

STAGES M2

Description Scientifique

ACRONYME et titre du projet : APETI

Apprentissage Profond non supervisé Et Topologique appliqué aux Images de télédétection micro-ondes

Noms et coordonnées des porteurs :

Pr Cécile Mallet
Université Paris-Saclay - UVSQ Campus
Laboratoire LATMOS
OVSQ Observatoire Versailles Saint Quentin en Yvelines
11 bd d'Alembert
78280 Guyancourt, France

Pr Mustapha Lebbah
Université Paris-Saclay - UVSQ Campus
Laboratoire DAVID
UFR des sciences
45 avenue des Etats-Unis,
78035 Versailles, France

Vibolroth Sambath (doctorante)
Université Paris-Saclay - UVSQ Campus
Laboratoire LATMOS
OVSQ Observatoire Versailles Saint Quentin en Yvelines
11 bd d'Alembert
78280 Guyancourt, France

Laboratoires ou équipes :

Équipe SPACE (Statistiques Processus Atmosphère Cycle de l'Eau) et axe transverse IA du LATMOS
Équipe ADAM (Ambient Data Access and Mining) de DAVID

Description (2 pages) :

Le travail de stage proposé est dans le cadre des travaux de thèse en cours de Sambath Vibolroth au LATMOS il s'agit de quantifier l'évolution des propriétés des systèmes nuageux et de la convection au cours des dernières décennies sous la pression du changement climatique. En effet, les changements dans les conditions thermodynamiques de l'atmosphère doivent se traduire par une modification des propriétés micro- et macro-physiques des

systèmes précipitants. Si de nombreuses études se sont penchées sur l'évolution des propriétés microphysiques des nuages et leur impact sur le bilan radiatif terrestre, peu de travaux portent sur l'évolution des propriétés macro-physiques.

Depuis maintenant 30 ans, les satellites défilants en orbite basse sont équipés de capteurs hyperfréquences passifs qui mesurent les températures de brillances au sommet de l'atmosphère. Ces mesures sont réalisées dans différents canaux entre 10 et 150 GHz (9 canaux seront utilisés), à ces fréquences, les principaux processus contributeurs sont l'émission par les gouttes de pluie, la diffusion par les hydrométéores glacés et l'émission par la vapeur d'eau (principal gaz intervenant dans l'effet de serre). L'originalité du travail repose sur l'utilisation des observations brutes (températures de brillance) qui contiennent de manière implicite l'information pertinente sur le cycle de l'eau atmosphérique, mais combinée avec d'autres variables. Les études réalisées classiquement se fondent sur des produits géophysiques plus ou moins élaborés tels que les produits IMERG ou TMPA (voir par ex. [1]) qui ont l'avantage de fournir une grandeur physique directement interprétable, mais qui d'une part n'exploitent pas forcément toute l'information contenue dans les données brutes et d'autre part contiennent des hypothèses souvent contraignantes. L'approche globale envisagée n'est cependant pas sans poser des problèmes d'interprétation mais elle évite ainsi les erreurs introduites lors de la phase d'inversion.

L'approche globale envisagée consiste à exploiter l'apport des algorithmes de Deep Learning proposés récemment dans le domaine de la segmentation automatique des images en computer vision. Il s'agit de segmenter les observations et après interprétation des différents segments d'étudier leur évolution temporelle en lien avec l'évolution du climat. Les images de télédétection spatiale micro-ondes sont multi-canaux et présentent des structures non linéaires et multi-échelles particulières. Les pixels de différentes classes ne sont généralement pas linéairement séparables, les structures spatiales ont un pouvoir discriminant limité. La plupart des méthodes basées sur des algorithmes de Deep Learning montrent des performances spectaculaires dans le domaine de la computer vision mais les adaptations concluantes aux domaines de la télédétection micro-ondes restent rares. En raison des différences entre champs de températures de Brillance et images naturelles l'application directe ne donne pas de résultats satisfaisants comme l'ont constaté Zhai et al. 2021 [2] dans les applications aux images hyperspectrales.

Les travaux menés jusqu'ici avec une approche de segmentation non-supervisée basée sur celle de Kim et al. [3] ont permis d'obtenir une segmentation qui permet une comparaison qualitative aux paramètres géophysiques comme la précipitation, les nuages, les glaces, etc. Mais il y a une difficulté à attribuer et interpréter de manière quantitative les classes obtenues dans cette segmentation car elles ne sont pas simplement liées avec un seul paramètre géophysique. Un lien quantitatif est nécessaire pour ensuite arriver à étudier l'évolution des systèmes précipitants.

L'objectif du stage consiste à introduire une notion de topologie dans l'apprentissage des réseaux de neurones profond du type de ce qui est fait dans l'apprentissage des cartes

topologique de Kohonen (SOM : Self Organizing Map) pour favoriser l'émergence d'un groupe de clusters contenant les situations précipitantes qui nous intéressent. Des apprentissages semi-supervisés par des observations des radars spatiaux colocalisés pourront être envisagés.

Mustapha Lebbah recruté récemment au laboratoire DAVID a notamment travaillé sur l'apprentissage profond non supervisé (DESOM : Deep Embedded SOM [4]). Le modèle DESOM est un modèle basé sur des réseaux neurones combinant l'apprentissage de représentations via un auto-encodeur (AE) et l'auto-organisation des prototypes via une carte topologique SOM. L'apprentissage de la représentation (l'espace latent) et le clustering topologique est réalisé d'une manière conjointe. Ces approches basées sur l'apprentissage simultané surpassent les approches en deux étapes où la réduction de dimension de l'AE et le clustering sont effectués séparément. La carte SOM induit une forme de régularisation pour améliorer la qualité de la quantification et de la topologie dans l'espace latent. En plus de l'étude de l'état de l'art de l'apprentissage profond non supervisé, l'autre objectif du stage est d'étudier l'adaptation du modèle DESOM à la modélisation des images de télédétection micro-ondes et de manière plus générale comment introduire la conservation de la topologie dans la segmentation non supervisée par les algorithmes de deep-learning.

Les trois candidats interviennent ensemble cette année dans la formation de Master TRIED de l'université Paris-Saclay. Ce sujet de stage permettra d'initier et de développer une collaboration entre le LATMOS qui utilise des approches par apprentissage pour l'analyse et la modélisation des données issues de la télédétection spatiale et le laboratoire DAVID, particulièrement l'équipe ADAM (Ambient Data Access and Mining), dont l'expertise concerne la modélisation des données hétérogènes.

Le sujet est en lien avec l'axe 1 (des données aux connaissances) et avec l'axe 2 (apprentissage et IA) de DataIA puisqu'il s'agit d'extraire des connaissances à partir des données complexes et particulièrement bruitées. Les images du rayonnement naturel observé par des capteurs passifs (radiomètres) en orbite basse, appelé Température de brillance, présentent une géométrie particulière et un bruit de mesure important. L'atmosphère observée est un système complexe qui est caractérisé par une variabilité extrême. Le modèle d'apprentissage profond envisagé combine l'apprentissage de la représentation (l'espace latent) et le clustering topologique. Le domaine d'application envisagé est celui du climat puisqu'il s'agit de quantifier à l'échelle du globe les impacts sur le cycle de l'eau de l'évolution du climat. L'approche développée sera d'un grand intérêt pour l'analyse des données issues de la télédétection spatiale quel que soit l'application envisagée.

Bibliographie

[1] J. Wang, R.F. Adler, G.J. Huffman, and D. Bolvin, 2014: An Updated TRMM Composite Climatology of Tropical Rainfall and Its Validation. *J. Climate*, **27**, 273– 284, <https://doi.org/10.1175/JCLI-D-13-00331.1>

[2] H. Zhai, H. Zhang, P. Li and L. Zhang, "Hyperspectral Image Clustering: Current achievements and future lines," in IEEE Geoscience and Remote Sensing Magazine, vol. 9, no. 4, pp. 35-67, Dec. 2021, doi: 10.1109/MGRS.2020.3032575.

[3] Unsupervised learning of image segmentation based on differentiable feature clustering. W Kim, A Kanezaki, M Tanaka - IEEE Transactions on Image Processing, 2020

[4] Deep Embedded Self-Organizing Map for Joint Representation Learning and Topology-Preserving Clustering. Forest, Florent; Lebbah, M.; Azzag, H.; and Lacaille, J. Neural Computing and Applications. 2021. <https://github.com/FlorentF9/DESOM>

STAGES M2 Annexe financière

Récapitulatif

Titre du Projet	Noms des Porteurs	Montant du Financement demandé
Apprentissage Profond non supervisé Et Topologique appliqué aux Images de télédétection micro-ondes	Cécile Mallet Mustapha Lebbah Vibolroth Sambath	3900 euros (6mois au niveau de gratification habituel)

Durée et dates envisagées du projet :

6 mois de Avril à Septembre

Etablissement en charge de la gestion des fonds :

Laboratoire LATMOS
11 bd d'Alembert
78280 Guyancourt, France

Noms et Contacts des gestionnaires administratifs et financiers du projet :

A préciser