Spatio-temporal modeling of urban extreme rainfall events at high resolution

Chloé Serre-Combe [1] & Nicolas Meyer [2] & Thomas Opitz [3] & Gwladys Toulemonde [4]

[1] *Univ. Montpellier, CNRS, IMAG, Inria Lemon, France,* *chloe.serre-combe@umontpellier.fr*
[2] *Univ. Montpellier, CNRS, IMAG, Inria Lemon, France,* *nicolas.meyer@umontpellier.fr*
[3] *INRAE, BioSP, Avignon, France,* *thomas.opitz@inrae.fr*
[4] *Univ. Montpellier, CNRS, IMAG, Inria Lemon, France,* *gwladys.toulemonde@umontpellier.fr*

Résumé. La modélisation des précipitations est essentielle pour analyser les risques d'inondation. Nous proposons un cadre intégrant un réseau de pluviomètres de Montpellier et des données radar de Météo-France pour modéliser la distribution et la dépendance spatio-temporelle des précipitations à haute résolution temporelle et fine échelle spatiale. La loi de Pareto généralisée étendue (EGPD) est utilisée pour représenter en univarié simultanément les précipitations modérées et extrêmes sans choix de seuil, simplifiant ainsi l'estimation des paramètres. La dépendance spatio-temporelle est capturée par un processus de r-Pareto avec une structure de dépendance gaussienne et un variogramme spatio-temporel non séparable avec advection, permettant des simulations plus réalistes de précipitation. Une approche de vraisemblance composite, basée sur des indicateurs de dépassement conjoints bivariés, est utilisée pour estimer les paramètres du variogramme. Validée par simulations, cette approche sera appliqué aux données de Montpellier pour ensuite servir de base à la construction d'un générateur stochastique de précipitations.

Mots-clés. Théorie des valeurs extrêmes, EGPD, haute résolution spatio-temporelle, processus de r-Pareto, advection

Abstract. Precipitation modeling is of great interest for flood risk analysis. We propose a framework integrating a rain gauge network in Montpellier and Météo-France radar data to model the distribution and spatio-temporal dependence of rainfall at high temporal resolution and fine spatial scale. The Extended Generalized Pareto Distribution (EGPD) is used for univariate modeling, capturing both moderate and extreme rainfall without threshold selection, simplifying parameter estimation. Spatio-temporal dependence is modeled using an r-Pareto process with a Gaussian dependence structure and a non-separable variogram with advection, enabling more realistic precipitation simulations. A composite likelihood approach based on bivariate joint exceedance indicators is used for variogram parameter estimation. Validated through simulations, this approach will be applied to Montpellier data as a basis for developing a stochastic precipitation generator.

Keywords. Extreme value theory, EGPD, high spatio-temporal resolution, r-Pareto process, advection

1 Context and study area

Montpellier, with its Mediterranean climate and urbanization, faces heavy rainfall events leading to flooding. Its proximity to the Cévennes mountains exacerbates flood risks, highlighting the need for precise rainfall simulations for flood assessment and urban planning. The localized nature of these events underscores the importance of high-resolution spatio-temporal rainfall modeling.

(a) City district division

(b) Rain gauges

Figure 1: Location of the 17 rain gauges of the OMSEV rain gauges network

The study focuses on the OMSEV[1] rain gauge network, which includes 17 stations in the Verdanson catchment, a tributary of the Lez (see Figure 1). Rainfall data, provided by HydroScience Montpellier (Finaud-Guyot et al., 2023), spans from 2019 to 2022 with minute-level measurements, aggregated into 5-minute intervals to reduce errors while maintaining high temporal resolution. Due to this fine scale, only 1.2% of the data consists of non-zero values. The station spacing ranges from 77 to 1531 meters, highlighting the fine spatial resolution. Additionally, the COMEPHORE dataset from Météo France (Tabary et al., 2012) will be considered. This is reanalysis data from 1997 to 2021, providing hourly cumulative rainfall at a 1 km^2 resolution by combining radar and gauge measurements for better spatial consistency. The main objective is to model rainfall's spatio-temporal variability in the Montpellier area and simulate realistic rainfall events using a stochastic precipitation generator, integrating both datasets.

2 Modeling univariate margins

To model the marginal distribution of positive rainfall, the Extended Generalized Pareto Distribution (EGPD) is used. This distribution avoids explicit threshold selection and efficiently captures both moderate and extreme events. The EGPD is a transformation of the Generalized Pareto Distribution (GPD), which models threshold exceedances with shape parameter

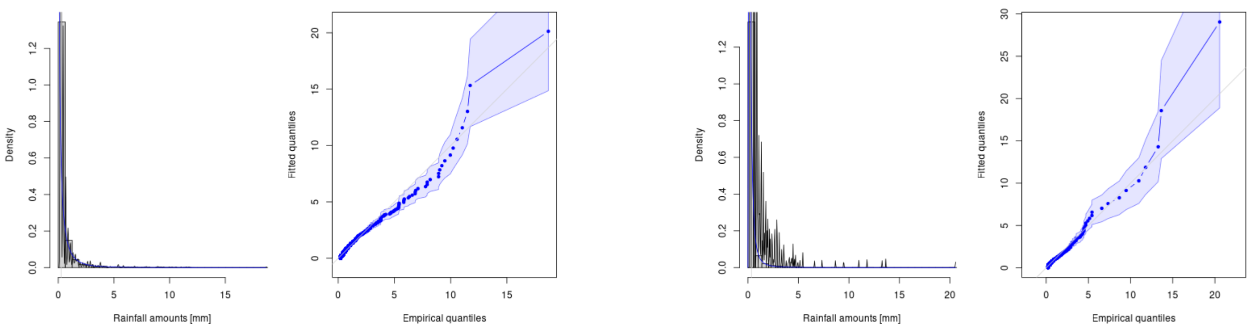
[1] Observatoire Montpelliérain et au Sud de l'Eau dans la Ville

ξ and a scale parameter σ_u, with u a fixed threshold u. The cumulative distribution function is approximated, for a rainfall intensity $X_{\boldsymbol{s}}$ at location $\boldsymbol{s} \in \mathbb{R}^2$, as

$$\mathbb{P}\left(X_{\boldsymbol{s}} - u > y | X_{\boldsymbol{s}} > u\right) \approx \overline{H}_\xi\left(\frac{y}{\sigma_u}\right) = \begin{cases} \left(1 + \xi \frac{y}{\sigma_u}\right)_+^{-1/\xi} & \text{if } \xi \neq 0\,, \\ e^{-\frac{y}{\sigma_u}} & \text{if } \xi = 0\,, \end{cases}$$

with $a_+ = \max(a, 0)$. A continuous function $G(x) = x^\kappa$ is used to modify the GP distribution, where κ controls the moderate part of the distribution. Then, the upper tail still follows a GP distribution, while the lower tail is based on a GP distribution for negative values with an upper bound of 0, leading to a power-law behavior.

The EGPD is fitted to the OMSEV data, with small left-censoring adjustments applied at each station to improve model performance, according the Root Mean Square Error (RMSE) as a goodness-of-fit measure (Haruna et al., 2023). Good fits are obtained for all rain gauges, as shown in the quantile-quantile plot in Figure 2. So, the EGPD model is a good candidate for modeling the marginal distribution of rainfall at each station.



(a) CRBM (b) CNRS

Figure 2: EGPD fitting on two rain gauges of the OMSEV network with the density estimation (left) and the quantile-quantile plot (right).

3 Spatio-temporal dependence modeling

Let $\boldsymbol{X} = \{X_{\boldsymbol{s},t} \mid (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}\}$ be a strictly stationary isotropic random process that represents the rainfall field over the study area. The spatio-temporal dependence is modeled by an r-Pareto process, as described by de Fondeville and Davison, 2018, which is an alternative to max-stable processes for threshold exceedances.

Let's consider a nonnegative and 1-homogeneous risk function $r(\boldsymbol{X}) = X_{\boldsymbol{s}_0,t_0}$. As described in Dombry et al., 2024, the r-exceedances of $\boldsymbol{X}$, defined by $\{X_{\boldsymbol{s},t} \mid r(\boldsymbol{X}) > u\}$, converge in distribution to an r-Pareto process $\boldsymbol{Y} = \{Y_{\boldsymbol{s},t}, (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}\}$ when u goes to infinity. This process is defined by a Gaussian process $\boldsymbol{W}$, associated with a variogram γ, that captures the spatial and temporal dependence. The r-Pareto process is given by

$$Y_{\boldsymbol{s},t} = u R_{\boldsymbol{s},t} e^{W_{\boldsymbol{s},t} - W_{\boldsymbol{s}_0,t_0} - \gamma_r(\boldsymbol{s},t)}\,,$$

where $R_{\boldsymbol{s},t}$ follows a univariate standard Pareto distribution, and $\gamma_r(\boldsymbol{s}, t) = \gamma(\boldsymbol{s} - \boldsymbol{s}_0, t - t_0)$. Our goal is to model the dependence structure of $\boldsymbol{Y}$, which is directly determined by the dependence structure of $\boldsymbol{W}$, given by the spatio-temporal variogram γ. This is achieved by examining the space-time extremogram, which measures the extreme spatio-temporal dependence by analyzing the simultaneous excesses for pairs of locations with spatial and temporal lags. Let $\Lambda_{\mathcal{S}} \subset \mathbb{R}^2$ and $\Lambda_{\mathcal{T}} \subset \mathbb{R}_+$ be sets of spatial and temporal lags. The extremal dependence measure, or extremogram, is defined for $\boldsymbol{h} \in \Lambda_{\mathcal{S}}, \tau \in \Lambda_{\mathcal{T}}$ as (Coles et al., 1999):

$$\chi(\boldsymbol{h}, \tau) = \lim_{q \to 1} \chi_q(\boldsymbol{h}, \tau), \quad \text{with} \quad \chi_q(\boldsymbol{h}, \tau) = \mathbb{P}(X^*_{\boldsymbol{s},t} > q \mid X^*_{\boldsymbol{s}+\boldsymbol{h},t+\tau} > q),$$

where $q \in [0, 1[$ and for any $\boldsymbol{s} \in \mathcal{S}$ and any $t \in \mathcal{T}$, $X^*_{\boldsymbol{s},t}$ is the uniform rank transformation of $X_{\boldsymbol{s},t}$. For r-Pareto processes, the r-extremogram χ_r can be defined as the extremogram of the r-exceedances of our rainfall process $\boldsymbol{X}$, which will be directly the marginals of the process. The dependence structure of r-Pareto processes, which is inherited from the Gaussian framework, leads to a relationship between the r-variogram and the r-extremogram (Davis et al., 2013 and Coles et al., 1999):

$$\chi_r(\boldsymbol{h}, \tau) = 2\left(1 - \phi\left(\sqrt{\frac{1}{2}\gamma_r(\boldsymbol{h}, \tau)}\right)\right), \ \boldsymbol{h} \in \Lambda_{\mathcal{S}}, \ \tau \in \Lambda_{\mathcal{T}}.$$

where ϕ is the standard normal distribution function.

4 Estimation of the variogram parameters

4.1 Variogram structure

To model spatial-temporal dependencies in rainfall, we use a non-separable variogram of a fractional Brownian motion, which accounts for advection, which is the horizontal transport of air masses. The variogram, incorporating a velocity vector $\boldsymbol{V} \in \mathbb{R}^2$, is given by:

$$\gamma(\boldsymbol{h}, \tau) = 2\left(\beta_1 \|\boldsymbol{h} - \tau \boldsymbol{V}\|^{\alpha_1} + \beta_2 |\tau|^{\alpha_2}\right), \quad \boldsymbol{h} \in \Lambda_{\mathcal{S}}, \tau \in \Lambda_{\mathcal{T}},$$

with $\beta_1 > 0, \beta_2 > 0, 0 < \alpha_1, \alpha_2 \leq 2$. If $\boldsymbol{V} = \boldsymbol{0}$, the variogram is separable in space and time.

4.2 Composite likelihood approach

The variogram parameters $\boldsymbol{\Theta} = (\beta_1, \beta_2, \alpha_1, \alpha_2, \boldsymbol{V})$ are estimated by maximizing the composite likelihood function. In spatio-temporal processes, the dependence structure of extremes can be characterized by the total number of joint exceedances across site pairs separated by a spatial lag $\boldsymbol{h}$ and a temporal lag τ:

$$K_{\boldsymbol{h},\tau}(u) = \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in \mathcal{N}(\boldsymbol{h})} \sum_{(t_i, t_j) \in \mathcal{N}(\tau)} \mathbb{1}_{\{X_{\boldsymbol{s}_i,t_i} > u, X_{\boldsymbol{s}_j,t_j} > u\}}.$$

where $\mathcal{N}(\boldsymbol{h})$ and $\mathcal{N}(\tau)$ are the sets of all pairs of sites separated by the spatial lag $\boldsymbol{h} \in \Lambda_{\mathcal{S}}$ and all pairs of time indices separated by the temporal lag $\tau \in \Lambda_{\mathcal{T}}$, respectively.

For considered r-Pareto processes, the indicator of joint exceedances becomes $k_{\boldsymbol{s},t} = \mathbb{1}_{\{X_{\boldsymbol{s}_0,t_0}>u, X_{\boldsymbol{s},t}>u\}}$ for $(\boldsymbol{s}, \boldsymbol{s}_0) \in \mathcal{N}(\boldsymbol{h})$ and $(t, t_0) \in \mathcal{N}(\tau)$. Then, since t_0 is fixed, the total number of temporal pairs within temporal lag τ is 1, while the number of spatial pairs corresponds to the number of pairs of sites separated by $\boldsymbol{h}$ from the site $\boldsymbol{s}_0$. When u goes to infinity, the total of joint exceedances follow a binomial distribution, with the number of trials equal to the number of spatial pairs and the probability of success is directly given by the r-extremogram and the marginal probability of exceedance above u, which is $\mathbb{P}(X_{\boldsymbol{s}_0,t_0} > u) = 1$. Then, the composite log-likelihood to maximize is given by

$$l_C(\boldsymbol{\Theta}) \propto \sum_{(\boldsymbol{s}_0,t_0)\in\mathcal{R}} \sum_{(\boldsymbol{s},\boldsymbol{s}_0)\in\mathcal{N}(\boldsymbol{h})} \sum_{(t,t_0)\in\mathcal{N}(\tau)} k_{\boldsymbol{s},t} \log \chi_{r,\boldsymbol{\Theta}}(\boldsymbol{h}, \tau) + (1 - k_{\boldsymbol{s},t}) \log(1 - \chi_{r,\boldsymbol{\Theta}}(\boldsymbol{h}, \tau)).$$

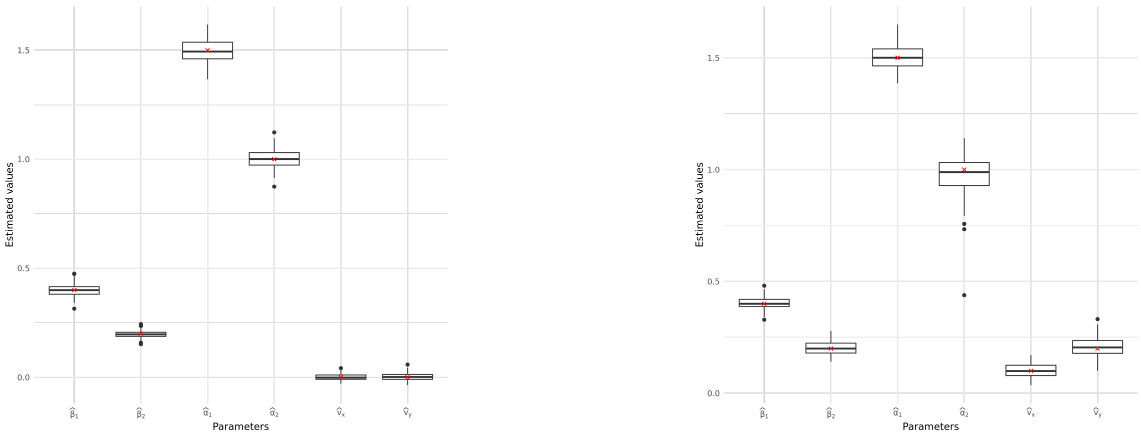
where $\mathcal{R}$ denotes the set of conditioning spatio-temporal points for which there is an exceedance, where no two points $(\boldsymbol{s}, t)$ and (s', t') in $\mathcal{R}$ are within the same spatial and temporal neighbourhood. For real data applications, the optimization is initialized following the approach of Buhl et al., 2019, using a separable variogram and Weighted Least Squares estimation of the parameters.

4.3 Model validation on simulations

The simulations of r-Pareto processes were performed by adapting the methodology of the exact simulation of spatial max-stable processes proposed by Dombry et al., 2016 and the algorithm of Leber, 2015 for simulating spatio-temporal Brown-Resnick processes. The algorithm was modified to incorporate a Pareto random variable and it simulates only one Gaussian process. Simulations were conducted on a regular grid with a threshold $u = 1$, defining the r-exceedances. The risk function $r(\boldsymbol{X}) = X_{\boldsymbol{s}_0,t_0}$ was used, conditionning the process on the first site and time step. The variogram parameter estimation was conducted on 100 simulated r-Pareto processes with 1000 replicates each, producing excellent results, as illustrated in the boxplots of Figure 3, with constant advection and without advection.

5 Outlook

The proposed approach will be applied to real COMEPHORE data to estimate an advection vector for each extreme episode, where observed wind speed data will be used as a covariate to include event-specific information on advection. The resulting advection will then be used as the constant advection parameter to optimize the other variogram parameters on OMSEV data. Finally, this process will allow the simulation of realistic rainfall events, improving the analysis of flood risks in the Montpellier area. To facilitate its use and reproducibility, we aim to develop it as an `R` package. The code is currently available on GitHub: chloesrcb/generain.



(a) Without advection.

(b) With advection.

Figure 3: Estimations of variogram parameters with maximum likelihood optimization of 100 simulated r-Pareto processes with 1000 replicates each, with 25 sites and 30 time observations. The true parameters are given by red crosses: $\mathbf{\Theta} = (0.4, 0.2, 1.5, 1, (0, 0))$ (left) and $\mathbf{\Theta} = (0.4, 0.2, 1.5, 1, (0.1, 0.2))$ (right).

References

Buhl, S., Davis, R. A., Klüppelberg, C., & Steinkohl, C. (2019). Semiparametric estimation for isotropic max-stable space-time processes. *Bernoulli*, *25*.

Coles, S., Heffernan, J., & Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*.

Davis, R. A., Klüppelberg, C., & Steinkohl, C. (2013). Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society*, *42*(3).

de Fondeville, R., & Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, *105*(3).

Dombry, C., Legrand, J., & Opitz, T. (2024). *Pareto processes for threshold exceedances in spatial extremes*.

Dombry, C., Engelke, S., & Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, *103*(2), 303–317.

Finaud-Guyot, P., Guinot, V., Marchand, P., Neppel, L., Salles, C., & Toulemonde, G. (2023). Rainfall data collected by the HSM urban observatory (OMSEV).

Haruna, A., Blanchet, J., & Favre, A.-C. (2023). Modeling intensity-duration-frequency curves for the whole range of non-zero precipitation: A comparison of models. *Water Resources Research*, *59*(6).

Leber, D. (2015). *Comparison of simulation methods of brown-resnick processes*.

Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., & Laurantin, O. (2012). A 10-year (1997—2006) reanalysis of quantitative precipitation estimation over france: Methodology and first results. *IAHS-AISH publication*, (351).