

M2 / Ecole : Fiche de stage Année 19_20

Titre du stage : Classification automatique des zones nuageuses appliquée aux prévisions numériques et aux observations satellites.

Nom et statut du (des) responsable (s) de stage :

Laure Raynaud (DR/CNRM/GMAP/RECYF)
Lucie Rottner-Peyrat (DR/CNRM/GMAP/RECYF)
Matthieu Sorel (DSM/CS/DC)

Coordonnées (téléphone et email) du (des) responsable (s) de stage :

Laure Raynaud – laure.raynaud@meteo.fr – 05 61 07 96 48
Lucie Rottner-Peyrat – lucie.rottner@meteo.fr – 05 61 07 98 01
Matthieu Sorel – matthieu.sorel@meteo.fr – 05 61 07 83 59

Sujet du stage :

Les opérateurs satellites (CNES, AIRBUS Defence and Space, etc.) ont besoin d'avoir la connaissance la plus fine du couvert nuageux afin d'optimiser la prise de vue d'images de la terre depuis des satellites à orbite basse. Une classification des zones nuageuses (dissocier une perturbation, d'un ciel de traîne, d'une zone de ciel clair) constituerait une nouvelle donnée d'entrée dans l'algorithme de priorisation des scènes d'acquisition du satellite.

Il s'agit donc d'identifier les objets suivants dans les données météo :

- zones de ciel couvert ;
- zones intermédiaires (structures nuageuses entourées de ciel clair) ;
- zones de ciel clair ;

Des travaux récents ont permis de mettre en œuvre un algorithme de classification de l'organisation spatiale des précipitations (Hamidi et al., 2019). Chaque point de grille est affecté à l'une des trois catégories suivantes : pas de précipitations, précipitations continues, précipitations intermittentes. Les structures spatiales des champs de nébulosité et de précipitations présentant des ressemblances assez fortes, on propose dans ce travail d'adapter l'algorithme existant pour les pluies au cas de la nébulosité.

Cet algorithme utilise la méthode des forêts aléatoires (Breiman et al., 2001), qui permettent d'obtenir des modèles prédictifs de classification. Il s'agira en premier lieu de construire les bases d'apprentissage et de validation permettant d'entraîner et d'évaluer le modèle. Plusieurs tests seront ensuite effectués afin de déterminer les prédicteurs les plus informatifs et d'optimiser les réglages du classifieur. Des évaluations objectives et subjectives de la méthode seront réalisées, sur des domaines géographiques différents, afin de s'assurer de la pertinence des résultats obtenus.

Le stagiaire travaillera d'abord à partir des données de la prévision numérique (champs de nébulosité totale et/ou image satellite prévue).

L'algorithme développé pourra être testé sur quelques images satellites observées (canal visible et/ou infrarouge et/ou masque nuageux) pour en estimer les performances.

Si le temps le permet, le stagiaire pourra dissocier la zone de ciel de traîne en deux sous-zones :

- zones nuageuses très denses (nuages entourées de quelques zones de ciel clair) ;
- zones nuageuses peu denses (ciel clair ponctué de quelques nuages) ;

Compétence requises :

- Excellentes compétences en Python (manipulation de données scientifiques)
- Excellentes connaissances des méthodes de machine learning (méthodes d'apprentissage statistiques).
- Connaissance en deep learning appréciées.

Références

Leo Breiman, « Random Forests », *Machine Learning*, 45, 5-32.

Yamina Hamidi, La détection et la caractérisation automatiques de la texture des précipitations, Rencontres R&D. http://www.umr-cnrm.fr/IA_180619/presentations/11-Pres-Hamidi-180619.pdf