Analyse statistique de la pluviométrie et de son impact sur le dimensionnement des réseaux d'assainissement

Marie Boutigny

DOCTORAT BRETAGNE LOIRE MATHSTIC

Université de Bretagne Occidentale

Eau du Ponant
Société Publique Locale

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Mathématiques et leurs Interactions*

Par
Marie BOUTIGNY

Analyse statistique de la pluviométrie et de son impact sur le dimensionnement des réseaux d'assainissement

Application à Brest Métropole dans le cadre du projet MEDISA

Thèse présentée et soutenue à Brest, le 30 août 2021
Unité de recherche : Laboratoire de Mathématiques de Bretagne Atlantique

Rapporteurs avant soutenance :

Grégoire MARIETHOZ	Professeur à l'Université de Lausanne, Institut des dynamiques de la surface terrestre
Marie-George TOURNOUD	Professeur à HydroSciences Montpellier, UMR 5569

Composition du Jury :

Président :	Valérie MONBET	Professeure, Université de Rennes 1, IRMAR
Examinateurs :	Grégoire MARIETHOZ	Professeur, l'Université de Lausanne, Institut dynamiques de la surface terrestre
	Marie-George TOURNOUD	Professeur à HydroSciences Montpellier, UMR 5569
	Philippe NAVEAU	Directeur de Recherche, LSCE, CNRS Gif sur Yvette
	Anne CUZOL	Docteure, Université Bretagne Sud, Vannes
	Aurore CHAUBET	Ingénieure, Eau du Ponant SPL
Dir. de thèse :	Benoît SAUSSOL	Professeur, Université de Bretagne Occidentale, UMR 6205 LMBA
Co-dir. de thèse :	Pierre AILLIOT	Maître de Conférences, Université de Bretagne Occidentale

Acknowledgement

Je tiens tout d'abord à remercier mes encadrants Pierre Ailliot, Aurore Chaubet, Benoît Saussol et Antoine Sinquin.

Ensuite mon entourage tout au long de la thèse : famille, amis, colocataires et autres proches en tout genre.

Une mention spéciale à tous ceux qui m'ont aidée dans les moments difficiles, bien sûr Guillaume mais aussi Valérie Monbet, Denis Allard, et tous ceux à qui j'ai pu parler.

A Bartok.

SOMMAIRE

INTRODUCTION

0.1 Cadre de la thèse

Eau du Ponant est une Société Publique Locale en charge entre autre de la collecte et du traitement des eaux usées notamment sur le territoire de Brest Métropole. Pour assurer le transport des eaux usées vers la station d'épuration, deux types de réseaux existent : 1) les réseaux dits séparatifs où deux réseaux coexistent, un pour drainer l'eau pluviale vers le milieu naturel (rivière, fleuve, mer, etc.) et l'autre pour diriger les eaux usées vers la station d'épuration, et 2) les réseaux unitaires où une seule conduite récolte à la fois les eaux usées et les eaux de pluie. Jusque dans les années 1970, les réseaux d'assainissement unitaires étaient mis en œuvre. Dans le cadre de leur renouvellement, ils sont rarement remplacés par des réseaux séparatifs pour différentes raisons (coût, emprise foncière, capacité de transit). La Figure 1 montre la répartition des réseaux séparatifs et unitaires sur le territoire de Brest Métropole.

Lorsqu'il pleut faiblement, les eaux de pluie récoltées par le réseau unitaire sont traitées à la station au même titre que les eaux usées alors qu'elles sont « propres ». Elles ne sont donc pas infiltrées dans le sol et ne remplissent pas les nappes. Lors d'évènements pluvieux plus intenses, le réseau unitaire peut être saturé. Des déversoirs d'orages (DO) sont alors sollicités pour le décharger vers le milieu naturel. C'est ce qui s'appelle des déversements directs d'eaux usées non traitées. C'est le principal problème lié à l'utilisation de réseaux unitaires.

Les DO sont des ouvrages dont le fonctionnement correspond à celui montré en Figure 2. Par temps sec la hauteur d'eau dans la conduite est inférieure au seuil de déversement, tout le débit passe donc par la conduite principale vers la station d'épuration. Par temps de pluie si la hauteur dans la conduite passe au-dessus du seuil une partie du débit est déversée, c'est-à-dire qu'elle part dans la conduite vers l'exutoire. Le débit déversé contient un mélange d'eau de pluie et d'eau usée et l'exutoire donne généralement sur le milieu naturel.

Les déversements en milieu naturel constituent une source de pollution non négli-

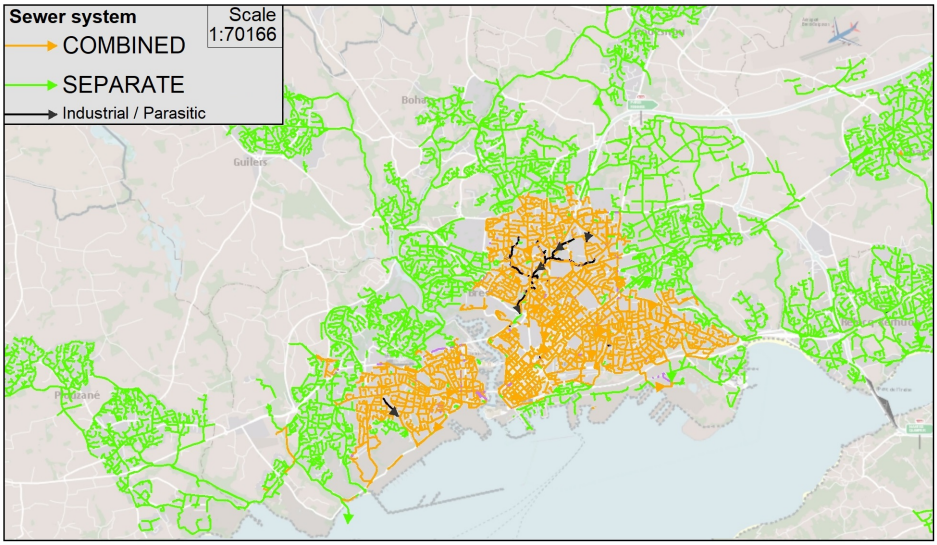


FIGURE 1 – Réseaux d'assainissement de Brest Métropole.

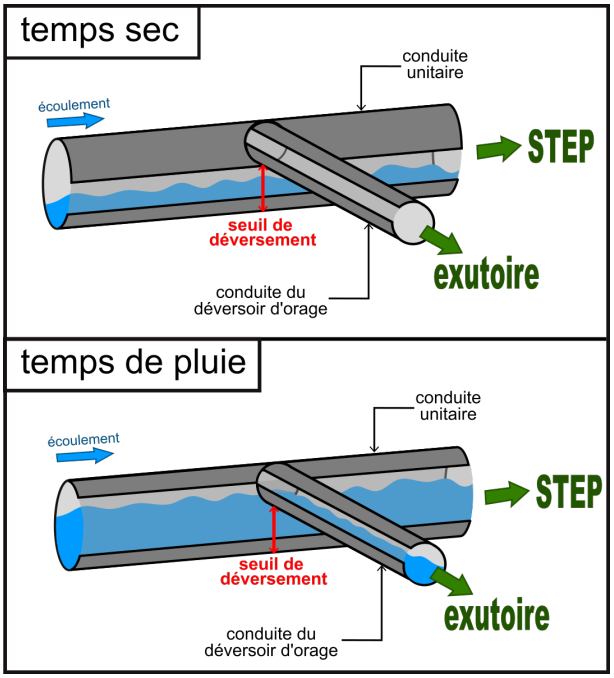


FIGURE 2 – Schéma de fonctionnement d'un déversoir d'orage. STEP : station d'épuration.

geable qui est soumise à la réglementation. L'évolution croissante des exigences de qualité au regard du milieu naturel et de ses usages (baignade, conchyliculture, impact sur les écosystèmes) et l'amélioration des connaissances sur le fonctionnement des réseaux d'assainissement (suivi, télégestion, études, modélisation, etc.) portent les collectivités à réfléchir sur l'impact des rejets de leurs systèmes d'assainissement. L'arrêté du 21 juillet 2015 relatif aux systèmes d'assainissement collectifs confirme la volonté de diminuer les déversements directs des systèmes unitaires vers le milieu naturel. Plus précisément cet arrêté donne trois possibilités pour l'établissement de la conformité du système de collecte : 1) il y a des déversements moins de 20 jours par an, 2) le volume déversé représente moins de 5% du volume collecté ou 3) le flux de pollution déversé représente moins de 5% du flux collecté. La note technique du 7 septembre 2015 relative à cet arrêté précise que la conformité sera appréciée sur la base de 5 années de mesures afin de prendre en compte la variabilité interannuelle.

Le projet MEDISA (Méthodologie de Dimensionnement des Systèmes d'Assainissement) est né dans le cadre de cette problématique. Ce projet de recherche et développement, labellisé par le pôle de compétitivité PMBA (Pôle Mer Bretagne Atlantique) et porté par Eau du Ponant, inclut des partenaires industriels et des laboratoires de recherche académiques dont le Laboratoire de Mathématiques Bretagne Atlantique (LMBA). Il a pour objectif de proposer aux collectivités une méthodologie permettant de les aider à choisir des scénarios d'aménagement réalistes nécessaires à la gestion des pollutions liées aux déversements d'un système d'assainissement par temps de pluie.

Le stockage des sur-volumes générés par les pluies puis la restitution à débit maîtrisé semble être une solution bien adaptée pour gérer les à-coups hydrauliques et réduire les déversements d'eaux usées non traitées au milieu naturel. La réalisation d'ouvrages de stockage est d'autant plus difficile et onéreuse que les réseaux unitaires se trouvent principalement dans des contextes urbains denses. Leur dimensionnement doit donc être optimisé. En parallèle, une démarche de déconnexion de surfaces actives peut être conduite afin de traiter le problème à la source (suppression de toute ou d'une partie des eaux pluviales transitant dans les réseaux unitaires).

Le projet MEDISA propose de développer une méthode, mise en œuvre via une plateforme informatique, pour aider les collectivités à dimensionner de manière réaliste des ouvrages de gestion permettant de répondre aux contraintes environnementales et aux évolutions réglementaires. Pour cela, il est nécessaire de mener une réflexion globale sur le système (pluviométrie, système d'assainissement, pollution, milieu récepteur, faisabilité

économique et spatiale) et de se doter d'outils d'analyse intégrés.

Afin de prendre en compte tous les éléments liés à la problématique des déversements en milieu naturel, le projet vise à développer :

- un modèle hydraulique qui représente les réseaux d'assainissement et qui permet de simuler les volumes déversés,
- un modèle de concentration en polluant permettant d'estimer les flux de pollution contenus dans les volumes déversés,
- un modèle de dispersion de la pollution déversée dans la rade de Brest qui permet d'étudier l'impact sur le milieu et
- un outil d'aide à la décision permettant à une collectivité de choisir parmi différents scénarios d'aménagements celui qui répond à la réglementation et qui est le plus adapté selon ses critères (coût des travaux, impact environnemental, etc.).

La méthode développée dans le cadre du projet MEDISA reposera en grande partie sur un modèle hydraulique développé par Eau du Ponant pour décrire le fonctionnement du réseau d'assainissement. Parmi les forçages de ce modèle, les conditions météorologiques et notamment les précipitations jouent un rôle prépondérant. Les résultats obtenus vont donc dépendre fortement des forçages météorologiques utilisés en entrée du modèle. L'approche usuelle consiste à dimensionner des ouvrages sur la base d'une pluie de projet de période de retour et de durée fixée (le plus souvent une pluie de type « Desbordes » (Desbordes & Raous, 1980) d'une durée de 1 ou 4 heures et de période de retour 10 ans). Cette approche est très sécurisante du point de vue du dimensionnement et par voie de conséquence très coûteuse. Pour dimensionner de façon plus juste les ouvrages de gestion, Eau du Ponant souhaite se baser sur des données collectées aux pluviomètres situés sur les bassins versants à gérer. Ainsi, se pose la question de la représentativité des données collectées au cours des dernières années pour procéder au dimensionnement de tels ouvrages. Enfin, la réglementation invite les maîtres d'ouvrages à considérer 5 années de précipitations pour évaluer la conformité du système de collecte (prise en compte de la variabilité interannuelle des pluies). Il semble nécessaire de vérifier que cette approche est justifiée dans le contexte brestois.

0.2 Systèmes de mesure

Plusieurs systèmes de mesure de la pluie existent, les plus courants étant les pluviomètres à bascule et les radars au sol, ce sont les mesures qui seront utilisées dans cette

thèse.

Les pluviomètres à auget basculant fonctionnent sur le principe suivant : à chaque fois que l'eau récoltée dans l'auget dépasse une certaine hauteur (le plus souvent 0.2 mm), il bascule et l'eau est vidée. L'appareil enregistre les bascules. Cette mesure est donc par nature discrète et peut entraîner des évènements « étranges », surtout à un pas de temps fin. La rosée peut suffire à faire basculer l'auget, faisant apparaître des mesures non nulles par temps sec, et si une pluie a une intensité inférieure à 0.2 mm par pas de temps de la mesure, on verra apparaître régulièrement des mesures à 0.2 mm, entrecoupées de zéros. Historiquement relevés à un pas de temps d'une journée, les pluviomètres sont aujourd'hui télé-relevés [1], ce qui permet d'avoir des mesures à un pas de temps de quelques minutes.

Les radars météorologiques sont des radars Döpler, qui mesurent l'atténuation des ondes qu'ils envoient dans l'atmosphère, donc la réflectivité. Ils ne mesurent pas directement la pluie mais celle-ci est déduite par une loi empirique, la loi Marshall-Palmer (Marshall & Palmer, 1948) en France. Un post-traitement assez lourd des données est nécessaire pour obtenir une estimation correcte de la pluviométrie. Les radars météorologiques ont généralement une grille spatiale de 1 km^2, et une résolution temporelle pouvant descendre jusqu'à 5 minutes.

D'autres moyens de mesure existent comme les pluvimates (e.g. Benoit et al., 2018), des pluviomètres à haute résolution qui comptent les gouttes d'eau, ce qui leur permet d'avoir un pas de temps inférieur à la minute. Les satellites peuvent également donner une estimation de la pluie, mais leur résolution spatio-temporelle est assez basse et la mesure est, comme dans le cas du radar, indirecte et nécessite une correction lourde. Enfin on peut citer l'utilisation des télécommunications : l'atténuation des ondes transmises permet une estimation de l'intensité de pluie. De nouveau ce système de mesure demande un post-traitement des données lourd et a le désavantage d'avoir une résolution spatiale qui varie en fonction de la couverture de la zone par le réseau de télécommunication.

0.3 Objectifs applicatifs de la thèse

La pluviométrie est un forçage central dans la problématique des déversements directs au milieu naturel, aussi le projet MEDISA a besoin de s'assurer que la chronique de pluie

1. Un pluviomètre télé-relevé est appelé un pluviographe, mais par abus de langage on utilisera indistinctement le terme pluviomètre, plus courant.

utilisée est la plus adaptée.

Objectif

L'objectif principal est de déterminer quelle donnée de pluie utiliser pour dimensionner le réseau d'assainissement de Brest Métropole afin de limiter les déversements en milieu naturel.

Le modèle hydraulique d'Eau du Ponant est originellement calé avec comme forçage pour la pluie un seul pluviomètre, celui situé sur le toit de l'hôtel de ville de Brest, qu'on appelera BMO (Brest Métropole Océane). Le calage consiste à comparer les débits simulés par le modèle hydraulique à la mesure en réseau sur des évènements de pluie et de temps sec. Plusieurs paramètres du modèle sont ajustables pour se rapprocher de la mesure. Le choix d'utiliser le pluviomètre de BMO est fait car il est considéré comme l'un des plus fiables du réseau d'Eau du Ponant, et car il est situé au milieu de la zone d'étude. Toutefois la question se pose de savoir si la pluie mesurée en ce point est représentative de toute la zone, s'il est donc réaliste de considérer une pluie constante sur 15 km^2. La littérature montre en effet que la distribution spatiale de la pluie peut fortement influencer à la fois le volume, le pic d'intensité et le temps de réaction du bassin versant (e.g. Arnaud et al., 2002 ; Benoit et al., 2018 ; Krajewski et al., 1991). Toutefois la plupart de ces études concerne soit les problématiques d'inondation soit des échelles spatiales très fines. Aussi même s'il y a une variabilité spatiale de la pluie, celle-ci a-t-elle une influence significative sur la sortie du modèle qui nous intéresse, c'est-à-dire les débits déversés ? Cette première question en appelle une plus générale : à quelles caractéristiques de la précipitation le modèle hydraulique est-il sensible ?

Différentes sources de données de précipitation sont disponibles : 1) Eau du Ponant et Brest Métropole entretiennent un réseau de 11 pluviomètres à auget répartis dans la zone d'intérêt. La plupart de ces pluviomètres sont en milieu urbain, souvent installés dans des cimetières ou sur les toits d'immeubles pour le centre ville. Ces pluviomètres ont été mis en place en 2010 et ont un pas de temps d'enregistrement de 3 minutes. 2) La station météorologique de Météo France située à Guipavas sur le site de l'aéroport contient un pluviomètre à auget, en place depuis 1945. A l'époque la mesure était journalière, aujourd'hui le pas de temps d'enregistrement est de 6 minutes. 3) Météo France dispose d'un réseau de radar météorologiques qui couvrent toute la France. Le plus proche est à

Plabennec, soit à 5 km de Brest. La donnée est disponible à un pas de temps de 5 minutes sur une grille de 1 km^2 depuis 2006.

Un premier travail est l'analyse de ces différentes données de pluie. En effet on a à la fois des instruments de mesure hétérogènes (pluviomètres vs. radar), une qualité de mesure hétérogène (des pluviomètres en ville ou non), une maintenance hétérogène (surveillance au jour le jour à Météo France, ce qui n'est pas le cas sur le réseau d'Eau du Ponant), un traitement hétérogène de la donnée (rien pour les pluviomètres, un algorithme très lourd pour le radar), et enfin un échantillonage spatiotemporel hétérogène (3, 5 ou 6 minutes pour le pas de temps, et 1 km^2, 1 ou 11 points pour la couverture spatiale). Il est donc nécessaire de comprendre ce que chaque source de donnée contient, en particulier quelles erreurs on peut y trouver et quelles propriétés de la précipitation sont correctement reproduites par la mesure.

Les différentes sources de données permettent de faire varier les propriétés du forçage du modèle et donc d'étudier la sensibilité de la sortie (les débits déversés) à la pluie.

Le deuxième point d'intérêt du projet MEDISA est la recommandation de l'arrêté du 21 juillet 2015, qui précise que la réglementation doit être respectée sur 5 années de manière à prendre en compte la variabilité interannuelle de la pluie. Nous nous posons donc la question du choix et de la représentativité de cette chronique. La notion de représentativité n'est pas plus discutée dans l'arrêté, mais on comprend que le choix des données doit être justifié. Dans la bibliographie sur les variables environnementales la notion de représentativité du climat et d'année typique est une question courante. Une première question est le choix entre données réelles et données simulées. La simulation de données grâce à un générateur aléatoire permet d'obtenir une base de données bien plus fournie que ce qu'on peut avoir avec de la mesure et de faire varier les caractéristiques de la chronique d'entrée de façon contrôlée. Un exemple de générateur stochastique pour éviter l'utilisation d'années typiques est montré dans Ailliot et al. (2020). L'utilisation d'un générateur aléatoire de conditions météorologiques en couplage avec une modèle hydraulique a été fait par exemple dans Caron et al. (2008) ou encore Breinl (2016). Toutefois ces références utilisent des modèles hydrauliques simplifiés qui permettent de faire tourner un grand nombre de séries.

Comme ça a été évoqué précédemment dans le cadre du projet MEDISA trois modèles sont impliqués : le modèle hydraulique, un modèle de flux et un modèle milieu. Le premier et le dernier ont un coût numérique très élevé, ce qui contraint fortement la longueur

et le nombre de séries que l'on peut tester comme forçage. Aussi pour l'application on considère seulement une année « typique » qu'on choisit parmi les années existantes car ça permet d'avoir accès aux autres forçages (vent, vagues, débit des rivières, etc.) et aux mesure in-situ sur le réseau d'assainissement. Cette année servira à tester les scénarios d'aménagement, mais à terme il faudra vérifier le choix du scénario sur 5 années, comme le demande la réglementation. Les données fournies devront être les plus réalistes possibles (minimiser les erreurs) et devront respecter les caractéristiques identifiées comme impactantes pour le modèle hydraulique.

Au vu de l'hétérogénéité des mesures disponibles, fournir les chroniques les plus réalistes possible demande de fusionner les données. La fusion de données est un champ largement couvert dans le cadre de la correction des biais du radar par les pluviomètres (voir e.g. Cecinati et al., 2017 ; Seo, 1998). Un autre piste, qui permet de prendre en compte les erreurs de plusieurs sources de mesure, est l'assimilation de données (Jones & Macpherson, 1997 ; Lien et al., 2013), qui consiste à reconstruire la « vraie » précipitation (non observée), à partir des différentes observations. Chaque source d'observation est définie avec une erreur associée, qui peut par exemple contenir un biais systématique nul ou non, ou avoir une covariance non nulle ou aucune dépendance spatiale, etc.

Que ce soit pour la génération de données pour l'étude de sensibilité ou l'assimilation de données pour tester les scénarios d'aménagements avec des chroniques de pluie les plus réalistes possible, il est nécessaire de passer par la modélisation statistique de la pluie (marginale puis multivariée).

0.4 Modélisation statistique

La question de la modélisation marginale de la pluie est une première étape vers la création d'un générateur aléatoire météorologique ou d'un modèle d'assimilation de données. La précipitation représente une variable clé dans de nombreux domaines d'étude, avec en premier lieu l'hydrologie et la météorologie, mais aussi l'agronomie ou encore les études d'impact (e.g. Bauer et al., 2015 ; Caseri et al., 2016).

Historiquement la précipitation est mesurée quotidiennement, aussi la littérature est abondante pour la distribution de la pluie accumulée à l'échelle journalière ou mensuelle. La distribution la plus populaire pour les intensités de pluie est la loi Gamma (Katz, 1999) qui donne généralement de bons résultats d'ajustement à ces échelles de temps. D'autres

distributions ont été utilisées et comparées, leurs performances dépendant fortement de la localisation de la mesure, ce qui n'est pas surprenant car la distribution de la pluie est fortement impactée par le climat local.

Pour ce qui est des précipitations accumulées sur des périodes plus courtes que la journée, la plupart de la littérature est développée pour des données horaires. De façon générale à l'échelle horaire l'occurrence de la précipitation (temps sec / temps de pluie) est modélisée séparément de l'intensité de pluie. En effet quand on descend en dessous de l'échelle journalière, le pic de la distribution en zéro dû aux mesures de temps sec devient trop important pour ajuster une loi complètement continue.

Les données de pluie à échelle très fine (quelques minutes) sont relativement récentes et la recherche est encore active pour la modélisation de leur distribution. Comme le précise Schilling (1991) ou encore Berne et al. (2004), les modèles hydrologiques ont besoin de données à l'échelle (temporelle mais aussi spatiale) la plus fine possible. En particulier, tester la sensibilité et la robustesse des modèle hydrauliques urbains nécessite de décrire et de simuler la précipitation à ces échelles de quelques minutes. Pour citer l'étude de Berne et al. (2004) menée à Marseille, donc dans un climat méditerranéen :

> « According to the results, hydrological applications for urban catchments of the order of 1000 ha require a temporal resolution of about 5 min and a spatial resolution of about 3 km. For urban catchments of the order of 100 ha, it becomes a resolution of about 3 min and 2 km, that common operational networks or radars cannot provide. »

Sur la zone de Brest Métropole il y a environ 1300 ha de bassin versant unitaire, répartis en deux bassins de collecte globaux (950 et 350 ha). Ces bassins sont sous divisés dans le modèle hydraulique, qui est calé sur des points qui ont en amont entre 10 et 500 ha de bassin de collecte. Pour les points de calage on est donc plutôt dans le cas de la deuxième recommandation : 3 minutes et 2 km.

Le pas de temps utilisé de base pour le modèle hydraulique est celui a priori idéal de trois minutes car il correspond au pas de temps de la mesure en réseau (donc de la mesure des débits déversés). Par contre pour la grille spatiale une seule chronique est utilisée pour toute la zone, ce qui est très éloigné de la recommandation.

A l'échelle infra-horaire, la plupart des mesures sont nulles et on observe une forte discrétisation due à la précision de la mesure. Même si la majorité des intensités sont faibles, on trouve tout de même quelques intensités très fortes qui rendent la partie positive de la distribution fortement asymétrique à droite. La Figure 3 montre un exemple de dis-

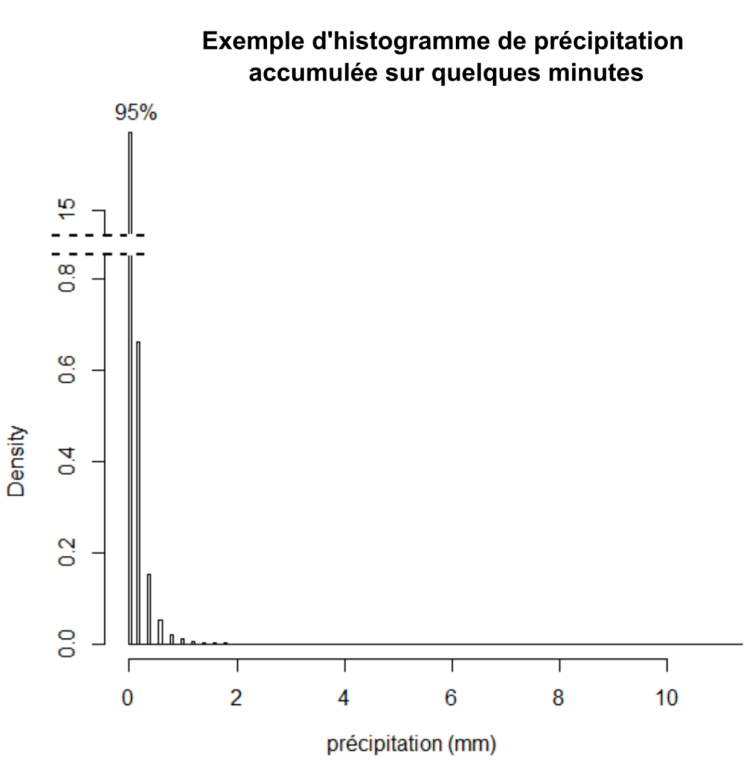


FIGURE 3 – Exemple de distribution infra-horaire de précipitation (pluviomètre de Météo France à la station Guipavas, 2006-2017, 6 minutes).

tribution de précipitation accumulée sur quelques minutes (mesurée par un pluviomètre proche de Brest). On constate que la discrétisation des données est très visible. En effet dans certaines régions, comme celle qui sera notre objet d'étude, quand on descend à un pas de temps de quelques minutes, près de 97% des mesures positives sont inférieures à 1 mm. Une précision de 0.2 mm a donc un impact très fort sur la distribution.

Caractéristiques de la distribution marginale de la pluie

En résumé les précipitations présentent plusieurs caractéristiques qui les rendent particulièrement difficiles à modéliser. Tout d'abord la pluie a par nature un caractère intermittent, dû au fait qu'il ne pleut pas toujours. C'est une variable largement non Gaussienne quand on est en dessous de la mesure mensuelle. La distribution est fortement asymétrique à droite, ce qui signifie que la plupart des intensités sont faibles. Les intensités fortes sont plus rares mais des évènements extrêmes peuvent survenir. On pense en effet spontanément aux inondations, des phénomènes extrêmes qui ont des enjeux très forts et qui font l'objet d'études de risque et d'études d'impact. Finalement le moyen de mesure donne lui-même une dernière caractéristique à la pluie qui complexifie sa modélisation : la discrétisation. Avec une précision de généralement 0.2 mm les données sont très fortement discrétisées, et ce caractère doit être pris en compte dans la modélisation.

Une approche classique en modélisation statistique pour gérer la non-Gaussianité est d'utiliser une transformation pour normaliser les données. En effet ceci permet de réutiliser tout ce qui a été développé pour le cas Gaussien. Dans le cadre des variables météorologiques, il y a un intérêt particulier dans les modèles existants pour les processus Gaussiens spatio-temporels, basés sur une modélisation des structures d'ordre 1 et 2 du processus, ainsi que les algorithmes de simulation associés (e.g. Hussain et al., 2010).

Pour normaliser la précipitation, prendre la racine carrée a été proposé par Panofsky et al. (1958). La transformation Box-Cox (Box & Cox, 1964) a aussi largement été utilisée (e.g. Cecinati et al., 2017). Avec ces transformations, l'occurrence de la précipitation peut être intégrée dans la transformation à l'échelle journalière, et à l'échelle horaire elle est généralement modélisée séparément avec une loi de Bernoulli. Cette gestion de l'occurrence peut être compliquée à étendre au cas multivarié.

Afin de modéliser le temps sec conjointement à l'intensité, une approche classique pour modéliser la précipitation consiste à utiliser des processus Gaussiens transformés (Allcroft & Glasbey, 2003 ; D. Wilks, 1998) (qu'on appellera aussi modèles meta-Gaussiens). Les modèles méta-Gaussiens consistent à appliquer une censure, ce qui permet de créer une

composante discrète en zéro (conditions sans pluie), suivie d'une transformation puissance ou exponentielle, ce qui crée une composante positive avec un skewness positif (Allard & Bourotte, 2015). Le modèle s'écrit alors

$$Y = 0 \times \mathbb{1}_{X<0} + \psi(X) \times \mathbb{1}_{X\geq 0}, \quad \text{avec } X \sim \mathcal{N}(\mu, 1),$$

où $\mathbb{1}_A$ est la fonction indicatrice égale à 1 si la condition A est vraie, et égale à zéro sinon. La pluie est notée Y, X est une variable aléatoire Gaussienne de moyenne μ et de variance 1, et $\psi : [0, +\infty[\rightarrow [0, +\infty[$ est une fonction croissante généralement appelée anamorphose dans la littérature.

La plupart des modèles méta-Gaussiens sont développés pour des données horaires et ne conviennent donc pas aux données avec un pas de temps de quelques minutes dont on dispose. Dans cette thèse nous développons un nouveau modèle méta-Gaussien, il aura notamment pour objectif de convenir à une grande variété de pas de temps, de quelques minutes à un mois. Les propriétés des queues (inférieure et supérieure) de la distribution des précipitations positives seront utilisées pour développer l'anamorphose

$$(x) = y_m + \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2},$$

où y_m est la valeur minimale qui peut être mesurée. Chaque paramètre de l'anamorphose proposée est lié à une partie différente de la distribution : μ décrit la probabilité de pluie (occurrence), α est lié à la forme de la distribution proche de zéro, ξ contrôle la queue de la distribution (valeurs extrêmes) et enfin σ est un paramètre d'échelle.

Les modèles méta-Gaussien se prêtent particulièrement bien à l'assimilation de données (Lien et al., 2013) car ils définissent par nature une variable latente. Bien que ni l'assimilation de données ni le générateur aléatoire n'aient pas pu être faits à cause des contraintes de temps du projet MEDISA et de la thèse, le modèle méta-Gaussien développé a été étendu au cas multivarié. L'ajustement multivarié a notamment posé la question de l'estimation de la structure d'ordre deux d'un champ Gaussien à partir du champ Gaussien censuré. La méthode de maximum de vraisemblance est comparée à des alternatives basées sur les moments.

0.5 Plan

Chapitre 1 Le Chapitre 1 vise à introduire les données qui ont été utilisées tout au long de la thèse. Les différents types de mesures (pluviomètres à bascule et radars météorologiques) sont documentés et les erreurs de mesure ainsi que les biais sont présentés, via une étude bibliographique et une comparaison des sources de données disponibles. Enfin le climat de la zone d'étude est brièvement décrit.

Chapitre 2 Il a été vite évident que les modèles classiques de distribution des précipitations n'étaient pas adaptés aux données disponibles. En effet le court pas de temps (quelques minutes) donne beaucoup d'importance à certaines caractéristiques de la pluviométrie qui ne sont pas toujours prises en compte dans les modèles classiques. En se basant sur les propriétés de la distribution des précipitations à petite échelle temporelle, un modèle est développé et étudié en détails sur les données présentées dans le Chapitre 1. Ce modèle fait partie de la classe des modèles Gaussiens censurés transformés, qui ont la particularité de contenir une variable latente Gaussienne.

Chapitre 3 Le modèle pour les marges développé dans le Chapitre 2 est étendu à la modélisation multivariée dans le Chapitre 3. Ce chapitre soulève notamment des difficultés liées à l'ajustement multivarié de Gaussiennes censurées, difficultés qui sont assez peu évoquées dans la littérature.

Chapitre 4 Dans le Chapitre 4, les besoins du modèle hydraulique sont étudiés. Il en ressort une nécessité de corriger les données disponibles, ce qui mène à une étude bibliographique montrant que les méthodes classiques semblent insuffisantes. Plusieurs chroniques de pluie seront construites en cherchant à répondre aux besoin du modèle assainissement, et les sorties obtenues seront étudiées pour analyser la sensibilité du modèle aux précipitations.

Chapitre 1

Présentation des données

Les trois premières sections de ce chapitre visent à introduire les différentes sources de données qui seront utilisées tout au long de la thèse. Le climat de la zone, les systèmes de mesure et leurs erreurs associées seront abordées, puis les sources de données seront comparées.

1.1 Le pluviomètre de Météo France

1.1.1 La station

La station météorologique de Météo France la plus proche est située à Guipavas, au Nord-Ouest de Brest (coordonnées géographiques 48.44°N, 4.41°O, cf. Figure 1.6). En service depuis 1945, elle mesure au départ la pluie avec un pas de temps journalier. Dans les années 90 cela passe à un pas de temps horaire puis en 2006 à 6 minutes. Ainsi deux séries de données ont été récupérées auprès de Météo France : 1) les données historiques journalières de 1945 à 2018 et 2) les données à 6 minutes de 2006 à 2018. Les données historiques s'étendant sur plus de 70 ans, elle serviront surtout à se faire une idée du climat brestois, tandis que les données à 6 minutes s'approchent plus du pas de temps qui nous intéresse.

Les données de la station de Météo France sont considérées comme très fiables. En effet la précision du pluviomètre est de ±5% pour les données à 6 minutes, les instruments de mesure sont surveillés quotidiennement et les données corrigées en temps réel en cas de problème (bouchage du pluviomètre, opération de maintenance, etc.). De plus la localisation de la station est choisie pour respecter des critères assez stricts. La zone doit être dégagée, elle doit pouvoir recevoir des vents de toutes directions et on notera notamment que dans l'idéal la station doit être à une distance de quatre fois (minimum deux fois) la hauteur des obstacles à proximité, donc à 32 mètres d'une maison standard. Ainsi les stations Météo France sont généralement situées en zones semi-rurales, ou dans

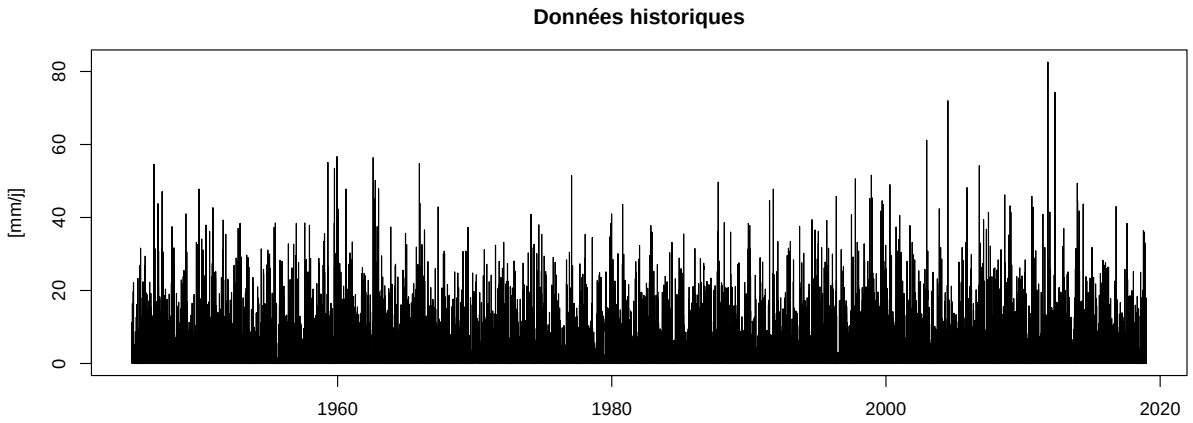


FIGURE 1.1 – Données journalières historiques de Météo France à la station Guipavas (1945-2018)

des endroits avec peu d'obstacles hauts comme à proximité des aéroports (c'est le cas à Guipavas).

Remarque : Dans la ville de Brest les immeubles font autour de 50 mètres de haut, une station en ville devrait donc se trouver à 200 mètres de tout immeuble et à 32 mètres de toute maison, ce qui apparaît vite impossible.

1.1.2 Données historiques journalières

Comme ça a été dit précédemment, les données historiques journalières permettent de donner une idée du climat de la zone d'intérêt. La série entière des données journalières est montrée en Figure 1.1.

L'impact du changement climatique sur les précipitations n'est pas aussi marqué que sur d'autres variables climatiques. L'intensification des précipitations extrêmes est probablement le phénomène le plus marqué (e.g. Gordon et al., 1992 ; O'Gorman, 2015). Les autres variations incluent principalement une baisse de la fréquence et du cumul de précipitation (Trenberth, 2011).

Le rapport de Météo France sur les données DRIAS 2020 (Soubeyroux et al., 2020) donne des projections à horizon 100 ans pour les précipitations en France métropolitaine. Une augmentation des pluies extrêmes est prévue pour toute la France, ce qui montre l'importance des distributions permettant de bien modéliser les queues lourdes. Pour ce qui est des cumuls de pluie, ce rapport montre une augmentation des cumuls en hiver (surtout marquée pour le nord de la France), et une baisse en été (surtout dans le sud).

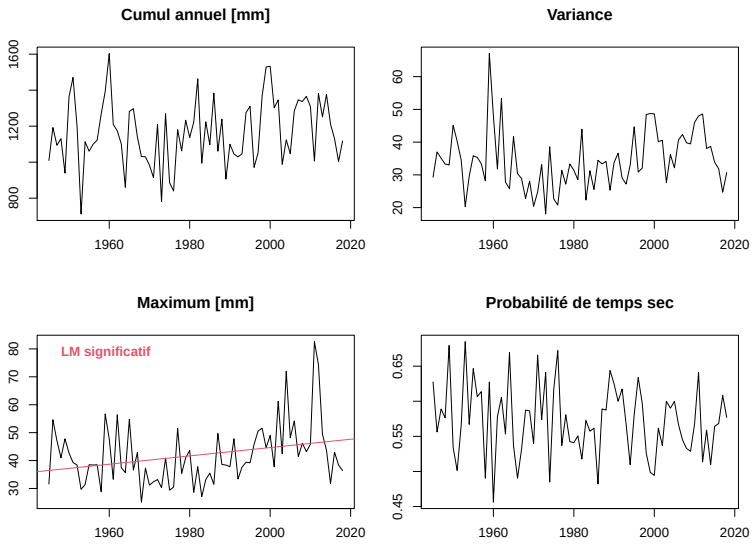


FIGURE 1.2 – Statistiques descriptives annuelles des données journalières historiques (1945-2018)

La Figure 1.2 montre l'évolution de quelques statistiques descriptives annuelles depuis 1945. Les seules évolutions significatives en regardant simplement avec un modèle linéaire :

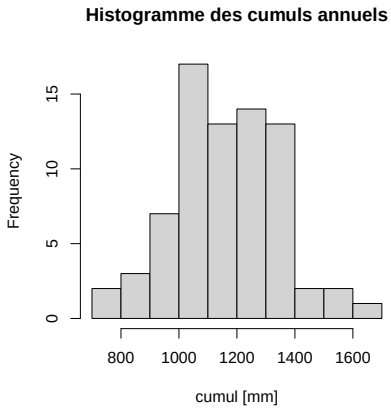
— au global : le maximum journalier annuel augmente,

— mois par mois : la probabilité et la longueur des périodes de temps sec augmente en septembre, le cumul mensuel augmente en juillet et le maximum augmente en octobre.

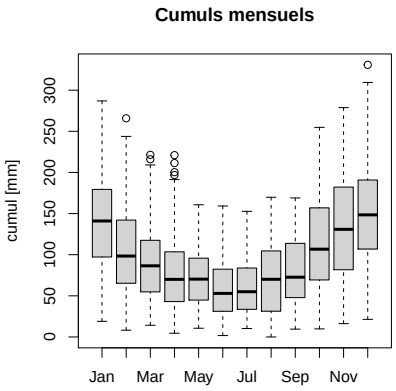
En conclusion la précipitation est une variable pour laquelle le changement climatique est pour l'instant assez peu marqué dans la région qui nous intéresse. Les évolutions visibles dans nos données concernent surtout les maximums, donc les évènements extrêmes. Cette observation est confirmée par les projections climatiques, mais les évènements extrêmes ne sont pas forcément les plus importants pour notre application : les fortes pluie entraîneront de toute façon des déversements, et les eaux usées seront dans ces cas très diluées. Le projet MEDISA cherche surtout à gérer les pluies peu intenses, qu'on pense plus problématiques du point de vue de la pollution.

Il a donc été décidé de considérer les données dont on dispose comme stationnaires.

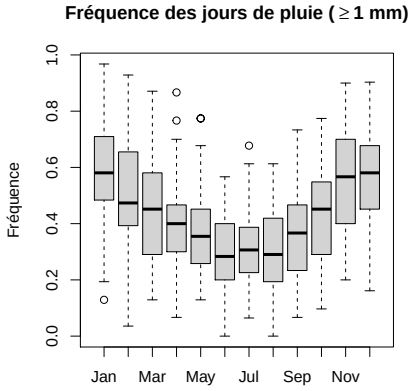
La Figure 1.3 montre les distributions des statistiques évoquées précédemment. Le climat brestois est constitué d'une moyenne de 160 jours de pluie par an (soit environ deux jours sur cinq) pour un cumul moyen de 1.2 mètres. La variabilité inter-annuelle est



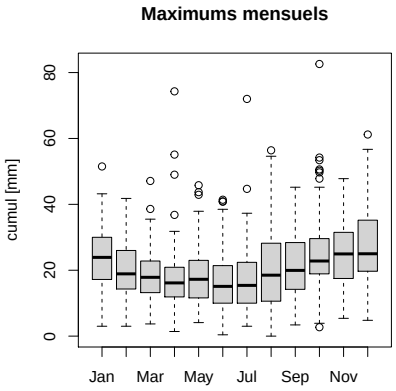
(a) Histogramme des cumuls annuels



(b) Boxplot des cumuls mensuels



(c) Boxplot des fréquence des jours de pluie



(d) Boxplot des maximums journaliers mensuels

FIGURE 1.3 – Caractéristiques des précipitations brestoises sur les données historiques (1945-2018)

assez forte avec un ratio entre l'année la plus pluvieuse et l'année la plus sèche de 2.3 (Fig. 1.3a).

Les cumuls mensuels (Fig. 1.3b) et les fréquences des jours de pluie (Fig. 1.3c) sont marqués par une forte variation saisonnière avec un facteur d'environ 2 entre l'été et l'hiver : 1/3 jour de pluie l'été contre 2/3 l'hiver, et 60 mm par mois l'été contre 140 mm l'hiver.

Les maximums journaliers mensuels (Fig. 1.3d) ont peu de variation saisonnière. Ils sont légèrement plus importants en hiver, ce qui s'explique par le fait que les maximums des cumuls journaliers ne montrent pas les intensités maximales des évènements (qui durent généralement quelques heures), une pluie soutenue qui dure toute la journée aura donc un cumul journalier plus fort qu'un orage d'une heure.

1.1.3 Données à 6 minutes

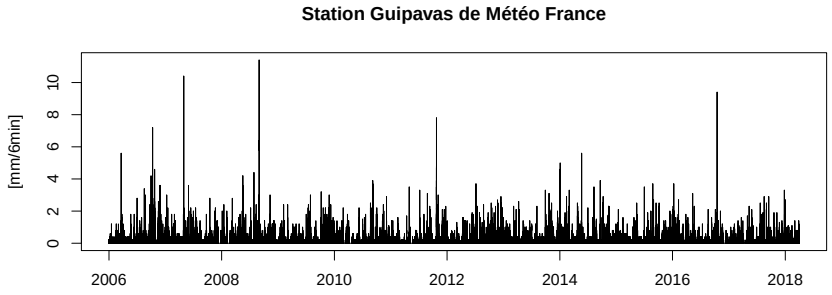
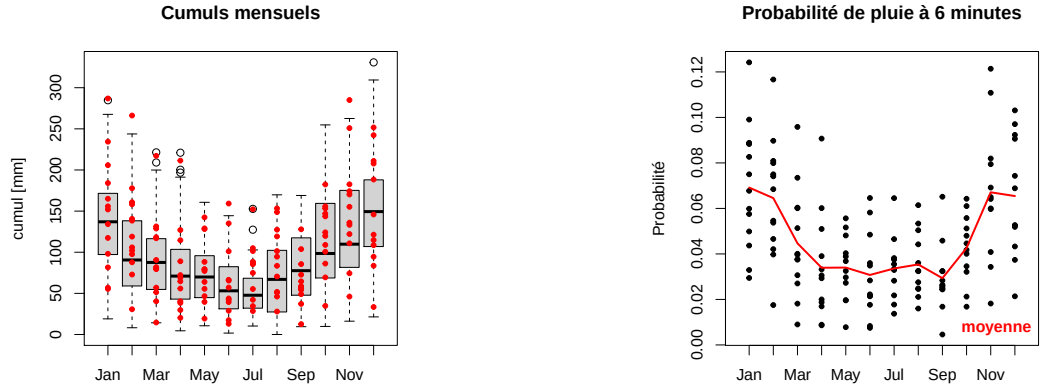


FIGURE 1.4 – Chronique de la station Guipavas Météo France à 6 minutes de 2006 à 2018

La Figure 1.4 montre l'entièreté de la série de la station Guipavas de Météo France au pas de temps 6 minutes. Comparée aux séries qu'on pourra voir dans la section suivante avec les données du réseau d'Eau du Ponant, la chronique est très propre au sens où à première vue on ne peut pas repérer d'erreur ou de donnée aberrante.

Les données à 6 minutes ne concernant que les années 2006 à 2018, il est possible de comparer rapidement ces 12 dernières années à la « climatologie », c'est-à-dire aux données historiques dont on dispose. Tout d'abord on retrouve bien environ 150 jours de pluie et 1.2 mètres de pluie annuelle. La Figure 1.5a montre où se situent les cumuls mensuels des 12 dernières années par rapport aux reste des données historiques, ce qui permet de constater que les 12 dernières années semblent plutôt représentatives du climat de la zone.



(a) Boxplots des cumuls mensuels des données historiques (1945-**2005**). Les points montrent les cumuls mensuels des données à 6 minutes 2006-2018.

(b) Probabilité de pluie à 6 minutes. Les points montrent la moyenne par mois et par année, et la ligne la moyenne par mois.

FIGURE 1.5 – Caractéristiques des données des pluviomètres à 6 minutes (2006-2018)

La Figure 1.5b montre la probabilité de pluie à 6 minutes, qui apparaît plus forte en hiver (7% contre 4% en été). A l'échelle journalière, la probabilité d'observer une mesure nulle (un jour de temps sec) est autour de 30%. A 6 minutes, il y a autour de 95% de mesures nulles, ce qui rend la modélisation statistique du temps sec très différente de ce qu'on peut faire pour des données journalières, ou même horaires.

1.2 Les pluviomètres d'Eau du Ponant

Brest Métropole et Eau du Ponant (EDP) entretiennent 12 pluviomètres sur la zone de Brest (15*20 km environ), leurs positions sont montrées en Figure 1.6.

On dispose des données de 9 pluviomètres sur la période 2010-2014. Fin 2014 un pluviomètre est arrêté (Gouesnou, au Nord en Figure 1.6) et 3 sont mis en route début 2015 (Tram, Tromeur et St Marc). Sur la période 2015-2020 il y a donc 11 pluviomètres.

Jusqu'en 2014 les données sont vérifiées et corrigées manuellement en temps réel par Eau du Ponant, on a donc sans surprise des données assez propres, mais on retrouve quand même quelques problèmes. Après 2014 il n'y a plus aucun post proccessing des données. On récupère les cumuls bruts, dont on montre un exemple en Figure 1.7. Les appareils sont des pluviomètres à bascule d'une précision de 0.2 mm, la donnée retournée est le nombre de bascules depuis la mise en place du pluviomètre. L'intensité de pluie est

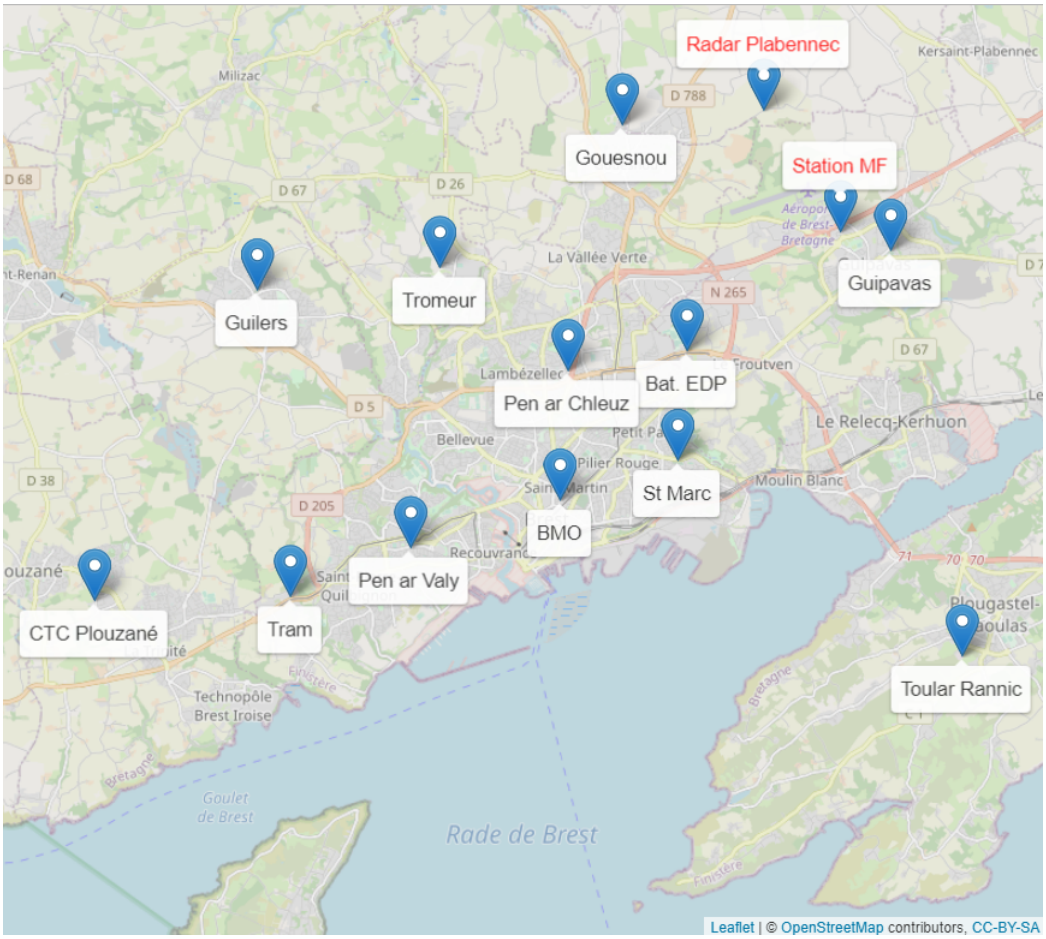


FIGURE 1.6 – Localisation des pluviomètres de Brest Métropole (noir) et de la station de Météo France (rouge).

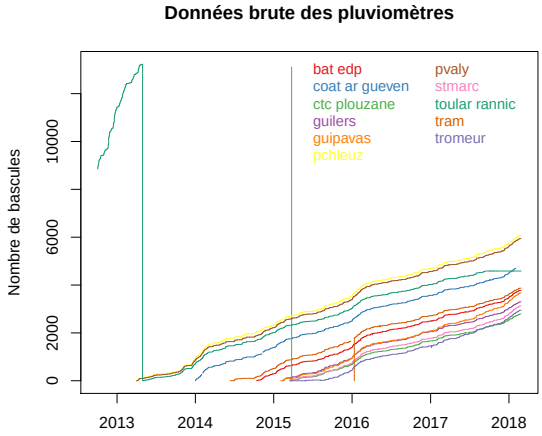


FIGURE 1.7 – Données brutes issues des pluviomètres : nombre de bascules depuis le début de la chronique

récupérée en calculant la différence entre deux pas de temps et en multipliant par 0.2 mm. Le pas de temps de la mesure est de 3 minutes.

Sur la Figure 1.7 on repère beaucoup d'anomalies évidentes : des cumuls remis à zéro (e.g. Toular Rannic), des mesures ponctuelles à zéro (e.g. Guipavas) et des longues périodes sans aucune bascule (e.g. Toular Rannic). En récupérant les intensités avec les écarts entre deux pas de temps, on voit apparaître de nouvelles anomalies. La Figure 1.8 montre l'exemple de la chronique obtenue au pluviomètre Bat EDP (Fig 1.6). On trouve des valeurs négatives, de longues périodes de temps sec (plusieurs mois) ou encore des périodes suspectes avec beaucoup de très fortes intensités. Sur d'autres pluviomètres on peut aussi trouver des valeurs extrêmes au milieu d'une période de temps sec.

Un premier traitement des données a été fait à la main. Seules les valeurs indéniablement aberrantes ont été déclarés non disponibles (NA), comme par exemple les données négatives, les très longues séquences de zéros (plusieurs mois), etc.

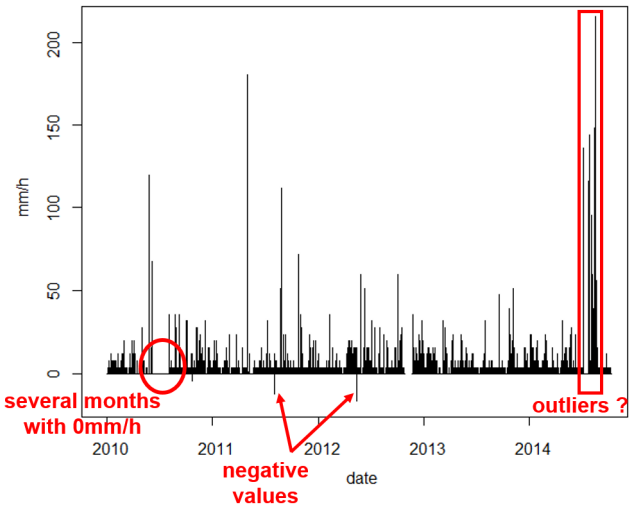


FIGURE 1.8 – Pluviomètre bat edp : données aberrantes

1.3 Les données radar

1.3.1 Fonctionnement du radar météorologique

Principe général

Les radars météorologiques sont des radars Döpler qui envoient des ondes électro-magnétiques directionnelles dans l'atmosphère et qui en mesurent la puissance rétro diffusée (en Watts). La réflectivité radar (en décibel Z) est proportionnelle au produit du carré de la distance cible-radar et de la puissance rétro diffusée. C'est cette réflectivité qui est finalement convertie en intensité de pluie (en mm/h) grâce à des lois paramétriques empiriques.

Pour plus de détails sur le fonctionnement des radars météorologiques, on pourra se référer à Doviak et al. (1979) ou encore Bringi et Chandrasekar (2001).

L'efficacité de l'interaction entre les ondes envoyées et les gouttes de pluie dépend de la longueur d'onde utilisée (λ), ce qui fait qu'il existe trois types de radar, chacun ayant ses avantages et inconvénients.

— **Bande X** ($\lambda = 2.5$ cm, portée 50 km)

Le radar de bande X est adapté aux zones de montagne, car son installation est facilité par sa petite taille (antenne de 2 m de diamètre contre 3.5 à 6.5 m pour les autres bandes) et sa faible portée n'est pas un problème étant donné qu'elle y est d'abord limitée par le relief. Son principal inconvénient est que le signal est fortement affaibli, ce qui fait que s'il y a une seconde cellule pluvieuse derrière la première, elle risque de ne pas être détectée.

— **Bande C** ($\lambda = 5$ cm, portée 200 km)
Le radar de bande C est adapté pour les plaines, particulièrement pour les pluies stratiformes. C'est un bon compromis entre les deux autres bandes mais il n'est pas adapté aux zones avec du relief.

— **Bande S** ($\lambda = 10$ cm, portée 200 km)
Le radar de bande S est spécifique aux zones tropicales ou à l'arc méditerranéen car il est particulièrement adapté aux pluies intenses. En contrepartie il y a avec cette bande un bruit atmosphérique et électronique important.

Le réseau des 33 radars métropolitains français est montré en Figure 1.9, avec la répartition entre les différentes bandes. On notera que même si la portée des radars de bande S et C est de 200 km, pour des fins hydrologiques on considère 100 km comme une limite raisonnable.

Les données radar sont lourdement post traitées. On retrouve dans Tabary (2007) une description détaillée de la méthodologie utilisée sur le réseau de radar de Météo France. Harrison et al. (2000) donne un autre exemple des corrections possibles pour les données radar, en explicitant ce qui est fait sur le réseau des radars de Grande Bretagne.

Avec un radar météorologique les erreurs de mesure peuvent être dues :

— aux échos fixes (bâtiments), qui sont supprimés par traitement statistique ;
— à la bande brillante : les précipitations proches de l'isotherme 0°C vont donner une réflectivité intense (car présence de glace, flocons) ;
— à une propagation anormale sous certaines conditions atmosphérique : le signal peut se courber et atteindre le sol, faisant croire à une précipitation ;
— à des précipitations qui n'atteignent pas le sol ;
— à des précipitations qui se trouvent sous le faisceau du radar.

La Figure 1.10 montre les principaux problèmes rencontrés par le radar, et la façon de les traiter. Le radar mesure la réflectivité à plusieurs hauteurs pour une même dis-

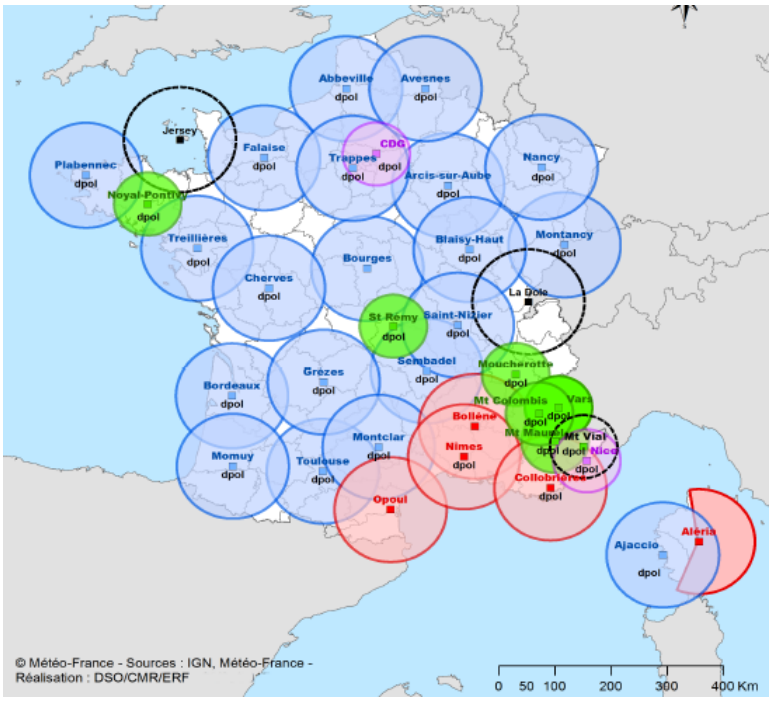


FIGURE 1.9 – Le réseau de radars de Météo-France en métropole (situation au 8 octobre 2020) : bande S en rouge, bande C en bleu ou noir et bande X en vert ou violet. Source : © Météo-France

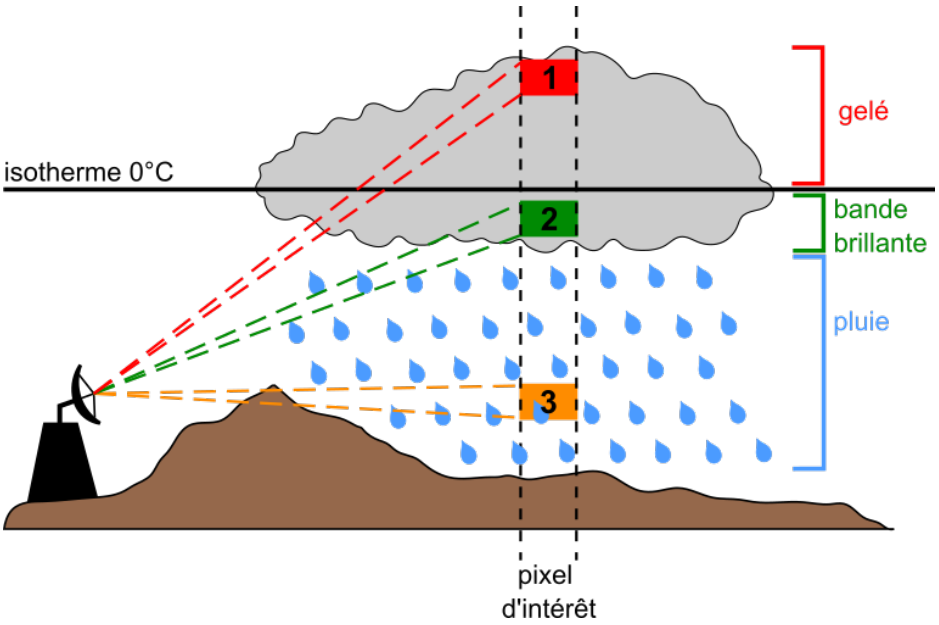


FIGURE 1.10 – Schéma expliquant le poids des pixels en fonction de la hauteur du faisceau

tance, et un poids est appliqué à chaque hauteur afin de calculer la réflectivité du pixel d'intérêt. Au-dessus de l'isotherme 0°C et dans la bande brillante (hauteur 1 et 2) on a la présence de glace et de neige qui créent des réflectivités intenses qui ne sont pas dues à des précipitations réelles. La hauteur 3 montre le problème des échos fixes, et on voit bien que ce sont les hauteurs entre 2 et 3 qui devraient avoir le poids le plus important.

Une fois la réflectivité Z corrigé, l'estimation de l'intensité de pluie R se fait le plus souvent à partir de la loi Marshall Palmer (Marshall & Palmer, 1948) : $Z = aR^b$.

La forme de cette loi paramétrique est principalement adaptée au latitudes moyennes et aux pluies synoptiques, il existe d'autres relations pour d'autres types de pluie : orage, neige, pluie tropicale.

> « La plus couramment employée est la distribution de Marshall-Palmer, qui présuppose une diminution exponentielle du nombre de gouttes de pluie comprises par leur taille dans un petit intervalle donné de variation autour d'un diamètre lorsque ce dernier augmente. [...] L'intensité de précipitation R [...] fait dépendre du diamètre des particules [...] la répartition du nombre de ces particules ainsi que l'expression de leur vitesse de chute. Alors, on peut montrer que dans les conditions les plus courantes, les grandeurs Z et R sont liées par une loi puissance de la forme $Z = aR^b$ dans laquelle les constantes a et b sont déterminées empiriquement ou bien déduites de l'expression de la

répartition granulométrique. » [1]

Les paramètres a et b sont estimés empiriquement pour la métropole française : $Z = 200R^{1.6}$

1.3.2 Données fournies par Météo France

Notre zone d'intérêt se situe à Brest, le radar qui nous concerne est donc le radar de bande C de Plabennec (Fig. 1.6 page 29 et Fig. 1.9). Des données sont disponibles depuis juillet 2006 au pas de temps 5 minutes, sur une grille de 1 km^2. Météo France fournit deux quantités : le cumul de pluie sur 5 minutes en mm (qui est une moyenne sur 1km ^{2}), et un « code qualité », qui correspond à un nombre entre 0 et 100 mais qui n'est pas à prendre comme un pourcentage de confiance. Ils considèrent toutefois d'expérience des seuils qui leur permettent de jauger la qualité de la donnée. Pour les citer : *« La donnée est jugée bonne si la qualité est supérieure à 84 (valeur communément admise) mais cela ne veut pas dire que la donnée est inutilisable si le code de qualité est inférieur à 84 ; dans ce cas, il est probable que l'estimation de lame d'eau soit sous-estimée. Par contre, ne pas utiliser de données dont le code de qualité est inférieur à 70. »*

Le descriptif donné par Météo France (en Annexe A) explique de façon simplifiée les différentes étapes de post-traitement de la donnée de réflectivité. On y trouve aussi le listing des changements dans l'algorithme de traitement, on répète ici ce qui concerne le radar de Plabennec :

- Mars 2007 : Ajustement horaire par les pluviomètres.
- Août 2009 :
  - Correction de l'atténuation par les gaz,
  - Code qualité en fonction de l'altitude → plus de poids aux élévations les plus basses,
  - Augmentation de la correction de la sous-estimation du faisceau à grande distance.
- 2012 : A Plabennec, le radar en bande S est remplacé par un radar en bande C.
- Février 2012, pour les radars à diversité de polarisation :
  - Élimination des échos d'air clair,

1. http://www.meteofrance.fr/publications/glossaire?articleId=153496

    - Correction de l'atténuation par les pluies.
- Novembre 2017 :
    - Le champ d'advection utilisé est le champ 2PIR,
    - Le format du produit évolue pour permettre (usage futur) de préciser le type d'ajustement aux pluviomètres utilisé.

En conclusion le produit radar a beaucoup évolué depuis 2006, et les données sont par nature hétérogènes, l'algorithme de post-traitement de la réflectivité ne retraitant pas les données passées. En récupérant un extrait des données depuis 2006, il a été constaté que jusqu'en 2012 les codes qualité dans la zone de Brest sont souvent très faibles ou non indiqués. Ceci ajouté au fait qu'en 2012 on a un changement du type de radar, il a été choisi de travailler sur les données radar de 2013 à 2020.

Un exemple d'image radar et des codes qualité associés est montré en Figure 1.11. La zone récupérée couvre tout le Finistère mais la zone d'étude est bien plus réduite (cadre en pointillés). Le code qualité est globalement plus faible lorsqu'on s'éloigne du centre du radar (triangle), mais on observe une zone juste autour du point du radar de Plabennec où il est plus faible (même s'il reste supérieur à 84).

La Figure 1.12 montre la répartition du code qualité en fonction des seuils donnés par Météo France. Cette image, représentative de ce qu'on trouve dans les données de 2013 à 2020, montre qu'on peut avoir quelques pixels dans notre zone d'intérêt qui ne soient « pas fiables, mais quand même utilisables » mais la grande majorité de la zone est considérée comme fiable par Météo France.

Sur la zone d'étude quelques pixels ont rapidement été repérés car ayant un comportement très différents des autres. Ils présentent un cumul de pluie très faible comparé à l'entièreté de la zone, comme c'est visible en Figure 1.13. Cette différente n'est pas explicable par le nombre de données manquantes légèrement plus élevé. Les 10 pixels étranges se trouvent au sud-ouest du centre du radar de Plabennec et ne semblent à première vue pas correspondre à des endroits où se trouveraient des obstacles très hauts ou au centre-ville de Brest. La différence disparaît à partir de mai 2018, et on voit alors apparaître une zone autour du centre de radar où les cumuls mensuels sont de nouveau plus faibles (Fig. 1.13), ce qui est cette fois-ci explicable par la présence du radar. En conclusion avant mai 2018 on a 10 pixels au sud-ouest de Plabennec qui ne sont pas fiables et qui sont dans

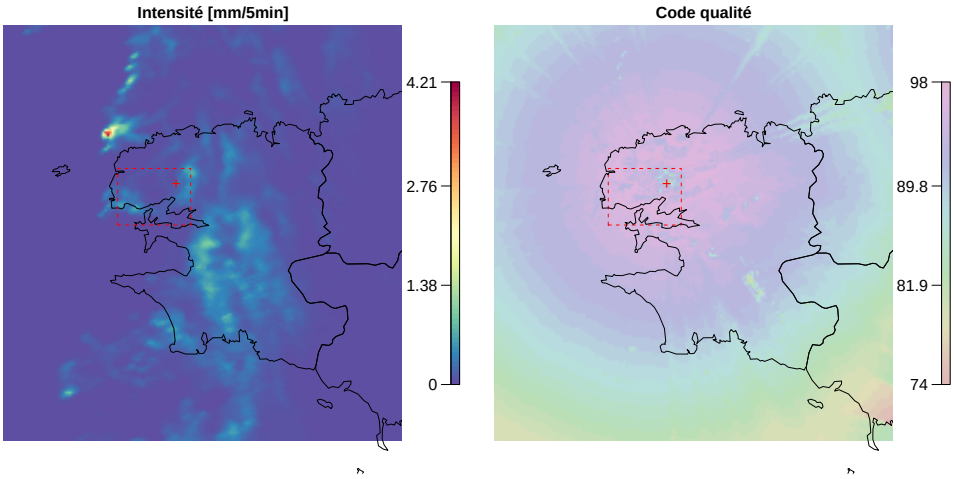


FIGURE 1.11 – Exemple d'image radar en janvier 2019 (18/01/2019 14h45 UTC). Le cadre en pointillés représente la zone d'étude et le + montre le point du radar de Plabennec.

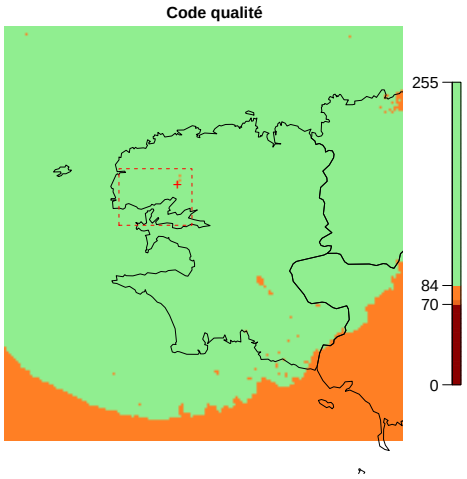


FIGURE 1.12 – Répartition du code qualité selon les critères donnés par Météo France. En vert la donnée est fiable, en orange elle est utilisable mais pas fiable et en rouge elle n'est pas utilisable. Le cadre en pointillés représente la zone d'étude et le + montre le point du radar de Plabennec.

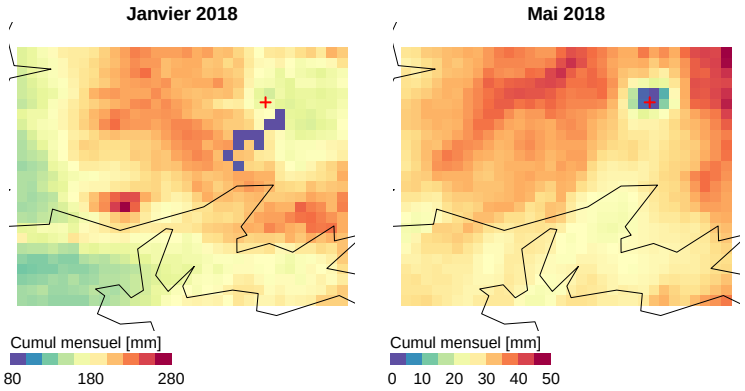


FIGURE 1.13 – Cumul mensuel et nombre de données manquantes en janvier sur les données radar à 5 minutes. La croix rouge montre le centre du radar.

la zone d'étude. Après mai 2018, la zone directement autour du radar n'est pas fiable non plus mais comme on le verra au Chapitre 3, elle ne fait pas directement partie de la zone d'étude.

Enfin on note qu'en 2016 on a autour du radar de Plabennec énormément de données manquantes en juin et juillet : jusqu'à 5000 par mois alors que sur les autres années on est plutôt à quelques dizaines de données manquantes par mois.

1.4 Comparaison des différentes sources de données

Dans cette section on va chercher à continuer à décrire le climat de la zone, notamment en vérifiant si les propriétés trouvées dans les données journalières historiques se retrouvent dans les autres sources. La cohérence des données au niveau instantané sera aussi étudiée : est-ce que les différentes sources voient passer les mêmes évènements ? Enfin la structure spatiale peut être estimée sur les pluviomètres sur réseau EDP et sur le radar, aussi il est intéressant de les comparer.

Les quatre jeux de données disponibles qui ont été présentés dans les sections précédentes seront appelés : MF journalier et MF 6 minutes pour les données du pluviomètre de la station de Météo France, Radar pour les données du radar météorologique de Plabennec et enfin EDP pour le réseau des pluviomètres d'Eau du Ponant.

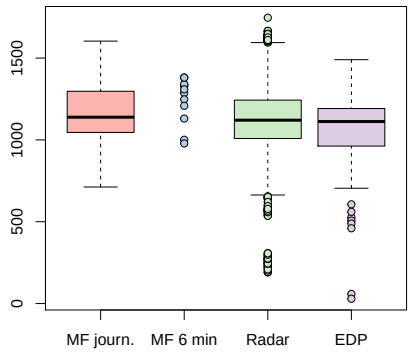


FIGURE 1.14 – Cumuls annuels dans les différentes sources de données.

1.4.1 Caractéristiques des précipitations brestoises

Climat

Les premières statistiques descriptives qu'on peut regarder car elles ne dépendent a priori pas de la résolution temporelle sont les cumuls annuels et mensuels.

Les cumuls annuels sont montrés en Figure 1.14, ils permettent de constater la cohérence très globale des données, et de confirmer qu'il y a autour de 1.2 mètres de pluie par an dans le secteur de Brest.

La Figure 1.15 montre des boxplots des cumuls mensuels pour chaque source de données. Les boxplots ne sont donc pas basés sur le même nombre d'observations : 1) pour le radar on a 200 points fois 7 ans, 2) pour la station MF à 6 minutes on a 1 point fois 12 ans, 3) pour les données journalières historiques on 1 point fois 74 ans et enfin 4) pour les données EDP on a entre entre 9 et 11 points fois 10 ans. Il faut donc garder en tête que l'étalement du boxplot n'est pas forcément représentatif de la variabilité annuelle des cumuls mensuels. Toutefois ce graphique permet de vérifier la cohérence très globale des différentes données. L'hiver toutes les sources de données sont globalement d'accord, mais l'été seuls les pluviomètres sont cohérents. En effet le radar montre des cumuls plus faibles que les autres sources de juin à septembre.

Pour ce qui est de l'occurrence de la pluie, les données ont été agrégées à la journée et à 30 minutes, qui est le plus petit pas de temps commun entre les 3 sources de données infra-journalières (dans ce dernier cas les données journalières ne sont donc pas utlisées). A la journée le seuil pour définir le temps sec est de 1 mm, et à 30 minutes il est de

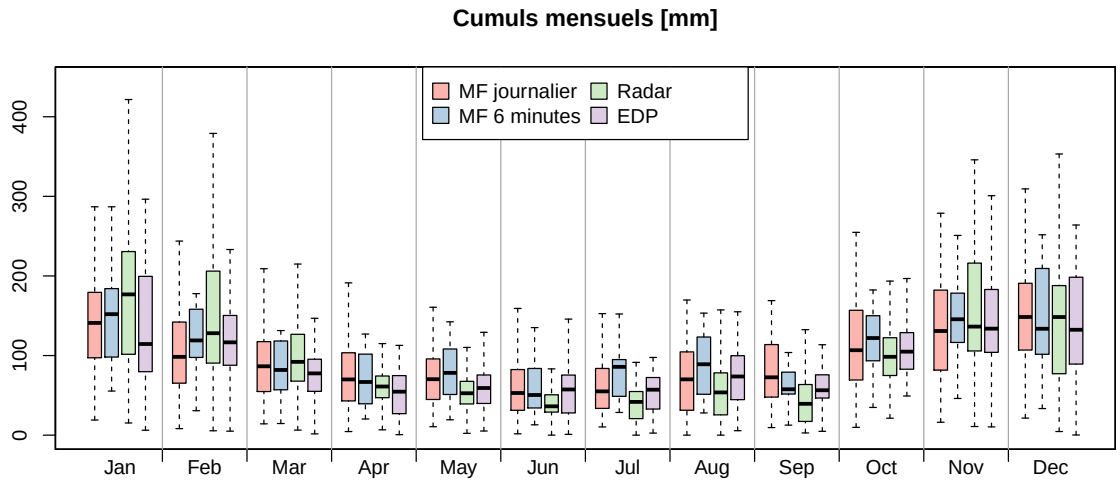


FIGURE 1.15 – Cumuls mensuels dans les différentes sources de données. Attention les boxplots ne sont pas réalisés sur le même nombre de points.

0.2 mm (car le fonctionnement des pluviomètres à bascule fait qu'une mesure nulle est en fait une mesure inférieure à 0.2 mm). Les résultats sont montrés en Figures 1.16 et 1.17.

Pour ce qui est des jours de pluie on constate que le radar contient beaucoup plus de temps sec que les autres sources de données en été, ce qui est cohérent avec la baisse de cumul observé en Figure 1.15. Comme ça a été dit la fréquence des jours de pluie est autour de 0.35 en été et 0.5 en hiver, ce qui fait sur l'année autour de 150 jours de pluie.

La probabilité de pluie à 30 minutes est assez similaire entre les deux mesures de pluviomètres (celui de Météo France à 6 minutes et le réseau d'EDP à 3 minutes). Le radar lui a une probabilité de pluie toujours plus faible. Pourtant si on regarde les mesures exactement égale à zéro, on aura là des probabilité similaires entre radar et pluviomètres, avec un peu plus de mesures nulles dans les pluviomètres en hiver. Une piste d'explication possible sont les évènements à intensité faible et constante. Les effets d'accumulation dans les pluviomètres vont retranscrire ces évènements par une période de temps « sec » avec sporadiquement des mesures à 0.2 mm. Le radar peut lui voir la vraie intensité et contiendra donc beaucoup de mesures faibles là où le pluviomètre aura surtout des zéro. C'est ce qui semble se passer en hiver, mais en été la probabilité d'avoir un zéro est la même dans les deux types de mesure. En été les intensités entre 0 et 0.2 mm du radar correspondent donc à des mesures supérieures à 0.2 mm chez les pluviomètres.

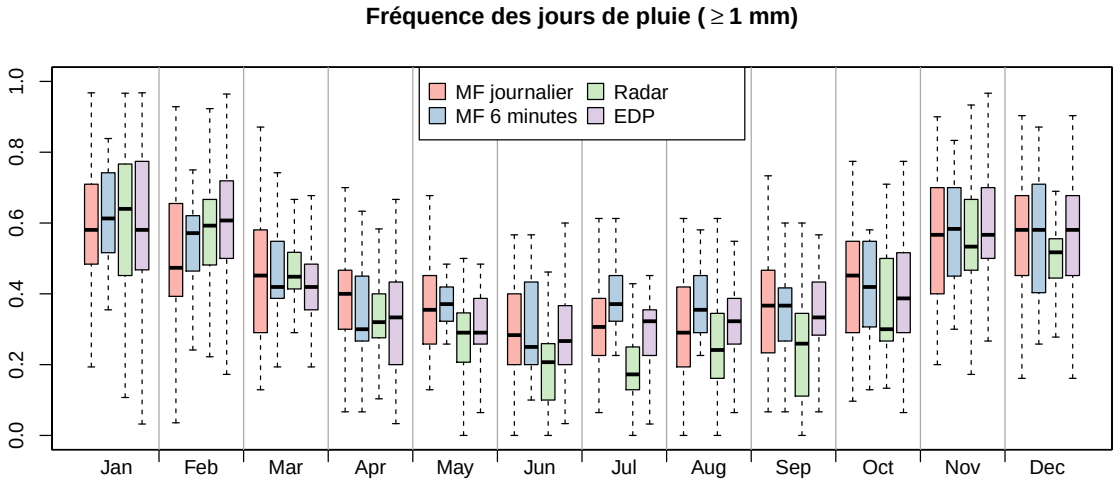


FIGURE 1.16 – Fréquence des jours de pluies (seuil 1 mm) dans les différentes sources de données. Attention les boxplots ne sont pas réalisés sur le même nombre de points.

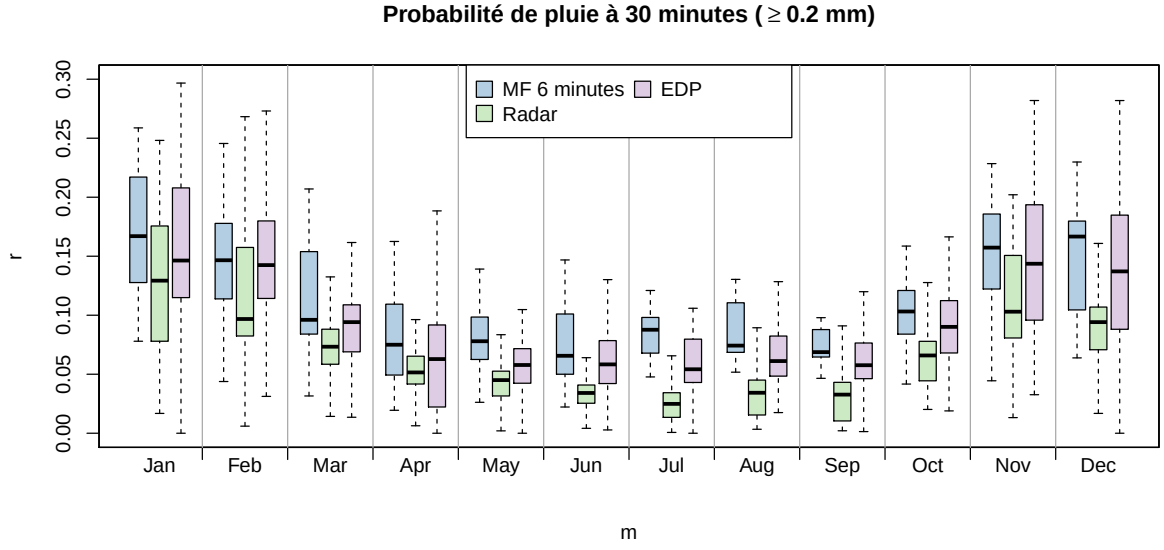


FIGURE 1.17 – Probabilité de pluie à 30 minutes (seuil 0.2 mm) dans les différentes sources de données. Attention les boxplots ne sont pas réalisés sur le même nombre de points.

Évènements pluvieux

La délimitation des chroniques de précipitation en évènements de temps sec et de temps de pluie peut être faite de nombreuses façon. Les deux points de discussion principaux sont la définition d'un seuil pour séparer le sec de la pluie et la durée de l'intervalle sec minimum qui doit apparaître pour considérer deux évènements de pluie comme séparés (e.g. Bracken et al., 2008 ; Dunkerley, 2015). Le seuil pour définir le temps sec est généralement zéro, mais une valeur positive peut être utilisée, notamment dans le cas des données radar (Schleiss et al., 2011). Pour l'intervalle sec minimum la valeur communément admise est de 6 heures mais elle varie beaucoup en fonction des études (de 3 minutes à 1 jour) et c'est souvent un choix arbitraire reposant sur la connaissance du climat de la zone d'étude et sur l'application visée. Dunkerley (2015) réalise une bibliographie des valeurs utilisées pour l'intervalle minimum, et démontre l'impact de cette valeur sur le nombre et les caractéristiques des évènements pluvieux. Des exemples d'études pour déterminer la durée de l'intervalle minimum sont Veneziano et Lepore (2012), qui la fait dépendre de la durée de l'évènement de pluie, ou encore Bonta et Rao (1988) qui utilise l'auto-corrélation en avançant que le but est d'identifier des évènements indépendants.

Dans notre cas la connaissance des données et les essais ont poussé à utiliser les règles suivantes :

1. Le temps sec est défini par zéro pour les pluviomètres, et par un seuil de 0.2 pour le radar.
2. L'intervalle sec minimum est de 3 heures : si une période de temps sec de moins de 3 heures interrompt une période de pluie, elle est intégrée dans l'évènement de pluie.
3. Si une période de temps sec est interrompue par une seule mesure de pluie, celle-ci est intégrée dans l'évènement sec.
4. Les données manquantes (NA) sont considérées comme une catégorie à part : dès qu'il y a un NA l'évènement en cours s'arrête, et dès qu'il y a de nouveau une donnée disponible l'évènement de NA s'arrête.

Pour chaque évènement on calculera les statistiques classiquement utilisées (Merz et al., 2006) suivantes : durée (h), cumul (mm), maximum (mm/h), intensité moyenne (mm/h).

L'algorithme permettant de définir ces évènements et d'en calculer les caractéristiques

	Durée [h]	Cumul [mm]	Maximum [mm/h]	Moyenne [mm/h]
Temps de pluie				
Radar	1.3 *(0.2-9)*	3 *(0.5-19.6)*	7.7 *(3.1-36.8)*	2.9 *(0.7-9.2)*
EDP	3 *(0.1-16.3)*	3 *(0.4-21.6)*	8 *(4-32)*	1.4 *(0.4-6.7)*
MF	3.1 *(0.2-16.8)*	3 *(0.4-22.4)*	6 *(2-21.3)*	1.3 *(0.4-4.9)*
Temps sec				
Radar	11 *(0.1-217)*	1.1 *(0-6.7)*	2 *(0-7.3)*	0.1 *(0-0.4)*
EDP	18 *(3.2-174)*	0.4 *(0-4.2)*	4 *(0-8)*	0 *(0-0.2)*
MF	17 *(3.8-183)*	0.4 *(0-2.6)*	2 *(0-6)*	0 *(0-0.1)*

TABLE 1.1 – Caractéristiques des évènements (toute l'année), quantiles : 50% *(5%-95%)*

mentionnées est disponible sur Github [2].

Les caractéristiques globales (sur toute l'année) des évènements de temps de pluie et de temps sec sont données en Table 1.1. Les deux données des pluviomètres donnent des résultats semblables, le radar diffère un peu plus sur certains points : les périodes de temps sec contiennent plus de pluie (cumul), ce qui s'explique par le choix d'un seuil supérieur à la précision des données, et la durée des évènements semble plus courte. Bien qu'on mélange ici été et hiver on peut se faire une idée de l'évènement pluvieux typique de la région, avec une durée de 3 heures, une intensité moyenne entre 1 et 2 mm/h avec un pic autour de 7 mm/h. Les pluies sont donc la plupart du temps plutôt courtes et peu intenses. Pour ce qui est des périodes de temps sec, la plupart sont plutôt courtes, avec une médiane inférieure à une journée, et on a l'été quelques périodes plus longues qui peuvent monter à 8 jours. Cette durée peut paraître courte mais il faut garder en tête qu'on a choisi d'interrompre les évènements de temps sec dès qu'on a plus d'une mesure de pluie, aussi ce n'est pas forcément représentatif des durées des sécheresses au sens agronomique.

Le nombre moyen d'évènements par mois est montré en Figure 1.18, et les durées sont représentées en Figures 1.20 et 1.19. On remarque une forte présence de données manquantes dans le radar en juin-juillet, qui va avec ce qui a été vu en Section 1.3.2 sur les données manquantes en 2016. Ces évènements de NA sont très courts et créent une asymétrie entre le nombre d'évènements de temps sec et de temps de pluie. La Figure 1.19 montre une chute drastique des périodes de temps sec en juin et dans une moindre mesure en juillet. On peut penser que ce soit une des raisons de la faible durée des évènements du radar notée en Table 1.1.

2. https://github.com/mbtgy/tools, cf. fonction « `blocs` »

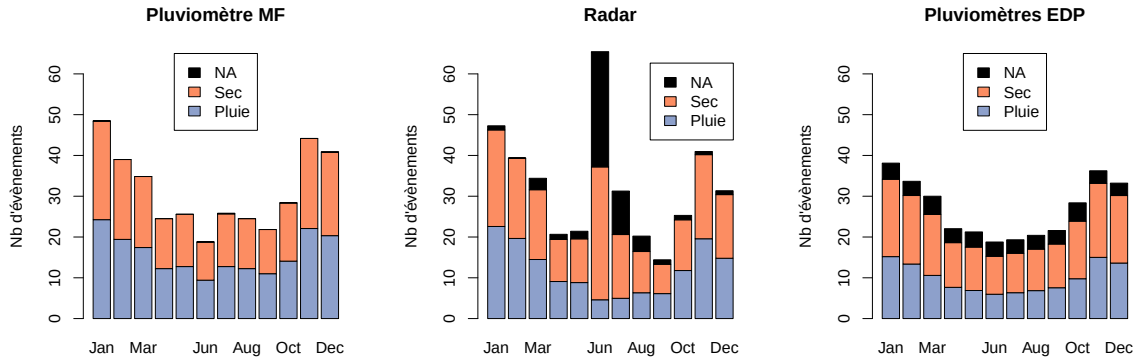


FIGURE 1.18 – Nombre d'évènements dans les différentes sources de données

En dehors de ces NA, les trois sources de données donnent des résultats similaires pour les nombres d'évènements : autour de 15 évènements de chaque type (sec / pluie) par mois, avec plus d'évènements en hiver qu'en été, ce qui s'explique par les longues périodes de temps sec en été (Fig 1.19). La durée des périodes de temps sec montre une forte variation saisonnière, contrairement aux évènements pluvieux qui ont environ la même durée toute l'année, de l'ordre de quelques heures. On trouve toutefois légèrement plus de pluies très longues en hiver.

Si les deux données de pluviomètres sont très cohérentes sur la durées des pluies, le radar donne toute l'année des évènements de pluie plus courts. Cette différence semble due au choix de fixer le seuil pour radar à 0.2 mm, car en prenant zéro elle disparaît, mais le zéro du radar n'est pas sensé correspondre au zéro des pluviomètres. On peut penser que la différence est en partie due au fait que la fin des évènements passe souvent par une période où l'intensité est inférieure à la précision des pluviomètres, mais elle peut y apparaître partiellement par effet d'accumulation dans l'auget. Le radar n'ayant pas cet effet d'accumulation, le seuil de 0.2 mm fait disparaître la fin des évènements. On peut aussi émettre l'hypothèse qu'on voit là la conséquence de ce qui avait été observé sur les probabilité de pluie en Section 1.4.1.

Pour ce qui est des autres caractéristiques, les évènements pluvieux ont une intensité moyenne plus élevée en été (juillet-septembre notamment). Toutefois le cumul reste constant, ce qui montre que l'augmentation de l'intensité moyenne en été est compensée par la baisse de la durée. Les maximums varient peu au cours de l'année, même si on note une légère augmentation de la queue de distribution des maximums sur les mois de juillet à décembre.

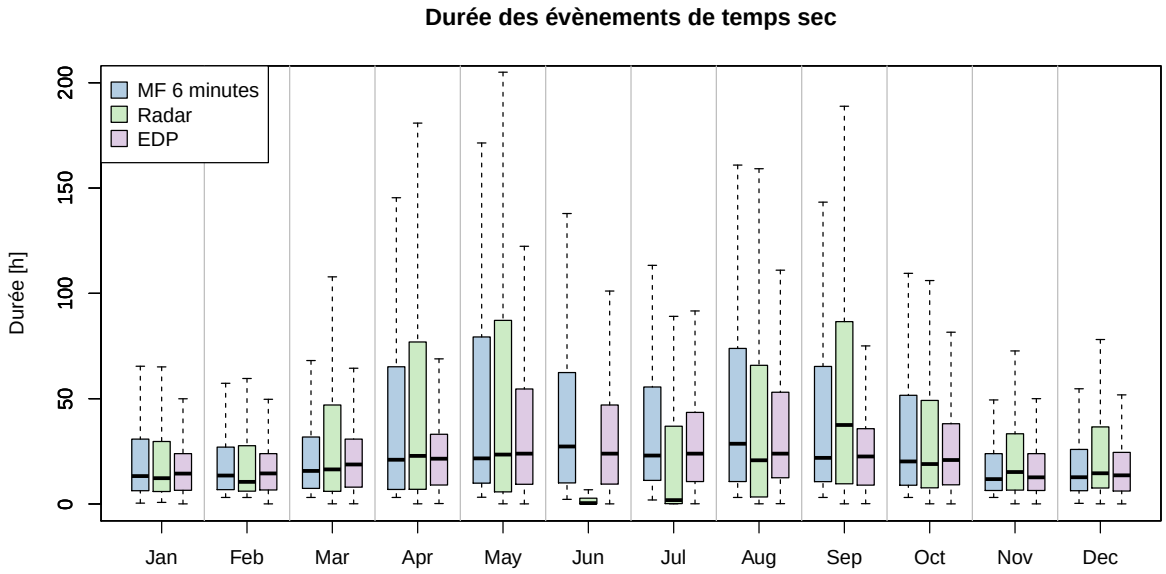


FIGURE 1.19 – Durée des évènements de temps sec dans les différentes sources de données

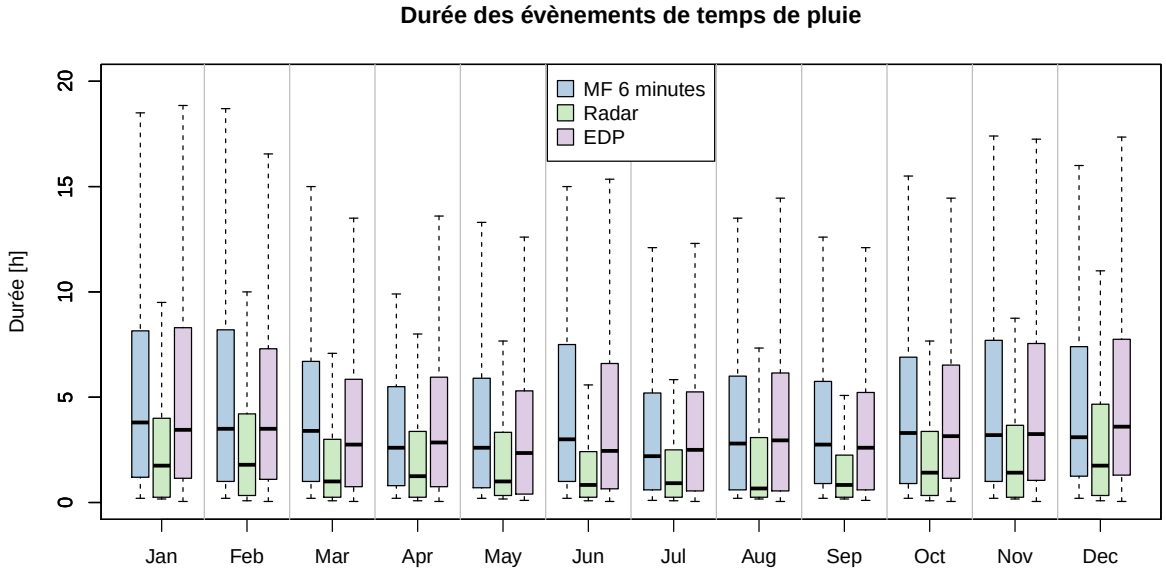


FIGURE 1.20 – Durée des évènements de temps de pluie dans les différentes sources de données

En conclusion les évènements pluvieux typiques de la région brestoise sont plutôt courts (autour de 3 heures), ce qui correspond au climat océanique très changeant, et montrent assez peu de variations saisonnières. L'intensité moyenne des évènements est entre 1 et 2 mm/h, et en été on trouve des pluies à la fois un peu plus courtes et plus intenses. Les périodes de temps sec présentent une forte variabilité saisonnière, avec une durée d'un jour en moyenne l'été contre 12 heures l'hiver.

1.4.2 Mesure instantanée

Dans cette section on va chercher à vérifier la simultanéité des données. Pour ça on va simplement mesurer la corrélation entre les différentes sources : à 1 jour, 30 minutes et 15 minutes. Spatialement on a des points qui se recoupent uniquement entre le radar et les pluviomètres. Pour l'échelle journalière on peut ignorer l'écart spatial entre les points étant donné que comme on l'a vu précédemment les évènements pluvieux durent généralement quelques heures. A 30 et 15 minutes on aura donc seulement la corrélation entre chaque pluviomètre et le pixel radar qui le contient.

Plus le pas de temps est faible plus il y a de temps sec et par voie de conséquence plus il est « facile » d'avoir une forte corrélation. On calculera donc les corrélations uniquement sur les mesures où au moins une des sources de données mesure une intensité supérieure à zéro.

journ.	MF	Radar
Radar	0.69 *(0.65-0.74)*	
EDP	0.67 *(0.34-0.79)*	0.74 *(0.36-0.87)*

30 min	Radar
MF	0.72
EDP	0.54 (0.29-0.73)

15 min	Radar
EDP	0.47 *(0.28-0.64)*

TABLE 1.2 – Corrélation à différentes échelles d'agrégation temporelle entre les données : moyenne *(min-max)* ou corrélation entre les deux points les plus proches.

A l'échelle journalière les corrélations sont de l'ordre de 0.7. Il n'est pas étonnant que la corrélation la plus faible soit entre les pluviomètres EDP et celui de Météo France, car les pluviomètres ont des erreurs spatialement indépendantes. Le radar a lui une correction

qui fait intervenir un champ d'advection, il est donc plus lisse et a des erreurs spatialement dépendantes. Il est étonnant de noter que la corrélation entre le radar et le réseau Eau du Ponant est plus forte que celle entre le radar et le pluviomètre de Météo France, qui est pourtant celui utilisé dans la correction des données radar.

Pour ce qui est du réseau d'Eau du Ponant, on a bien une baisse de la corrélation entre pluviomètres et radar avec la diminution pas de temps, la baisse est assez forte ce qui peut être dû à la discrétisation qui est beaucoup plus visible sur des pas de temps court.

1.4.3 Structure spatiotemporelle

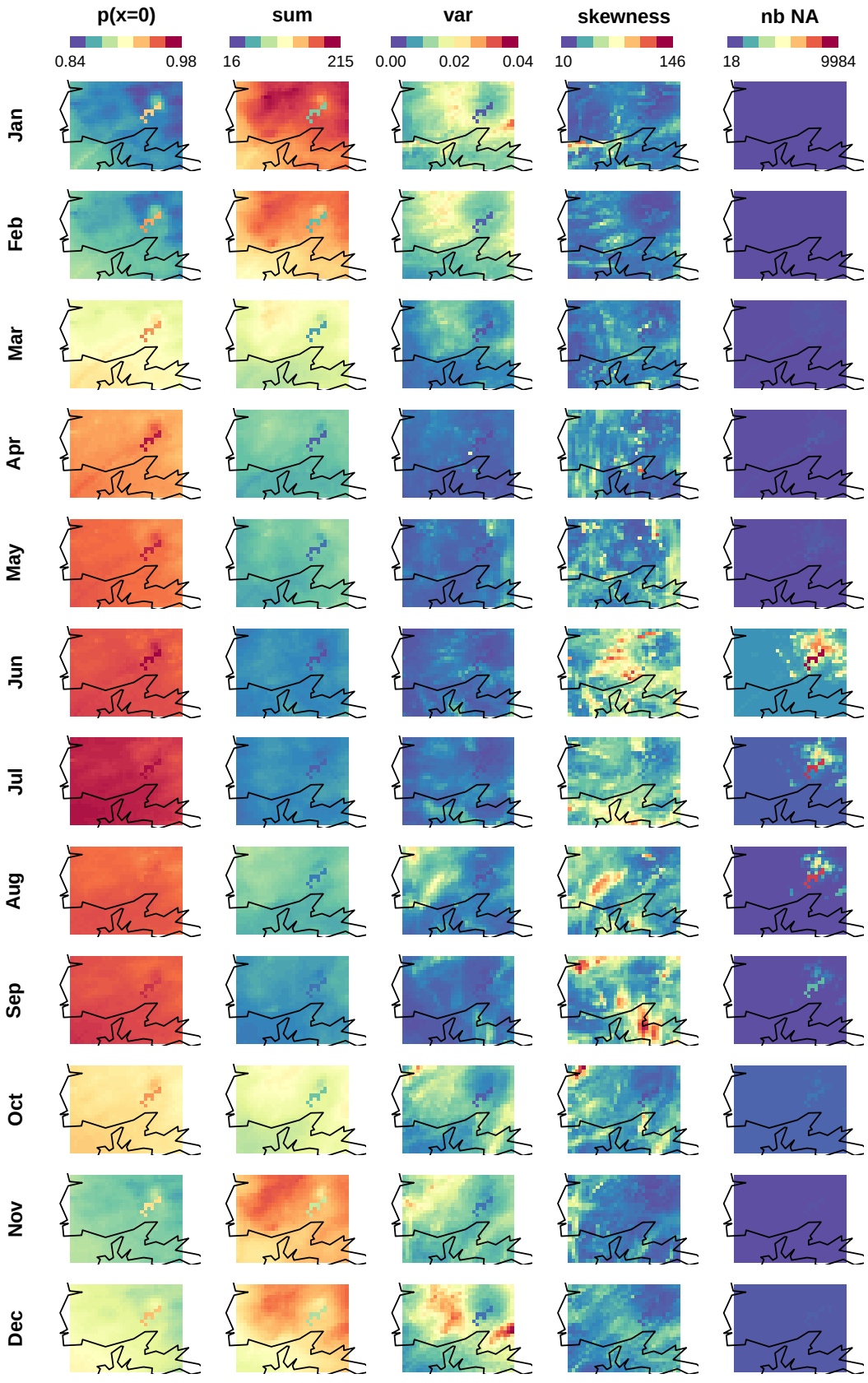
Tout d'abord on peut grâce au radar étudier l'évolution de la structure spatiale des statistiques descriptives au cours de l'année, c'est ce qui est montré en Figure 1.21. On retrouve la variation de la probabilité de temps sec observée en Figure 1.17, et on constate que la probabilité de pluie augmente quand on s'avance dans les terres, ce qui peut s'expliquer par le fait qu'en zone côtière la météorologie est complexe :

> « Cette frontière est en effet le cadre de phénomènes particulier [...], qui se produisent à cause d'un changement très marqué d'humidité, de température, de rugosité et d'ensoleillement entre les deux milieux. » [3]

La démarcation qui suit le trait de côte est surtout marquée l'hiver, ce qui peut s'expliquer par le fait qu'il y ait simplement plus de pluie ou par le fait que ce soient les mois où on a le plus de fronts de pluie venant de la mer. Les mêmes constats sont fait pour le cumul mensuel : on observe une variation spatiale plus marquée en hiver qui suit le trait de côte. On en déduit que dans les terres, il pleut plus et plus souvent.

La variance est globalement plus élevée l'hiver, ce qui est cohérent avec le fait que les pluies conséquentes ont plutôt lieu l'hiver et donc ces mois-ci la distribution est plus remplie au niveau des intensités moyennes à fortes. L'impact des valeurs extrêmes sur la variance semble modéré, même si l'été on a quelques tâches (par exemple en août) qu'on retrouve très marquées dans l'asymétrie (moment d'ordre 3). Les cartes d'asymétrie sont très irrégulières et on voit ressortir quelques tâches qui sont liées à un seul évènement. De plus il est attendu que l'estimateur de l'asymétrie soit très instable car comme ce sera discuté au Chapitre 2, la queue de distribution des données est lourde et les moments d'ordre trois voire deux sont potentiellement infinis.

3. https://fr.wikipedia.org/wiki/M%C3%A9t%C3%A9orologie_c%C3%B4ti%C3%A8re



GURE 1.21 – Statistiques descriptives des données radar à 5 minutes

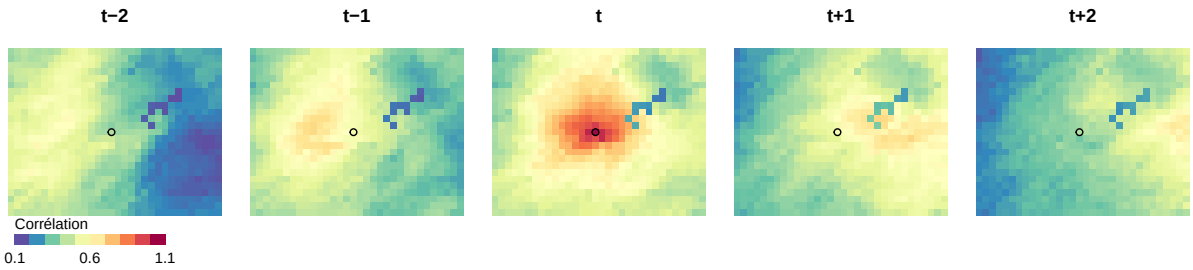


FIGURE 1.22 – Corrélations entre le pixel central (point) et le reste de la zone à différents décalages temporels, calculées en décembre.

Finalement le fort nombre de données manquantes de juin à août autour de Plabennec est en fait uniquement présent jusqu'en 2016. D'après la fiche produit transmise par Météo France (Annexe A) il n'y a rien de particulier en 2016 mais il reste très probable que ce changement soit dû à une évolution du produit.

La Figure 1.22 montre les cartes de corrélation à différents décalages temporel entre la carte et le point central de la zone d'étude, représenté par un point. Les données étant à leur résolution la plus fine on va ici de $t-10$ min à $t+10$ min. On constate que d'une part les cartes sont très lisses, ce qui est cohérent avec ce qui a déjà été observé. La corrélation baisse à mesure qu'on s'éloigne du point central (au temps t), et à première vue cette baisse ne semble pas dépendre de la direction. La direction principale des précipitations de l'ouest vers l'est se retrouve clairement dans le déplacement de la zone de plus forte corrélation.

La Figure 1.23 montre l'évolution de la corrélation entre deux points de la zone en fonction de la distance les séparant. Les données sont agrégées à 15 minutes (le plus petit pas de temps commun des deux mesures) afin de représenter les résultats du radar et des pluviomètres sur le même graphique. La corrélation baisse beaucoup plus lentement avec la distance dans le radar que dans les pluviomètres, ce qui montre de nouveau que les données radar sont plus lisses spatialement que les données des pluviomètres. La dépendance spatiale est globalement plus faible dans les pluviomètres. On note aussi qu'à une distance donnée, la corrélation entre les pluviomètres a une variance beaucoup forte qu'entre les pixels du radar. Il semble donc difficile d'estimer la structure spatiale des précipitations avec les données des pluviomètres d'Eau du Ponant.

Remarque : La deuxième « banane » des données radar avec une corrélation plus faible correspond aux dix pixels étranges qui ont été abordés en Section 1.3.2.

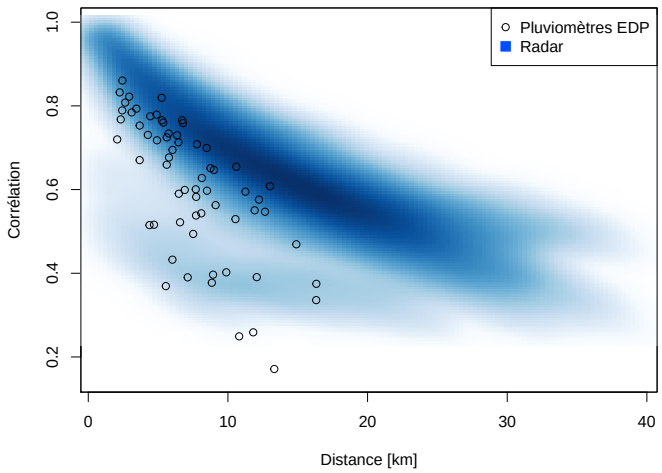


FIGURE 1.23 – Corrélation en fonction de la distance entre les points sur les données radar et les pluviomètres d'Eau du Ponant à 15 minutes.

1.5 Conclusion

	Période couverte	Résolution temporelle	Résolution spatiale	Précision
Données historiques MF	1945-2018	1 jour	1 point	0.1 mm avant 1990, 0.2 mm après (?)
Pluviomètre MF	2006-2018	6 min	1 point	0.2 mm
Pluviomètres EDP	2010-2019	3 min	9 ou 11 points	0.2 mm
Radar MF	2013-2019	5 min	grille 1 km^2	0.01 mm

TABLE 1.3 – Résumé des caractéristiques des données disponibles

On dispose de quatre jeux de données, dont les caractéristiques techniques sont résumées en Table 1.3. La station Météo France, qui nous donne des données journalières depuis 1945 et des données à 6 minutes depuis 2006, fournit les données les plus fiables. Toutefois cette donnée n'est disponible qu'en un point qui n'est pas situé dans le bassin d'intérêt, ce qui ne permet pas d'étudier la variabilité spatiale des précipitations, un point crucial du projet MEDISA. Le réseau des pluviomètres d'Eau du Ponant donne une meilleure résolution spatiale avec une dizaine de points, mais le principal défaut de ce jeu de donnée est la forte présence d'erreurs de mesure. Certaines erreurs sont facilement identifiables mais la donnée reste moins fiable que celle de Météo France. Il est à noter que tous les pluviomètres ne présentent pas la même fiabilité, car ils ne sont pas tous installés dans les mêmes conditions. Les erreurs de mesure des pluviomètres sont a priori

indépendantes, contrairement à celles de la dernière source de données, le radar météorologique de Météo France. En effet ce dernier offre la meilleure résolution spatiale et la meilleure précision (0.01 mm contre 0.2 mm pour les pluviomètres), mais c'est un produit avec un algorithme de traitement lourd qui crée une dépendance spatiale dans les erreurs de mesure. Bien que les radars météorologiques soient généralement considérés comme représentant moins bien la « vraie » intensité de pluie, ils restent la meilleure source d'information pour la structure spatiotemporelle des précipitations, et en particulier pour le déplacement des cellules pluvieuses.

En conclusion aucun des jeux de données n'est utilisable en l'état pour analyser la sensibilité du modèle hydraulique d'Eau du Ponant, mais de ce chapitre on peut tirer les caractéristiques de la précipitation bien représentées par chaque jeu de données.

Chapitre 2

Modélisation de la précipitation de l'échelle infra-horaire à l'échelle journalière avec un modèle méta-Gaussien à queues lourdes

---

La modélisation marginale de la précipitation à une échelle temporelle fine est la première étape pour aller vers la construction d'un modèle d'assimilation de données ou un générateur aléatoire. De plus modéliser correctement les marges permet de mieux comprendre les caractéristiques des données de précipitation sur la région brestoise.

La première section de ce chapitre est un article soumis au journal Water Resources Research, dont le pré-print est aussi disponible aussi en ligne (Boutigny et al., 2021).

2.1 Modelling rainfall from sub-hourly to daily scale with a heavy tailed meta-Gaussian model

2.1.1 Introduction

Precipitation is a key variable for many environmental studies such as hydrology of course but also agronomy, meteorology, climatology, etc. There is an abundant literature on the subject showing the interest for accurate rainfall models, and until nowadays modelling of the distribution of precipitation is still discussed, especially when considering rainfall accumulated on short periods of time (sub-hourly). A particular impetus for this work was the need for realistic rainfall conditions to be used as input to an urban hydrological model. The model requires rainfall conditions with a time step of 3 minutes

whereas the sources of rainfall data may be available at other time step (e.g. 5 minutes for the radar, 6 minutes or even daily for rain gauges). In such situation it is useful to have a parametric model for the rainfall distribution which is interpretable and flexible enough to describe rainfall accumulated over different periods of time.

Precipitation is a tricky variable to model, due to its discrete component (dry/wet measurements) and to the shape of the positive part of the distribution. Most of the measurements are small intensities but extreme rainfall can occur, especially when considering sub-hourly rainfall, creating a heavily left skewed distribution on a positive support with a peak in zero.

The marginal distribution of rainfall at monthly to daily time scale has been widely studied. Most of the time the occurrence of precipitation (dry/wet measurement) is modelled separately from the intensity (amount of precipitation on a wet measurement). The most popular distribution for positive daily precipitation is probably the gamma distribution (see e.g. Castellvı et al. (2004), Katz (1999) et D. S. Wilks (1999)), which is a pretty good fit for precipitation at this scale. However many other distributions have been used, for example mixed exponential (D. S. Wilks, 1999), Weibull (Castellvı et al., 2004), log-normal (Shoji & Kitaura, 2006), etc. Comparisons have been made for specific data sets, for example Woolhiser et Roldan (1982) ranked the mixed exponential first and the gamma second, Liu et al. (2011) ranked the log normal first, then mixed exponential, gamma and finally exponential, and Selker et Haith (1990) compares single-parameter distributions. The performance of these different distributions is very dependent on the location of the meteorological stations as the climate strongly impacts the distribution of rainfall.

Another usual strategy is to make the data Gaussian, using for example a square root (Panofsky et al., 1958). The Box-Cox transformation (Box & Cox, 1964) is widely used in the statistical literature to transform non-Gaussian variables into Gaussian variables, and has been used for precipitation (Cecinati et al., 2017 ; Hussain et al., 2010). Tukey's g-and-h transformation, introduced by Tukey (1977), can be used to gaussianize heavy tailed distributions (Goerg, 2015), which is interesting for precipitation. For example Xu et Genton (2017) use it to gaussianize precipitation. Using the the log transformation leading to the log-normal distribution is another example of that strategy. The idea lying behind this method is that it can be easier to work with Gaussian data, and it is especially popular among people working with space-time modelling where many methods have been developed for the Gaussian case, as it is the case in Hussain et al. (2010). The discrete part

of the distribution may be neglected when working on rainfall accumulated over long time period such as monthly data, since the probability of having a dry measurements is very low. However for daily and sub-daily scales, the peak of dry measurement is generally too important to be neglected. A usual strategy when working with univariate distribution is to model it separately using a Bernoulli variable but this approach is difficult to generalise to multivariate (e.g. spatio-temporal) data sets.

Transformed censored Gaussian models (also known as meta-Gaussian models) have been developed to keep on working in a Gaussian framework for daily and sub-daily scales. The idea is to define a latent Gaussian variable - that does not need to have a physical interpretation -, to use a left censorship to reproduce the dry measurements and to transform the rest of the distribution to match rainfall intensities. One advantage of this approach is that it allows using the many probabilistic results and statistical methodologies developed for the Gaussian framework. It allows for example building multivariate, temporal, spatial or spatio-temporal models or using Kalman-like algorithm for rainfall and this has lead to many applications in hydrology. For example, it has been used for rainfall disaggregation (Allard & Bourotte, 2015 ; Allcroft & Glasbey, 2003 ; Guillot & Lebel, 1999), downscaling and model correction (Maraun et al., 2010 ; Rebora et al., 2006 ; Zhao et al., 2017), short term or spatial prediction (Benoit et al., 2018 ; Sigrist et al., 2012), building stochastic weather generators (Ailliot et al., 2009 ; Bardossy & Plate, 1992 ; Kleiber et al., 2012), data assimilation (Lien et al., 2013) or merging different data sources (Cecinati et al., 2017).

Another difficulty that arises as the observation time step decreases is the fact that the peaks of the events are better measured and hence the strong intensities are more present in the observations. As a consequence, when working at daily to sub-daily scales, it has been found that the positive part of the distribution is heavy tailed and precipitation has been studied in the context of extreme value theory : for example Papalexiou et al. (2013) and Katz (1999) fitted a Generalized Pareto (GP) distribution to rainfall data. The main drawback of extreme value theory is that it studies threshold exceedances and not the full distribution. However recent models such as the one of Naveau et al. (2016) have been developed to model the full distribution. In this paper extreme value theory is applied to the lower tail as well as is upper tail, allowing to model the entire distribution. Such questions about the shape of the lower and upper tails have not been studied in the context of meta-Gaussian models : this is one of the objectives of the present paper. Consequently the model of Naveau et al. (2016) will be used as benchmark for the original

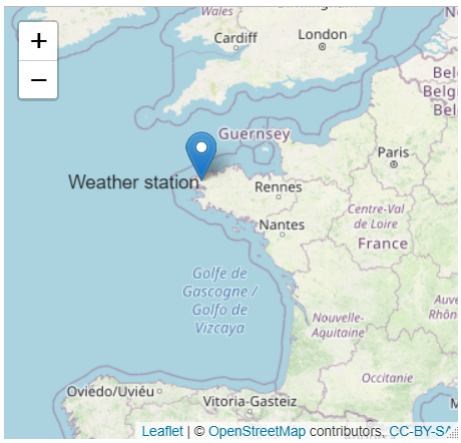


FIGURE 2.1 – Location of the weather station (Guipavas)

model presented in the paper.

The paper focuses on developing a new meta-Gaussian model based on the desired properties for the upper and lower tails of the distribution. In particular the model should be able to produce heavy tails in order to fit precipitation data at small time steps. It should also be flexible enough to fit rainfall accumulated on a wide range of time scales.

The data set that will be used throughout the paper are introduced in Section 2.1.2. The proposed model is justified and presented in Section 2.1.3, along side with other models that will be used for comparison. Finally results are shown in Section 2.1.4, including a comparisons between the different models and a test of the flexibility of the models by varying the time step.

2.1.2 Data

In this section the data used throughout the paper is be presented and the properties of the precipitation distribution raised in the introduction are illustrated. The data set is a 12 years series of precipitation measured at Guipavas, France (geographical coordinates 48.45°N, 4.38°W) provided by Météo France. Guipavas is located in the North-West part of France, close to the city of Brest (cf. Figure 2.1). Its climate is influenced by oceanic conditions, characterised by a low temperature amplitude and alternation of rainy frontal systems coming from the Atlantic and high pressure systems which bring dryer conditions, with a mean annual precipitation of 1200 mm and a wet day (≥ 1 mm) frequency of about 2 days out of 5.

Precipitation is available at a 6 minutes time step [1] from 2006 to 2017 and in order to remove the seasonal components, a focus is made on the three months of Autumn, i.e. October, November and December. Figure 2.2 shows as an example the time series for 2006. Figure 2.3 shows the histogram of precipitation for the whole series, with the entire distribution on the left, the wet measurements only in the middle and a focus on low and moderate intensities (between 0.2 and 2 mm) on the right.

The measurement device is a tipping bucket gauge with a 0.2 mm precision. Hence the data present a strong discretization, visible in Figure 2.3 on the right. In Figure 2.2 many 0.2 mm measurements can be observed in what seem to be dry periods (especially in October). They can be due to the dew that is sometimes sufficient for the gauge to toggle. A drizzle with an intensity lower than 0.2mm/6min can also make 0.2 measurement appear more or less regularly in a "dry" period.

The histogram on the left panel of Figure 2.3 shows a strong peak in zero, the rest of the histogram being almost invisible. It is expected at a small time step and a fully continuous distribution obviously can not handle the observed frequency of dry measurement (0.94). As for the positive rainfall (Fig. 2.3, left and middle), the distribution is strongly skewed as most measurements are low intensities. The positive rainfall have a 99.9% quantile of 3.2 mm. The observations above this quantile are highlighted on the left histograms using stars and maximum of the series is 9.4 mm. Some on these heavy rainfall events are also visible in Figure 2.2.

To sum up the goal is to model precipitation at a sub-hourly scale, hence to have a strongly skewed distribution with a discrete component in zero and the ability to produce heavy tails. The next section discusses the choice of such model.

2.1.3 Models

Meta-Gaussian Models

A classical approach for modelling rainfall, sometimes called meta-Gaussian model, is to assume that rainfall amounts Y can be linked to a Gaussian variable according to

$$Y = 0 \times \mathbb{1}_{X<0} + \psi(X) \times \mathbb{1}_{X\geq 0}, \quad \text{with } X \sim \mathcal{N}(\mu, 1), \tag{2.1}$$

where $\mathbb{1}_A$ is the indicator function equal to 1 if condition A is true, and equal to 0

1. Data can be found for free online at https://donneespubliques.meteofrance.fr/, but only at 1 hour time step.

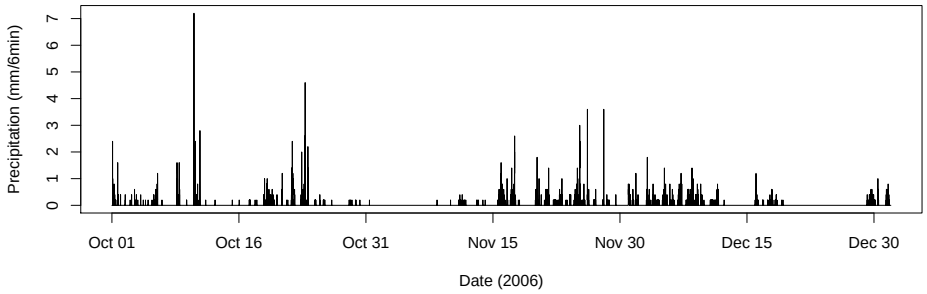


FIGURE 2.2 – Example of precipitation series in Autumn 2006 in Guipavas.

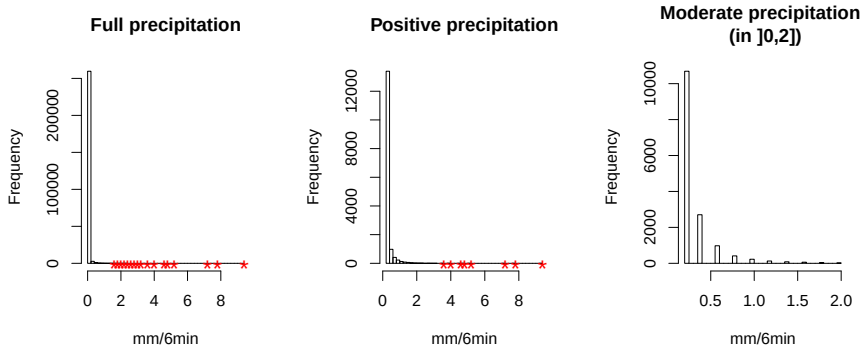


FIGURE 2.3 – Histograms of precipitation at a 6 minutes time step (2006-2017, Guipavas). Whole distribution on the left, positive part in the middle, low intensities (between 0.2 and 2 mm) on the right. The stars represent the observations above the 99.9% quantile of the represented distribution (i.e. 1.4 mm on the left, and 3.2 mm in the middle).

otherwise. Y denotes the rainfall, X is a Gaussian random variable with mean μ and variance 1 and $\psi : [0, +\infty[\to]0, +\infty[$ is an increasing function which is generally refereed to as the anamorphosis in the literature. The operation of such model is schematised in Figure 2.4. The censorship in 0 produces dry conditions with a proportion linked to the mean of the Gaussian (step 1 in Figure 2.4), whereas the transformation ψ acts on the positive part of the distribution which corresponds to wet conditions (step 2 in Figure 2.4).

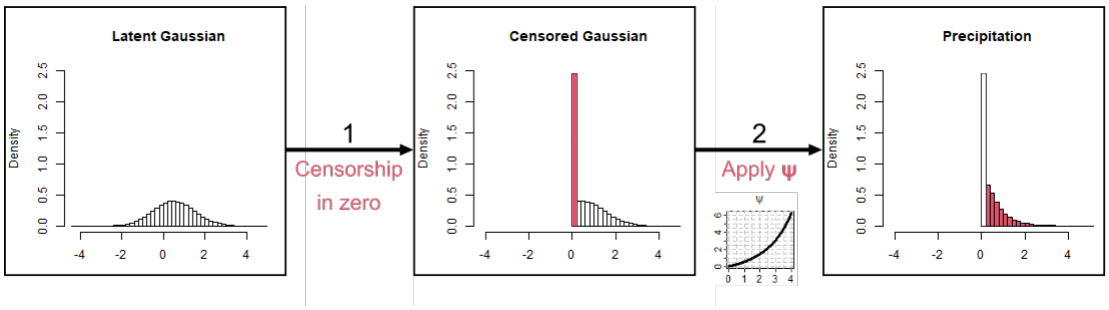


FIGURE 2.4 – Schematic functioning of a meta-Gaussian model. The coloured areas in the histograms represent the part of the distribution that just has been modified by the process.

The cumulative distribution function (cdf) of such model can be written as

$$F_Y(y) = \begin{cases} \Phi(\psi^{-1}(y) - \mu) & \text{if } y > 0 \\ \Phi(-\mu) & \text{if } y = 0 \end{cases} \tag{2.2}$$

where Φ is the cdf of the standard normal distribution.

Remark that this meta-Gaussian model is general since any positive random variable Y which has a discrete component at the origin like precipitation can be written as (2.1) using

$$(x) = F_Y^-(\Phi(x - \mu)) \tag{2.3}$$

where $\mu = -\Phi^{-1}(P(Y = 0))$ and F_Y^- denotes the quantile function of Y (generalized inverse function of the cdf F_Y of Y). Plugging a non-parametric estimate of the quantile function of Y in (2.3) allows building non-parametric estimates of ψ, see e.g. Cecinati et al. (2017) et Lien et al. (2013). The dots in Figure 2.5 show the estimate obtained on the particular data set introduced in Section 2.1.2. The shape of ψ near zero is linked to the small precipitations : a horizontal tangent at the origin means that they are more low rainfall than expected low values in the truncated Gaussian. The probability of low rainfall increases if ψ is flatter at the origin. The growth speed is linked to the upper tail :

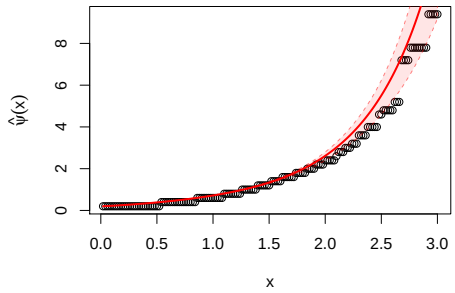


FIGURE 2.5 – Non-parametric estimate of the anamorphosis function based on (2.3) (dots). The plain curve corresponds to the proposed parametric model fitted to the data (see Section 2.1.4), and the dotted line is its 95% confidence interval computed on 500 non parametric bootstrap replicates.

a convex-exponential shape indicates that the tail is heavier than a Gaussian one.

However, parametric approaches are generally favoured in the applications and many models have been proposed in the literature. The most classical is the one of Bardossy et Plate (1992) et Sigrist et al. (2012)

$$(x) = \sigma x^{1/\alpha}, \tag{2.4}$$

but other transformation have been proposed such as Allcroft et Glasbey (2003) which uses a quadratic power function, Rebora et al. (2006) which uses a simple exponential or finally in Allard et Bourotte (2015)

$$(x) = \sigma_1(\exp(\sigma_2 x^{1/\alpha}) - 1) \tag{2.5}$$

is used. To force the resulting distribution to match a specific distribution the inverse of a cdf can also be used, as it is the case with the Gamma distribution in Kleiber et al. (2012).

Transformation (2.4) being the most commonly used, it will be a point of comparison and will be referred to as the classical meta-Gaussian model.

Low and Heavy Rainfall Modelling with Meta-Gaussian Models

The choice of an appropriate anamorphosis function for a particular application is typically a trade-off between model complexity, versatility, tractability and interpretability.

In this section, it is advocated that the properties of lower and upper tails of the positive part of the rainfall distribution may also provide interesting insights.

Different studies have shown that rainfall at daily or sub-daily scales are generally heavy tailed (see e.g. Papalexiou et al. (2013) and references therein). In this situation, should be chosen such that the transformed Gaussian variable defined by (2.1) is tail equivalent with a Pareto distribution with positive shape parameter ξ. According to Appendix C, this holds true if and only

$$\lim_{x\to\infty} \frac{x\psi(x)}{\psi'(x)} = \frac{1}{\xi}. \tag{2.6}$$

The first function that comes to mind which satisfies (2.6) is $x \mapsto \exp\frac{\xi x^2}{2}$. By re-writting ψ as

$$(x) = \exp\frac{\xi x^2}{2} \exp u(x)$$

- which is always possible - condition (2.6) becomes

$$\lim_{x\to\infty} \frac{u'(x)}{x} = 0.$$

This condition seems easier to work with as it allows understanding that loosely speaking, the anamorphosis should increase "like" the function $x \mapsto \exp\frac{\xi x^2}{2}$ when $x \to +\infty$ to get heavy tail distributions. In particular, one can verify that most of the anamorphosis that can be found in the literature - including the classical meta-Gaussian model (2.4) introduced previously - do not satisfy condition (2.6) and hence is not suitable for modelling rainfall with heavy tail. An exception is the model (2.5) of Allard et Bourotte (2015) which is tail equivalent with a Pareto distribution if and only if $\alpha = \frac{1}{2}$.

Naveau et al. (2016) advocated, using arguments of the extreme value theory applied to low rainfalls, that the lower part of the distribution of the positive amount should approximately follow a power-law, i.e. satisfy

$$\lim_{y\downarrow 0} \frac{F_Y(y) - F_Y(0)}{y^\alpha} = C$$

for some positive constant C and shape parameter $\alpha > 0$. Remark that this condition holds true in particular for the Gamma distribution (with shape parameter α) which is often used to model daily rainfalls. Using (2.2) to derive a first order expansion of F_Y

close to 0, it can be shown that this holds true if and only

$$(x) = x^{\frac{1}{\alpha}} K(x) \tag{2.7}$$

with K such that $\lim_{x \downarrow 0} K(x)$ exists and is strictly positive. This condition is verified by most of the anamorphosis that can be found in the literature, including the classical meta-Gaussian model (2.4) obviously. Remark however that for model (2.5) the same parameter α controls the shape of the distribution for low and heavy rainfall, and that is not possible to create an heavy tailed distribution with a power shape parameter different from $\alpha = \frac{1}{2}$ for low rainfalls.

To conclude, for the distribution to have a Pareto upper tail and a lower tail that follows a power law, the anamorphosis ψ should be chosen such that conditions (2.6) and (2.7) are satisfied.

Generalized Pareto Meta-Gaussian Model

Based conditions (2.6) and (2.7), this paper advocates the use of the simplest anamorphosis function that satisfies both condition, i.e.

$$(x) = \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2} \tag{2.8}$$

with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^{+*}$, $\alpha \in \mathbb{R}^{+*}$ and $\xi \in \mathbb{R}$. The distribution of the random variable Y defined through (2.1) with $X \sim \mathcal{N}(\mu, 1)$ and ψ given by (2.8) will be referred to as the GP meta-Gaussian distribution with parameter $(\mu, \sigma, \alpha, \xi)$.

The expected roles of those four parameters result from the analytical study of a meta-Gaussian model (2.1) with anamorphosis (2.8). μ is directly related to the dry probability through (2.2), σ is a scale parameter, α controls the shape of the lower tail and ξ the shape of the upper tail.

If $\xi > 0$, the distribution is tail equivalent with a Pareto distribution with shape parameter ξ. It implies in particular that $E[X^p] = +\infty$ if $p > \frac{1}{\xi}$.

The case $\xi = 0$ corresponds to the classical meta-Gaussian model (2.4). A negative ξ can seem counter intuitive for modelling rainfall as it creates an upper bound to the distribution, but when considering rainfall accumulated on a long period (several days) the fitted model naturally goes for negative ξ, as it will be shown in Section 2.1.4 (Fig. 2.10).

When $\xi < 0$, ψ is strictly monotonic increasing only on the interval $(0, x_{sup})$ with

$$x_{sup} = \sqrt{\left|\frac{-1}{\min(\alpha\xi, 0)}\right|}. \tag{2.9}$$

For negative ξ, the GP meta-Gaussian distribution is thus defined by applying (2.1) with given by (2.8) to the Gaussian variable $X \sim \mathcal{N}(\mu, 1)$ truncated at x_{sup}. Remind that truncation means that values above x_{sup} are not observed - unlike a censorship, which is used to create the dry component, where the observations above the bound take the value of the bound. The support of the distribution is $[0, y_{sup}]$ with

$$y_{sup} = \sigma \quad \frac{e^{-1}}{\max(-\alpha\xi, 0)}\Bigg)^{\frac{1}{2\alpha}}$$

the upper bound in the precipitation domain. Note that when $\xi \geq 0$ the bounds become $x_{sup} = y_{sup} = +\infty$, so those notations can be used for $\xi \in \mathbb{R}$. When the Gaussian is truncated above x_{sup}, the cdf (2.2) must be corrected by the probability of truncation (cf. Appendix B).

Another advantage of the GP meta-Gaussian transformation is the possibility to derive an analytical expression for the inverse of

$$^{-1}(y) = \sqrt{\frac{1}{\alpha\xi}W \quad \alpha\xi\left(\frac{y}{\sigma}\right)^{2\alpha}\Bigg)} \tag{2.10}$$

where W denotes the Lambert W function (Goerg, 2015) defined as the inverse of the function $x \mapsto x \log x$, which is available in usual statistical software. Remark that this is also possible for the classical meta-Gaussian model (2.4). Using (2.2) an analytical expression can be derived for the cdf and the probability density function (pdf). This allows in particular to compute easily the likelihood function and fit the model to data (see Section 2.1.4). Analytical expressions for the finite moments can also be derived, which is not the case for many meta-Gaussian models that can be found in the literature. To our knowledge the classical transform (2.4) is the only other meta-Gaussian model with analytical moments.

Expressions for the GP meta-Gaussian model pdf, cdf, quantile function and moments can be found in Appendix B.

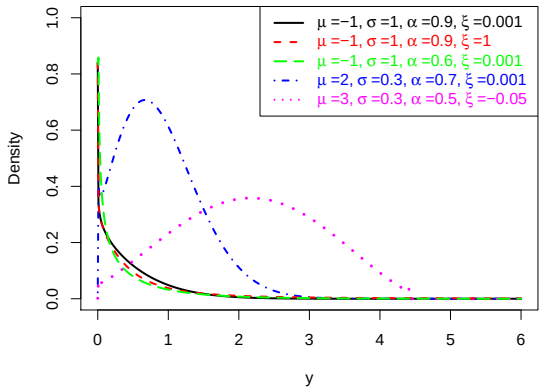


FIGURE 2.6 – Examples of distributions obtained with different parametrizations of the GP meta-Gaussian model.

Extended Generalized Pareto Model

As it was said in the introduction, the model of Naveau et al. (2016) will be a point of comparison and will be referred to as the extended GP model. This paper is the one that was used in Section 2.1.3 to justify the power shape in 0 for the proposed anamorphosis. The specificity of the extended GP is that extreme value theory is applied to the upper tail but also the lower tail, allowing to avoid threshold selection. Note that here the lower tail does not include the dry component of precipitation, and that the model is developed only for strictly positive rainfall amounts Y_+. The general model proposed in Naveau et al. (2016) is defined as

$$Y_+ = \sigma H_\xi^{-1}(U^{1/\alpha}) \tag{2.11}$$

where U follows a standard uniform distribution and H_ξ is the cdf of a GP distribution with shape parameter ξ.

Remark that the transform that is first applied to U, here a power function, is discussed in Naveau et al. (2016) and other functions are tested. In our case the power function - which is the simplest one - was the one to perform the best on the data set presented in Section 2.1.2.

The cdf of Y_+ can be related to the cdf of U in the same way that the cdf of the GP meta-Gaussian model can be related to the Gaussian cdf, hence one can write

$$F_{Y_+}(y) = \{H_\xi(y/\sigma)\} \tag{2.12}$$

2.1.4 Detailed Example of Inference

The previous section introduced three models : first a classical meta-Gaussian model, i.e. (2.1) with anamorphosis (2.4), then the proposed GP meta-Gaussian model, i.e. (2.1) with anamorphosis (2.8) and finally the extended GP model, i.e. (2.11) for strictly positive rainfall.

Those three models will be tested and compared in this section on the data presented in Section 2.1.2.

Inference Method : Dealing with Discretization

With the meta-Gaussian model as written in (2.1) the dry measurements are supposed to be created by the censorship in 0 and hence controlled by μ - the mean of the latent variable. But as the anamorphosis is written in (2.8), it can produce values that are lower than 0.2, the minimal value that can be measured - that will be noted y_m - and when discretizing those values will become zeros. In other words zeros are produced by the censorship (controlled by μ) and also by the anamorphosis (controlled by $\{\sigma, \alpha, \xi\}$). Therefore in order to have separable parameters y_m needs to be introduced in $\psi(x) = y_m + \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2}$. Note the ψ^{-1} and y_{sup} are consequently modified.

A similar reasoning can be applied to the extended GP model : in (2.11) when $U \to 0$ we have $Y_+ \to 0$, when the minimal value that can be observed is 0.2. Hence y_m is also introduced in $Y_+ = y_m + \sigma H_\xi^{-1}(U^{1/\alpha})$.

The introduction of y_m in the models greatly improves the results for all three models : a completely different solution is chosen for the parameters and the fit is better for the whole distribution.

For the meta-Gaussian models, the likelihood is usually computed directly from the continuous density (see Appendix B), however it has been noticed that taking into account discretization significantly improves the results. It is valid for the data considered in this study, i.e. rain gauge measurements with a bucket that tips every 0.2 mm, but more surprisingly a significant improvement was also observed when working with radar data with a 0.01 mm precision for the same area (see Figure 2.11 and discussion in Conclusion).

The discrete likelihood is based on the functioning of a tipping bucket rain gauge, which means that $P(G = g) = P(g \leq Y < g + step)$, G being the discrete measurement, Y being the continuous rainfall and *step* being the precision of G, i.e. 0.2 mm in our case. Therefore the discrete log likelihood for both meta-Gaussian models is based on the cdf

(2.2) :

$$\log \mathcal{L}_\theta^{MG} = n_0 \log(\Phi(-\mu)) + \sum_{i:\ y_i>0} \log\left\{\Phi(\psi^{-1}(y_i + step) - \mu) - \Phi(\psi^{-1}(y_i) - \mu)\right\}, \quad (2.13)$$

where n_0 is the number of dry measurements. For the classical meta-Gaussian ψ is given in (2.4) and $\theta = \{\mu, \sigma, \alpha\}$, and for the GP meta-Gaussian model, ψ is given in (2.8) and $\theta = \{\mu, \sigma, \alpha, \xi\}$.

The discrete log likelihood of the extended GP model is based on its cdf (2.12) in the same way :

$$\log \mathcal{L}_\theta^{extGP} = \sum_{i:\ y_i>0} \log\left\{(H_\xi[(y_i + step)/\sigma])^\alpha - (H_\xi[y_i/\sigma])^\alpha\right\},$$

with $\theta = \{\sigma, \alpha, \xi\}$.

The moments could also be used to infer the parameters of the GP meta-Gaussian (see Appendix B). Remark however that the usual method of moment is tricky to implement when working with heavy tail distributions since some moments are infinite. Furthermore it may also be sensitive to the discretization of the data.

Results

Figure 2.7 shows the results obtained with of the three models adjusted to 6 minutes rain gauge data with discrete likelihood. First of all a focus is made on the proposed GP meta-Gaussian model, hence the second column. The global fit of the model, observable in the quantile-quantile (QQ) plot on the bottom, is very satisfying. The density (top) shows a quasi perfect fit for low intensities. For medium intensities (4-6 mm) a slight deviation can be observed on the QQ plot, the model producing more of these values that what is present in the data. Remark that the GP meta-Gaussian model was tested on many data sets and it was observed that it is not a recurrent issue. Finally the tail of the distribution is very well reproduced by the model, and is quite heavy as the tail parameter is 0.45. It means that moments of order greater than 2.2 are infinite. Hence even when simply computing the variance of some precipitation series at a fine scale (minutes), one must remember that the variance of the estimator is likely to be infinite.

Those results can also be found in Figure 2.5, where the plain curve represents the estimated transformation function ψ, and the dots are the empirical one. ψ is almost perfectly estimated for low intensities and the good reproduction of extremes is satisfying.

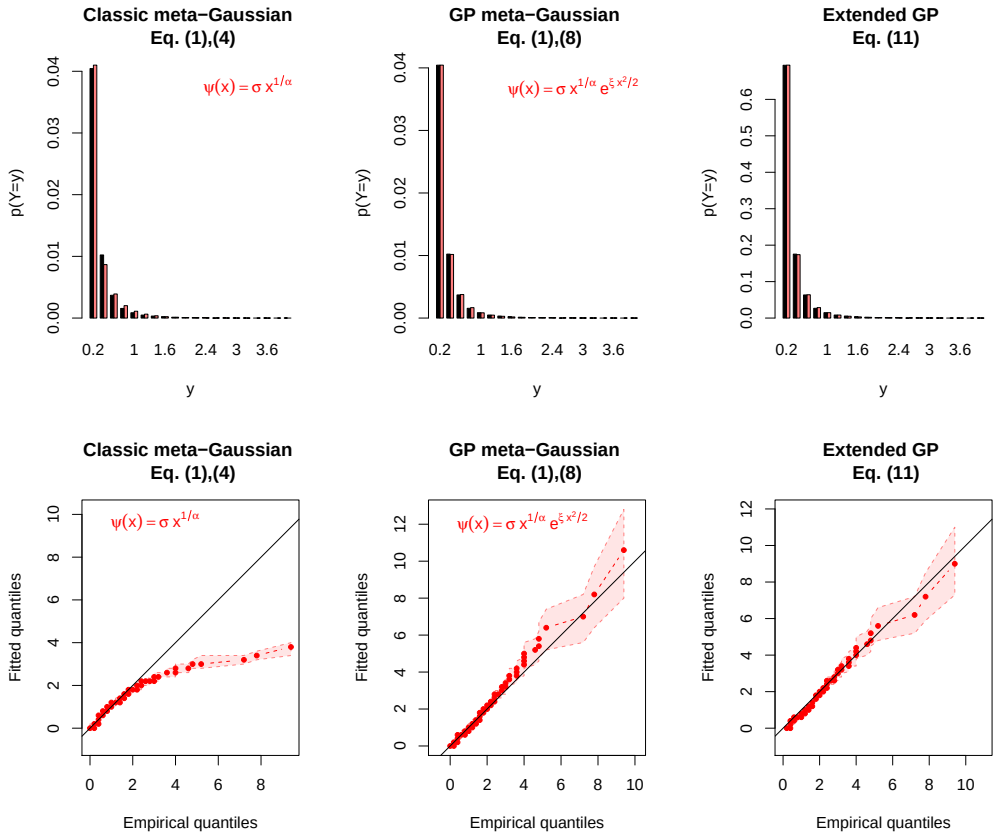


FIGURE 2.7 – Fitting the three models to the whole series. 1st row : density of low intensities (≤ 4 mm), empirical in black, model in red. 2nd row : quantile-quantile plot of the full distribution, empirical versus fitted quantile. The light area gives the 95% intervals computed with 500 non parametric bootstrap replicates.

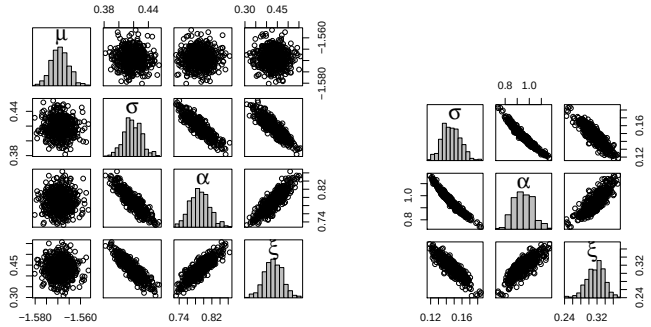


FIGURE 2.8 – Correlation between parameters, computed with 500 non parametric bootstrap replicates. GP meta-Gaussian on the left, extended GP on the right. Histograms of the parameters are shown in the diagonal.

Figure 2.8 (left) shows the correlation between the parameters of the GP meta-Gaussian model. μ, that controls the dry component of the distribution, is completely uncorrelated with the parameters of the transformation, which is satisfying. However there is a quite strong correlation between the other parameters. It will be an important point to keep that in mind when trying to interpret the parameters.

As expected the classical meta-Gaussian model is unable to reproduce the tail of the distribution (Fig. 2.7, first column). The lower tail is also affected, and even though it is not too bad the other models do better even for low intensities. The parameters are completely uncorrelated for this model, but the poor fit explains why the classical meta-Gaussian model will not be further discussed in the following section.

Finally the extended GP and the GP meta-Gaussian models give very similar results in terms of goodness of fit (Fig. 2.7, second and third column). The parameters of the extended GP model seem to be slightly more correlated than the ones obtained for the proposed model (Fig. 2.8, right). It was also noted that the parameters were quite close in terms of values, which will be investigated in the next section.

Note that meta-Gaussian model with transform (2.5) was also tested : it gave a satisfying fit, as it only unhooked in the very end of the tail, but the quality of fit was not the main problem with this model. In Section 2.1.3 it was already mentioned that the low and strong intensities are controlled by the same parameter, and the correlation between parameters showed a quasi direct link between the three parameters of the transform.

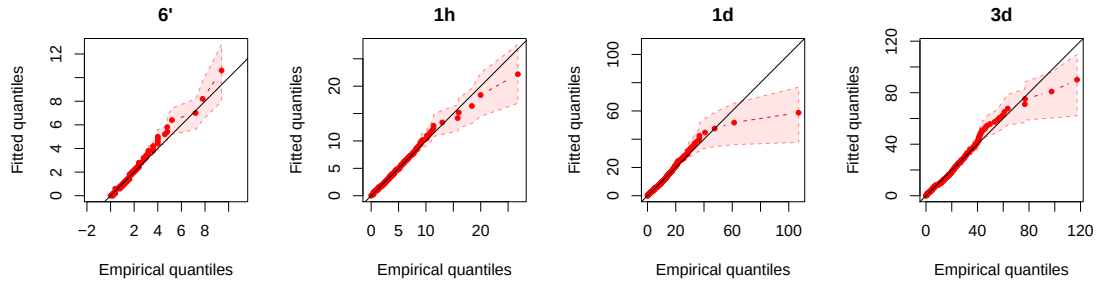


FIGURE 2.9 – Fitting the GP meta-Gaussian distribution at various time lags (6 minutes, 1 hour, 1 day and 3 days). The light area gives the 95% intervals computed with 500 non parametric bootstrap replicates.

Time Lag

Accumulating precipitation data over various periods of time allows exploring the physical meaning of the parameters - if they have one - and checking the flexibility of the model. As it was said through the paper, the GP meta-Gaussian distribution aims at modelling precipitation at a wide range of time scales, from sub-hourly to daily rainfall. The extended GP model will also be used in this section, and as it was already mentioned the classical meta-Gaussian model will not be further discussed.

The GP meta-Gaussian model fitted for data ranging from 6 minutes up to daily scale gave very satisfying results (see the QQ plots in Figure 2.9), which demonstrates that the model is flexible enough to reproduce the distribution of rainfall at a wide range of time lags. Remark that the extended GP model is not shown in Figure 2.9 but the QQ plots obtained were so similar that they were almost indistinguishable.

Figure 2.10 shows the evolution of the model parameters with time aggregation (note that the time axis is non linear). The first thing to notice is the fact that the evolution of the parameters is smooth and monotone. Note that with other meta-Gaussian models such smoothness and monotony was not observed.

μ is increasing with aggregation, which is expected as there are less and less dry measurements. σ is the global scale parameter, hence it is expected to increase as well. The tail parameter ξ is decreasing which is coherent with the intuition that averaging random variable will tend to "gaussianize" them and produce distributions with lighter tails. Plus practical knowledge of precipitation in the considered region tells us that with a larger time steps there are less extreme values.

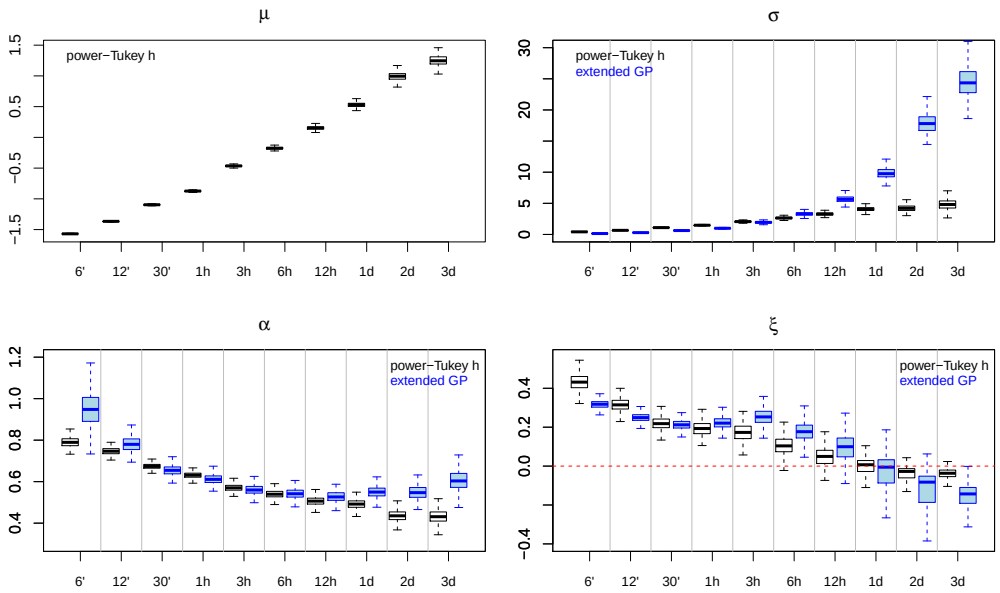


FIGURE 2.10 – Parameter estimation at various time lags, boxplots computed on 500 non parametric bootstrap replicates. The empty boxes are parameters obtained with the GP meta-Gaussian distribution and the fulled ones with the extended GP.

As α decreases the distribution becomes more "peaky" at the origin. The reason why α decreases is not straightforward, but the rainfall accumulated over a given time period is the sum of a random number (because of the dry measurements) of correlated (because of the temporal dependence) random variables and hence it may have a complicated behaviour. Maybe α controls not only the lower tail but has also a strong impact on the medium intensities. Another lead is that it might be due to the strong correlation between the parameters (Fig. 2.8). This reminds us that caution is needed and one should not over interpret the parameters.

As the GP meta-Gaussian model and the extended GP one are closely related, it is also interesting to check if the parameters have the same role in both models, and hence the same evolution with time aggregation. Figure 2.10 shows the parameter estimation of the extended GP is shown in blue. From 6 minutes to 1 day the parameters of both models are similar and have the same variation. After 1 day they diverge to different solutions, and it could be explained by the fact that after one day a non-zero mode starts to emerge on the histogram of the data. However both models still have the same quality of fit.

2.1.5 Conclusion

The goal of the GP meta-Gaussian distribution is to extend the class of meta-Gaussian models to small time steps - several minutes. Properties of the lower and upper tails motivates the choice of the transformation, using extreme value theory to derive two conditions for the anamorphosis. The proposed GP meta-Gaussian model is tractable and analytical expressions exist for the pdf, cdf, quantile function and for the moments.

Results are very satisfying for a wide range of time steps - from 6 minutes to several days, demonstrating the flexibility of the model. Even though the data presented in this article did not allowed to aggregate further (only 12 years of data), the model was also tested up to monthly scale and gave similarly good results.

Comparison with a classical meta-Gaussian model shows what the proposed transform brings to this class of models : a better fit at small time scale due to its capacity to produce heavy tails. The GP meta-Gaussian model is quite similar to the extended GP model (Naveau et al., 2016) in terms of construction but also in terms of performance. The advantage of the meta-Gaussian model is its link with the latent Gaussian that allows using methods developed for Gaussian data (multivariate, spatiotemporal models, Kalman-like algorithm, etc.).

The evolution of parameters with aggregation is very interesting for several reasons. First it demonstrates the similarities between the GP meta-Gaussian and the extended GP model. Second the smooth and monotonic evolution makes it possible to interpret the roles of the parameters, even though α is a bit harder to interpret than the other parameters. Finally when seeing such smoothness in the variation one can wonder if a unique model with parameters varying with time aggregation could be developed. Having such model would be very interesting as it would mean that hourly or daily rainfall could say something about the parameters values at a few minutes time scale.

A final point that was not central in the paper but that is very important for practical applications is the role of discretization. To demonstrate the performance of the discrete likelihood against the continuous one, Figure 2.11 shows the results of the optimisation on simulated data with various discretization precisions and various "time scales", i.e. various parameter sets that were the ones found in Section 2.1.4. What is shown is the 0.95 quantile of the absolute error between the true parameters and the ones estimated by the two likelihoods. Both likelihoods could be used for data with 0.01 precision but for higher discretization steps the discrete likelihood performs way better than the continuous

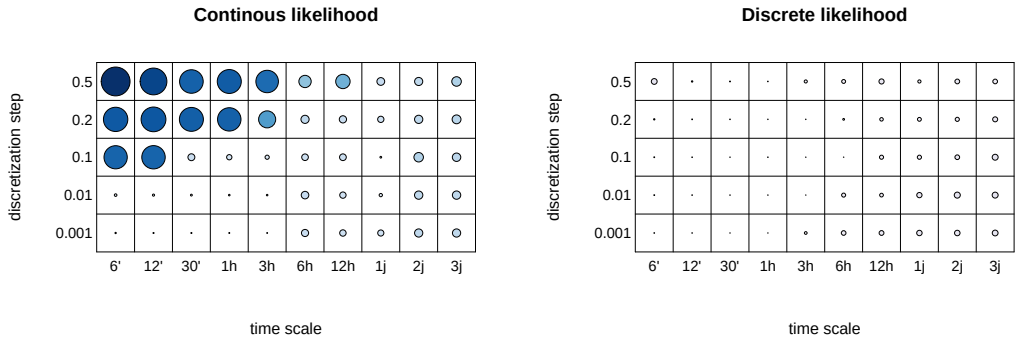


FIGURE 2.11 – Quantile 0.95 of parameters error between the truth and both likelihood estimates (the size and colour of the dots are on a log scale).

one, even with hourly data. Note that most of the parameter error is due to α and ξ. For the models that were tested in this paper, the discrete likelihood was a efficient and easy way to deal with the discretization issues.

2.2 Autres tests

Dans l'article nous nous sommes concentrés sur la variation des paramètres du modèle avec l'agrégation temporelle, et pour ça les données du pluviomètre de Météo France situé à Guipavas ont été utilisées. Comme ça a été présenté dans le Chapitre 1 nous avons à notre disposition d'autres données, notamment les 11 pluviomètres d'EDP ou les données radar qui permettent d'étudier la variation spatiale des paramètres. Les 2 types de mesures (pluviomètres et radar) permettront aussi de tester différentes discrétisation, et enfin il sera aussi intéressant d'étendre l'analyse à l'année entière afin d 'étudier l a variabilité saisonnière des paramètres.

2.2.1 Structure spatiale

GP méta-Gaussien

Pour rappel le modèle GP méta-Gaussien consiste à décrire la pluie à partir d'une variable latente Gaussienne à travers (2.1), et à utiliser comme anamorphose (2.8). Ce modèle a été ajusté par maximum de vraisemblance en utilisant la vraisemblance discrétisée (2.13). Pour les pluviomètres on a $y_m = step = 0.2$ et pour le radar on a $y_m = 0.02$ et $step = 0.01$.

Pour étudier la structure spatiale des paramètres, il a été choisi de se concentrer en premier lieu sur deux mois de l'année : janvier et août. Il est attendu que ces deux mois auront des ajustements très différents, avec en janvier beaucoup plus de pluie, notamment des évènements stratiformes, et en août des évènements convectifs orageux plus intenses et concentrés spatialement.

La structure spatiale de la pluie est mieux représentée dans les données radar, aussi les ajustements ont été faits sur le radar à 5 minutes, et les paramètres obtenus sont présentés en Figure 2.12. Comme ça a été évoqué au Chapitre 1, quelques pixels du radar proches de la station de Plabennec présentent beaucoup de données manquantes et ont souvent des valeurs qui apparaissent étranges sur les images radar. On retrouve facilement ces pixels sur les cartes de la Figure 2.12, il est en effet attendu que ces erreurs / données manquantes aient un impact sur l'ajustement du modèle.

La première chose qui saute aux yeux et qu'on attendait est la différence entre les paramètres μ du mois de janvier et d'août. Il y a bien beaucoup plus de temps sec (μ très négatif) en août qu'en janvier. Spatialement, μ est réparti de la même façon sur les deux mois, on trouve une occurrence de pluie plus forte dans les terres, avec une variation spatiale qui suit le trait de côte.

Le paramètre de queue ξ est aussi très différent entre les deux mois. En janvier on a une carte assez lisse avec des valeurs plutôt faibles, ce qui correspond à l'idée qu'en hiver la plupart des évènements sont longs, aussi la distribution est moins vide entre la masse et le maximum. En août les queues de distribution sont plus lourdes, surtout dans les terres. On retrouve l'effet du trait de côte qui peut s'expliquer par la complexité de la météorologie zone côtière. Enfin la carte de ξ en août est plutôt irrégulière, en effet les orages très intenses étant rares et spatialement peu étendu toute la zone d'étude n'est pas touchée également par les orages sur seulement 8 ans de données. On note que les valeurs de ξ peuvent facilement dépasser 1, ce qui signifierait que l'espérance des données est infinie. On peut s'interroger sur le réalisme de ce résultat.

Les deux autres paramètres σ et α sont plus difficiles à interpréter. Tout d'abord on note (surtout pour août) la corrélation entre α et ξ. Les variations spatiales des paramètres ne semblent pas évidemment liées à des statistiques descriptives des données (Fig 1.21 page 48). On peut toutefois noter qu'en janvier σ est fort quand la probabilité d'avoir une mesure en dessous de 0.2 est faible et que le cumul mensuel est fort. Autrement ça correspondrait à des endroits avec peu de faibles intensités mais beaucoup d'intensités

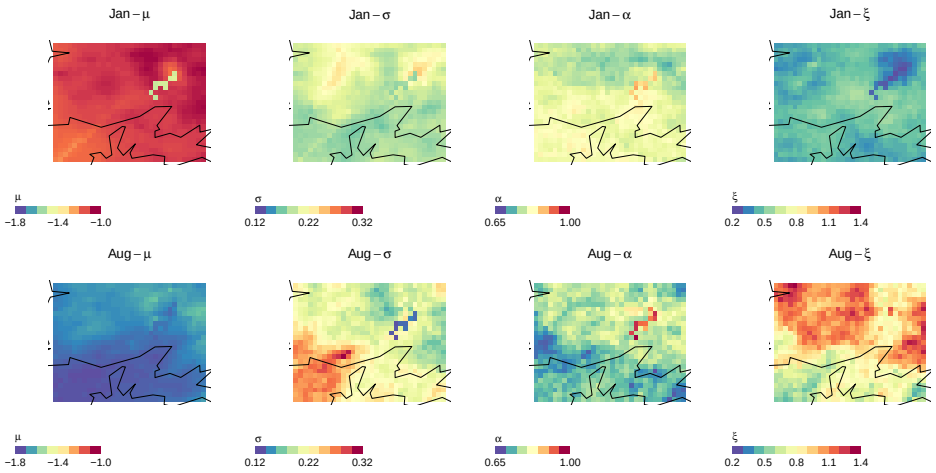


FIGURE 2.12 – Carte des estimations des paramètres du modèle GP méta-Gaussien pour janvier et août sur les données de radar à 5 minutes.

moyennes (ξ est plutôt faible).

En ajustant le modèle de la même façon sur les 13 pluviomètres du réseau d'Eau du Ponant (des données à 3 minutes donc), on obtient la Figure 2.13. On se concentre cette fois-ci sur janvier. En comparant avec les résultats du radar tout d'abord on retrouve certaines variations dues au pas de temps qui ont été mises en évidence en Section 2.1.4 : μ est plus faible car plus de temps sec et ξ est plus élevé. Toutefois sur les deux autres paramètres on ne retrouve pas vraiment ces variations.

Tout d'abord on peut noter le cas du pluviomètre de Gouesnou (point le plus au Nord), qui semble trouver un ajustement très différent des autres pluviomètres : ξ est négatif et on a beaucoup plus de temps sec que dans les autres pluviomètres. La qualité de l'ajustement étant bonne, on en conclut que la différence vient des données en elle-mêmes. En effet le pluviomètre de Gouesnou est arrêté assez vite (2014), on a donc seulement 4 ans de données, et il a été noté que vers la fin de son fonctionnement le pluviomètre semble mesurer des cumuls de pluie beaucoup plus faibles que ses voisins.

La deuxième chose qui saute aux yeux est que les variations spatiales des paramètres sont beaucoup moins lisses avec les pluviomètres qu'avec le radar. Deux pluviomètres proches n'ont pas forcément les mêmes jeux de paramètres. Les erreurs de mesure des pluviomètres sont a priori indépendantes, contrairement au pixels du radar où la correction (notamment l'utilisation du champ d'advection 2PIR) crée nécessairement une cohérence spatiale. On peut émettre l'hypothèse que l'installation des pluviomètres (en haut d'un immeuble versus dans un cimetière) peut jouer sur la distribution observée. On peut aussi

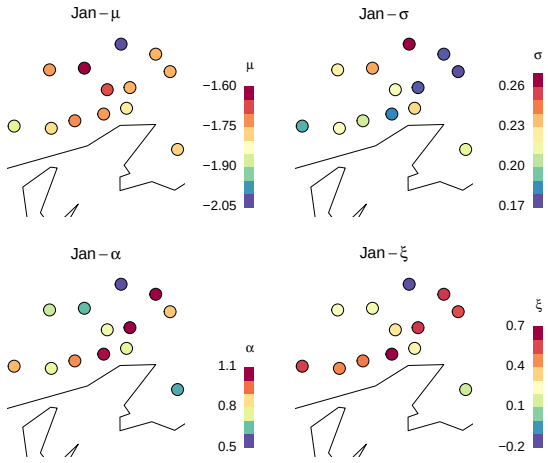


FIGURE 2.13 – Paramètres du modèle GP méta-Gaussien estimés en janvier sur les données des pluviomètres à 3 minutes.

penser à la forte discrétisation (0.2 mm) qui pourrait entraîner une plus grande sensibilité aux fortes valeurs (pour les faibles intensités il n'y a que quelques points à ajuster).

On retrouve de nouveau la forte corrélation entre les estimateurs, ce qui permet de rappeler qu'il ne faut pas sur-interpréter la valeur des paramètres.

Autres transformations

Ici on se concentre sur les données radar (car ce sont celles qui donnent une meilleure structure spatiale), et sur le mois de janvier. On souhaite tester les autres modèles méta-Gaussiens évoqués :

— Le modèle **power** avec la transformation (2.5), qui est le cas particulier $\xi = 0$ du modèle GP meta-Gaussien et

— le modèle **power-exp** avec la transformation (2.4).

Les paramètres ajustés sont montrés en Figure 2.15. Les cartes de paramètres sont globalement aussi lisses qu'avec le modèle GP méta-Gaussien, même si pour la transformation power (qui correspond au cas particulier $\xi = 0$) quelques pixels ressortent. On constate aussi que μ donne les mêmes résultats quelques soient la transformation, ce qui est attendu puisque ce paramètre à toujours le même rôle et on a montré en Figure 2.8 page 68 qu'il n'était pas ou peu corrélé aux autres paramètres. Pour compléter cette figure sur la dépendance entre les paramètres des modèles meta-Gaussien, la Figure 2.14 montre la corrélation entre les estimateurs des paramètres du modèle power-exp.

Sur les cartes les paramètres de la transformation power apparaissent de nouveau très

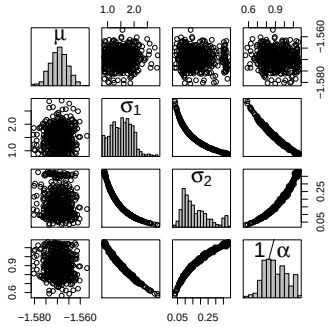


FIGURE 2.14 – Corrélation entre les paramètres du modèle méta-Gaussien power-exp, calculée sur 300 échantillons bootstrap non paramétriques

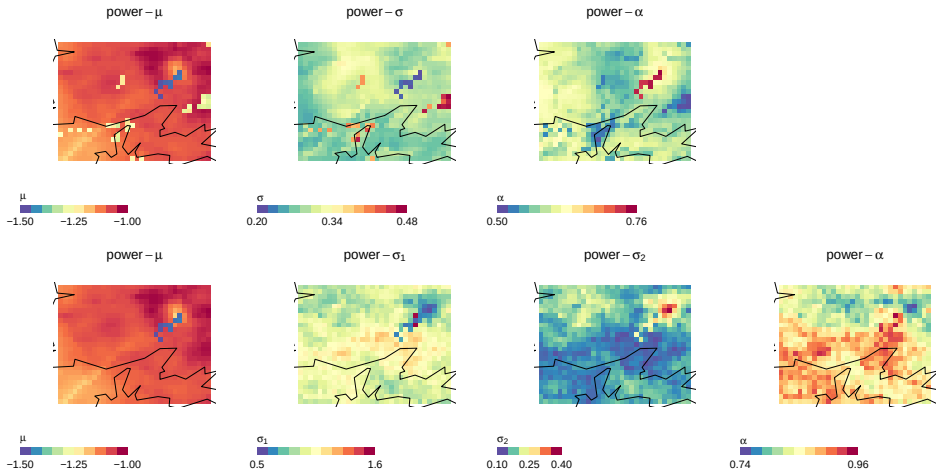


FIGURE 2.15 – Carte des paramètres des modèles méta-Gaussiens « power » et « power-exp » estimés en janvier sur les données radar à 5 minutes.

peu corrélés, alors que pour la transformation power-exp on voit bien l'impact du lien quasi direct entre σ_1 et σ_2.

Une estimation de la qualité d'ajustement a été faite à travers la distance d'Anderson Darling (Anderson & Darling, 1952), qu'on peut écrire

$$\sum_{y \in A} \frac{(F_{emp}(y) - F(y))^2}{F(y)(1 - F(y))}$$

où F_{emp} est la cdf empirique des observations Y et F est la cdf théorique (modèle). y parcourt le support des deux distributions, aussi dans notre cas il a été choisi de prendre $A = \{0, y_m, y_m + step, y_m + 2 * step, ..., \max(F^{-1}(99.99\%), \max(Y))\}$. Cette distance est

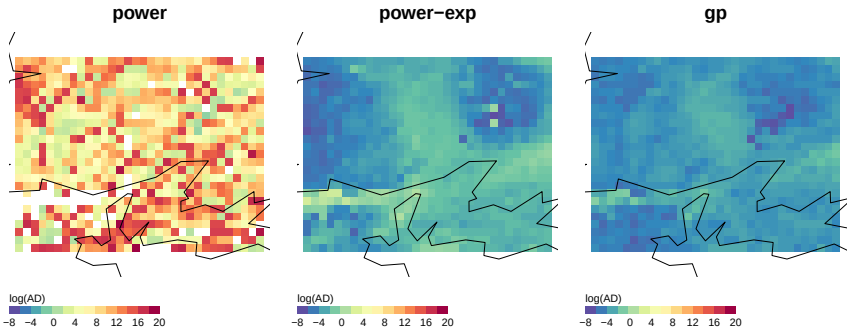


FIGURE 2.16 – Distance d'Anderson Darling pour les 3 modèles méta-Gaussien ajustés sur les données radar à 5 minutes en janvier (échelle logarithmique).

souvent utilisée pour la précipitation (Arshad et al., 2003 ; Khudri & Sadia, 2013, e.g.), notamment dans le cadre des valeurs extrêmes car elle donne plus de poids aux queues de la distribution que les distances comme Cramer Von Mises ou Kolmogorov–Smirnov.

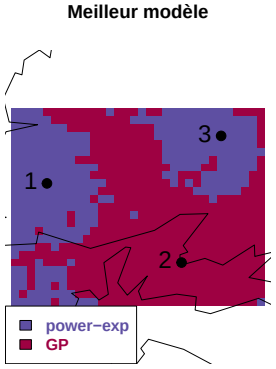
Les résultats de la distance d'Anderson Darling sont montrés en Figure 2.16 (échelle logarithmique). La Figure 2.17a montre pour chaque pixel quel est le meilleur modèle du point de vue de l'AIC, et enfin la Figure 2.17 montre pour trois points du radar un QQ plot permettant de comparer la qualité d'ajustement sur les trois modèles.

Tout d'abord on peut remarquer que tous les modèles reproduisent correctement les faibles intensités, inférieures à 1 mm/5 min (Fig. 2.17b, 2.17c, 2.17d).

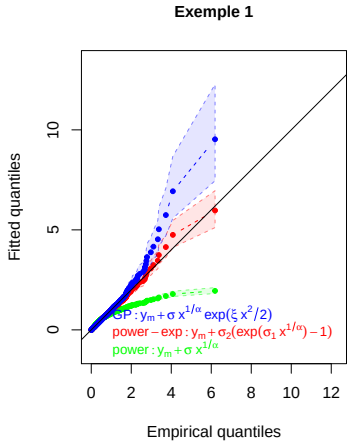
De la même façon que ce qu'on avait pu voir sur les données du pluviomètre de Météo France, la simple transformation puissance n'est pas adaptée aux données à un pas de temps de quelques minutes. La distance d'Anderson Darling (Fig. 2.16) est largement plus élevée, et le QQ plot (Fig. 2.17b, 2.17c, 2.17d) montre que le modèle power « décroche » très vite et ne parvient pas à créer suffisamment de fortes valeurs.

Les modèles power-exp et GP donnent des résultats assez proches : le QQ plot (Fig. 2.17c) montre que le modèle power-exp est un peu moins bon sur les intensités moyennes à fortes pour le point pris en exemple. Il y a des zones où la distance d'Anderson Darling (Fig. 2.16) est plus forte avec le modèle power-exp, notamment sur la presqu'île de Plougastel et au-dessus de la Penfeld. A ces endroits ξ est assez fort (Fig. 2.12) et en vérifiant les ajustements des QQ plots on a pu constater que les 2 modèles avaient du mal à reproduire la queue de la distribution dans ces zones (Fig 2.17c).

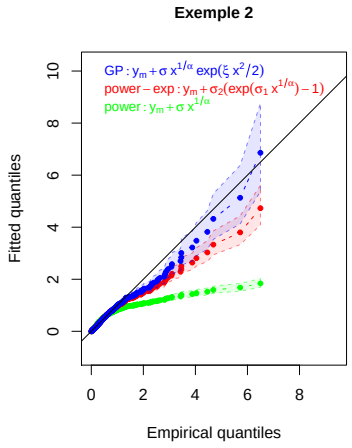
La transformation power-exp marche un peu mieux que GP dans deux zones : 1. près du centre radar et 2. à l'ouest de la zone d'étude, ce qui correspond à des endroits où ξ est relativement faible pour le modèle GP. Une explication possible est que le modèle



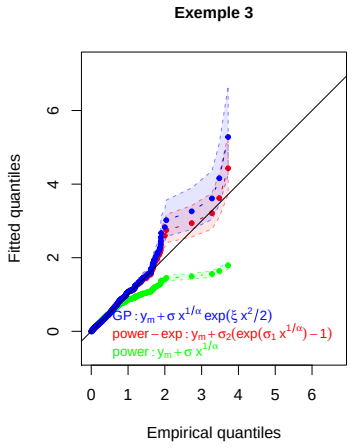
(a) Meilleur modèle méta-Gaussien du point de vue de l'AIC.



(b) QQ plot des 3 modèles au point 1.



(c) QQ plot des 3 modèles au point 2.



(d) QQ plot des 3 modèles au point 3.

FIGURE 2.17 – Comparaison de la qualité d'ajustement de trois modèles méta-Gaussien sur les données radar à 5 minutes. Les intervalles sont calculés sur 300 échantillons bootstrap non paramétriques.

power-exp a moins de difficultés à reproduire les queues comme attendu par la théorie (cf. commentaire dans le papier : si $\xi = 0$, la queue supérieure n'impose pas de contrainte sur la valeur de α).

Ces zones sont cohérentes avec le choix du meilleur modèle par l'AIC (Fig. 2.17a). Le modèle power-exp est meilleur à l'ouest et dans le quart nord-est autour de Guipavas. Le modèle GP proposé est meilleur partout ailleurs.

2.2.2 Variations saisonnières

Les Sections 2.2.1 et 2.1 ayant montré l'intérêt du modèle GP méta-Gaussien proposé par rapport aux autres transformations, pour cette section il a été choisi d'abandonner les modèles méta-Gaussiens power et power-exp.

Après avoir étudié les variations spatiales des paramètres, il est intéressant de passer aux variations temporelles et plus précisément saisonnières. On a déjà pu commencer à voir les différences été/hiver avec la Figure 2.12 page 74, ici on va s'intéresser à tous les mois. Les résultats des paramètres estimés par le modèle GP méta-Gaussien sur le radar à 5 minutes sont donnés pour chaque mois en Figure 2.18.

μ, le paramètre qui définit la probabilité de temps sec, est probablement le plus facile à interpréter. La variation sur l'année est très lisse, avec peu de temps sec en hiver, beaucoup plus en été et des transitions lisses en automne et printemps. Le mois le plus sec semble être juillet ($\mu = -1.8$ donc probabilité de pluie de 4%) et le plus humide janvier (probabilité de pluie de 13%).

ξ montre la présence d'orages en été avec une distribution avec une queue très lourde sur les mois de juin à septembre. Les mois d'hiver, plutôt caractérisés par des fronts stratiformes plus étendus dans le temps et l'espace donnent des valeurs de ξ plus faibles associées à une probabilité de pluie plus forte.

L'étendue spatiale des évènements en été versus en hiver se retrouve dans le fait que pour les quatre paramètres, les cartes sont plus lisses en hiver et plus irrégulières en été.

Comme d'habitude σ et α sont plus difficiles à interpréter. Les variations sont moins lisses sur l'année (cf. mai juin juillet pour σ), et les cartes sont assez irrégulières. Aucun lien évident avec les statistiques descriptives présentées en Figure 1.21 page 48 n'a été trouvé.

Le modèle GP méta-Gaussien a aussi été ajusté mois par mois sur les données des pluviomètres à trois minutes, les résultats sont présentés en Figure 2.19. Les conclusions

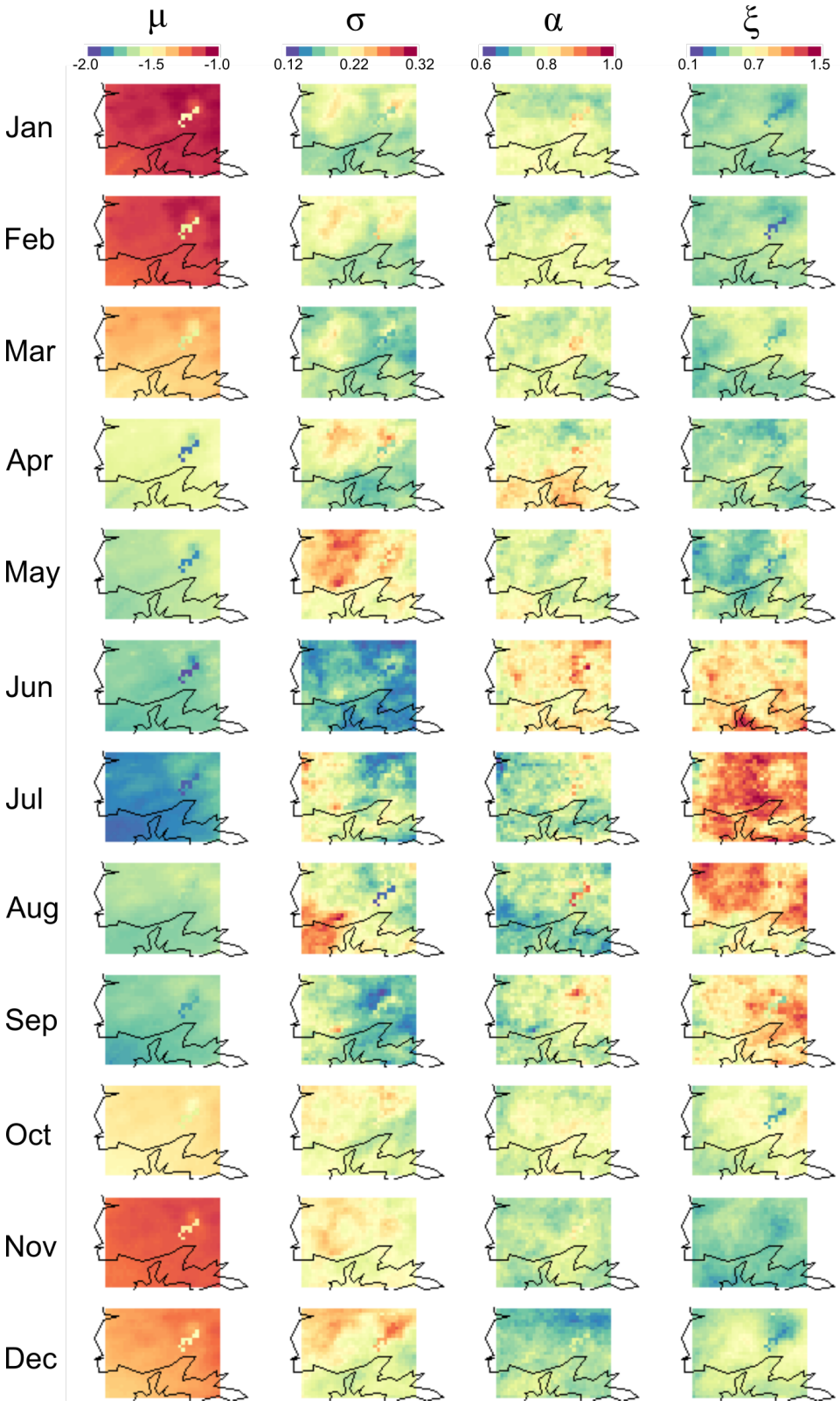


FIGURE 2.18 – Carte des paramètres du modèle GP méta-Gaussien estimés pour chaque mois sur les données de radar à 5 minutes.

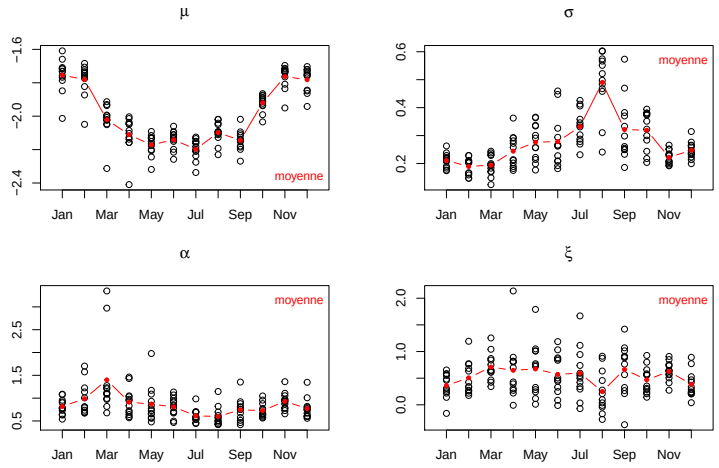


FIGURE 2.19 – Paramètres du modèle GP méta-Gaussien estimés par mois sur les données des pluviomètres à 3 minutes.

sont similaires pour μ mais ça s'arrête là. La variabilité annuelle pour les trois autres paramètres est plutôt faible comparée à la variabilité spatiale. On ne retrouve pas du tout l'augmentation de ξ en été, ce qui est assez surprenant, à la place on a plutôt une augmentation de σ qui n'est pas aussi présente dans les données radar. On rappelle qu'il y a une corrélation négative entre les estimateurs de ξ et σ qui peut être un explication. α varie assez peu, ce qui est cohérent avec les résultats sur les données radar.

Afin de pouvoir mieux comparer les données radar et pluviomètres, les ajustements du modèle GP méta-Gaussien ont été fait sur les données agrégées à 15 minutes (le plus petit pas de temps commun des deux mesures). Les paramètres ajustés aux points des pluviomètres sont montrés en Figure 2.20, on constate que la dispersion des pluviomètres est beaucoup plus forte que celle du radar et que les comportements de σ et ξ sur les mois d'été sont très différents. Les orages sont donc gérés différemment : par une augmentation de l'échelle pour les pluviomètres, et par une augmentation de la queue pour le radar.

On remarque aussi que si les variations de μ sont les mêmes pour les deux sources de données, le paramètre d'occurrence de la pluie est toujours plus élevé dans les données radar. Il y a donc moins de temps sec dans les données radar, ce qui est cohérent avec le fait que dans les données des pluviomètres, les zéros correspondent en fait à des intensités inférieures à 0.2 mm.

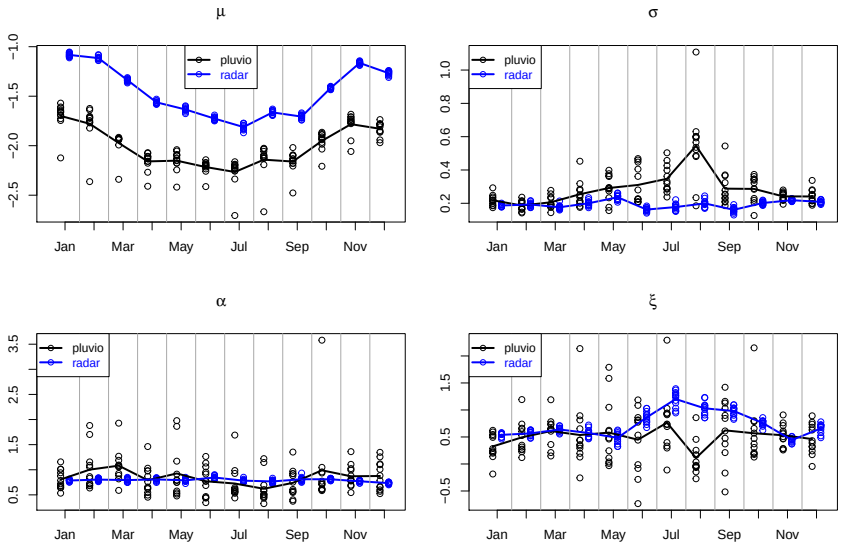


FIGURE 2.20 – Paramètres du modèle GP méta-Gaussien estimés sur les données des pluviomètres et du radar à 15 minutes.

2.3 Conclusion

Dans ce chapitre un nouveau modèle pour la distribution marginale de la précipitation a été développé. Il fait partie de la classe des modèles méta-Gaussiens, qui ont l'avantage de lier la pluie à une variable latente Gaussienne, ce qui permet d'utiliser toutes les méthodes statistiques développées pour les lois normales. On pense notamment aux méthodes d'assimilation de données, qui peuvent aider à corriger les différentes sources de données dont on dispose en prenant en compte leurs erreurs.

La particularité du modèle GP méta-Gaussien développé est la possibilité de générer des lois à queues lourdes, ainsi que sa versatilité qui lui permet de s'ajuster à des données de l'échelle mensuelle jusqu'à quelques minutes. Il a aussi été montré que le modèle était adapté aux pluviomètres et aux données radar, la discrétisation étant prise en compte dans la vraisemblance.

Un autre point qui a été beaucoup été abordé dans ce chapitre est l'interprétabilité des paramètres du modèle. Un réflexion théorique sur les propriétés des précipitations a mené à la définition d'un modèle dont chaque paramètre est relié à un partie différente de la distribution. Toutefois les estimateurs des paramètres présentent une forte dépendance (qui a été retrouvée les autres modèles qui ont pu être testés), ce qui complique l'interprétation des paramètres. On a quand même pu expliquer en partie la variation des

paramètres (surtout le paramètre de temps sec et le paramètre de queue) avec la saison, l'aspect spatial et le pas de temps des données.

Il a été constaté que la variation des paramètres avec l'agrégation temporelle était monotone et suffisamment lisse pour qu'on puisse envisager d'estimer par exemple les paramètres de la loi à quelques minutes à partir de données de pluie à l'échelle horaire. On pourrait donc avoir des informations sur la probabilité de pluie ou la lourdeur de la queue (donc les évènements extrêmes), ce qui peut s'avérer très intéressant dans des contextes applicatifs.

Chapitre 3

Extension du modèle méta-Gaussien au cas multivarié

3.1 Introduction

Ce chapitre cherche à étendre le modèle GP méta-Gaussien proposé au Chapitre 2 au cas multivarié. Ce modèle fait partie de la classe des modèles méta-Gaussiens qu'on définit par

$$Y_i = 0 \times \mathbb{1}_{X_i<0} + \psi(X_i) \times \mathbb{1}_{X_i\geq 0}, \quad \text{avec } X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

où $\mathbb{1}_A$ est la fonction indicatrice égale à 1 si la condition A est vraie, et égale à zéro sinon. Les n variables de pluie sont notées $Y = \{Y_i\}_{i\in 1:n}$, X est un champ Gaussien de vecteur de moyennes $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$, et enfin l'anamorphose est $: [0, +\infty[\to [0, +\infty[$, une fonction croissante. Dans le cas du modèle GP méta-Gaussien l'anamorphose proposée est

$$(x) = y_m + \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2},$$

où y_m est la valeur minimale qui peut être mesurée. Chaque paramètre est lié à une partie de la distribution : μ décrit l'occurrence de la pluie, α est lié à la forme de la distribution proche de zéro, ξ contrôle la queue de la distribution et σ est un paramètre d'échelle.

Il y a deux sets de paramètres à estimer dans le cas multivarié : les paramètres des anamorphoses et les paramètres de la covariance du champ Gaussien. La littérature donne plusieurs façons de procéder, dont les deux principales sont citées ci-après.

1. EMV : Estimation du maximum de vraisemblance avec tous les paramètres.
2. Inference functions for margins (IFM) (e.g. Allard & Bourotte, 2015 ; Joe & Xu, 1996) :
   (a) Ajuster les paramètres des marges,

(b) envoyer les données dans le domaine Gaussien avec ces paramètres,

(c) estimer les paramètres de dépendance sur le champ Gaussien censuré et

(d) éventuellement ré-estimer les paramètres marginaux en prenant en compte la dépendance.

3.2 Estimation de tous les paramètres (EMV)

De la même façon qu'on a utilisé la discrétisation des données pour établir une vraisemblance discrète au Chapitre 2, on va définir G le champ de pluie discrétisé par $\mathbb{P}(G = g) = \mathbb{P}(g \leq Y < g + step)$. Pour chaque pas de temps on peut séparer le vecteur des observations en $g = (g_S, g_P)$ avec d'un coté les variables qui observent du temps sec et de l'autre celles qui observent du temps de pluie. La probabilité d'une observation s'écrit alors

$$\begin{aligned} \mathbb{P}(G = g) &= \mathbb{P}(g_S, g_P) \\ &= \mathbb{P}(x_S \leq 0, \psi^{-1}(g_P) \leq x_P < \psi^{-1}(g_P + step)) \end{aligned} \tag{3.1}$$

où x_S et x_P sont les observations du champ Gaussien associées à g_S et g_P respectivement. La fonction de répartition et la fonction de survie d'un vecteur gaussien peuvent s'exprimer en fonction de probabilités d'Orthant et peuvent être estimée numériquement. Le package R `mvtnorm` utilise par défaut l'algorithme décrit dans Genz (1992). Ainsi on peut calculer la log vraisemblance discrète à partir de (3.1).

Une vraisemblance peut aussi être écrite dans le cas continu, cette option est développée en Annexe D.

3.3 Estimation séparée de la dépendance et des marges (IFM)

3.3.1 Estimation de la dépendance avec la méthode des moments

Soit un couple gaussien $X = (X_1, X_2) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$ et les couples censurés et tronqués à gauche en h associés $X^+ = (X_1^+, X_2^+), X^{+*} = (X_1^{+*}, X_2^{+*})$ définis par

$$\begin{aligned} x_i^+ =& x_i * \mathbb{I}_{x_i \geq h} + h * \mathbb{I}_{x_i < h} \\ X_i^{+*} =& \{x_i\}_{(i,j):(x_i,x_j)\geq(h,h)} \end{aligned}$$

où $(i,j) \in \{1,2\}^2$. Dans le cas tronqué les variables ne sont observées que si elles sont toutes les deux supérieures à h, et dans le cas censuré les valeurs inférieures à h deviennent h. La Figure 3.1 montre plusieurs exemples de couples Gaussiens censurés et tronqués avec différentes valeurs de dépendance. On constate qu'on n'a plus beaucoup d'observations dans le cas tronqué quand la corrélation est très négative et que h est très élevé (Fig. 3.1a). On peut même ne plus avoir d'observations du tout quand $\rho \to -1$ car il devient très peu probable d'avoir les deux variables qui soient supérieures à h.

On veut estimer ρ à partir de X^+ et/ou X^{+*}. On calcule donc

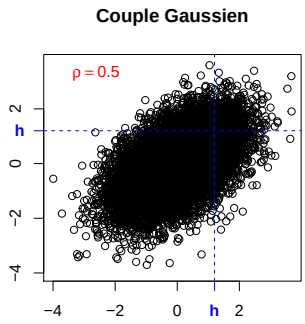
$$\mu_{11}^+ = E(X_1^+ X_2^+)$$

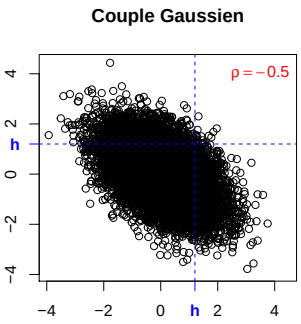
$$C^+ = \mu_{11}^+ - E(X^+)^2$$

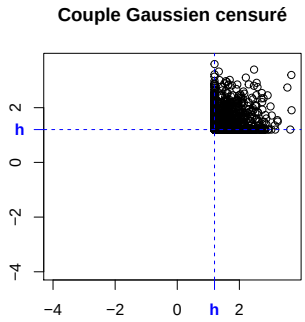
empiriquement (et idem avec X^{+*}) et avec les formules présentées en Annexe E, qui sont tirées de Muthen (1990). En minimisant la différence on estime $\hat{\rho}$.

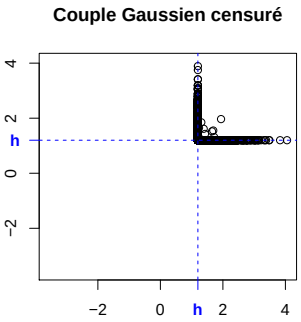
Sauf indication contraire on prendra h=1.2, proche de ce qu'on a généralement dans les données de pluie à un pas de temps de quelques minutes ($\mu = -1.2$).

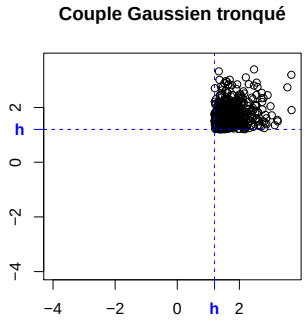
Si on trace les estimateurs évoqués (μ_{11}^{+*}, μ_{11}^+, C^+ et C^{+*}) en fonction de ρ, on obtient la Figure 3.2. Comme on peut s'y attendre lorsque les variables sont tronquées et que ρ est très négatif les estimateurs ne sont plus utilisables car on s'attend à n'avoir quasiment plus aucune observation. Même sans prendre en compte ce problème au-delà d'environ ρ=0.3 la fonction $\rho \mapsto \mu_{11}^{+*}$ n'est pas bijective, ce qui rend ce moment inutilisable pour estimer la

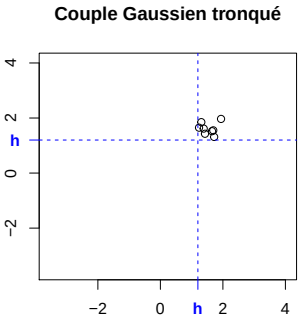












(a) Cas d'une corrélation $\rho = 0.5$ (b) Cas d'une corrélation $\rho = -0.5$

FIGURE 3.1 – Exemples de couples Gaussiens censurés et tronqués

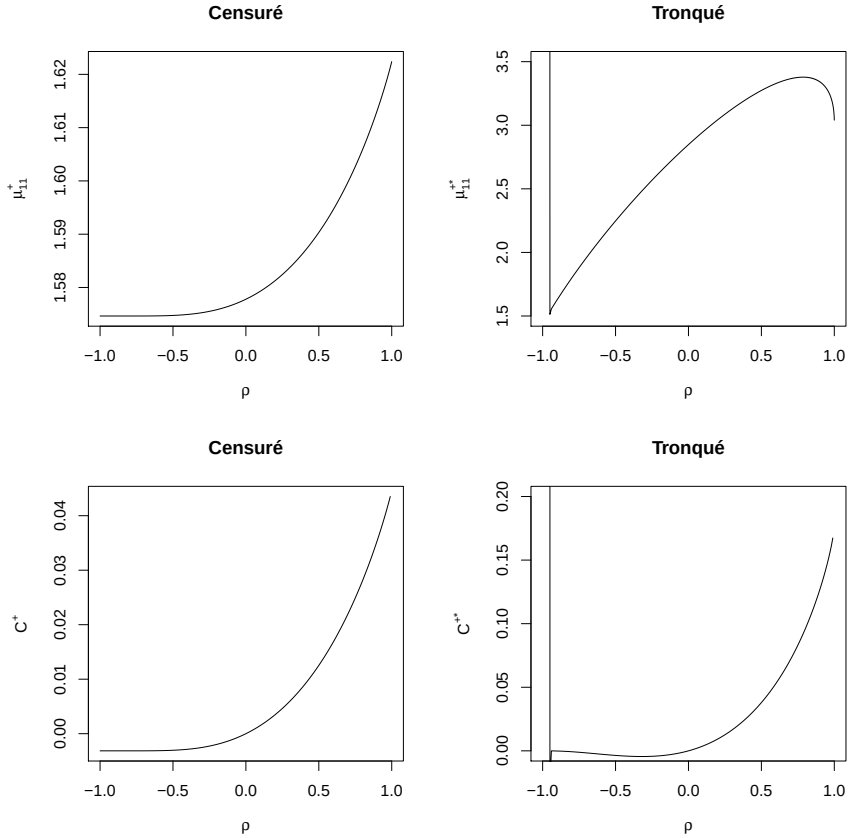


FIGURE 3.2 – Première ligne : moments μ_{11}^{+*} et μ_{11}^{+} en fonction de ρ. Deuxième ligne : covariance censurée et tronquée en fonction de ρ

dépendance. De la même façon l'estimateur de la covariance des données tronquées n'est pas utilisable pour les ρ négatifs.

Les estimateurs des données censurées μ_{11}^{+} et C^{+} sont équivalents. Les fonctions sont bijectives et peuvent donc être utilisées pour estimer ρ mais on s'attend à avoir une mauvaise estimation des dépendances fortement négatives car quand $\rho \to -1$ la courbe s'aplatit. Pour la suite on choisira d'utiliser l'estimateur de la covariance des données censurées pour estimer la dépendance par

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \, (C^{+} - Cov(X^{+}))^2.$$

3.3.2 Estimation de la dépendance par maximum de vraisemblance

L'estimation des paramètres de dépendance d'un champ Gaussien censuré peut être faite grâce à la vraisemblance par paires (Allcroft & Glasbey, 2003 ; Durbán & Glasbey, 2001).

Soit un couple de variables Gaussiennes censurées $X^+ = (X_1^+, X_2^+)$ tel que $X_i^+ = h \times \mathbb{1}_{X_i<h} + X_i \times \mathbb{1}_{X_i \geq h}$, avec $X \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ et $\mathbf{\Sigma} = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. La dépendance ρ peut être estimée en maximisant la vraisemblance $\mathcal{L}(X;\rho) = \prod p(x;\rho)$ où

$$p(x;\rho) = \begin{cases} \Phi_2((h,h)', \mathbf{\Sigma}) & \text{si } x_1 = x_2 = h \\ \phi(x_i)\Phi\left(-\frac{\rho}{\sqrt{1-\rho^2}} x_j\right) & \text{si } x_i > h \text{ et } x_j = h \\ \phi_2(x, \mathbf{\Sigma}) & \text{si } x_1 > h \text{ et } x_2 > h \end{cases}, \tag{3.2}$$

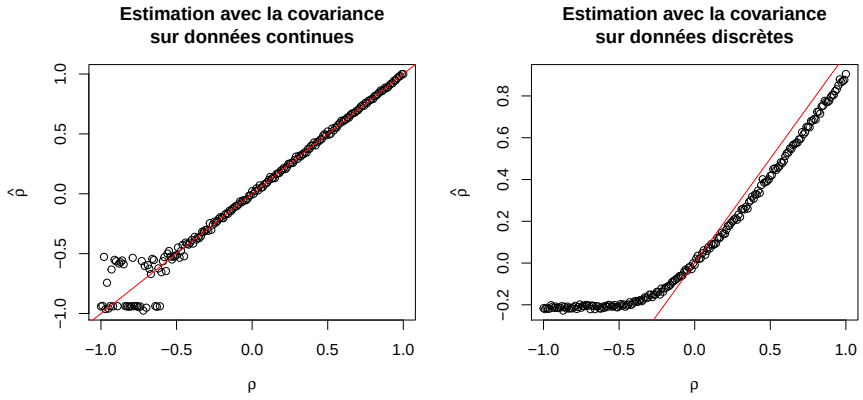
avec ϕ, la densité et la cdf d'une variable Gaussienne et $\phi_{2,2}$ la densité et la cdf d'un couple Gaussien.

3.3.3 Comparaison des deux estimateurs

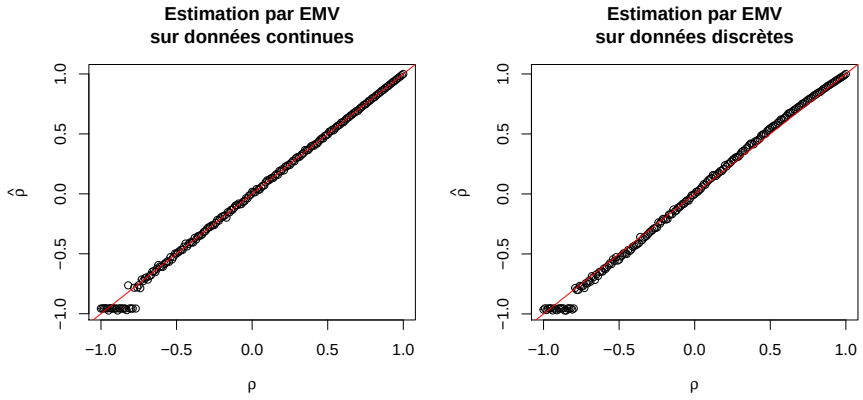
Dans cette section pour différentes valeurs de ρ un couple Gaussien de taille 10^4 est simulé puis censuré en 1.2. La dépendance est estimée avec la covariance censurée ou avec le maximum de vraisemblance. On souhaite aussi étudier l'effet de la discrétisation, on arrondit donc les données à 0.2 avant de ré-estimer la dépendance.

La Figure 3.3a permet de comparer les performances de la méthode des moments et la Figure 3.3b celles du maximum de vraisemblance.

On constate que sur les données continues les deux estimateurs donnent des résultats satisfaisants. Pour les dépendances très négatives, la covariance censurée commence à poser problème pour $\rho < -0.5$ alors que le maximum de vraisemblance reste utilisable jusqu'à $\rho > -0.75$ environ. La discrétisation a un impact beaucoup plus fort sur l'estimateur de la covariance que sur le maximum de vraisemblance. On observe un biais pour les dépendances positives et une stagnation à -0.2 pour les dépendances négatives. Pour la suite, on utilisera donc le maximum de vraisemblance, et il faudra garder en tête que pour les données fortement discrétisées, l'estimateur est légèrement biaisé pour les dépendances d'intensité moyenne (surestimation autour de 0.5, sous-estimation autour de -0.5).



(a) Méthode des moments avec la covariance censurée



(b) Maximum de vraisemblance

FIGURE 3.3 – Estimation de la dépendance sur données continues et discrètes (arrondies à 0.2)

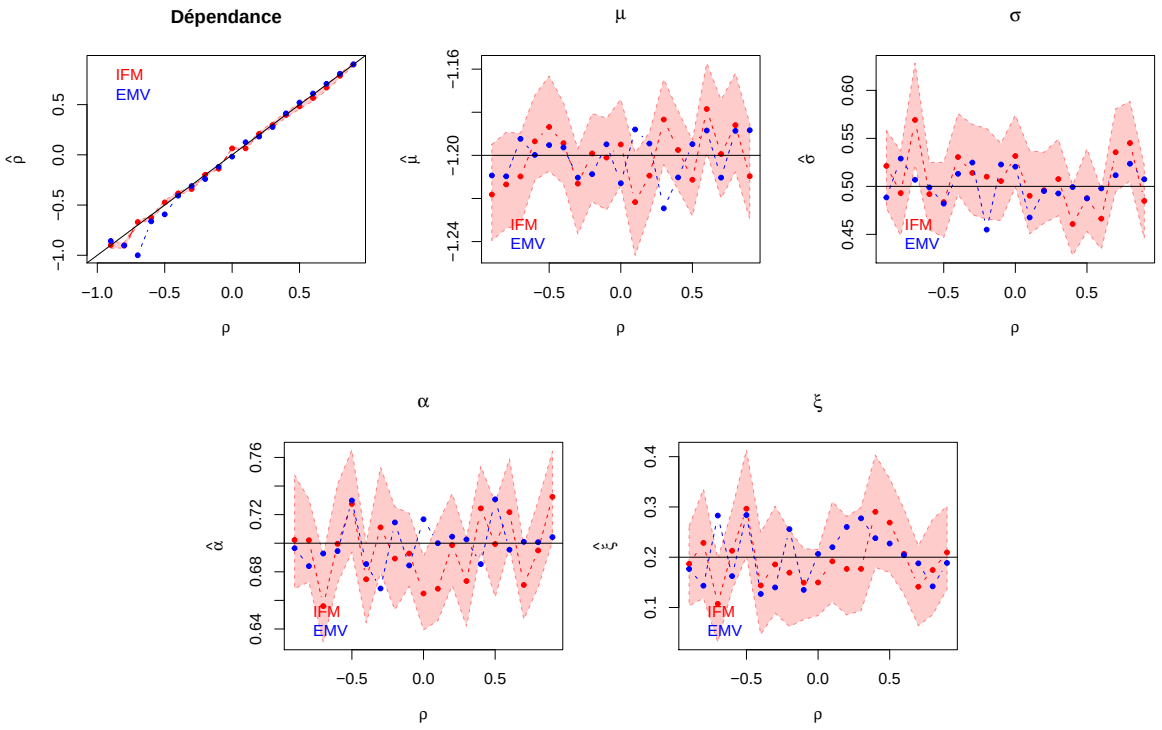


FIGURE 3.4 – Estimation des paramètres du modèle GP méta-Gaussien par IFM et EMV en fonction de la dépendance. L'intervalle contient 95% des valeurs calculées sur 100 échantillons bootstrap non paramétriques. La ligne pleine représente les vraies valeurs des paramètres.

3.4 Tests sur simulations

Dans cette section, on simule un couple Gaussien de moyenne $\mu = -1.2$, de variance 1 et de corrélation ρ. Les variables de « précipitation » associées sont produites en appliquant le modèle GP méta-Gaussien (2.8) avec comme paramètres $\{\sigma = 0.5, \alpha = 0.7, \xi = 0.2\}$. Ces paramètres ont été choisis pour ressembler à ce qu'on retrouve souvent dans nos données à des pas de temps de quelques minutes. Les échantillons simulés contiennent 10^4 observations. La Figure 3.4 montre les paramètres estimés par les deux méthodes présentées précédemment (IFM et EMV). On constate que sur ces simulations les deux méthodes donnent des résultats similaires et très satisfaisants. La dépendance est particulièrement bien estimée avec peu de variabilité. Bien que l'estimation des paramètres des marges montre plus de variabilité, l'ajustement des marges a été vérifié sur les QQ plots et densités et s'est avéré quasiment parfait pour les deux méthodes.

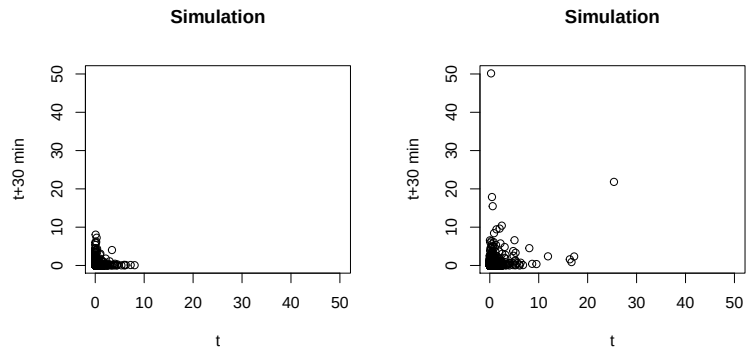


FIGURE 3.5 – Gauche : Couple $(Y_t, Y_{t+30\text{min}})$ des données radar à 5 minutes au point de BMO en février. Droite : Simulation d'un couple GP méta-Gaussien, paramètres estimés par EMV.

La variabilité de l'estimation par EMV n'a pas été faite pour des raisons de temps de calcul.

3.5 Test sur les données

Afin de minimiser l'impact de la discrétisation, il a été choisi de travailler sur les données radar. De plus les simulations ayant été présentées pour des couples ayant les mêmes distributions marginales, on prendra comme couple $(Y_t, Y_{t+30\text{min}})$. Finalement on choisit de travailler avec le point du radar qui se situe au-dessus du pluviomètre de BMO et on se concentre sur le mois de février. Les Figures 3.5 et 3.6 (gauche) montre le couple $(Y_t, Y_{t+30\text{min}})$. On observe de fortes concentrations le long des axes pour les fortes valeurs, ce qui montre qu'une intensité forte à l'instant t est suivie par une précipitation faible voire nulle 30 minutes après.

Les résultats obtenus par IFM (estimation des marges puis maximum de vraisemblance dans le domaine Gaussien) sont montrés en Figure 3.6. Ceux obtenus par EMV (estimation de tous les paramètres par maximum de vraisemblance) sont montrés en Figure 3.5.

Tout d'abord on remarque que l'estimation des distributions marginales avec l'EMV est assez mauvaise : on obtient beaucoup plus de fortes valeurs. Ce phénomène est vérifiable en Figure 3.7 où on montre la qualité d'ajustement des marges avec les deux méthodes. Il n'est pas étonnant que les marges soient bien reproduites par l'IFM étant donné qu'on a déjà pu montrer que le modèle GP méta-Gaussien univarié donnait de bons ajustements

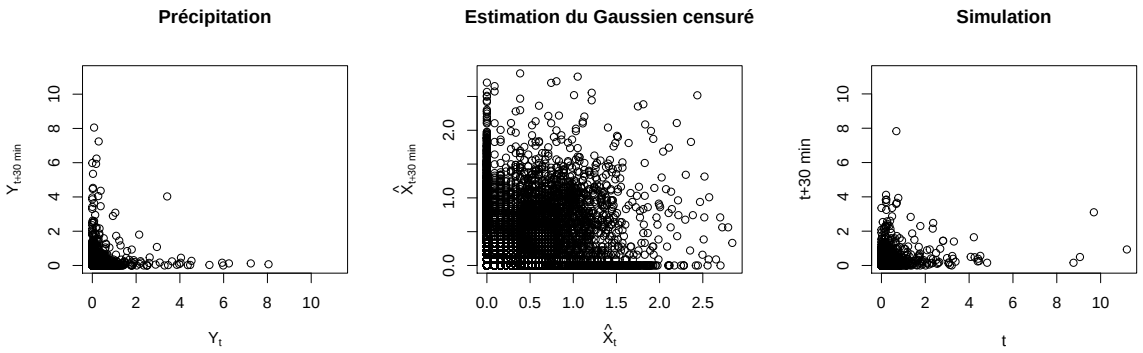


FIGURE 3.6 – Étapes de l'ajustement par IFM. Gauche : Couple $(Y_t, Y_{t+30\text{min}})$ des données radar à 5 minutes au point de BMO en février. Milieu : Estimation du couple Gaussien censuré grâce à l'estimation des paramètres des marges. Droite : Simulation d'un couple GP méta-Gaussien après estimation de la dépendance.

sur les données radar. On constate que les faibles intensités sont correctement reproduites par les deux méthodes, mais la queue de distribution est trop lourde avec les paramètres ajustés par l'EMV. En effet le seul paramètre très différent entre les deux méthodes d'estimation est ξ, qui est estimé à 0.6 par l'IFM et à 0.9 par l'EMV.

Dans le cas de l'IFM le principal problème est la surestimation de la dépendance : sur la Figure 3.6 on voit que les simulations ne reproduisent pas les concentrations des points sur les axes. La Figure 3.8 permet de mieux comparer les données aux simulations en zoomant sur les intensités modérées. Même si la différence est moins marquée que sur les intensités extrêmes on constate bien la surestimation de la corrélation par l'IFM. Les concentrations sur les axes sont sous-estimées, des points qui devraient être collés aux axes (une variable vaut 0, l'autre est moyenne) se retrouvent légèrement décollés, ce qui crée un surplus d'observations où une variable est faible et l'autre est modérée.

En conclusion les ajustements sur les données ne sont pas vraiment satisfaisants, ce qui peut s'expliquer soit par un problème d'estimation soit par une mauvaise spécification du modèle. Les résultats sur simulations semblent indiquer la deuxième option, ce qui pourrait par exemple vouloir dire qu'une copule Gaussienne n'est pas adaptée pour décrire la dépendance dans le cas étudié.

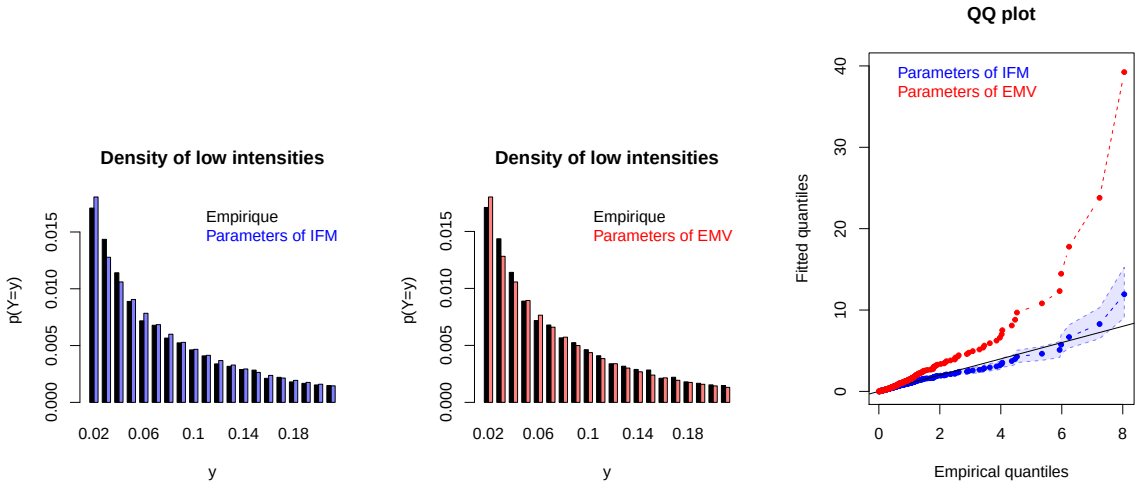


FIGURE 3.7 – Qualité de l'ajustement des marges. Gauche et milieu : densités des intensités faibles (< 0.2). Droite : QQ plot, l'intervalle 95% (surface colorée) est calculé sur 100 échantillons bootstrap non paramétriques.
Remarque : le bootstrap n'est fait que dans le cas de l'IFM pour des raisons de temps de calcul.

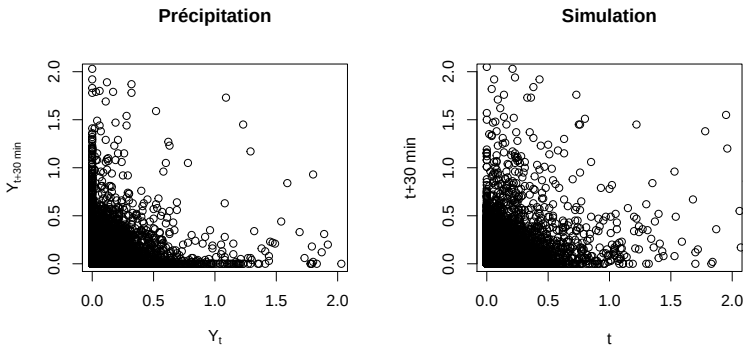


FIGURE 3.8 – Zoom de la Figure 3.6 sur les intensités faibles à moyennes (<2 mm/5 min)

Les délais n'ayant pas permis de pousser plus loin l'analyse ni de chercher des extensions aux modèles méta-Gaussiens qui soient plus adaptées à la dépendance des précipitations, l'assimilation de données a été abandonnée. Afin de produire le champ de précipitation le plus réaliste possible, une méthodologie plus simple a été développée, en se basant sur l'idée venant du Chapitre 1 d'utiliser la structure spatiotemporelle du radar et la distribution des pluviomètres.

Chapitre 4

Génération de chroniques de pluie pour étudier la sensibilité du modèle hydraulique

4.1 Introduction

Le modèle hydraulique décrivant le fonctionnement du réseau d'assainissement a historiquement comme entrée pour la précipitation la chronique du pluviomètre de BMO. Cette entrée est donc mesurée en un seul point, à un pas de temps de 3 minutes, avec une précision de 0.2 mm par un pluviomètre à auget. Au vu de la taille des bassins versants, l'étude de Berne et al. (2004) recommande une résolution spatio-temporelle de 3 minutes par 2 km^2. Cette étude est menée à Marseille et n'est donc pas directement transposable au cas de la région brestoise mais ça permet de donner un ordre de grandeur.

Dans ce chapitre on cherche à répondre aux deux besoins du projet MEDISA : 1) étudier la sensibilité du modèle hydraulique à la pluviométrie et 2) fournir des chroniques d'entrée pour le dimensionnement des réseaux d'assainissement.

Après une étude bibliographique des méthodes existantes de fusion de données et de sélection d'année typique, une méthode sera développée pour générer des données de pluie avec différentes caractéristiques (échantillonage spatio-temporel, distribution, discrétisation), afin tester la sensibilité du modèle à ces caractéristiques.

Ainsi la méthode présentée ici sera une méthodologie de fusion de données basée sur les résultats obtenus au Chapitre 1, qui ont montré que les données radar donnaient une bonne représentation de la structure spatio-temporelle alors que les pluviomètres avaient une distribution marginale de la pluie généralement considérée comme plus fiable.

Les résultats de l'analyse de sensibilité permettront de proposer les chroniques à utiliser comme entrée du modèle hydraulique. Une comparaison des déversements simulés par le modèle sera alors faite avec la mesure en réseau afin de valider la chronique.

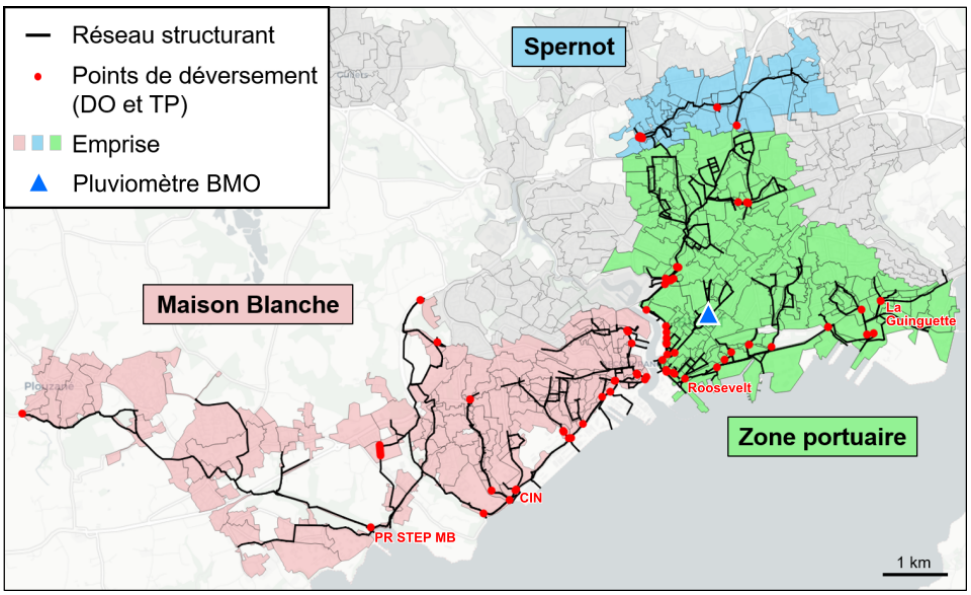


FIGURE 4.1 – Carte simplifiée du réseau d'assainissement.

Finalement cette étude permettra de conseiller Eau du Ponant sur un éventuel recalage du modèle et sur les données de précipitation à utiliser pour tester différents scénarios d'aménagement.

4.1.1 Présentation du modèle hydraulique

Le modèle hydraulique représente le fonctionnement des réseaux d'assainissement sur Brest Métropole. Une carte simplifiée du réseau est montrée en Figure 4.1. La zone d'étude est étendue sur environ 14 km sur l'axe O-E et 8 km sur l'axe N-S pour une surface de collecte de 1300 ha de bassins unitaires. Eau du Ponant travaille avec le logiciel Infoworks ICM, qui permet notamment de simuler la réaction du réseau à des conditions météorologiques données.

Seul le réseau structurant est modélisé, c'est celui qui tracé en Figure 4.1. Ce choix est justifié par l'objectif du modèle, qui est de modéliser les déversements au niveau des déversoirs d'orage (DO).

Les hypothèses fortes du modèle développé par Eau du Ponant sont :

— La modélisation des eaux de nappes par une année moyenne et non pas par un modèle de sol,

— la représentation de l'occupation des sols par seulement 2 types (perméable / imperméable),

— l'absence de ré-injection des volumes déversés dans le réseau.

L'occupation des sols est en effet généralement décrite de façon très détaillée, par exemple dans le rapport publié par Le Grand Lyon (2008) ou encore celui d'AMODIAG Environnement (2012) il y a une vingtaine de types de sols. Le choix drastique de passer à seulement deux types relève d'une volonté de simplifier le suivi et la compréhension du calage. La multiplication des paramètres de calage pourrait entraîner un sur-pprentissage du modèle. La petite taille des sous bassin versants justifie également cette simplification, et il a été observé qu'avec cette simplification les résultats obtenus étaient tout aussi satisfaisants. Finalement le ruissellement est modélisé par la méthode de Desbordes (Desbordes, 1974), qui est classiquement utilisée en France et au Royaume-Uni.

Le modèle est calé de l'amont vers l'aval, puis au niveau global. Il y a 85 points de calage instrumentés : 7 bassins d'orage, 48 déversoirs d'orage et 20 postes de relevage. L'instrumentation mesure la hauteur d'eau dans la conduite, hauteur qui est ensuite convertie en débit en prenant en compte la structure de l'ouvrage.

Les chroniques de pluies utilisées comprennent des chroniques de calage et des chroniques de validation, à chaque fois par temps de pluie et par temps sec. Pour faire coïncider la mesure aux résultats du modèle, plusieurs paramètres sont modulables. Pour le calage par temps sec, les paramètres sont le fonctionnement des pompes, les eaux claires parasites permanentes (infiltration des eaux des nappes phréatiques dans les réseaux) à travers le débit moyen annuel et les eaux usées à travers le nombre d'habitants. Il a été considéré que 95% des volumes d'eau potable consommés sont rejetés dans le réseau d'assainissement (zone fortement urbanisée). Par temps de pluie, les paramètres de calage sont la surface active des bassins versants et le coefficient de propagation de la pluie sur les surfaces imperméables.

Enfin le modèle hydraulique a un pas de temps de calcul de une minute, mais il est calé sur des données à 3 minutes, qu'il s'agisse des données de pluie ou de la mesure en réseau.

4.1.2 Besoins et données disponibles

Actuellement le calage du modèle est réalisé avec les données de pluie du pluviomètre considéré comme le plus fiable de la zone, celui de l'hôtel de ville BMO. La même pluie est donc appliquée sur l'ensemble de la zone, et la question est de savoir si il est pertinent

d'utiliser une donnée mesurée jusqu'à 10 km du point considéré. De plus, d'expérience il est connu que selon la direction du front deux pluies de même intensité n'auront pas les mêmes conséquences sur le réseau, et cette information ne peut être prise en compte avec le système actuel. Dans le logiciel Infoworks, la grille la plus fine est celle des sous bassins versants (délimités par les lignes grises en Figure 4.1), autrement dit il est possible rentrer une chronique par sous bassin versant.

La question principale est donc celle de la spatialisation : la spatialisation des pluies a-t-elle un impact sur les déversements en milieu naturel ? Plus généralement c'est la question du choix de la chronique qui se pose, c'est-à-dire quelle entrée de pluie permet de reproduire au mieux la mesure en réseau et donc à quelle(s) caractéristique(s) de la pluie le modèle hydraulique est-il sensible ?

Au vu des données disponibles (Chapitre 1), on souhaite tester la sensibilité du modèle :

— à l'échantillonnage spatial (1 point, 11 points ou une grille de 1 km^2),

— à l'échantillonnage temporel (3 ou 5 minutes),

— à la discrétisation (0.2 mm ou 0.01 mm),

— à la distribution marginale (mesure issue du radar ou d'un pluviomètre).

Pour tester ces effets en les séparant correctement les uns des autres, il apparaît nécessaire de pouvoir mélanger les données des pluviomètres et les données du radar. Comme ça été démontré au Chapitre 1, chaque source de donnée contient des erreurs de mesure et nécessite d'être corrigée.

4.2 État de l'art sur la fusion de données

Les techniques de fusion de données visent à utiliser différentes sources de données pour reconstruire un jeu de données qui tire profit de toutes les sources. Dans la littérature de nombreuses méthodes visant à fusionner les données de radar aux données de pluviomètres existent. Historiquement le but est principalement de corriger les biais du radar avec les relevés des pluviomètres au sol, considérés comme plus fiables.

La méthode la plus simple et la plus classique est le Mean Field Biais (Chumchean et al., 2006 ; Seo, 1998), qui a pour but de corriger les différences spatialement systématiques entre radar et pluviomètres. Le biais est défini comme

$$B_t = \frac{\sum_{i=1}^{n_t} G(i,t)}{\sum_{i=1}^{n_t} R(i,t)}$$

en notant $R(i,t)$ le pixel du radar à la position i accumulé sur l'heure t, $G(i,t)$ le pluviomètre à la position i accumulé sur l'heure t, et n_t est le nombre de pluviomètres qui ont un cumul non nul sur l'heure t. L'image radar entière à l'heure t peut ensuite simplement être corrigée avec B_t, comme c'est le cas dans Chumchean et al. (2006), mais B_t peut aussi être modélisé comme un processus aléatoire comme c'est le cas dans Smith et Krajewski (1991).

Une variation de cette méthode consiste à estimer un biais non uniforme spatialement, ce qu'on appelle généralement la correction de biais local (Seo & Breidenbach, 2002). La correction locale nécessite une bonne couverture de la zone par les pluviomètres.

De nombreuses approches de la fusion de données sont basées sur le krigeage. Le krigeage est une technique d'interpolation spatiale originellement développée dans le domaine de la géostatistique par Matheron (1965). Le krigeage consiste en une combinaison linéaire des points disponibles avec un poids déterminé par un modèle variographique, c'est-à-dire un modèle donnant une mesure du lien entre deux points qui est généralement une fonction de la distance qui les sépare. Le cokrigeage est une extension du krigeage au cas multivarié, c'est la méthode utilisée historiquement pour la fusion radar - pluviomètres (Azimi-Zonooz et al., 1989 ; Krajewski, 1987). Le problème de cette méthode est qu'elle nécessite de calculer la covariance entre la « vraie » précipitation et à la fois les pluviomètres et le radar. Dans Krajewski (1987) cette covariance est approximée sur la base de la matrice de covariance du radar, mais ça reste peu satisfaisant. Aussi les références plus récentes utilisent le krigeage avec dérive externe (Haberlandt, 2007). La dérive externe est le radar qui sert comme information d'arrière plan pour estimer la variable principale qui sera l'interpolation spatiale des pluviomètres. Une version du krigeage avec dérive externe basée sur des indicatrices (Isaaks & Srivastava, 1989) est aussi utilisée dans C. Berndt et al. (2014), l'idée étant de transformer le radar en indicatrices avec plusieurs seuils et de réaliser le krigeage avec dérive externe sur chaque indicatrice. Cela permet d'obtenir une densité de probabilité pour le champ interpolé des pluviomètres. Une autre technique est le conditional merging (Ehret, 2003 ; Sinclair & Pegram, 2005), qui consiste à combiner le champ des pluviomètres interpolé par krigeage ordinaire avec l'information sur la variabilité entre les points des pluviomètres issue du radar. C. Berndt et al. (2014), Ehret (2003) et Goudenhoofdt et Delobbe (2009) comparent les différentes méthodes basées sur le krigeage. Le conditional merging semble être la meilleure option pour les données à 10 minutes, le krigeage avec dérive externe est plus adapté aux données journalières mais

aussi plus sensible à la qualité des données radar que sa version sur indicatrices ou que le conditional merging.

Une des difficultés liées au krigeage des données de précipitation est de savoir comment la non gaussianité des données impacte l'interpolation. D'après Matheron (1973) si la distribution est asymétrique alors les estimateurs du krigeage sont sensibles aux quelques très fortes valeurs, mais la normalité n'est pas nécessaire pour le krigeage. Cette question est souvent traitée en transformant les données pour les normaliser, ce qui est un approche classique de la modélisation de la précipitation de façon générale, comme ça a été montré au Chapitre 2. Les transformations les plus courantes sont sûrement la racine carrée, proposée par Panofsky et al. (1958), et la transformation Box-Cox (Box & Cox, 1964) utilisée par exemple dans Cecinati et al. (2017) et Hussain et al. (2010). Ces transformations sont comparées dans le cadre de l'utilisation du krigeage pour la fusion des données de pluie dans Cecinati et al. (2017). Cette étude conclue que la gaussianisation des données améliore le krigeage et note que la transformation Box-Cox dépend de la variabilité spatiale et temporelle de la précipitation.

Il existe encore de nombreuses autres méthodes de fusion des données de radar et de pluviomètres, comme par exemple des modèles non paramétriques pour lier la réflectivité radar à l'intensité de pluie (au lieu de la loi Marshall Palmer), associés à une interpolation des pluviomètres basés sur les copules (Wasko et al., 2013). Enfin on notera la méthode de Todini (2001), qui formule le krigeage dans un cadre bayésien. Ce papier pose la question de la comparaison de la mesure ponctuelle du pluviomètre à celle moyennée sur 1 km^2 du radar en utilisant un krigeage par bloc.

Toutes les méthodes de fusion présentées visent à corriger les données radar par les pluviomètres qui sont considérés comme mesurant la « vraie » pluie. Par conséquent elles ne permettent pas de prendre en compte de potentielles erreurs des pluviomètres. Or comme ça a été montré au Chapitre 1 le réseau de pluviomètres entretenus par Eau du Ponant ne peut pas être considéré comme fiable.

L'assimilation de données est un champ de recherche qui vise à corriger les sorties de modèles climatiques par des observations réelles. Les observations ne sont pas alors considérées comme la vérité, les méthodes statistiques permettent de prendre en compte l'incertitude à la fois du modèle et de la mesure. Des exemple de méthodes d'assimilation de données de précipitation sont Otsuka et al. (2016), Lien et al. (2013) ou encore Lien

et al. (2016). Le cadre de l'assimilation se mêle particulièrement bien à celui des modèle méta-Gaussiens, qui par nature définissent une variable latente,

D'après notre analyse des données et la littérature le radar donne une bonne estimation de la structure spatio-temporelle de la pluie. De plus, on ne peut pas obtenir une information fiable pour cette caractéristique à partir des pluviomètres à cause de l'échantillonnage spatial et des erreurs de mesure. Pour ce qui est de la loi marginale on peut s'attendre à ce qu'en sélectionnant les pluviomètres les plus fiables on obtiennent de meilleurs résultats qu'avec la distribution des données radar.

Conclusion sur la fusion de données

Pour fusionner les deux sources de données on cherche à construire un jeu de données ayant la structure spatio-temporelle d'un radar et la loi marginale d'un pluviomètre.

4.3 Choix d'une année pour le modèle hydraulique

Comme ça a été expliqué le projet MEDISA a besoin d'une ou 5 année(s) représentative(s). Ces années pourraient être fictives, autrement dit simulées avec un générateur aléatoire de conditions météorologiques. Toutefois il y a peu d'intérêt à faire un simulateur car on a besoin de peu d'années, et prendre une année réelle a des avantages. En effet le modèle milieu du projet simule la propagation des volumes déversés dans la rade de Brest et a comme forçage le vent et les vagues, il aurait donc fallu développer un générateur aléatoire de vent, vagues et précipitations. De plus prendre une année réelle permet de comparer les résultats obtenus à la mesure en réseau.

Il a été décidé de commencer les tests sur une année réelle entre 2014 et 2019, commune à tous les modèles, afin de pouvoir tester facilement de nombreux scénarios d'aménagements. Plus tard dans le projet, ces cinq années (toujours réelles) seront utilisées pour valider la robustesse du scénario choisi, conformément à la réglementation.

Deux jeux de données seront utilisés pour choisir l'année à utiliser.

— Données pluviomètres Eau du Ponant : 2010-2019, 3 min, pluviomètre de BMO
— Données historiques Météo France : 1945-2018, 1 jour, Guipavas

Les données historiques de Météo France donnent une bonne idée de la climatologie, de la distribution de la pluie (Section 1.1.2 page 24). Avoir des données sur une aussi longue période est très utile pour définir si une année est plutôt « normale » ou « exceptionnelle » mais il faut garder en tête que globalement on a un peu plus de pluie à Guipavas qu'à BMO. Ça va surtout être visible en agrégeant beaucoup (cumuls annuels).

Il est notamment important de choisir une année « caractéristique », qui ne va ni sous-dimensionner ni sur-dimensionner les ouvrages.

4.3.1 La notion d'année typique

Le choix d'une année de référence (ou année typique) est une problématique qui se pose souvent dans les domaines où les coûts numériques des modèles sont importants et où il n'est pas possible de faire tourner plus d'une année pour de nombreux scénarios. La bibliographie qui suit a été faite dans le cadre de l'écriture de Ailliot et al. (2020), elle concerne donc principalement des applications dans le domaine des systèmes énergétiques. Les années typiques ont aussi été utilisées pour des applications en hydrologie, par exemple dans Null et Viers (2013).

Une année typique est définie par la concaténation de mois typiques, où chaque mois est sélectionné parmi une base de données de façon à être représentatif des conditions climatiques de ce mois. Un des premiers travaux sur ce sujet est celui de Benseman et Cook (1969) qui utilise la comparaison des moyennes mensuelles du paramètre météorologique d'intérêt (e.g. irradiance solaire, température, vitesse du vent, etc.) pour choisir les mois typiques.

L'approche la plus utilisée est la méthode « Sandia », présentée dans I. J. Hall et al. (1978) pour l'énergie solaire, qui utilise la distance de Finkelstein-Schafer (Finkelstein & Schafer, 1971) basée sur les fonctions de répartition. De nombreuses autres méthodes existent, comme par exemple la méthode Danoise (Lund, 1995) développée pour la conception d'immeubles, qui calcule une distance basée sur la moyenne et l'écart-type. Une variation de la méthode Danoise est présentée dans Festa et Ratto (1993) pour l'irradiance solaire. Elle inclut une métrique basée sur les distributions via la distance de Kolmogorov-Smirnov.

Afin de gérer plusieurs variables météorologiques simultanément, l'approche la plus commune consiste à utiliser une somme pondérée des distances mesurées pour chaque variable météorologique.

La question de chercher la série la plus typique revient à mesurer une dissimilarité

entre deux séries temporelles de tailles différentes. De nombreuses mesures de distances existent, le package R `TSdist` (Mori et al., 2016) en contient une bonne quantité. On peut catégoriser les distances en fonction de ce qu'elles mesurent. Les distances correspondant aux méthodes classiques pour les années typiques sont généralement des distances basées sur la distribution (comparaison de moyenne, méthode « Sandia »). Dans cette catégorie d'autres distances souvent utilisées sont Cramer Von Mises (Darling, 1957), Anderson Darling (Anderson & Darling, 1952) qui est son adaptation pour les valeurs extrêmes, Kolmogorov Smirnov (Darling, 1957) et Kullback Leibler (Kullback & Leibler, 1951). Ces distances pouvant être multivariées, on peut faire intervenir plusieurs variables différentes ou une variable et son lag (X_t et $X_{t-1},\ t \in 2, ..., n$), ce qui nous amène à la seconde catégorie : les distances basées sur la structure temporelle. On y retrouve des distances basées sur la corrélation ou sur des modèles auto-régressifs. Ensuite des mesures élastiques peuvent être utilisées, elles visent à aligner au mieux les séries, comme le dynamic time warping (D. J. Berndt & Clifford, 1994), LCSS (plus longue séquence commune) (Vlachos et al., 2002) ou encore la distance de Fréchet (Fréchet, 1906). Enfin on note quelques distances plus originales, comme celles basées sur la compression de données (Keogh et al., 2004) ou l'encodage de la série en chaîne de caractères (Lin et al., 2003).

Un précédent travail (Boutigny, 2017, disponible sur HAL) a cherché à comparer les capacités de ces différentes méthodes à sélectionner une année typique de précipitation permettant de reproduire les nombres de jours déversements produits par un système très simplifié (un bassin versant et un bassin d'orage avec un trop-plein). Des résultats on peut retenir 4 distances, qui sont les plus consistantes et qui donnent des années typiques moyennes à conservatrices (i.e. nombre de jours de déversement élevé) : Cramer Von Mises, Anderson Darling, Kullback Leibler et la distance basée sur la corrélation croisée (Liao, 2005)

$$\sqrt{\frac{1 - CC(x, y, 0)^2}{\sum_k (1 - CC(x, y, k)^2)}}$$

où CC est la fonction de corrélation croisée entre x et y au décalage k.

Un inconvénient attendu lors de l'usage d'une année typique construite par les méthodes classiques est que l'année typique ne reproduit pas la variabilité inter-annuelle. Ce problème peut être traité en utilisant un générateur aléatoire, mais dans notre cas c'est une année réelle et entière qui doit être sélectionnée. La concaténation de mois typique n'a pas été envisagée afin de simplifier la coordination du travail des différentes parties

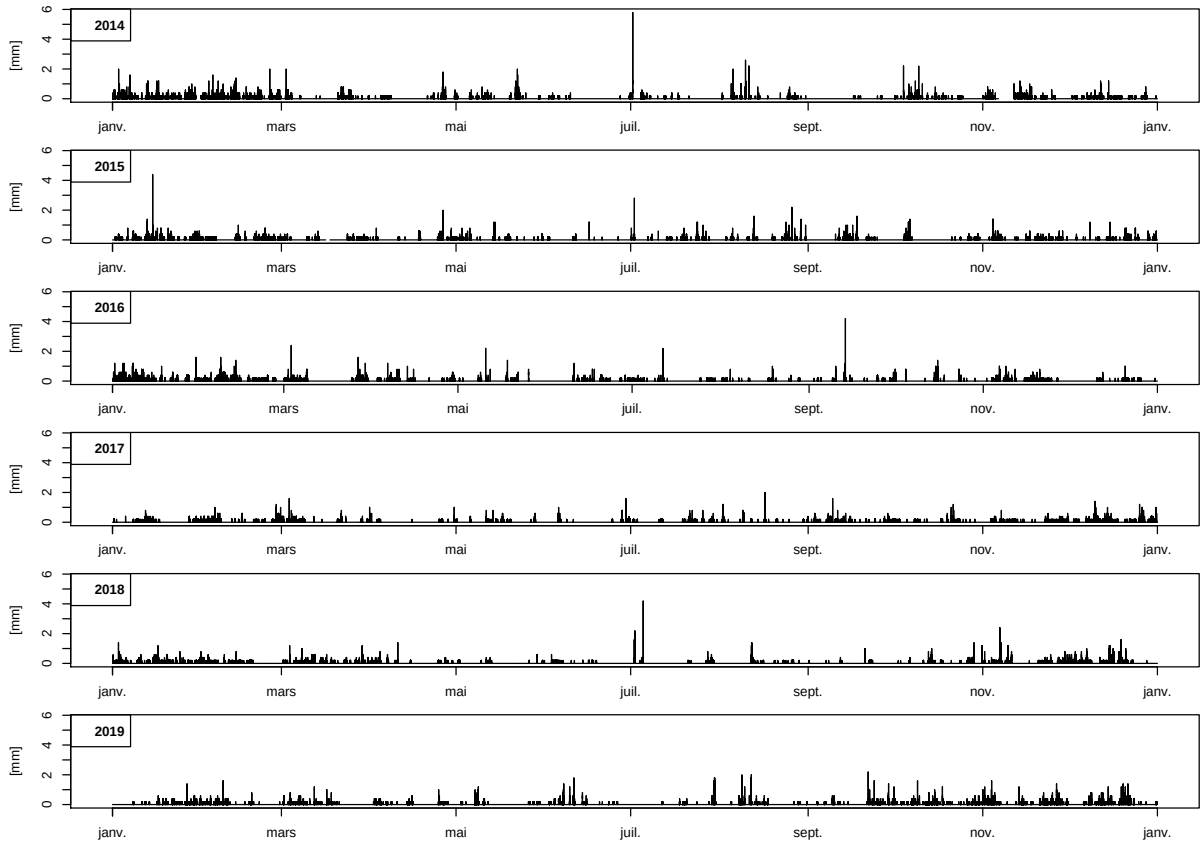


FIGURE 4.2 – Les six années possibles à un pas de temps de 3 minutes (pluviomètre de BMO).

du projet MEDISA. On pourra aussi argumenter que la concaténation pose le problème des ruptures entre les fin et début des différents mois choisis.

4.3.2 Vue globale

Les six dernières années sont présentées en entier au pas de temps trois minutes sur la Figure 4.2. Même s'il est difficile d'évaluer les années à partir de ce graphique il est déjà possible de repérer des années où il pleut particulièrement peu (2017) et les pics d'intensité (on remarque que 2019 et 2017 sont les seules années à ne pas en avoir). Les périodes où il semble pleuvoir le plus (début 2014, début 2016, novembre 2019) sont aussi repérables.

Ces impressions sont cohérentes avec les cumuls annuels, visibles en Figure 4.3 à gauche. L'histogramme montre les données historiques de Météo France (74 ans), et les points montrent les six années disponibles à BMO. Pour donner une idée de la variabilité

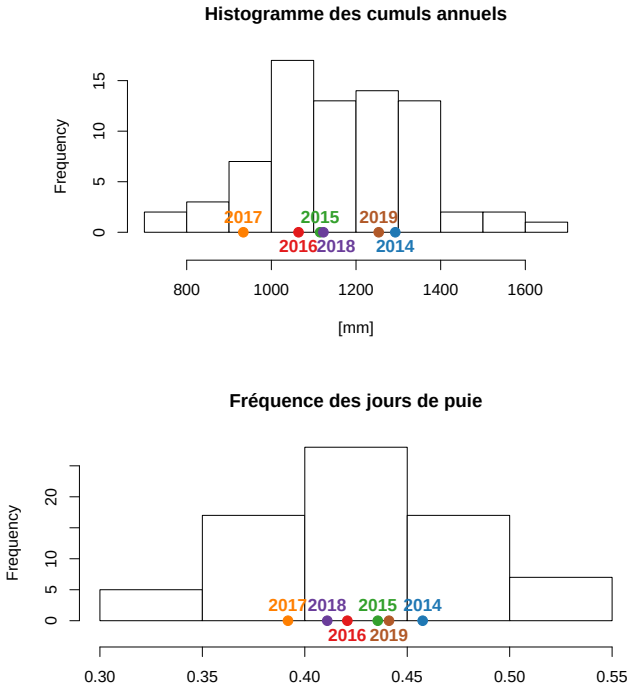


FIGURE 4.3 – Histogramme sur les données historiques 1945-2018 des cumuls annuels (gauche) et de la fréquence des jours de pluie (droite) à la station Météo France Guipavas. Les points en couleur correspondent aux données des 6 dernières années à BMO.

interannuelle de la pluviométrie, le ratio entre l'année la plus pluvieuse et l'année la moins pluvieuse est de 2.3 dans les données historiques. Dans les 6 années possibles, ce ratio est de seulement 1.4. On s'attend donc à ce que la variabilité interannuelle ait un impact très important sur le dimensionnement, ce qui explique que la réglementation demande d'utiliser 5 ans de données. 2014 est bien l'année la plus pluvieuse en termes de cumuls annuels, et on a tout de suite envie d'écarter 2017, bien trop sèche. La fréquence des jours de pluie est représentée de la même façon en Figure 4.3 à droite. Un jour est considéré comme pluvieux si le cumul journalier est supérieur ou égal à 1 mm.

Ces graphiques montrent que les années 2014 et 2019 semblent être les plus pluvieuses. De l'autre côté l'année 2017 est clairement la plus sèche, ce qui permet de la sortir des possibilités dès maintenant. 2014 est certes une année très pluvieuse, la plus pluvieuse des 6 possibles, mais ce n'est pas non plus une année « exceptionnelle » au regard de la climatologie obtenue avec les données historiques de Météo France. Aucune des années disponibles ne semble être exceptionnelle. Les statistiques annuelles restant une information

très globale, la prochaine section se fera au niveau mensuel.

4.3.3 Vue mensuelle

Tout d'abord on revient sur les distances utilisées pour la sélection d'année typique. On rappelle qu'on a sélectionné trois distances sur les distributions : Cramer Von Mises (CVM) qui mesure la somme des écarts entre les cdf (fonctions de répartition), Anderson Darling (AD) qui fait la même chose mais avec un poids donnant plus d'importance aux extrêmes, et Kullback Leibler (KL) qui est basée sur les pdf (densités). Une quatrième distance, basée sur la corrélation croisée, mesure la bonne représentation de la structure temporelle.

Les résultats par mois des distances entre chaque année entre 2014 et 2019 et la série entière (2010-2019) sont montrés en Figure 4.4. On constate que les distances CVM et AD donnent des résultats très similaires, ce qui est assez surprenant. Les mois de janvier et février 2014 montrent une distribution très différente des autres années (CVM, AD et KL fortes), ce qui semble confirmer le côté exceptionnel de ce début d'année très pluvieux. Si on repère les mois avec les distances les plus fortes, on aurait envie d'exclure : 2014 (jan., fév., sep.), 2015 (juin), 2016 (déc., juil.), 2017 (avr.), 2019 (oct.). La seule année restante serait alors 2018. Toutefois on note que les années très sèches (2015 et 2017) ne ressortent pas tant que ça sur les graphiques, or on sait que le risque de sous-dimensionner avec ces années est très fort, l'utilisation de ces distances n'est donc pas suffisant.

Un premier point d'intérêt concerne les cumuls mensuels et la fréquence des jours de pluie (cumul journalier supérieur à 1 mm), visibles en Figure 4.5. En 2018 il y a peu de précipitations par rapport à la norme sur l'été, et l'hiver est plutôt moyen malgré deux mois fortement pluvieux (déc. et jan.). En 2016 malgré un début d'année très chargé le reste de l'année est plutôt faible par rapport aux normales saisonnières. En 2019 le début d'année est plutôt calme mais il y a des cumuls élevés à la fin de l'année avec notamment deux mois avec une fréquence des jours de pluie très forte. Finalement en 2014 il y a des forts cumuls répartis sur toute l'année (jan., fév., mai, août). L'hiver 2014 est particulièrement pluvieux, même comparé aux 74 ans de données historiques, il y a notamment en février quasiment 100% de jours de pluie.

Ensuite le maximum de chaque mois permet de mieux représenter les pics d'intensité donc par exemple les orages. Pour ça les données historiques peuvent être utilisées, mais ce sera le cumul journalier le plus fort du mois uniquement. Les données à trois minutes

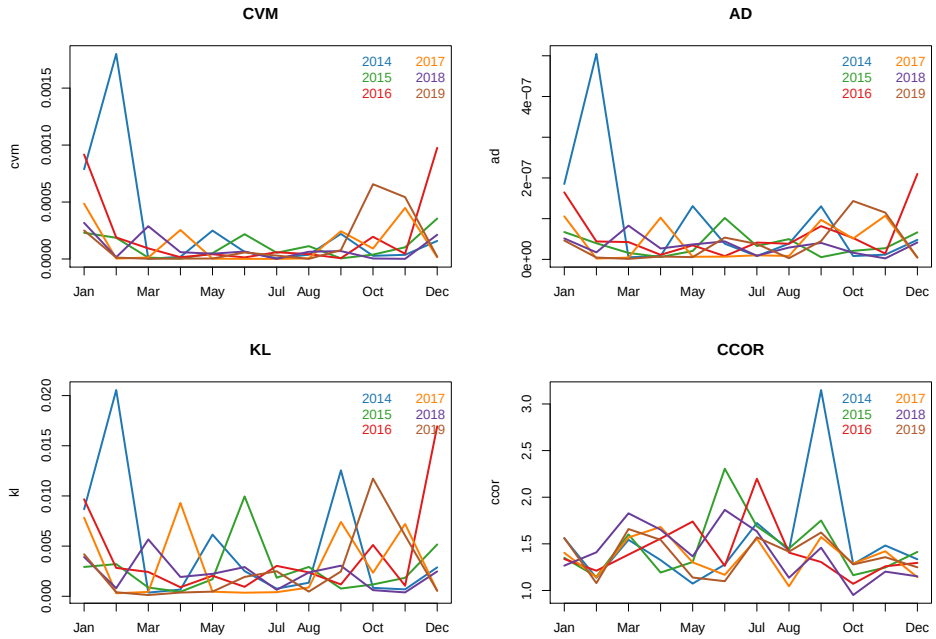


FIGURE 4.4 – Distances entre une année et la série entière au pluviomètre de BMO à 3 minutes : Cramer Von Mises (CVM), Anderson Darling (AD), Kullback Leibler (KL) et la distance basée sur la corrélation croisée (CCOR).

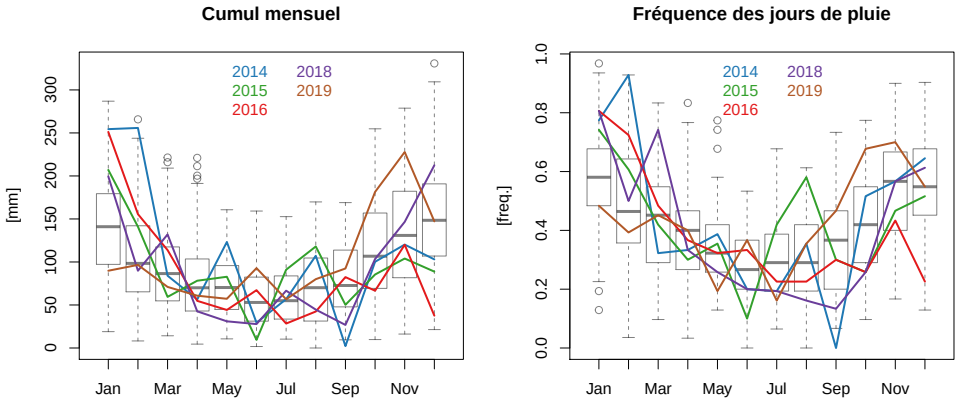


FIGURE 4.5 – Cumuls mensuels (gauche) et fréquence des jours de pluie (droite). Les boxplots montrent les données historiques 1945-2018 et les lignes les 6 dernières années à BMO.

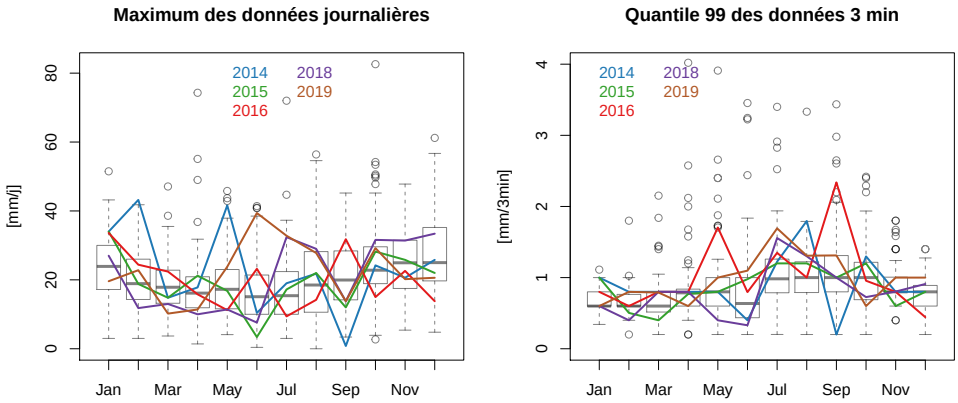


FIGURE 4.6 – Maximums journaliers (gauche) et quantile 99% de la partie positive des données à trois minutes (droite). Les boxplots montrent les données historiques 1945-2018 à droite et les 9 ans de données des pluviomètres du réseau d'Eau du Ponant. Les lignes représentent les 6 dernières années à BMO.

peuvent aussi être utilisées, mais il n'y a que neuf ans de données (donc moins représentatif de la climatologie). De plus les pluviomètres du réseau Eau du Ponant peuvent contenir des erreurs, il a donc été choisi de calculer le quantile 99% de la partie positive de la distribution au lieu du maximum. Les deux options sont montrées en Figure 4.6. Notez que pour le quantile 99% le graphique est zoomé, il y a une dizaine points compris entre 4 et 10 mm/3 min.

Le premier constat est que les données journalières et celles à trois minutes ne donnent pas du tout les mêmes informations quant aux pics d'intensités. Les données journalières représenteront mieux les évènements longs comme des fronts stratiformes alors que les données à trois minutes montreront plutôt les pics des évènements convectifs.

— 2018 a des maximums journaliers assez fort d'octobre à décembre, alors que les quantiles 99% à 3 minutes ne sont pas si forts que ça. Ça laisse penser à des évènements stratiformes de type front venant de l'ouest, avec un pic d'intensité pas si fort mais un évènement plutôt long. En juillet il y a un pic d'intensité avec un évènement particulièrement court, mais le pic n'est pas si fort que ça. En dehors de ce pic l'année 2018 a plutôt des maximums d'intensité faibles à moyens par rapport aux « normales » du mois.

— 2014 cumule en hiver à la fois des maximums journaliers très importants et des quantiles 99% à 3 minutes aussi forts par rapport aux « normales » du mois. On a donc à faire à des évènements qui ont à la fois un pic d'intensité fort et qui sont

aussi des évènements qui s'étalent dans le temps. Le mois de mai montre aussi un cumul journalier extrême et enfin en juillet il y un gros orage avec un évènement de 40 minutes pour 17 mm tombés (Fig. 4.2).

— 2019 est une année plutôt moyenne pour les quantiles 99% des intensités à 3 minutes, ce qui était attendu en voyant la série temporelle (Fig. 4.2). En juillet il y a tout de même un évènement assez intense qui s'étale sur une journée. Les maximums journaliers plutôt moyens sauf en été, l'été ne semblent donc pas présenter d'orage mais plutôt des évènements assez longs.

— Pour ce qui est des autres années, au niveau des quantile 99% à 3 minutes l'année 2016 montre 2 mois avec des évènements très intenses, et les données journalières semblent montrer des fronts hivernaux plutôt forts. 2015 est très moyenne à faible, que ce soit sur le maximum journalier ou sur les pics à 3 minutes.

4.3.4 Recommandations pour le choix d'une année

Un récapitulatif de tout ce qui a été observé sur les six années possibles est donné en Tableau 4.1. Ce tableau montre que 2016, 2018 et 2019 semblent être les années les plus appropriées. Toutefois elles ont toutes leurs défauts, on retiendra que :

— Si on veut une période très difficile : 2016 (hiver) ou 2019 (automne)

— Si on veut des orages : 2016

— Si on veut des gros fronts intenses : 2018 ou 2019

En rajoutant les informations sur les volumes déversés mesurés ou simulés, il est possible de constater que ça correspond bien à ce qui est dit sur les différentes années (2017-2015 volume faible, 2014 volume fort). 2019 est plus chargée en déversements, et 2016 et 2018 sont plutôt équivalentes. Il est peut-être possible de réfléchir de plusieurs façons :

— Quels sont les évènements qui devront être gérés ? orages ou fronts → 2016 ou 2018/2019

— Quelles sont les années les plus « classiques », c'est-à-dire qu'est-ce qui se retrouve souvent ? → un début d'automne avec deux mois de pluie constante en 2019 n'est peut-être pas aussi courant qu'un hiver bien chargé (2016/2018).

Dans tous les cas aucune des trois années ne saute aux yeux comme étant la plus adaptée, il sera alors d'autant plus important de tester le dimensionnement calculé avec l'année choisie sur les 5 années demandées par la réglementation.

Remarque : si il est facile de respecter de la réglementation, travailler avec 2014 permettrait alors d'obtenir un dimensionnement plus résilient.

Année	Rapport à la climatologie (74 ans MF)	Orages	Fronts	Autres remarques	Possible ?
2014	Forte (mais pas extrême)	Juil.-août, très intenses	Hiver très chargé	Les fronts hivernaux ont aussi des pics d'intensité forts pour des mois d'hiver. Les évènements intenses sont nombreux et répartis sur toute l'année.	Pour sécuriser
2015	Moyenne à faible	Aucun	Janv.	Moyenne sur toute l'année	Non
2016	Moyenne à forte	Mai, sept.	Janv. (fév.)	Les fronts hivernaux ne sont pas très intenses, ce sont des évènements moyens mais tout le temps (fréquence de jours de pluie très forte).	Oui
2017	Trop sèche			Risque de sous-dimensionner très important	Non
2018	Moyenne à forte	Peu nombreux, juil.	Déc.-janv.	L'été est très sec, et il y a en mars des pluies faibles mais en continu. L'année présente un peu de tout mais rien de très marqué.	Oui
2019	Moyenne à forte	Pas vraiment, juil.	Automne très chargé	C'est surtout l'automne qui est très pluvieux, avec une fréquence de pluie très élevée. Le début d'année est sec et l'été moyen.	Oui

TABLE 4.1 – Récapitulatif et conclusions sur le choix d'une année.

Retour des modélisateurs Le choix a été d'utiliser l'année 2018. Les pics orageux sont plutôt une problématique inondations, pour l'application ce sont surtout les gros fronts stratiformes qui sont adaptés pour une gestion des rejets et donc le projet de dimensionnement. La chronique sur 5 années intègre 2018/2019 mais aussi une année « orageuse » avec 2016 donc parait adaptée. La simulation de 2014 ne permet pas de respecter « facilement » la réglementation et est potentiellement trop sécuritaire ou avec un rapport coût bénéfice très important par rapport à 2018. Elle sera donc simulée pour avoir un critère de résilience du scénario et voir comment le scénario réagit à cette année très pluvieuse, sans pour autant être un critère de dimensionnement.

4.4 Spatialisation des données des pluviomètres à l'échelle des sous bassins versants

Les 8 pluviomètres répartis sur le territoire de Brest constituent une donnée spatialisée, au pas de temps de la mesure en réseau, soit 3 minutes. L'échantillonnage spatial n'est toutefois pas aussi détaillé que celui du radar (1 km^2). La qualité des données peut poser problème car les pluviomètres ne sont pas corrigés lors des opérations de maintenance et ils peuvent contenir beaucoup d'erreurs de mesure. On s'attend tout de même à ce que ce soit les données spatialisées les plus proches de la chronique utilisée pour le calage du modèle, étant donné que la donnée provient du même type de mesure (pluviomètre à bascule avec une précision de 0.2 mm).

Au vu de la qualité des données et de la forte discrétisation, la piste de l'interpolation des pluviomètres pour créer un champ plus détaillé n'a pas été explorée. Il a été choisi de simplement attribuer un pluviomètre à chaque sous bassin versant. Pour cela considérer le pluviomètre le plus proche n'est pas forcément le plus adapté, car la direction principale des fronts (SW $\rightarrow$ NE) et l'orographie (côte, Penfeld) peuvent jouer. Une façon simple de prendre en compte ces informations est de se concentrer uniquement sur la pluie. En effet avec les données radar toute la zone de Brest est couverte, il est donc possible d'extraire les chroniques des pixels dans lesquels se situent les pluviomètres. Chaque pixel de la zone pourra alors être comparé aux pixels des pluviomètres et se verra attribué un des pluviomètres. Ainsi la zone d'influence de chaque pluviomètre pourra être estimée.

Pour la comparaison des chroniques il a été choisi d'utiliser une simple corrélation, le but étant de savoir si globalement les évènements pluvieux arrivent au même moment. Les zones d'influence des pluviomètres ainsi calculées sont représentées en Figure 4.7. Les zones sont d'un seul bloc, ce qui est satisfaisant, et on peut voir à certains endroit l'effet du trait de côte (presqu'île de Plougastel, pointe de Roscanvel) ce qui montre l'intérêt de ne pas utiliser la distance.

4.5 Traitement des données radar

Comme ça a été dit en Section 4.2, on cherche à construire un jeu de données ayant 1) la structure spatio-temporelle d'un radar, car on ne peut pas obtenir une information fiable sur la structure spatio-temporelle de la précipitation à partir des pluviomètres à cause de l'échantillonnage spatial et des erreurs de mesure, et 2) l'échantillonnage temporel et la

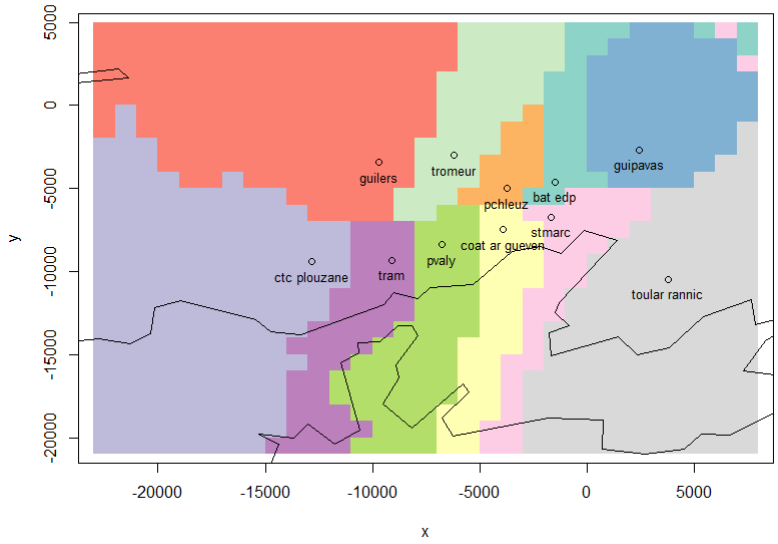


FIGURE 4.7 – Carte des zones d'influence des pluviomètres.

loi marginale d'un ou des pluviomètres, pour se rapprocher de ce qui a servi à caler le modèle hydraulique.

Les données radar étant à 5 minutes, on aura déjà un premier travail pour les passer à 3 minutes. Ensuite il est reconnu dans la littérature que les données radar nécessitent une correction pour pouvoir être comparées aux pluviomètres. Classiquement la correction du radar est faite par le calcul de biais, comme ça a été présenté en Section 4.2. Un des problèmes majeurs de ces méthodes est la non prise en compte des erreurs des pluviomètres. Or comme ça a été évoqué au Chapitre 1 le seul pluviomètre qui soit complètement fiable sur notre zone d'étude est le pluviomètre de Météo France à Guipavas, ce qui exclut les méthodes basées sur le krigeage.

4.5.1 Etape 1 : passage à 3 minutes

Les images radar sont réputées pour la reproduction du déplacement des cellules pluvieuses, ce qui a été confirmé sur nos données (Chap. 1). La vitesse de déplacement entre deux images permet d'interpoler entre ces deux pas de temps (e.g. Atencia et al., 2011).

Vitesse de déplacement

Dans la production du produit radar, la vitesse de déplacement est déjà utilisé dans les étapes de correction (cf. Chapitre 1 et Tabary, 2007) sous forme de champ d'advection. Il sert notamment à corriger la non simultanéité des mesures (le radar met 5 minutes à faire 360°) et l'effet de sous-échantillonage (estimation du cumul sur 5 minutes). Ce champ d'advection est estimé en cherchant le maximum de corrélation entre l'image déplacée et l'image d'origine, comme dans Tuttle et Foote (1990).

D'autres méthodes existent pour estimer un champ d'advection. On notera notamment la méthode du flux optique (Dérian et al., 2013). L'idée est d'utiliser une décomposition en ondelettes afin d'estimer le champ d'advection à différentes résolutions. Les vitesses sont estimées en minimisant l'erreur au carré entre l'image déplacée et l'image d'origine.

Enfin une méthode assez courante, notamment utilisée pour le champ d'advection 2PIR de Météo France est le suivi des cellules pluvieuses. L'algorithme 2PIR identifie les cellules avec un seuil, et pour chaque cellule le maximum de corrélation est cherché dans l'image suivante. Il y a donc autant de vitesses de déplacement que de cellules, et la vitesse de chaque pixel est calculée avec une somme pondérée par la distance du pixel aux cellules. Il existe de nombreux algorithmes permettant de suivre les cellules pluvieuses, comme par exemple le Storm Cell Identification and Tracking algorithm (SCIT) basé sur le suivi des centroïdes (Johnson et al., 1998), ou encore CELLTRACK (Kyznarová & Novák, 2009).

L'algorithme 2PIR étant facile à implémenter, il a été testé sur les images radar mais a donné beaucoup d'incohérences, notamment dues à la sortie de cellules pluvieuses de la zone observable.

Au lieu de chercher un champ d'advection, il est aussi possible de ne calculer qu'une seule vitesse pour toute l'image, comme c'est fait dans Luini et al. (2011). La taille de la zone de travail (environ 15 par 10 km) laisse penser que cette méthode pourrait suffire.

Détails de la méthode (cf. Figure 4.8) :

— Le maximum de corrélation est cherché entre l'image au temps $t + 1$ et l'image au temps t déplacée. Sur le schéma les cadres bleus et rouges sont les images déplacées correspondant aux vitesses de déplacement représentées par les flèches de même couleur. Le cadre rouge est le plus pertinent car on aura une meilleure corrélation avec l'image au temps t.

— Afin d'éviter les effets de bords, les zones extraites des images radar sont plus grandes que la zone de travail de 10 pixels tout autour. La vitesse maximum qui pourra être obtenue sera donc de 10 km/5 min soit 120 km/h. Sur le schéma la

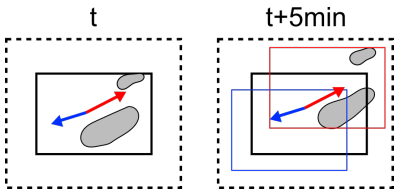


FIGURE 4.8 – Schéma explicatif de la recherche de la vitesse de déplacement. La zone d'intérêt est en trait plein noir, et la zone de recherche en pointillés.

zone d'intérêt est en trait plein, la zone de recherche en pointillés.

— Si au moins une des deux images est constituée entièrement de zéros la vitesse est non renseignée (NA), et si les deux le sont la vitesse est nulle.

Les codes R permettant d'estimer la vitesse sont disponibles sur disponible sur Github [1].

Il est attendu que les vitesses de déplacement soient cohérentes avec le vent. Des données de vent sont disponibles dans la base de données MeteoNet [2] (Larvor et al., 2020), qui couvre le quart NO de la France (et le quart SE) sur trois ans, de 2016 à 2018.

Les données mesurées à la station de Guipavas de vent moyen à 10 m au pas de temps six minutes ont ainsi été récupérées pour l'année 2018. La cohérence entre le déplacement des cellules pluvieuses et le vent à 10 m est surtout attendue au niveau des tendances, le vent comme les vitesses de déplacement estimées sont donc agrégés à trois heures. Pour ce faire, une moyenne est calculée sur les deux composantes en longitude et en latitude.

La Figure 4.9 se focalise sur le mois de janvier afin de voir en détails ce qu'il se passe. La vitesse et la direction (recalculées à partir des composantes) sont montrées pour le vent et le déplacement. La cohérence globale entre ces deux mesures est très satisfaisante, même si la vitesse du vent est plus lisse que celle du déplacement. Remarque : les directions manquantes correspondent aux cas où le déplacement est nul pour les deux composantes.

Pour avoir une idée plus globale le Tableau 4.2 donne la corrélation entre le vent et le déplacement, composante par composante. Cette corrélation est calculée sur le temps de pluie uniquement, et on travaille par saison donc par série de trois mois afin d'augmenter la longueur de la série. Bien que les corrélations soit légèrement plus faibles sur les mois d'été, elles restent très satisfaisantes toute l'année. L'estimation du déplacement des cellules pluvieuses reflète donc bien une réalité physique.

1. https://github.com/mbtgy/tools, cf. fonction « `motion` »
2. METEO FRANCE - Données originales téléchargées depuis https://meteonet.umr-cnrm.fr/, mis à jour le 30 janvier 2020

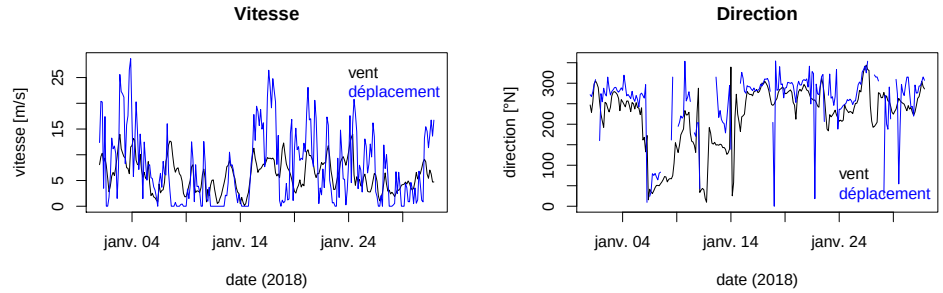


FIGURE 4.9 – Vitesse et direction du vent (noir) et des déplacements estimés (bleu) en janvier 2018.

	JFM	AMJ	JAS	OND
Composante N/S	0.62	0.56	0.51	0.58
Composante E/O	0.65	0.58	0.41	0.59

TABLE 4.2 – Corrélation entre le vent et la vitesse de déplacement sur les deux composantes (NS, OE), en fonction des saisons.

Il est ensuite intéressant d'étudier la capacité de prédiction du déplacement estimé. Il a été choisi d'utiliser la vitesse au temps t pour déplacer l'image $t+1$ et ainsi prédire l'image $t+2$. Comme il a été remarqué que le vent et le déplacement estimé étaient très proches, le vent est aussi utilisé pour déplacer l'image afin de montrer l'intérêt du déplacement. La persistance correspond à prédire l'image $t+2$ par l'image $t+1$ non déplacée. L'erreur quadratique moyenne

$$RMSE = \sqrt{\frac{1}{n}\sum_i (obs_i - mod_i)^2}$$

est calculée par image et les résultats obtenus sont présentés en Figure 4.10 pour le mois de janvier. L'estimation du déplacement améliore la prédiction, par rapport à la persistance mais aussi par rapport au vent. L'amélioration est surtout visible sur les fortes erreurs : la médiane baisse peu voire augmente, mais la moyenne (qui est plus influencée par les fortes valeurs) baisse. De plus, le quantile 99% est de 0.2 avec le vent ou la persistance quand il est de 0.1 avec le déplacement. Des résultats similaires sont observables sur les autres mois.

Les déplacements obtenus étant très saccadés au pas de temps 5 minutes, plusieurs filtres visant à les lisser ont été testés. De la même façon que précédemment c'est la capacité de prédiction qui a été étudiée. Il en est ressorti que les meilleurs résultats étaient

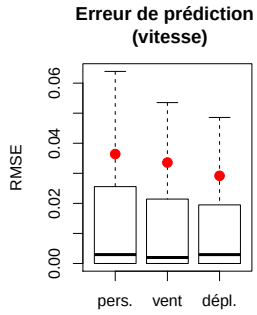


FIGURE 4.10 – Boxplots des erreurs de prédiction (RMSE) pour le mois de janvier, les points rouges montrant la moyenne.

obtenus en utilisant la médiane mobile des déplacements aux temps $t-2$ à $t+2$, donc sur 25 minutes. Toutefois l'amélioration reste légère par rapport au déplacement instantané, les deux options seront donc conservées pour la suite. Elle seront notées v_t et v_{med}.

Interpolation

L'interpolation, qui est schématisée en Figure 4.11, peut être faite dans deux sens. L'image au temps t est déplacée avec $1/5$ ème de la vitesse aux temps $t+1$ min, $t+2$ min, $t+3$ min et $t+4$ min. De la même façon en partant de l'image $t+5$ min une autre estimation de ces images interpolées est possible, en utilisant $-1/5$ ème de la vitesse.

La moyenne des deux interpolations est ensuite calculée. Afin de revenir au pas de temps 3 minutes, les images sont ensuite cumulées. Passer par une interpolation à 1 minute avant de cumuler permet de s'assurer que les images initiales sont présentes dans le produit final. De plus les fortes intensités seront ainsi mieux représentées.

Le code R permettant d'interpoler les images de 5 à 1 minutes est disponible sur Github [3].

Afin de vérifier que la méthode d'interpolation utilisée est valable, une chronique radar à 10 minutes est extraite des données en prenant une mesure sur deux. En utilisant la méthode expliquée précédemment (estimation du déplacement et interpolation), une chronique à 5 minutes est obtenue et peut donc être comparée aux données de base. Il est de nouveau possible de comparer les différentes vitesses utilisables : le vent, le déplacement v_t et la médiane sur 25 minutes du déplacement v_{med}. Pour la persistance la vitesse est

3. https://github.com/mbtgy/tools, cf. fonction « `interp` »

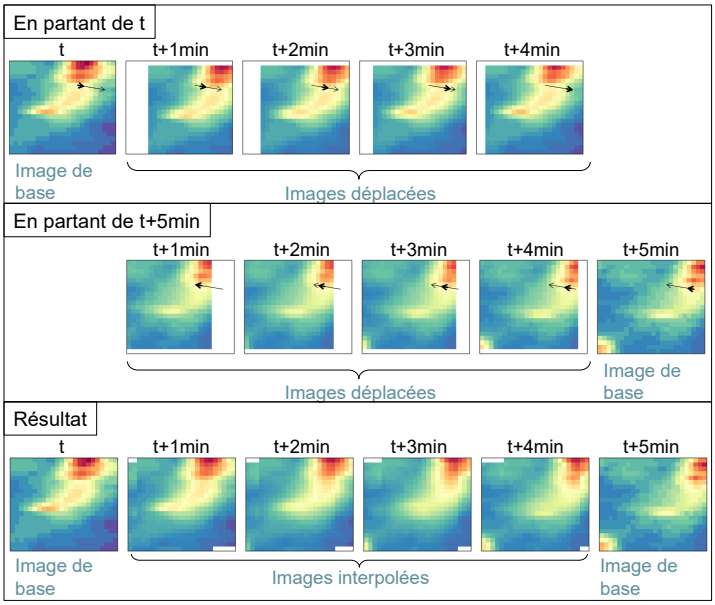


FIGURE 4.11 – Explication de l'interpolation.

égale à zéro, ce qui revient à prédire l'image $t+5$ min par l'image t. Les tests sont fait sur le mois de janvier et l'erreur de prédiction (RMSE) obtenue est montrée en Figure 4.12. Utiliser le déplacement des cellules pluvieuses donne de bien meilleurs résultats que la persistance et que le vent. Le déplacement instantané semble être légèrement plus adapté mais de nouveau la différence entre v_t et v_{med} n'est pas flagrante, les deux options seront donc testées sur la méthode dans sa globalité.

4.5.2 Etape 2 : correction en distribution

Une différence importante entre les données radar et les pluviomètres est la discrétisation, mais appliquer une simple discrétisation au radar n'est pas satisfaisant. Par exemple la probabilité de temps sec n'est plus respectée, elle augmente, ce qui est due à l'asymétrie de la distribution de la pluie. De plus l'interpolation à 1 minute a tendance à créer des images plus lisses, avec notamment plus de faibles intensités, le phénomène pourrait donc être amplifié.

Une méthode classique pour envoyer une distribution sur une autre est le quantile-quantile (QQ) mapping (e.g. Boé et al., 2007 ; Dobler et al., 2012). Le principe est le suivant : si on note X la variable aléatoire ayant pour cdf F_1, alors appliquer $x \mapsto F_2^{-1}(F_1(x))$ à X permet d'obtenir une variable aléatoire ayant pour cdf F_2. Les cdf sont généralement empiriques mais elles peuvent aussi être paramétriques. Ici c'est la version empirique qui

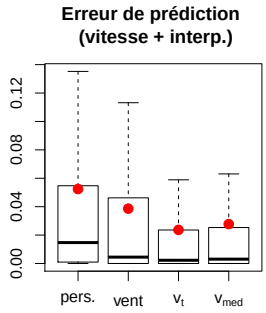


FIGURE 4.12 – Boxplots des erreurs de prédiction (RMSE) en recréant la chronique à 5 minutes à partir d'une chronique à 10 minutes du mois de janvier. Les points rouges montrent la moyenne.

sera utilisée, principalement car ça a été la plus rapide à mettre en œuvre. La méthode basée sur des lois paramétriques, qui utiliserait donc la loi développée dans le Chapitre 2, a été envisagée mais la plus petite mesure observable pose un problème car elle n'est pas la même dans les deux jeux de données. En effet utiliser un jeu de paramètre ajusté avec un $y_m = 0.2$ pour générer des données avec $y_m = 0.01$ changera la distribution obtenue.

Pour ce qui est de la cdf visée, il a été choisi de prendre un seul pluviomètre pour toute la zone. Le produit final sera donc

$$R_{tr} = F_P^{-1}(F_R(R))$$

où R est le radar interpolé à 3 minutes, F_R sa cdf empirique et F_P la cdf empirique du pluviomètre de référence choisi P.

Tous les pluviomètres seront testés et les cdf seront évaluées mois par mois. Si la chronique à laquelle sera appliquée la transformation concerne uniquement 2018, la transformation peut être apprise sur une chronique plus longue. Utiliser une longue série devrait permettre une meilleure estimation de la transformation. Comme évoqué dans le Chapitre 1 le radar est exploitable de 2014 à 2019 (la même plage temporelle sera prise pour les données des pluviomètres), il y a donc deux options : 1 mois ou 6 mois de données pour apprendre le QQ mapping.

4.5.3 Résultats

Dans la méthode expliquée dans les deux sections précédentes, il reste trois points à définir.

— La vitesse : instantanée v_t ou lissée v_{med}
— L'apprentissage du QQ mapping : sur 1 mois ou 6 mois
— Le choix du pluviomètre pour le QQ mapping : 8 possibilités

Toutes les possibilités seront testées et les pixels obtenus au niveau des pluviomètres seront comparés aux chroniques des pluviomètres. Plusieurs indicateurs seront calculés, qu'on peut classer en deux types :

— Les indicateurs de comparaison instantanée, avec tout d'abord des mesures de la synchronisation à travers la corrélation et une mesure de l'accord temps sec / temps de pluie. Ensuite pour les mesures d'erreur instantanée seront utilisées l'erreur quadratique (RMSE) et la RMSF (Craciun & Catrina, 2016)

$$RMSF = \sqrt{\frac{1}{n}\sum \quad 10\log\frac{mod}{obs}\Bigg)^2}$$

qui mesure les erreurs multiplicatives et qui contrairement à la RMSE est peu sensible aux fortes valeurs.
— Les indicateurs de comparaison en distribution avec des mesures sur la distribution à travers la moyenne et le maximum.

Tous ces indicateurs seront calculés par mois, pour chacun des 8 pluviomètres disponibles et pour chaque combinaison des trois points de la méthode évoqués à l'instant.

Tout d'abord les 3 choix de méthodologie sont comparés de façon globale. En Figure 4.13 sont montrés l'effet la vitesse utilisée et l'effet de la longueur de la série d'apprentissage du QQ mapping, et en Figure 4.14 est montré l'effet du pluviomètre utilisé pour le QQ mapping. Pour ce qui est de la comparaison entre v_t et v_{med}, aucun des indicateurs ne montre une réelle différence. L'utilisation d'un seul mois pour l'apprentissage du QQ mapping donne des erreurs instantanées (RMSE et RMSF) légèrement plus faibles, mais c'est surtout au niveau des indicateurs de la distribution que la différence est flagrante : utiliser 1 mois permet de mieux reproduire la distribution. Le pluviomètre utilisé pour le QQ mapping (Fig. 4.14) a peu d'impact sur les mesures de synchronisation (accord temps sec / temps de pluie, corrélation). Avec les erreurs instantanées et les mesures

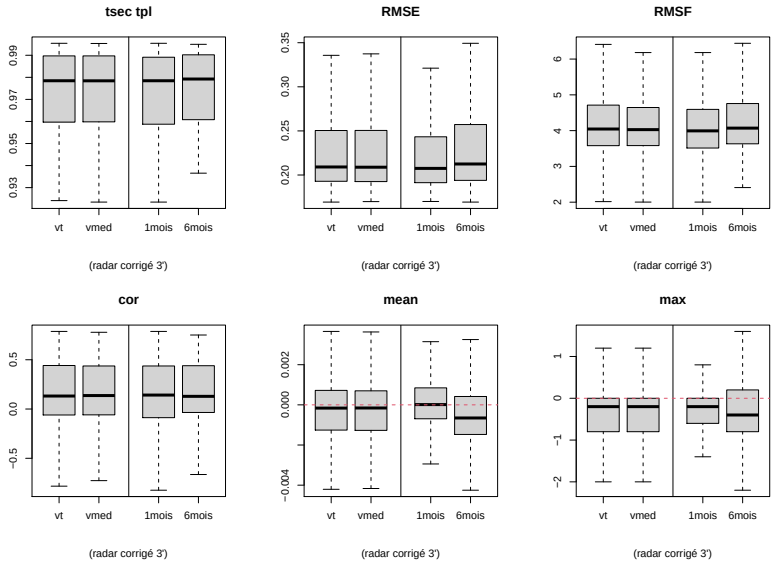


FIGURE 4.13 – Boxplots des différents indicateurs. A gauche comparaison entre v_t et v_{med} pour l'interpolation et à droite comparaison entre 1 mois et 6 mois pour le QQ mapping.

sur la distribution, certains pluviomètres semblent être à écarter : Guipavas et Bat EDP notamment.

En mettant les six indicateurs à minimiser et en calculant la moyenne comme indicateur global, la meilleure combinaison des 3 points de méthodologie est la suivante : v_t, 1 mois et le pluviomètre Tram. Pour avoir un point de comparaison, ces mêmes indicateurs seront calculés entre les pluviomètres et le radar brut (au plus petit pas de temps commun : 15 min) et le radar interpolé mais pas corrigé. Ainsi la Figure 4.15 permet de voir si la méthodologie améliore vraiment le radar. L'interpolation, qui a tendance à lisser les chroniques et à créer des faibles intensités, conduit en conséquence à une dégradation de l'accord temps sec / temps de pluie et une augmentation de la corrélation. De la même façon la RMSE a tendance à légèrement augmenter et la RMSF à baisser (surtout entre l'étape d'interpolation et de correction), ce qui laisse à penser que la correction en distribution élimine beaucoup de petites erreurs. Pour les indicateurs de la distribution, la moyenne est légèrement mieux reproduite et pour le maximum il n'y a pas vraiment de différence flagrante.

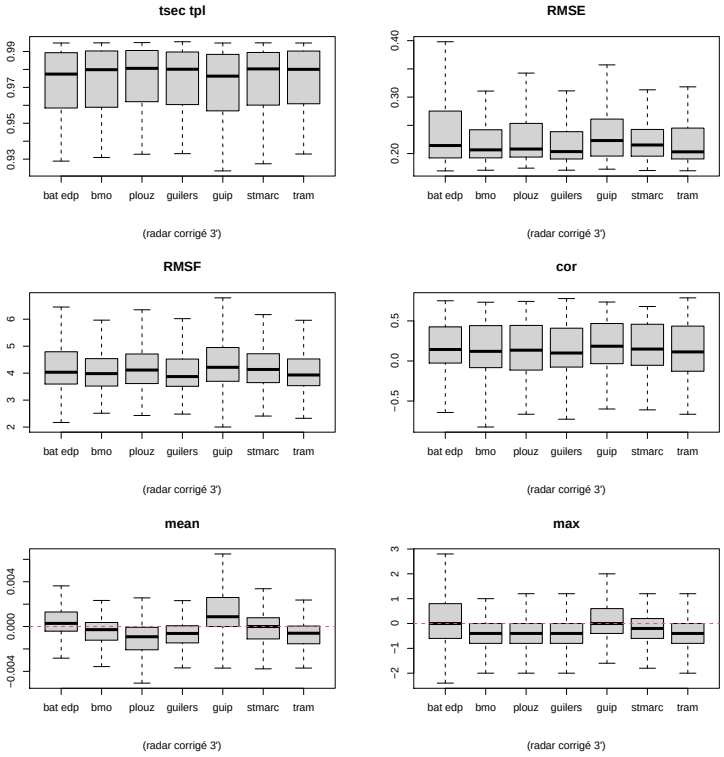


FIGURE 4.14 – Boxplots des différents indicateurs, effet du pluviomètre utilisé pour le QQ mapping.

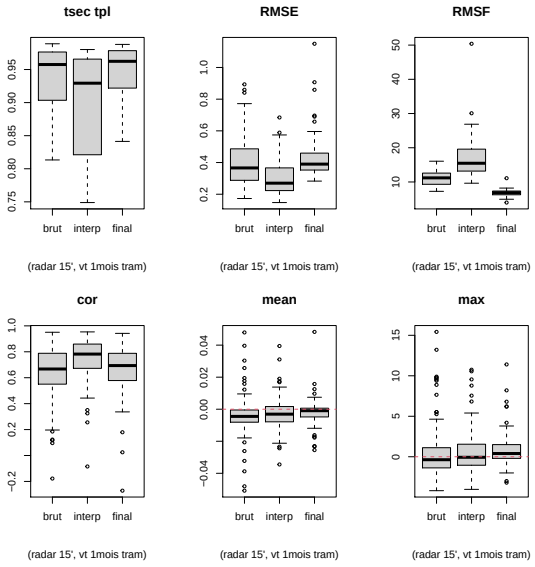


FIGURE 4.15 – Boxplots des différents indicateurs, comparaison du radar brut, radar interpolé et radar interpolé et corrigé (chroniques accumulées à 15 minutes).

4.6 Analyse de sensibilité

Comme ça a été dit plusieurs fois, on souhaite étudier la sensibilité du modèle hydraulique aux différentes caractéristiques de la pluie. L'analyse de sensibilité est un champ de recherche à part entière, et on pourra se référer à J. W. Hall et al. (2009) pour une étude des différentes méthodologies et de leurs utilisations pour les modèle hydrauliques. Dans notre cas l'analyse de sensibilité n'a pas été le centre de la thèse, aussi elle sera conduite de façon assez simple : chaque caractéristique de la pluie sera testée séparément et les effets seront étudiés par comparaison des sorties sur des statistiques descriptives.

4.6.1 Effets testés

Les caractéristiques de la précipitation dont on souhaite tester l'effet sur le modèle hydraulique sont : l'échantillonnage spatial (1 point, 11 points ou une grille de 1 km^2), l'échantillonnage temporel (3 ou 5 minutes), la discrétisation (0.2 mm ou 0.01 mm) et la distribution marginale (mesure issue du radar ou d'un pluviomètre). On utilisera les chroniques suivantes

(0) chronique de référence : radar interpolé

(1) effet de la spatialisation : la chronique de référence (0) mais uniquement aux points des pluviomètres d'Eau du Ponant (11 points)

(1bis) effet de la spatialisation : la chronique de référence (0) mais uniquement au point BMO

(2) effet de la distribution : chronique de référence (0) corrigée en distribution (Section 4.5.2)

(3) effet de la discrétisation : chronique de référence (0) arrondie à 0.2 mm

(3bis) effet de la discrétisation : chronique de référence (0) avec le fonctionnement d'une pluviomètre à bascule de précision 0.2 mm

(4) effet de l'échantillonnage temporel : chronique de référence (0) agrégée à 6 minutes

Pour chaque test, le but a été de changer une seule chose à la fois par rapport à la chronique de référence. La seule exception est la correction en distribution, qui est celle décrite en Section 4.5.2. La transformation QQ mapping envoie donc les données dans le domaine des pluviomètre, elle va par conséquent discrétiser les données à 0.2 mm. Il pourrait être envisagé d'utiliser un QQ mapping paramétrique avec le modèle présenté au

Chapitre 2, mais comme ça a été évoqué la plus petite mesure observable (y_m) pose un problème.

La discrétisation à été reproduite dans les chroniques test de deux façons. Tout d'abord un simple arrondi, puis une méthode qui permet de reproduire les effets d'accumulation dans un pluviomètre à bascule. Cette deuxième méthode peut s'écrire

$$g_{j+1} = \left(\left\lfloor \sum_{i=1}^{j+1} y_i \right\rfloor_{0.2} - \left\lfloor \sum_{i=1}^{j} y_i \right\rfloor_{0.2} \right)$$

où $\lfloor . \rfloor_{0.2}$ est la partie entière à une précision de 0.2, y_i est la précipitation non discrétisée à l'instant i et g_i est la précipitation reproduisant le fonctionnement d'un pluviomètre à bascule de précision 0.2 mm à l'instant i.

Tester l'effet de la spatialisation avec un seul point revient à appliquer une pluie constante sur tout le territoire. Utiliser 11 points revient à ce qui a été fait pour les données des pluviomètres en Section 4.4 et donc à mettre une pluie constante dans les zones d'influence montrées en Figure 4.7 en page 114.

Remarque : Afin de ne pas pas être affectés par les pixels au comportement étranges situés au sud du Plabennec (cf. Section 1.3.2 page 35), les résultats seront montrés à partir de mai 2018.

4.6.2 Présentation des sorties

Les 6 chroniques ont été utilisées en entrée du modèle hydraulique, et les débits de sorties ont été récupérés en quatre points dont les positions sont montrées sur la carte du réseau en Figure 4.1 page 98, ce sont les points rouges qui sont nommés. Ces points ont été choisis car ils sont à la fois parmi les points de mesure les plus fiables et car ils couvrent le territoire : Roosevelt est proche du centre (donc du pluviomètre de BMO utilisé pour le calage), La Guinguette est à l'est de la zone d'étude et CIN et Maison Blanche sont à l'ouest. Si les trois premiers points sont des déversoirs d'orages, Maison Blanche est un poste de relevage de la station d'épuration du même nom. Même si l'objet d'étude de MEDISA sont les déversements, il a été choisi de garder ce point car la mesure y est considérée comme très fiable. Les débits simulés (ou mesurés) à ce point sont donc très différents des autres DO, ce qu'on constate sur la Figure 4.16 qui montre des exemples d'évènements obtenus aux 4 points de calage : les simulations à Maison Blanche

contiennent beaucoup de paliers, aussi la chronique a été lissée sur une heure (moyenne glissante), et on voit nettement le fonctionnement des pompes (environ 6 heures).

La Figure 4.16 montre les débits obtenus avec les différentes chroniques d'entrée. Les formes des évènements sont globalement semblables entre les différentes entrées, mais on peut noter des écarts non négligeables. A Roosevelt (proche de BMO), les chroniques avec une information spatiale dégradée restent très proche de la chronique de référence. Des différences apparaissent notamment dans l'exemple pris pour CIN : en utilisant uniquement la grille spatiale des pluviomètres, les déversements sont fortement atténués, et avec uniquement le point de BMO le premier déversement a complètement disparu et le deuxième est décalé. Ce genre de décalage est typiquement ce à quoi on pouvait s'attendre avec la dégradation de l'information spatiale. Sur l'exemple de Roosevelt (qui est le plus proche de BMO), les principaux écarts sont dans la chronique où la distribution a été modifiée et celle qui a été discrétisée. La discrétisation semble fortement baisser voire faire disparaître les déversements dans tous les exemples, mais uniquement dans la version où les données sont arrondies. Quand l'accumulation typique des pluviomètres est reproduite, les différences avec la chronique de référence sont alors très faibles.

4.6.3 Vue globale

La Figure 4.17 montre quelques statistiques globales sur les sorties des différentes chroniques aux quatre points de calage : le volume annuel, la variance et la probabilité de déversement (cette dernière ne concernant pas le poste de relevage de Maison Blanche). On constate que si la discrétisation est prise en compte comme un arrondi les trois statistiques sont fortement impactées. Il a moins de déversement en quantité et en fréquence et la variance baisse également. Ce qu'on voyait dans les exemples se confirme donc et peut s'expliquer par le fait que la distribution de la précipitation étant asymétrique, arrondir reviendra toujours à créer plus de petites intensités (et moins de fortes intensités). Ces résultats montrent deux choses : 1) la précision des pluviomètres n'est pas une dégradation par rapport à celle du radar grâce à leur fonctionnement et 2) l'étape de correction en distribution qui a été expliquée en Section 4.5.2 ne convient pas car elle crée une discrétisation sans reproduire l'effet d'accumulation qu'on trouve dans les pluviomètres.

Le deuxième effet qu'on voit se démarquer en Figure 4.17 est la distribution. Comme ça a été mentionné, cet effet est « mélangé » à celui de la discrétisation. Toutefois l'effet est plutôt à l'inverse, ce qui peut s'expliquer par le fait que la correction marginale ne permet pas de reproduire le côté intermittent qu'on trouverait avec une discrétisation

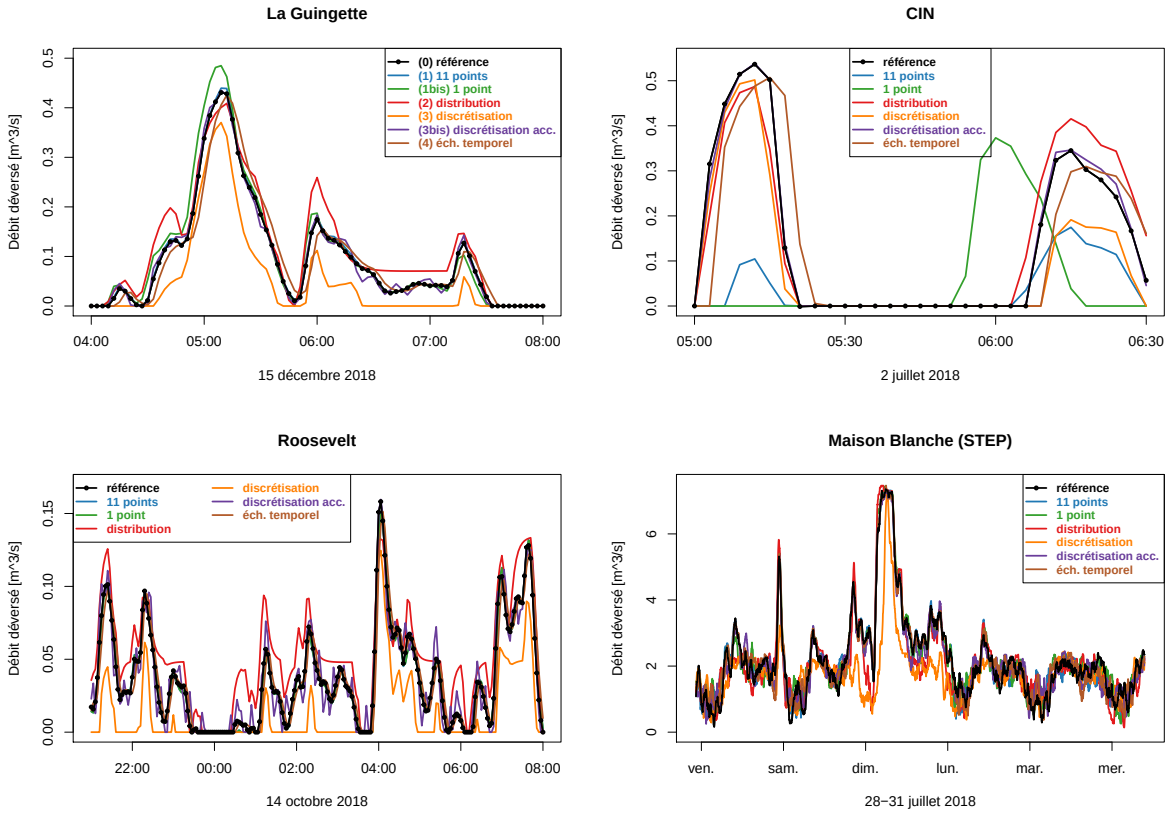


FIGURE 4.16 – Exemples d'évènements de déversement obtenus avec les différentes entrées. « Discrétisation acc. » : discrétisation accumulée, soit la chronique (3bis).

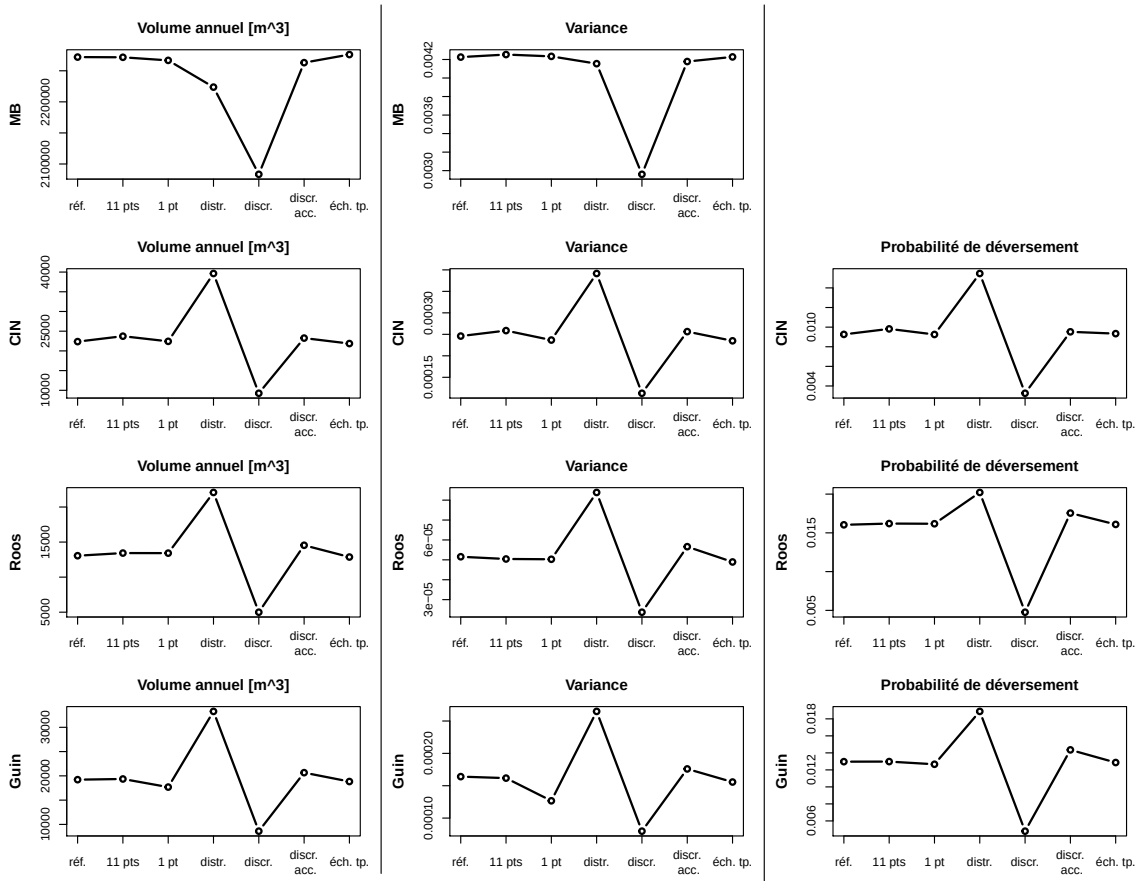


FIGURE 4.17 – Statistiques globales des déversements obtenus avec les différentes entrées

prenant en compte l'accumulation. Ici on retrouvera plutôt des blocs qui valent soit 0 soit 0.2 pour les faibles intensités, ce qui peut créer plus de déversements. Une autre hypothèse est bien sûr que c'est la forme de la distribution en elle-même qui est modifiée et qui crée plus de déversements.

Ces résultats se confirment avec la RMSE (erreur quadratique avec la série de référence), montrée en Figure 4.18. En plus de ce qui a été dit on note aussi un fort impact de la dégradation de la spatialisation pour les deux DO éloignés du pluviomètre de BMO. Cette différence n'étant pas visible sur les statistiques annuelles, on peut avancer l'hypothèse que ce qui est impacté est principalement le moment où les évènements arrivent, la synchronisation autrement dit. Toutefois l'augmentation de la RMSE n'est présente que si on utilise un seul point, aussi la grille des 11 pluviomètres d'Eau du Ponant semble être une résolution suffisante pour le modèle hydraulique. Les résultats de l'échantillonnage temporel ne semble pas montrer de différence entre 3 et 6 minutes.

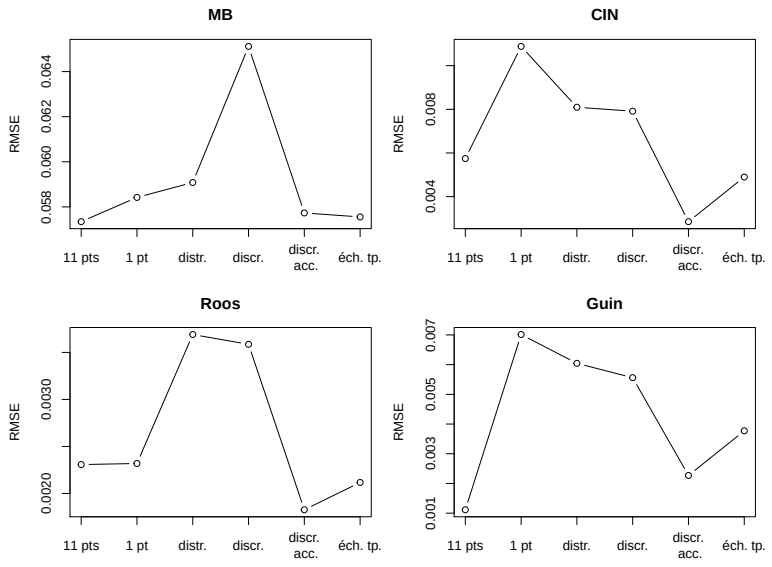


FIGURE 4.18 – Erreur quadratique moyenne entre le test de référence et les différents tests

La corrélation a été calculée entre les différentes sorties et la chronique de référence, au pas de temps 3 minutes et à 1 heure. Les résultats en Figure 4.19 montrent qu'en effet la spatialisation a un fort impact sur la synchronisation : une baisse de corrélation est notamment marquée à La Guinguette et à CIN (les deux DO éloignés de pluviomètre de BMO) dans le cas où on n'utilise qu'un point. De nouveau la grille des 11 pluviomètres semble suffisante, mais pas un seul point. Cet impact est autant visible sur les données à 3 minutes qu'à une heure. Pour tous les points la corrélation à une heure est légèrement affectée par la modification de distribution et par la discrétisation lorsqu'elle ne prend pas en compte l'accumulation.

L'effet très marqué de l'arrondi des données à 0.2 à Maison Blanche peut s'expliquer par la présence de paliers dans les données qui peuvent être plus fortement impactés par la discrétisation. Finalement la corrélation globalement faible à Maison Blanche à 3 minutes s'explique aussi par la nature des données qui contiennent beaucoup d'effets de seuils et de variations dues aux pompes.

4.6.4 Variations saisonnières

On se demande ici si certains effet sont plus marqués suivant la période de l'année considérée. Il faudra garder en tête qu'on a une seule année sur laquelle se baser.

Les volumes maximums et probabilités de déversement mensuels sont montrés en Fi-

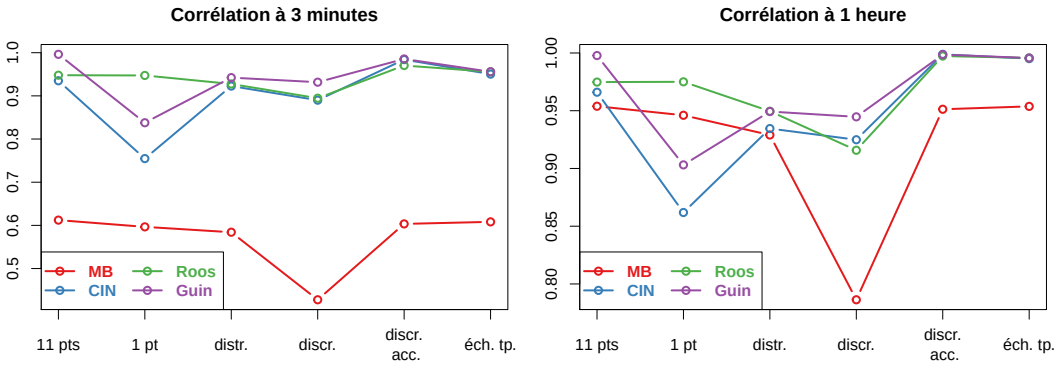


FIGURE 4.19 – Corrélation entre le test de référence et les différents tests

gures 4.20, 4.22 et 4.21. L'échantillonnage temporel ne semble pas avoir d'effet visible sur aucune de ces statistiques, quelque soit le mois de l'année.

L'autre effet qui reste globalement toujours très faible est celui de la discrétisation lorsque celle-ci est faite en reproduisant le fonctionnement d'un pluviomètre à bascule. La discrétisation avec un simple arrondi affecte les volumes et probabilités de déversement (Fig. 4.20 et 4.21) fortement toute l'année. L'hiver la différence avec la référence est plus marquée, ce qui peut s'expliquer par le fait que c'est surtout dans ces périodes que les longues pluies à faibles intensité se produisent. En effet si l'intensité est inférieure à 0.1 mm, ces précipitations disparaissent et une grande partie des effets d'accumulation dans le réseau peut disparaître. Une autre explication est simplement le fait que les déversement sont plus nombreux et plus importants l'hiver.

Les maximums mensuels (Fig. 4.22) sont surtout affectés par la spatialisation. On peut voir apparaître pour la première fois des différences importantes entre la grille des pluviomètres et la référence à Roosevelt et à CIN, notamment sur les mois d'été, ce qui semble indiquer qu'une grille de 11 points ne suffit pas pour bien représenter certains évènements orageux. En utilisant un seul point les maximums sont fortement changés sur tous les déversoirs. On s'attend à ce que les maximums soient sous-estimés, car un seul point de mesure à de fortes chances de manquer le pic d'intensité d'un évènement pluvieux qui est généralement peu étendu spatialement. Mais on note que les maximums peuvent aussi être légèrement surestimés en hiver, auquel cas on peut penser que quand le pic est perçu par le point de BMO il est appliqué à toute la zone, ce qui amplifie le maximum de débit déversé.

La distribution a un impact plus faible sur les maximums (Fig. 4.22) mais il est tout de même marqué pour certains mois d'été (août, septembre). Étant donné que sur ces mois

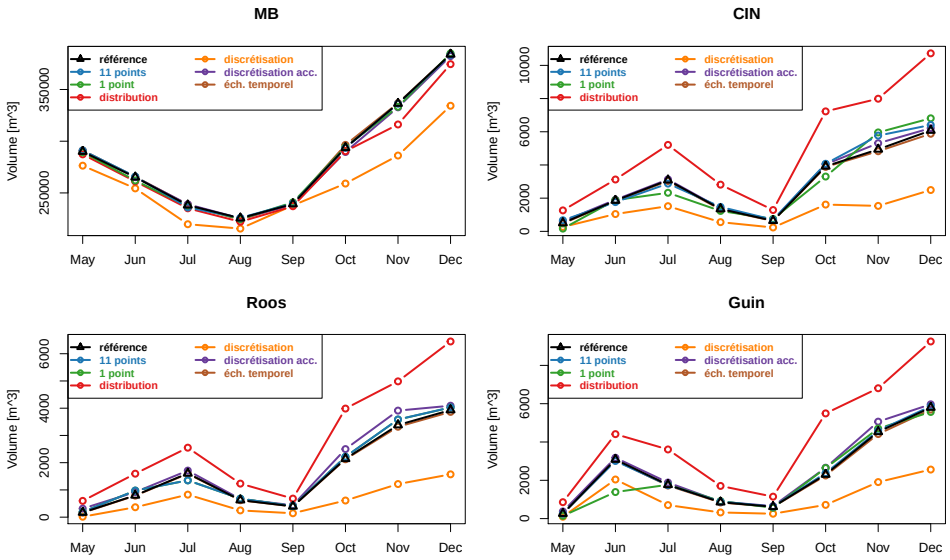


FIGURE 4.20 – Volumes mensuels obtenus avec les différentes chroniques d'entrée

l'effet de la discrétisation est lui aussi non négligeable, on ne peut de nouveau pas savoir à quoi attribuer cette différence. Pour le mois de juillet à Roosevelt, la discrétisation n'a pas d'effet, on peut donc dire que l'effet de la distribution est bien dû uniquement a une queue de distribution dans ce cas plus lourde dans la chronique de référence que dans le pluviomètre utilisé pour le QQ mapping.

4.6.5 Conclusion sur la sensibilité du modèle hydraulique

Un premier point qui ressort de cette analyse de sensibilité est l'importance de la façon de prendre en compte la discrétisation. Le fonctionnement des pluviomètres fait que leur précision ne conduit pas à une dégradation de la donnée. Il apparaît donc très important de reproduire ce fonctionnement si on est amené à discrétiser les données radar, ce qui n'est actuellement pas le cas avec la méthode décrite en Section 4.5.2.

La distribution a aussi un effet non négligeable, mais il n'est pas clair si cet effet est dû à la forme de la distribution ou simplement à la discrétisation et au fait que la transformation ne prenne pas en compte l'aspect dynamique des précipitations.

Au niveau de la spatialisation, utiliser la grille des pluviomètres semble suffisant, par contre un seul point n'est pas suffisant et a notamment un impact fort sur la reproduction des pics de déversement (intensité et synchronisation). Les volumes mensuels et annuels sont eux peu impactés par la résolution spatiale.

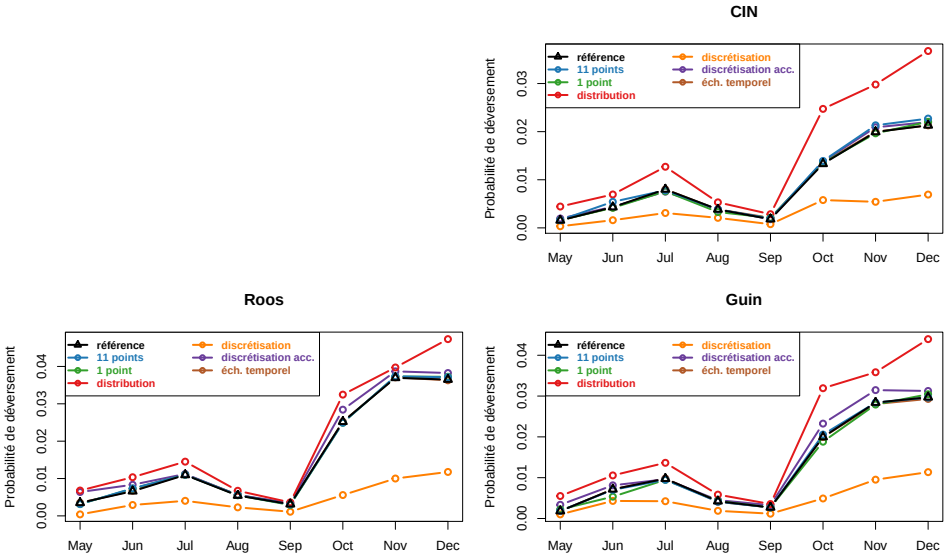


FIGURE 4.21 – Probabilité de déversement mensuelles obtenues avec les différentes chroniques d'entrée

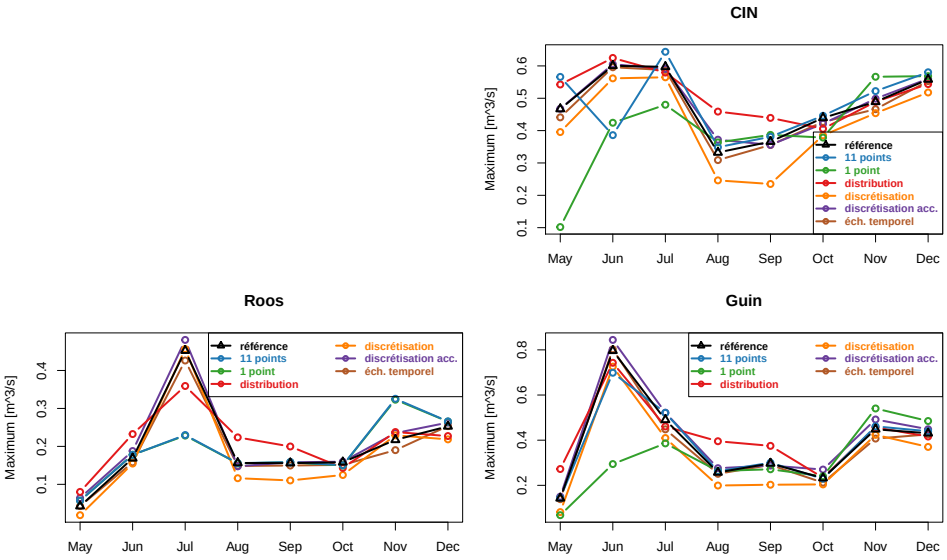


FIGURE 4.22 – Débits déversés maximums mensuels obtenues avec les différentes chroniques d'entrée

Finalement l'échantillonnage temporel ne montre pas de différence entre des données à 3 minutes et des données à 6 minutes, ce qui semble indiquer qu'il serait possible d'utiliser les données radar à 5 minutes, ou encore le pluviomètre de Météo France qui est à 6 minutes.

D'un point de vue pratique, il ne parait pas envisageable d'utiliser les données radar corrigées (cf. Section 4.5) en entrée d'un modèle calé sur le pluviomètre de BMO, car on a montré un effet de la spatialisation et de la correction en distribution.

L'interpolation ne semble a priori pas nécessaire pour la sensibilité du modèle. Toutefois pour le QQ mapping empirique il est nécessaire d'avoir des données au même pas de temps. Il pourrait être envisagé d'utiliser le QQ mapping paramétrique avec le modèle présenté au Chapitre 2, car il a été montré que la variation des paramètres de modèle GP méta-Gaussien avec l'agrégation temporelle était suffisamment lisse et monotone pour qu'on puisse envisager d'extrapoler les paramètres à d'autres pas de temps que celui de la mesure. Mais comme ça a déjà été dit la discrétisation et notamment la plus petite mesure observable restent un problème qui empêche d'utiliser un QQ mapping paramétrique.

Dans l'idéal il faudrait travailler avec des données les moins dégradées possibles, et il serait donc recommandé de recaler le modèle hydraulique avec des données : avec une précision de 0.01 mm (radar) ou une précision à 0.2 mm qui reproduise le fonctionnement des pluviomètres, avec une grille spatiale au moins aussi détaillée que celle des 11 pluviomètres d'Eau du Ponant et un pas de temps d'au moins 6 minutes. Pour la distribution entre le radar et les pluviomètres on est pas surs qu'il y ait une « bonne » et une « mauvaise » solution, même si la littérature s'accorde plutôt à accorder plus fiabilité à la distribution d'un pluviomètre. Il faut principalement que le calage et les tests soient fait sur des données ayant la même distribution.

4.7 Comparaison aux mesures sur le réseau hydraulique

Étant donné qu'on dispose de mesure des débits déversés et qu'on a pu construire différents jeux de données de précipitation dont on a montré que certaines caractéristiques impactaient fortement les déversements, il est intéressant de regarder si la mesure en réseau est mieux reproduite en utilisant comme forçage les jeux de données proposés en

Section 4.4 et 4.5.

Toutefois il faut garder en tête que le modèle hydraulique n'a pas été recalé car c'est un processus très long. Il est donc calé sur les données du pluviomètre de BMO, il est attendu que cette option soit favorisée puisqu'il est possible que le calage de certains paramètres du modèle permettent de corriger les erreurs dues au fait que la chronique de pluie ne soit « pas la bonne ». En forçant le modèle avec des données radar, les différences qui seront observées ne pourront pas uniquement être attribuées à l'aspect spatial mais pourront aussi être dues à une différence de distribution.

On dispose des quatre mêmes points de mesure que pour l'analyse de sensibilité, on rappelle qu'ils ont été choisis car considérés comme fiables et répartis spatialement de façon à représenter l'ensemble du territoire. Pour ces quatre points il y donc d'un coté la mesure en réseau et de l'autre les résultats du modèle hydraulique forcé avec 1) le pluviomètre BMO, 2) les 8 pluviomètres spatialisés, 3) le radar brut à 5 minutes ou 4) le radar corrigé.

Remarque : Ces quatre forçages sont différents de ceux utilisés pour l'analyse de sensibilité, ils correspondent aux produits décrits en Sections 4.4 et 4.5.

Deux exemples de séries de déversement sont montrées en Figure 4.23. La Guinguette, tout comme les deux autres points de mesure qui ne sont pas montrés, est un déversoir d'orage. La plupart du temps il y a donc un débit nul déversé, et lors des forts évènements pluvieux les déversements ont une distribution asymétrique à gauche (Fig. 4.24a). Maison Blanche n'est pas un DO mais un point de mesure en entrée de station d'épuration, la série est donc très différente des trois autres points. D'autre part la sortie de modèle à Maison Blanche n'est disponible qu'à un pas de temps de 10 minutes. Ce point a été conservé car c'est une mesure considérée comme très fiable, mais elle ne pourra pas être utilisée de la même façon que les DO.

4.7.1 Incertitude de la mesure

Les différents points de mesure ont été étudiés par l'entreprise 3D Eau dans le but d'estimer l'incertitude de la mesure. Leur étude prend en compte l'erreur de mesure de hauteur et l'erreur amenée par la loi hauteur-débit. Un exemple d'incertitude obtenue pour les déversements positifs est représentée en Figure 4.24a. L'incertitude est très forte pour les faibles débits, et le saut visible vers 0.6 m^3/s est dû à un changement de loi hauteur-débit.

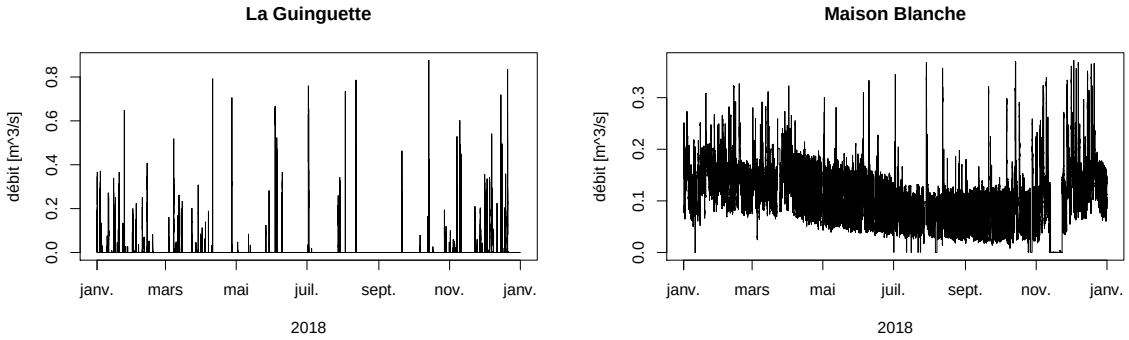
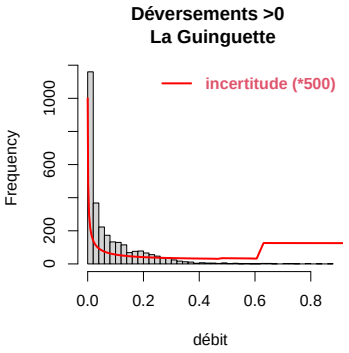
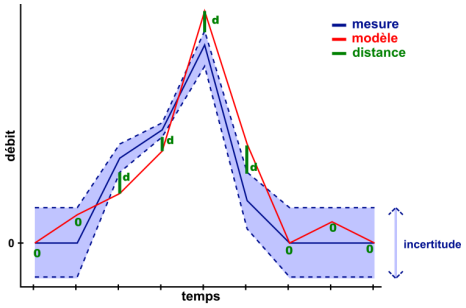


FIGURE 4.23 – Déversement au DO La Guinguette et débit d'entrée de la STEP Maison Blanche en m^3/s.



(a) Histogramme des déversements positifs à La Guinguette. La ligne rouge montre l'incertitude de la mesure.



(b) Schéma de la prise en compte des incertitudes dans le calcul d'une distance instantanée entre la mesure et le modèle.

FIGURE 4.24 – Gestion de l'incertitude

Pour les zéros, étant donné qu'on ne peut pas appliquer un pourcentage d'incertitude, il a été choisi de procéder comme suit :

— Si la mesure est entre deux zéro, l'incertitude est nulle.
— S'il y a un non nul autour de la mesure, on prend l'incertitude du plus petit débit de la table d'incertitudes fournie par 3D Eau.

Lors du calcul d'une distance instantanée entre la mesure et le résultat du modèle, l'incertitude peut être prise en compte de deux façons. En premier lieu, elle peut être utilisée comme un poids, mais cette méthode donnerait un poids important aux forts débits et minimiserait l'importance des faibles débits, ce qui n'est pas forcément souhaitable. C'est donc la deuxième méthode qui sera utilisée : l'idée est de prendre la distance la plus

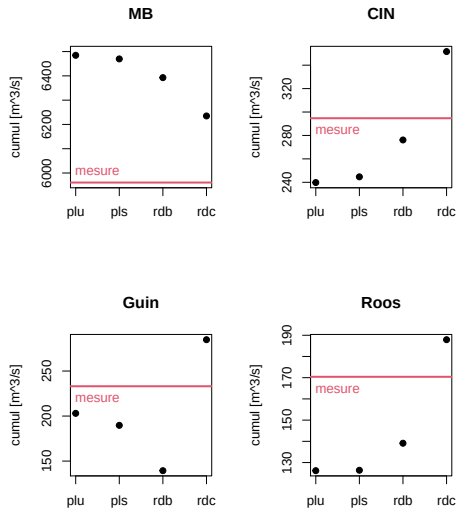


FIGURE 4.25 – Cumuls annuels des déversements aux différents points de mesure. « plu » : le pluviomètre BMO, « pls » : les 8 pluviomètres, « rdb » : le radar brut à 5 minutes et « rdc » : le radar corrigé.

faible entre le modèle et l'intervalle de la mesure défini par l'incertitude. La Figure 4.24b illustre cette méthode.

4.7.2 Résultats

Une mesure assez importante en modélisation des réseaux d'assainissement est le volume annuel déversé. La Figure 4.25 montre donc les cumuls annuels obtenus avec les différents forçages en 2018 (NB : à Roosevelt, les 20 derniers jours de l'année ne sont pas comptés à cause d'erreurs de mesure). C'est le radar corrigé qui semble être le plus à même de reproduire le cumul annuel. A part à La Guinguette, le radar brut et les pluviomètres spatialisés représentent aussi une amélioration par rapport au pluviomètre unique de BMO. Cette information va donc dans le sens d'une utilité de la prise en compte de l'aspect spatial de la pluie mais comme on va le voir par la suite l'amélioration n'est visible que sur le cumul annuel.

La Figure 4.26 donne tout d'abord à gauche l'erreur quadratique (RMSE). L'incertitude est prise en compte comme ça a été expliqué (Fig. 4.24b), et on constate que le pluviomètre unique reproduit mieux la chronique de mesure que les autres forçages proposés. Les pluviomètres spatialisés donnent des résultats assez proches, ce qui peut laisser

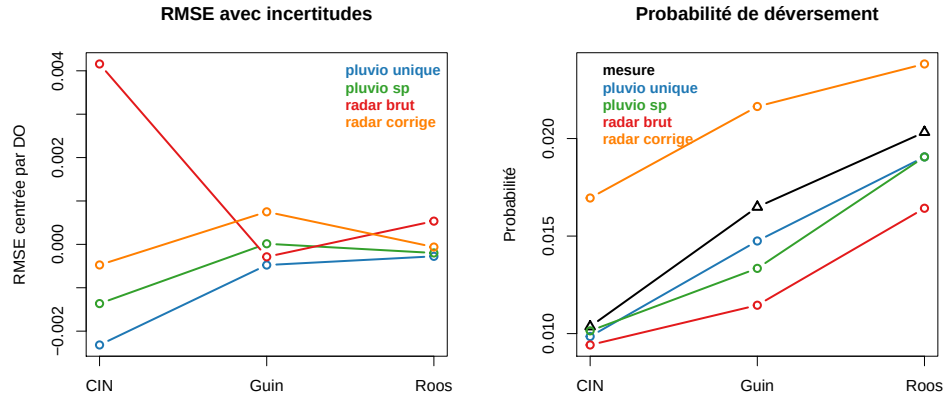


FIGURE 4.26 – RMSE centrée par point de mesure (gauche) et probabilité de déversement (droite), dans les deux cas en prenant en compte l'incertitude.

penser que l'écart peut être dû au fait que le modèle est calé sur le pluviomètre unique. Le radar brut donne un très mauvais résultat au déversoir d'orage de CIN, ce qui est surprenant car le volume annuel était plutôt bon. On peut supposer que c'est la synchronisation qui n'est pas bonne.

La Figure 4.26 montre aussi à droite les probabilités de déversement obtenues avec la mesure (en noir) et avec les forçages. L'incertitude est prise en compte en considérant que les deux sont d'accord si la mesure est égale à 0 et que le modèle est inférieur à l'incertitude du plus petit débit de la table fournie par 3D Eau (ou au contraire que la mesure est supérieure à 0 et le modèle supérieur à cette incertitude). Il est intéressant de noter que les forçages issus des pluviomètres ont tendance à légèrement sous-estimer la probabilité de déversement. Etant donné que 1) les pluviomètres spatialisés font le même score que le pluviomètre de BMO et 2) l'analyse de sensibilité n'a pas montré d'impact de la discrétisation quand on mime le fonctionnement d'un pluviomètre, on peut émettre l'hypothèse que cette sous-estimation vient du calage du modèle hydraulique.

Le radar brut et le radar corrigé donnent des résultats très différents. On constate que les volumes et probabilités de déversement sont globalement trop faibles avec le radar brut (Fig. 4.25 et 4.26 (droite)), et la RMSE est notamment très forte à CIN. Etant donné qu'on a pu montrer avec l'analyse de sensibilité que la résolution temporelle entre 3 et 5 minutes n'avait pas d'impact sur les sorties, on sait que ces différences entre radar brut et corrigé sont dues à la correction en distribution par le QQ mapping. On retrouve donc l'augmentation des volumes et des probabilités de déversement montrée avec l'analyse de

sensibilité. Cette augmentation due au QQ mapping est plutôt à l'avantage du produit radar corrigé.

4.8 Conclusion

Conclusion pour le projet MEDISA : sans re-calage du modèle hydraulique, le pluviomètre unique est recommandé car les résultats avec les autres forçages s'éloignent de la mesure. Le calage étant un processus très long et qui a dû être bouclé tôt dans le projet, le re-calage n'a pas pu être fait. Il a toutefois été montré que l'amélioration par la prise en compte de la spatialisation était un phénomène plutôt marginal, qui concerne surtout les pics d'intensité. La reproduction des pics n'est pas centrale dans le contexte de dimensionnement qui vise surtout à gérer les faibles pluies.

Les pluviomètres spatialisés donnent des résultats très proches du pluviomètre unique, ce qui donne un bon espoir qu'ils puissent être meilleurs s'il n'y avait pas l'avantage dû au fait que le calage est fait sur le pluviomètre unique. Il serait intéressant de caler le modèle sur ce forçage (comme en plus c'est une donnée disponible facilement par la suite pour EDP). Mais il faudrait alors gérer les erreurs dans ces pluviomètres, qui ne sont pas aussi fiables q ue B MO. L 'analyse d e s ensibilité m ontre q ue l 'augmentation d e l a résolution spatiale permettra principalement de mieux représenter les pics des déversements (en intensité et en synchronisation).

Les données radar détériorent trop les sorties pour être utilisées telles quelles. Il est difficile de savoir si c'est dû au calage ou au forçage en lui-même, car les effets de la correction en distribution et de la discrétisation n'ont pas pu être correctement séparés dans l'analyse de sensibilité. Il est intéressant de noter que les modélisateurs en charge du calage du modèle hydraulique ont fait le retour suivant : sur les évènements utilisés pour le calage les données radar corrigées améliorent beaucoup les résultats, et elles ont permis de mieux comprendre ce qui se passait dans le modèle hydraulique. Toutefois au niveau d'une année entière on ne retrouve pas cette amélioration, et en ajoutant à ça le fait que les données radar sont plus difficilement accessibles (payantes, et une correction doit être faite), il n'est pas conseillé à Eau du Ponant de refaire le calage avec des données radar.

Conclusion

Résumé et discussion

Un premier travail de cette thèse a consisté à étudier les données de précipitation de la zone de Brest. La station météorologique de Météo France située Guipavas dispose d'observations journalières depuis 1945 et d'observations à un pas de temps de 6 minutes depuis 2006. L'instrument de mesure est un pluviomètre à bascule de précision 0.2 mm. Les données ne présentent a priori pas ou peu d'erreurs de mesure, car elles sont corrigées au jour le jour. Les données journalières s'étendant sur une longue période (74 ans), elles permettent de mieux étudier la distribution de la pluie et de se familiariser avec le climat de la zone d'étude.

Eau du Ponant entretient un réseau de 11 pluviomètres à bascule de précision 0.2 mm et avec un pas de temps de 3 minutes répartis sur Brest Métropole. Ils sont donc souvent installés en ville, sur des toits ou dans des cimetières. Les données sont disponibles à partir de 2010 et contiennent beaucoup d'erreurs de mesure : depuis 2014 elles ne sont plus corrigées en temps réel et tous les pluviomètres de ce réseau n'ont pas la même qualité de données.

Météo France dispose d'un réseau de radars météorologiques qui permet une évaluation de la précipitation sur un grille spatiale de 1 km^2 à un pas de temps de 5 minutes avec une précision de 0.01 mm. Ces données sont un produit très transformé qui est généralement considéré comme moins fiable que les pluviomètres pour la représentation de l'intensité de pluie. Toutefois ces données permettent une meilleure estimation de l'aspect spatial de la pluie, notamment en représentant particulièrement bien les déplacements des cellules pluvieuses, bien visibles sur les images radar.

L'étude des différentes sources de données a mis en évidence l'intérêt de créer un produit qui les combine en tirant le meilleur de chaque source et en prenant en compte les erreurs spécifiques aux instruments de mesure.

> Un produit « idéal » aurait notamment la structure spatio-temporelle du radar et la distribution d'un pluviomètre.

Dans le but de créer un modèle d'assimilation de données pour combiner les sources de données, une première étape a été d'étudier la distribution marginale des précipitations à un pas de temps de quelques minutes. Les modèles les plus classiquement utilisés (loi Gamma, log-normale, exponentielle, etc.) ont été testés et ont montré leurs limites, ne pouvant en particulier pas reproduire la queue de la distribution. Parmi les modèles existants dans la littérature la classe des modèles méta-Gaussiens a paru particulièrement intéressante car en liant la précipitation à une variable latente normale ces modèles permettent d'utiliser toutes les méthodes statistiques développées dans le cas Gaussien. Ces modèles se marient particulièrement bien au cadre de l'assimilation de données et permettraient par exemple d'utiliser des méthodes comme les filtres de Kalman.

De nombreux modèles méta-Gaussiens ont été testés sur les différentes données. La plupart ne donnaient pas un bon ajustement, principalement au niveau de la queue de la distribution. Afin de développer un modèle méta-Gaussien qui conviennent au données à un pas de temps de quelques minutes, les propriétés théoriques des queues inférieures et supérieures de la distribution ont été étudiées. Le modèle proposé est

Modèle GP méta-Gaussien

$$Y = 0 \times \mathbb{1}_{X<0} + \psi(X) \times \mathbb{1}_{X\geq 0}, \quad \text{avec } X \sim \mathcal{N}(\mu, 1) \text{ et}$$

$$(x) = y_m + \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2},$$

où $\mathbb{1}_A$ est la fonction indicatrice égale à 1 si la condition A est vraie, et égale à zéro sinon. La pluie est notée Y, X est une variable aléatoire Gaussienne de moyenne μ et de variance 1, y_m est la valeur minimale qui peut être mesurée.

Ce modèle a pour particularité d'avoir une queue de distribution équivalente à une loi de Pareto généralisée et une forme puissance en zéro, d'où le nom GP méta-Gaussien. Bien que théoriquement chaque paramètre soit lié à une partie différente de la distribution une assez forte dépendance entre les estimateurs des paramètres a été notée, ce qui limite l'interprétation des résultats, notamment pour les paramètres σ et α. Cette dépendance a toutefois été retrouvée dans tous les modèles adaptés aux précipitations à cette échelle de temps.

Le modèle a été ajusté sur de nombreux jeux de données, avec différentes échelles de temps (de quelques minutes à un mois), sur toute la zone d'étude et sur toutes les sources

de données. L'inférence a montré l'importance de prendre en compte la discrétisation des données à un pas de temps aussi faible, aussi une vraisemblance discrète a été utilisée. Les moments de ce modèle étant disponibles la méthode des moments a aussi été abordée, mais elle posait le problème du choix de l'ordre des moments, certains étant infinis ils était nécessaire d'avoir un première estimation du paramètre de queue ξ. La bonne performance de la vraisemblance discrète a mené à ne pas pousser plus loin l'utilisation des moments.

L'extension du modèle GP méta-Gaussien au cas multivarié a été partiellement abordée. Différentes méthodes d'inférence basées sur la vraisemblance ont été testées et ont donné des résultats satisfaisants sur des simulations. En passant sur les données radar, le modèle a montré ses limites avec notamment des difficultés à reproduire la dépendance d'un couple $(Y_t, Y_{t+30\text{min}})$. L'utilisation de la méthode des moments a aussi été envisagée pour estimer la dépendance d'un couple Gaussien censuré, mais il a été montré que cette méthode était biaisée par la discrétisation des données.

Par manque de temps le modèle d'assimilation de données qui était envisagé n'a pas pu être développé, aussi l'idée de créer un jeu de données ayant la structure spatiotemporelle du radar et le distribution des pluviomètres a été reprise avec une méthode plus rapide à mettre en œuvre :

1. Estimation du déplacement des images radar par maximum de corrélation
2. Interpolation des images à un pas de temps de 3 minutes
3. Correction en distribution par QQ mapping avec comme référence un pluviomètre

Avec cette méthodologie il a été possible de produire plusieurs jeux de données afin de tester la sensibilité du modèle hydraulique à différentes caractéristiques. L'importance de reproduire la discrétisation des pluviomètres a été démontrée : un simple arrondi à 0.2 mm dégrade fortement les résultats mais si l'effet d'accumulation qui se produit dans un pluviomètre à bascule est reproduit, alors la discrétisation n'a peu ou pas d'impact sur le modèle hydraulique. Ce résultat rend compliquée l'utilisation du produit corrigé basé sur le radar car la correction en distribution par QQ mapping ne permet ni de conserver la précision du radar ni de reproduire les effets d'accumulation.

Le deuxième résultat important de l'analyse de sensibilité est l'impact de la spatialisation de la pluie sur les sorties du modèle. Il a été montré qu'en utilisant une pluie

constante pour toute la zone les maximums de débit déversés étaient fortement impactés, mais qu'utiliser une grille de 8 pluviomètres apportait déjà une bonne amélioration.

Conclusions pour le projet MEDISA

Comme ça a été évoqué l'utilisation du produit corrigé à partir du radar est compromise par la façon dont la discrétisation est appliquée. L'autre possibilité était l'utilisation des pluviomètres du réseau d'Eau du Ponant, sous réserve d'une correction des erreurs de mesure. Toutefois 1) l'utilisation d'une chronique spatialisée demande de recaler le modèle hydraulique sur des chroniques spatialisées, ce qui n'était pas envisageable dans les délais du projet, et 2) l'impact de la spatialisation concerne surtout les maximums de débit déversé, ce qui est plutôt une problématique pour les inondations mais pas vraiment important dans le cadre des déversements. En effet l'objectif du projet est surtout de mieux gérer les petites pluies.

> En conclusion la précision spatiale des données ayant surtout un effet à la marge des besoins du projet MEDISA, il a été conseillé de continuer à travailler avec le pluviomètre de BMO appliqué à toute la zone.

Perspectives

Pour la partie applicative de la thèse une piste d'amélioration pour le modèle hydraulique d'Eau du Ponant est l'utilisation du réseau de pluviomètres pour forcer le modèle hydraulique. Ce changement nécessiterait une correction des données des pluviomètres, ce qui est déjà un des objectifs d'Eau du Ponant.

Pour ce qui est de la correction des données radar par les pluviomètres, une option qui a été assez peu explorée est celle du QQ mapping paramétrique basé sur le modèle GP méta-Gaussien. Comme ça a pu être évoqué la variation des paramètres avec le pas de temps montrée au Chapitre 2 laisse penser qu'il serait possible d'estimer les paramètres à 5 minutes à partir des données à 3 minutes des pluviomètres. Toutefois reste le problème de la discrétisation, qui dans la transformation QQ mapping est appliquée à chaque pas de temps sans reproduire les effets d'accumulation qui existent dans les pluviomètres. Plusieurs solutions pourraient être envisagées : 1) créer des pseudo-pluviomètres avec les données radar en reproduisant cet effet d'accumulation et de bascule à 0.2 mm avant la

correction en distribution ou 2) désagréger les pluviomètres afin d'obtenir des données plus continues sur lesquelles ajuster le QQ mapping.

Le problème de la fusion des différentes sources de données pourrait être formulé dans le cadre de l'assimilation de données. On définit un champ Gaussien latent de moyenne $\boldsymbol{\mu}$ et de covariance $\boldsymbol{\Sigma}$ et la transformation $\Psi : x \mapsto 0 \times \mathbb{1}_{x<0} + \psi(x) \times \mathbb{1}_{x\geq 0}$. Le modèle espace-état s'écrit alors

$$\begin{cases} X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ R = \Psi(X + \delta), & \delta \sim \mathcal{N}(\boldsymbol{\mu}_R, \boldsymbol{\Sigma}_R) \\ G_{EDP} = \Psi(X + \epsilon), & \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2_{EDP}) \\ G_{MF} = \Psi(X + \eta), & \eta \sim \mathcal{N}(0, \sigma^2_{MF}). \end{cases}$$

R, G_{EDP} et G_{MF} sont respectivement le radar, les pluviomètres d'Eau du Ponant et le pluviomètre de Météo France. δ, ϵ et η sont leurs erreurs de mesure respectives. Elles ont chacune des caractéristiques qui peuvent être déduites de ce qu'on connaît des données. Le radar peut avoir un biais non nul ($\boldsymbol{\mu}_R$) et des erreurs de mesure spatialement dépendantes ($\boldsymbol{\Sigma}_R$). Les pluviomètres n'ont a priori pas de biais, aussi leurs erreurs de mesure sont centrées. Les erreurs sont spatialement indépendantes et leurs variance peut varier suivant les points : on s'attend à ce qu'elle soit faible pour Météo France, et on peut imaginer avoir des mesures de qualité différentes pour les différents pluviomètres du réseau d'Eau du Ponant.

Comme ça a été montré au Chapitre 3, il faudrait alors étudier de plus près la possibilité d'utiliser une copule Gaussienne pour des données de précipitation et un point important sera la modélisation de la covariance du champ Gaussien latent et des erreurs du radar.

Annexe A

Documentation fournie par Météo France concernant les données radar

---

Attention, cette fiche concerne les premières données commandées à Météo France, celles du radar individuel de Plabennec. Les données finalement utilisées sont celles de la mosaïque qui combine tous les radars du territoire français. La zone d'intérêt est très proche du radar de Plabennec, on s'attend donc à ce que les autres radars n'ait que peu ou pas d'impact sur les données de notre zone, mais les corrections indiquées dans cette fiche sont probablement incomplètes.

Fiche synthétique Produit de Données Spatialisées

« Lame d'eau individuelle radar »

Date de mise à jour 06/02/2018

INTITULÉ ET STATUT DU PRODUIT

- Estimation en mm du cumul de précipitations en 5 minutes à partir des mesures d'un radar local, accompagnée en chaque pixel d'un code qualité dynamique.

- Produit opérationnel au T1 2018

DONNÉES DE BASE

- Toutes les données issues des différents tours d'antenne aux différents angles de site balayés toutes les 5 minutes par le radar local.

- Cartes de niveau moyen des échos fixes par tour d'antenne.

- Cartes statistiques de masques (observés et simulés en utilisant un Modèle Numérique de Terrain).

- Carte d'interdiction des élévations basses (en vue d'éliminer les échos d'éoliennes et les échos de mer).

- Altitude de l'isotherme 0°C issue de la BDAP.

- Champs d'advection 2PiR.

MÉTHODE DE CONSTITUTION DU PRODUIT

L'algorithme mis en oeuvre pour produire la lame d'eau comprend les traitements suivants:

- Correction des masques causés par le relief, les arbres ou les bâtiments,
- Correction des sous-estimations à grande distance liées à l'altitude du faisceau,
- Correction des sur-estimations hivernales liées aux bandes brillantes,
- Correction de l'atténuation par les précipitations
- Correction de l'atténuation par les gaz.
- Correction des cumuls par les pluviomètres
- Réduction des zones d'échos fixes par un compositage multi-sites en chaque pixel ainsi que par un bouchage par advection,
- Elimination des discontinuités par une combinaison linéaire pondérée des différents sites disponibles,

- Le produit lame d'eau individuelle comprend une information dynamique de qualité
  - Chaque lame d'eau individuelle est systématiquement accompagnée d'une carte de codes qualité,
  - Les codes qualité dépendent de l'altitude de la mesure par rapport au sol, de la correction

du taux de masquage, de l'âge de l'advection, et pour la bande X, de la correction de l'atténuation par les précipitations. La qualité est une fonction décroissante de l'altitude, de l'âge de l'advection, et des corrections appliquées.

- Les codes qualité proposés ont été mis en relation avec des scores de long-terme entre radar et pluviomètres,

- Le traitement des échos fixes est dynamique:
  - Il faut noter que les zones d'échos fixes peuvent varier en localisation et en extension au cours du temps. Sur une certaine période, un pixel peut être classé tantôt comme écho fixe et tantôt comme écho de précipitations,
  - Comme mentionné plus haut, les zones d'échos fixes sont très réduites avec les lames d'eau radar. Toutefois, pour certains radars ayant peu d'angles d'élévation et / ou situés en région montagneuse, ces zones d'échos fixes ne sont pas totalement éliminées,
  - Les échos fixes résiduels sont toujours signalés dans l'image de lame d'eau par le code de valeur manquante (65535).

- Le taux de pluie R est estimé à partir de la réflectivité Z à l'aide d'une loi empirique de type Z=a*R^b:
  - En métropole, c'est la loi de Marshall Palmer qui est utilisée: a=200 et b=1.6
  - En Martinique et Guadeloupe: a=150 et b=1.5
  - A la Réunion et en Nouvelle Calédonie: a=300 et b=1.35

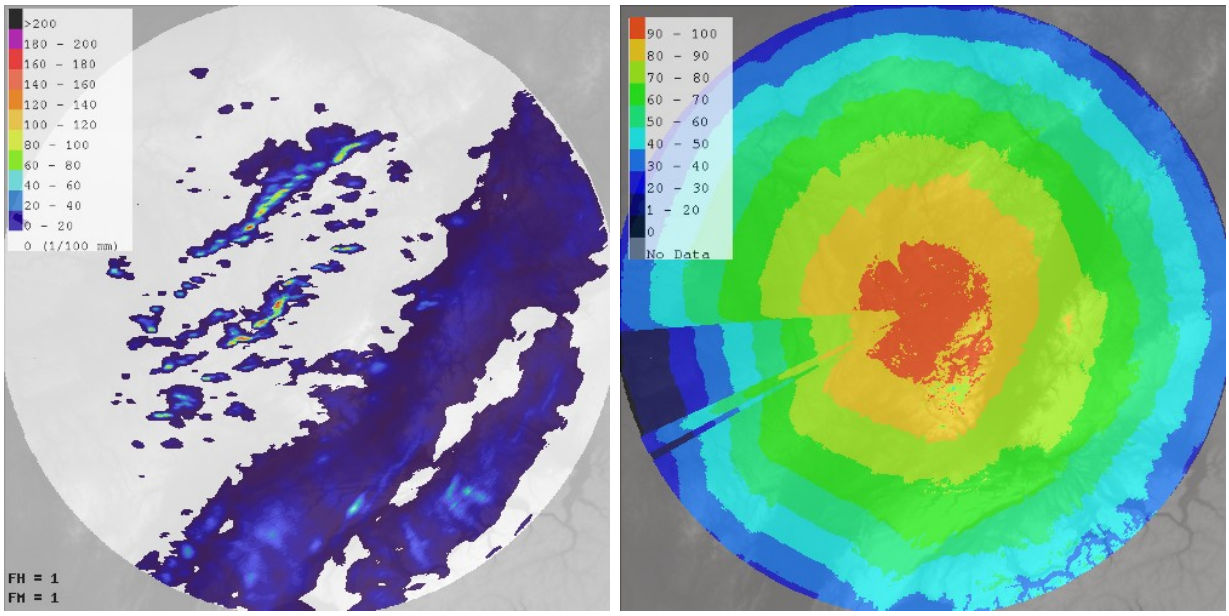


Fig.1:lame d'eau individuelle 5mn de Nancy le14 Septembre 2017, et codes de qualité compris, entre 0 (mauvais) et 100 (excellent)

CONTENU DÉTAILLÉ DU PRODUIT

- Chaque fichier BUFR représentant une lame d'eau individuelle contient 4 types d'informations:
  - La valeur en chaque pixel du cumul de précipitation en centièmes de mm
  - La valeur en chaque pixel du code qualité

- La valeur en chaque pixel de l'ajustement aux pluviomètres: champ non utilisé pour l'instant
- Le type de précipitation (stratiforme, convective): champ non renseigné pour l'instant

- Résolution spatiale: 1x1 km^2
- Domaine : grille cartésienne régulière de données sur 512 * 512 km^2
- Projection: radar local, assimilée à gnomonique ou azimutale équidistante centrée sur la position du radar.
- Résolution temporelle: 5 minutes

RECOMMANDATIONS CONCERNANT L'UTILISATION DU PRODUIT

- DSO/CMR fournit des cumuls avec possiblement des valeurs manquantes. Pour cumuler des lames d'eau sur une certaine période T, DSO/CMR préconise la méthode suivante :
  - Pour des cumuls sur des durées allant jusqu'à 15 minutes:
    - En cas d'absence d'un pixel 5 minutes: le code qualité du pixel 5 minutes manquant est forcé à 0, le cumul 15 minutes associé à ce pixel correspond à 1.5 fois le cumul 10 minutes. Le code qualité associé au cumul 15 minutes correspond à la moyenne des codes qualité 5 minutes.
    - En cas d'absence de plus d'un pixel 5 minutes, le cumul 15 minutes est considéré comme manquant à ce pixel, le code qualité associé est alors manquant.
  - Pour des cumuls sur des durées supérieures à 15 minutes: la recommandation est de remplacer les valeurs manquantes (65535) par des 0 (mm) dans les produits que l'on veut cumuler et de les prendre en compte dans le cumul. Si l'une des valeurs à cumuler est manquante, son code qualité est forcé à 0 et le code qualité du cumul est alors égal à la moyenne des codes qualité.

LIMITES DU PRODUIT

- La lame d'eau individuelle 5 minutes tend à sur-estimer les pluies faibles et à sous-estimer les pluies fortes.

- Des cas de fausses alarmes sont possibles, liés notamment à des éoliennes dont la présence génère des échos qui résistent à l'algorithme de détection dynamique des échos fixes.

- Possibilité de non détection de précipitations hivernales (neige) très faibles et/ou faiblement développées verticalement

EVOLUTIONS DU PRODUIT

- Mars 2007:ajout du module d'ajustement par les pluviomètres.

- Août 2009:
  - Ajout de la correction de l'atténuation par les gaz,
  - Modification du calcul du code qualité en fonction de l'altitude de la mesure afin de donner plus de poids aux élévations les plus basses,
  - Augmentation de la correction de la sous-estimation du faisceau à grande distance,

- Février 2012, pour les **radars à diversité de polarisation**:
  - Elimination des échos d'air clair,
  - Correction de l'atténuation par les pluies.

- Juillet 2015 pour les radars en bande X :
  - Correction de l'atténuation par le radôme mouillé ,
  - Le taux de pluie sur les pixels de forte réflectivité est estimé à partir de l'atténuation spécifique Kdp
  - les codes de qualité prennent en compte la correction de l'atténuation par les précipitations.

- Novembre 2017:
  - le champ d'advection utilisé est le champ 2PIR,
  - le format du produit évolue pour permettre (usage futur) de préciser le type d'ajustement aux pluviomètres utilisé.

Annexe B

Some Theoretical Properties of the GP Meta-Gaussian Distribution

The density, cdf and quantile function of a meta-Gaussian model as defined in (2.1) are :

$$f_Y(y) = c \times \begin{cases} \phi_\mu\left(\psi^{-1}(y)\right)/\psi'\left(\psi^{-1}(y)\right) & \text{if } y > 0 \\ \Phi_\mu(0) & \text{if } y = 0 \end{cases},$$

$$F_Y(y) = c \times \begin{cases} \Phi_\mu(\psi^{-1}(y)) & \text{if } y > 0 \\ \Phi_\mu(0) & \text{if } y = 0 \end{cases},$$

$$F_Y^-(u) = \begin{cases} \psi(\Phi_\mu^{-1}(u/c)) & \text{if } u > \Phi_\mu(0) \\ 0 & \text{if } u = \Phi_\mu(0) \end{cases},$$

with ϕ_μ and Φ_μ respectively the density and cdf of a normal distribution with mean μ. c is the normalisation constant that deals with the probability of truncation when $\xi < 0$ with the GP meta-Gaussian transform. Hence $c = 1$ for the classical transform (2.4), and for the GP meta-Gaussian transform (2.8) $c = 1/\Phi_\mu(x_{sup})$, with x_{sup} the upper bound in the Gaussian domain as defined in (2.9).

An explicit writing of the moments was found for the GP meta-Gaussian distribution. Let us write Y_+ the wet measurements.

$$\begin{aligned} E(Y_+^p) &= \frac{1}{\sqrt{2\pi}(1-\Phi(-\mu))} \int_0^{+\infty} (x)^p \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} dx \\ &= \frac{\sigma^p}{\sqrt{2\pi}(1-\Phi(-\mu))} \exp\left(-\frac{\mu^2}{2}\right) \int_0^{+\infty} x^{p/\alpha} \exp\left\{-\frac{1-\xi p}{2}x^2 + \mu x\right\} dx \end{aligned}$$

By identification in Gradshteyn et Ryzhik (2007) (equation 1 of section 3.462, page

365), with $\gamma = -\mu$, $\nu - 1 = p/\alpha$ and $\beta = (1 - \xi p)/2$,

$$E(Y_+^p) = \frac{\sigma^p (1-\xi p)^{-\frac{1}{2}(\frac{p}{\alpha}+1)}}{\sqrt{2\pi}(1-\Phi(-\mu))} \exp\left\{\frac{\mu^2}{2} \quad \frac{1}{2(1-\xi p)} - 1\right)\right\} \Gamma\left(\frac{p}{\alpha}+1\right) D_{-(\frac{p}{\alpha}+1)} \quad -\frac{\mu}{\sqrt{1-\xi p}}\right)$$

Γ is the Gamma function and D_ν can be expressed with the $W_{\kappa,\mu}(.)$ Whittaker's function (Gradshteyn et Ryzhik (2007), section 9.240, page 1028) as

$$D_\nu(z) = 2^{\frac{1}{4}+\frac{\nu}{2}} W_{\frac{1}{4}+\frac{\nu}{2},-\frac{1}{4}} \quad \frac{z^2}{2}\right) z^{-1/2}.$$

The expression of $E(Y_+^p)$ is valid under conditions on p that depend on ξ :

$$\begin{aligned} &\text{for } \xi > 0, \quad -\alpha < p < 1/\xi \\ &\text{for } \xi < 0, \quad -\alpha < p \text{ and } 1/\xi < p \\ &\text{for } \xi = 0, \quad -\alpha < p \end{aligned}$$

Note that for $\xi \leq 0$, $p > 0$ is a sufficient condition, whereas for $\xi > 0$ first estimation of ξ is needed to determine to moments that can be computed, which is due to the fact that some moments are infinite.

Annexe C

Pareto Tail for Meta-Gaussian Models

Proposition 1. *Let Z be any positive absolutely continuous random variable with pdf f_Z and with a Pareto survival function $\overline{F}_Z$. Let X be any standardized normal distributed random variable, and let's define the positive random variable*

$$Y \overset{d}{=} \quad (X),$$

where $\overset{d}{=}$ means equality in distribution and $\psi(.)$ represents a continuous and increasing function from the real line to $[0, \infty)$. The two random variables Z and Y are tail-equivalent if and only if

$$\lim_{x \to \infty} \frac{x\psi(x)}{\psi'(x)} = \frac{1}{\xi}, \tag{C.1}$$

where ξ corresponds the common positive GP shape parameter of Z.

Proof of Proposition 1 : ϕ and $\overline{\Phi}$ denote the pdf and survival function of X, a standardised normal distributed random variable.

Recall that Z and Y are tail-equivalent, i.e.

$$\lim_{y \to \infty} \frac{\overline{F}_Z(y)}{\mathbb{P}[Y > y]} = c \in (0, \infty),$$

This condition is satisfied if they have the same tail index. Assuming a Pareto tail with positive shape parameter ξ for Z implies that Z is regularly varying with index $1/\xi$. Proposition A.3.8(b) from Embrechts et al. (2013) recalled that this regular variation type is equivalent to

$$\lim_{z \to \infty} \frac{z \times f_Z(z)}{\overline{F}_Z(z)} = \frac{1}{\xi}.$$

Hence, to show that Y and Z are tail equivalent, one needs to determine under which

condition it can be written that

$$\lim_{z\to\infty} \frac{z \times f_Y(z)}{\overline{F}_Y(z)} = \frac{1}{\xi}.$$

where f_Y and $\overline{F}_Y$ denote the pdf and survival function of Y, respectively.

By construction, the survival function of Y equals to

$$\overline{F}_Y(z) = \mathbb{P}[X > \quad ^{-1}(z)] = \overline{\Phi}\left[\psi^{-1}(z)\right],$$

The density of Y is

$$f_Y(z) = \left(\psi^{-1}(z)\right)' \phi\left[\psi^{-1}(z)\right].$$

Then one can write

$$\frac{z \times f_Y(z)}{\overline{F}_Y(z)} = \left(z \times \quad ^{-1}(z) \times \left(\psi^{-1}(z)\right)'\right) \times \quad \frac{\phi\left[\psi^{-1}(z)\right]}{^{-1}(z)\overline{\Phi}\left[\psi^{-1}(z)\right]}\right).$$

Mill's ratio tells us that the ratio in the last bracket goes to one as $^{-1}(z)$ goes to ∞ (i.e. as z grows). Hence, the condition

$$\lim_{z\to\infty} \left(z \times \quad ^{-1}(z) \times \left(\psi^{-1}(z)\right)'\right) = \frac{1}{\xi}, \tag{C.2}$$

is equivalent to

$$\lim_{z\to\infty} \frac{z \times f_Y(z)}{\overline{F}_Y(z)} = \frac{1}{\xi}.$$

This is equivalent to have tail equivalence between Z and Y.

Changing variables with $z = \psi(x)$, $x = \psi^{-1}(z)$ and $\left(\psi^{-1}(z)\right)' = dx/dz$, condition (C.2) is equivalent to condition (C.1).

This is the necessary and sufficient condition on $\psi(.)$ to build a Pareto random variable of tail index ξ from a standardized normal random variable X.

Annexe D

Estimation par maximum de vraisemblance d'un champ GP méta-Gaussien continu

Pour chaque pas de temps on réorganise les n variables de pluie $Y = (Y_S, Y_P)'$ pour séparer le temps sec et le temps de pluie. La matrice de covariance du champ Gaussien associé $X = (X_S, X_P)'$ s'écrit alors $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{SS} & \mathbf{\Sigma}_{SP} \\ \mathbf{\Sigma}_{PS} & \mathbf{\Sigma}_{PP} \end{pmatrix}$, on note le vecteur des moyennes du champ Gaussien $\boldsymbol{\mu} = (\boldsymbol{\mu_S}, \boldsymbol{\mu_P})'$.

On a alors la distribution conditionnelle de X_S conditionnellement à X_P qui est un champ Gaussien de moyenne $\boldsymbol{\mu}_c = \boldsymbol{\mu}_S + \mathbf{\Sigma}_{SP}\mathbf{\Sigma}_{PP}^{-1}(X_p - \boldsymbol{\mu}_P)$ et de covariance $\mathbf{\Sigma}_c = \mathbf{\Sigma}_{SS} - \mathbf{\Sigma}_{SP}\mathbf{\Sigma}_{PP}^{-1}\mathbf{\Sigma}_{PS}$.

En utilisant de théorème de Bayes on peut faire apparaître cette distribution conditionnelle et ainsi écrire la log vraisemblance d'un pas de temps comme

$$\log\left(\Phi_{\boldsymbol{\mu}_c,\mathbf{\Sigma}_c}(Y_S)\right) + \log\left(\phi_{\boldsymbol{\mu}_P,\mathbf{\Sigma}_{PP}}(Y_P)\right) - \sum_{y_P \in Y_P} \log\Big(\psi'(\psi^{-1}(y_p))\Big).$$

Toutefois comme on l'a montré au Chapitre 2 les résultats sont fortement impactés par la discrétisation et on s'attend à ce que ce soit aussi le cas en multivarié.

Annexe E

Moments des Gaussiennes tronquées et censurées

Soit un couple de variables Gaussiennes

$$x = (_1, x_2) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

On note le couple censuré x^+ et le couple tronqué x^{+*}, ils sont définis par

$$\begin{array}{rcl} x_i^{+*} = & x_i \times \mathbb{1}_{b_i \leq x_i \leq a_i} \\ x_i^{+} = & x_i \times \mathbb{1}_{b_i \leq x_i \leq a_i} + 0 \times \mathbb{1}_{x_i < b_i \text{ ou } x_i > a_i} \end{array} .$$

Les troncatures et les censures se font donc à gauche en b_i et à droite en a_i pour la variable i. Dans le cas tronqué x_i^{+*} n'est pas observé hors de l'intervalle $[b_i, a_i]$, et dans le cas censuré on observe soit a_i soit b_i en fonction du côté de l'intervalle où on se trouve.

On notera

$$p = p(a_1, b_1, a_2, b_2; \rho) = \int_{b_1}^{a_1} \int_{b_2}^{a_2} \phi(x_1, x_2; \rho) dx_2 dx_1$$

On s'intéresse surtout aux troncatures et censures unilatérales à gauche et aux moments d'ordre $(r, s) \in [(1, 0), (2, 0), (1, 1)]$, avec r et s interchangeables : $\mu_{r,s} = E(x_1^r x_2^s)$

Toutes les formules qui vont suivre sont tirées de Muthen (1990).

E.1 Troncature

La formule générale est

$$\begin{aligned}
pE(x_i^{+*}x_j^{+*};a_1,b_1,a_2,b_2) = &\begin{Bmatrix}1\\ \rho\end{Bmatrix} p - \begin{Bmatrix}1\\ \rho\end{Bmatrix} a_i\phi(a_i)[\Phi((a_j-\rho a_i)c)-\Phi((b_j-\rho a_i)c)]\\
&+\begin{Bmatrix}0\\ c^{-1}\end{Bmatrix}\phi(a_i)[\phi((a_j-\rho a_i)c)-\phi((b_j-\rho a_i)c)]\\
&+\begin{Bmatrix}1\\ \rho\end{Bmatrix} b_i\phi(b_i)[\Phi((a_j-\rho b_i)c)-\Phi((b_j-\rho b_i)c)]\\
&-\begin{Bmatrix}0\\ c^{-1}\end{Bmatrix}\phi(b_i)[\phi((a_j-\rho b_i)c)-\phi((b_j-\rho b_i)c)]\\
&-\begin{Bmatrix}\rho\\ 1\end{Bmatrix}\rho a_j\phi(a_j)[\Phi((a_i-\rho a_j)c)-\Phi((b_i-\rho a_j)c)]\\
&+\begin{Bmatrix}c^{-1}\\ 0\end{Bmatrix}\rho\phi(a_j)[\phi((a_i-\rho a_j)c)-\phi((b_i-\rho a_j)c)]\\
&+\begin{Bmatrix}\rho\\ 1\end{Bmatrix}\rho b_j\phi(b_j)[\Phi((a_i-\rho b_j)c)-\Phi((b_i-\rho b_j)c)]\\
&-\begin{Bmatrix}c^{-1}\\ 0\end{Bmatrix}\rho\phi(b_j)[\phi((a_i-\rho b_j)c)-\phi((b_i-\rho b_j)c)],
\end{aligned}$$

où $\begin{Bmatrix}u\\ v\end{Bmatrix} = \begin{cases} u & \text{si } i=j \\ v & \text{sinon}\end{cases}$, $c=(1-\rho^2)^{-\frac{1}{2}}$, et $(i,j)\in(1,2)$.

Avec $\phi(+\infty)=\phi(-\infty)=0$, $\Phi(-\infty)=0$, $\Phi(+\infty)=1$ et enfin $x\phi(x)\to 0$ quand $x\to\pm\infty$ on pourra simplifier ces expressions dans le cas des troncatures unilatérales.

E.1.1 Moments $\mu_{1,0}$, $\mu_{0,1}$

On notera $L(h,k;\rho)=p(\infty,h,\infty,k;\rho)=\mathbb{P}(x_i>h\cap x_j>k)$

$$E(x_i^{+*};\infty,h,\infty,k)=\frac{\phi(h)}{L(h,k;\rho)}\quad 1-\Phi\left[\frac{k-\rho h}{\sqrt{1-\rho^2}}\right]\Bigg)+\frac{\rho\phi(k)}{L(h,k;\rho)}\quad 1-\Phi\left[\frac{h-\rho k}{\sqrt{1-\rho^2}}\right]\Bigg)$$

$$E(x_i^{+*};\infty,h,\infty,h) = \frac{\phi(h)}{L(h,h;\rho)}(1+\rho) \quad 1-\Phi\left[\frac{(1-\rho)h}{\sqrt{1-\rho^2}}\right]\Bigg)$$

$$E(x_i^{+*};\infty,0,\infty,0) = \frac{1+\rho}{2\sqrt{2\pi}L(0,0;\rho)}$$

E.1.2 Moments $\mu_{1,1}$, $\mu_{2,0}$, $\mu_{0,2}$

On notera $L(h,k;\rho) = p(\infty,h,\infty,k;\rho) = \mathbb{P}(x_i > h \cap x_j > k)$

$$\begin{aligned}
E((x_i^{+*})^2;\infty,h,\infty,k) = 1 + \frac{1}{L(h,k;\rho)} \Big\{ & h\phi(h)(1-\Phi[(k-\rho h)c]) + \rho^2 k\phi(k)(1-\Phi[(h-\rho k)c]) \\
& + \frac{\rho}{c}\phi(k)\phi[(h-\rho k)c]\Big\} \\
E(x_i^{+*}x_j^{+*};\infty,h,\infty,k) = \rho + \frac{1}{L(h,k;\rho)} \Big\{ & \rho h\phi(h)(1-\Phi[(k-\rho h)c]) + \rho k\phi(k)(1-\Phi[(h-\rho k)c]) \\
& + \frac{1}{c}\phi(h)\phi[(k-\rho h)c]\Big\}
\end{aligned}$$

$$\begin{aligned}
E((x_i^{+*})^2;\infty,h,\infty,h) =& 1 + \frac{\phi(h)}{L(h,h;\rho)} \left\{ (1+\rho^2)h \quad 1-\Phi\left[\frac{1-\rho}{\sqrt{1-\rho^2}}h\right]\right) + \rho\sqrt{1-\rho^2}\phi\left[\frac{1-\rho}{\sqrt{1-\rho^2}}h\right]\right\} \\
E(x_i^{+*}x_j^{+*};\infty,h,\infty,h) =& \rho + \frac{\phi(h)}{L(h,h;\rho)} \left\{ 2\rho h \quad 1-\Phi\left[\frac{1-\rho}{\sqrt{1-\rho^2}}h\right]\right) + \sqrt{1-\rho^2}\phi\left[\frac{1-\rho}{\sqrt{1-\rho^2}}h\right]\right\}
\end{aligned} \tag{E.1}$$

$$\begin{aligned}
E((x_i^{+*})^2;\infty,0,\infty,0) =& 1 + \frac{\rho\sqrt{1-\rho^2}}{2\pi L(0,0;\rho)} \\
E(x_i^{+*}x_j^{+*};\infty,0,\infty,0) =& \rho + \frac{\sqrt{1-\rho^2}}{2\pi L(0,0;\rho)}
\end{aligned}$$

E.2 Censure

La formule générale est

$$\begin{aligned}
E((x_1^+)^r(x_2^+)^s;a_1,b_1,a_2,b_2,\rho) =& p(b_1,-\infty,+\infty,a_2;\rho)b_1^r a_2^s \\
&+ p(a_1,b_1,+\infty,a_2;\rho)E((x_1^{+*})^r;a_1,b_1,+\infty,a_2,\rho)a_2^s \\
&+ p(+\infty,a_1,+\infty,a_2;\rho)a_1^r a_2^s \\
&+ p(b_1,-\infty,a_2,b_2;\rho)E((x_2^{+*})^s;b_1,-\infty,a_2,b_2,\rho)b_1^r \\
&+ p(a_1,b_1,a_2,b_2;\rho)E((x_1^{+*})^r(x_2^{+*})^s;a_1,b_1,a_2,b_2,\rho) \\
&+ p(+\infty,a_1,a_2,b_2;\rho)E((x_2^{+*})^s;+\infty,a_1,a_2,b_2,\rho)a_1^r \\
&+ p(b_1,-\infty,b_2,-\infty;\rho)b_1^r b_2^s \\
&+ p(a_1,b_1,b_2,-\infty;\rho)E((x_1^{+*})^r;a_1,b_1,b_2,-\infty,\rho)b_2^s \\
&+ p(+\infty,a_1,b_2,-\infty;\rho)a_1^r b_2^s
\end{aligned}$$

On récupère les censures unilatérales en utilisant le fait que si une intégrale de p tend vers zéro (ses deux bornes tendent vers $\pm\infty$), elle le fait plus vite que u^r ou u^s quand $u \to +\infty$, car $r, s \leq 2$.

Dans le cas de la censure à gauche, on a $a_1 = a_2 = +\infty$, $b_1 = h$, $b_2 = k$ et on peut alors écrire

$$\begin{aligned}
E((x_1^+)^r(x_2^+)^s;\infty,h,\infty,k,\rho) =& p(h,-\infty,+\infty,k;\rho)E((x_2^{+*})^s;h,-\infty,+\infty,k,\rho)h^r \\
&+ L(h,k;\rho)E((x_1^{+*})^r(x_2^{+*})^s;+\infty,h,+\infty,k,\rho) \\
&+ \Phi(h,k;\rho)h^r k^s \\
&+ p(+\infty,h,k,-\infty;\rho)E((x_1^{+*})^r;+\infty,h,k,-\infty,\rho)k^s
\end{aligned}$$

E.2.1 Moments $\mu_{1,0}$, $\mu_{0,1}$

NB : $p(a,b,c,-\infty;\rho) + p(a,b,+\infty,c;\rho) = \int_a^b \int_{-\infty}^{+\infty} \phi(x,y;\rho)dxdy = \Phi(b) - \Phi(a)$

$$\begin{aligned}
E(x_1^+;+\infty,h,+\infty,k) =& \phi(h) + h\Phi(h) \\
E(x_1^+;+\infty,0,+\infty,k) =& \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

E.2.2 Moments $\mu_{1,1}$, $\mu_{2,0}$, $\mu_{0,2}$

$$
\begin{aligned}
E((x_1^+)^2; +\infty, h, +\infty, k) =& 1 + h\phi(h) + (h^2 - 1)\Phi(h) \\
E((x_1^+)^2; +\infty, 0, +\infty, k) =& \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
E(x_1^+ x_2^+; +\infty, h, +\infty, k) =& \rho L(h, k; \rho) + hk\Phi(h, k; \rho) + \sqrt{1-\rho^2}\phi(h)\phi\left(\frac{k - \rho h}{\sqrt{1-\rho^2}}\right) \\
&+ k\phi(h)\Phi\left[\frac{k - \rho h}{\sqrt{1-\rho^2}}\right] + h\phi(k)\Phi\left[\frac{h - \rho k}{\sqrt{1-\rho^2}}\right] \\
E(x_1^+ x_2^+; +\infty, h, +\infty, h) =& \rho L(h, h; \rho) + h^2\Phi(h, h; \rho) + \sqrt{1-\rho^2}\phi(h)\phi\left(\frac{1-\rho}{\sqrt{1-\rho^2}}h\right) \\
&+ 2h\phi(h)\Phi\left[\frac{1-\rho}{\sqrt{1-\rho^2}}h\right] \qquad \text{(E.2)} \\
E(x_1^+ x_2^+; +\infty, 0, +\infty, 0) =& \rho L(0, 0; \rho) + \frac{\sqrt{1-\rho^2}}{2\pi}
\end{aligned}
$$

BIBLIOGRAPHIE

Ailliot, P., Boutigny, M., Koutroulis, E., Malisovas, A. & Monbet, V., (2020), Stochastic weather generator for the design and reliability evaluation of desalination systems with Renewable Energy Sources, *Renewable Energy*, *158*, 541-553.

Ailliot, P., Thompson, C. & Thomson, P., (2009), Space–time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions, *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, *58 3*, 405-426.

Allard, D. & Bourotte, M., (2015), Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, *Stochastic environmental research and risk assessment*, *29 2*, 453-462.

Allcroft, D. J. & Glasbey, C. A., (2003), A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation, *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, *52 4*, 487-498.

AMODIAG Environnement, (2012), Etude complémentaire de modélisation du réseau eaux pluviales en fonction de la topographie pour la ville d'Yvetot.

Anderson, T. W. & Darling, D. A., (1952), Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes, *The annals of mathematical statistics*, 193-212.

Arnaud, P., Bouvier, C., Cisneros, L. & Dominguez, R., (2002), Influence of rainfall spatial variability on flood prediction, *Journal of Hydrology*, *260 1-4*, 216-230.

Arshad, M., Rasool, M. & Ahmad, M., (2003), Anderson Darling and Modified Anderson Darling Tests for, *Pakistan Journal of Applied Sciences*, *3 2*, 85-88.

Atencia, A., Mediero, L., Llasat, M. & Garrote, L., (2011), Effect of radar rainfall time resolution on the predictive capability of a distributed hydrologic model, *Hydrology and Earth System Sciences*, *15 12*, 3809-3827.

Azimi-Zonooz, A., Krajewski, W., Bowles, D. & Seo, D., (1989), Spatial rainfall estimation by linear and non-linear co-kriging of radar-rainfall and raingage data, *Stochastic Hydrology and Hydraulics*, *3 1*, 51-67.

Bagnall, A., Bostrom, A., Large, J. & Lines, J., (2016), The great time series classification bake off : an experimental evaluation of recently proposed algorithms, *Extended Version. CoRR, abs/1602.01711.*

Bardossy, A. & Plate, E. J., (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water resources research*, *28 5*, 1247-1259.

Bauer, P., Thorpe, A. & Brunet, G., (2015), The quiet revolution of numerical weather prediction, *Nature*, *525 7567*, 47-55.

Benoit, L., Allard, D. & Mariethoz, G., (2018), Stochastic Rainfall Modeling at Sub-kilometer Scale, *Water Resources Research*, *54 6*, 4108-4130.

Benseman, R. & Cook, F., (1969), Radiation in New Zealand-Standard year and radiation on inclined slopes, *New Zealand Journal of Science*, *12 4*, 696.

Berndt, C., Rabiei, E. & Haberlandt, U., (2014), Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios, *Journal of Hydrology*, *508*, 88-101, https://doi.org/https://doi.org/10.1016/j.jhydrol.2013.10.028

Berndt, D. J. & Clifford, J., (1994), Using dynamic time warping to find patterns in time series., *KDD workshop*, *10 16*, 359-370.

Berne, A., Delrieu, G., Creutin, J.-D. & Obled, C., (2004), Temporal and spatial resolution of rainfall measurements required for urban hydrology, *Journal of Hydrology*, *299 3-4*, 166-179.

Boé, J., Terray, L., Habets, F. & Martin, E., (2007), Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, *International Journal of Climatology : A Journal of the Royal Meteorological Society*, *27 12*, 1643-1655.

Bonta, J. V. & Rao, A. R., (1988), Factors affecting the identification of independent storm events, *Journal of Hydrology*, *98 3-4*, 275-293.

Boutigny, M., (2017), La notion d'année typique en météorologie.

Boutigny, M., Ailliot, P., Chaubet, A., Naveau, P. & Saussol, B., (2021), Modelling rainfall from sub-hourly to daily scale with a heavy tailed meta-Gaussian model, *Earth and Space Science Open Archive*, 71, https://doi.org/10.1002/essoar.10506575.1

Box, G. E. & Cox, D. R., (1964), An analysis of transformations, *Journal of the Royal Statistical Society : Series B (Methodological)*, *26 2*, 211-243.

Bracken, L., Cox, N. & Shannon, J., (2008), The relationship between rainfall inputs and flood generation in south–east Spain, *Hydrological Processes : An International Journal*, *22 5*, 683-696.

Breinl, K., (2016), Driving a lumped hydrological model with precipitation output from weather generators of different complexity, *Hydrological Sciences Journal*, *61 8*, 1395-1414.

Bringi, V. N. & Chandrasekar, V., (2001), *Polarimetric Doppler weather radar : principles and applications*, Cambridge university press.

Caron, A., Leconte, R. & Brissette, F., (2008), An improved stochastic weather generator for hydrological impact studies, *Canadian Water Resources Journal*, *33 3*, 233-256.

Caseri, A., Javelle, P., Ramos, M.-H. & Leblois, E., (2016), Generating precipitation ensembles for flood alert and risk management, *Journal of Flood Risk Management*, *9 4*, 402-415.

Castellvı, F., Mormeneo, I. & Perez, P., (2004), Generation of daily amounts of precipitation from standard climatic data : a case study for Argentina, *Journal of hydrology*, *289 1-4*, 286-302.

Cecinati, F., Wani, O. & Rico-Ramirez, M. A., (2017), Comparing Approaches to Deal With Non-Gaussianity of Rainfall Data in Kriging-Based Radar-Gauge Rainfall Merging, *Water Resources Research*, *53 11*, 8999-9018.

Champeaux, J.-L., Cheze, J.-L. & Tabary, P., (2012), The French operational radar network and products, *7th European Conference on Radar in Meteorology and Hydrology (Toulouse, France), 24-29 June 2012*.

Chan, A., (2016), Generation of typical meteorological years using genetic algorithm for different energy systems, *Renewable Energy*, *90*, 1-13.

Chen, L. & Ng, R., (2004), On the marriage of lp-norms and edit distance, *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, 792-803.

Chiles, J.-P. & Delfiner, P., (2009), *Geostatistics : modeling spatial uncertainty* (T. 497), John Wiley & Sons.

Chumchean, S., Sharma, A. & Seed, A., (2006), An integrated approach to error correction for real-time radar-rainfall estimation, *Journal of Atmospheric and Oceanic Technology*, *23 1*, 67-79.

Craciun, C. & Catrina, O., (2016), An objective approach for comparing radar estimated and rain gauge measured precipitation, *Meteorological Applications*, *23 4*, 683-690.

D'amato, N. & Lebel, T., (1998), On the characteristics of the rainfall events in the Sahel with a view to the analysis of climatic variability, *International Journal of Climatology : A Journal of the Royal Meteorological Society*, *18 9*, 955-974.

Darling, D. A., (1957), The kolmogorov-smirnov, cramer-von mises tests, *The Annals of Mathematical Statistics*, *28 4*, 823-838.

Dérian, P., Héas, P., Herzet, C. & Mémin, E., (2013), Wavelets and optical flow motion estimation, *Numerical Mathematics : Theory, Methods and Applications*, *6*, 116-137.

Desbordes, M., (1974), *Réflexions sur les méthodes de calcul des réseaux urbains d'assainissement pluvial* (thèse de doct.), Université des Sciences et Techniques du Languedoc.

Desbordes, M. & Raous, P., (1980), Fondaments de l'élaboration d'une pluie de projet urbaine. Méthodes d'analyse et applicationa la station Montpellier Bel Air, *La Météorologie, VI série*, *20 21*, 317-326.

Dobler, C., Hagemann, S., Wilby, R. & Stötter, J., (2012), Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed, *Hydrology and Earth System Sciences*, *16 11*, 4343-4360.

Doviak, R. J., Zrnic, D. S. & Sirmans, D. S., (1979), Doppler weather radar, *Proceedings of the IEEE*, *67 11*, 1522-1553.

Dunkerley, D., (2008), Identifying individual rain events from pluviograph records : a review with analysis of data from an Australian dryland site, *Hydrological Processes : An International Journal*, *22 26*, 5024-5036.

Dunkerley, D., (2015), Intra-event intermittency of rainfall : An analysis of the metrics of rain and no-rain periods, *Hydrological Processes*, *29 15*, 3294-3305.

Durbán, M. & Glasbey, C., (2001), Weather modelling using a multivariate latent Gaussian model, *Agricultural and Forest Meteorology*, *109 3*, 187-201.

Ehret, U., (2003), *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar, Mitteilungen Instit.*

Embrechts, P., Klüppelberg, C. & Mikosch, T., (2013), *Modelling extremal events : for insurance and finance* (T. 33), Springer Science & Business Media.

Esling, P. & Agon, C., (2012), Time-series data mining, *ACM Computing Surveys (CSUR)*, *45 1*, 12.

Fan, J. & Zhang, W., (2004), Generalised likelihood ratio tests for spectral density, *Biometrika*, *91 1*, 195-209.

Festa, R. & Ratto, C. F., (1993), Proposal of a numerical procedure to select reference years, *Solar Energy*, *50 1*, 9-17.

Finkelstein, J. M. & Schafer, R. E., (1971), Improved goodness-of-fit tests, *Biometrika*, *58 3*, 641-645.

Fréchet, M. M., (1906), Sur quelques points du calcul fonctionnel, *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, *22 1*, 1-72.

Genz, A., (1992), Numerical computation of multivariate normal probabilities, *Journal of computational and graphical statistics*, *1 2*, 141-149.

Goerg, G. M., (2015), The Lambert way to Gaussianize heavy-tailed data with the inverse of Tukey'sh transformation as a special case, *The Scientific World Journal*, *2015*.

Gordon, H., Whetton, P., Pittock, A., Fowler, A. & Haylock, M., (1992), Simulated changes in daily rainfall intensity due to the enhanced greenhouse effect : implications for extreme rainfall events, *Climate Dynamics*, *8 2*, 83-102.

Goudenhoofdt, E. & Delobbe, L., (2009), Evaluation of radar-gauge merging methods for quantitative precipitation estimates, *Hydrology and Earth System Sciences*, *13 2*, 195-203.

Gradshteyn, I. S. & Ryzhik, I. M., (2007), *Table of integrals, series, and products* (A. Jeffrey & D. Zwillinger, Éd. ; 7th), Academic press.

Guillot, G. & Lebel, T., (1999), Disaggregation of Sahelian mesoscale convective system rain fields : Further developments and validation, *Journal of Geophysical Research : Atmospheres*, *104 D24*, 31533-31551.

Haberlandt, U., (2007), Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, *Journal of Hydrology*, *332 1-2*, 144-157.

Hall, I. J., Prairie, R., Anderson, H. & Boes, E., (1978), *Generation of a typical meteorological year* (rapp. tech.), Sandia Labs., Albuquerque, NM (USA).

Hall, J. W., Boyce, S. A., Wang, Y., Dawson, R. J., Tarantola, S. & Saltelli, A., (2009), Sensitivity analysis for hydraulic models, *Journal of Hydraulic Engineering*, *135 11*, 959-969.

Harrison, D., Driscoll, S. & Kitchen, M., (2000), Improving precipitation estimates from weather radar using quality control and correction techniques, *Meteorological Applications : A journal of forecasting, practical applications, training techniques and modelling*, *7 2*, 135-144.

Huang, Y., (2011), International Weather for Energy Calculations (IWEC Weather Files) Users Manual.

Hussain, I., Spöck, G., Pilz, J. & Yu, H.-L., (2010), Spatio-temporal interpolation of precipitation during monsoon periods in Pakistan, *Advances in water resources*, *33 8*, 880-886.

Isaaks, E. H. & Srivastava, M. R., (1989), *Applied geostatistics.*

Ismailaja, N. et al., (2015), Comparing the efficiency of CID distance and CORT coefficient for finding similar subsequences in time series.

Joe, H. & Xu, J. J., (1996), The estimation method of inference functions for margins for multivariate models.

Johnson, J., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D. & Thomas, K. W., (1998), The storm cell identification and tracking algorithm : An enhanced WSR-88D algorithm, *Weather and forecasting*, *13 2*, 263-276.

Jones, C. & Macpherson, B., (1997), A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model, *Meteorological Applications : A journal of forecasting, practical applications, training techniques and modelling*, *4 3*, 269-277.

Katz, R. W., (1999), Extreme value theory for precipitation : sensitivity analysis for climate change, *Advances in water resources*, *23 2*, 133-139.

Keogh, E. & Kasetty, S., (2003), On the need for time series data mining benchmarks : a survey and empirical demonstration, *Data Mining and knowledge discovery*, *7 4*, 349-371.

Keogh, E., Lonardi, S. & Ratanamahatana, C. A., (2004), Towards parameter-free data mining, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 206-215.

Khudri, M. M. & Sadia, F., (2013), Determination of the best fit probability distribution for annual extreme precipitation in Bangladesh, *Eur J Sci Res*, *103 3*, 391-404.

Kleiber, W., Katz, R. W. & Rajagopalan, B., (2012), Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, *Water Resources Research*, *48 1.*

Krajewski, W. F., (1987), Cokriging radar-rainfall and rain gage data, *Journal of Geophysical Research : Atmospheres*, *92 D8*, 9571-9580.

Krajewski, W. F., Lakshmi, V., Georgakakos, K. P. & Jain, S. C., (1991), A Monte Carlo study of rainfall sampling effect on a distributed catchment model, *Water Resources Research*, *27 1*, 119-128.

Kullback, S. & Leibler, R. A., (1951), On information and sufficiency, *The annals of mathematical statistics*, *22 1*, 79-86.

Kyznarová, H. & Novák, P., (2009), CELLTRACK—Convective cell tracking algorithm and its use for deriving life cycle characteristics, *Atmospheric research*, *93 1-3*, 317-327.

Larvor, G., Berthomier, L., Chabot, V., Le Pape, B., Pradel, B. & Perez, L., (2020), MeteoNet, an open reference weather dataset by METEO FRANCE [Available online at https://meteonet.umr-cnrm.fr/].

Le Grand Lyon, (2008), Méthode pour le dimensionnement des ouvrages de stockage.

Liao, T. W., (2005), Clustering of time series data—a survey, *Pattern recognition*, *38 11*, 1857-1874.

Lien, G.-Y., Kalnay, E. & Miyoshi, T., (2013), Effective assimilation of global precipitation : Simulation experiments, *Tellus A : Dynamic Meteorology and Oceanography*, *65 1*, 19915.

Lien, G.-Y., Miyoshi, T. & Kalnay, E., (2016), Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecast system, *Monthly Weather Review*, *144 2*, 643-661.

Lin, J., Keogh, E., Lonardi, S. & Chiu, B., (2003), A symbolic representation of time series, with implications for streaming algorithms, *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2-11.

Liu, Y., Zhang, W., Shao, Y. & Zhang, K., (2011), A comparison of four precipitation distribution models used in daily stochastic models, *Advances in Atmospheric Sciences*, *28 4*, 809-820.

Luini, L., Jeannin, N., Capsoni, C., Paraboni, A., Riva, C., Castanet, L. & Lemorton, J., (2011), Weather radar data for site diversity predictions and evaluation of the impact of rain field advection, *International Journal of satellite communications and networking*, *29 1*, 79-96.

Lund, H., (1995), The Design Reference Year user's manual, Thermal Insulation Laboratory, *Technical University of Denmark : Lyngby.*

Mahalanobis, P. C., (1936), On the generalized distance in statistics, *Proceedings of the National Institute of Sciences (Calcutta)*, *2*, 49-55.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M. et al., (2010), Precipitation downscaling

under climate change : Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, *48 3*.

Marshall, J. S. & Palmer, W. M. K., (1948), The distribution of raindrops with size, *Journal of Atmospheric Sciences*, *5 4*, 165-166, https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2

Marteau, P.-F. & Gibet, S., (2015), On recursive edit distance kernels with application to time series classification, *IEEE transactions on neural networks and learning systems*, *26 6*, 1121-1133.

Matheron, G., (1965), *Traite de geostatistique appliquee. Tome II. Le krigeage*, Editions B.R.G.M., https://books.google.fr/books?id=J7-3swEACAAJ

Matheron, G., (1973), The intrinsic random functions and their applications, *Advances in applied probability*, *5 3*, 439-468.

McDonald, J. B. & Turley, P., (2011), Distributional characteristics : Just a few more moments, *The American Statistician*, *65 2*, 96-103.

Meischner, P., (2005), *Weather radar : principles and advanced applications*, Springer Science & Business Media.

Merz, J., Dangol, P. M., Dhakal, M. P., Dongol, B. S., Nakarmi, G. & Weingartner, R., (2006), Rainfall-runoff events in a middle mountain catchment of Nepal, *Journal of hydrology*, *331 3-4*, 446-458.

Mori, U., Mendiburu, A., Lozano, J. & Mori, M. U., (2015), Package 'TSdist'.

Mori, U., Mendiburu, A. & Lozano, J. A., (2016), Distance Measures for Time Series in R : The TSdist Package, *R JOURNAL*, *8 2*, 451-459.

Muthen, B., (1990), Moments of the censored and truncated bivariate normal distribution, *British Journal of Mathematical and Statistical Psychology*, *43 1*, 131-143.

Naveau, P., Huser, R., Ribereau, P. & Hannart, A., (2016), Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, *Water Resources Research*, *52 4*, 2753-2769.

Null, S. E. & Viers, J. H., (2013), In bad waters : Water year classification in nonstationary climates, *Water Resources Research*, *49 2*, 1137-1148.

O'Gorman, P. A., (2015), Precipitation extremes under climate change, *Current climate change reports*, *1 2*, 49-59.

Otsuka, S., Kotsuki, S. & Miyoshi, T., (2016), Nowcasting with data assimilation : A case of global satellite mapping of precipitation, *Weather and Forecasting*, *31 5*, 1409-1416.

Panofsky, H. A., Brier, G. W. & Best, W. H., (1958), *Some application of statistics to meteorology*, Earth ; Mineral Sciences Continuing Education, College of Earth and ...

Papalexiou, S., Koutsoyiannis, D. & Makropoulos, C., (2013), How extreme is extreme ? An assessment of daily rainfall distribution tails., *Hydrology & Earth System Sciences*, *17 1*.

Rebora, N., Ferraris, L., von Hardenberg, J. & Provenzale, A., (2006), RainFARM : Rainfall downscaling by a filtered autoregressive model, *Journal of Hydrometeorology*, *7 4*, 724-738.

Schilling, W., (1991), Rainfall data for urban hydrology : what do we need ?, *Atmospheric Research*, *27 1-3*, 5-21.

Schleiss, M., Jaffrain, J. & Berne, A., (2011), Statistical analysis of rainfall intermittency at small spatial and temporal scales, *Geophysical research letters*, *38 18*.

Schoof, J., (2008), Application of the multivariate spectral weather generator to the contiguous United States, *Agricultural and forest meteorology*, *148 3*, 517-521.

Selker, J. S. & Haith, D. A., (1990), Development and Testing of Single-Parameter Precipitation Distributions, *Water resources research*, *26 11*, 2733-2740.

Seo, D.-J., (1998), Real-time estimation of rainfall fields using radar rainfall and rain gage data, *Journal of Hydrology*, *208 1*, 37-52, https://doi.org/https://doi.org/10.1016/S0022-1694(98)00141-3

Seo, D.-J. & Breidenbach, J., (2002), Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements, *Journal of Hydrometeorology*, *3 2*, 93-111.

Shoji, T. & Kitaura, H., (2006), Statistical and geostatistical analysis of rainfall in central Japan, *Computers & Geosciences*, *32 8*, 1007-1024.

Sigrist, F., Künsch, H. R., Stahel, W. A. et al., (2012), A dynamic nonstationary spatio-temporal model for short term prediction of precipitation, *The Annals of Applied Statistics*, *6 4*, 1452-1477.

Sinclair, S. & Pegram, G., (2005), Combining radar and rain gauge rainfall estimates using conditional merging, *Atmospheric Science Letters*, *6 1*, 19-22.

Smith, J. A. & Krajewski, W. F., (1991), Estimation of the mean field bias of radar rainfall estimates, *Journal of Applied Meteorology and Climatology*, *30 4*, 397-412.

Soubeyroux, J.-M., Bernus, S., Corre, L., Drouin, A., Dubuisson, B., Etchevers, P., Gouget, V., Josse, P., Kerdoncuff, M., Samacoits, R. & Tocquer, F., (2020), *Les nou-*

velles projections climatiques de référence DRIAS-2020 pour la métropole (rapp. tech.), Météo France, CNRM, Cerfacs et IPSL.

Tabary, P., (2007), The new French operational radar rainfall product. Part I : Methodology, *Weather and forecasting*, *22 3*, 393-408.

Todini, E., (2001), A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements, *Hydrology and Earth System Sciences*, *5 2*, 187-199.

Trenberth, K. E., (2011), Changes in precipitation with climate change, *Climate Research*, *47 1-2*, 123-138.

Tukey, J. W., (1977), Modern techniques in data analysis, *Proceedings of the NSF-Sponsored Regional Research Conference*, *7*.

Tuttle, J. D. & Foote, G. B., (1990), Determination of the boundary layer airflow from a single Doppler radar, *Journal of Atmospheric and oceanic Technology*, *7 2*, 218-232.

Veneziano, D. & Lepore, C., (2012), The scaling of temporal rainfall, *Water Resources Research*, *48 8*.

Vlachos, M., Kollios, G. & Gunopulos, D., (2002), Discovering similar multidimensional trajectories, *Data Engineering, 2002. Proceedings. 18th International Conference on*, 673-684.

Vlček, O. & Huth, R., (2009), Is daily precipitation Gamma-distributed ? : Adverse effects of an incorrect use of the Kolmogorov–Smirnov test, *Atmospheric Research*, *93 4*, 759-766.

Wagner, R. A. & Fischer, M. J., (1974), The string-to-string correction problem, *Journal of the ACM (JACM)*, *21 1*, 168-173.

Wagner, S. & Wagner, D., (2007), *Comparing clusterings : an overview*, Universität Karlsruhe, Fakultät für Informatik Karlsruhe.

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P. & Keogh, E., (2013), Experimental comparison of representation methods and distance measures for time series data, *Data Mining and Knowledge Discovery*, 1-35.

Wasko, C., Sharma, A. & Rasmussen, P., (2013), Improved spatial prediction : A combinatorial approach, *Water Resources Research*, *49 7*, 3927-3935.

Watterson, I. & Dix, M., (2003), Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution, *Journal of Geophysical Research : Atmospheres*, *108 D13*.

Wilks, D. S., (1999), Interannual variability and extreme-value characteristics of several stochastic daily precipitation models, *Agricultural and forest meteorology*, *93 3*, 153-169.

Wilks, D., (1998), Multisite generalization of a daily stochastic precipitation generation model, *journal of Hydrology*, *210 1-4*, 178-191.

Woolhiser, D. A. & Roldan, J., (1982), Stochastic daily precipitation models : 2. A comparison of distributions of amounts, *Water resources research*, *18 5*, 1461-1468.

Xu, G. & Genton, M. G., (2017), Tukey g-and-h random fields, *Journal of the American Statistical Association*, *112 519*, 1236-1249.

Yang, L., Wan, K. K., Li, D. H. & Lam, J. C., (2011), A new method to develop typical weather years in different climates for building energy use studies, *Energy*, *36 10*, 6121-6129.

Zhang, H., Ho, T. B., Zhang, Y. & Lin, M.-S., (2006), Unsupervised feature extraction for time series clustering using orthogonal wavelet transform, *Informatica*, *30 3*.

Zhao, T., Bennett, J. C., Wang, Q., Schepen, A., Wood, A. W., Robertson, D. E. & Ramos, M.-H., (2017), How suitable is quantile mapping for postprocessing GCM precipitation forecasts ?, *Journal of Climate*, *30 9*, 3185-3196.

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UBO
Université de Bretagne Occidentale

Titre : Analyse statistique de la pluviométrie et de son impact sur le dimensionnement des réseaux d'assainissement

Mot clés : Précipitations, Statistiques, Modélisation, Hydraulique urbaine, Assainissement

Résumé : Dans les zones urbaines, où une part importante des réseaux d'assainissement est unitaire, des déversements d'eaux usées en milieu naturel peuvent avoir lieu lors d'événements pluvieux. Le projet MEDISA vise à établir une méthodologie de dimensionnement des réseaux d'assainissement afin de mieux les gérer. La pluviométrie est un forçage central pour les modèles hydrauliques qui décrivent les réseaux d'assainissement. Le dimensionnement dépend fortement de ce forçage, il est donc nécessaire d'étudier la sensibilité du modèle à différents facteurs afin de déterminer les forçages à utiliser. Les modèles hydrauliques nécessitent des données à un pas de temps de quelques minutes. A cette échelle la distribution marginale de la pluie est caractérisée par 1) une forte occurrence de temps sec, 2) une discrétisation non négligeable et 3) une forme asymétrique à queue lourde. Un modèle méta-Gaussien basé sur ces caractéristiques est développé, pouvant convenir à des données allant de quelques minutes à l'échelle mensuelle et présentant l'avantage de lier chaque paramètre à une partie différente de la distribution. Une analyse des données de pluie disponibles dans la zone d'étude est réalisée. Les pluviomètres et les radars météorologiques contiennent des erreurs de mesure et ne sont pas représentatifs de la vraie pluie, aussi une méthode de fusion de données est développée pour tirer profit des deux sources de données et tester la sensibilité du modèle hydraulique.

Title: Statistical analysis of precipitation and its incidence on sewerage networks sizing

Keywords: Precipitation, Statistics, Modelling, Urban hydraulics, Sewerage

Abstract: In urban areas, where an important part of the sewerage system is combined, waste water dumping can occur during rainy weather. The MEDISA project aims at developing a methodology for designing sewerage networks in order to better handle water dumpings. Precipitation is a central forcing for hydraulic models that describe sewerage networks. The optimal design is strongly dependant on this forcing, hence studying the sensitivity of the hydraulic model to different factors is necessary to determine the forcing to be used. Hydraulic models require data at few minutes time step. At this scale the marginal distribution of precipitation is characterised by 1) a strong occurrence of dry weather, 2) a non negligible discretization and 3) an asymmetric shape with a heavy tail. A meta-Gaussian model based on those characteristics is developed, capable of fitting data from few minutes to monthly scale and with the advantage of linking each parameter to a different part of the distribution. An analysis of precipitation data available in the area is executed. Both rain gauges and meteorological radar can contain measurement errors and are not representative of the true rainfall. A merging technique is therefore developed to take advantage of both data sources and assess the sensitivity of the hydraulic model.