Discovering the Most Predictive Data Modalities for Algal Bloom Forecasting

Kadir Tolga Gökçe - 5039169
K.T.Gokce@student.tudelft.nl

Responsible Professor: Dr. Jan van Gemert (j.c.vangemert@tudelft.nl)
Supervisors: Attila Lengyel (a.lengyel@tudelft.nl),
Robert-Jan Bruintjes (r.bruintjes@tudelft.nl)

1. Background

- An algal bloom is identified as a rapid increase in common algae (phytoplankton) abundance in water bodies and when the growth of algae populations is out of control, the damages on ecosystems are severe [1].
- Chlorophyll-a concentration of the water body can be used as the direct indicator of algal blooms.
- Remote sensing is the process of detecting and monitoring the physical characteristics of an area.
- Using remotely sensed data with machine learning makes it possible to forecast harmful algal blooms.
- Multiple types of environmental measurements are used together to forecast algal blooms.
- Each environmental measurement data used to forecast algal blooms is referred to as a "data modality".

Figure 1: Damage of a harmful algal bloom on fish population.

2. Research question

- *Which input modality is the most predictive for estimating chlorophyll-a concentrations for water bodies in Uruguay?*

3. Methodology

- A Linear classifier is used as the machine learning model to predict chlorophyll-a concentrations for Palmar water reservoir in Uruguay.
- Input data is constructed in data cube (combination of multiple data modalities) structure containing the remotely sensed measurements of 11 different environmental factors, as demonstrated in Figure 2, that are relevant for detecting algal bloom events. Furthermore, the data is clipped, normalized, and the missing values are replaced with mean of the data, as the pre-processing steps.

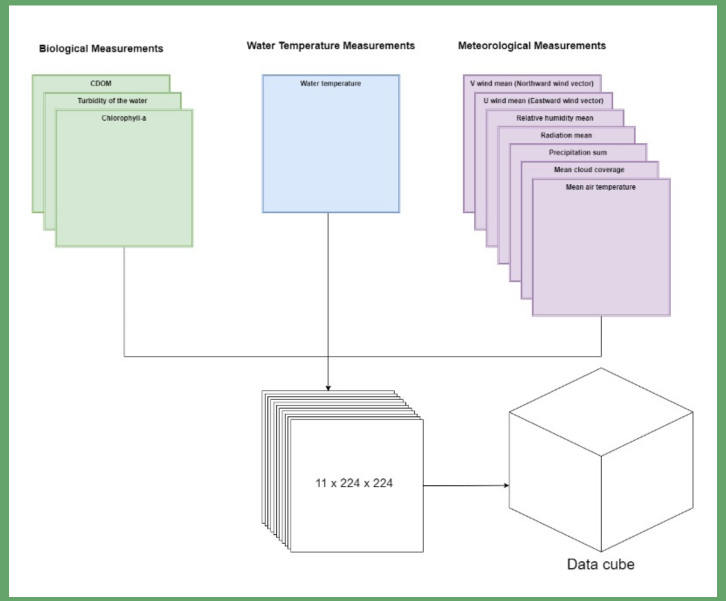

Figure 2: Construction of data cubes with 11 data modalities.

- Chlorophyll-a concentrations belonging to the day after when the input data is obtained, are used as the ground truth labels to be able to forecast the next day. The ground truth values are assigned to the intervals of **0-10** ug/L, **10-30** ug/L, **30-75** ug/L, and **75+** ug/L.
- The machine learning model is trained with each data modality individually against the chlorophyll-a concentrations and tested with unseen data so that the accuracy metrics are collected to produce the results.

4. Results

Accuracy scores calculated for each data modality and when all data modalities are combined are presented in Table 1 and Table 2.

Data Modality	Chlorophyll-a	Turbidity	CDOM	Water Temperature	Mean Air Temperature	Mean Cloud Coverage
Accuracy Score	27.47%	**34.33%**	25.87%	22.56%	17.67%	27.66%

Table 1: Accuracy scores belonging to 6 data modalities for predicting chlorophyll-a concentrations.

Data Modality	Precipitation Sum	Radiation Mean	Relative Humidity Mean	U Wind Mean	V Wind Mean	All Modalities Combined
Accuracy Score	30.71%	**34.86%**	23.38%	31.95%	31.14%	27.63%

Table 2: Accuracy scores belonging to 5 data modalities and when all data modalities are combined for predicting chlorophyll-a concentrations.

5. Conclusion & Limitations

- Turbidity of water and radiation mean are the most predictive single data modalities with accuracy scores around 34%, which is even higher than when all modalities are combined, however, using individual data modalities fail to predict all intervals of chlorophyll-a and they are more meaningful when combined.
- The linear classifier does not produce accurate results and is not suitable for algal bloom forecasting.
- More informative evaluation metrics such as confusion matrix should be utilized for better comparison of the results.

References

- [1] R. Santoleri. "Year-to-year variability of the phytoplankton bloom in the southern Adriatic Sea (1998–2000): Sea-viewing Wide Field-of-view Sensor observations and modeling study". In: Journal of Geophysical Research 108.C9 (2003). DOI: 10 .1029 / 2002jc001636. URL: http://dx.doi.org/10.1029/2002jc001636.
- Figure 1: https://ioc.unesco.org/news/toxic-algal-blooms-cause-more-economic-damage-aquaculture-any-greater-storm
- Background image: Foster, Joanna M. (20 November 2013). "Lake Erie Is Dying Again, And Warmer Waters And Wetter Weather Are To Blame". ClimateProgress. Archived from the original on 3 August 2014. Retrieved 3 August 2014.