

Utilizing of FerryBox data in combination with other information sources for forecasting of cyanobacterial blooms

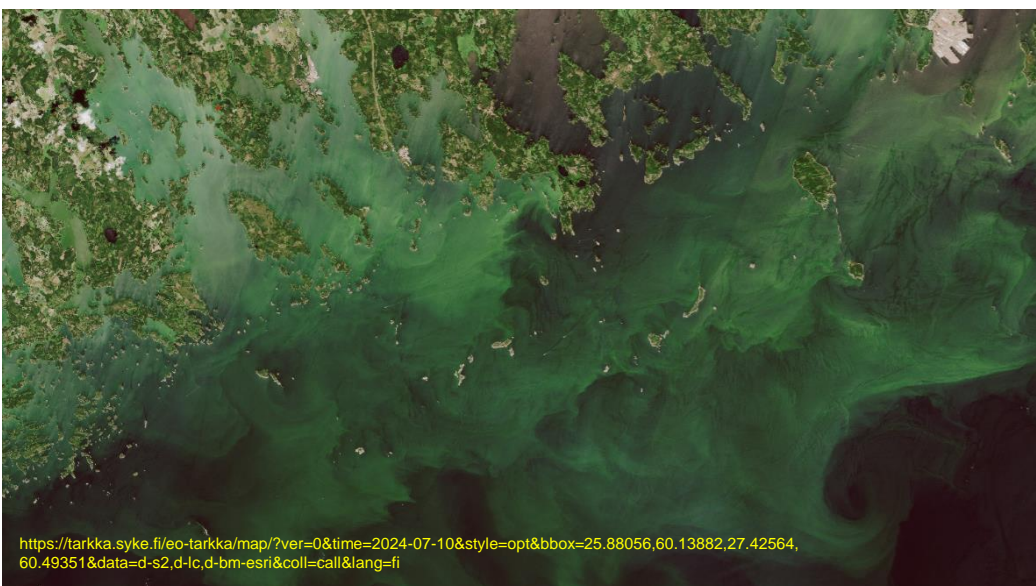
Heikki Peltonen
Finnish Environment Institute (Syke)

12th FerryBox Workshop
1-2 October 2024



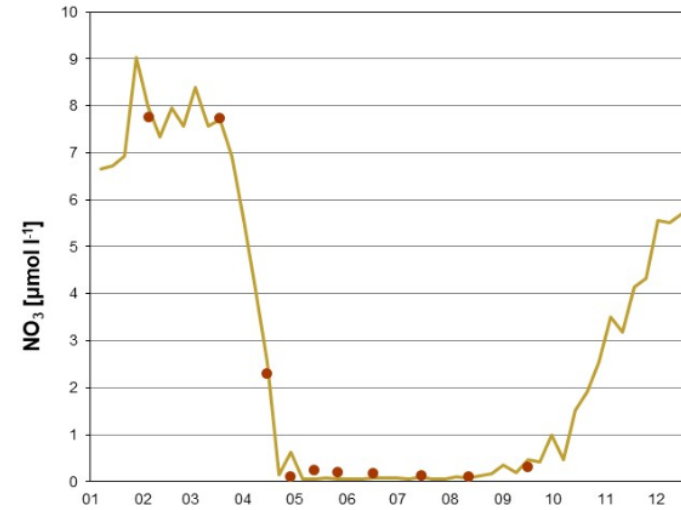
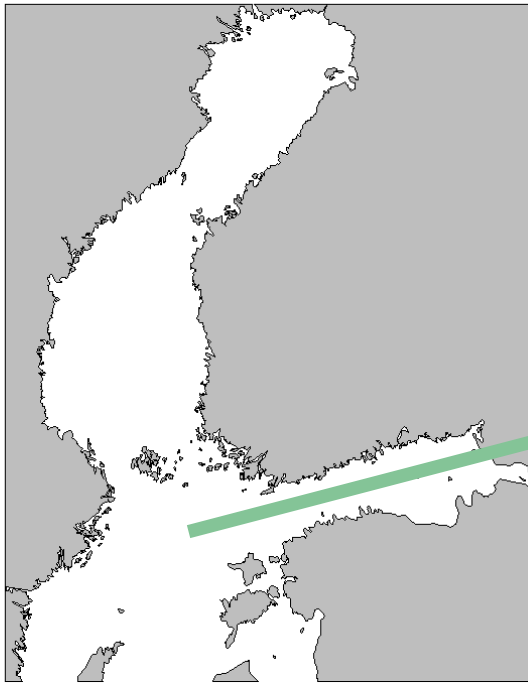
Background

- Cyanobacteria, also known as blue-green algae, produce harmful summertime blooms in large areas on the Baltic Sea
- For uses of marine space and resources, for management and for science it is of interest to know bloom risks in different areas and learn about the mechanisms contributing to the bloom intensity
- The current study integrated multiple data sources, including FerryBox Alg@line data and analyses with machine learning for forecasting of cyanobacterial blooms in summer 2024 in the Northern basins of the Baltic Sea

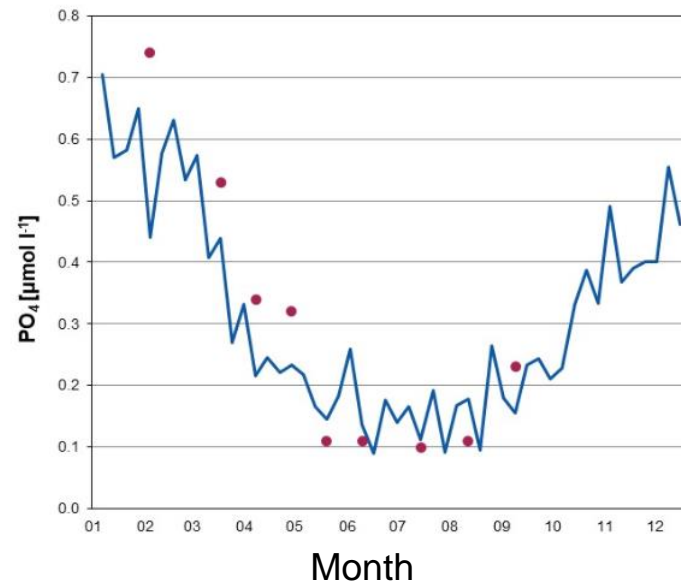


Nutrient concentrations in surface layer after spring bloom are assumed to control the cyanobacterial bloom level in summer

[Time series from Ferrybox monitoring - Marinefinland.fi](https://marinefinland.fi)



● = 2024 monthly mean
— = long-term mean

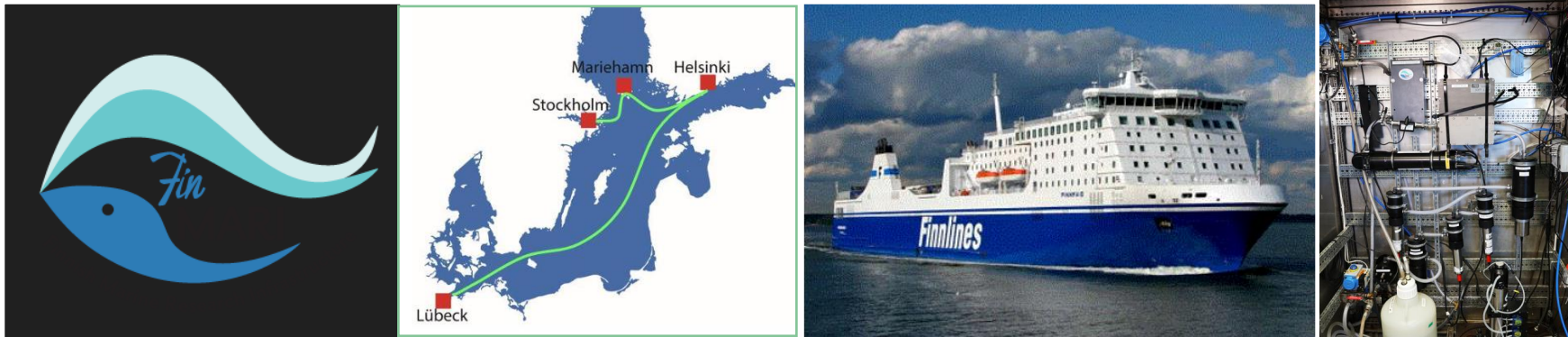


Datasets, model fitting

- Water quality monitoring Finland, Sweden 2003-2023 + Alg@line ferrybox m/s Finnmaid
→ DIN and DIP concentrations → GAM-interpolation → nutrient concentration raster maps
- Satellite instruments, reflectance data, bloom summaries classified in four levels, 2003-2023 (Syke Tarkka service)
- Weather data from 2003-2023, Copernicus Climate Change Service reanalysis, raster maps, 2m monthly temperature averages (May, June, July, August) (+ solar irradiation and wind)

Datasets, forecasts

- Water quality monitoring Finland, Sweden + Alg@line ferrybox *m/s Finnmaid*, *m/s Serenade*) → DIN and DIP from winter 2023-24 → concentration raster maps
- Weather forecast for the spring and summer 2024, Copernicus Services, seasonal temperature forecasts with 51 models → temperatures in May



Pictures: Katri Kuuppo

Modelling methods

Machine learning models, *models suitable for analyses on categorical response data*

- *on 4 bloom levels – low, moderate, considerable, high*



Multinomial logistic regression and Random Forest

Explanatory variables - logistic regression: DIP, DIN, t2m_May, Lat, ID, Year

- random forest: DIP, DIN, t2m_May, Lat, Long, Year

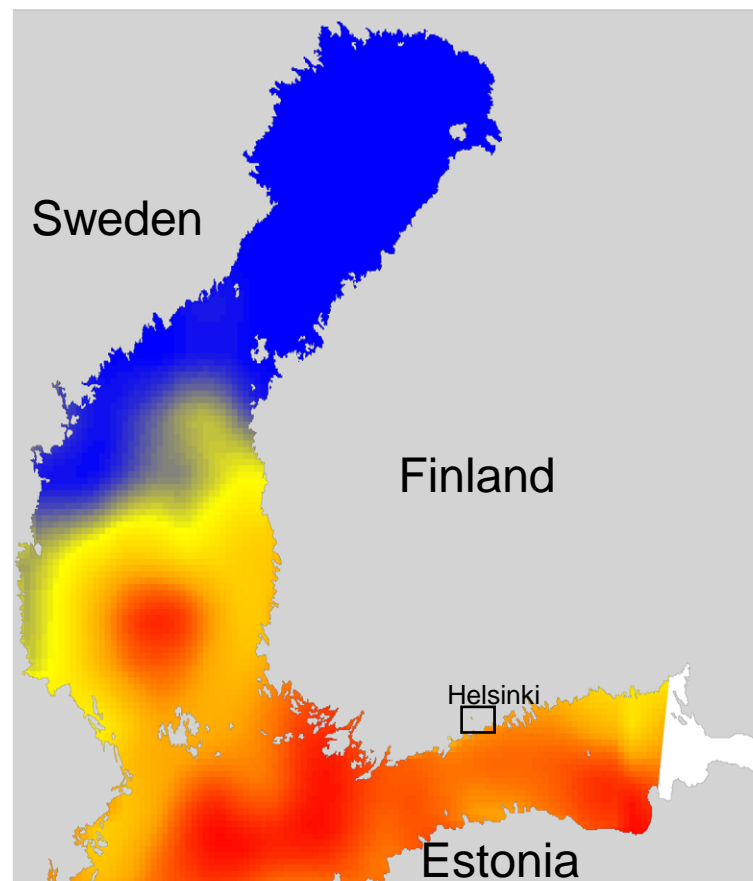
Forecasts for the summer 2024

Logistic Regression and Random Forest models learned from data + datasets for forecasting

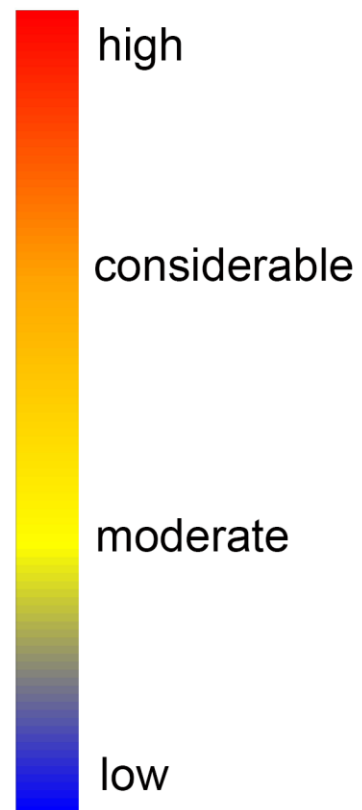
→ ensemble forecast

The forecast for the summer 2024 (published as a press release)

Summer 2024

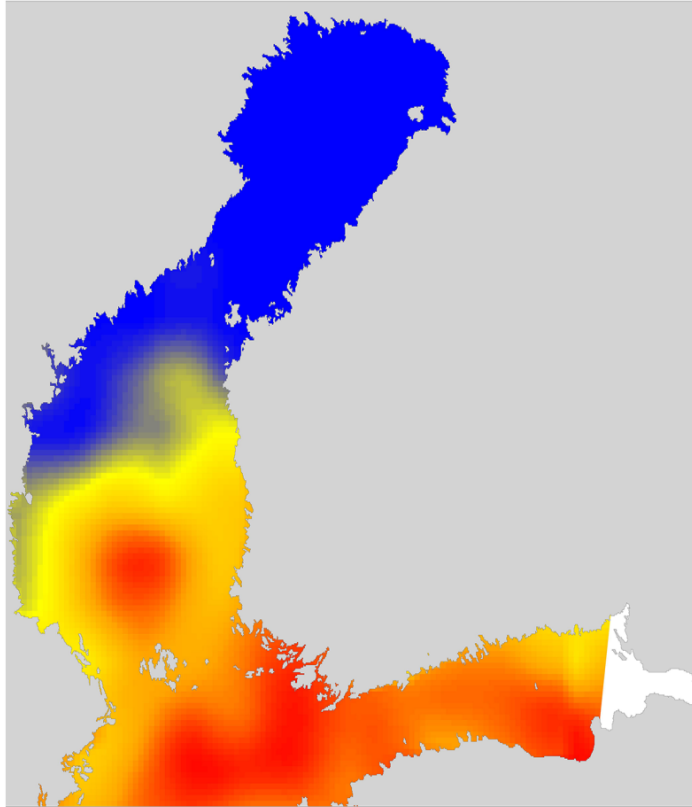


Risk of blue-green
algae blooms



The forecast and the observed blooms

Summer 2024



Risk of blue-green algae blooms

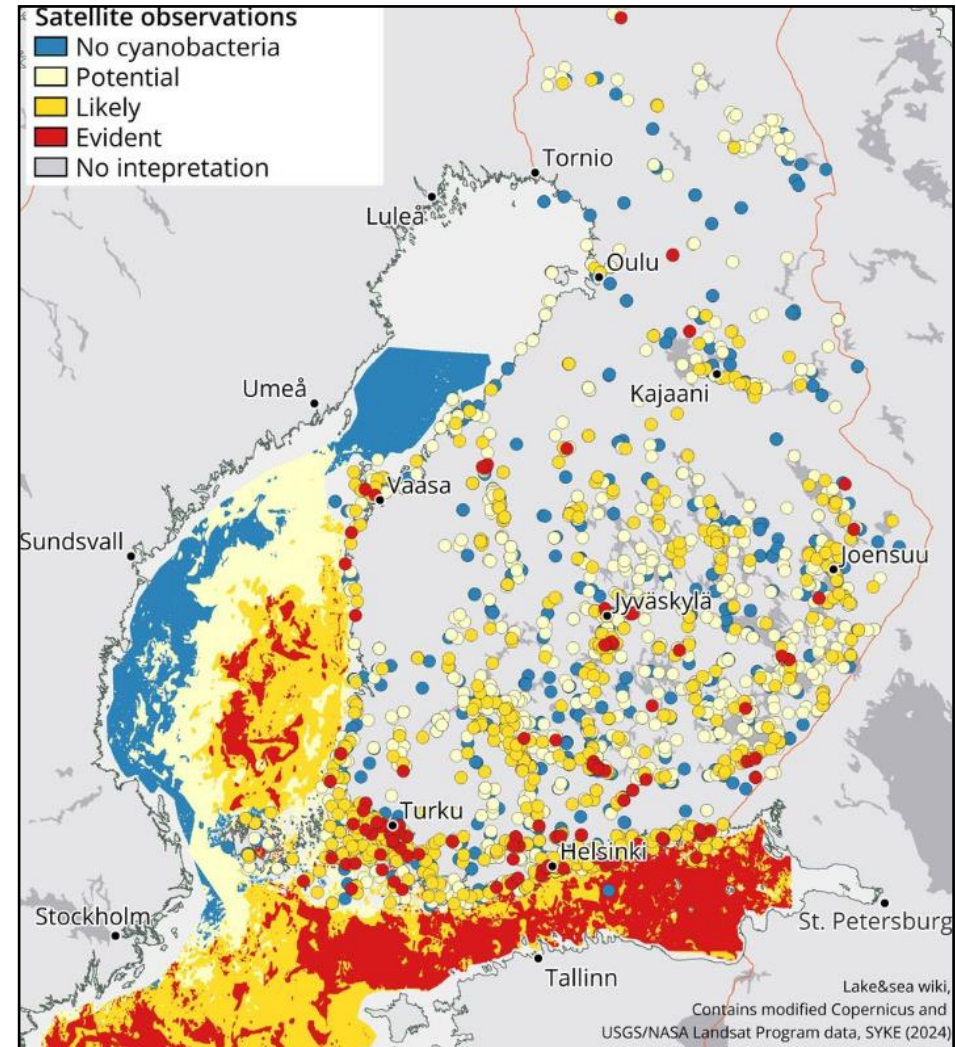


high

considerable

moderate

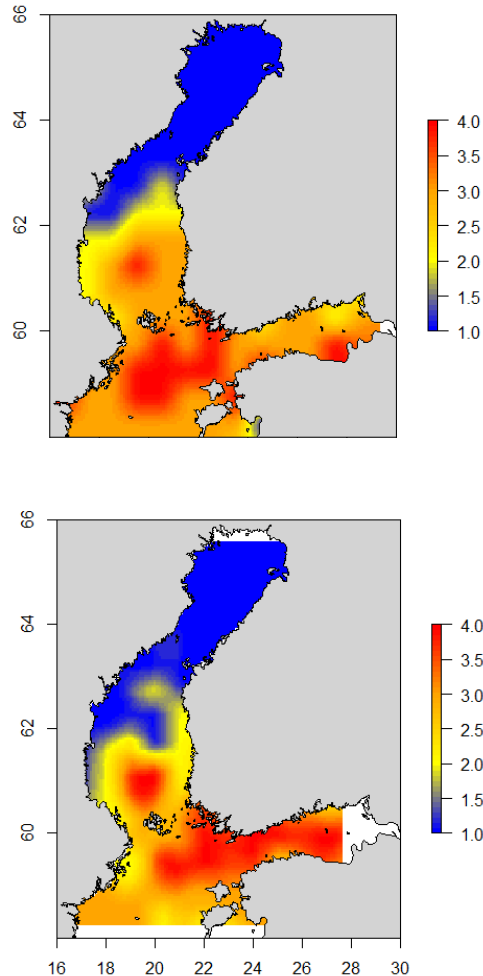
low



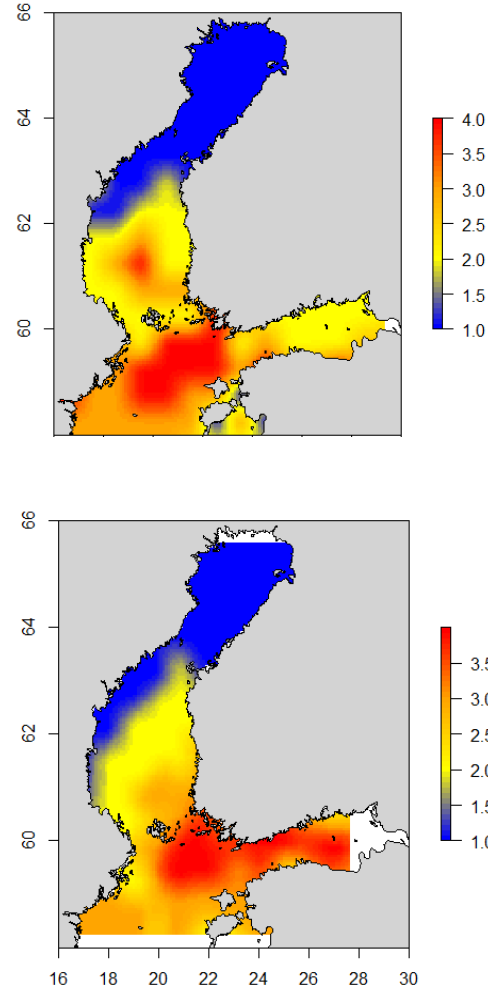
Modelling of blooms - support to management, climate change adaptation, marine ecology research (new hypotheses) etc.

Example – bloom risk forecasts with average, higher and lower spring temperatures

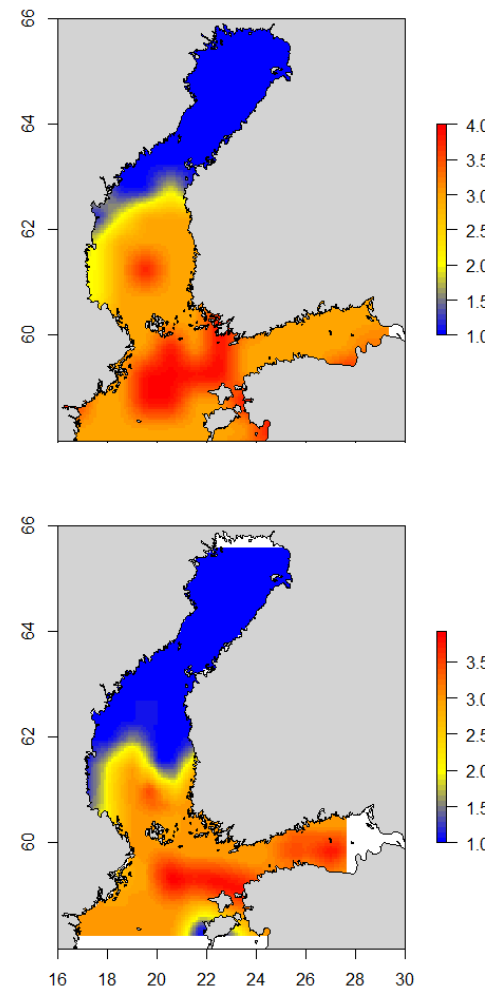
May, temperature + 0



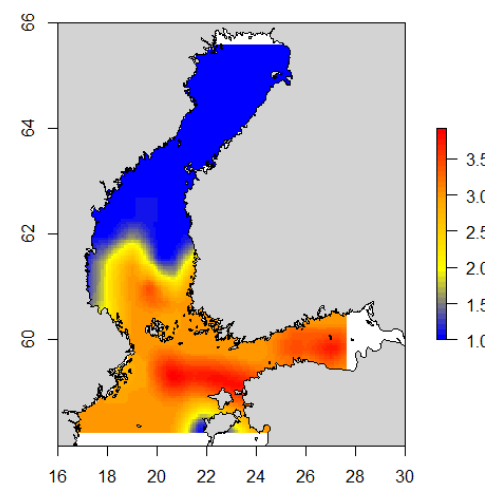
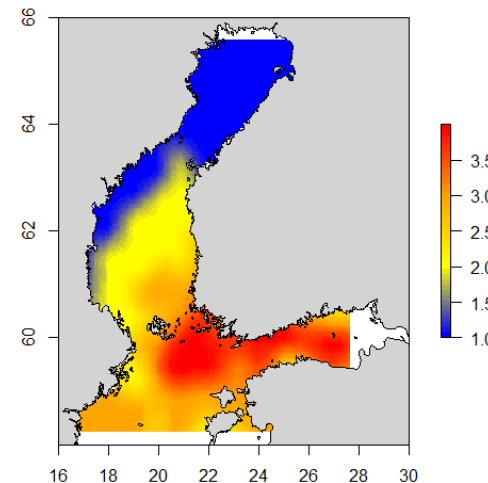
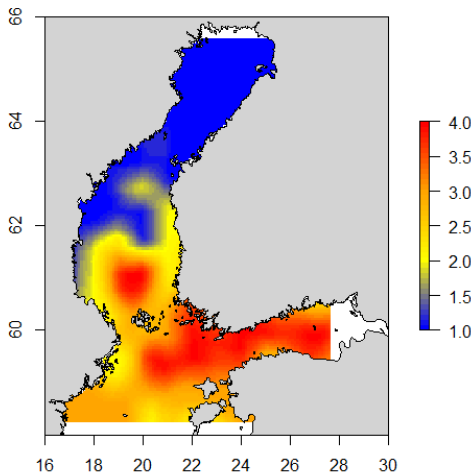
May, temperature + 2



May, temperature - 2



Upper row: Logistic regression



Lower row: Random forest

Next development stages?

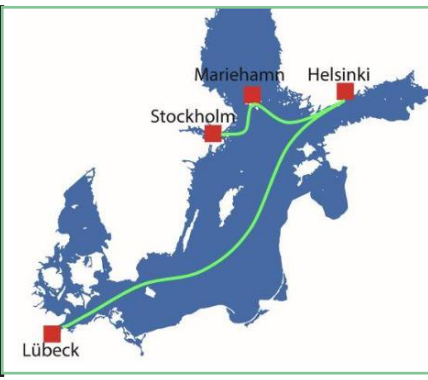
More data sources e.g. SEADATANET-CTD-data, more Alg@line data...

Other machine learning methods, enhanced seeking of best methods

Short term (e.g. weekly) forecasts during summer?

Automated updating with new information?

Dynamic biogeochemical modelling? (currently suitable models not available)



Pictures: Katri Kuuppo

Several people in Syke contributed to the work

- Sebastian Ehrhart, Petri Maunula, Jukka Seppälä – **Alg@line**
- Jenni Attila, Sakari Väkevä – SYKE Tarkka, satellite data
- Elina Miettunen – Copernicus Climate Change Service
- Antti Räike – VESLA+SHARKWEB, marine monitoring data
- Harri Kuosa, Sirpa Lehtinen – algal bloom ecology
- Jouni Lehtoranta, Samuli Korpinen – project management