



HAL
open science

Modélisation statistique de précipitations urbaines à fine échelle spatio-temporelle

Chloé Serre-Combe, Nicolas Meyer, Thomas Opitz, Gwladys Toulemonde

► To cite this version:

Chloé Serre-Combe, Nicolas Meyer, Thomas Opitz, Gwladys Toulemonde. Modélisation statistique de précipitations urbaines à fine échelle spatio-temporelle. JDS 2023 - 54e Journées de Statistique de la SFds, SFdS, Jul 2023, Bruxelles, Belgium. hal-04374113

HAL Id: hal-04374113

<https://hal.science/hal-04374113>

Submitted on 8 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MODÉLISATION STATISTIQUE DE PRÉCIPITATIONS URBAINES À FINE ÉCHELLE SPATIO-TEMPORELLE

Chloé Serre-Combe¹ & Nicolas Meyer² & Thomas Opitz³ & Gwladys Toulemonde⁴

¹ *Univ. Montpellier, CNRS, IMAG, Inria, France, chloe.serre-combe@umontpellier.fr*

² *Univ. Montpellier, CNRS, IMAG, Inria, France, nicolas.meyer@umontpellier.fr*

³ *INRAE, BioSP, Avignon, France, thomas.opitz@inrae.fr*

⁴ *Univ. Montpellier, CNRS, IMAG, Inria, France, gwladys.toulemonde@umontpellier.fr*

Résumé. Pour l'analyse du risque d'inondation, la modélisation des précipitations s'avère être d'un grand intérêt. On propose une modélisation de la distribution des pluies mesurées, à une haute résolution spatiale et temporelle, par l'observatoire urbain de Montpellier depuis 2019. À notre connaissance, il n'y a pas eu d'études de ce type en France concernant les épisodes méditerranéens à une échelle spatio-temporelle aussi fine, en milieu urbain. Pour cette modélisation, nous considérons les pluies modérées et intenses. Pour contourner le choix d'un seuil souvent délicat en statistique des extrêmes, on utilise une loi univariée de type *Extended Generalized Pareto* pour notre modélisation. Cette famille de distribution permet aussi de limiter le nombre de paramètres à estimer. Enfin, une analyse de la dépendance entre les différents pluviomètres du réseau de mesures montre, via des indices d'autocorrélation extrême, sa variabilité entre les sites en lien avec leur localisation et avec la temporalité des mesures.

Mots-clés. Statistiques des extrêmes, modélisation de précipitations, EGPD, haute résolution spatio-temporelle, dépendance extrême

Abstract. For flood risk analysis, precipitation modeling is of great interest. We propose modeling the distribution of rainfall measured at a high spatial and temporal resolution by the urban observatory of Montpellier since 2019. To our knowledge, there has been no such study in France concerning Mediterranean episodes at such a fine spatio-temporal scale and in an urban environment. For our modeling approach, we consider moderate and intense rainfall. To avoid the explicit threshold selection that is often delicate in statistics of extremes, we use the *Extended Generalized Pareto* distribution for our modeling. This family of distributions also allows us to limit the number of parameters to be estimated. Finally, an analysis of the extremal dependence between the different rain gauges of the measurement network via indices of extremal autocorrelation shows its variability between the sites in relation to their spatial distances and to the temporality of the measurements.

Keywords. Statistics of extremes, rainfall modeling, EGPD, high spatial and temporal resolution, extremal dependence

1 Introduction

La modélisation des précipitations est un enjeu statistique majeur dans le cadre de la prévention des risques naturels, notamment en milieu urbain où des inondations peuvent entraîner des dégâts considérables. Ces dernières peuvent être des conséquences d'événements pluvieux intenses, ou bien de périodes de pluies plus modérées mais d'une durée élevée. La forte variabilité spatio-temporelle de données de pluies, même à très fine échelle, rend cependant la modélisation de ce type de données particulièrement délicate (voir [Benoit and Mariethoz \(2017\)](#)).

L'objectif de ce travail est de modéliser à une fine échelle spatio-temporelle des événements pluvieux à Montpellier, ville connue pour ses épisodes méditerranéens. On s'appuie pour cela sur des données issues d'un réseau de 17 pluviomètres sur la zone du bassin versant du Verdanson, dans la partie nord de la ville. Plus d'informations sur ces données sont disponibles dans [Finaud-Guyot et al. \(2023\)](#).

La distance entre les différents pluviomètres est de l'ordre de l'hectomètre (voir Figure 1). Les données ont été acquises à partir de 2019, année d'installation des premiers pluviomètres, et l'enregistrement de la quantité de pluie tombée est effectuée toutes les minutes. Pour diminuer l'effet d'éventuelles erreurs de mesures, on agrège les données à des pas de temps de 5 minutes ce qui nous permet de conserver une échelle temporelle particulièrement fine. Cette haute résolution spatio-temporelle implique une très faible proportion de valeurs non nulles dans les données recueillies, de l'ordre de 1.2% des données.

Nous nous intéressons à trois sites en particulier : le premier correspond à celui situé sur le toit de Polytech (pluviomètre 3 sur la Figure 1), le second à celui positionné sur le bâtiment 35 de la Faculté des Sciences (pluviomètre 10) et le dernier est situé au CNRS (pluviomètre 6). Focalisons-nous, dans un premier temps, sur le couple de pluviomètres situés à Polytech et au CNRS. Les Figures 2a et 2b illustrent l'évolution des hauteurs de pluie sur ces deux sites. On observe une autocorrélation temporelle ainsi qu'une certaine saisonnalité avec notamment des pluies plus intenses en automne. Des périodes avec des hauteurs disparates apparaissent toutefois en 2021. La Figure 2c indique qu'une corrélation croisée plutôt faible est présente entre les deux séries, alors que les sites sont géographiquement très proches. Ceci motive donc l'analyse des précipitations à une échelle spatio-temporelle aussi fine.

Les intensités de pluies que nous étudions peuvent être séparées en trois différents régimes : les périodes sèches, les pluies modérées et les pluies à forte intensité. Pour les intensités élevées, la modélisation des précipitations s'appuie sur la théorie des valeurs extrêmes. Cette dernière propose des résultats théoriques pour les dépassements de seuil. Les deux méthodes les plus populaires de la théorie des valeurs extrêmes sont la méthode des maxima par blocs et la

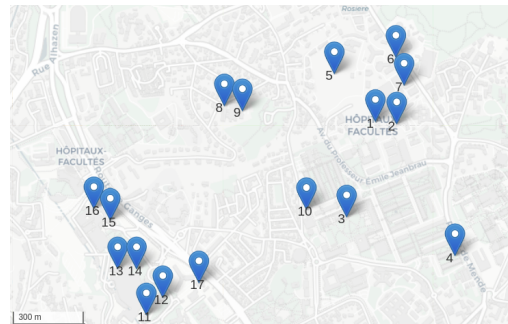


FIGURE 1 – Localisation des 17 pluviomètres (Leaflet, OpenStreetMap).

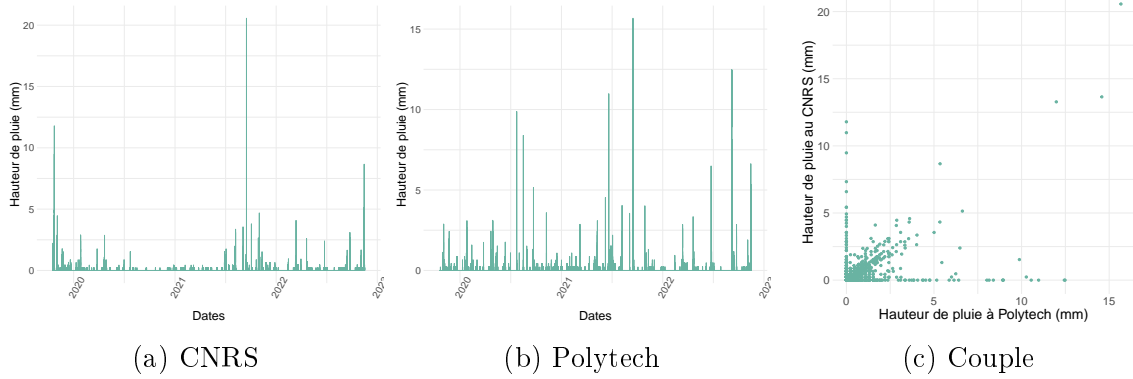


FIGURE 2 – Mesures des intensités de pluies des deux pluviomètres considérés

méthode POT pour *Peaks-Over-Threshold*. La première consiste à extraire les maxima de chaque période ou bloc partitionnant équitablement les observations (voir [Gumbel \(1958\)](#)). La seconde méthode repose sur l'approximation de la distribution des dépassements de seuil par une loi de Pareto Généralisée (GPD pour *Generalized Pareto Distribution*) décrite par [Pickands III \(1975\)](#). Dans notre cas, il y a relativement peu de données car la temporalité est assez courte. C'est pourquoi nous privilégions la seconde méthode, plus flexible dans ce contexte. En effet, l'approche POT permet de choisir un seuil relativement bas pour retenir un nombre suffisamment élevé de dépassements pour l'inférence statistique des paramètres. Enfin, il existe des extensions intéressantes mobilisant tout le spectre des données que nous utiliserons ici.

Le travail présenté dans ce papier s'articule autour de deux axes. Tout d'abord, on cherche à proposer un modèle univarié qui tient compte de la totalité du spectre de pluie. Ensuite, on étudie l'intérêt d'un modèle de dépendance spatiale et temporelle.

2 Modèle univarié pour les précipitations

Dans un cadre univarié, notons Y la variable aléatoire donnant les hauteurs de pluie en millimètre mesurées en un site donné. L'approche de dépassement de seuil de la théorie des valeurs extrêmes consiste à modéliser la probabilité d'excès à l'aide d'une GPD de paramètre de forme ξ et de paramètre d'échelle σ_u dépendant du choix de seuil u . Lorsque $\xi > 0$, nous avons des queues de distributions lourdes, typiques de la distribution des précipitations. La fonction de répartition de cette loi est notée H_ξ et nous avons l'approximation

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

où $a_+ = \max(a, 0)$.

Cependant, une telle modélisation ne permet pas de prendre en compte le reste de la distribution des pluies et nécessite un choix de seuil qui peut être difficile à déterminer et pour lequel les résultats sont très sensibles (voir [Scarrott and MacDonald \(2012\)](#)). Il est alors intéressant de chercher une modélisation pour la totalité du spectre de pluies. Plusieurs modèles ont été étudiés à cette fin, par exemple avec une loi gamma ([Katz \(1977\)](#)) ou bien avec des modèles de mélanges comme celui de la Pareto hybride qui relie une loi gaussienne et une GPD ([Carreau and Bengio \(2009\)](#)). Cependant, ces modèles sont limités : la loi gamma ne donne pas une queue assez lourde pour les valeurs extrêmes d'intensité de pluies et le modèle de Pareto hybride peut entraîner des valeurs négatives incompatibles avec la modélisation de données de pluies. Enfin, ces deux modèles nécessitent un choix de seuil qui est critique lors de la modélisation.

L'article de [Naveau et al. \(2016\)](#) introduit le modèle de la GPD étendue ou l'*Extended Generalized Pareto Distribution* (EGPD) qui tient compte de la distribution complète des précipitations sans passer par l'étape de sélection d'un seuil. L'idée de ce modèle est de chercher à contrôler les deux queues de distribution, droite et gauche, selon des principes de la théorie des valeurs extrêmes. Pour les fortes valeurs nous avons un comportement suivant une GPD de paramètres (ξ, σ) . Le comportement pour les faibles valeurs est basé sur la GPD en considérant la variable $-Y$ avec 0 en borne supérieure. Ce choix pour la queue basse nous ramène à un comportement de fonction puissance. Une distribution adéquate pour les précipitations doit alors satisfaire

$$F_Y(y) \approx \begin{cases} 1 - c \times \overline{H}_\xi\left(\frac{y}{\sigma}\right) & \text{pour } y \text{ assez grand,} \\ c \times y^\kappa & \text{pour } y \text{ proche de 0.} \end{cases}$$

L'idée est alors de modifier une loi de Pareto avec une fonction continue G de sorte que la nouvelle loi obtenue modélise correctement les pluies d'intensité faible mais aussi élevée. On définit ainsi l'EGPD via la fonction de répartition

$$F_Y(y) = G\left(H_\xi\left(\frac{y}{\sigma}\right)\right), \quad y \in \mathbb{R}. \quad (1)$$

La fonction de répartition (1) est basée sur différentes propositions paramétriques de G . Quatre différentes formes ont été proposées correspondant à quatre différents modèles d'EGPD de paramètres $(\xi, \sigma, \kappa, \delta)$. Nous nous concentrons sur les deux formes suivantes :

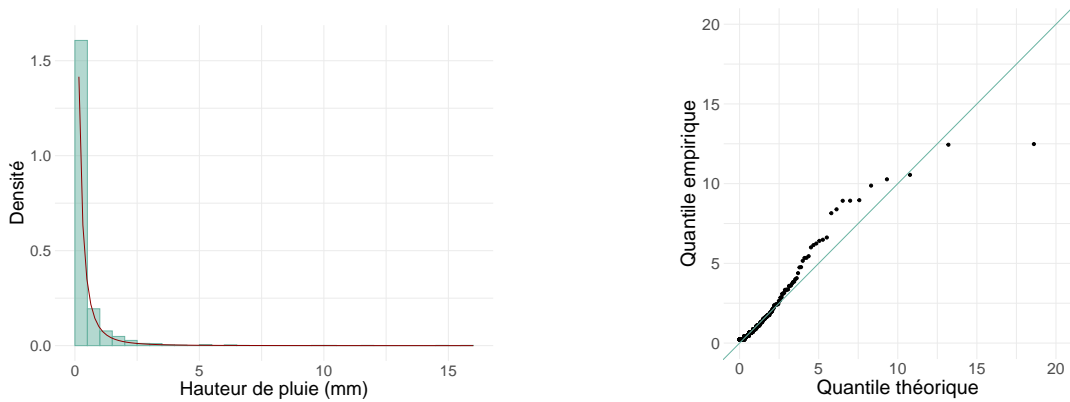
$$G_1(x) = x^\kappa, \quad (2)$$

$$G_2(x) = \left[1 - Q_\delta\left((1-x)^\delta\right)\right]^{\kappa/2}, \quad (3)$$

où Q_δ désigne la fonction de répartition d'une Beta de paramètres $(1/\delta, 2)$. Le premier modèle est un cas particulier du second avec $\delta = 1$.

Pour les deux modèles d'EGPD, ξ contrôle la forme de la queue haute de la distribution et σ est le paramètre d'échelle. Le paramètre κ , quant à lui, permet de contrôler la forme de la distribution pour les faibles valeurs. Concernant le second modèle, l'ajout du paramètre δ donne plus de flexibilité sur le comportement des valeurs centrales de la distribution.

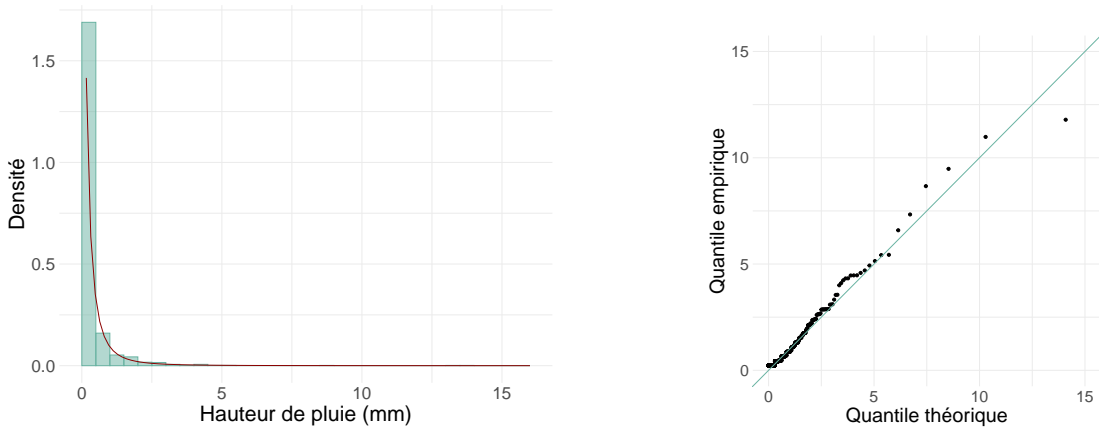
En ajustant ces modèles sur les données de pluies de Montpellier, des critères visuels, tels que les QQ -plots, montrent que la préférence va au premier modèle d'EGPD. On obtient une estimation des paramètres pour ce modèle qui est comparable d'un site à l'autre ayant le même nombre d'années de mesures. L'ajout du paramètre δ pour avoir le second modèle n'améliore pas les résultats. Nous avons pour la station de Polytech l'estimation des paramètres du premier modèle d'EGPD suivante : $\hat{\kappa} = 0.56$, $\hat{\sigma} = 0.26$ et $\hat{\xi} = 0.51$. En regardant sur six autres stations comparables, on constate des estimations sensiblement similaires. Nous ajusterons alors une unique EGPD avec les paramètres estimés pour Polytech et nous obtenons les graphiques des Figures 3 et 4. Ces derniers nous montrent la bonne qualité d'ajustement de ce modèle pour les sites de Polytech et du CNRS.



(a) Histogramme et densité de l'EGPD en rouge

(b) QQ -plot

FIGURE 3 – Qualité d'ajustement au premier modèle d'EGPD avec les paramètres estimés $\hat{\kappa} = 0.56$, $\hat{\sigma} = 0.26$ et $\hat{\xi} = 0.51$ sur les données mesurées à Polytech



(a) Histogramme et densité de l'EGPD en rouge

(b) QQ -plot

FIGURE 4 – Qualité d'ajustement au premier modèle d'EGPD avec les paramètres estimés $\hat{\kappa} = 0.56$, $\hat{\sigma} = 0.26$ et $\hat{\xi} = 0.51$ sur les données mesurées au CNRS

3 Dépendances spatiale et temporelle

Lorsque l'on se place dans un cadre bivarié, il est intéressant d'analyser la dépendance extrême entre deux variables aléatoires correspondant à un couple de pluviomètres. En statistique des extrêmes, une mesure de dépendance utilisée est le paramètre $\chi \in [0, 1]$, qui correspond à la limite

$$\chi = \lim_{u \rightarrow 1} \chi(u), \quad \text{avec} \quad \chi(u) = \mathbb{P}(U > u \mid V > u),$$

où les variables aléatoires U et V suivent des lois uniformes sur $[0, 1]$ (Coles et al. (1999)). Ce paramètre donne une mesure de la dépendance extrême entre les variables U et V puisqu'il correspond à la probabilité asymptotique que U prenne de grandes valeurs sachant que V excède un certain seuil u . Le cas $\chi = 0$ correspond à l'indépendance asymptotique, tandis que $\chi = 1$ signifie qu'il y a une dépendance parfaite. La valeur de χ nous donne ainsi un indice de la force de la dépendance extrême entre les deux variables étudiées étant marginalement transformées.

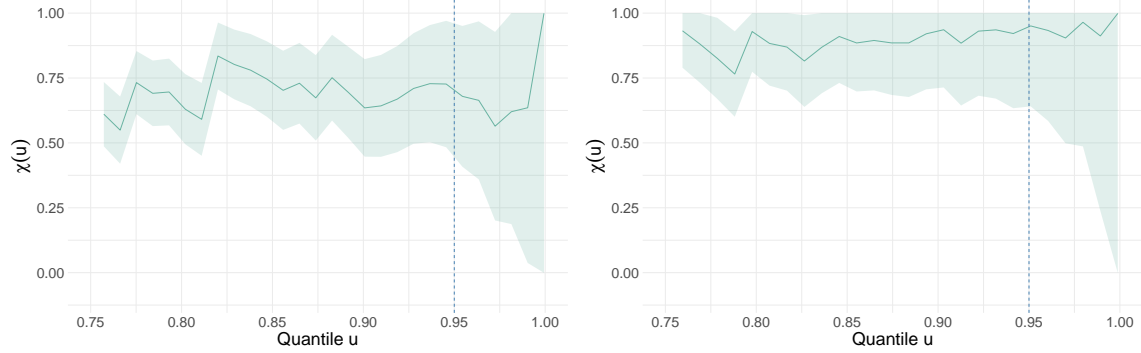
Sur des données réelles, la valeur de χ s'estime donc après avoir transformé empiriquement les marginales en loi uniforme et en prenant un seuil u suffisamment grand (typiquement un quantile d'ordre 95%).

La Figure 5 représente l'évolution de $\chi(u)$ en fonction de différents seuils u à partir des probabilités empiriques pour deux couples de pluviomètres. Le premier couple considéré correspond à des sites éloignés d'environ 1 kilomètre : Polytech et CNRS. Le second couple contient des pluviomètres avec une plus forte proximité (moins de 300 mètres de distance) : Polytech et le bâtiment 35 de la Faculté des Sciences. Nous observons une dépendance quasi parfaite lorsque les sites sont plus proches alors que la dépendance est moins forte pour les pluviomètres plus distants. Ainsi, nous constatons des comportements différents en fonction de la position des couples bien que nous soyons à très fine échelle.

Finalement, nous avons un résumé de cette information sur la dépendance pour l'ensemble des couples de pluviomètres avec un quantile fixé à $u = 0.95$, qui est un choix raisonnable pour l'estimation de la valeur asymptotique à la vue des graphiques de la Figure 5.

Nous avons le même constat sur l'ensemble des couples de sites en estimant la valeur du χ asymptotique pour un quantile $u = 0.95$ qui semble raisonnable à la vue des graphiques de la Figure 5. Nous retrouvons les χ obtenus et la matrice des distances sur la Figure 6. La matrice des valeurs de χ (Figure 6b) montre que nous n'avons pas une dépendance parfaite entre les différentes stations bien que nous soyons à une fine échelle spatiale. Les sites les plus proches ont tendance à avoir une dépendance plus importante, mais les différences entre les valeurs estimées de χ sont moins marquées et plus bruitées que celles entre les distances (Figure 6a).

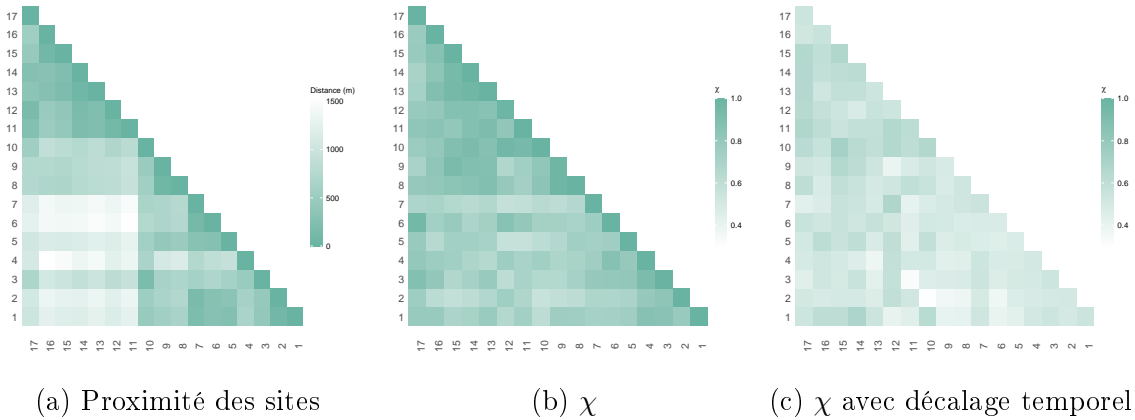
Enfin, la considération d'un décalage temporel de 5 minutes entre chaque site permet de prendre en compte le déplacement de la perturbation dans le temps. La Figure 6c donne la nouvelle estimation du χ avec ce décalage et, naturellement, on observe une diminution de la dépendance quand on ne regarde plus nos observations au même instant.



(a) Sites distants d'environ 1 kilomètre
(Polytech et CNRS)

(b) Sites distants de moins de 300 mètres
(Polytech et bâtiment 35)

FIGURE 5 – Courbe du $\chi(u)$ avec son intervalle de confiance en fonction du quantile u pour les deux couples de stations avec en pointillés le quantile retenu pour l'estimation de la valeur asymptotique



(a) Proximité des sites

(b) χ

(c) χ avec décalage temporel

FIGURE 6 – Matrice des distances et valeurs de la mesure de dépendance χ pour tous les couples de pluviomètres avec et sans décalage temporel de 5 minutes

4 Conclusion et perspectives

Une première analyse montre l'importance de considérer la spatialité et la temporalité des pluies lorsque nous sommes à fine échelle. Nous avons un premier modèle en univarié d'EGPD qui ne prend pas encore en considération des phénomènes spatiaux ou temporels. Pour tenir compte de ces phénomènes, une possibilité envisagée est d'intégrer aux paramètres du modèle des covariables temporelles cycliques, telles que le quantième du mois, mais aussi des covariables correspondant aux coordonnées spatiales des pluviomètres. Nous envisageons cette intégration avec des modèles GAMLSS (pour *Generalized Additive Model for Location Scale and Shape*) introduits par [Carrer and Gaetan \(2022\)](#). Ce type de modélisation repose sur une classe de modèles GAM (*Generalized Additive Model*) permettant d'intégrer des covariables de façon non-linéaire dans plusieurs des paramètres du modèle EGPD.

Nous présentons ici une approche exploratoire multi-site pour modéliser la dépendance. Une prochaine étape sera de prendre en compte la dépendance spatiale et temporelle avec une modélisation spatio-temporelle utilisant, par exemple, des champs gaussiens.

Les premiers résultats de ce travail s'inscrivent dans un projet plus ambitieux visant à réaliser, avec une haute résolution spatio-temporelle, un générateur stochastique de précipitations. Ce dernier devra ensuite alimenter un modèle d'écoulement et permettre d'étudier le risque d'inondation.

Références

- Benoit, L. and Mariethoz, G. (2017). Generating synthetic rainfall with geostatistical simulations. *Wiley Interdisciplinary Reviews : Water*, 4(2) :e1199.
- Carreau, J. and Bengio, Y. (2009). A hybrid pareto model for asymmetric fat-tailed data : the univariate case. *Extremes*, 12(1).
- Carrer, N. L. and Gaetan, C. (2022). Distributional regression models for extended generalized pareto distributions. *arXiv preprint arXiv :2209.04660*.
- Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2 :339–365.
- Finaud-Guyot, P., Guinot, V., Marchand, P., Neppel, L., Salles, C., and Toulemonde, G. (2023). Rainfall data collected by the HSM urban observatory (OMSEV), will be published at <https://doi.org/10.23708/67lc36>; available by sending an email at pascal.finaud-guyot@umontpellier.fr.
- Gumbel, E. J. (1958). *Statistics of extremes*. Columbia Univ. Press.
- Katz, R. W. (1977). Precipitation as a chain-dependent process. *Journal of Applied Meteorology (1962-1982)*, pages 671–676.
- Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131.
- Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical journal*, 10(1) :33–60.