

ANALYSE SPATIALE DE DONNÉES HÉTÉROGÈNES DE PRÉCIPITATION, APPLICATION À L'HYDROLOGIE URBAINE

Marie Boutigny ^{1,2} & Pierre Ailliot ¹ & Benoît Saussol ¹ & Antoine Siquin ²

¹ *Eau du Ponant, 210 Boulevard François Mitterrand 29490 Guipavas*

² *Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest*

Corresponding author : marie.boutigny@eauduponant.fr

Résumé. Dans les zones urbaines, où une part importante des réseaux d'égouts est unitaire, des déversements d'eaux usées peuvent avoir lieu lors d'événements pluvieux. Les ouvrages de collecte et de stockage d'eaux usées sont généralement dimensionnés grâce à des modèles hydrologiques qui décrivent le fonctionnement des réseaux. Les conditions météorologiques telles que l'évapotranspiration et surtout les précipitations sont des forçages centraux pour de tels modèles. L'approche usuelle consiste à prendre jusqu'à 5 ans de données météorologiques, considérées comme "représentatives" et observées en un site proche du système. Cela permet d'estimer la distribution statistique des déversements avec un coût de calcul raisonnable. Afin d'estimer la sensibilité du modèle au choix des conditions météorologiques "représentatives", on propose de développer un générateur stochastique de conditions météorologiques (Wilks & Wilby (1999), Ailliot et al. (2015)) afin de simuler un grand nombre (e.g. plusieurs siècles) de séries météorologiques spatio-temporelles réalistes. Ces séries artificielles peuvent être utilisées en entrée du modèle hydrologique afin d'estimer la sensibilité du modèle à différentes choses comme par exemple la longueur de la série ou la spatialisation de la pluie, et finalement de déterminer les forçages météorologiques à utiliser dans les études visant à dimensionner les systèmes d'assainissement.

Mots-clés. Statistique spatiale, précipitation, générateurs aléatoires de conditions météorologiques, fusion de données, hydrologie urbaine

Abstract. In urban areas, where an important part of the sewerage system is combined, waste water dumping can occur during rainy weather. Water collection and storage systems are generally designed using hydrological models which describe the functioning of the sewerage systems. Weather conditions such as evapotranspiration and most of all precipitation are very important forcing factors for such models. An usual approach to take into account the variability of weather conditions consists in forcing the hydrological model with up to 5 years of meteorological data, considered as "representative" and observed in a site close to the system location. This permits to estimate the statistical distribution of water dumping with reasonable computational cost. In order to assess the sensitivity of the method to the choice of the "representative" meteorological conditions, we propose to develop a stochastic weather generator (Wilks & Wilby (1999), Ailliot et al. (2015)) to simulate a high number (e.g. several centuries) of realistic spatio-temporal meteorological series. These artificial weather conditions can then be used as input to the hydrological model. This will allow us to estimate its sensitivity to different things, such

as the length of the input time series or rainfall spatialization, and finally determine the forcing weather conditions to be used in studies aiming at designing sewage systems.

Keywords. Spatial statistics, rainfall, stochastic weather generator, data merging, urban hydrology

1 Contexte

Les systèmes d’assainissement forment des réseaux complexes et peuvent être de 2 types (cf. figure 1) : unitaire quand les eaux usées et les eaux de pluies sont récupérées par le même réseau, séparatif quand elles sont traitées différemment. Dans les zones urbaines, particulièrement en centre ville, la majorité du réseau est en unitaire, ce qui peut entraîner en cas de fortes pluies des saturations du réseau. Un mélange d’eau de pluie et d’eaux usées est alors déversé directement dans le milieu naturel. Le projet MEDISA, porté par Eau du Ponant, vise à établir une méthodologie de dimensionnement des réseaux d’assainissement afin de mieux gérer les déversements en milieu naturel.

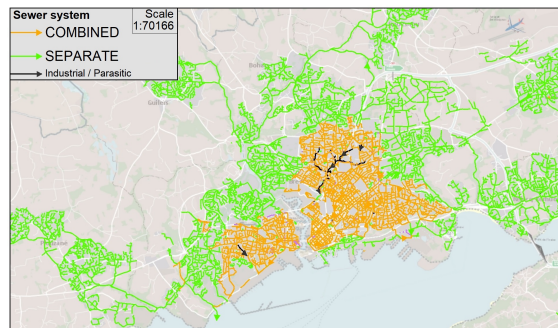


FIGURE 1 – Réseau d’assainissement de Brest Métropole

Un modèle hydrologique décrit le fonctionnement du système d’assainissement de Brest, prenant en entrée les conditions météorologiques (précipitation et évapotranspiration) et donnant les chroniques de déversement aux différents points de rejet du réseau. La méthode usuelle de dimensionnement des réseaux consiste à faire tourner le modèle hydrologique avec une série météorologique de 5 ans considérée comme typique.

Le projet MEDISA souhaite étudier la sensibilité du modèle hydrologique à différents facteurs et déterminer les forçages météorologiques à utiliser pour dimensionner les systèmes d’assainissement. Pour cela on propose de développer un générateur stochastique de conditions météorologiques, qui permettra de générer un grand nombre de séries spatio-temporelles réalistes. Utilisées en entrée du modèle hydrologique, elles permettront d’étudier la sensibilité à différents facteurs tels que la longueur de la série ou la spatialisation de la pluie, et donc de déterminer les forçages à utiliser.

Parmi les challenges à développer un tel générateur stochastique de conditions météorologiques, on peut citer

1. les observations : différentes sources de données sont disponibles, donnant des informations différentes et ayant chacune leurs forces et faiblesses,
2. la résolution : une résolution temporelle fine est exigée par le modèle hydrologique (3 à 15 minutes), et une résolution spatiale fine (avec un réseau dense de pluviomètres et des données radar sur 1km^2) est aussi nécessaire pour reproduire les événements convectifs,
3. la nécessité de reproduire des statistiques complexes qui peuvent impacter les déversements comme la durée des périodes sèches et humides, les déplacements des cellules pluvieuses avec des vitesses et directions variables, ou encore la relation entre plusieurs variables météorologiques (précipitation, vent, évapotranspiration),
4. la nature non gaussienne des précipitations,
5. la taille des jeux de données spatio-temporels (données radar).

2 Données

Trois sources de données sont disponibles.

1. Pluviomètres :

- (a) Météo France fournit depuis 2006 des données à un pas de temps de 6 minutes mesurées à la station de Guipavas. Ces données sont considérées comme très fiables car corrigées en temps réel.
 - (b) Eau du Ponant possède 11 pluviomètres répartis sur Brest, en fonctionnement depuis 2010 pour la plupart, à un pas de temps de 3 minutes. Les données sont brutes et peuvent donc contenir beaucoup d'erreurs de mesure, des valeurs aberrantes en cas d'opération de maintenance par exemple (cf. figure 2). Ces pluviomètres offrent une couverture spatiale satisfaisante et fonctionnent de manière indépendante.
2. **Radar** : le radar de Plabennec de Météo France couvre la région d'étude, il fournit des données depuis 2006 à un pas de temps de 5 minutes, sur une grille de 1 km^2 . Les données radar sont traditionnellement considérées comme moins fiables que les pluviomètres, la relation entre la réflectivité mesurée et l'intensité de pluie n'étant pas parfaite. Elles ont tendance à lisser les données car on a une moyenne sur 1km^2 , et enfin l'algorithme de post-traitement du radar a beaucoup évolué depuis 2006, rendant cette base de données très hétérogène. Toutefois le radar donne une information essentielle d'une part sur la structure d'ordre deux de la zone d'étude et d'autre part sur les déplacements des cellules pluvieuses.

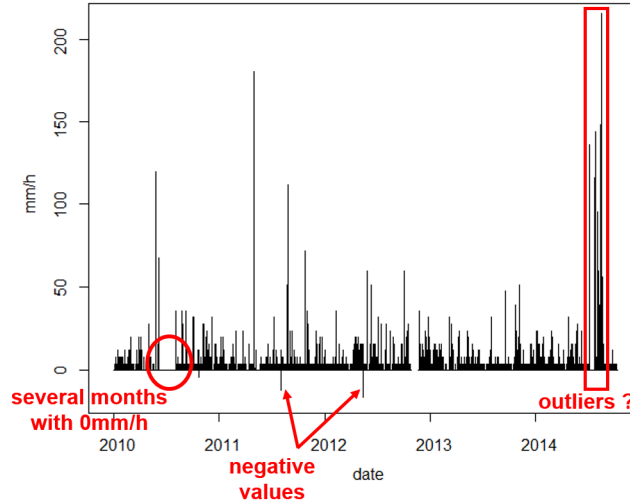


FIGURE 2 – Exemple de chronique d’un pluviomètre d’Eau du Ponant

Ces données hétérogènes semblent difficilement exploitables pour construire directement un générateur aléatoire de conditions météorologiques. Une étape préliminaire de fusion de données est utile pour homogénéiser ces données en corrigeant les erreurs de mesure et en interpolant les données sur une grille régulière avec une discrétisation spatio-temporelle adaptée à l’hydrologie urbaine. Nous allons nous focaliser sur cette étape de fusion de données dans cet exposé.

3 Modèle

Nous proposons d’utiliser un modèle hiérarchique pour fusionner les jeux de données décrits dans le paragraphe précédent.

Modélisation de la distribution spatiale des précipitations.

Un processus gaussien latent censuré et transformé, similaire à celui présenté dans Benoit et al. (2018), Allcroft & Glasbey (2003), est utilisé pour modéliser la distribution spatiale des précipitations. Plus précisément, nous supposons qu’il existe un champ spatial gaussien X

$$X \sim N(\mu, \Sigma), \quad X = (X_1, \dots, X_n)'$$

Le passage du champ gaussien X aux intensités de précipitation Z se fait ensuite via une transformation Ψ qu’on écrit comme suit en un point s

$$Z(s) = \Psi(X(s)) = 0 * \mathbb{I}_{X(s) \leq 0} + \psi(X(s)) * \mathbb{I}_{X(s) \geq 0}.$$

La censure du champ gaussien en 0 permet de reproduire les zones de temps sec, et la fonction $\psi(\cdot)$, souvent appelée anamorphose, transforme la partie positive en intensité de pluie. Plusieurs formes sont proposées par la littérature, par exemple un mélange de puissance et fonction quadratique dans Allcroft & Glasbey (2003) ou un mélange de puissance et d'exponentielle dans Allard & Bourotte (2015). Dans notre cas, à l'échelle qui nous intéresse, c'est cette dernière qui est la plus adaptée. On supposera donc $\psi(x) = b(e^{ax^c} - 1)$ avec a et b des paramètres inconnus.

Modélisation des erreurs d'observation.

Les erreurs d'observations des pluviomètres et du radar sont modélisées par des modèles additifs dans le domaine gaussien décrits plus précisément ci-dessous.

- L'erreur d'observation du radar a une structure spatiale, et est éventuellement de moyenne non nulle (par exemple négative, lorsque le radar a tendance à sous-estimer les intensités de pluie). On écrit donc l'équation d'observation comme suit

$$R(s) = \Psi(X(s) + \epsilon_R(s)), \quad \epsilon_R \sim N(\mu_R, \Sigma_R)$$

avec $R(s)$ l'observation radar au point s .

- Pour les pluviomètres, la mesure est discrète à cause leur fonctionnement à bascule (précision de 0.2mm) et nous avons trouvé que la prise en compte de cette discrétisation dans le modèle permet d'améliorer significativement la qualité de l'estimateur maximum de vraisemblance. Les erreurs d'observations des différents pluviomètres pouvant être considérées comme indépendantes, on choisit un bruit i.i.d., de moyenne nulle car on n'attend pas de biais systématique. L'équation d'observation est donc la suivante :

$$G(s) = 0.2 \left\lfloor \frac{\Psi(X(s) + \epsilon_G(s))}{0.2} \right\rfloor, \quad \epsilon_G(s) \sim N(0, \sigma_G^2) \text{ (i.i.d.)}$$

avec $G(s)$ l'observation donnée par le pluviomètre au site s et $\lfloor \cdot \rfloor$ la partie entière.

Inférence statistique

L'estimation des paramètres du modèle se fait par maximum de vraisemblance, avec une construction similaire à celle présentée dans Benoit et al. (2018). Notons $Y = (Y_{pluie}, Y_{sec})$ l'ensemble des observations séparées en deux blocs Y_{pluie} et Y_{sec} correspondant respectivement aux observations strictement positives et aux observations égales à 0. Les observations étant indépendantes conditionnellement au champ gaussien X (ou $X + \epsilon_R$ pour le radar), on verra apparaître schématiquement les trois termes ci-dessous dans la vraisemblance

1. la probabilité de temps de pluie, c'est-à-dire $p_X(\psi^{-1}(Y_{pluie}))$,
2. un facteur $1/\psi'(\psi^{-1}(Y_{pluie}))$ dû au changement de variable dans l'écriture de la densité des observations,
3. la probabilité de temps sec, $\mathbb{P}(Y_{sec}|X)$, qui s'exprime avec la fonction de répartition de la loi normale multivariée (probabilité d'Orthant) et qui peut être estimée numériquement.

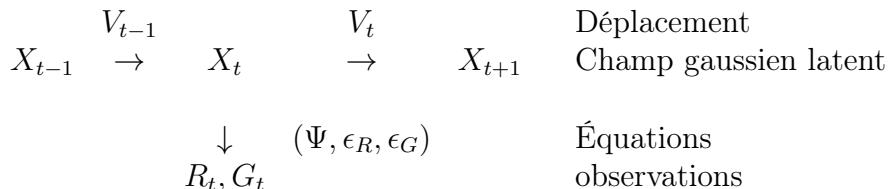
Pour les pluviomètres, la discrétisation rendra la vraisemblance différente et fera apparaître la fonction de répartition de la loi normale au lieu de sa densité.

Une fois les paramètres estimés, l'étape de fusion/reconstruction consiste à prédire le champ latent Y , qui décrit les vraies précipitations, conditionnellement aux observations disponibles.

4 Résultats

Dans cet exposé, nous présenterons différents résultats numériques obtenus avec le modèle hiérarchique décrit dans le paragraphe précédent sur les données de précipitation sur la région brestoise. Nous comparerons ces résultats avec d'autres méthodes de fusion de données qui ont été proposées dans la littérature.

Nous discuterons également une extension spatio-temporelle du modèle précédent, dans laquelle le champ latent sera modélisé en utilisant une dynamique markovienne qui inclut un terme d'advection (déplacement) estimé à partir des images radar, schématisée par le diagramme acyclique orienté ci-dessous.



Références

- Ailliot, P., Allard, D., Monbet, V. & Naveau, P. (2015), 'Stochastic weather generators : an overview of weather type models', *Journal de la Société Française de Statistique* **156**(1), 101–113.
- 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.
- Benoit, L., Allard, D. & Mariethoz, G. (2018), 'Stochastic rainfall modeling at sub-kilometer scale', *Water Resources Research* **54**(6), 4108–4130.
- Wilks, D. S. & Wilby, R. L. (1999), 'The weather generation game : a review of stochastic weather models', *Progress in Physical Geography* **23**(3), 329–357.