

# Environnement et IA

## MAIA workshop, 23 mai 2024



MAIA  
Mastering Artificial Intelligence Applications

# Les présents

Anton Sokolov : Physique Chimie LPCA/ULCO : dynamique atmosphérique, groupe vert

Felipe Artigas, LOG/ULCO : ecologie microbienne, dynamique phytoplanctonique

Luncida Lanoy : M2 sciences de la mer, ULCO-LOG, classification supervisée de cyto en flux.

Perrine Meynard, LPCA/ULCO : doct. turbulence dans la couche superficielle, classification supervisée, features - SVM, ANN, kNN, QDA

Emilie Poisson Caillault, LISIC, ULCO : apprentissage incremental, frugal, clustering

Arnaud Liefhooghe, LISIC, ULCO : optimisation multiobjectif, classification.

Sébastien Verel, LISIC, ULCO : optimisation

**PhD en cours** : “Modèles généraux d’apprentissage profond pour l’émulation d’images hyperspectrales. Application à la télédétection”, Chedly Ben Azizi, dir. Claire Guilloteau, Gilles Roussel

# Quelques exemples en cours

Anton Sokolov, Périne Meynard

- Travaux chimie avec collègues :  
besoin de IA pour modéliser des propriétés d'arrangement d'atomes,  
générer des configurations plus grandes.
- Travaux atmosphere :  
pb dynamique atmosphérique, changement climatique, pb optimisation, comment  
détecter des événements dans les scans LIDAR ?  
*détection panache industrie à partir de scan lidar (7 km)*
- These Périne :  
Analyse de textures du champs de vent.  
Pas assez de données, beaucoup de features (700).  
399 scans extraits sur 40 000.  
3 catégories de structures (bien droites, floues, aucune visible).
  - saisonnalité 21-22, 4 TP Haralick

# Quelques travaux en cours

Lanoy Lucinda :

Random Forest (RF) for classification of flow cytometer, classes très déséquilibrées

Rclustool. cytométrique -> json.

Peltic 7 millions de données, sampling 500 000 particles.

30 000 où chaque groupe doit contenir au moins 10% de particules par groupe

Beaucoup de données, mais grandement non bien répartie (parfois peu de données par classe)

-> Utilisation de stratification.

-> Analyse des paramètres : paramétrage du RandomForest,

-> Analyse de l'importance des prédicteurs/features

# Quelques défis à travailler

## **Fiabilité des données :**

- Etiqueteurs/experts fiables (diversité des avis, consensus à trouver)
- Ajout des conditions d'acquisition (capteurs) ?

## **Peu de données, ou classes déséquilibrées**

- Frugalité de l'entraînement, transformation de l'ens.
- Echantillonnage adaptatif, etc.

## **Sélection de variables**

- But de l'explication physique du phénomène, peut-on aller plus loin (équations, etc.)
- Importance des variables, confiance de la prédiction

## **Sélection de (méta)-paramètres, sélection de méthode d'apprentissage**

- Travail d'automatisation par non-expert, augmenter la confiance sur le résultat

-> Lancer un premier workshop pour réunir les collègues (UPJV, Artois, ULCO)