Internship offer

Physics Informed Machine Learning for lake pollution forecast

Adviser: David Métivier is permanent INRAE junior researcher (*chargé de recherche*).

Location: UMR MISTEA (Mathematics, Informatics, and Statistics for Environment and Agronomy), 2 Pl. Pierre Viala, 34000 Montpellier, France

Teasing: Build a fancy Machine Learning model that is aware of the laws of physics, a.k.a. Physics Informed Machine Learning (PIML), for real life environmental application.

Contact: Please send your application with a CV and a few motivation lines to: david.metivier@inrae.fr.

Context

More than half of the freshwater lakes and rivers of the world are polluted. The World Health Organization (WHO) 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: they are primary producers, some of them can fix the atmospheric nitrogen, others form symbiotic relationships with other organisms e.g., plants, fungi, etc. However, their proliferation can also be harmful for aquatic ecosystems. Indeed, the population of cyanobacteria can grow very quickly and accumulate on the water surface, forming scums (see picture). This phenomenon, known as algal bloom, has important economic, ecological and health consequences. Climate change and rising temperatures will amplify algal blooms [AOZ+09].

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 managers to anticipate these restrictions and alert users. Such a forecasting tool has been developed for the experimental lake of Champs-sur-Marne (Paris area) and can be seen on `https://balneau-leesu-rec.enpc.fr/#!/dashboard` where three variables of interest are predicted three days ahead: water temperature, phytoplankton concentration and cyanobacteria concentration.

Objectives

The objective of this internship will be to improve the quality of the forecast using methods based both on data and physics modelling.

Dynamical models for hydrodynamics, such as rivers or lakes, benefit from a huge literature and well understood physics phenomenon. However, they are often hard to calibrate and slow to run. On the other hand, data based methods are often very fast to train and operate. In the case of deep learning, they can in theory learn arbitrary complex phenomenon given enough data. This is currently not suited for many tasks in fluid mechanics problems where abundance of data is limited. Enforcing physics knowledge into a data based is referred to as Physics Informed Machine Learning. It facilitates the model training, e.g. reducing the number of data required. Moreover, it imposes the resulting model to respect some physics laws.

We would like to build a PIML to predict the future distribution of the variable of interests (water temperature, cyanobacteria concentration, etc) conditionally on the past observations. This tool will allow critical prediction such as, **in three days with high probability, the lake will not be suited for swimming due to dangerous cyanobacteria concentration**.

The model will be tested and trained on the experimental lakes like the in Champs-sur-Marne or Créteil (Paris area) where high-frequency measurements (meteorology, water temperature, water quality) are collected by the LEESU (Water, Environment, and Urban Systems Laboratory, École des Ponts ParisTech) since 2017.

Starting point Many review articles and tutorials are available, they will be the starting point of the internship to familiarize the student with machine learning and PIML before application with real lake data. Recent examples of PIML for climate [KMA+] or hydrological [FTH, BW] applications will also guide the internship.

Required skills

The applicant should be a master student with an interest in Machine Learning and a taste for modeling and numerical simulation. David Métivier is a **Julia** user. Julia is a relatively new programming language built using all the good ideas from older languages (C/C++, Python, Matlab, R etc.). It is fast and easy to learn. It is compatible with the standard machine learning packages like TensorFlow and PyTorch. Julia is particularly well adapted for Physics Informed Machine Learning. We can discuss coding languages together. The first semester will be a good occasion to discover Julia.

Knowledge of ecology, biogeochemistry and/or French is not required.

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.

Terms of the internship

The terms of the internship are defined by the IDIL Master Program in Modeling Biological & Environmental Systems (MoBiEn).

References

[AOZ+09] Rita Adrian, Catherine M. O'Reilly, Horacio Zagarese, Stephen B. Baines, Dag O. Hessen, Wendel Keller, David M. Livingstone, Ruben Sommaruga, Dietmar Straile, Ellen Van Donk, Gesa A. Weyhenmeyer, and Monika Winder. Lakes as sentinels of climate change. *Limnology and oceanography*, 54(6):2283–2297, November 2009.

[BW] Daan Bertels and Patrick Willems. Physics-informed machine learning method for modelling transport of a conservative pollutant in surface water systems. 619:129354.

[FTH] Dongyu Feng, Zeli Tan, and QiZhi He. Physics-Informed Neural Networks of the Saint-Venant Equations for Downscaling a Large-Scale River Model. 59(2):e2022WR033168.

[KMA+] K. Kashinath, M. Mustafa, A. Albert, J-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and null Prabhat. Physics-informed machine learning: Case studies for weather and climate modelling. 379(2194):20200093.