

STAGE MASTER 2 - INGENIEUR

Apprentissage de la pluie pour la validation du produits IMERG

Mots clefs : Machine Learning, Pluie, Auto-Encodeurs (AE), Variational Auto-Encoder (VAE), Topographic VAE, IMERG, Météo-France

Encadrants : Aymeric Chazottes (aymeric.chazottes@latmos.ipsl.fr) et Laurent Barthes (laurent.barthes@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 plus aussi co-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 un 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

Contexte du Stage :

Dans un contexte de changement climatique, l'estimation des champs précipitants depuis l'espace est un enjeu primordial. La mission GPM (Global Precipitation Measurement) a pour objectif de fournir des produits de précipitations à l'échelle du globe à partir d'une constellation de satellites [1]. Nous nous intéresserons en particulier au produit (image) IMERG (Integrated Multi-satellitE Retrievals for GPM) qui fournit des observations de précipitations à l'échelle globale avec une résolution temporelle de 30 minutes et une résolution spatiale d'une dizaine de kilomètres sous nos latitudes, le rendant ainsi particulièrement adapté pour des études à l'échelle régionale.

La qualité de restitution de ce produit composite/multi-capteurs dépend de plusieurs facteurs inhérents à la mesure, à la configuration géométrique et à la constellation de satellites. Plusieurs études ont déjà été réalisées pour évaluer la qualité de ce produit. Les cartes de précipitation IMERG y sont comparées à des observations de pluviomètres ou de radars météorologiques [2][3]. Des éléments importants comme l'heure de passage des satellites ou la nature des capteurs utilisés (pour obtenir le produit) ne sont pas pris en compte par ces études.

Sujet et objectif du Stage :

Dans le cadre de ce stage nous nous concentrerons sur une validation des produits/images IMERG au-dessus de la France métropolitaine. L'objectif étant de réaliser une validation au niveau du pixel mais aussi à la résolution temporelle du produit. Pour cela, nous aurons besoin de références (produits pluie Météo-France) mais aussi d'un algorithme (T-VAE). La qualité des produits IMERG (images de résolutions spatiale de 10km et temporelle de 30min) sera basée sur un produit pluie de référence (images de résolutions spatiale de 1km et temporelle de 5min) issu du réseau de radars météorologiques ARAMIS fourni par Météo France. Pour cela, on pourra se baser sur une base de quatre années d'observation d'une zone de 800x800 km² centrée sur la France. (Le travail préliminaire de colocalisation des produits IMERG et des produits pluie de Météo-France a déjà été réalisé.)

Les produits IMERG et Météo-France étant de natures différentes il est nécessaire d'établir une correspondance entre les deux. En tirant parti des données colocalisées, l'objectif est d'obtenir un espace de représentation commun à ces deux produits. Dans les faits, cet espace commun sera l'espace latent associé à un Auto-Encodeur (AE).

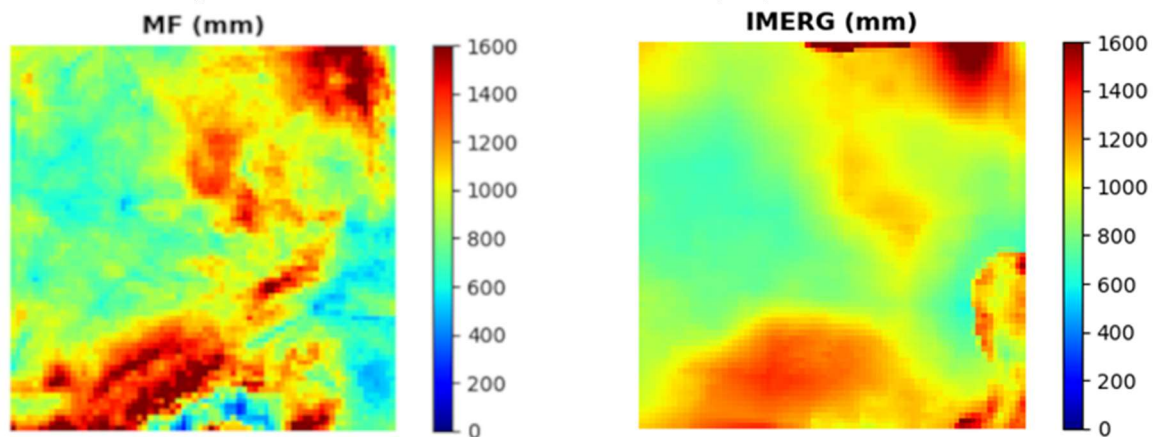


Figure 1 : Deux images de pluie. La première, à gauche, est issue de mesures de la pluie à partir de radars sols (produit de pluie de référence Météo-France). La seconde, à droite, est issue de mesures de la pluie faites de l'espace (produit IMERG)

Dans un premier temps, il s'agira d'apprendre les images Météo-France pour obtenir un espace de représentation compressé et cohérent de ces données. On s'intéresse à un Auto-Encodeur Variationnel (VAE pour l'anglais Variational Auto-Encoder). Ce modèle génératif permet d'obtenir un espace latent dont l'organisation a été le sujet de travaux précédents (et accessibles en début de stage). Le jeu d'images Météo-France étant constitué de séries temporelles, il serait regrettable de ne pas en tirer parti lors de l'apprentissage de l'espace latent. Ce sera l'objectif de ce stage. Des premiers tests seront conduits sur l'apport d'un Topographic-VAE [4] (T-VAE). Comme il a été conçu pour prendre en compte la « topographie » associée aux séries temporelles il permettra d'obtenir un espace latent structuré pour pouvoir prendre en compte la dynamique des précipitations.

Dans un second temps, il s'agira de travailler à générer des images/produits IMERG à partir de l'espace latent appris sur les images Météo-France. Pour cela, on se basera sur les données colocalisées (Météo-France et IMERG). Cette étape permettra de mieux comprendre les similarités et les différences entre les deux types de produits. Cela sera aussi l'occasion de mettre en place une ou des métriques adaptées à la validation des produits IMERG.

Finalement, on terminera par la projection des produits IMERG sur l'espace latent précédemment obtenu (i.e. appris sur les données Météo-France) ce qui permettra de mettre en place la validation des produits IMERG souhaitée.

Méthodes et outils : Il faudra se familiariser avec l'état de l'art des AE/VAE via une bibliographie (quelques articles fournis en début de stage). Il s'agira de prendre en main un modèle existant et déjà implémenté dans pytorch (Topographic VAE) pour cela on pourra utiliser, dans un premier temps, les jeux de données de l'article (MNIST et dSprites). Dans un second temps, il faudra appliquer l'algorithme au jeu de données d'images Météo-France et mettre en œuvre quelques modifications mineures du modèle pour permettre la validation des produits IMERG. Si le temps le permet, il pourra être demandé de mettre en œuvre d'autres modèles de la littérature pour comparer leurs performances aux résultats déjà obtenus.

Bibliographie :

[1] NASA precipitation Measurement Mission <https://pmm.nasa.gov/gpm>

[2] Sungmin, O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., and Petersen, W. A.: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria, *Hydrol. Earth Syst. Sci.*, 21, 6559-6572, <https://doi.org/10.5194/hess-21-6559-2017>, 2017.

[3] LINDA BOGERD, AART OVEREEM, HIDDE LEIJNSE, AND REMKO UIJLENHOET : A Comprehensive Five-Year Evaluation of IMERG Late Run Precipitation Estimates over the Netherlands, *Journal of Hydrometeorology*, 2021, 22, 1855–1868

[4] Keller, T. Anderson, and Max Welling. "Topographic vaes learn equivariant capsules." *Advances in Neural Information Processing Systems* 34 (2021): 28585-28597.

- Lien vers l'article : <https://arxiv.org/abs/2109.01394>

- Lien vers le code-source du modèle : <https://github.com/akandykeller/TopographicVAE.git>

- Présentation succincte : <https://youtu.be/8QJmO6u0SwM?feature=shared>