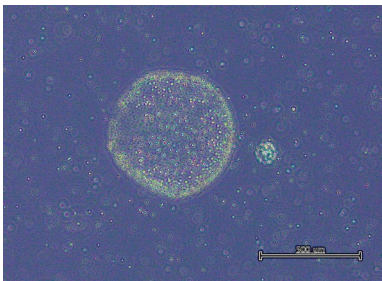


Apprentissage des interactions entre espèces de phytoplancton et intégration des connaissances préalables

Nous nous intéressons aux modèles d'écologie marine ayant pour but d'étudier et d'anticiper les variations de populations de phytoplancton¹ au cours des saisons. Ces variations ont un fort impact sur les services écologiques rendus par les océans, comme la régulation du climat, la qualité des eaux et la production de ressources vivantes, issues de la pêche notamment. La compréhension et la prédiction de ces systèmes revêt donc un intérêt particulier.



Colonies formées par l'algue Phaeocystis.

([Wikipédia](#))

Efflorescence (surpopulation) de phytoplancton vue du ciel.

([NASA](#))



Pour cela, une approche répandue est l'utilisation de réseaux de neurones afin de construire des modèles prédictifs². Cette approche est néanmoins souvent qualifiée de « boîte noire », car les mécanismes internes du modèle sont très difficiles à comprendre, à l'opposé des approches dites « explicables³ » qui fournissent une explication des prédictions produites.

Ce projet rentre dans le cadre d'une collaboration entre le laboratoire d'informatique et d'automatique (CRISAL) à Villeneuve-d'Ascq et le Laboratoire d'Océanographie et de Géosciences à Wimereux.

Objectifs du projet

L'approche utilisée ici [1] est dite symbolique, plutôt que statistique : elle consiste à apprendre un ensemble de règles logiques qui constitue un modèle prédictif explicable. Un travail préalable avait permis d'appliquer cette méthode au jeu de données de Suivi Régional des Nutriments [2] qui

- 1 Le *phytoplancton* est l'ensemble des espèces, généralement unicellulaires, vivant suspension dans l'eau et se nourrissant uniquement grâce à la photosynthèse.
- 2 Un modèle est dit *prédictif* s'il permet de prédire l'état futur du système avec assez peu d'erreur. Dernièrement l'apprentissage par réseaux de neurones a donné de bons résultats en matière de prédictibilité car ces modèles sont bien adaptés pour la reconnaissance de motifs.
- 3 Un modèle est dit *explicable* s'il peut produire une explication pour chaque résultat qu'il retourne (décision, prédiction, etc.). L'explicabilité est fondamentale dans certains domaines comme la médecine, où il faut pouvoir expliquer un diagnostic ou une prise de décision. Les modèles de réseaux de neurones sont dits « boîtes noires » car ils sont basés sur un très grand nombre d'opérations mathématiques qui ne sont pas interprétables par un être humain.

contient des mesures de paramètres environnementaux (salinité de l'eau, ensoleillement, taux d'azote...) et des comptages de populations de phytoplancton.

L'objectif de ce stage est de repartir sur ces bases et d'améliorer la qualité des résultats en plusieurs étapes :

- Compréhension de la méthode d'apprentissage LFIT et prise en main de l'existant [3, 4].
- Intégration dans le pipeline de connaissances biologiques préalables sur les espèces de phytoplancton étudiées, telles que la température idéale de reproduction ou les nutriments consommés.
- Amélioration de l'inférence du graphe des interactions obtenu à partir du programme logique appris, pour faciliter la lecture des résultats.
- Dialogue avec le partenaire écologue concernant les hypothèses faites, les traitements, le modèle appris et les prédictions effectuées.

Si le temps de stage le permet, il sera aussi possible d'étendre davantage le travail :

- Intégration de nouvelles espèces de phytoplancton depuis les données.
- Comparaison des deux approches d'apprentissage de LFIT (complète et gloutonne) en cherchant notamment à optimiser le résultat de l'approche gloutonne pour le rapprocher de l'approche complète.

Prérequis

- Maîtrise du langage Python ou d'un langage de script similaire
- Appétence pour les mathématiques et intérêt pour l'apprentissage machine
- Aucune connaissance préalable en biologie ou en écologie n'est requise

Ce sujet est destiné au niveau master 2, et pourra faire l'objet d'une poursuite en doctorat.

Contacts

Maxime FOLSCHETTE : Maxime.Folschette@centralelille.fr — Bâtiment ESPRIT, bureau S3.45

Cédric LHOSSAINE : Cedric.Lhoussaine@univ-lille.fr — Bâtiment ESPRIT, bureau S3.35

Les réunions de projet se dérouleront au bâtiment ESPRIT (Cité Scientifique de Villeneuve-d'Ascq)

Références

- [1] Tony Ribeiro, Maxime Folschette, Morgan Magnin et Katsumi Inoue. [Learning any memory-less discrete semantics for dynamical systems represented by logic programs](#). *Machine Learning*, Springer, 2021. DOI [10.1007/s10994-021-06105-4](https://doi.org/10.1007/s10994-021-06105-4) — Bibliothèque pyLFIT : <https://github.com/Tony-sama/pylfit>
- [2] SRN - Regional Observation and Monitoring program for Phytoplankton and Hydrology in the eastern English Channel. SRN dataset. SEANOE, 2022. DOI [10.17882/50832](https://doi.org/10.17882/50832)
- [3] Omar Ikne. Automatic Construction of Biological Interaction Models within Marine Ecosystems using LFIT framework. Master 1 internship report, Université de Lille, 2021.
- [4] Omar Ikne, Maxime Folschette, Tony Ribeiro. [Automatic Modeling of Dynamical Interactions Within Marine Ecosystems](#). 1st International Joint Conference on Learning & Reasoning (IJCLR'2021), 2021.