Combining Earth Observation and Machine Learning for Cyanobacterial Bloom Forecasting

Daniel Atton Beckmann (supervised by Dr Ian Jones, Dr Peter Hunter, Dr Evangelos Spyrakos)
University of Stirling
daniel.atton.beckmann@stir.ac.uk
www.hydronationscholars.scot

Scottish Government gov.scot
Hydro Nation Scholars Programme

UNIVERSITY of STIRLING

Scottish Funding Council
Comhairle Maoineachaidh na h-Alba

Motivation

- Cyanobacteria (blue-green algae) can be toxic and scum-forming, posing a risk to ecosystems and public health
- Continued nutrient enrichment of water bodies from agriculture, industry and urbanisation alongside climate change will likely lead to more frequent and severe cyanobacterial blooms in many areas

Previous Work

- Earth Observation (EO) data has been demonstrated to be effective at monitoring chlorophyll-a (chl-a) concentrations as a proxy for phytoplankton
- This has potential to allow for widespread monitoring of algal blooms but can only inform bloom management in a retrospective way

The Aim: Cyanobacterial Bloom Forecasting

- Forecasting cyanobacterial blooms is highly desirable to provide pre-warning to society and enable management processes to be activated in advance
- Previous forecasting attempts have typically focused on lakes and reservoirs that are regularly monitored, but there is a need to understand monitoring and modelling requirements in more detail so that useful forecasts can be implemented for many more lakes
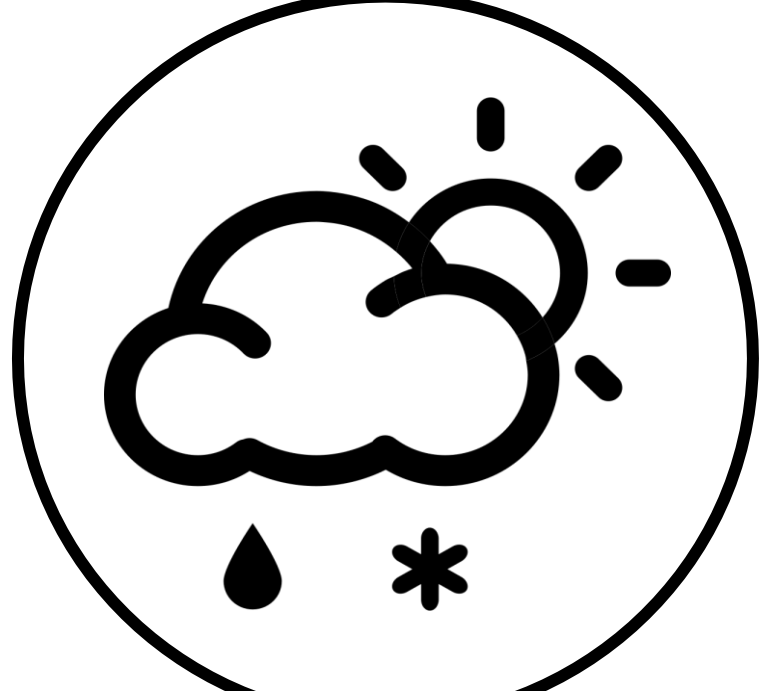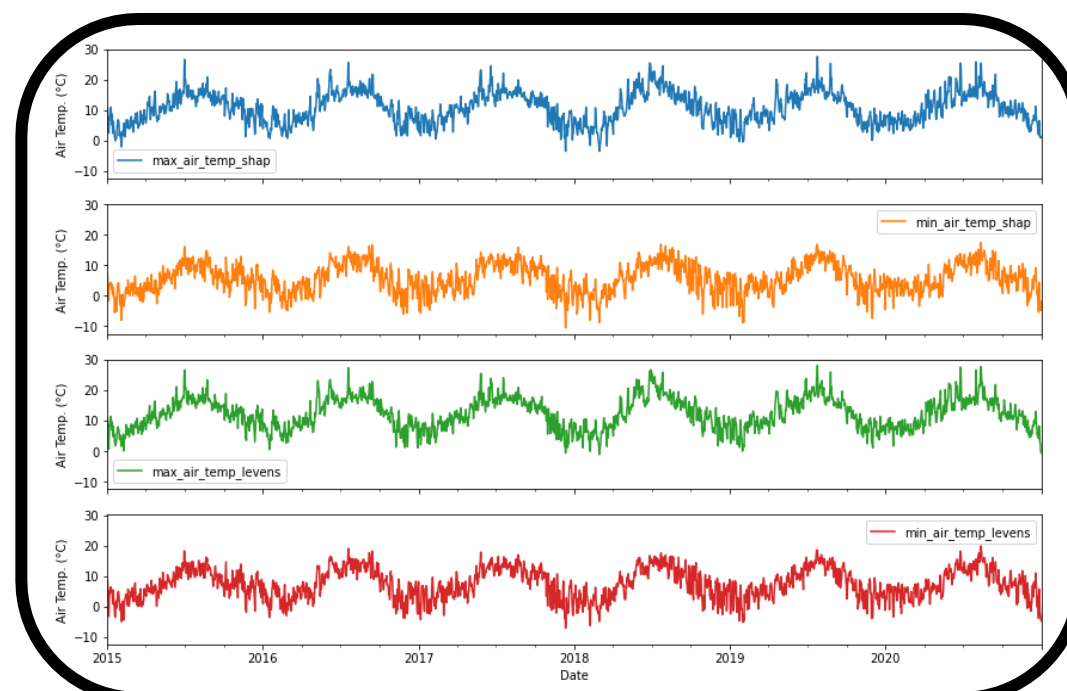
Satellite Remote Sensing
- Data from ESA Sentinel-2 and Planet Super-Doves
- Can estimate chl-a or phycocyanin concentration as a proxy for cyanobacteria
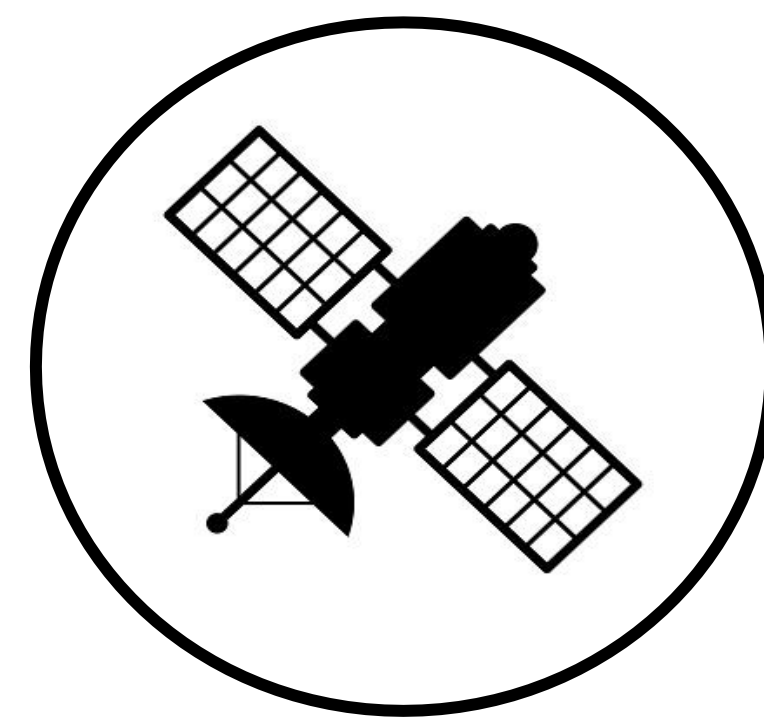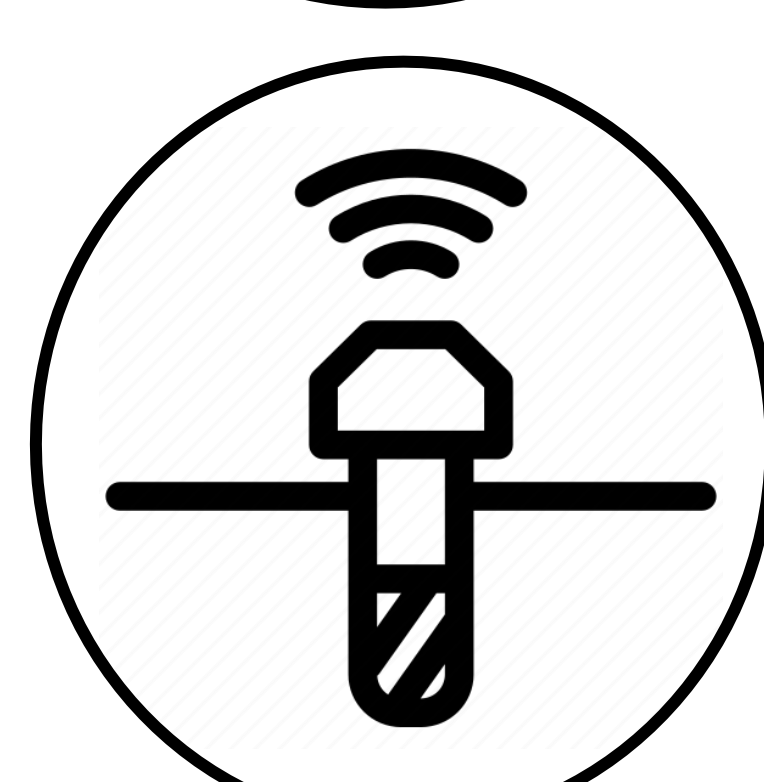
Meteorological Data
- Data from weather stations and forecasts
- E.g. precipitation, wind speed, air temp.

Automated In-situ Sensors
- In-situ reflectance data will be used to calibrate and validate EO data
- Other sensors (eg. water temperature, fluorescence) may be used for monitoring

Data Driven Modelling
- Machine learning models (eg. Neural networks, LSTM, SVM, random forest)
- Training on historical data from multiple lakes
- Potential to investigate coherence of blooms between multiple lakes, which may allow for forecast performance improvements
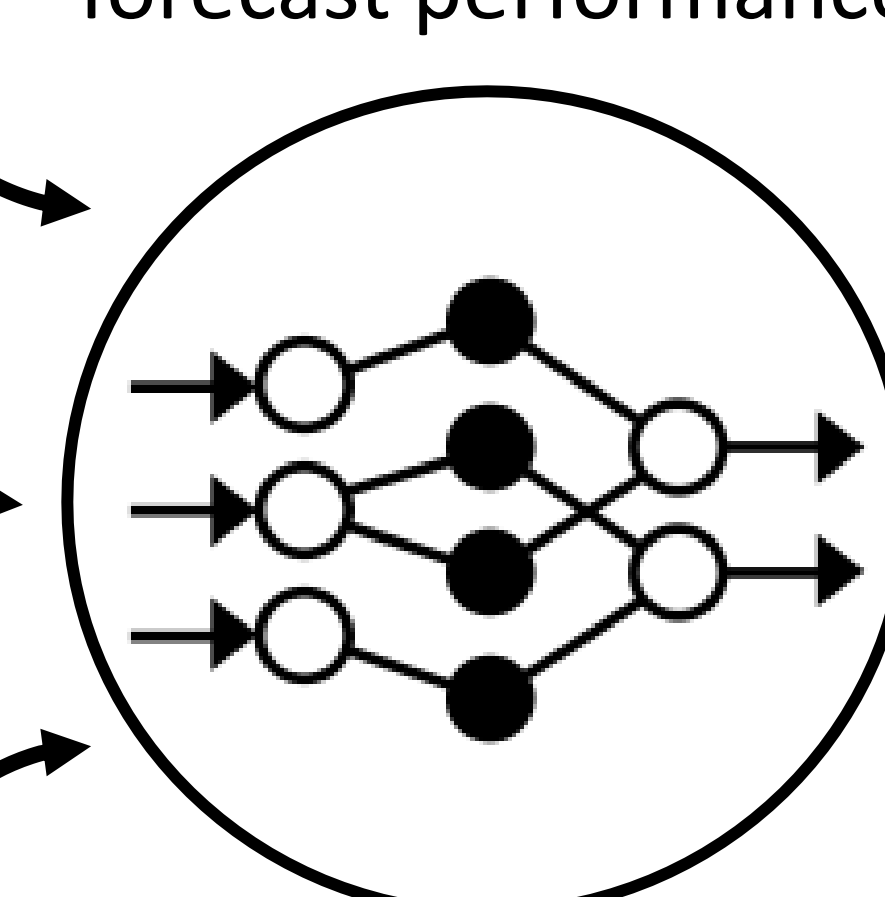
Short-term Bloom Forecasts
- Forecasts for chl-a concentration and bloom likelihood, up to 2 weeks into the future
- Initial forecast validation using hind-casting, looking to implement a real-time forecast
- Possibility to investigate forecasting taste & odour issues associated with cyanobacterial blooms
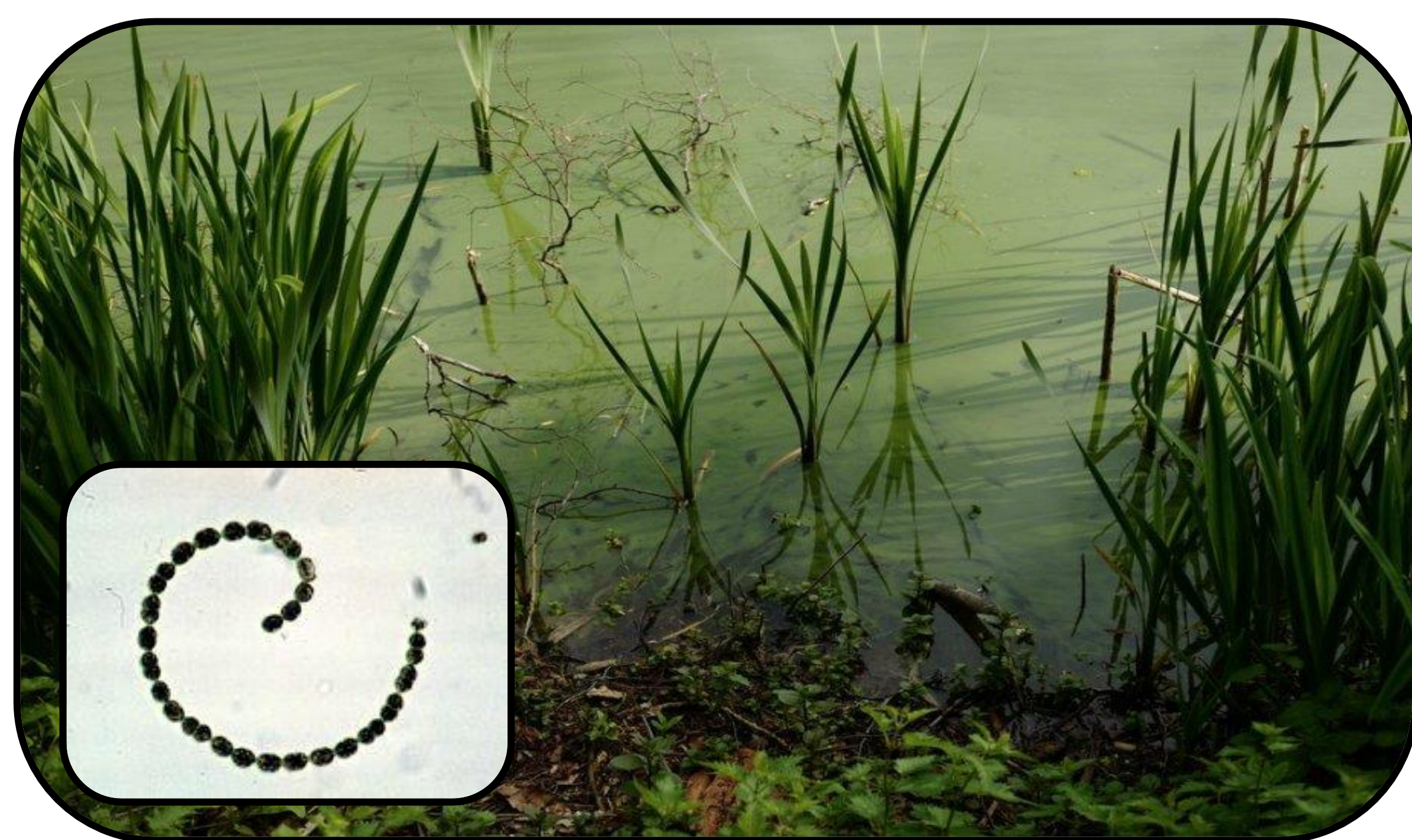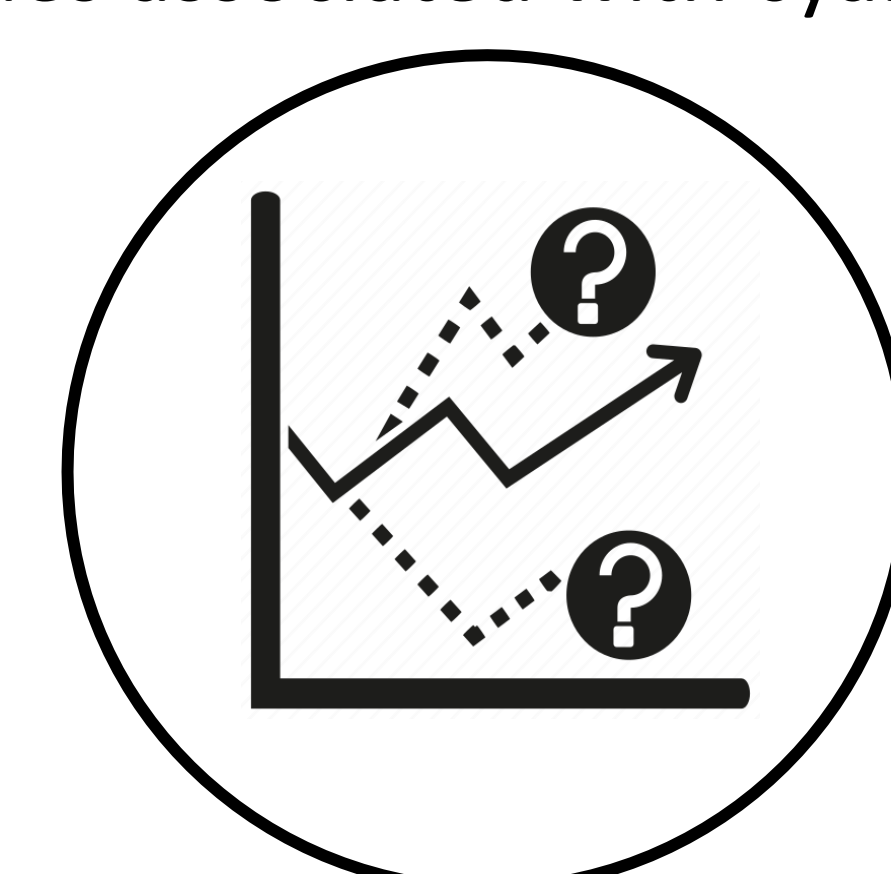
Cyanobacterial bloom in Airthrey Loch (University of Stirling) and close-up of Anabaena, a common cyanobacteria genus (Wikipedia)

## Detecting Cyanobacteria with High Resolution Satellite Constellations

- In-situ hyperspectral reflectance measurements have been matched with *Planet SuperDoves* satellite images from Airthrey Loch (University of Stirling)
- Preliminary analysis indicates that the Superdoves data is likely suitable for retrieval of water quality parameters associated with cyanobacteria (*chlorohyll-a, phycocyanin*)
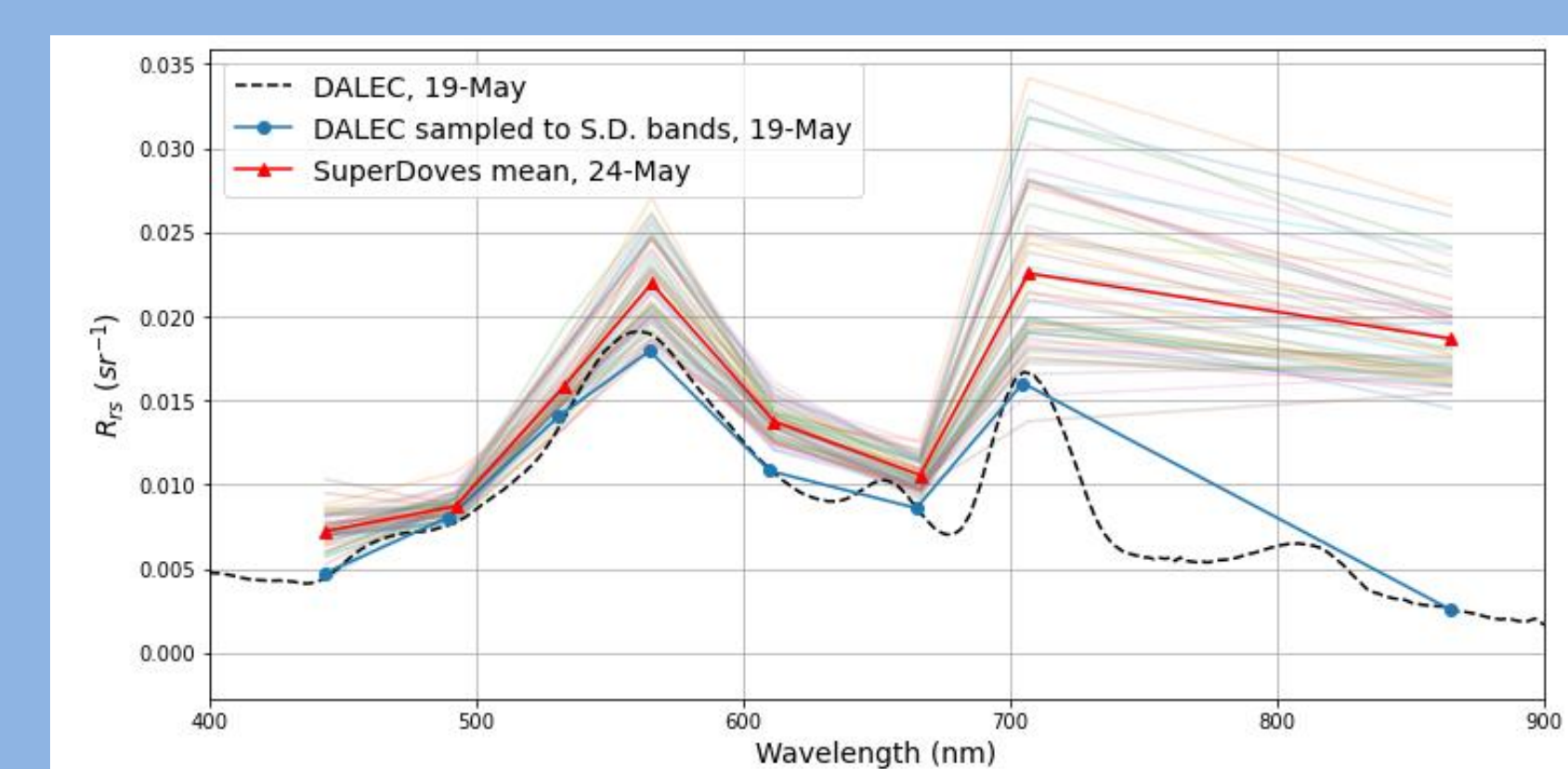
This indicates the potential of new satellite constellations to monitor cyanobacterial blooms in small water bodies at higher temporal resolutions than previously possible

Forecasting Cyanobacterial Blooms

- Forecasting work has been carried out with historical data from Blelham Tarn, a small eutrophic lake in the English Lake District
- A paper is currently in draft which focuses on understanding how important long-term datasets are for training forecasts, and a demonstration of best-practice for benchmarking machine learning forecasts
- Future work will focus on expanding this study with data from other lakes, and evaluating the importance of different parameter sets
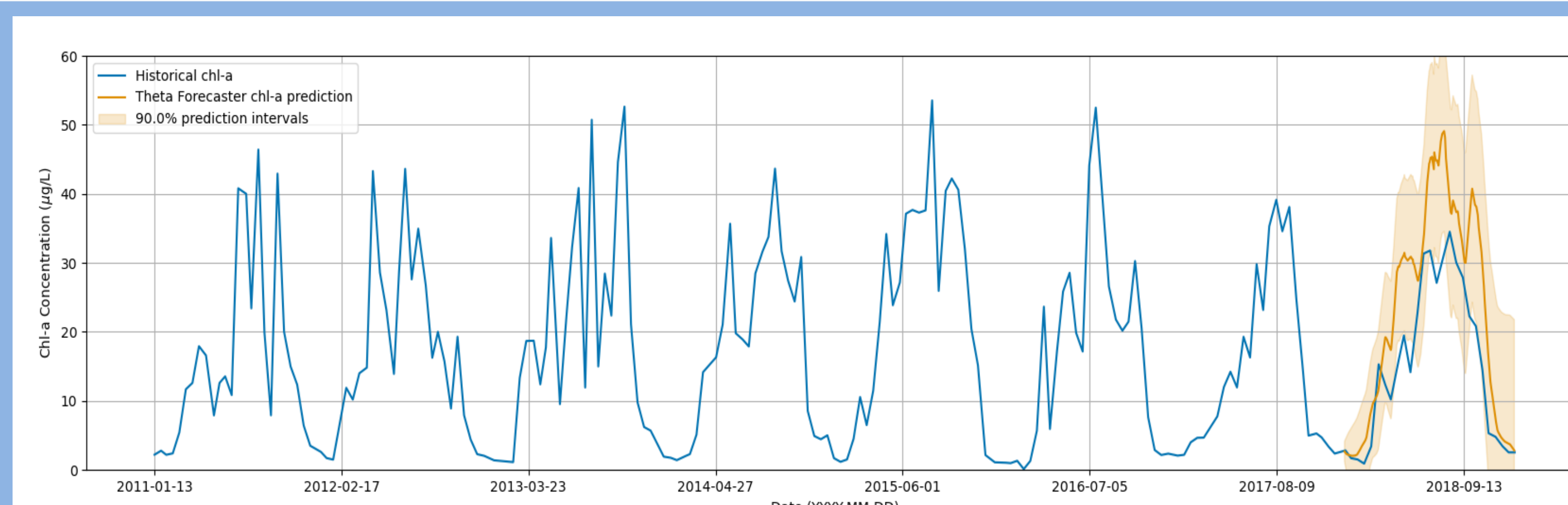
Example forecast: univariate 'Theta Forecaster' applied to historical chlorophyll-a data from Blelham Tarn