Accelerating Scientific Discovery of Complex Scientific Applications with Process-Guided Deep Learning: Aquatic Eco-Dynamics in Lakes

Supervisors: Professor Matthew Piggott (Imperial College London), Dr R. Iestyn Woolway (University of Reading), Professor Stephen Maberly (UK CEH)

The success of Machine Learning in many applications where large-scale data is available has led to a growing anticipation of similar accomplishments in scientific disciplines. The use of data science is particularly promising in scientific problems involving processes that are not completely understood. However, a purely data-driven approach to modelling can be problematic, as it can create a complex model that is neither generalizable beyond the data on which it was trained nor is it physically interpretable. This problem becomes worse when there is not enough training data, which is common in many scientific applications. A model that is guided by explainable theories stands a better chance at safeguarding against learning spurious patterns from the data that lead to non-generalizable performance. In this project, the PhD candidate will develop a framework that uses the unique capability of Machine Learning to automatically build models from data, without ignoring the treasure of accumulated scientific knowledge and the existing sophisticated numerical models this is encoded within. Specifically, this project will use a new hybrid modelling approach, where scientific knowledge and data science models are integrated synergistically, to advance scientific discovery in a complex scientific domain. This project will focus on the field of Aquatic Eco-Dynamics, a study of the interrelations between physics, biogeochemistry and biota in aquatic environments. A method known as Process-Guided Deep Learning (PGDL) will be used in this project to predict the response of aquatic systems to local and global pressures, which is a research area currently at a tipping point where PGDL can have a transformative effect.

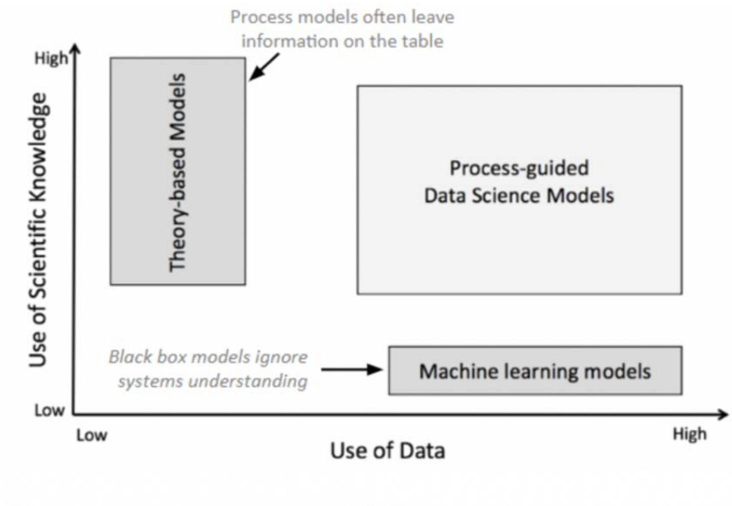

Figure: A new paradigm for scientific discovery that combines theory with data science models

The ultimate goal of this project is to use PGDL to simulate the influence of climatic forcing on lake ecosystems worldwide using a wealth of observational data, including novel satellite observations recently generated within the European Space Agency's Climate Change Initiative Lakes project (https://climate.esa.int/en/projects/lakes/), long-term observations from some of the best monitored lakes in the world, notably from lakes located in the English Lake District (monotired by project partners UK CEH; https://www.ceh.ac.uk/our-science/monitoring-site/lake-observatories), and observations from further afield by collaborating with scientists associated with the Global Lake Ecological Observatory Network (https://gleon.org). Process-based modelling will be conducted using the 1D General Lake Model (https://github.com/AquaticEcoDynamics/GLM) and the 3D unstructured, adaptive mesh model Thetis (https://thetisproject.org/).

Figure: ESA satellite image from the Envisat satellite showing an algal bloom in Lake Erie, North America.

The PhD candidate will be given training in numerical modelling, data science and machine learning, and the utilisation of remote sensing data from lakes, including satellite and drone images, which will provide unprecedented detail of lake processes and will feed into the PGDL framework.

Working closely with the supervisors, the PhD researcher will use the new innovative and sophisticated data and modelling techniques to explore the responses of aquatic ecosystems to climate change.