Satellite-Based Analysis of Water-Color and Dissolved Organic Content Within Inland Lakes

Aisha Malik[1], Fahmeda Khanom[2], Touheda Khanom[2], Marzi Azarderakhsh[3], Charles T Driscoll[4], Kevin C Rose[5]

[1]CUNY Hunter College (REU Scholar), [2]CUNY NYC College of Technology (IUSE Scholar), [3]CUNY NYC College of Technology, Department of Construction Management & Civil Engineering, [4]Syracuse University, Department of Civil and Environmental Engineering, [5]Rensselaer Polytechnic Institute, Department of Biological Sciences

Research Objective

We seek to explore spatial and temporal patterns of water chemistry data that leverage historic and long-term monitoring programs with satellite remote sensing imagery.

The research is centered on assessing changes in water color as an indicator of water quality beyond the selected sampling lakes through a future Survey of Climate change and Adirondack Lake Ecosystems (SCALE) Pilot Program.

Bear Pond

Study Region

- Adirondack Park (Upstate NY)
  - One of US' largest protected wilderness areas → protected by the Clean Air Act to control anthropogenic impacts on lake health
  - Previously experienced acidification in its lakes → Adirondack Lake Survey (ALS) collected extensive field data on water chemistry, fish populations, and more, revealing the lake recovery
  - Concerns regarding algal blooms, rising lake temperatures, and other threats to lake ecosystems continue to exist in this region.

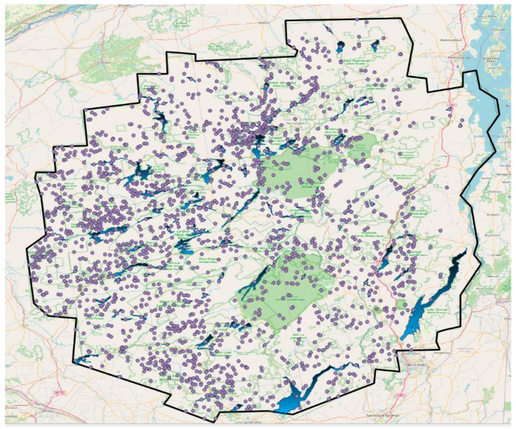

Figure 1. Location of sampled lakes in Adirondack Park in 1984-86

Datasets

Satellites series used:

- **LANDSAT-5 & 7**
  - 30m pixel resolution
  - 16-day temporal resolution
  - Landsat 5 (1984 - 2013)
  - Landsat 7 (1999 - 2022)

In-situ Data: Total > 1,000 lakes

1469	ALS
28	AEAP
41	EPA TIME
51	NY CSLAP
173	ELS
70	EMAP
86	NY LCI
54	ALTM

Water quality data available in LAGOS

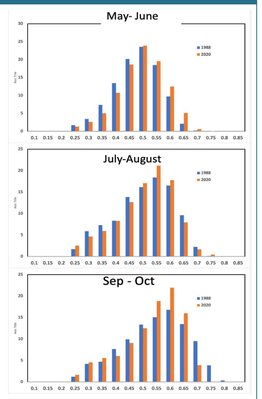

Figure 3. Comparison of NDWI index between Landsat 5 (1988) vs Landsat 7 (2020). Higher color change in Sep-Oct

Existing Algorithms

Table 1. Existing equations selected for re-testing with Landsat 5 & 7 for CDOM prediction

Model	Equation form	Coefficients	R^2	RMSE	Lake Type (size, depth, etc)	CDOM	R^2	RMSE
Brezonik et al. (2015)	$\ln(_{a440}) = a_1 + a_2$ (B2) + a_3(B2/B5) *Landsat 8*	2.304, − 255.88, − 0.2542	0.53	1.32	Depth: 1.8 - 39.0 meters	N/A	0.00	15.99
Olmanson et al. (2016)	$\ln(_{a440}) = a_1 + a_2$(B1/B4) + a_3 (B2/B4) *Landsat 8*	29.2, 109.3, − 143.2	0.83	0.44	N/A	Low	0.06	8.38
Martins et al. (2018)	$a_{CDOM}(485) = a_1 + a_2$ (B4/B1) *Landsat 5*	− 0.5986, 5.5510	0.91	0.78	N/A*	High*	0.00	20479
Olmanson et al. (2020)	$\ln(_{a440}) = a_1(R_{rs}(B4)/ R_{rs}(B3)) + a_2(R_{rs}(B5)/ R_{rs}(B3)) + a_3$ *Landsat 8*	0.42, 1.79, 6.07	0.85	0.49	Size: ≥ 4 ha	Low	0.02	0.72
Koll-Egyed et al. (2021)	$\ln(CDOM_{(a440)}) = a_1 - a_2$ (ln(B3/B4)) - a_3 (ln(B2) *Landsat 8*	3.65, 2.91, 0.41	0.47	0.73	Size: ≥ 10 ha	Low	0.00	0.67
This study:	CDOM = $a_1 - a_2$(B2/B3) + a_3 (B3/B4) *Landsat 8*	20.3, 10, 2.4						

Table 2. Band ratios are tested with significant lakes achieving an R^2 >0.5

Lake Name	Martins et al.	Kutser et al.	Brezonik et al.	Menken et al.	Ficek et al.	Griffin et al.	Simple ratio			
	B4/B1	B2/B3	B1+B1/B4	B3/B2	B2/B3	B3+B2/B1	B1/B2	B2/B4	Max depth (m)	Color Range
FOREST LAKE-W	0.52							0.51	9	10-60
TROUT POND		0.62		0.68	0.62				8.5	3-160
JENNY LAKE			0.64			0.68			7	1-140
WINDOVER LAKE			0.63			0.63			10	14-62
SECOND POND						0.69			4	13-35
LINCOLN POND							0.54		7.7	5-40
ADIRONDACK LA	0.62							0.65	7	10-35
Middle Pond	0.58							0.54	3.4	10-100

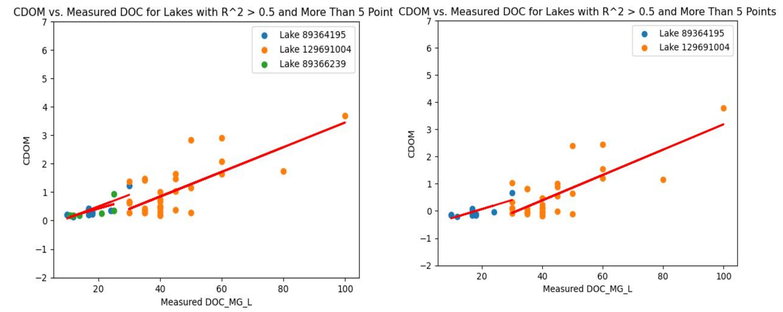

Figure 5. Lake predicted color graphs where the Green/NIR band outperforms the Blue/NIR band ratio with R^2 = 0.65 (Equation used: Martins et al.)

Methodology

- Implemented several **machine-learning** Boosting and Bagging techniques using different input features from surface reflectance:
  - **RandomForest**
  - **AdaBoost**
  - **XGBoost**
  - **GradientBoosting**
  - **SVR**
  - **MLPRegressor**
- Model utilizes 70% of data from each lake selected for training, 30% for testing performance
- To find the best performing model, we examined the impact of lake classification, atmospheric correction algorithms, and lake water depths on the model's performance

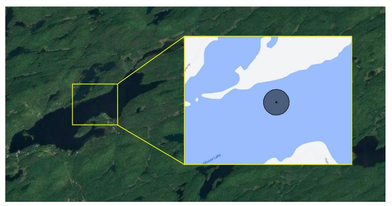

Figure 4. Improved satellite data processing by creating a 100-meter buffer around sampled lake centroids for greater accuracy.

Results (Random Forest)

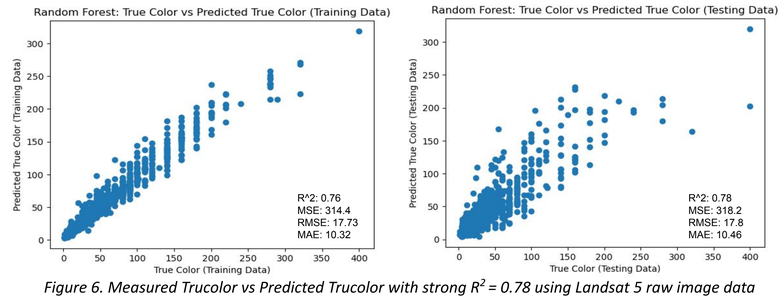

Figure 6. Measured Trucolor vs Predicted Trucolor with strong R^2 = 0.78 using Landsat 5 raw image data

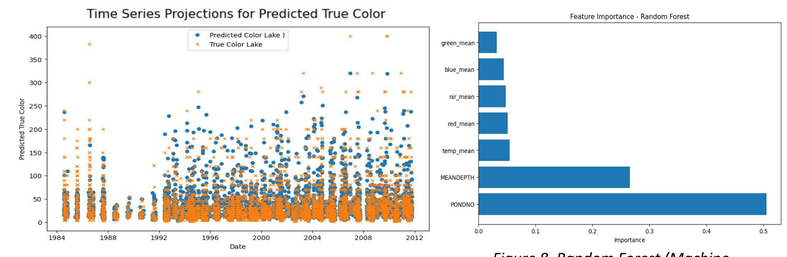

Figure 7. Time series of predicted and measured trucolor for all lakes depicts an increasing trend over time

Figure 8. Random Forest (Machine Learning) feature significance

Figure 9. Time series of predicted Lake Color depicts a gradual upward trend over time using both Landsat 5 & 7 for Big Moose Lake and Sunday Pond

Discussion

- The R^2 values generated by the Landsat 5 band ratios B2/B4 and B1/B4 indicate that they are suitable for lakes with low and high concentrations of CDOM.
- The band ratio, B2/B4, achieved the highest R^2 value of 0.69. Although this indicates a good correlation, we are expecting R^2 values above 0.80.
- According to our current findings, the selected band ratios work well for lakes with low maximum depth (of < 7 meters).
- The results show better correlation of the raw images than the Surface reflectance product (R^2 = 0.78 vs. 0.69).
- In the Random Forest model, lake depth and lake ID have the largest feature importance in comparison to Landsat 5 individual bands which could lead to statistical bias.

Conclusions

- Satellite based color estimation is capable of studying the long-term trends in water quality parameters that better explain the regional trends and help with selecting the sampling lakes that represent the Adirondack region
- A data gap exists beyond the studied lakes. Despite this limitation, extrapolation to other lakes indicates machine learning model scalability.
- Future efforts will involve the creation of CDOM maps across various lakes to provide a visual representation of CDOM distribution
- We also aim to strengthen our current models and utilize neural networks to advance our machine learning method.

References

1. Brezonik, Olmanson, L. G., Finlay, J. C., & Bauer, M. E. (2015). Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. Remote Sensing of Environment, 157, 199–215. https://doi.org/10.1016/j.rse.2014.04.033
2. Koll-Egyed, Cardille, J. A., & Deutsch, E. (2021). Multiple Images Improve Lake CDOM Estimation: Building Better Landsat 8 Empirical Algorithms across Southern Canada. Remote Sensing (Basel, Switzerland), 13(18), 3615–. https://doi.org/10.3390/rs13183615
3. Kutser, Pierson, D. C., Kallio, K. Y., Reinart, A., & Sobek, S. (2005). Mapping lake CDOM by satellite remote sensing. Remote Sensing of Environment, 94(4), 535–540. https://doi.org/10.1016/j.rse.2004.11.009
4. Olmanson, Brezonik, P. L., Finlay, J. C., & Bauer, M. E. (2016). Comparison of Landsat 8 and Landsat for regional measurements of CDOM and water clarity in lakes. Remote Sensing of Environment, 185, 119–128. https://doi.org/10.1016/j.rse.2016.01.007
5. Olmanson, Page, B. P., Finlay, J. C., Brezonik, P. L., Bauer, M. E., Griffin, C. G., & Hozalski, R. M. (2020). Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. The Science of the Total Environment, 724, 138141–138141. https://doi.org/10.1016/j.scitotenv.2020.138141
6. Zhang, Wang, S., Zhang, B., Zhang, F., Shen, Q., Wu, Y., Mei, Y., Qiu, R., & Li, J. (2022). Analysis of the water color transitional change in Qinghai Lake during the past 35 years observed from Landsat and MODIS. Journal of Hydrology. Regional Studies, 42, 101154–. https://doi.org/10.1016/j.ejrh.2022.101154

Acknowledgement

This research project was supported by NY Department of Environmental Conservation Grant #DEC01-C01714GG-3350000 (NYS-DEC) , NSF GRANT# AGS-2023174 (NSF-IUSE), and NSF Grant AGS-1950629 (REU). Special appreciation to the New York State SCALE (Survey of Climate Change and Adirondack Lake Ecosystems) project for sharing the in-situ data. The statements contained within the poster are not the opinions of the funding agency or the U.S. government but reflect the author's opinions.