Physics Informed Machine Learning for Lake Pollution Forecast

David Métivier
david.metivier@inrae.fr
Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie

Goal

Build a fancy Machine Learning model that is aware of the laws of physics, for a real life environmental application with real data.

Context

More than half of the freshwater lakes and rivers of the world are polluted. The World Health Organization declared microbial hazards, such as toxic cyanobacteria, to be "of public health importance". Cyanobacteria are photosynthetic bacteria playing a key role in the life cycle. However, their proliferation is a significant environmental problem with important economic, ecological and health consequences. Under "favorable" conditions, the cyanobacteria population can grow very quickly and accumulate on the water surface, forming scums, this phenomenon known as algal bloom (see picture) will likely be amplified by climate change.
In this context, it is important to develop some operational tools for the management of lake ecosystems. The objective is to propose a short-term prediction tool that can be used as a warning system. The forecasting tool will enable better anticipation of algal blooms, alerting lake managers and users.

Advisers

David Métivier is permanent INRAE junior researcher (*chargé de recherche*). He first studied theoretical physics and got his Ph.D. working in Dynamical Systems and Statistical Physics. Then, he progressively shifted his research focus to environmental issues using Statistics modeling.
This internship is the sequel of two projects focusing exclusively either on dynamical systems approaches or Machine Learning models.

Collaboration

- Céline Casenave (INRAE - MISTEA) is a researcher expert on lake modelling Dynamical Systems and control.
- Brigitte Vinçon-Leite (École des Ponts ParisTech - LEESU) is a researcher expert on the lake and cyanobacteria modelling.

Julia Language

The Julia programming language plays a crucial role in our project. Julia offers a powerful modern ecosystem for scientific computing, providing high-performance numerical libraries and a user-friendly syntax. It is said to "Walk Like Python; Run Like C". With its ability to express mathematical equations concisely, Julia enables us to implement the physics-based models and seamlessly integrate them into the machine learning framework.

Some advantages of Julia:

- Syntax.
- High-performance computing capabilities
- Easy integration/combination with existing code and libraries
- Built-in support for parallel and distributed computing

Physics Informed Machine Learning (PIML)

We propose to use PIML to integrate the governing equations of lake pollution dynamics, including cyanobacterial growth, into the learning process. By incorporating the underlying physical laws, we can guide the machine learning models to make predictions consistent with the known physics, improving their accuracy and interpretability.

Required skills

- Interest in Machine Learning
- A taste for modeling and numerical simulations.
- Concern for environmental issues.

The first semester will be a good occasion to discover Physics Informed Machine Learning, lake dynamics and Julia.