Internship offer

Model Informed Deep Learning to forecast water quality in lakes

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

When: Starting February 2025 (flexible) for 4 to 6 months.

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

Contact: Please send your application with a CV and a few motivational lines to: david.metivier@inrae.fr. You can add link to scientific (unrelated) projects you did, if any. Don't hesitate to ask questions about the internship (**do read the whole announcement first**!).

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 as they can very quickly accumulate on the water surface, forming scum. This phenomenon, known as an algal bloom, has important economic, ecological and health consequences. Climate change and rising temperatures will amplify algal blooms. Hence, it is important to develop forecasting tools that can be used as warning systems.

Internship objectives and expected work

This internship aims to develop a predictive model that combines data-driven and physics-based approaches to make reliable forecasts based on real-world data.

Detailed "physical models" describing lake hydrodynamics (e.g., water temperature) and "ecodynamics" (e.g., relationships among oxygen levels, bacterial concentrations, and temperature) can offer accurate and comprehensive modeling but can be challenging to calibrate and combine. In contrast, data-driven methods are very flexible and model agnostic e.g. adding a new variable just require changing the input shape of the Machine Learning model. Deep Learning, in particular, can theoretically capture complex dynamics given sufficient data. However, lake data are often very limited. Additionally, working with time series introduces challenges compared to training independent samples. Incorporating known some piece of models into a neural network can enable more efficient and robust training as well as reducing the data requirements while ensuring that predictions respect "physical laws".

Different approaches have already been considered (during past internships at MISTEA) like standard RNN, NeuralODE, Physics Informed Neural Networks and GAN for time series. There are still many open questions and research direction for example:

- Explore NeuralODE variations like Neural Stochastic Differential Equation (NeuralSDE).
- Improve current models and select the best architecture.
- Study the effect of preprocessing time series before using them deep learning models

The exact research questions will be discussed with the applicant before the internship, taking into account his/her skills.

Required Skills

- Knowledge in Machine Learning, mathematical modeling and statistics.
- Skills and interest in scientific programming. The programming language will be **Julia** (no prior knowledge required, but at least look up what it is). It is very well suited for Physics Informed Deep Learning.
- Interest in environmental and/or climate issues.

Terms of the internship

The intern will work at the MISTEA lab mostly with David Métivier and also will collaborate with experts on lake models Céline Casenave (MISTEA) and Brigitte Vinçon-Leitesuch (LEESU). The student receives a "*gratification*" around 600€/month. The canteen is subsidized. We can help students find housing, but we sadly cannot pay for it, nor pay for the trip to Montpellier.