Assessing and Forecasting Chlorophyll Abundances in Minnesota Lakes using Remote Sensing and Statistical Approaches

Benjamin Von Korff[1], Bryce Hoppie[1], Fei Yuan[1], Leif Olmanson[2], Benjamin Page[3] , Guarionex Salivia[1]

[1]Minnesota State University Mankato [2]University of Minnesota Twin Cities, [3]United States Geological Survey

Introduction

Harmful Algae Blooms (HAB)s can negatively impact water bodies by the reduction of aesthetic value, oxygen depletion, and for some algae (mainly cyanobacteria), toxin production (cyanotoxins). However, the MPCA lacks a routine cyanobacteria monitoring approach, relying on incident-based sampling and a message to lake users of "when in doubt, best keep out". In this study, three approaches - 1) Unmanned Aerial Vehicle (UAV) remote sensing 2) time series forecasting 3) analysis of meteorological impacting blooms were explored for addressing the HAB monitoring and forecasting needs in Minnesota.

Shiraishi et al. [2] developed a new chlorophyll prediction index using RGB values called Visible Water Residence Index (VWRI) :

$$\frac{\mathrm{GREEN} - \mathrm{BLUE} - \mathrm{RED}}{\mathrm{GREEN} + \mathrm{BLUE} + \mathrm{RED}}$$

VWRI has been studied in the laboratory, has not been tested in the field until this study.

One common technique for monitoring HABs is using time series forecasting of high frequency chlorophyll data. However, multi-parameter sondes are expensive ($10,000+). Xiao et al. [2] developed a Wavelet-coupled Artificial Neural Network (WANN) which allowed for chlorophyll forecasting based on a single parameter (chlorophyll-a). We compared the accuracy of a univariate time series approach to a multivariate chlorophyll-a forecast to recommend the lowest possible monitoring setup for chlorophyll-a forecasting in Minnesota lakes.

Southern Minnesota observed a relatively low algal bloom frequency in 2019, and anecdotally this appeared to coincide with a wet spring and summer. However, the impact of meteorological factors on the algae blooms in Minnesota lakes requires further study.

The following research questions were examined:

1) How does the accuracy for UAV algae mapping compare between VWRI and normalized difference vegetation index (NDVI)?

2) Can a univariate time series forecast perform as well as a multivariate forecast?

3) Can meteorological factors explain interannual variability in chlorophyll during July/August in Bass Lake, Faribault County, MN?

4) Is there a correlation between precipitation and temperature and chlorophyll-a in lakes across the Western Cornbelt Plains?

Methods

A weather station and Hydrolab DS5 sonde was installed at Bass Lake in Faribault County, MN from May-October 2019 and 2020. Four flights were conducted at Bass Lake during the fall of 2020 using a Phantom 4 drone (DJI) with a Sentera NDVI Single Sensor modification. Drone imagery was stitched together using Drone Deploy cloud software. Chlorophyll was measured at between 7-25 data points within the flight area using a YSI series 6 sonde equipped with a chlorophyll-a probe. A buffer was created around sampling points to reflect potential error in the GPS location. Next, regression between chlorophyll-a and the mean VWRI or NDVI in the buffer areas was performed in R.

Univariate time series forecasting of chlorophyll-a from the deployed sonde was performed using autoregressive integrated moving average (ARIMA) in R, using 67% of the dataset for training, and 33% for testing. Root mean squared error (RMSE) was used to measure model performance. Univariate and multivariate forecasting was also performed using Long Short Term Memory (LSTM) with Python. Multiple regression was performed in R to examine the relationship between chlorophyll-a and other chemical and physical water quality or meteorological parameters to select predictors for the multivariate forecast. LSTM was performed both as a univariate forecast, and multivariate forecast.

Median chlorophyll-a during July/August 2017-2020 at Bass Lake were downloaded from the University of Minnesota Lake Browser. Precipitation, temperature, and snowfall data was downloaded from the National Climate Data Center (NCDC), using either Global Historical Climatology Network (GHCN) or Community Collaborative Rain, Hail, and Snow network (CoCoRaHS). Data was downloaded from stations near approximately 160 lakes in the Western Corn Belt Plain (WCBP) region of Minnesota. A multiple regression was performed to examine the impact of meteorological factors on chlorophyll-a broadly across the WCBP region. Precipitation factors included AN (annual precipitation), JJA (June-July-August precipitation), MAM (March-April-May precipitation), annual snowfall, JJAT (June-July-August average temperature), MAMT (March-April-May average temperature), and the natural log of the precipitation factors. Separately, multiple regression was performed for Bass Lake to examine the impacts of precipitation and temperature on inter-annual variability in chlorophyll-a at a local scale to Bass Lake.

Results

On 9/14/2020, both NDVI (Fig. 1) and VWRI (Fig. 2) had strong, positive relationships with chlorophyll-a (NDVI: $R_{adj}^2 = 0.986$, p = 4.12e-11, VWRI: $R_{adj}^2 = 0.866$, p = 3.07e-07). On 9/18/2020, results were similar but VWRI had the strongest relationship (VWRI: $R_{adj}^2 = 0.799$, p = 0.000296, $R_{adj}^2 = 0.665$, p = 0.00774). On 9/26/2020, NDVI ($R_{adj}^2 = 0.785$, p = 3.06e-06) showed a stronger relationship than VWRI ($R_{adj}^2 = 0.467$, p = 0.00212). On 10/8/2020, both NDVI and VWRI were positively correlated to chlorophyll-a, with VWRI ($R_{adj}^2 = 0.797$, p = 0.00427) showing a stronger relationship than NDVI ($R_{adj}^2 = 0.631$, p = 0.0203). In general, the VWRI maps appeared to show much more variability in algal abundance and highlighted patches of algae poorer than the NDVI maps (Fig. 3). After combining the data from all 4 flights on Bass Lake, NDVI still showed a positive correlation with chlorophyll-a ($R_{adj}^2 = 0.797$, p = 2.98e-12) when fit with a 3rd order polynomial, while VWRI did not have a significant correlation with chlorophyll-a ($R_{adj}^2 = 0.027$, p = 0.130).

The simplest regression model was predicting chlorophyll-a using turbidity (p = 0.000143, negative correlation), specific conductivity (p = 2.90e-06, negative correlation), water temperature (p = 0.000756, positive correlation), the past 60 days (2 months) of precipitation (p = 6.02e-05, positive correlation), with a R_{adj}^2 of 0.343 (p = 1.36e-12). The RMSE for the 2019 and 2020 multivariate forecasts (Fig. 4) was 9.702 and 2.791, respectively. The univariate LSTM had a lower RMSE in both 2019 and 2020: 1.197 and 1.767, respectively. ARIMA had the lowest RMSE of the three model approaches, with a RMSE of 1.160 in 2019, and 0.936 in 2020. The 2020 10-day ARIMA forecast is from earlier in the season than the LSTM forecasts.

After stepwise deletion of non-significant variables, the 2019 chlorophyll-a model for Western Corn Belt Plains lakes was left with 2 variables (Fig. 5), AN19 and JJA19, both of which become non-significant after stepwise deletion (p > 0.05 for each variable). However, for the 2020 model, MAM20 (p = 0.0192 and JJA20 (p = 0.0014), and both negatively correlated to median August chlorophyll-a. The overall p value for the 2020 model was p = 0.00596, however, the R_{adj}^2 was extremely low (0.0519).

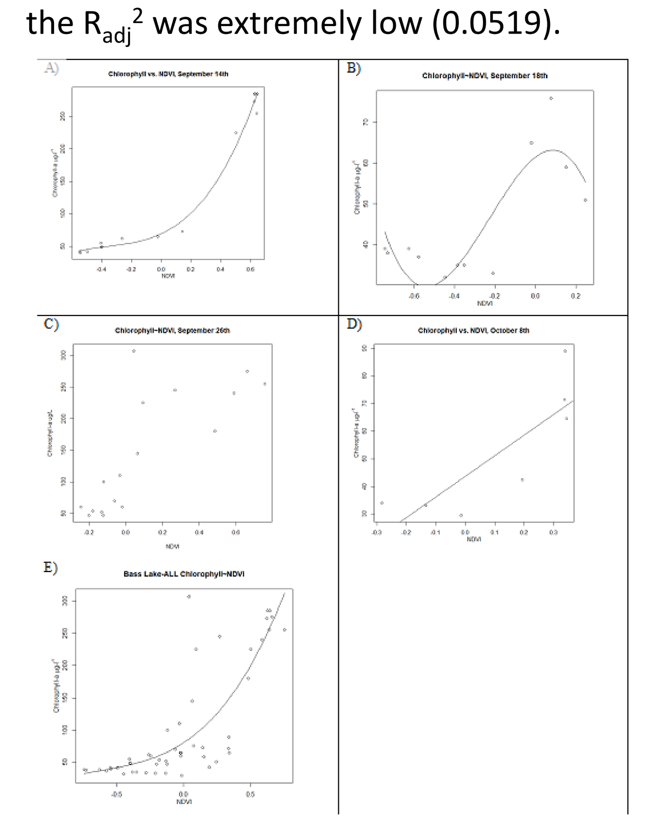

Figure 1. NDVI vs. Chlorophyll-a at Bass Lake in Faribault County, MN on A) 9/14/2020, B) 9/18/2020, C) 9/26/2020, D) 10/8/2020 and E) Combining all data.

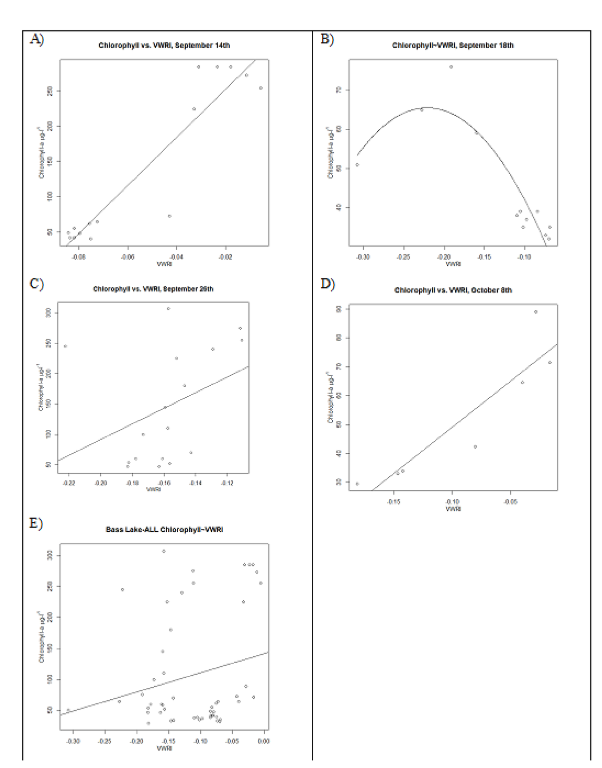

Figure 2. VWRI vs. Chlorophyll-a at Bass Lake in Faribault County, MN on A) 9/14/2020, B) 9/18/2020, C) 9/26/2020, D) 10/8/2020 and E) Combining all data.

A multiple regression of maximum temperature, and 1-week, 2-week, 30-day, and 60-day precipitation totals for predicting median August chlorophyll-a at Bass Lake showed 2 variables to be significant: 2 week precipitation totals (positive correlation, p = 0.0158), and 60 day precipitation totals (negative correlation, p = 0.0301).

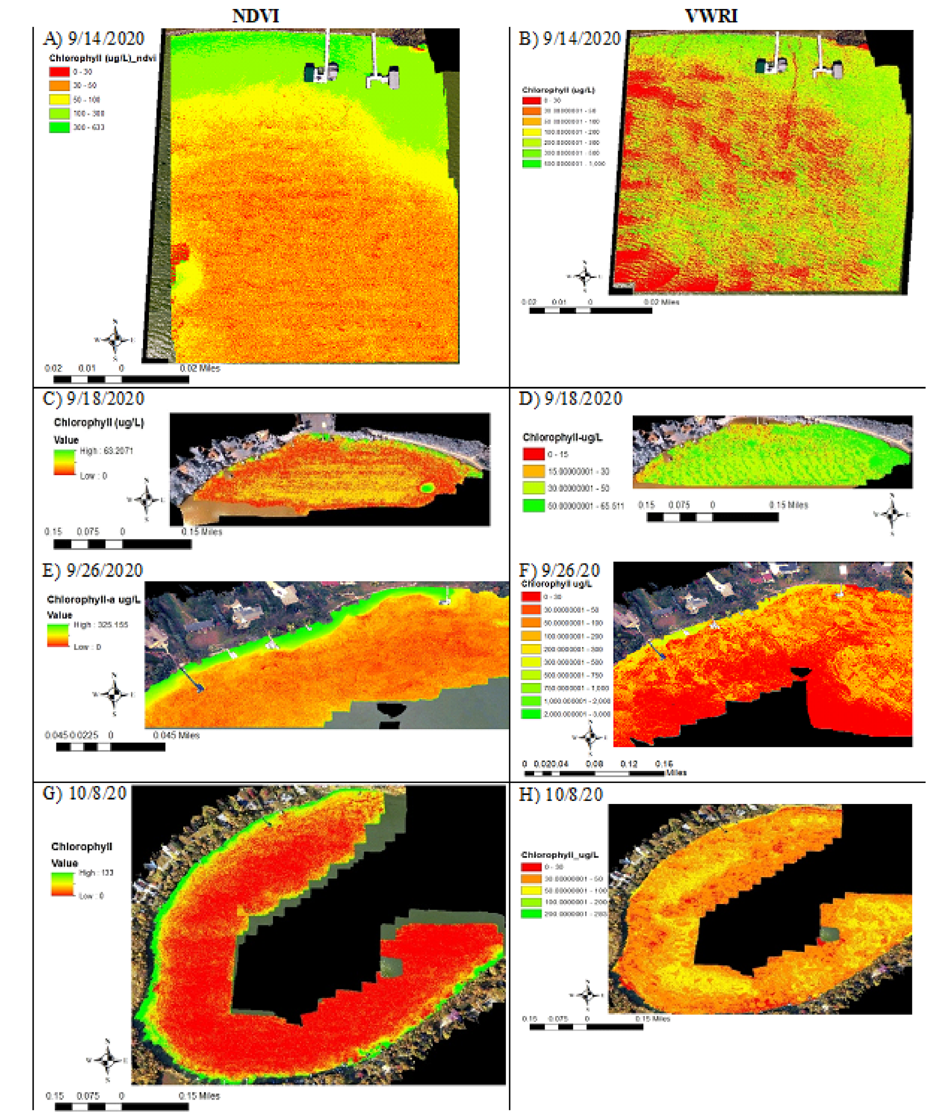

Figure 3. Model of chlorophyll-a at Bass Lake on 9/14/2020 using A) NDVI and B) VWRI; on 9/18/2020 using C) NDVI and D) VWRI; on 9/26/2020 using E) NDVI and F) VWRI; and on 10/8/2020 using G) NDVI and H) VWRI.

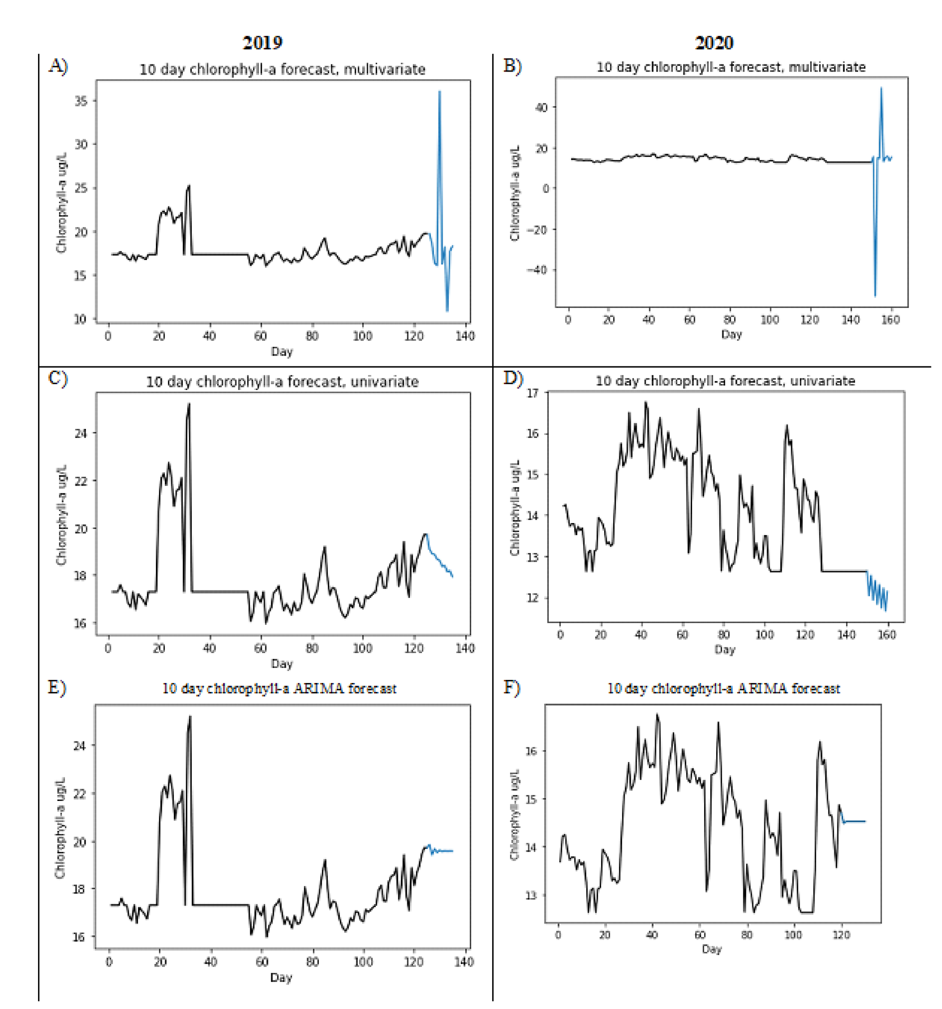

Figure 4. 10-day LSTM multivariate time series forecasts of chlorophyll-a in A) 2019 and B) 2020. Univariate LSTM forecasts are shown in C) 2019 and D) 2020. ARIMA forecasts are shown in E) 2019 and F) 2020.

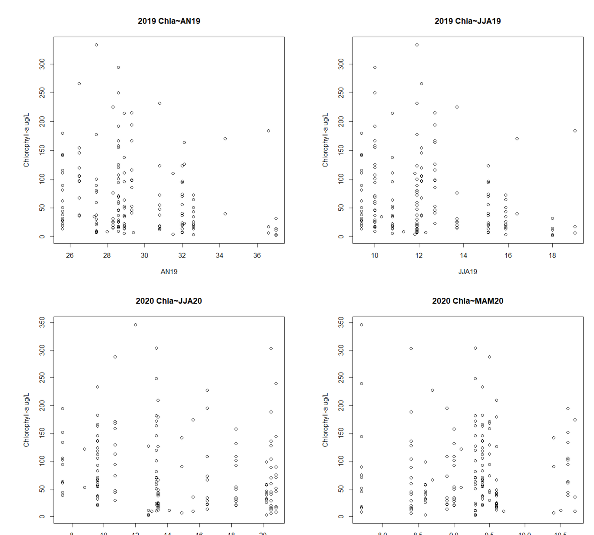

Figure 5. Precipitation vs. chlorophyll-a in 160 WCBP lakes. Chlorophyll-a in 2019 and 2020 vs. annual precipitation in 2019 (AN19), June-July-August precipitation in 2019 (JJA19), June-July-August precipitation in 2020 (JJA20), March-April-May precipitation in 2020 (MAM20). Other precipitation and temperature variables were removed via stepwise deletion. Only the 2020 model was significant (R2= 0.0519, P = 0.00596).

Discussion/Conclusions

Our results suggest that NDVI outperformed VWRI. However, VWRI appeared to be a viable method for chlorophyll mapping in Bass Lake. Advantages of NDVI included providing less patchy maps of algal blooms and highlighting algae better, generally increased and more reliable model performance. In addition, the strong correlation between NDVI and chlorophyll-a when data were combined across multiple flight dates suggests that it may be possible to apply a single regression equation across multiple flights using NDVI, which could reduce time needed in the field to create algal bloom maps, and reduce laboratory costs for chlorophyll-a sampling. Whereas, if VWRI is used, it is recommended that sampling is performed every flight, and a new regression relationship used for each flight.

The results from this study confirm previous suggestions from Xiao et al. [13] that univariate forecasts of chlorophyll-a can be performed successfully at a much lower cost than a multivariate forecasts. While ARIMA showed low forecasting error, ARIMA was problematic due to producing relatively flat forecasts. Our results show that in Bass Lake, our univariate LSTM forecast outperformed the multivariate forecast, both in RMSE and in reliability of the forecast. Since a YSI or hydrolab multiparameter sonde ranges in cost from $10,000+, the results from this study suggest a single parameter probe (chlorophyll-a) should be used for monitoring/forecasting chlorophyll in Minnesota lakes where excessive algae biomass and HABs are the primary concern.

These results suggest that chlorophyll-a may be positively correlated with 2-week precipitation totals in Bass Lake, and negatively correlated with 60-day precipitation totals. If low residence times due to dry conditions cause algae to develop, algae could become nutrient starved later in the summer, and a short pulse of nutrients could potentially allow a bloom to continue and develop further. The lack of strong trends in the relationship between chlorophyll-a and meteorological patterns for WCBP lakes suggests that a broad scale analysis may not be feasible, and that other confounding factors are present such as watershed area, land use, trophic state, slope, or presence of point sources.

Acknowledgements

Thanks to The Hoppie Lab and the Water Resources Center at Minnesota State University, Mankato, for collaboration and equipment sharing. This project received funding assistance from a Graduate Research Fellowship at Minnesota State University Mankato. Thanks to the Bass Lake Association for assisting with finding a dock for the monitoring station.

References

[1] Haruhiro Shiraishi. 2018. New Index for Estimation of Chlorophyll-a Concentration in Water with RGB Value. *International Journal of Engineering & Technology* 18, 6, 10-16.

[2] Xi Xiao, Junyu He, Haomin Huang, Todd R. Miller, George Christakos, Elke S. Reichwaldt, Anas Ghadouani, Shengpan Lin, Xinhua Xu, and Jiyan Shi. 2017. A novel single-parameter approach for forecasting algal blooms. *Water research* 108 (January 2017), 222-231