

## Proposition de stage de Master 2 (6 mois)



Laboratoire Environnement Ressources  
Centre Ifremer Manche-Mer du Nord  
Boulogne-sur-Mer



Laboratoire d'Informatique Signal et Image de la  
Côte d'Opale  
Université du Littoral Côte d'Opale  
Calais



### Encadrants principaux :

Dr. Guillaume WACQUET (Ifremer, ODE/COAST/LER-BL, [Guillaume.Wacquet@ifremer.fr](mailto:Guillaume.Wacquet@ifremer.fr))

Pr. Émilie POISSON-CAILLAULT (ULCO, LISIC, [Emilie.Poisson@univ-littoral.fr](mailto:Emilie.Poisson@univ-littoral.fr))

### Laboratoire d'accueil :

Laboratoire Environnement Ressources (LER-BL)

IFREMER, Centre Manche-Mer du Nord

150 Quai Gambetta, 62321 Boulogne-sur-Mer

## Développement d'une architecture de fusion multimodale pour la prédiction des blooms de phytoplancton à partir de mesures *in situ*, d'images satellites et de sorties de modèles numériques

### Problématique et objectifs scientifiques

Les blooms de phytoplancton, et plus particulièrement les efflorescences toxiques/nuisibles (Harmful Algal Blooms, HAB), représentent aujourd'hui un enjeu majeur pour la gestion durable des écosystèmes marins et la sécurité des activités côtières (aquaculture, santé publique). Ces épisodes de prolifération algale peuvent provoquer des perturbations profondes des réseaux trophiques marins, altérer la qualité de l'eau et engendrer des mortalités massives d'organismes aquatiques. Ils posent également des risques sanitaires pour la consommation de produits de la mer (conchyliculture, pêche), menacent la santé publique par l'exposition directe aux toxines, et génèrent des impacts économiques considérables sur les filières aquacoles, touristiques et maritimes.

La compréhension et la prévision de ces phénomènes sont particulièrement complexes car elles reposent sur une combinaison de processus multi-échelles, nécessitant l'intégration de sources de données hétérogènes, telles que les mesures *in situ*, les observations satellitaires et les sorties de modèles numériques. Cependant, ces données peuvent potentiellement présenter des résolutions temporelles et spatiales différentes, être partielles (données manquantes) ou encore susciter des incertitudes importantes concernant leur qualité et leur fiabilité.

Dans ce contexte, la fusion multimodale des données apparaît comme une stratégie particulièrement prometteuse. En combinant de manière cohérente l'information issue de différentes modalités, elle permet de tirer parti des forces de chaque source tout en compensant leurs limites respectives. Une telle approche favorise la construction de modèles plus précis et plus robustes, capables de capturer la dynamique complexe des blooms de phytoplancton et de réduire les incertitudes associées aux prédictions.

## **Méthodologie**

L'objectif du stage est de concevoir et développer une architecture modulaire (R ou Python) dédiée à l'intégration et à la fusion de données hétérogènes issues du milieu marin. Cette architecture permettra de charger, prétraiter et synchroniser les différentes sources de données à disposition, puis d'y intégrer les modèles prédictifs envisagées dans une logique de fusion multimodale. Trois grandes familles de stratégies de fusion seront explorées :

- **Fusion amont** (*Early Fusion*), consistant à concaténer les différentes modalités de données en entrée du modèle d'apprentissage, afin de capturer les corrélations les plus brutes entre sources.
- **Fusion embarquée** (*Intermediate Fusion*), où des représentations spécifiques sont extraites de chaque modalité, puis combinées dans un espace latent commun permettant une intégration plus flexible et hiérarchisée de l'information.
- **Fusion décisionnelle** (*Late Fusion*), qui repose sur l'agrégation des prédictions issues de modèles distincts, afin de comparer et combiner leurs apports respectifs.

Une attention particulière sera portée à l'intégration d'une couche de cross-attention, permettant d'ajuster de manière adaptative le poids accordé à chaque source d'informations, en fonction du contexte. Cette étape est essentielle pour gérer les cas où une source de données pourrait présenter des lacunes (ex : couverture nuageuse affectant les images satellites, capteur *in situ* défaillant, etc.).

L'évaluation de ces approches sera menée à l'aide d'indicateurs quantitatifs adaptés (ex : erreur de prédiction sur les concentrations de phytoplancton, scores de précision/rappel sur la détection de blooms, métriques de robustesse, ...). Les performances obtenues feront l'objet d'une analyse critique approfondie, permettant non seulement de comparer les différentes stratégies de fusion, mais aussi de discuter leur capacité de généralisation à d'autres contextes marins et leur robustesse vis-à-vis de la variabilité naturelle des écosystèmes. Enfin, des pistes d'amélioration et perspectives seront dégagées, notamment en termes d'optimisation des architectures et de transposabilité à d'autres contextes ou applications marines.

## **Compétences requises et qualités personnelles**

- Niveau Master 2 en informatique/data science
- Bonne maîtrise de R et/ou Python pour le traitement de données (*machine/deep learning*)
- Intérêt pour l'écologie marine pour la compréhension du sujet, de la littérature associée, ainsi que pour l'exploitation des résultats
- Rigueur et méthode, sens de l'organisation, adaptabilité,
- Bonne compréhension de l'anglais
- Capacité à travailler en équipe multidisciplinaire.

**Candidatures** (CV + lettre de motivation) à transmettre pour le **30/11/2025 au plus tard**  
à [Guillaume.Wacquet@ifremer.fr](mailto:Guillaume.Wacquet@ifremer.fr)