Improving red tide detection machine learning models with ship movement data

Lab seminar 02 March 2023

Dominic Soh (B4)

Motivation

Damage caused by harmful algal blooms (HABs)

Need for **monitoring and pre-emptive measures**

Sustainability of monitoring efforts

- Resource intensity
- Consistency

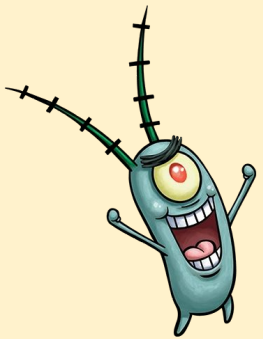
Idea to incorporate shipping data from thesis advisor, Prof. Takeuchi

Research Question and hypothesis

Question: Can maritime vessel movement be used to improve Red tide prediction models?

Hypothesis: Introduction of maritime vessel movement as training parameters will increase Red tide prediction accuracy.

What are HABs (Harmful Algal Blooms)?



Phytoplankton + Excessive nutrients = Algal blooms

Phytoplankton:

- inhabit ocean waters all around the world.
- Species can vary based on region.

Proliferate based on various environmental conditions like:

- dissolved oxygen, temperature
- nitrogen
- phosphorous

Benign VS Harmful

- Some blooms algal are relatively benign (but obstructs sunlight when bloomed etc.)
- Some directly kill fish by blocking their gills or poisoning birds when in sufficient concentration.

Detection of HABs

Red tide Detection Index (RDI)

$$\mathrm{RDI} = \left(\frac{1}{R_{\mathrm{rs}}(\lambda_1)} - \frac{1}{R_{\mathrm{rs}}(\lambda_2)} \right) \times R_{\mathrm{rs}}(\lambda_3)$$

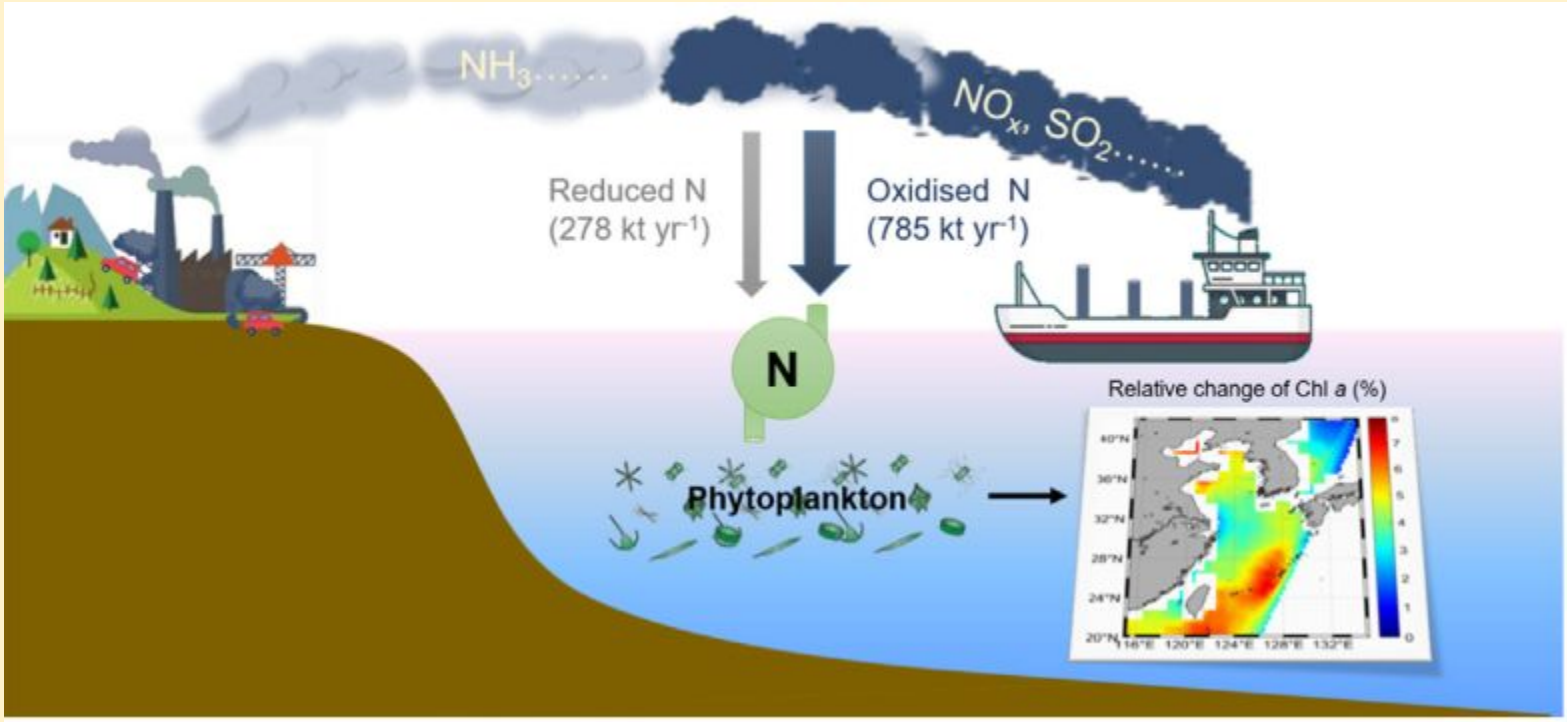
Band 1: should be maximally sensitive to absorption by chlorophyll a (660-690nm)

Band 2: Modified to 550-570nm from 710-730nm

Band 3:740–760 nm, considering the scattering effects of particulate matter.

Threshold value: > 0.175 signifies positive ID for algal bloom (Chl a ~10µg/L)

How do ships affect phytoplankton growth?



Approach

Features of interest: euphotic depth (ED), Secchi disk depth, chlorophyll-a, chlorophyll-gsm, chlorophyll-giop, diffuse attenuation coefficient (Kd_490), SST, FLH, particulate backscattering coefficient at 547 nm (bbp_547_giop), and turbidity index. (MODIS)

New feature: Global shipping lane traffic data (AIS) / nighttime light data that can determine ship profile

Model: GBDT? ANN? DBN? Particular model is TBD but the focus is not on the type of model but the use of shipping lane traffic data as a new feature.

Detection: Red tide Detection Index (RDI)

Why use Machine Learning (ML)?

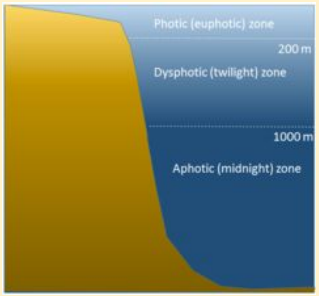
relationship between phytoplankton growth + environmental variables: complicated and nonlinear

traditional methods for forecasting like time series regression analysis is insufficient - eg: ARIMA (autoregressive integrated moving average)

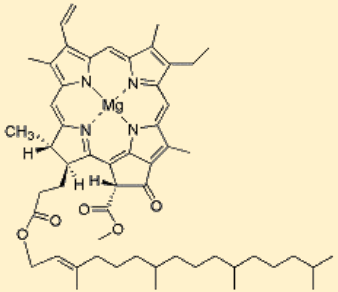
ARIMA

flawed as a linear correlation structure is assumed among the time series values, meaning NO nonlinear patterns can be captured by the ARIMA model

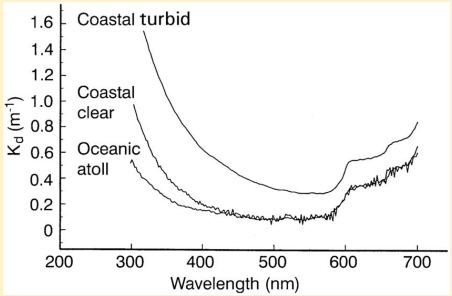
Feature explanation



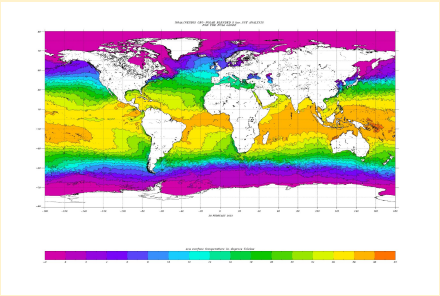
Euphotic depth /
Secchi disk depth (m)



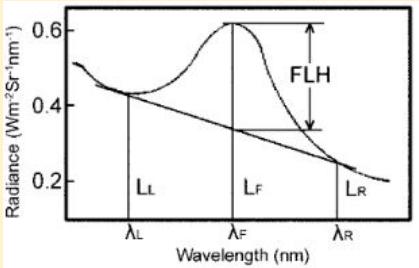
Chlorophyll a (mg/m3)



Diffuse attenuation
coefficient (Kd_490; m−1)



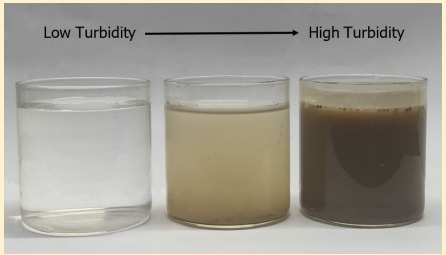
Sea surface temperature (◦ C)



Fluorescence line height

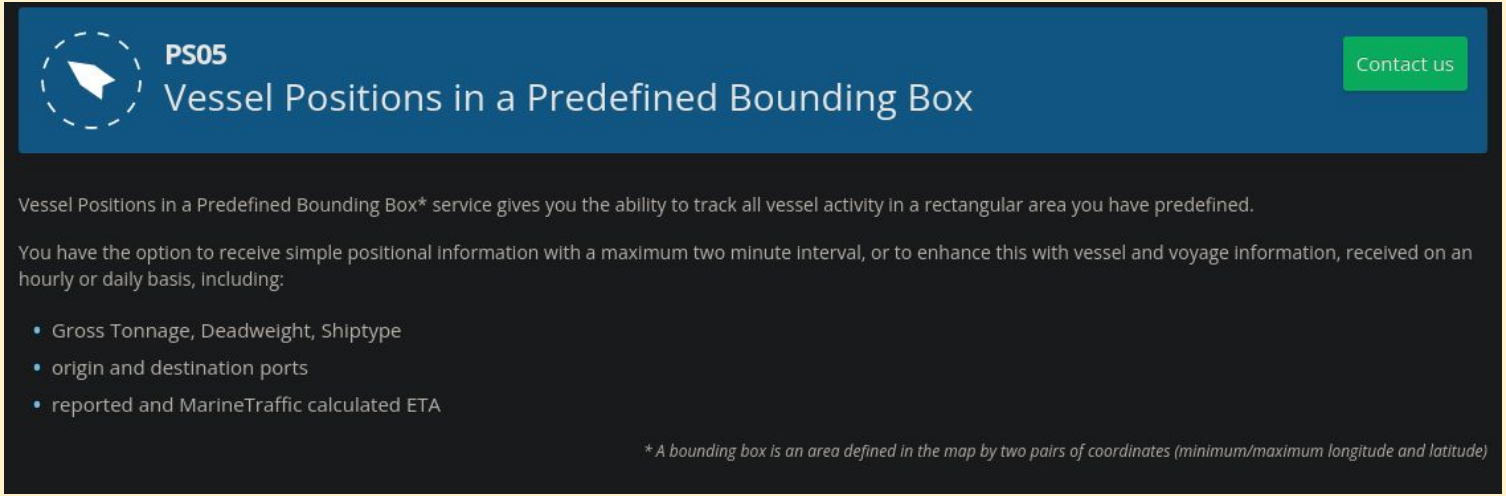
$$b_{bp}(\lambda_0) = b_{bp}(55x) = \frac{u(\lambda_0) \times a(\lambda_0)}{1 - u(\lambda_0)} - b_{bw}(55x)$$

Particulate backscattering
coefficient



Turbidity index

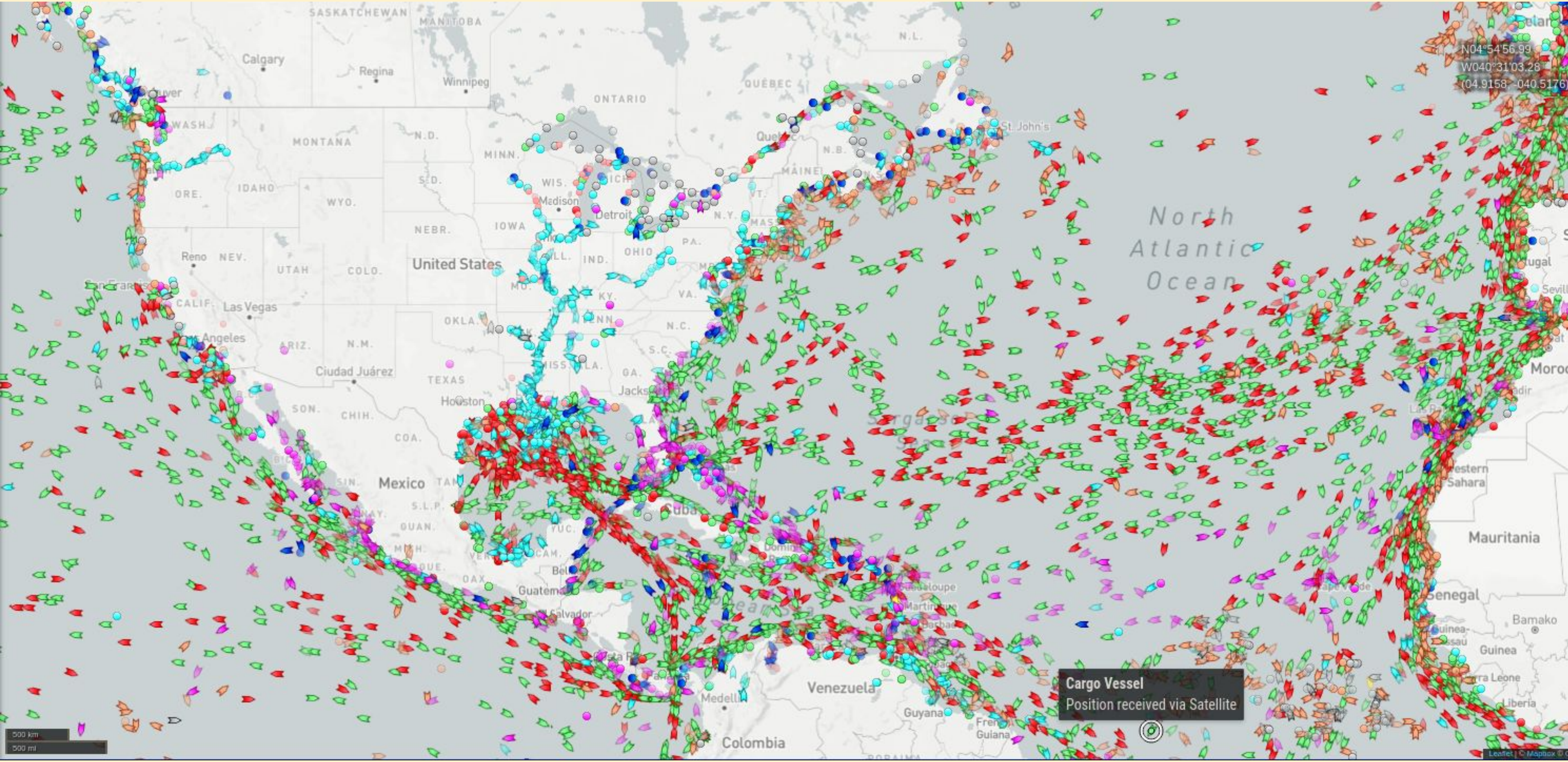
How would shipping lane data be used?



Data collected from AIS/nighttime light can determine the size of ship that passed and the number of ships passing through.

We can classify the ships into several categories

*(assuming that **size of ship** is proportional to **emissions**)*



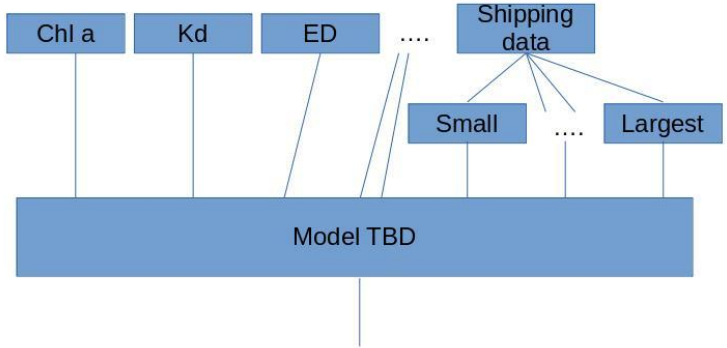
SASKATCHEWAN
MANITOBA
Calgary
Regina
Winnipeg
ONTARIO
QUÉBEC
N.L.
St. John's
WASH.
MONTANA
N.D.
MINN.
IDAHO
ORE.
WYO.
S.D.
WIS.
Madison
Detroit
N.Y.
NEBR.
IOWA
PA.
OHIO
IND.
Reno
NEV.
UTAH
COLO.
United States
MO.
KY.
VA.
CALIF.
Las Vegas
OKLA.
N.C.
S.C.
ARIZ.
N.M.
Ciudad Juárez
TEXAS
GA.
Houston
SON.
CHIH.
COA.
Mexico
North Atlantic Ocean
Cuba
Venezuela
Colombia
Guyana
Mauritania
Senegal
Bamako
Guinea
Liberia
Cargo Vessel
Position received via Satellite
N04°54'56.99
W040°31'03.28
(04.9158, -040.5176)
500 km
500 mi

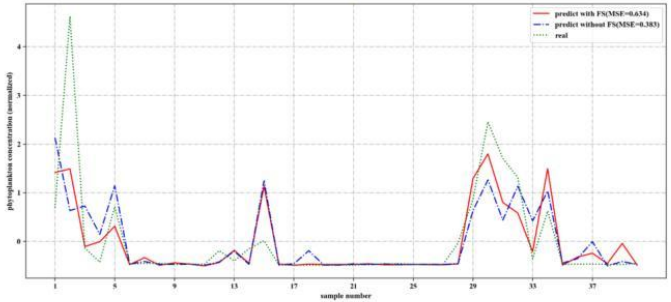
Analysis to do:

Decide on maximum distance from region of interest for ships to be counted.

Integrate number and type of ships as a new feature and compare prediction accuracy with base model.

Training to be done via supervised learning using RDI for cross validation of algal blooms.





Some prediction

Questions and Future work

Formulate methodology:

- Select a region of interest to analyse.
- Determine distance from ROI for ships to be considered to have passed through.
- Choose type of machine learning model.

Research:

- Read more on machine learning theory.
- Understand more on characteristics and data from MODIS Aqua/Terra.
- Investigate more on how HAB occurs and forcing mechanisms.