

Sujet Stage M2

Reconstruction de la biomasse phytoplanctonique dans l'océan global : incertitudes liées aux phases d'apprentissages des méthodes de deep learning

Encadrants : Elodie Martinez, Thomas Gorgues et J. Roussillon (LOPS), R. Fablet et L. Drumetz (IMT-Atlantique)

Contact : elodie.martinez@ird.fr

Lieu : LOPS/IUEM (Brest), 1^{er} semestre 2021

Sujet :

Le phytoplancton (micro algue peuplant la couche supérieure éclairée de l'océan) alimente la chaîne alimentaire océanique et régule les niveaux de dioxyde de carbone océanique et atmosphérique grâce à la fixation photosynthétique du carbone. Tandis que les cycles saisonniers et interannuels de la biomasse phytoplanctonique sont maintenant relativement bien caractérisés, de larges incertitudes persistent quant à sa variabilité décennale, notamment dû aux manques d'observations sur ces échelles de temps. L'essor des méthodes d'apprentissage profond (deep learning) est donc apparu comme une opportunité pour palier à ce problème d'observations afin de permettre d'améliorer la compréhension de la dynamique des écosystèmes phytoplanctoniques des dernières décennies.

Tandis que les approches par apprentissage profond sont maintenant couramment utilisées pour reconstruire des séries temporelles de variables physiques, la reconstruction de la variabilité spatio-temporelle de la biomasse phytoplanctonique n'en est qu'à ses prémices. Celle-ci étant fortement contrôlée par des processus de la dynamique océanique, de récents travaux au LOPS, et en collaboration avec l'IMT Atlantique, ont permis d'initier cette approche en reconstruisant la concentration en chlorophylle (Chl, proxy de la biomasse phytoplanctonique) issue des observations satellites uniquement à partir de variables physiques (température de la mer, vitesse des courants et des vents, etc...) et d'approches de deep learning (Martinez et al., 2020; in rev). Dans ces approches une première phase d'entraînement sélectionne de manière aléatoire en temps et en espace un échantillon de données utilisées pour l'apprentissage. Dans une deuxième phase, la variable d'intérêt (ici la Chl) peut être reconstruite sur la totalité des données des prédicteurs physiques disponibles (ie sur toute la zone et la période d'étude).

Le stagiaire utilisera 1 ou 2 méthodes opérationnelles (du type Multi-Layer Perceptron ou Convolutional Neural Network) déjà éprouvées dans le cadre de la thèse de J. Roussillon (LOPS/IMT-A). L'objectif du stage sera d'explorer l'impact des stratégies d'échantillonnage (cf première phase) sur la reconstruction de la Chl. Par exemple, au lieu de sélectionner les données d'apprentissage de façon aléatoire, des choix spatiaux ou/et temporels seront imposés et leurs impacts sur la Chl évalués. Ces résultats permettront également d'évaluer l'importance relative de processus physiques et biogéochimiques impliqués dans différents modes de variabilité spatio-temporelle du phytoplancton à l'échelle de la planète.

Le stage se déroulera au LOPS, sur le site de l'IUEM, à Brest, et bénéficiera d'un environnement scientifique stimulant avec des interactions avec l'Institut Mines Télécom-Atlantique.

Outils de développements : python, git, keras, PyTorch

Prérequis : programmation, notion de machine learning, aucune connaissance en écologie marine ou océanographie n'est requise pour ce stage mais serait un plus

Références :

- Martinez, E., T. Gorgues, , and R. Fablet (2020). Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach. *Front. Mar. Sci.*, 7:464. doi: 10.3389/fmars.2020.00464.
- Martinez, E., A. Brini, T. Gorgues, , and R. Fablet (**in revision**). Neural network approaches to reconstruct phytoplankton time-series in the global ocean. *Remote Sensing*