

STAGE MASTER 2 - INGENIEUR

Estimation de la distribution des tailles de gouttes de pluie comme un processus dynamique neuronal

Mots clefs : Machine Learning, Pluie, Gouttes de pluie, Drop Size Distribution (DSD), Processus Ponctuels, Temporal Point Process (TPP) , Neural TPP (NTPP), disdromètre, pluviomètre

Encadrant : Aymeric Chazottes (aymeric.chazottes@latmos.ipsl.fr)

Laboratoire : LATMOS (Equipe SPACE - Machine Learning et Cycle de l'eau)

Environnement de travail : Au sein de l'équipe SPACE du LATMOS le stage sera aussi encadré de façon plus collégiale par une équipe composée de quatre chercheurs experts en apprentissage statistique et/ou en processus physiques des précipitations.

Lieu : 11 Boulevard d'Alembert, 78280 Guyancourt

(accès par RER C, Train lignes N et U, Bus lignes 475 et 91-06)

Durée du stage : 6 mois avec début en Février/Mars/Avril

Rémunération : selon les règles du CNRS, soit environ 600€ net / mois + 50% carte Navigo

Possibilité de poursuite en thèse sur des sujets connexes sous réserve de financement

Moyens de calcul disponibles : cluster GPU

Description du sujet :

La pluie est un sujet à enjeux. Avec les extrêmes climatiques qui lui sont associés (allant de la sécheresse aux inondations) elle a des conséquences sociétales considérables. La pluie est aussi diverse. Avec sa prédiction, elle est au cœur de nombreux domaines scientifiques (météorologie, hydrologie, télécommunication, ...). Selon le sujet d'intérêt elle est considérée via des capteurs ayant des échelles et des procédés de mesure différents. Le plus souvent considérée comme une grandeur continue elle peut aussi être vue comme un ensemble discret de gouttes (distribuées dans l'espace et le temps). Il n'est pas toujours simple d'établir la correspondance entre les visions macro-physique (intensité de pluie) et microphysique (distribution de taille des gouttes).

L'interaction des ondes électromagnétiques avec l'atmosphère précipitante est influencée par la microphysique. La mesure d'un radar météorologique ou l'affaiblissement du signal lors d'une liaison de télécommunication sera donc liée à la répartition des gouttes dans l'atmosphère. L'hydrologue voudra lui connaître des volumes d'eau mesurés au niveau du sol. Dans cette perspective la pluie sera résumée par les mesures d'un pluviomètre. On la considérera comme un processus continu de taux précipitant (i.e. le volume d'eau d'un flux de gouttes captées par une surface sur un intervalle de temps).

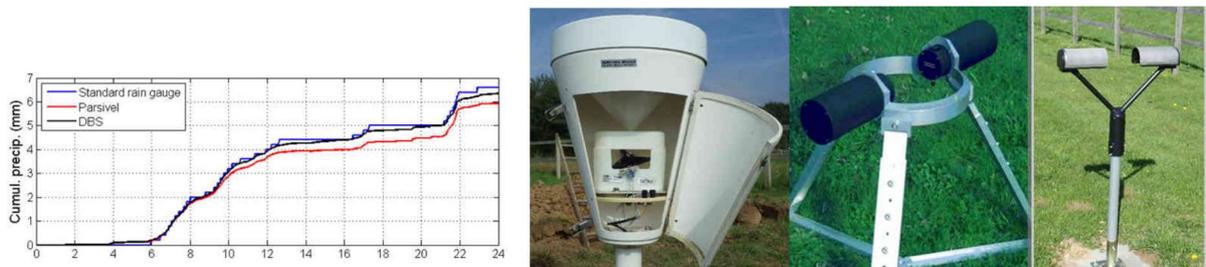


Figure 1 : Cumuls de pluie mesurés par différents capteurs de pluie (2 disdromètres et un pluviomètre).

Quelle que soit la perspective la pluie est un phénomène complexe, “non stationnaire” et “extrême”. Considérer la pluie comme un processus discret permet de la représenter plus

finement. On a ainsi une meilleure caractérisation des systèmes précipitants. Cela permet aussi de mieux contrôler, expliquer et interpréter les paramètres du modèle retenu. C'est la raison pour laquelle le travail proposé pour ce stage repose sur la mesure de la pluie à la résolution la plus fine et la plus détaillée (microphysique). Il est important de comprendre que le modèle incorpore aussi la représentation continue à plus grande échelle. Cette dernière peut être vue comme une intégration de ses sorties aux résolutions « macroscopiques ».

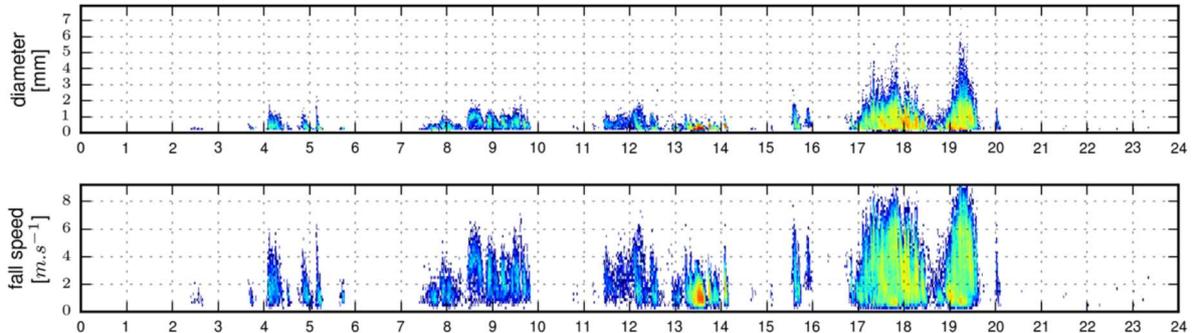


Figure 2 : Deux séries temporelles de pluies présentant la pluie comme une succession d'histogrammes de tailles de gouttes ou comme une succession d'histogrammes de vitesses de gouttes. (Pour une succession d'intervalles de temps on représente via un code couleur le nombre de gouttes correspondant à des intervalles de diamètres/vitesses de gouttes.)

Il existe de nombreux appareils de mesure de la pluie, nous nous intéresserons plus spécifiquement aux mesures fournies par un disdromètre. Cet appareil mesure les gouttes une à une dans une zone très localisée de l'espace. Notre jeu de données de départ sera donc constitué de triplets correspondant à une taille (D_i), une vitesse (V_i) et un temps d'arrivée (t_i). Généralement ce genre de série temporelle est modélisé par un processus ponctuel temporel (TPP pour l'anglais Temporal Point Process) avec marque (m_i). La marque est l'information qui vient en plus du temps (t_i), ici la taille et/ou la vitesse de la goutte. Il existe une littérature conséquente sur la modélisation stochastique de la pluie et de ses gouttes [1]. Il en découle une multitude de modèles dont certains incorporent un a priori sur la nature de la pluie [2] alors que d'autres sont purement stochastiques [3]. Ces différents modèles sont basés sur des hypothèses fortes qui conviennent mal à un phénomène aussi versatile que la pluie. Les contraintes/hypothèses du modèle doivent permettre une modélisation appropriée de la pluie et notamment assurer l'obtention de propriétés macroscopiques cohérente avec les mesures de la pluie. Il semble important de proposer une modélisation à même de prendre en compte les connaissances a priori tout en restant fidèle à la mesure. L'objectif du stage est de développer un modèle dont on puisse valider et comprendre les paramètres

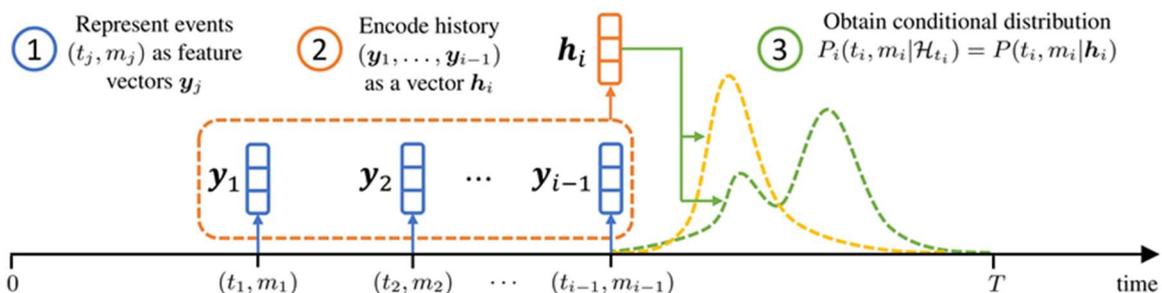


Figure 3: Modélisation neuronale d'un processus ponctuel temporel [4]. Le modèle apprend des probabilités conditionnelles (3) caractérisant les données à partir d'une série temporelle d'événements (1) i.e. les couples (t_i, m_i) . Pour cela, il doit connaître l'historique des événements précédents ainsi que l'état courant du modèle (2).

Récemment des modèles neuronaux NTPP (pour Neural TPP) [4][5] adaptés aux processus ponctuels inhomogènes avec marques ont été proposés. Ces modèles ont des architectures [5][6][7] leur permettant de prendre en compte les spécificités de la pluie tout en ajustant les propriétés propres aux séries temporelles apprises.

Déroulement du stage :

Dans un premier temps, il faudra se familiariser avec les données et des modèles. La prise en main du module EasyTPP permettra une implémentation aisée des processus ponctuels (TPP). En parallèle, il faudra s'appropriier une bibliographie succincte (fournie en début de stage). La prise en main des modèles neuronaux se fera via l'apprentissage de quelques jeux de données synthétiques de référence. La transition se fera naturellement vers l'apprentissage d'un jeu de données de séries temporelles de gouttes de pluie. Il s'agit d'une base de données contenant plusieurs années de séries temporelles acquises par un disdromètre. Des mesures de capteurs colocalisés (autre disdromètre mais aussi pluviomètre) permettront de valider les résultats. Si le temps le permet cette validation permettra aussi de caractériser la variabilité de la pluie et de caractériser les différences de précision entre les différents capteurs. Une possible extension du modèle pourra aussi être envisagée pour prendre en compte l'apprentissage simultané de deux types de capteurs.

Bibliographie :

- [1] Michele and Ignaccolo. New perspectives on rainfall from a discrete view. HP, 2013.
- [2] Lavergnat and Golé. A Stochastic Raindrop Time Distribution Model. JAM, 1998.
- [3] Waymire and Gupta. The mathematical structure of rainfall representations. WRR, 1981.
- [4] Shchur et al. Neural temporal point processes : A review. arXiv :2104.03528, 2021.
 - l'article : <https://arxiv.org/abs/2104.03528>
 - Présentation succincte : <https://youtu.be/J7qH7i0EyfU?feature=shared>
- [5] Xue et al. Easytp: Towards open benchmarking the TPP. arXiv:2307.08097, 2023.
 - l'article : <https://arxiv.org/abs/2307.08097>
 - le code-source : <https://github.com/ant-research/EasyTemporalPointProcess.git>
- [6] Lin et al. Exploring generative neural temporal point process. arXiv:2208.01874, 2022.
- [7] Bae et al. Meta temporal point processes. arXiv:2301.12023, 2023.

Résumé en anglais :

By considering droplets Rain can be viewed as a temporal point process. There are numerous stochastic models to characterize rain events. They rely on strong hypotheses and hence lack the flexibility to properly model rain measurements. The neural Temporal Point Processes [4][5] are flexible and they are able to surpass what is done with the actual stochastic rain models.