

## Filière Métiers de la Recherche

### Predicting species distribution from environmental data: the CNN-SDM approach

Jérémy Fix

jeremy.fix@loria.fr

#### - Sujet en lien avec atelier des transitions : OUI / NON

Les activités ont un impact sur l'environnement. Depuis la révolution industrielle, l'augmentation de la pollution, la déforestation, l'utilisation de produits phytosanitaires, cet impact est négatif sur la biodiversité. Pour pouvoir accompagner des politiques de conservation, un enjeu est d'être en mesure de prédire l'état des populations d'espèces à partir de données mesurables à distance, sans avoir à se rendre sur les lieux pour réaliser des observations. C'est dans ce contexte que s'inscrit cette étude en cherchant à construire des modèles prédictifs de présence d'espèces à partir de données satellites et de modèles biophysiques.

#### - Contexte de l'étude

Le travail est réalisé dans le cadre de l'ANR SmartBiodiv, projet financé dans le cadre du challenge ANR IA-Biodiv (<https://www.challengeia-biodiv.fr/>). Dans cet appel à projets, 3 consortium combinants chercheurs en intelligence artificielle et chercheurs en écologie, travaillent sur les apports de l'IA pour l'étude de la biodiversité. L'un de ces sujets porte sur la capacité à prédire la présence ou l'abondance d'espèces à partir de données environnementales. Les modèles ainsi développés pourraient être appliqués pour estimer la présence d'espèce au-delà des mesures d'observations in-situ plus difficiles à mettre en œuvre et d'autre part conduire à définir des politiques de conservation pour rétablir la biodiversité. L'estimation de la présence ou de l'abondance d'espèces repose sur la notion de « Species Distribution Models » (SDM) pour lesquels plusieurs approches existent. La resurgence récente du deep learning a conduit des chercheurs à proposer des approches reposant sur des réseaux de neurones pour adresser ce problème [Morand et al.(2023)].



#### - Objectif

On se propose d'étudier la capacité qu'ont les modèles de deep learning à répondre à la problématique de la prédiction de présence d'espèces en testant d'une part une approche appliquée sur la prédiction de présence d'oiseaux et de poisson et en transférant la méthodologie sur la prédiction d'abondance de plankton.

#### - Déroulement

Le travail débutera par la prise en main du code et la reproduction des résultats des auteurs. Puis, dans un second temps, la méthodologie sera transférée sur la prédiction de plankton. Pour l'application au plankton, on pourra utiliser les données REPHY mais qui sont peu nombreuses ou les données simulées par CMEMS par exemple utilisée dans le challenge <https://www.kaggle.com/competitions/temporal-prediction-of-plankton>.

#### - Outils et méthodes

Pour réaliser cette étude, les données environnementales sont fournies par l'initiative européenne Copernicus (<https://data.marine.copernicus.eu/products>), via le wrapper python geoenrich (<https://github.com/morand-g/geoenrich>). La mise en œuvre des modèles de deep learning pourra être facilitée par l'utilisation de malpolon

(<https://github.com/plantnet/malpolon>). En particulier, le code de [Morand et al.(2023)] est disponible sur Zenodo [Morand et al.(2023b)]. Les modèles sont entraînés en Pytorch sur GPU. Une expérience préalable, au minimum en deep learning est nécessaire.

- Livrables attendus
  - o Poster en anglais
  - o Article de vulgarisation
  - o Codes de l'expérience
- Références bibliographiques

Gaétan Morand, Alexis Joly, Tristan Rouyer, Titouan Lorieul, Julien Barde (2023). Predicting species distributions in the open oceans with convolutional neural networks. *bioRxiv* 2023.08.11.551418; doi: <https://doi.org/10.1101/2023.08.11.551418>

Morand, Gaétan. (2023b). Deep-SDMs in the open oceans - CODE (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.8222155>