Internship offer

Physics Informed Neural Networks for lake pollution forecast

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

When: Sometime in 2024 for 4 to 6 months.

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 Neural Networks (PINN), for a real life environmental application.

Contact: Please send your application with a CV and a few motivational lines to: david.metivier@inrae.fr. Don't hesitate to ask questions about the internship (but do read everything).

Context

More than half of the freshwater lakes and rivers in 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 that play a key role in the life cycle. 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 scum (see picture). This phenomenon, known as an algal bloom, has important economic, ecological and health consequences. Climate change and rising temperatures will amplify algal blooms [1].

Hence, it is important to develop 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 help to anticipate and alert users. A basic example has been deployed for the experimental lake of Champs-sur-Marne (Paris area) and can be seen on this website where three variables of interest are predicted three days ahead: water temperature, phytoplankton concentration and cyanobacteria concentration.

Scientific objectives

The goal of this internship is to build a forecast model using methods based both on data and physics modeling and to use it on real data. If the model works well, one can imagine having it tested in real time in a real lake.

"Physics models" for the hydrodynamics and biology of rivers or lakes are very detailed and precise, however, they are often hard to calibrate and slow to run. On the other hand, data-based methods, a.k.a. Machine Learning, are often very fast to train and operate. In the case of Deep Learning, they can, in theory, learn arbitrary complex phenomena given enough data. However, this is not suited for many tasks in hydrodynamics and biology, where the abundance of data is limited. Moreover, generating temporal data presents new challenges compared to generating images. Enforcing physics knowledge into a neural network is referred to as Physics Informed

Neural Network. It facilitates the model training and reduces the amount of data required. Most of all, it forces the resulting model to respect some physics laws.

We would like to build a PINN to predict the future distribution of the variables of interest in our lake problem (water temperature, cyanobacteria concentration, etc) conditionally on the past observations. This tool will allow critical predictions 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 Champs-sur-Marne or Créteil (Paris area) where high-frequency measurements (meteorology, water temperature, and 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 PINN before the application with real lake data. Recent examples of PINN for climate [2] or hydrological [3, 4] applications will also guide the internship.

Moreover, David Métivier mentored an internship in summer 2023 on the same topic without the physics informed part, i.e. only the neural network approach. This will help to get started quickly.

Required skills

The applicant should be a master's 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 suited for Physics Informed Neural Networks.

Collaboration

The student will collaborate and interact with Céline Casenave[1], Isabelle Sanchez[2] and Brigitte Vinçon-Leite[3]

Terms of the internship

The internship should last from 4 to 6 months. The student receives a "*gratification*" of around 600€/month. We can help students find housing, but we cannot pay for it, nor pay for the trip to Montpellier. (Sadly)

References

[1] Rita Adrian et al. "Lakes as Sentinels of Climate Change". In: *Limnology and oceanography* 54.6 (Nov. 2009), pp. 2283–2297.

[2] K. Kashinath et al. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194 (Feb. 15, 2021), p. 20200093.

[3] Arka Daw et al. "Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling". In: *Knowledge Guided Machine Learning*. Chapman and Hall/CRC, 2022. ISBN: 978-1-00-314337-6.

[4] Daan Bertels and Patrick Willems. "Physics-Informed Machine Learning Method for Modelling Transport of a Conservative Pollutant in Surface Water Systems". In: *Journal of Hydrology* 619 (Apr. 1, 2023), p. 129354.

[1] Researcher expert on lake modeling, dynamic systems, and control INRAE - MISTEA.
[2] Engineer expert on data analysis INRAE - MISTEA.
[3] Researcher expert on the lake and cyanobacteria modelling École des Ponts ParisTech - LEESU.