

## Offre de Stage IPSL 2021

(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.](#) 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 : plus le score d'une carte IMERG est élevé, plus celle-ci possède les caractéristiques d'une carte Météo France. Pour cela, il s'agit 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 sera envisagée.

Dans un premier temps, une 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 des deux produits afin d'évaluer d'un point de vue statistique leur caractéristiques communes et leurs différences. Le cœur du stage consistera à développer un modèle neuronal de type auto-encodeur permettant de projeter les cartes Météo France dans un espace de

dimension réduit. Une « fonction de densité » (mixture de gaussienne par exemple) estimée sur les 2 années du dataset Météo France permettra de réaliser un « mappage » de cet espace. Ce mappage sera ensuite utilisé pour évaluer les cartes IMERG. Chaque carte du produit IMERG projetée dans l'espace réduit à l'aide de l'auto encodeur précédemment développé permettra de déduire une probabilité dépendant à la fois de la fréquence de l'événement météorologique observé et de la pertinence de la carte à représenter des cellules de pluie réalistes. Cette probabilité sera ensuite utilisée pour estimer un score. Cette étape permettant in fine de fournir un score pour n'importe quelle carte IMERG au-dessus de la France métropolitaine. Enfin, dans une dernière partie, il s'agira d'analyser les scores obtenus via ce modèle en fonction de divers paramètres (situation météorologiques, heure de passage des satellites, ...).

**Pré-requis** : Machine learning, analyse de données, statistiques

**Résumé en anglais** : Estimating precipitating fields from space using remote sensing satellites is a major challenge both for the overall characterization of precipitation and for its evolution in a context of climate change. The goal of this internship is to assess the relevance of IMERG precipitation maps obtained by remote sensing in relation to those obtained by Météo France's meteorological radar network.

Responsable du stage : Barthès Laurent, MCF UVSQ & Mallet Cécile, PR UVSQ  
Contact laurent.barthes@latmos.ipsl.fr

Laboratoire concerné : LATMOS

Equipe de recherche concernée (si pertinent) : SPACE (Statistiques, Processus, Atmosphère et Cycle de l'Eau)

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

Licence ou Master(s) où sera proposé le sujet : Master TRIED et autres Master en science des données, Ecoles d'ingénieurs

Thème scientifique de l'IPSL concerné : cycle de l'eau

Durée du stage : \_\_6\_\_ mois

Période : 01/03/2021 □ 30/09/2021

Est-il prévu une thèse dans le prolongement du stage ? Un sujet dans un domaine connexe sera proposé par l'équipe.



