Detection of Harmful Algal Blooms (HABs) in Lakes at the Regional Scale Using Satellite Remote Sensing and Machine Learning

Gabriel Abreu-Vigil[1,6], Grace Tan[2,6], **Alastor Sherbatov** [3,6], Fahmeda Khanom[4], Marzieh Azarderakhsh[4], Nicholas Smirnov[5,6], Dr. Hamid Norouzi[4,6], Dr. Reginald Blake[4,6]

[1]M.S. Geography at University of Oregon, [2]Mathematics Teacher at Science, Technology and Research Early College, [3]B.S.E. Mechanical Engineering at Columbia University School of Engineering and Applied Sciences, [4]CUNY New York City College of Technology, [5]Earl L. Vandermeulen High School, [6]NASA Goddard Institute for Space Studies

Abstract

Harmful algal blooms (HABs) in freshwater systems are being exacerbated due to climate change and pose significant ecological and human health risks, thus necessitating effective monitoring and management strategies. Traditional ground-based monitoring methods are limited by high cost, labor intensiveness, site-specificity and unforeseen disruptions such as the COVID-19 pandemic. In this study, we propose a remote sensing approach to address these challenges, enabling cost-effective and continuous monitoring of HABs in more than 35 lakes around New York State leveraging the existing in-situ data from various monitoring programs. We test four different algorithms derived from band relationships to obtain chlorophyll-A (Chl-a) as a proxy for HABs presence on Sentinel-2 Multispectral Instrument (S2-MSI) Surface Reflectance (SR) data. There was a large range in formula accuracy; however, all exhibited ample error, leading us to incorporate different machine learning techniques into an expanded study of several large lakes in NY State, including Lake Champlain. Various S2-SR bands were examined as features and their importance was estimated using the in-situ data. The best-performing algorithm for Lake Champlain, our test region, Gradient Boosting, performed better than others with R^2=0.85. This approach addresses the need for effective regional HABs detection and monitoring in freshwater systems through the incorporation of remote sensing and machine learning techniques.

Background

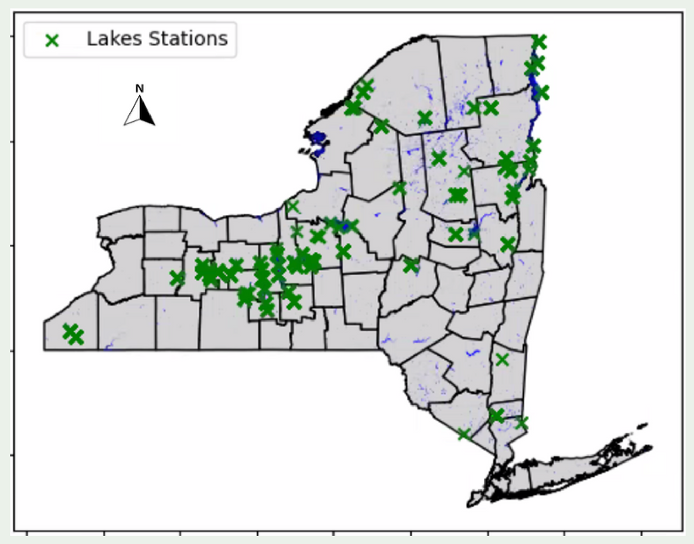
- Algae, cyanobacteria, and phytoplankton are simple organisms found worldwide, capable of blooming uncontrollably and producing toxins, leading to harmful algal blooms (HABs)
- Factors contributing to HABs include slow-moving water, sunlight, and excessive nutrients/fertilizer from runoff
- Other than producing toxins, HABs can cause eutrophication and hypoxia, negatively impacting many species and aquatic ecosystems
- Climate change and landcover changes exacerbate HAB frequently and severity

Landsat 8 image of Finger Lakes, July 2023

- Accurate identification and prediction of HABs in local water sources are essential.
- Current methods involve expensive and time-consuming field studies to collect samples and assess the danger of algal masses.
- Novel approaches, such as satellite sensing and machine learning, are being developed alongside fieldwork to optimize HABs identification and warning systems.

Methods

i. Find in-situ data
ii. Research algorithms
iii. Evaluate Chl-A algorithms for Cayuga Lake
iv. Expand lake dataset to other NY lakes
v. Incorporate machine learning; try several methods
iv. Compile analytics on ML algorithm & formula accuracies



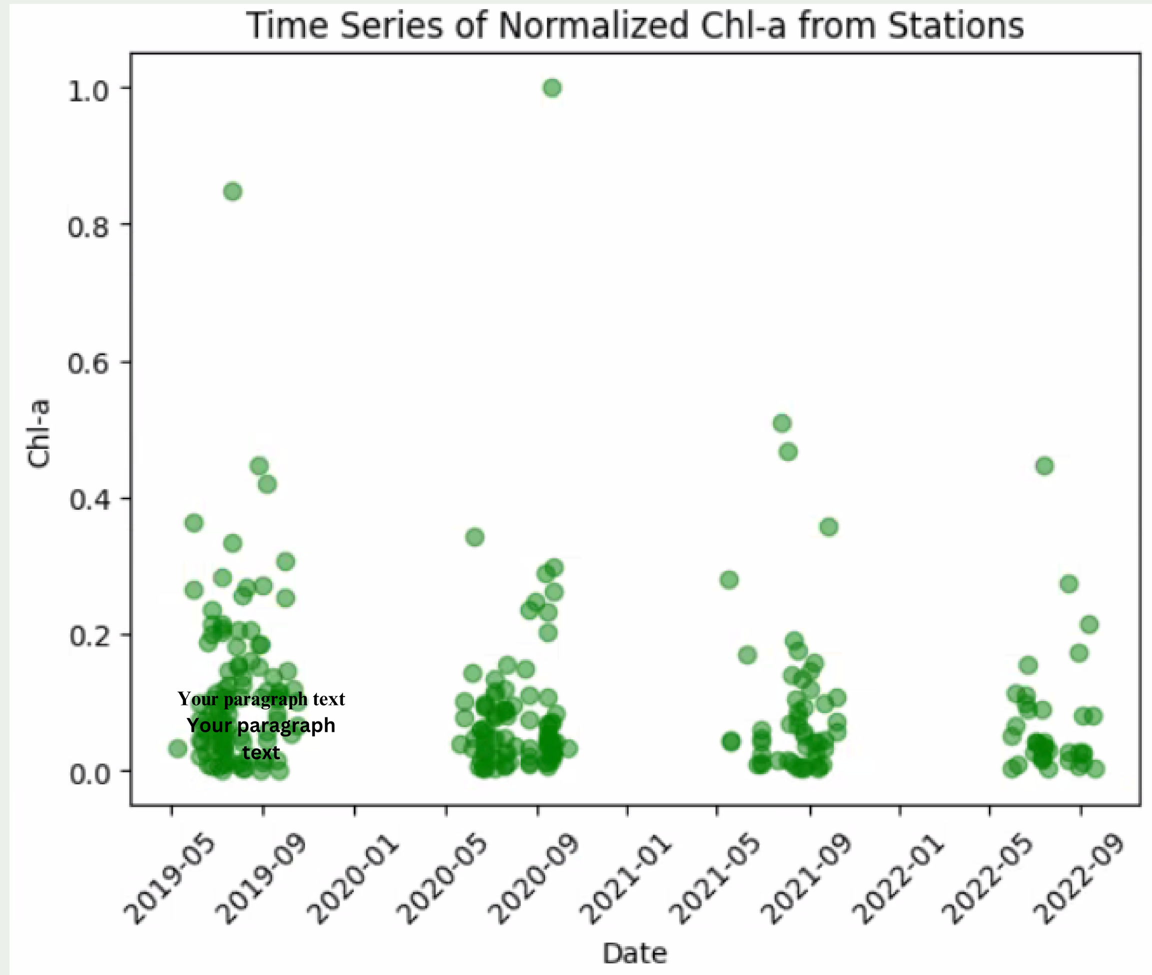
Chlorophyll-a Band Algorithms

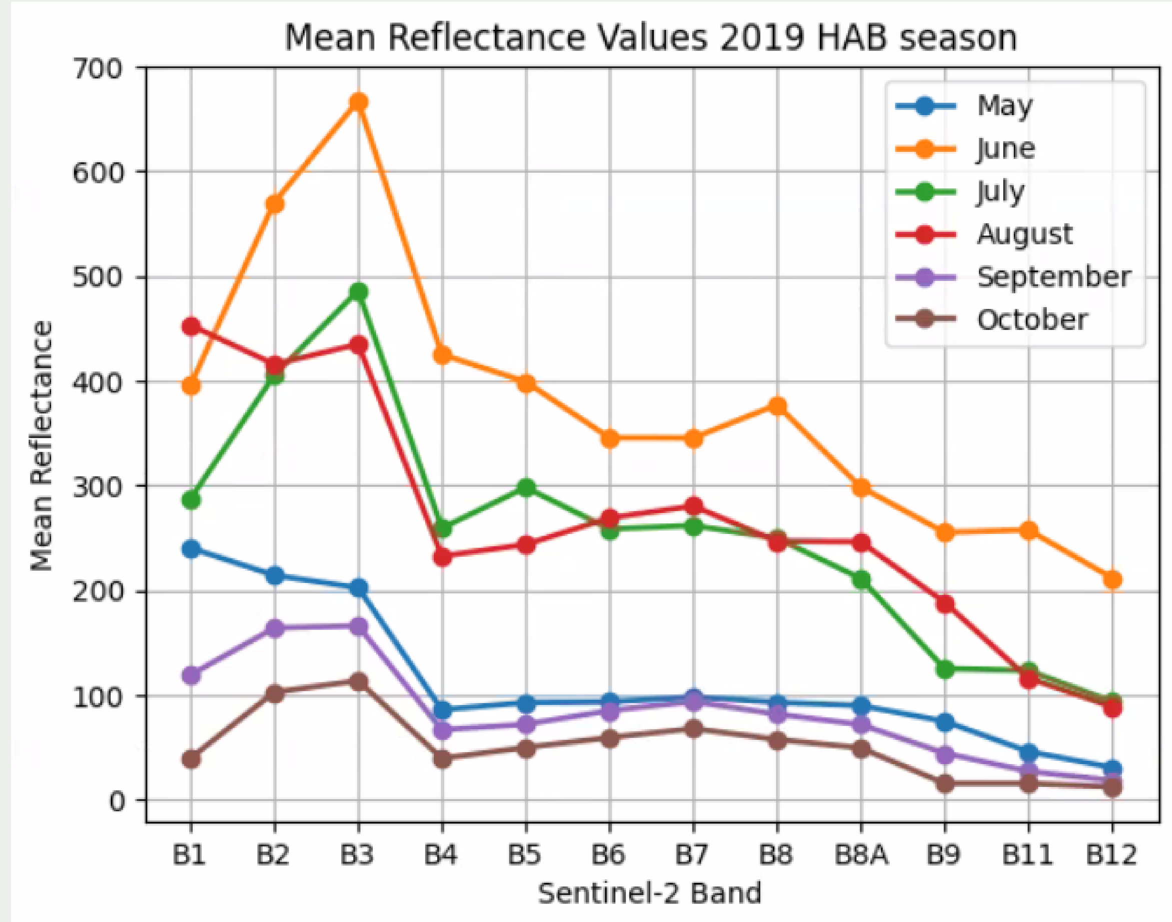
NDCI (Normalized Difference Chlorophyll Index)	**3BDA**	**2BDA**	**SABI** (Surface Algal Bloom Index)
Mainly for use in S2 due to the presence of red-edge band **Landsat:** $\frac{nir-red}{nir+red}$ **Sentinel 2:** $\frac{rededge-red}{rededge+red}$	*Mainly for use in Landsat* $\frac{blue-red}{blue+red}-green$	$\frac{nir}{red}$	$\frac{nir-red}{blue+green}$

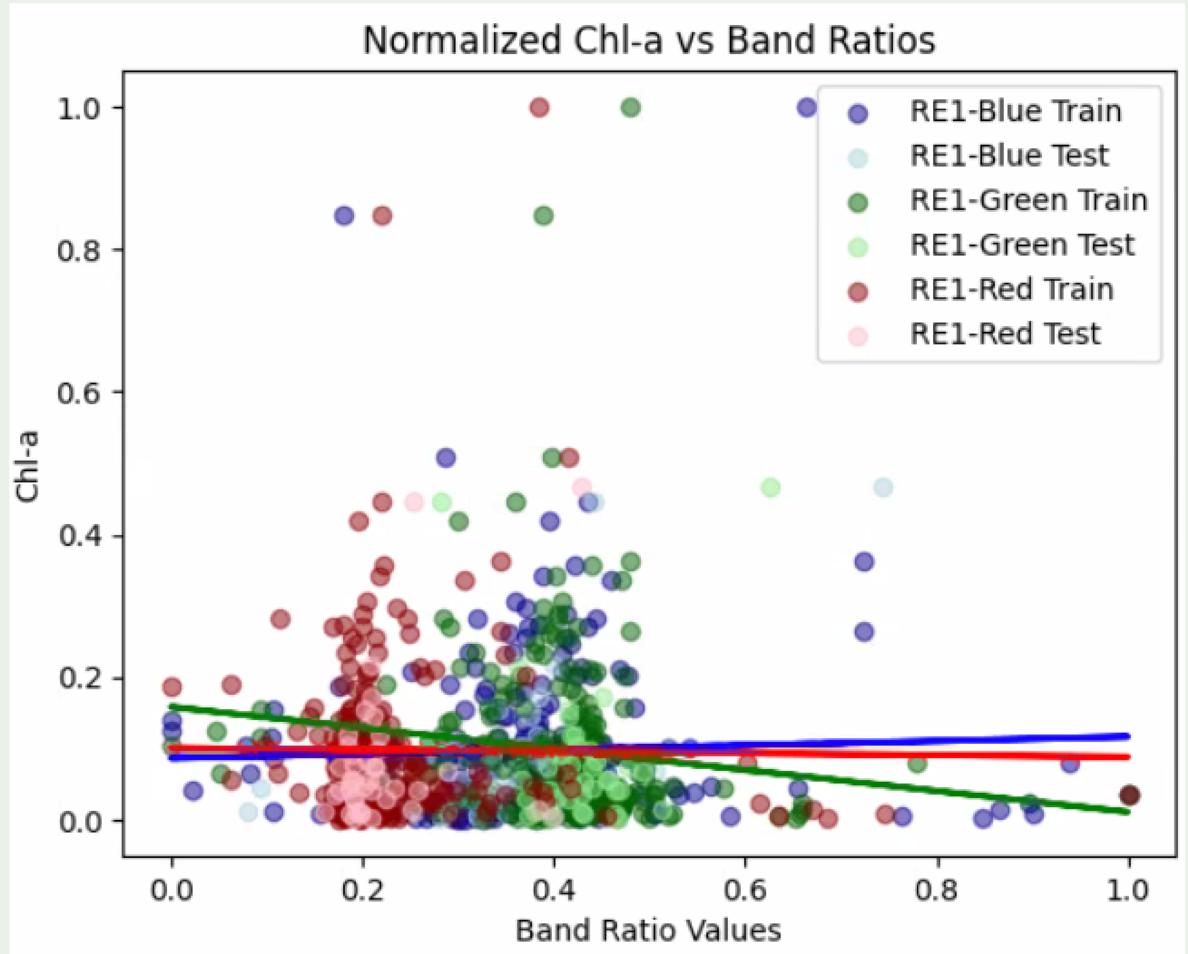
Machine Learning Algorithms

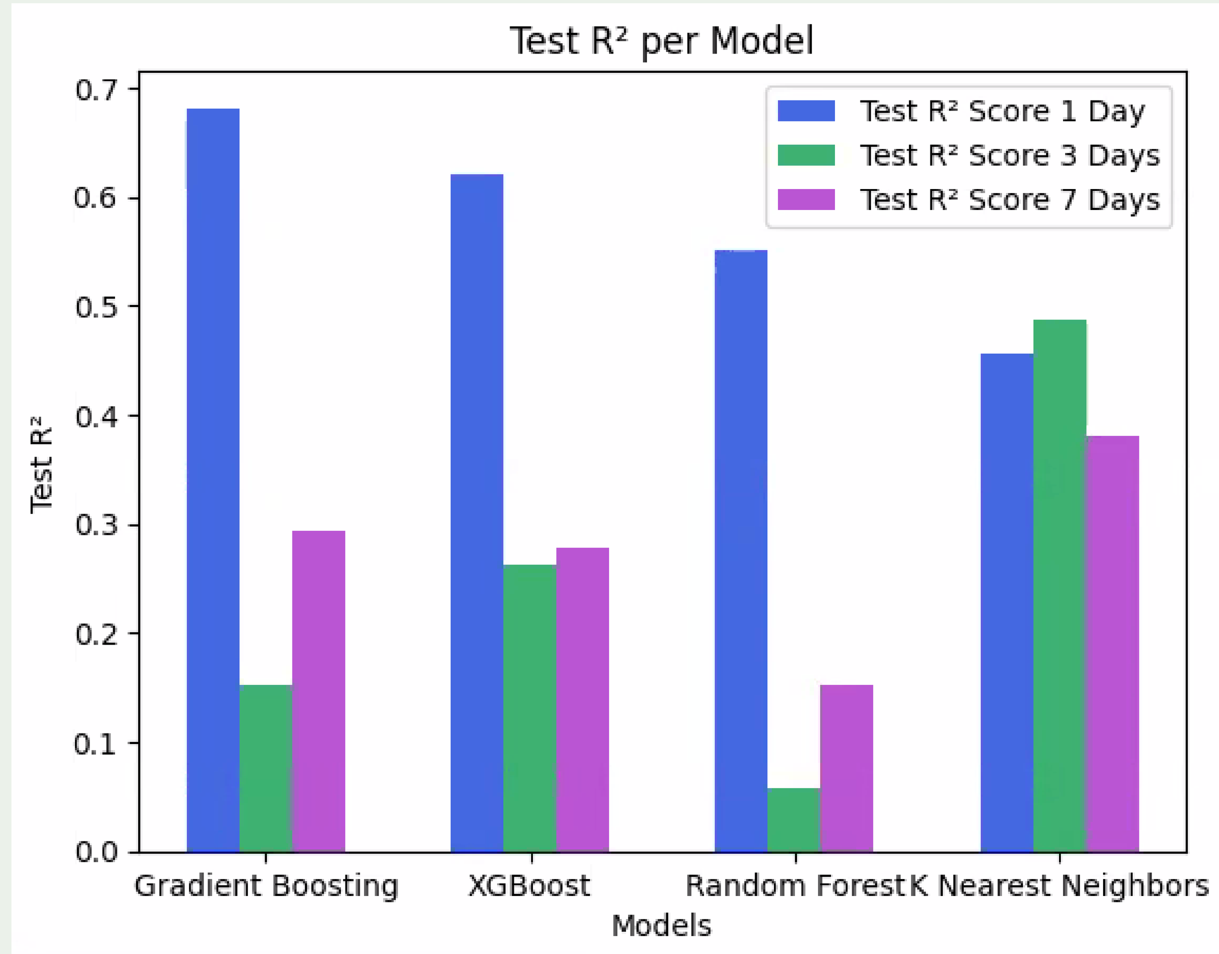
Random Forest (RF)
Gradient Boosting (GB)
AdaBoost (AB)
Neural Network (NN)
K Nearest Neighbors (KNN)
XGBoost (XGB)
Support Vector Regression (SVR)

Results









Conclusions

- HAB concentration does not follow a linear relationship
- Traditional ground-based monitoring methods have limitations, including high cost, labor intensity, site-specificity, and vulnerability to disruptions like the COVID-19 pandemic. Remote sensing offers a viable alternative for continuous and cost-effective monitoring of HABs

Discussion

- Mixing in-situ data from different sources is likely to lower final R^2 and accuracy
- While adding more lake in-situ data reduces the R^2, combining data from multiple sources can lead to error as there are different methods of data collection (volunteer sampling vs. consistent station monitoring)

Future Work

- Use MPI4Py/parallel computing to speed up reflectance retrieval
- Expand set of lakes to include NJ, PA, VT & other states
- Incorporate near-real-time in-situ data for more accurate and consistent data
- Create opportunities for collaboration between CCRI and CSLAP HAB researchers

References

Boucher, J., Weathers, K.C., Norouzi, H., & Steele, B. (2018). Assessing the effectiveness of landsat 8 chlorophyll a retrieval algorithms for regional freshwater monitoring. Ecological Applications, 8(4), 1044–1054.

Bramich, J., Bolch, C. J. S., & Fischer, A. (2021). Improved red-edge chlorophyll-a detection for Sentinel 2. Ecological Indicators, 120.

Caballero, I., Fernández, R., Escalante, O.M. et al. (2020). New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters. Scientific Reports 10 (1), 8743.

Chen, J., Zhu, W., Tian, Y.Q., Yu, Q., Zheng, Y., Huang, L. (2017). Remote estimation of colored dissolved organic matter and chlorophyll-a in Lake Huron using Sentinel 2 measurements. Journal of Applied Remote Sensing. International Society for Optics and Photonics 11 (3), 36007.

C

Acknowledgments

We want to thank Dr. Hamid Norouzi and Dr. Reginald Blake for their help and support throughout this research. We learned and benefited greatly from their expertise and guidance. We would also like to thank Fahmeda Khanom and Dr. Marzi Azar for their help in data analysis and computer programming. Finally, we want to thank Matthew Pearce for his invaluable assistance and encouragement throughout our CCRI research endeavors.