

## Offre de Stage IPSL 2020

(soutenu par le programme EUR IPSL--*Climate Graduate School*)

**Titre du sujet de stage :** Développement d'un algorithme de « scoring » par apprentissage profond pour l'évaluation de cartes de précipitations

**Description du sujet:** L'estimation des champs précipitants depuis l'espace à partir de satellites de télédétection est un enjeu primordial tant sur la caractérisation globale des précipitations que sur son évolution dans un contexte de changement climatique. 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 hétérogène de capteurs hyperfréquences [1]. Parmi ces produits, le produit IMERG a l'avantage de fournir des observations des précipitations à l'échelle globale avec une résolution spatio-temporelle intéressante pour des études à l'échelle régionale. La qualité de restitution de ce produit dépend de plusieurs facteurs, certains sont inhérents à la mesure (relation complexe entre températures de brillance issues des radiomètres spatiaux et intensité de pluie par exemple) alors que d'autres sont inhérents à la configuration de la constellation de satellites défilants. En effet le nombre de passage de chaque satellite sur une zone donnée est relativement limité, il n'est donc pas possible que toutes les cellules pluvieuses soient observées à tout instant (mais elles ont par contre pu être observées à des instants antérieurs sur une zone géographique voisine). L'algorithme de restitution IMERG utilise dans ce cas des champs d'advection permettant de calculer la position probable des cellules aux instants où ces dernières n'ont pas été observées. Ce procédé, bien qu'étant indispensable pour obtenir des cartes sans « trous » fournit une estimation de l'intensité de pluie des cellules d'autant plus dégradée que ces dernières ont été peu (ou pas) observées dans un passé proche. Ce type de défaut a par exemple été étudié par Sungmin et al. [2]. Par ailleurs, nous disposons sur la France métropolitaine des observations de précipitations réalisées à partir du réseau de radar météorologique de Météo France (ARAMIS). Ce réseau permet d'observer les précipitations avec une qualité bien supérieure à celle fournie par le produit IMERG, mais en contrepartie, ne permet pas une observation globale à l'échelle de la planète.

La finalité de l'étude est de réaliser un « scoring » du produit pluie IMERG en prenant comme référence le produit fourni par Météo France. Le but du stage est de développer un algorithme à base de réseaux neuronaux profonds permettant d'évaluer ce score. Le développement du modèle s'appuiera sur une base de deux années d'observation d'une zone de 800x800 km<sup>2</sup> centrée sur la France. Une approche à base d'auto-encodeur convolutionnel et de réseaux de type U-NET sera envisagée[2-7].

Dans un premier temps, une petite étude bibliographique sera menée suivie d'une prise en main de la base de données et de l'environnement informatique (Serveur GPU, framework Pytorch). On s'appuiera ensuite sur une analyse descriptive simple de cette base afin que définir le score le plus adéquat. Il s'agira ensuite de développer le modèle neuronal, puis d'analyser les scores obtenus via ce modèle en fonction de divers paramètres (situation météorologiques, présence des satellites, ...). Enfin un clustering des cartes IMERG projetées dans l'espace latent d'un réseau de type encodeur/décodeur suivi d'une étape de labellisation à partir des scores obtenus seront réalisés. Cette dernière étape permettant in fine de fournir un score pour de nouvelles cartes IMERG sans avoir à utiliser les cartes Météo France.

Au-delà du cas particulier de l'évaluation des produits IMERG ce stage se situe dans la thématique de l'extraction automatique de caractéristiques pertinentes pour la classification. Les améliorations apportées par un apprentissage combiné des caractéristiques et du clustering pourront être étudiées (DESOM)[9]

[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] X. Guo, X. Liu, E. Zhu, and J. Yin, Deep Clustering with Convolutional Autoencoders, [www.researchgate.net/publication/320658590\\_Deep\\_Clustering\\_with\\_Convolutional\\_Autoencoders](http://www.researchgate.net/publication/320658590_Deep_Clustering_with_Convolutional_Autoencoders)

[4] L. Zhang, L. Zhang and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," in IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 22-40, June 2016. doi: 10.1109/MGRS.2016.2540798

[5] M. M. Fard, T. Thonet, E. Gaussier, Deep k-Means: Jointly Clustering with k-Means and Learning Representations, 2018, arXiv : 1806.10069v1.

[6] K. Z. Dizaji, A. Herandi, C. Deng, W. Cai, H. Huang Deep Clustering via Joint Convolutional Autoencoder Embedding and Relative Entropy Minimization, 2017, arXiv : 1704.06327.

[7] X. Guo, L. Gao, X. Liu, J. Yin, Embedded Clustering with Local Structure Preservation, IJCAI, 2017

[8] Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas, U-Net: Convolutional Networks for Biomedical Image Segmentation, <https://arxiv.org/pdf/1505.04597.pdf>, 2015

[9] Deep Architectures for Joint Clustering and Visualization with Self-Organizing Maps, Florent Forest, Mustapha Lebbah, Jérôme Lacaille, conference LDRC Workshop PAKDD, 2019

#### **Résumé en anglais (5 lignes) :**

The estimation of precipitating fields from space is a crucial issue both for the global characterization of precipitation and its evolution in a context of climate change. The goal of the Global Precipitation Measurement (GPM) mission is to provide precipitation products from a heterogeneous constellation of microwave sensors. The quality of GPM-based products depends on several factors, some of which are inherent to the measurement while others are inherent in the configuration of the satellite constellation. The aim of the study is to carry out a scoring of the GPM IMERG rain product using the product supplied by Météo France's radar network using deep learning methods.

**Responsable du stage (Nom/prénom/statut) :** Cécile Mallet, MCF (Cecile.Mallet@latmos.ipsl.fr). Co-encadrement : Laurent Barthes, MCF

**Laboratoire concerné :** LATMOS

**Equipe de recherche concernée (si pertinent) :** SPACE

**Niveau du stage (Licence, M1, M2, internship) :** M2

**Licence ou Master(s) où sera proposé le sujet :** M2 TRIED/WAPE

**Thème scientifique de l'IPSL concerné :** Cycle de l'eau

**Durée du stage :** 6 mois

**Période :** 01/03/2020- au 30/09/2020

**Est-il prévu une thèse dans le prolongement du stage ? :** Réflexion en cours