y,n} \leq ny) = G(x, y)
\]

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

- If

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

\(\leftarrow \text{ultimately, normalized maxima of } X \text{ and } Y \text{ are independent.}\)

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

Otherwise, \((X, Y)\) are **Asymptotically Dependent (AD)**.
Dependence measures $\chi$ and $\overline{\chi}$

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

The $\chi$ parameter

$$\chi = \lim_{u \to 1} \mathbb{P} \left( F_Y(Y) > u \mid F_X(X) > u \right)$$

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

$$\equiv \lim_{u \to 1} \chi(u)$$

- $\chi > 0 \Rightarrow X$ and $Y$ are AD;
- $\chi = 0 \Rightarrow X$ and $Y$ are AI.

$\chi$ quantifies the strength of the extremal dependence.

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.
- $-1 \leq \overline{\chi} < 1 \Rightarrow X$ 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: \( x = (s, t), x' = (s, t + h_t) \)

\[ \chi_{x,x'}(u) \]

\[ \bar{\chi}_{x,x'}(u) \]
Space-time setup


- \( \{Z(x), x \in D\} \), **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)

$\rightarrow$ We explicitly model the original event data
Marginal modelling: Generalized Pareto (GP) distribution

- Distribution for (censored) exceedances: the GP distribution
- Asymptotic justification for \( u \to \tau_F \) (upper endpoint)

\[
\Pr(Z(x) - u \leq y | Z(x) > u) \approx 1 - \left(1 + \frac{\xi}{\psi} \frac{y}{\psi}\right)^{-1/\xi}_+
\]

\[
:= H(y; \xi, \psi), \quad y \geq 0
\]

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

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

\[ G(x) \sim \text{Gamma}(\alpha, \beta) \]

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

\[ P(Y(x) > 0 | G(x)) = \exp(-\kappa G(x)) \]

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

\[ \leadsto \text{joint space-time structure of the zero part and the positive part in the distribution of } Y(x) \]
Multivariate distribution over the threshold

Exploiting a direct connection between probabilities for $Y(\cdot)$ and $L_G(\cdot)$, we obtain:

$$\Pr(Z(x) > u) = E[\Pr(Z(x) > u \mid G(x))] = L_G(x)(\kappa)$$

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

$$\Pr(Z(x_1) > z_1, \ldots, Z(x_n) > z_n) = L_G(z - (u - \kappa))$$

Multivariate densities can be evaluated as soon as $L_G$ 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 $\omega$
- 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 $\{G(x), x \in \mathcal{D}\}$ 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 $\Gamma(\cdot)$ a Gamma RF defined on the space-time domain $\mathcal{D} = \mathbb{R}^2 \times \mathbb{R}_+$ such that

- for any set $A$, $\Gamma(A) := \int_A \Gamma(dx) \sim \text{Gamma}(\alpha(A), \beta)$;
- 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

$$\chi_{x,x'} = 0 \text{ and } \bar{\chi}_{x,x'} = \frac{c_2}{2c_0 - c_2} \geq 0$$
Application to rainfall data

- Hourly observations at 50 rainfall stations for the years 1993 to 2014 from September to November (54,542 hours)
- Moderately large dataset (50 × 54,542 observations)
Application to rainfall data

- Marginal distributions are not stationary in space

- 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.

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

Parameter units are kms for \( \gamma_1 \) and \( \gamma_2 \), radians for \( \phi \), hours for \( \delta \) and kms per hour for \( \omega_1 \) and \( \omega_2 \).
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.
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.
Compute

empirical estimates $\hat{p}_i(h_t)$ 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 $\partial s_i$ is the set of the four nearest neighbors of site $s_i$, $i = 1, \ldots, 50$.

Monte-Carlo estimates $\tilde{p}_i^{(j)}(h_t)$, $j = 1, \ldots, 200$.

Calculate site-specific root mean squared errors (RMSE)

$$\text{RMSE}_i(h_t) = \left\{ \frac{\sum_{j=1}^{200} (\tilde{p}_i^{(j)}(h_t) - \hat{p}_i(h_t))^2}{200} \right\}^{1/2},$$

and the resulting total $\text{RMSE}(h_t) = \sum_{i=1}^{50} \text{RMSE}_i(h_t)$. 

RMSE
<table>
<thead>
<tr>
<th></th>
<th>RMSE(0)</th>
<th></th>
<th></th>
<th>RMSE(1)</th>
<th></th>
<th></th>
<th>RMSE(2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q_{0.99}$</td>
<td>$q_{0.995}$</td>
<td>$q_{0.99}$</td>
<td>$q_{0.995}$</td>
<td>$q_{0.99}$</td>
<td>$q_{0.995}$</td>
<td>$q_{0.99}$</td>
<td>$q_{0.995}$</td>
</tr>
<tr>
<td>G1</td>
<td>2.614</td>
<td>2.096</td>
<td>1.901</td>
<td>1.643</td>
<td>1.475</td>
<td>1.496</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>2.605</td>
<td>2.072</td>
<td>1.907</td>
<td>1.626</td>
<td>1.477</td>
<td>1.480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>2.240</td>
<td>2.455</td>
<td>2.053</td>
<td>2.428</td>
<td>1.779</td>
<td>1.928</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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

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

\[
\begin{align*}
\{\max_i Z_i(x)\} & \quad \text{Max-stable} \\
\{\max(Z(x), u)\} & \quad \text{Gamma-Pareto processes} \\
\{Z(x) | \sup_{x \in D} Z(x) > u\} & \quad \text{Pareto processes} \\
\ell\text{-Pareto process} & \quad \{Z(x), x \in D | \ell(Z(x)) > u\}
\end{align*}
\]
Semi-parametric simulation method


Construction of standard space-time $\ell$-Pareto processes

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

\[ Z(s, t) := R Y(s, t) \]

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

- **Standardisation** \( \left\{ Z^*(s, t), s \in \mathcal{S}, t \in \mathcal{T} \right\} \) the Pareto standardised process.

- **Extraction**
  - Defining extreme episodes \( \rightarrow \) Cost functional \( \ell + \) threshold \( u \)
  - Select the \( m \) most extreme episodes
    \( \left\{ Z^*_{[i]}(s, t), s \in \mathcal{S}_i \subset \mathcal{S}, t \in \mathcal{T}_i \subset \mathcal{T} \right\}, \ i \in \{1, \ldots, m\} \)

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 \((mm)\).
- \(133\text{kms} \times 104\text{kms}\) grid with \(1\text{km}\) spatial resolution.
- Years: from 1997 to 2007. \(N = 87642\) hours time steps.

- \(\ell\): Space-time neighborhoods(15 kms, 24h)
- \(u = 0.99\)-quantile.
Some perspectives about urban flood risk study

Flow models

Risk indicators $f(H_s,t, V_s,t)$

Simulations of extreme events

Forcings

Stoch. Models

Data
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

Data → Simulations of extreme events → Flow models → Risk indicators

Forcings: Need high ST resolution

HR: Very time-consuming

Risk indicators: $f(H_s,t, V_s,t)$

Multivariate risk measures
Some references


THANK YOU!