

## PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DOTA-2020-01**

(à rappeler dans toute correspondance)

Lieu : Palaiseau

Département/Dir./Serv. : DOTA/MPSO

Tél. : 01 80 38 63 76

Responsable(s) du stage : S. Lefebvre

Email. : sidonie.lefebvre@onera.fr

### DESCRIPTION DU STAGE

Thématique(s) : analyse de données, deep learning, modélisation, nuages ,aérosols

Type de stage :  Fin d'études bac+5     Master 2     Bac+2 à bac+4     Autres

#### **Intitulé : Deep clustering pour l'évaluation de la transmission optique satellite-sol**

Sujet : L'établissement de liaisons optiques entre un satellite et le sol nécessite en premier lieu de connaître les occurrences des nuages sur le globe. La plupart des études se basent sur les statistiques de couverture nuageuse afin de déterminer les sites ou combinaison de sites de réception les plus favorables. Elles se limitent en général aux statistiques de transmission optique en ciel clair. Des études récentes ont orienté leurs travaux vers l'obtention de statistiques de transmission en présence de nuages optiquement fins, qui n'atténuent pas complètement le lien optique et permettent donc son établissement, même dégradé. La présence d'aérosols troposphériques est également un facteur d'atténuation qui est aujourd'hui mal documenté. Il apparaît donc important d'évaluer l'impact conjoint des aérosols et des nuages optiquement fins sur la transmission optique, en particulier en configuration LEO (Low Earth Orbit).

Il est possible, à l'aide de données lidar satellitaires, comme CALIOP qui est embarqué sur le satellite CALIPSO, d'accéder à des bases de données de grande taille de propriétés optiques des nuages et des aérosols, à l'échelle mondiale. Il faut ensuite exploiter ces bases de données pour construire des climatologies qui permettent de calculer des statistiques de transmission pour une liaison satellite-sol. Les travaux que nous avons menés ces 10 dernières années ont abouti à la mise au point d'un générateur de scènes basé sur une approche Monte Carlo, qui exploite l'ensemble des profils de la base, ici de taille réduite et provenant uniquement de CALIOP. Afin d'améliorer la représentativité de ce générateur, nous souhaitons pouvoir exploiter des bases de données de plus grande taille, issues de plusieurs satellites. Il ne sera alors plus possible de considérer l'ensemble des profils pour l'approche Monte Carlo.

Le but du stage est donc d'adapter des méthodes de clustering basées sur du deep learning aux profils verticaux atmosphériques associés aux aérosols et aux nuages, dans le but d'extraire des profils représentatifs. Ces profils représentatifs, pondérés par la probabilité d'occurrence du cluster associé, seront ensuite utilisés pour calculer les statistiques de transmission pour une liaison satellite-sol.

On constate en effet un essor important dans la littérature des méthodes de deep clustering depuis 2016, avec des performances bien supérieures à celles des méthodes classiques comme les k-means. L'idée est de combiner et d'optimiser conjointement un réseau de neurones profond pour extraire des caractéristiques non linéaires bien adaptées au clustering et une méthode de clustering classique comme les k-means [Guo2017][Fard2018] ou la régression logistique [Dizaji2017], appliquée dans l'espace des caractéristiques.

Cependant, la plupart des cas tests considérés pour évaluer les performances des méthodes de deep clustering sont des bases d'images comme MNIST ou de textes. Le ou la stagiaire devra donc sélectionner les 2 ou 3 approches les mieux adaptées à notre problématique, les mettre en œuvre et les transposer aux données de type profils atmosphériques verticaux, à partir des codes python fournis avec les publications.

La pertinence du clustering sera évaluée par comparaison avec des méthodes de clustering plus classiques (k-means, classification hiérarchique, propagation d'affinité) et en comparant les statistiques de transmission évaluées sur les profils représentatifs avec celles calculées sur l'ensemble de la base réduite que nous utilisons actuellement.

Les travaux pourront se poursuivre dans le cadre d'un sujet de thèse en collaboration entre le LATMOS

IPSL et l'ONERA.

### Bibliographie

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

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.

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

Est-il possible d'envisager un travail en binôme ? **Non**

### Méthodes à mettre en oeuvre :

- |   |  |
|---|--|
| <input checked="" type="checkbox"/> Recherche théorique | <input type="checkbox"/> Travail de synthèse                 |
| <input checked="" type="checkbox"/> Recherche appliquée | <input checked="" type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale        | <input type="checkbox"/> Participation à une réalisation     |

Possibilité de prolongation en thèse : **Oui**

**Durée du stage :** Minimum : 3 mois Maximum : 5 mois

Période souhaitée : 1<sup>er</sup> semestre 2020

### PROFIL DU STAGIAIRE

Connaissances et niveau requis :

Machine learning et deep learning, analyse de données, statistiques, python, R ou matlab

Des connaissances en transfert radiatif ou télédétection seraient un plus

Ecoles ou établissements souhaités :

Master 2 ou Ecole d'Ingénieur