2021

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

Ben Von Korff
Minnesota State University, Mankato

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

By

Ben Von Korff

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science at Minnesota State University, Mankato

Minnesota State University, Mankato

Mankato, Minnesota

July 2021

7/9/2021

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

Ben Von Korff

This thesis has been examined and approved by the following members of the student's committee.

Guarionex Salivia
Advisor

Bryce Hoppie
Committee Member

Fei Yuan
Committee Member

Acknowledgements

I would like to thank my advisor Dr. Guarionex Salivia and committee Dr. Bryce Hoppie and Dr. Fei Yuan for their guidance. This work would not have been possible without Bryce Hoppie and Owen Lott who collaborated on setting up trail camera, sonde, and weather station equipment, equipment sharing, and in water sampling and drone flights. Thank you to the Bass Lake Association for providing site access for drone flights, and a dock for the weather station. Thank you to The Water Resources Center at Minnesota State University Mankato and staff - Kendra Daby, Tyler Grupa, Kimberly Musser, Diane Wiley, for their support and assistance. Thank you to Dr. Mike Waters at Northern Kentucky University for running our trail camera imagery through their blue-green algae classifier. Thank you to Dr. Deon Van Der Merwe for discussing and offering advice on drone-algae mapping methods. Thank you to the MNSU graduate research grant program for providing financial support.

I would also like to thank my fiancée, Siyu Yang for her love, support, patience, encouragement, and help in the field (including blowing up the new inflatable kayak). Thank you to my parents Connie and Jerry Von Korff and brothers Michael and Joshua Von Korff for their love, support, and assistance with proofreading and watching practice thesis defenses.

List of Abbreviations

HAB	Harmful Algal Bloom
UAV	Unmanned Aerial Vehicle
VWRI	Visible Water Residence Index
NDVI	Normalized Difference Vegetation Index
BNDVI	Blue Normalized Difference Vegetation Index
NDCI	Normalized Difference Chlorophyll Index
RMSE	Root Mean Squared Error
BR	Band Ratio Algorithm
ARIMA	Autoregressive Integrated Moving Average
LSTM	Long Short Term Memory
WCBP	Western Corn Belt Plains
RGB	Red-Green-Blue

1 Table of Contents

Table of Figures

Table of Equations

Table of Tables

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

Ben Von Korff

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science at Minnesota State University, Mankato

Minnesota State University, Mankato
Mankato, Minnesota
July 2021

Abstract

Harmful algal blooms (HABs) can negatively impact water quality, lake aesthetics, and can harm human and animal health. However, monitoring for HABs is rare in Minnesota. Detecting blooms which can vary spatially and may only be present briefly is challenging, so expanding monitoring in Minnesota would require the use of new and cost efficient technologies. Unmanned aerial vehicles (UAVs) were used for bloom mapping using RGB and near-infrared imagery. Real time monitoring was conducted in Bass Lake, in Faribault County, MN using trail cameras. Time series forecasting was conducted with high frequency chlorophyll-a data from a water quality sonde. Normalized Difference Vegetation Index (NDVI) was generally well correlated to chlorophyll-a measured by a sonde (R^2 = 0.678 for all data from 5 flights, between 0.323-0.986 for individual flights), while Visible Water Residence Index (VWRI) showed a weaker and less consistent correlation with chlorophyll-a (R^2 = 0.027 for all data from 5 flights, between 0.17-0.866 for individual flights). While RGB cameras (trail cameras or UAVs) were useful for visual inspection and spotting blooms, these results suggest that quantitative remote sensing of chlorophyll in Minnesota Lakes should use near-infrared cameras at a minimum. Univariate time series forecasts using sonde chlorophyll-a data were compared using autoregressive integrated moving average (ARIMA) and machine learning techniques (LSTM, wavelet-LSTM). Chlorophyll-a was positively correlated to temperature and precipitation, while negatively correlated to conductivity and turbidity. Peak summer chlorophyll concentrations also appeared to be positively correlated to recent precipitation totals. 10-day chlorophyll-a forecasts using univariate LSTM and ARIMA outperformed a multivariate forecast (using conductivity, turbidity, temperature, and precipitation as predictors), suggesting that lower cost monitoring setups (a single chlorophyll probe) may be practical. To assist in understanding meteorological factors impacting interannual variability of blooms in Bass Lake, the relationship between peak summer chlorophyll-a (from Sentinel-2 satellite imagery) and temperature and precipitation were analyzed at Bass Lake. The impact of meteorological factors on patterns in chlorophyll-a for lakes in the Western Corn Belt Plains (WCBP) was also

examined, using Sentinel-2 imagery (imagery was available for 160 lakes in the WCBP during 2019 and 2020). Peak summer Chlorophyll-a at Bass Lake was positively correlated to 2-week precipitation totals, suggesting a potential role of precipitation induced nutrient loading in initiating blooms; a negative correlation between peak chlorophyll-a and 60-day precipitation totals also suggested that increased residence time during drier periods may be a driving factor as well. While a slight negative correlation between precipitation and peak summer chlorophyll-a was present in a larger scale analysis of 160 WCBP lakes, too many confounding factors were present to show the impact of precipitation on blooms in the larger scale analysis.

2 Introduction

2.1 Harmful Algal Blooms - Background of the Problem

Harmful algal blooms (HABs) are any species of algae which can harm humans or wildlife, whether by mechanical means, toxin production, or oxygen depletion. HABs can negatively impact water bodies by the reduction of aesthetic value, oxygen depletion, and for some algae (mainly cyanobacteria), toxin production (cyanotoxins) that can harm people and animals through exposure due to recreational use, drinking water contamination, or crop irrigation [62, 9]. Negative impacts of HABs on human health can range from minor problems such as the development of skin irritation, to more serious problems such as gastrointestinal, respiratory, or neurological symptoms [36]. For example, a recent drinking water contamination occurred in Lake Erie in 2014, closing the Toledo water supply for three days, and impacting over 400,000 people [31, 85]. There is some evidence that long term exposure to cyanotoxins may be linked to amyotrophic lateral sclerosis (ALS, Lou Gehrig's Disease) [8, 14, 26]. Dogs can become sick or die after drinking water or algal scums while swimming. Backer et al. [3] found over 300 documented cases of dog cyanotoxin poisoning in the past decade, although their review only included a small subset of likely cases. During 2016-2018 alone Roberts et al. [70] reported 389 and 413 cases of human and animal sickness, respectively, in 18 states.

HABs frequencies have been increasing [64], a trend which is expected to continue due to global climate change [31, 32]. Climate change has caused an increase in extreme precipitation events and droughts, which create an optimal environment for bloom formation [65]. Nutrient loading of phosphorus and nitrogen can stimulate algal blooms [19, 21]. An increase in anthropogenic nutrient loading to freshwater and coastal environments in the past 150 years has caused eutrophication, defined as the increase in supply of organic matter to a water body [81]. Effective management of HABs in the face of climate change will require a combination of nutrient management strategies (best management practices, phosphorus trapping, dredging), algae control strategies (such as use of algaecides); monitoring is also critical to both understand factors controlling bloom and toxin production as well as to protect lake users from toxin exposure [21, 65].

2.2 State of HAB Monitoring and Needs in Minnesota

There is a lack of established HAB and cyanotoxin monitoring programs across all states, tribes, and territories [9]. Brooks et al. [9] noted that while swimming beaches are commonly monitored for bacteria, they are not usually monitored for HABs or cyanotoxins, and that a lack of state and federal funding for HAB monitoring and research poses a barrier to effective HAB management. Brooks et al. [9] argued that there is a critical need for the global implementation and expansion of HAB monitoring efforts and technologies [9, 11]. Insufficient funding and a focus of federal funding towards coastal areas instead of inland waters, a lack of HAB toxin water quality standards, and a lack of standard methods for toxin measurement have all contributed to insufficient HAB

monitoring [9, 11]. Brooks et al. [9] also notes that because HABs are naturally occurring and impossible to prevent completely, HAB forecasting is a critical component of management. Coastal and marine HAB forecasting is listed as a major focus of National Oceanic and Atmospheric Administration [9, 11]. Spatial variability in blooms is a major challenge, and monitoring programs which use remote sensing may are important to address this challenge [21].

HAB monitoring in Minnesota often reactionary and limited to grab sampling; sampling following a dog's death is referred to as incident-based sampling [35]. Results may take days to be finalized due to time require to transport samples, conduct laboratory analyses, and analyze results [79], by which time negative impacts such as dog deaths could have occurred. The Minnesota Pollution Control Agency (MPCA) has carried out monthly monitoring of toxins in 12 targeted lakes in 2006 and 50 stratified-randomly selected lakes in 2007 in Minnesota [49]. An additional 87 lakes were added to the toxin dataset in 2012 [35]. However, Heiskary et al. [35] noted that sampling was insufficient to showing the range of toxin concentrations and relative risk of high toxin concentrations for individual lakes. In these studies, near-shore sampling was limited to a randomly selected site at only a subset of lakes. A complete sampling regime would have required more frequent sampling throughout the bloom season as well as targeted sampling in downwind, scum-rich areas. The MPCA lacks a routine cyanobacteria monitoring approach in Minnesota lakes, rather relying on lake users to recognize the presence of cyanobacteria blooms, and a message to lake users of "when in doubt, best keep out" [35].

Despite the history of substantial HAB monitoring relative to many states, Minnesota still perfectly fits the description of Brooks et al. [9] and Brooks et al. [11] as lacking a routine and established cyanobacteria monitoring program. Christensen et al. [20] carried out intensive monitoring in a reoccurring cyanobacteria bloom in Kabetogama Lake, in Voyagers National Park, and found the presence of cyanotoxins even prior to visible bloom formation. Intensive monitoring similar to Christensen et al. [20] seems critical in more Minnesota Lakes, especially in locations where frequent and reoccurring cyanobacteria blooms are of concern to lake users, or in areas where lakes are used as drinking water supplies. However, given national funding limitations for HAB monitoring and the lack of an existing monitoring program in Minnesota, developing cost effective monitoring strategies will be critical in order to increase the likelihood of increasing HAB monitoring in the state.

2.3 Three Approaches to Address HAB Monitoring Needs in Minnesota

In this study, three approaches were explored for addressing the HAB monitoring needs in Minnesota, with a goal of developing lower cost monitoring approaches and to improve our ability to forecast and predict when HABs will occur in Minnesota lakes. These approaches are detailed here, and followed by a literature review/background information related to each approach:

1) The use and cost-effectiveness of relatively recent remote sensing technologies were tested for the use of HAB detection, including trail cameras and unmanned aerial vehicles (UAVs). A novel, recently developed Red-Green-Blue (RGB) band ratio

algorithm (BR) was tested for monitoring and mapping algae called visible water residence index (VWRI). VWRI has been studied in the laboratory, but to our knowledge, has not been tested in the field prior to this study. VWRI was tested using three remote sensing devices – satellites (Sentinel-2 satellite), trail cameras (Spypoint Link Evo Verizon), and UAVs (Phantom-4) by comparing the effectiveness of VWRI to the more commonly used multispectral method, normalized difference vegetation index (NDVI). Our research intent is to determine if lower cost RGB devices (such as the Phantom-4's RGB camera) can effectively map algal blooms using VWRI, reducing the monitoring cost compared to high cost drones equipped with multi or hyperspectral cameras.

2) Time series forecasting of algal abundance: one common technique for monitoring HABs is to monitor chlorophyll-a continuously using a high frequency probe (typically a multi-parameter sonde attached to a buoy or dock), and to perform time series forecasting to predict algal abundance in the future. However, multi-parameter sondes are extremely expensive ($10,000+). The effectiveness of a multivariate chlorophyll-a forecast was compared to a univariate forecast, in order to make recommendations for the lowest possible cost data buoy/sonde setup for chlorophyll-a forecasting in Minnesota lakes.

3) The impact of meteorological factors on algal blooms were examined to assist in predicting algal blooms at a broader timescales in Minnesota lakes. In each of these three approaches, the study area was focused in Bass Lake, Faribault County, MN, although a second study site was also used during UAV testing (Little Rock Lake, Benton County, MN). Meteorological factors controlling interannual variability of algal blooms

were examined locally at Bass Lake, and in addition, correlations between meteorological factors and chlorophyll-a were examined more broadly across the Western Corn Belt Plains (WCBP) area of Minnesota. In addition, our testing of VWRI using Sentinel-2 imagery targeted lakes in the Twin Cities Metro area. A discussion of previous literature on HAB remote sensing, time series forecasting, and understanding meteorological factors controlling HABs, and a statement of key research questions follows.

2.4 Review of Background and Previous Literature for Three HAB Monitoring/Forecasting Approaches

2.4.1 1 - Remote Sensing of HABs- Background

Since algal blooms can occur sporadically and vary spatially in a lake, they require monitoring methods which address needs at both sufficient spatial and temporal scales [40, 42]. Algal blooms can range from occurring and disappearing rapidly over a period of a few days to lasting for months, therefore potentially requiring high frequency of monitoring to detect and track bloom dynamics [40, 74]. Grab samples are often insufficient to understand spatial variability of an algal bloom, and may be expensive and time consuming [13, 87]. Remote sensing technologies offer a low cost method to detect algal blooms over a large spatial scale, although the frequency, resolution, and spatial scale varies between methodologies. Remote sensing methods to detect algal blooms in inland waters generally falls into three categories: satellites, UAVs, and ground based sensors (GBS). These methods typically attempt to quantify either chlorophyll-a, a

measurement of algal biomass, or phycocyanin, a pigment specific to cyanobacteria (blue-green algae), by developing a model or correlation between measurements of the parameter of interest (chlorophyll or phycocyanin) and various BRs. Grab sampling or microscopy to confirm the presence of cyanobacteria species or toxin is still required for remote sensing monitoring [40].

2.4.1.1 Use of Satellites in HAB Remote Sensing

Sentinel-2 is a high resolution satellite which was launched in 2015 as part of the European Union's Copernicus program. A second Sentinel-2 satellite was launched in 2017; these two satellites are referred to as sentinel-2A and sentinel-2B, respectively. In addition to Sentinel-2 [66], satellites commonly used for mapping algal blooms include Sentinel-3 (3A launched in 2016 and 3B in 2018), Landsat-8, MERIS/OLCI, MODIS, and WorldView. Although WorldView 2/3 offers high spatial resolution, the imagery is not free. Sentinel-2 and Landsat-8 offers higher spatial resolution (20m for Sentinel-2 and 30m for Landsat-8) than MERIS, MODIS, or Sentinel-3 (250m for MODIS and 300m for MERIS). When mapping smaller or medium inland lakes, the higher resolution from Sentinel-2 or Landsat-8 may be required, while MERIS/MODIS/Sentinel-3 are generally more appropriate for mapping algal blooms in ocean environments. However, Landsat-8 is limited by a revisit frequency of 16 days. After the launch of Sentinel-2B in 2017, Sentinel-2A and 2B has a combined revisit frequency of 5 days. Satellite imagery from

Sentinel-2 and Landsat-8 are free. These satellite approaches are further limited by lack of image availability when clouds obscure the body of water of interest [98].

2.4.1.2 Remote Sensing Band Ratio Algorithms for HABs

Beck et al. [6] reviewed the accuracy of 12 different BRs, and found that normalized difference chlorophyll index (NDCI) [58] was generally one of the most accurate BRs across most satellites, where NDCI is:

Equation 1 - Normalized Difference Chlorophyll Index [58]:

$$\frac{B_{708} - B_{665}}{B_{708} + B_{665}}$$

NDCI is correlated to chlorophyll-a because chlorophyll-a has a reflectance peak near 700nm, and an absorption peak between 665-675nm. The BR uses narrow bands centered around the reflectance and absorption peaks, to prevent alteration of the reflectance spectra due to the impact of total suspended solids (TSS) and colored dissolved organic matter (CDOM) which occurs at lower wavelengths [58]. The denominator term in NDCI normalizes the difference between the reflectance and absorption peaks by the sum, which controls for uncertainties in the estimate of the reflectance, differences in solar azimuth, and atmospheric contributions at the wavelengths [58].

Phycocyanin is a photosynthetic pigment specific to cyanobacteria [7], and therefore is preferred for detecting cyanobacteria over chlorophyll when the species is unknown, since chlorophyll-a does not distinguish between green/brown/red algae and cyanobacteria. However, phycocyanin detection is less sensitive than chlorophyll-a via

remote sensing, and laboratory methodology for measuring phycocyanin are not as well standardized as chlorophyll-a, and therefore, when cyanobacteria is known to be present, chlorophyll is the preferred remote sensing metric for algal bloom mapping [76]. Phycocyanin BR require a narrow band to capture the absorption of phycocyanin, which occurs at 620nm [7]. Beck et al. [7] developed and compared phycocyanin detection algorithms: Sentinel-2 was found to have greater performance than Landsat-8 for phycocyanin mapping.

2.4.1.3 Use of UAVs in HAB Remote Sensing

Recent advances in UAV technology have caused researchers to examine their use in algal bloom mapping [40, 73, 91]. While spatial extent of algal bloom mapping via UAVs is limited compared to satellite imagery, UAVs offer advantages over satellites by being able to capture imagery even on cloudy days, typically providing higher resolution than satellite imagery, and potentially allowing for customization of camera band setup [40]. UAV technologies for algal bloom mapping can map on the scale of 10s of acres per flight for multi-rotor UAVs, or 100s of acres for fixed wing UAVs [87]. Because UAV mapping of algal blooms is a new field, standard methods are lacking [40]. Cameras used for UAV mapping range from red-green-blue (RGB), to broad band multispectral such as NDVI or blue normalized difference vegetation index (BNDVI) [87], to narrow-band multispectral and hyperspectral cameras [40]. Although hyperspectral cameras for UAV algae mapping are extremely expensive ($40,000+) they are useful for teasing out differences in spectral signatures between algal species [40].

UAV mapping setups at costs below $10,000 are generally limited to broad-band multispectral or RGB approaches. NDVI is a common BR for agriculture mapping, and the BR uses the following equation, where NIR is near infrared light, and red (ranging from approximately 600-700nm) is in the visible spectrum:

Equation 2 - Normalized Difference Vegetation Index [58]:

$$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$$

Van Der Merwe [87] found a correlation between BNDVI and blood packed cell volume, which can be used as a rapid estimate of algal biovolume in waters with low TSS and known presence of cyanobacteria. The BNDVI correlation was found to have a natural log based correlation due to saturation of BNDVI at high algae concentrations. BNDVI functions similar to NDVI except that it uses blue light (approximately 400-500nm) instead of red:

Equation 3 - Blue Normalized Difference Vegetation Index [87]:

$$\frac{\text{NIR} - \text{Blue}}{\text{NIR} + \text{Blue}}$$

2.4.1.4 Red-Green-Blue Band Ratio Algorithms

RGB BRs such as FLH-Violet have been used for chlorophyll mapping using satellites [6]. Li et al. [47] tested RGB algorithms (such as VARI) for measuring leaf area index (LAI). Recently, Shiraishi et al. [75] developed a new chlorophyll prediction index using RGB (red-green-blue light) values called Visible Water Resident Index, with green light occurring at approximately 500-600nm:

Equation 4 - Visible Water Residence Index [75]:

$$\frac{\text{Green} - \text{Blue} - \text{Red}}{\text{Green} + \text{Blue} + \text{Red}}$$

VWRI works by subtracting the blue and red light troughs from the green light peak in algae laden water, and the denominator normalizes the index. Shiraishi et al. [75] suggested that VWRI may allow for chlorophyll estimation using lower cost cameras such as smartphones and drones but their algorithm was not tested *in situ*.

One potential application of RGB devices includes GBS. Ground based remote sensing includes remote sensing devices such as a spectroradiometer, or even a smart phone. For example, Boddula et al. [10] developed a ground based spectroradiometer capable of monitoring algal blooms at the hyperspectral level at a single point, attached to a dock. One of the lowest cost ground based sensors which can be used for monitoring algal blooms is a cellular trail camera. The USGS has tested monitoring algal blooms using trail cameras by allowing for frequent visual inspection of a water body (Guy Foster, USGS, personal communication). However, it is unknown whether these cameras could provide quantitative estimates of algal abundance using RGB BRs such as VWRI. An ongoing study at Northern Kentucky University is developing a classifier to predict blue-green probability from trail camera imagery [90].

2.4.2 2 - Time Series Forecasting of HABs - Background

In addition to providing near-real time detection of HABs, continuous monitoring data from sondes can be used to forecast algal blooms. The intent of this study was to determine the fewest parameters to accurately forecast algal blooms to lower monitoring costs. Input parameters which are potentially important for chlorophyll-a forecasting

include nutrients (nitrogen and phosphorus), water temperature, pH, electrical conductivity, chlorophyll-a (measured in previous time steps), and meteorological data (wind speed, wind direction, air temperature, precipitation, barometric pressure, lighting conditions) [4, 15, 19]. In addition to current conditions, future weather forecasts (air temperature, wind speed/direction) can also be useful [77]. Which input parameters is important for forecasting can vary between water bodies or over time [95]. Appropriate input parameter selection is often carried out by stepwise testing, correlation analysis, or input minimization and maximization [45, 51, 95]. Liu et al. [50] found that nutrient parameters may not be required for short term forecasting [53].

Forecasting distance (the forecasting horizon) varies between studies, ranging from short term (1-3 days) [92], to longer term (up to 60 days) [24], but forecasting accuracy typically declines over time [32]. Du et al. [24] suggested that a multi-step NAR (non-linear autoregressive neural network) was capable of accurately forecasting chlorophyll-a with a horizon of 60 days.

A literature review of time series forecasting approaches was carried out to examine strengths and weaknesses of time series forecasting approaches for chlorophyll forecasting (Table 7 - Appendix). Comparing and contrasting approaches for specific water bodies is critical because what works in one region may not be as accurate in another [95]. Classical approaches (non-machine learning) to chlorophyll forecasting include autoregressive integrated moving average (ARIMA), multiple regression, exponential smoothing, and principal component analysis [17, 25, 51, 67]. Another non-machine learning alternative to forecasting chlorophyll-a is mathematical modeling [43,

71]. Some recent studies have found increased accuracy of classical approaches compared to machine learning approaches in general [55], however, for chlorophyll forecasting, machine learning forecasts generally have been found to outperform classical models [17, 67]. Chen et al. [16] found that ARIMA out-performed multiple linear regression forecasting, and Elhag et al. [25] found that exponential smoothing worked well but is only recommended for seasonal data (such as time series of monthly mean chlorophyll).

Neural networks are a popular technique in the literature for chlorophyll forecasting due to their non-linear nature [48, 50, 89, 97]. Artificial neural networks (ANN) are common for chlorophyll-a forecasting, and consist of a network of an input layer, hidden layer (1 or more layers), and an output layer [52, 92]. While most ANNs use correction with gradient descent called back propagation [54], self-organizing machine (SOM) uses competitive learning [52]. Recurrent neural networks (RNN) are unique in that they involve a loop in the hidden layer, and therefore work well when input data is sequential [44]; commonly used RNN's include Elman's RNN [88], long short term memory (LSTM) [17, 18], and non-linear autoregressive neural networks [22, 24]. Neural networks can also be combined with fuzzy logic, called co-active neuro fuzz inference systems [56]. Convolutional Neural Networks (CNN) are popular for time series forecasting, and work using kernel (a type of matrix) multiplication [5]. Besides this suite of neural network approaches, other major classes of machine learning approaches for chlorophyll forecasting include support vector machine regression (SVR) which use supervised learning to optimize the hyperplane of a regression [50, 51, 52, 61,

67], random forest, a supervised learning algorithm which uses decision trees [50, 53, 47], and genetic algorithms, which build models which are based off of the process of evolution [68, 94].

Of the diverse suite of machine learning approaches, artificial neural networks, CNN, and LSTM have been found to perform well for chlorophyll forecasting in general [5, 14, 18, 67, 92]. Random forest is a common forecasting approach, however, it is used for forecasting chlorophyll using a multivariate approach [93]. A combination of modeling techniques is often employed, and can increase accuracy over a single model. Examples in the literature include genetic algorithms coupled to SVR (GA-SVR), or ensemble models using the Bates-Granger approach [80]. Barzegar et al. [5] suggested that CNN and LSTM are best combined in a hybrid model to overcome limitations in the approach for forecasting low or high values.

Neural networks suffer from requiring high sample sizes and a potential for over-fitting, and work best with a high number of input parameters. However, time series data is often limited in sample size [26]. For example, Xiao et al. [92] used between approximately 100-300 daily averages of chlorophyll. However, Xiao et al. [92] developed a wavelet-coupled artificial neural network (WANN) which allowed for chlorophyll forecasting based on a single parameter (chlorophyll-a). Jeong et al. [44] found that an auto-regressive neural network model was possible using fewer input parameters than other empirical models. Another unique application of ANN was Tian et al. [86], which found that change in chlorophyll could be used as the ANN output instead of base chlorophyll to increase model accuracy. One weakness of machine learning

approaches for chlorophyll forecasting is that they do not perform well with non-stationary data, so data transformation is often required using a wavelet transform prior to model development [92].

2.4.3 3 - Understanding Meteorological Factors Controlling HABs

While time series forecasting can assist in predicting HABs at a short time scale, a good understanding of the environmental factors (hydrological, chemical, physical, biological) controlling algal blooms assists in predicting patterns in blooms. Southern Minnesota observed a relatively low algal bloom frequency in 2019 (including Bass Lake), and anecdotally this appeared to coincide with a wet spring and summer. Previous studies have found that meteorological factors such as precipitation, temperature, and wind can control the extent and timing of algal blooms. For example, Page et al. [59] found that a wet spring followed by a dry summer in Utah Lake likely triggered an algal bloom. One lake in Benton County, MN, Little Rock Lake, has been observed to have HAB problems in recent years, and DNR staff has observed that drier weather seems to prime the lake for algal blooms [41]. However, the impact of meteorological factors on the severity of an algal bloom season in Bass Lake, and other Minnesota lakes requires further study to confirm.

The potential impacts of precipitation on algal blooms is complicated in that algae can be limited by different nutrients in different bodies of water, and algal growth can be limited by other factors besides nutrient concentrations such as temperature, light, and residence time. While algae in freshwater environments such as Bass Lake are typically

limited by phosphorus (N:P ratios > 16:1), co-limition of both N+P (N:P ratios near 16:1) or even seasonal nitrogen limitation (N:P < 16:1) can occur [33, 82]. In addition, algal growth is usually enhanced by warmer temperatures [63], can be limited by the availability of light for photosynthesis [37], can be affected by water residence time due to algae requiring a stable environment to grow [1, 84, 34], and can be inversely related to discharge [1, 23, 28]. Zhang et al. [98] noted that wind is an important factor controlling blooms, with algal blooms more abundant in the direction opposite to the wind.

Since algal blooms may be limited by many factors, and the controlling factors of HABs may vary seasonally [98], explaining the precise relationship between meteorological factors such as precipitation and interannual variability in algal blooms is challenging. Page et al. [59] argued that the pattern of a wet spring followed by a dry summer may have resulted in high peak summer chlorophyll in Utah Lake because the wet spring supplied algae with the needed nutrients during the growing season, while the dry summer reduced the residence time, providing a stable environment in the lake for an algal bloom later in the summer when temperatures increased. Ho et al. [38] noted that the impact of meteorological factors on algal blooms are often understood locally, but are poorly understood at a broad scale, and have not been studied frequently. They examined the impact of temperature and precipitation on chlorophyll-a across 1,200 lakes in the United States. They found poor correlation between precipitation broadly across the U.S.; one potential reason is that precipitation can affect algal blooms in multiple ways: by

increasing nutrient runoff [57], by diluting nutrient concentrations [69] or by decreasing residence time [63].

2.5 Research Questions

The goals of this research is firstly to develop recommendations for proactive, low cost monitoring/forecasting using 2 approaches (remote sensing and forecasting) in Bass Lake, Faribault County, MN, and secondly to understand meteorological factors controlling HABs to make more broad scale interannual predictions for HABs in Minnesota Lakes. The following research questions were examined:

Remote Sensing

1) Can trail cameras be used to quantitatively estimate chlorophyll using RGB remote sensing equations?
2) Can trail cameras predict the likelihood of blue-green algae using existing blue-green-algae classifiers (Northern Kentucky University, Mike Waters)?
3) How does the accuracy for chlorophyll mapping compare between RGB and near-infrared BRs for satellite imagery (Sentinel-2), and for UAVs (Phantom 4 drone)?

Time Series Forecasting

4) Are classical approaches (ARIMA) or are machine learning approaches more accurate (LSTM)?

5) How far can time series forecasting accurately predict (forecasting horizon)?

6) Does data transformation (wavelet analysis) improve forecasting?

7) Which variables are most useful in a multivariate forecasting model in Bass Lake, Faribault County Minnesota?

8) Can a univariate forecast perform as well as a multivariate forecast?

Impact of Meteorological Factors

9) Can temperature, precipitation, or snowfall explain interannual variability in chlorophyll-a (measured by Sentinel-2 satellite) during July/August in Bass Lake?

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

3 Methods

Bass Lake is a shallow, 199 acre lake in Faribault County, MN, with a maximum depth of 20ft. It is part of the Le Sueur River Watershed, which is a major watershed in the Minnesota River Basin (Fig. 1). The lake has periodic algal blooms during the summer, which lake shore owners have expressed concern about. Bass Lake was chosen as the site to deploy water quality meters (sonde), weather station, and trail cameras (Fig. 1) due to these concerns about algae at the lake. In addition, Bass Lake was used as the main site for testing chlorophyll-a mapping using UAVs. Methods for 1) remote sensing, 2) time series forecasting, and 3) analyses of meteorological impacts on algal blooms are outlined in more detail below.

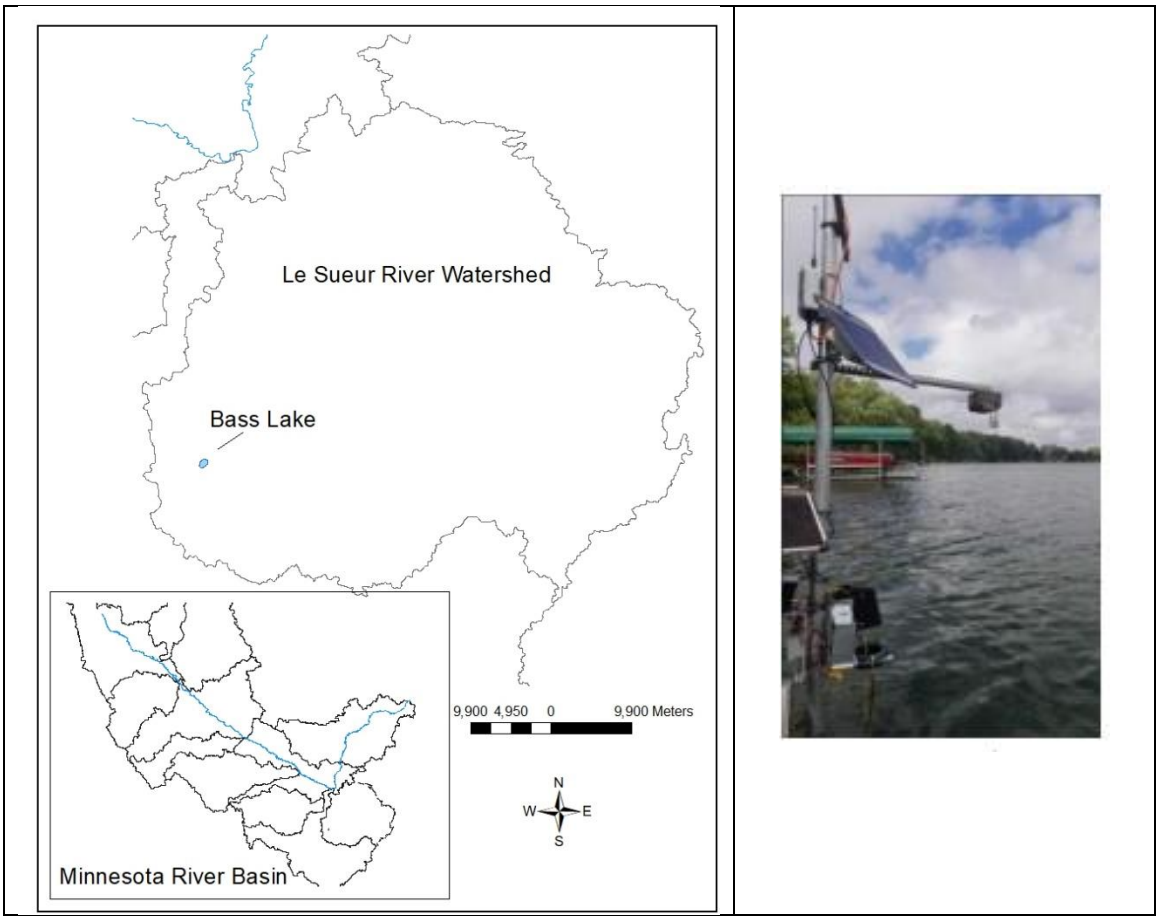

Figure 1. Left: Bass Lake, Faribault County, MN, and Right: monitoring station setup.

3.1 Remote Sensing Methods

3.1.1 Trail Cameras, Sonde, and Weather Station Setup at Bass Lake

Spypoint Link Evo Verizon cameras were installed along the western shore at Bass Lake in Faribault County, MN from May-October 2019 and 2020. The cameras were installed at the same location as a Hydrolab DS5 multiparameter sonde equipped with probes for specific conductivity, turbidity, water temperature, pH, dissolved oxygen, ORP (oxidation reduction potential) and chlorophyll-a, at a depth of approximately 1 meter below the surface. The sonde was replaced with a YSI Series 6 multiparameter

sonde when maintenance was required on the Hydrolab. A weather station measuring wind speed, wind direction, air temperature, barometric pressure, and total rainfall was installed at the same site. Sonde and weather station measurements were collected from May-October of 2019 and 2020.

The trail camera images at the station from 2019 were sent to Mike Waters at Northern Kentucky University, and run using HAB APP; HAB APP is a blue-green algae classifier developed by Mike Waters, which uses hue-saturation-value color histograms from RGB images to detect blue-green algae. HAB APP is a supervised learning classifier which has been trained to estimate the probability (from 0 to 1) that blue-green algae is present using RGB images (such as from trail cameras or smart phones). For each Bass Lake image, HAB APP used its machine learning model to estimate the probability of blue-green algae presence in the image.

VWRI from the trail camera was regressed against chlorophyll-a measured by the sonde to test for a correlation. Trail camera images were processed using python scripts to run geoprocessing in ArcMap (Raster Calculator). A regression of VWRI vs. chlorophyll was created by rounding image capture times to the nearest 15 minute interval to match Hydrolab readings. A small subset of images (< 100) were selected (using the Clip tool) which showed good image quality, using an area of interest away from the picture edge to examine the regression using the best images.

3.1.2 Sentinel-2 Satellite: comparing VWRI to Multispectral BR

The correlations between chlorophyll-a and both VWRI and NDCI were compared using Sentinel-2 Satellite imagery in Twin Cities Metro Area lakes. 34 lakes were selected which were sampled for chlorophyll-a by the Metropolitan Council on 5/15/2018 (Fig. 2, Appendix Table 8). NDCI can detect chlorophyll-a due to its reflectance peak in the near-infrared (NIR) range between 700-714nm and reflectance trough in the red region [58], and near-infrared indexes typically perform better than RGB indexes. While NDVI often includes broader band reflectance information from the red edge and red regions, NDCI used narrow bands in these regions. Sentinel-2's band 4 and 5 occur near the 665nm and 708nm locations used in the definition of NDCI [58]. Band 4 for Sentinel 2a and 2b occur at 664.9, and 664.6nm, respectively, while band 5 for Sentinel-2a and 2b occur at 704.1nm and 703.8nm, respectively. Previous studies of NDCI (using the Sentinel-2 bands near 705nm) have shown that NDCI is correlated to chlorophyll-a water sampling data in Minnesota Lakes [60]. Sentinel-2 imagery was downloaded from Earth Explorer and atmospherically corrected using Sen2COR in SNAP, ESA. 5-15-2018 was chosen because chlorophyll lab samples from 30+ lakes were obtained that day by the Metropolitan Council (the lab data was downloaded from Environmental Information Management System- EIMS). Sentinel 2 imagery was mosaicked together using ArcMap 10.8, and VWRI values were extracted (Extract Multi Values from Points) and matched to chlorophyll-a lab results. A similar approach was used to measure NDCI.

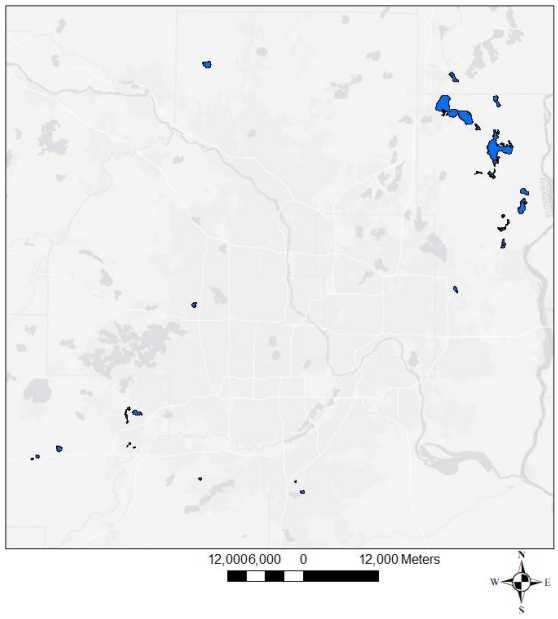

Figure 2. Map of Twin Cities Area, highlighting in blue the lakes monitored on 5/15/2018 by the Metropolitan Council.

3.1.3 UAV Flights with Sonde Chlorophyll-a Measurements, and Image Stitching

Four UAV flights were conducted at Bass Lake in Faribault County, MN during the fall of 2020. One flight was carried out at Little Rock Lake, in Benton County, MN. Little Rock Lake is a 1,311 acre lake, and was chosen due to reports of algal blooms on the lake. Little Rock is part of the Mississippi River- Sartell Watershed (Fig. 3).

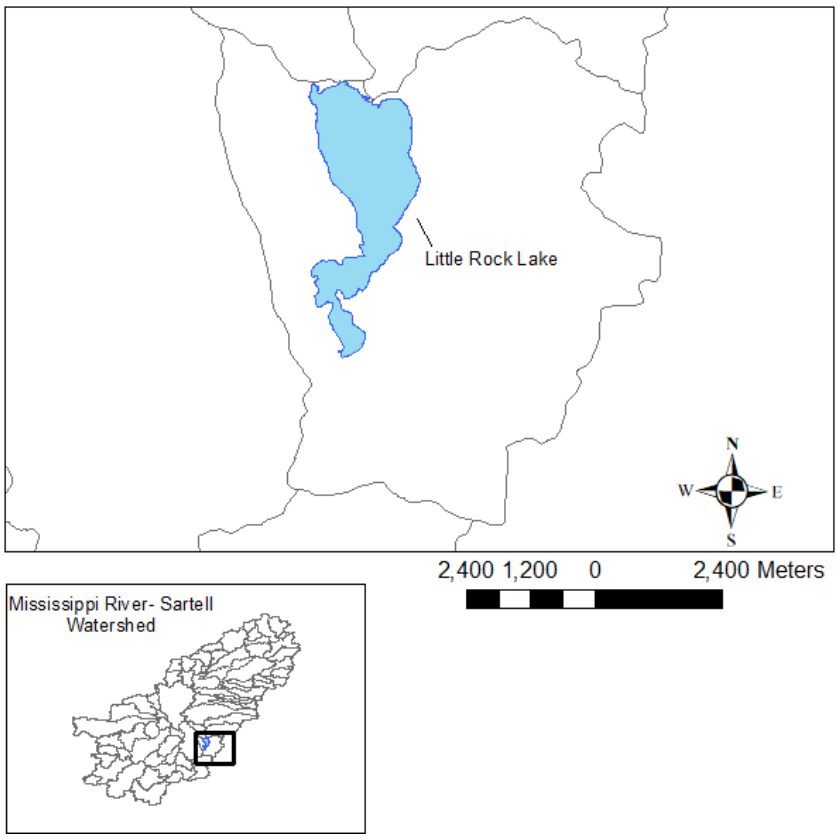

Figure 3. Little Rock Lake, Benton County, MN, part of the Mississippi River- Sartell Watershed Area.

A Phantom 4 drone (DJI) was used, with a Sentera NDVI Single Sensor modification. The Sentera modification attaches a second camera to the drone which measures red and NIR light. The drone flew at a height which varied from 150ft to 400ft depending on the wind conditions. Imagery was captured at a flight speed of 15mph with 75-80% overlap between images [87]. Drone Deploy cloud software was used to stitch images together. NDVI was calculated using an adjusted equation to account for differences in source radiation between points, where B3 and B1 are digital numbers in channels 3 and 1, respectively:

Equation 5 - Sentera Camera Normalized Difference Vegetation Index:

$$\frac{1.236 * \mathrm{B3} - 0.188 * \mathrm{B1}}{1.000 * \mathrm{B3} + 0.044 * \mathrm{B1}}$$

The coefficients for B3 and B1 are required for 2 reasons. First, B1 measures mainly red light but some near infrared (NIR) light, and B3 measures mainly NIR, but some red light. This equation removes the NIR from B3, and removes red from B1 to isolate the red band, and accounts for unequal irradiance in the red and NIR bands [72].

VWRI was calculated using bands from Phantom 4's RGB camera (CMOS) camera using B1 (red), B2 (green), B3 (blue):

Equation 6 - Visible Water Residence Index band math [75]:

$$\frac{\mathrm{B2} - \mathrm{B3} - \mathrm{B1}}{\mathrm{B2} + \mathrm{B3} + \mathrm{B1}}$$

Chlorophyll was measured at between roughly 7-25 data points within the flight area using a YSI series 6 sonde equipped with a chlorophyll-a probe as a measurement of relative chlorophyll-a. These measurements was carried out mostly from a kayak, and an android phone (Samsung S8) was used to collect GPS points. A smartphone GPS was chosen both to reduce costs for this project, as well as to test UAV mapping using low cost, commonly available devices. In some cases, a sample was collected using a 3 meter sampling rod from the shore or a dock, and the sonde was used to measure chlorophyll-a in the sample, in which case a GPS point was measured at shore and was adjusted by moving the GPS point 10ft perpendicular to the shore or dock from the GPS point. Between one to five drone flights were conducted each sampling day. Sampling was carried out as soon as possible after the flight, although flights took between

approximately 20 minute per flight, so drifting of algae around the lake between the flight and sampling could have occurred. Sonde-chlorophyll measurement locations were inspected in ArcMap to flag points collected near the edge of algal patches, which may have increased the potential for error from the effects of wind on algal or kayak drifting. Flights were conducted on low-wind days (10-20mph wind speeds) when possible to minimize presence of waves in drone imagery, and reduce error in GPS data.

Drone Deploy cloud software (pro version) was used to stitch overlapping imagery together. Sometimes drone deploy was unable to stitch all or a subset of the drone imagery, especially when the flight height was too low. For example, image stitching failed completely for one flight at Duck Lake (Blue Earth County, MN) in August 2020, when the drone was flown at 15 meters. Image stitching was successful over the area with YSI chlorophyll-a measurements for most flights. However, on September 18th the stitched RGB map contained a section near the shore where stitching failed, because the flight was too far from the shore for sufficient image overlap near the shore. However, the Sentera camera successfully stitched the area near the shore because the Sentera camera takes photos constantly while flying towards the starting destination. There were 2 RGB images collected near the shore on September 18th; one of these images was uploaded to ArcMap, and georeferenced to the stitched RGB imagery (choosing points at fixed locations such as boat docks) in order to collect remote sensing data at the chlorophyll measurements along the shore for the RGB map.

Drone imagery, including RGB and false color imagery, were edited using ArcMap 10.8. The individual red, green, and blue (RGB) channels were added separately

to ArcMap. The false color map from the Sentera camera's stitched map was geo-referenced to the RGB stitched map to align the maps, using fixed objects such as boat docks as geo-referencing control points. A feature class polygon was created and digitized manually to store a boundary which excluded trees and shadows along the shore, and locations with obvious sunspots; the stitched maps were masked (Extract by Mask tool) using the boundary layer. Raster calculator was used to calculate NDVI and VWRI using the above equations, with the adjustment of converting each digital number to a Float. A buffer was created around sampling points to reflect potential error in the GPS location. Error for phone GPS is generally reported to range between 5-15 meters, and depends on the availability of satellites and the location of the GPS point [27]. Additional error in the GPS location could occur due to wind, for a few reasons: first, the kayak drifted between sampling and collecting a GPS point especially on windy days. Secondly, wind resulted in movement of algae, sometimes rapidly across the lake, and therefore especially on windy days, there was a potential for the drone image to fail to match up perfectly with the sample location due to algal drift.

A 10m buffer was chosen as a conservative measurement, however, the buffer distance was reduced when the algae was extremely patchy, and it was possible to tell visually that a 10m buffer would likely be unrepresentative for a set of sample points; a 2.5 or 5m buffer was used in some cases. The buffer layer was clipped using the boundary polygon because in some cases, the buffer around a point extended onto the shore, a shadow, or boat dock. The Zonal Statistics as a Table tool was used to extract the

mean VWRI or NDVI in the buffer area for each sample point. A flow chart of the UAV chlorophyll-a mapping process is shown in the Appendix in Fig. 63.

3.1.4 Remote Sensing Regressions

Regression between chlorophyll-a measured by the YSI sonde in the UAV flight area and the mean VWRI or NDVI was performed in R. Regression was performed separately for each flight date, and a regression using all the data from 4 flights at Bass Lake was also tested. Using linear, logarithmic, or various polynomial fits to model chlorophyll-a from band ratio algorithms is common [58, 87]. First, a scatterplot was used to assess the relationship; scatterplots often showed a linear correlation between the predictors and chlorophyll-a, although at times, a linear trend was not present. When a polynomial fit the data well (a curved trend was present), second and third order polynomials were compared to a linear regression, and the choice of fit was made based on the best p value and R^2 values. However, in a some cases a polynomial fit resulted in unreasonable chlorophyll-a estimates: for example, a second order polynomial fit sometimes causes lower NDVI values having the highest chlorophyll concentrations. In these cases the alternative polynomial or linear regressions were chosen to allow for the regression equation to be used to predict chlorophyll-a across the entire range of NDVI or VWRI in the image, assuming that the lowest NDVI/VWRI values had the lowest chlorophyll-a concentrations. When the regression equation predicted negative chlorophyll-a concentrations (at low VWRI/NDVI values) chlorophyll-a was set to 0 using the Con tool in ArcMap, since a negative chlorophyll-a concentration is not possible.

3.1.4.1 Regression Diagnostics

Regression diagnostics were performed by examining the dataset to see if a linear, or polynomial fit appeared reasonable. Second, a scale-location plot was used to test the assumption of homogeneity of variance (homoscedasticity) of residuals. Normality of residuals was examined by looking at a histogram of the residuals, and also by performing a Shapiro-Wilks test, where p values < 0.05 were used to indicate non-normal data. It should be noted that the sample sizes for the regressions were often extremely small, so while the Shapiro-Wilks test was used as an indicator of normality, the results from this test were viewed with caution, and a higher sample size would likely be need to confirm the distribution. Potential outliers were examined using a plot showing Cook's Distance. When the data appeared to potentially violate the assumptions of normality or homoscedasticity, transformations of the chlorophyll-a variable were tested (log, natural log, cubed root, square root). In almost all cases, the assumption of homoscedasticity appeared to be violated, but transformation did not cause an improvement. Since transformation did not improve the situation, the regression analysis was continued without transformation. Outliers were generally not removed, however, when regression relationships were particularly poor, and contained a point which had a high Cook's Distance and showed evidence the sample was collected near the edge of a patch of algae, the results were presented both with and without the outlier. The many sources of potential error in lining up the sample points with the drone image suggest that there is a basis for outlier removal, especially when sampling points occurred just outside the edge of an algae bloom: either slight error in the GPS point, drifting of the kayak, or

movement of algae with the wind between the drone flight and sample collection could have caused an algae rich sample to be collected which lined up with an area outside the bloom in the drone photo.

3.2 Time Series Forecasting

Forecasting techniques generally include recursive, direct, joint, and recursive-joint strategies [24]. Recursive strategies which forecast a single day at a time suffer from the potential for accumulated error, while direct and joint strategies can be computationally intensive or may lose dependencies between data. Due et al. [24] suggested a recursive-joint strategy (forecasting multiple days at once rather than a single day or the full horizon) provides a balance between reducing accumulated error, reducing computation time, and retaining dependencies in the data. Therefore, a 2-day forecast was chosen using the recursive-joint approach suggested by Du et al. [24]. Models can use re-estimation or be without re-estimation; a model with re-estimation means that the forecasting model is re-trained each time a new timestep is predicted, while a model without re-estimation means that a single model is applied for all forecasting timesteps.

ARIMA is an autocorrelation model which used dependencies between an observation and lagged observations to predict future values [76]. It is parameterized with P, D, Q, where P is the autoregressive order, Q is the moving average order, and D is the differencing order. ARIMA requires the series to be stationary, so differencing may be required to reach stationarity. ARIMA was carried out in R using auto.arima, which automatically selects the p,d, and q parameters with the lowest Akaike Information Criterion (AIC), where a lower AIC shows a better fit [39]. The forecasting horizon was

10 days for ARIMA. A 10-day forecast was compared to the actual data in the test set, where 67% of the dataset was used as the training set, and 33% as the test set. Root mean squared error (RMSE) was used to measure model performance. RMSE is a common measure for model accuracy, which has the advantage of being sensitive to large errors, and is scaled to units of forecast values [76].

Long Short Term Memory (LSTM), is a recurrent neural network, where the hidden layer stores previous information, and the data moves through cells in sequence [76]. A unique strength of LSTM as a RNN is that it is capable of using long sequences in the model. [76]. LSTM is a back propagating neural network, meaning that the weights are changed based on the model output and then re-run [76]. Data moves through LSTM using gates: data input to a cell is determined at a Forget gate, the data is stored in a Memory gate, and output from an Output gate [76]. LSTM was performed using Python 3.0 in the Anaconda environment (Spyder), using the Keras package, and following methods modified from Brownlee [12]. LSTM was performed both as a univariate forecast, and multivariate forecast. The univariate forecasting was conducted using a 2-day forecast approach, where longer forecasts were carried out by adding the first 2 forecasted days to the sample history to forecast the next 2 days [24]. Since predictor variables such as conductivity, temperature, were not forecasted in the multivariate model, forecasted chlorophyll could not be added to the history to produce the next forecast; therefore, the multivariate forecast used a separate model to forecast each future timestep (days 1 through 10).

Hyperparameters for machine learning models such as LSTM are defined as parameters which are set prior to training the model. Hyperparameter selection was tested thoroughly on the multivariate LSTM model. The following hyperparameters were tested: number of input days, number of epochs, batches, number of hidden layers, and number of network nodes in each hidden layer. Each variable was changed over a relatively fine scale while keeping the others constant. Next, a coarser scale grid search was performed to test the best combinations of hyperparameters, where the ranges of each hyperparameter in the coarse scale grid search were chosen based off of the results from the fine scale searches. Hyperparameter testing was performed separately on the 2019 and 2020 datasets, and the lowest RMSE was used as the indicator of the best model setup. An RMSE was produced for each multivariate forecast as an average of 10 repetitions, producing an average RMSE for the day-1 forecast, the day-2 forecast… the day-10 forecast. These average RMSEs were averaged together, to produce an estimate of the average RMSE for the hyperparameter setup. The best hyperparameters for the multivariate forecast (epochs, batches, input days, network nodes) were tested in the univariate forecast (rather than doing another grid search for the univariate forecast), and it was found that the univariate forecast had a relatively low RMSE. Therefore, given time constraints, hyperparameter testing was not performed for the univariate forecast.

The impact of a wavelet transformation on the performance of the univariate forecast was examined. The wavelet transformation was performed using Matlab's Wavelet Toolbox module to convert the chlorophyll-a series into a low frequency

approximation series A', and a high frequency detail Series D' [92]. A discrete wavelet transformation was used:

Equation 7 - Discrete Wavelet Transformation [30, 80]:

$$\varphi_{\mathrm{m,n}}(t) = \frac{1}{\sqrt{a_0^m}} \varphi \frac{t - nb_0 a_0^m}{a_0^m}$$

where m and n are the dilation and transformation parameters, and $a_0 > 1$, and $b_0 > 0$ are the dilation step and location parameters, respectively [30, 80]. A univariate LSTM forecast was performed separately on each approximation and detail series, and all the approximation and detail series with the 10-day forecasts were summed together to reconstruct the chlorophyll-a series with 10-day forecast.

3.2.1 Multivariate Forecast: Determination of Controlling Factors

Multiple regression was performed in R to examine the relationship between chlorophyll-a measured by the DS5 or YSI sonde (installed at the trail camera station in 2019 and 2020 – see trail camera section above for sonde and weather station setup methods) and other chemical and physical water quality or meteorological parameters (including conductivity, pH, water temperature, turbidity, ORP, air temperature, PAR, wind direction, wind speed, barometric pressure, precipitation). Stepwise regression was used to remove predictor variables which were not significant until the simplest multiple regression model was obtained. Multiple precipitation parameters were tested including the previous week, 2-week, 30-day, and 60-day precipitation totals. Variance inflation factor (VIF) was calculated to test for the presence of multi-collinearity, and VIFs 10 or above were considered a violation of this assumption. Variables which were correlated to

chlorophyll in the multiple regression forecast were used as inputs to the multivariate LSTM forecasts in order to test if a single parameter chlorophyll probe with a univariate forecast could perform comparably to a higher cost monitoring approach (see above).

3.3 Impact of meteorological factors on algal blooms

Page et al. [60] developed methodology for atmospherically correcting and processing Sentinel-2 maps to map variety of water quality products including chlorophyll-a (using NDCI). Chlorophyll-a maps created using these methods were available online from the University of Minnesota Lake Viewer through a temporary web app (Leif Olmanson and Ben Page, University of Minnesota) [60]. Median chlorophyll-a during July/August 2017-2020 were downloaded from all available maps at Bass Lake from the lake viewer app. July and August concentrations were chosen similar to the approach in Ho et al. [38], since peak chlorophyll-a shows the final impact of nutrient loading in a lake. Meteorological data (precipitation, temperature, and snowfall) 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). The nearest station to each lake was used. Data was downloaded from approximately 160 lakes in the Western Corn Belt Plain (WCBP) region of Minnesota. A multiple regression was performed to examine the impact of precipitation and snowfall on chlorophyll-a broadly across the WCBP region. Precipitation factors were chosen following methods in Ho et al. [38], including 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 three precipitation factors: (ln(JJA), ln(MAM), and ln(AN)). Ho et al. [38] chose these precipitation variables from a list of 27 climate change indices (Expert Team on Climate Change Detection and Indices, http://etccdi.pacificclimate.org/ list_27_indices.shtml), and noted that these indices are supported by studies linking precipitation to nutrient loading [78]. 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.

4 Results

4.1 Remote Sensing Results

4.1.1 Trail Camera Results

The slope of the relationship between VWRI and chlorophyll-a for trail camera images was negative, although there was no significant correlation between VWRI and chlorophyll-a ($R^2 = 0.0014$, $p = 0.00162$, Fig. 4) for 1,200 of the trail camera images. The regression appeared to be significantly influenced by two outlier values: most of the VWRI values were between -0.4 and -0.2, while the 2 outliers had a VWRI between 0.2-0.4. The regression had a violation of the assumption of normality which was not possible to improve much using transformations. Removal of the outliers had little effect on the strength of the relationship (Fig. 5, Adjusted $R^2 = 0.0121$, $p = 0.00112$).

Image quality of Spypoint images was generally poor due to reflection, shadows, and turbulence, and the majority of images were affected by these problems. The small subset of selected trail camera imagery also showed a poor relationship between VWRI and chlorophyll-a (Fig. 6). The slope switched to a positive slope, however, the adjusted R^2 was extremely low (R^2 = -0.0125, p = 0.4708). When the 2 outliers were removed, the relationship between chlorophyll-a and VWRI improved slightly using this subset, but was almost nonexistent (R^2 = 0.04). Processing the imagery using raster calculator was time consuming, so in the future C# is recommended for this analyses.

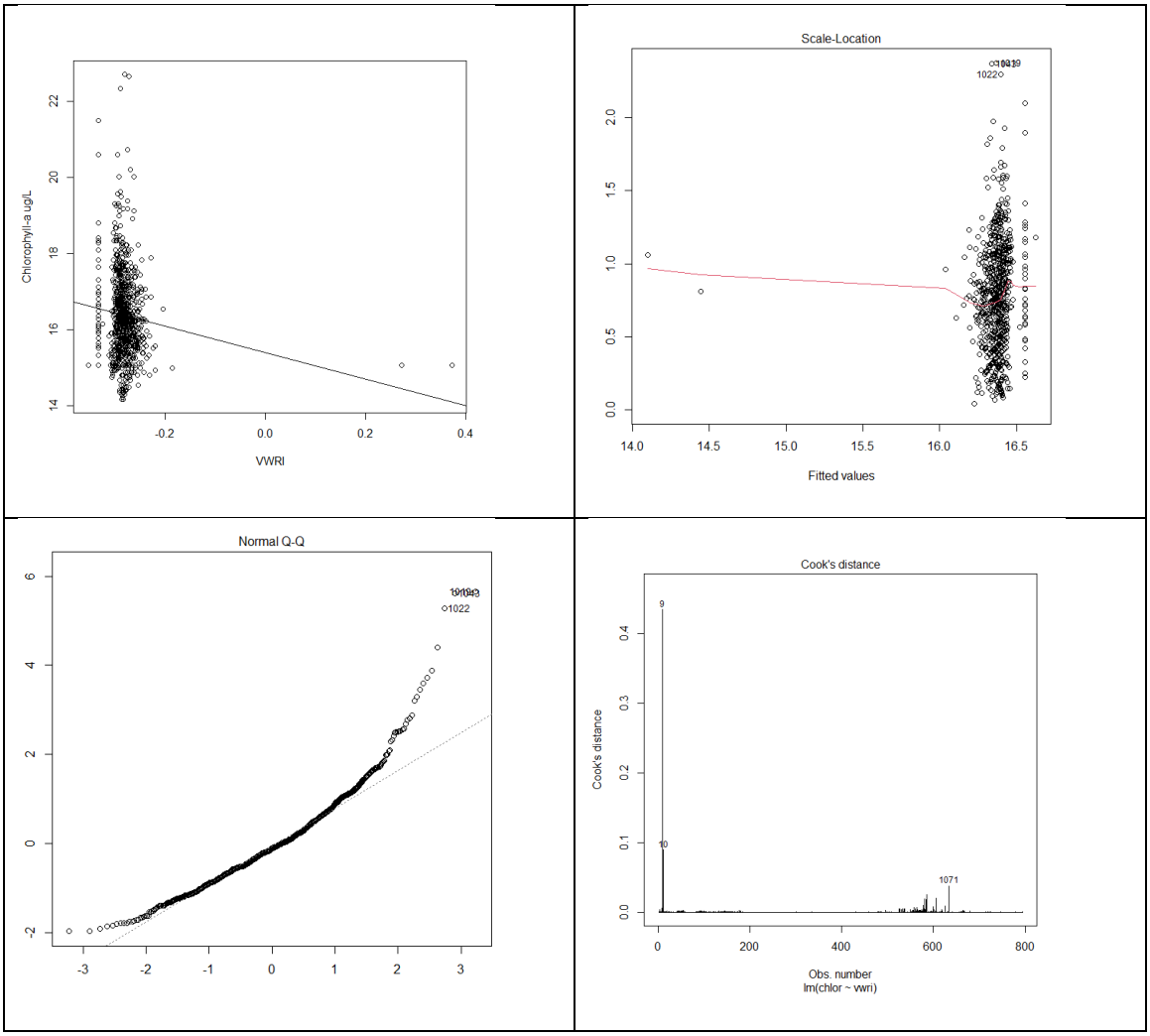

Figure 4. Chlorophyll-a from sonde vs. VWRI from 1,200 trail camera images. Adjusted $R^2 = 0.0112$, $p = 0.00162$).

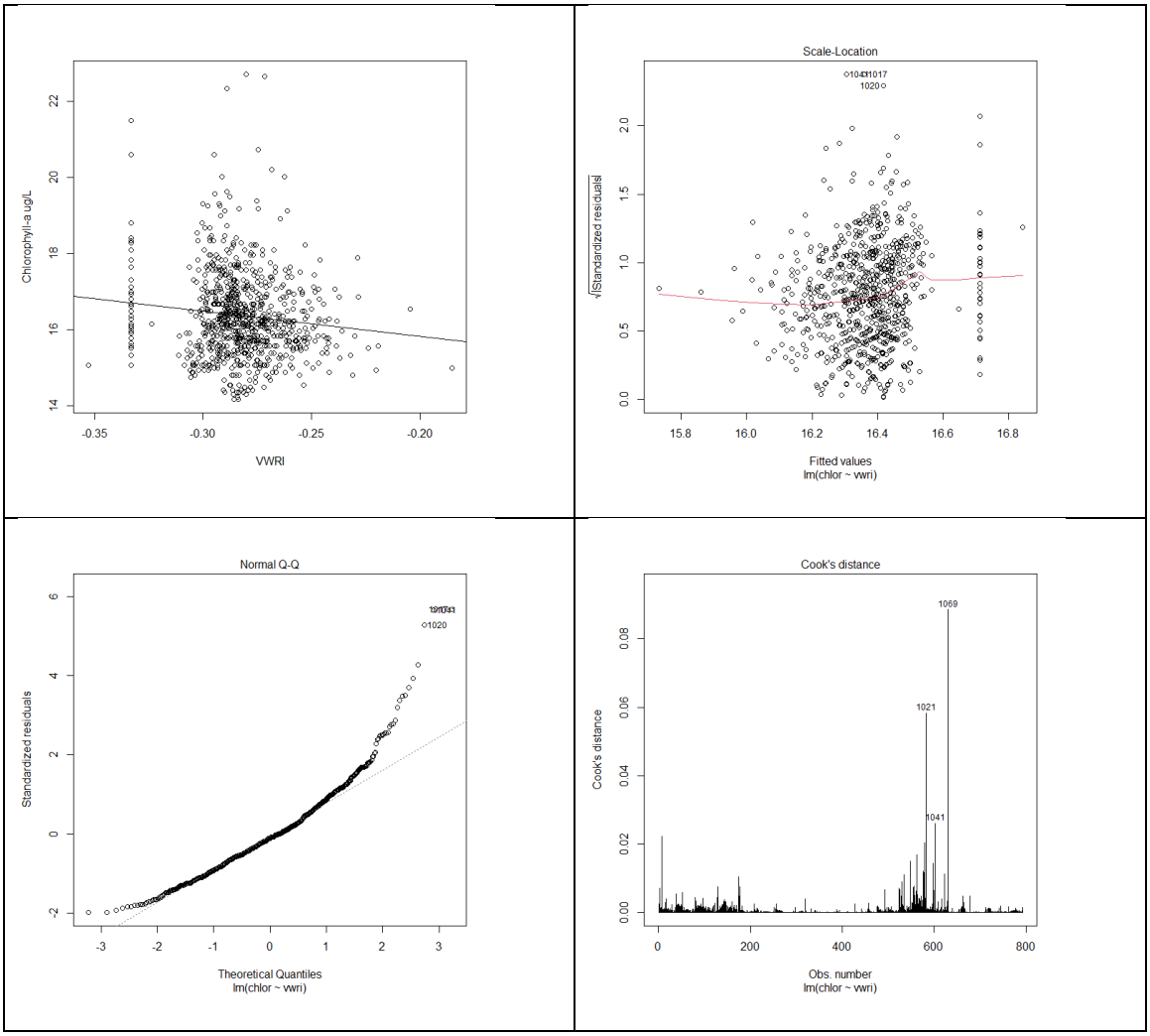

Figure 5. Chlorophyll-a from sonde vs. VWRI from 1,200 trail camera images with 2 outliers removed (Upper Left), and regression diagnostic plots. Adjusted $R^2 = 0.0121$, $p = 0.00112$).

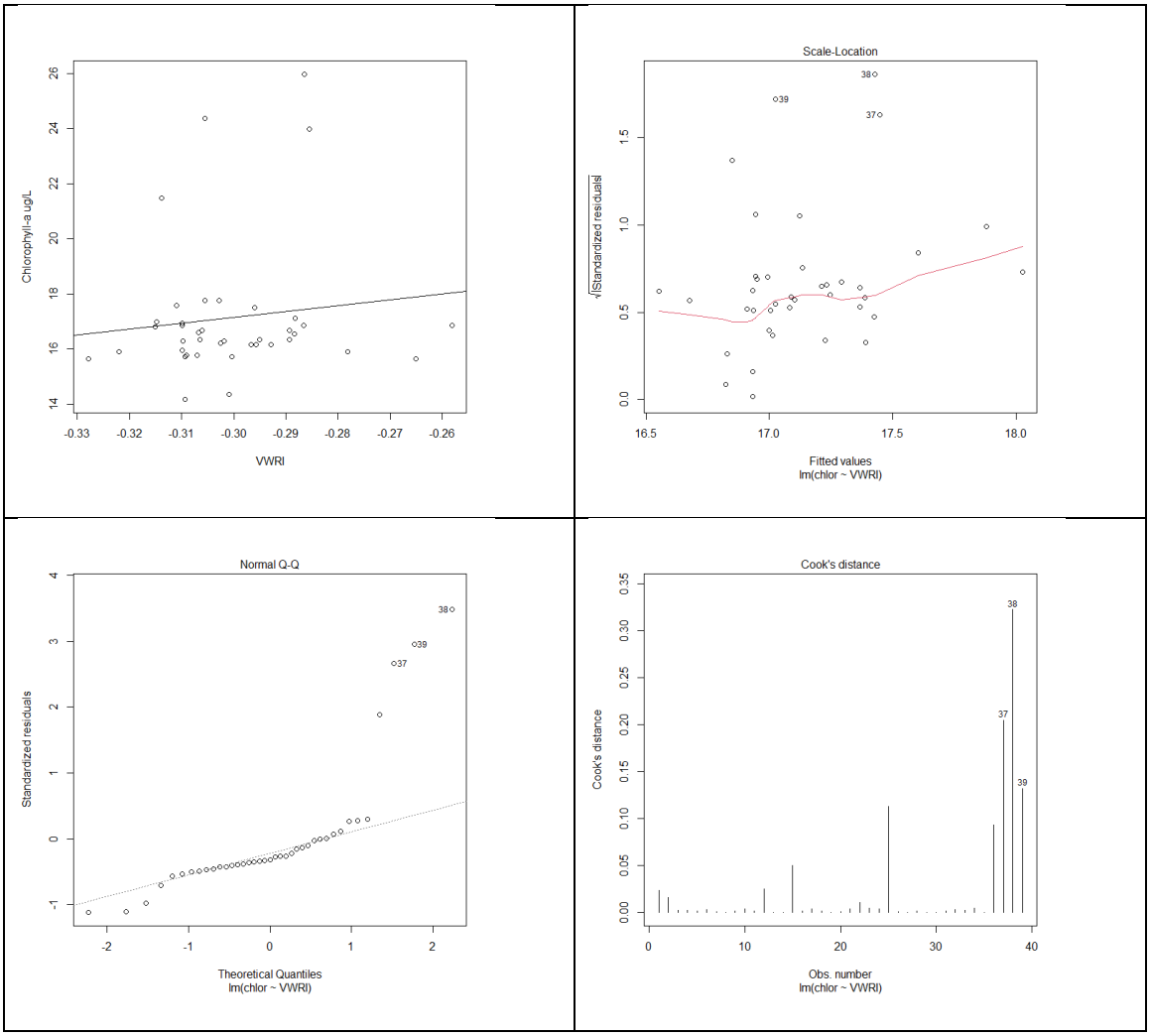

Figure 6. Chlorophyll-a vs. VWRI for a selected subset of images from Bass Lake, 2019 data; upper right and lower plots: regression diagnostic plots. Adjusted R^2 = -0.0125, p = 0.4708.

The University of Kentucky classifier showed significant variability in the probability of blue-green algae presence for Bass Lake (Fig. 7). From May to November 2019, the classifier showed fluctuating probability of blue-green algae ranging from 0 to 1 throughout the summer, and in addition, most days contained a probability of 1.00 (100%) at some point. The presence of a 100% probability of blue-green algae

throughout the entire monitoring season (points at 1.00 probability in Fig. 7) would suggest that at some point every day, blue-green algae was present at the trail camera site. However, these results did not match visual observations of the images and of the lake during field work, where the presence of blue-green algae was not observed in general until later in the summer and during the fall.

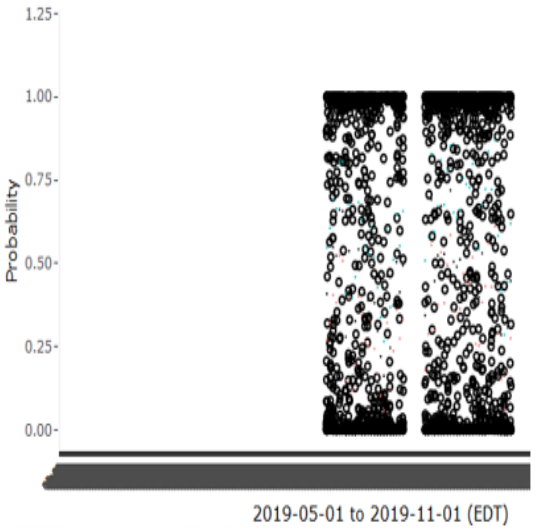

Figure 7. Blue green algae probability (0 to 1.00) from May 2019 to October 2019 from trail camera images, using a University of Kentucky classifier (graph created by Mike Waters, University of Kentucky).

4.1.2 Sentinel-2 Satellite Results

VWRI had a weak, positive correlation with chlorophyll-a (Adjusted R^2 = 0. 0.171, P = 0.00961, Fig. 8), while NDCI had a strong, positive correlation with chlorophyll-a (Adjusted R^2 = 0.65, P = 1.19e-08, Fig. 9). There was a lack of samples with eutrophic chlorophyll-a concentrations: chlorophyll-a generally ranged between 0-

50 ug/L for most of the dataset, with 2 samples > 100 µg/L. NDCI also had a lack of intermediate data, with NDCI ranging from -.1 to 0.2 for all but three samples; therefore the NDCI regression was missing samples with NDCI from 0.2-0.4, in general. The 2 high concentration samples likely drove the positive correlation to a degree. In fact, removing the top three model outliers from the Cook's Distance plot, including the 2 highest chlorophyll-a samples, reduced the R^2 to 0.248 ($p = 0.00301$, Lower Left of Fig. 9). Despite the lower correlation when 2 outliers are removed, the NDCI plot shows less variability than the VWRI plot, and suggests the possibility of two clusters of lakes which could have separate slopes.

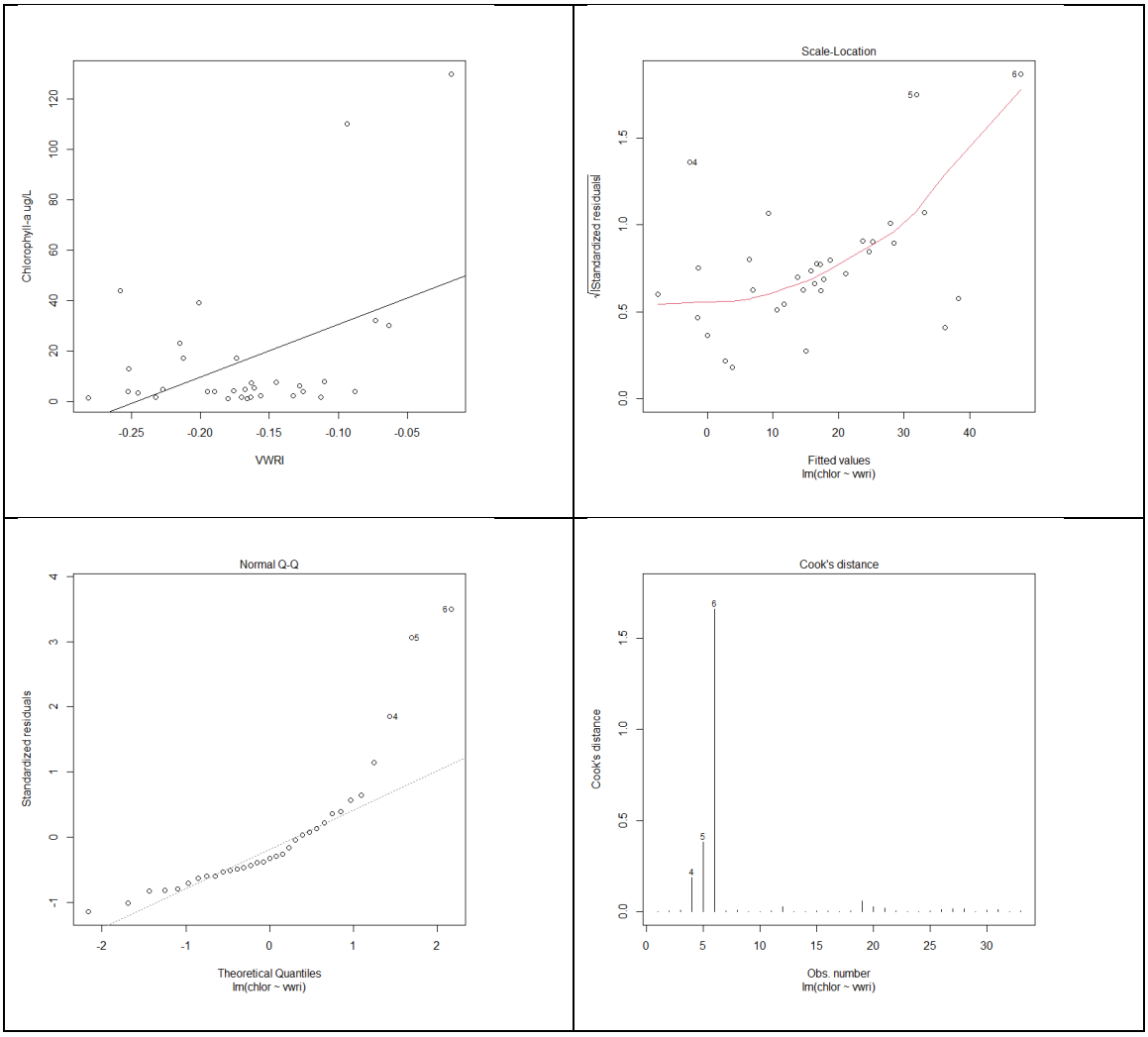

Figure 8. VWRI vs. Chlorophyll for over 30 lakes in the Twin Cities area (Upper Left); Met Council chlorophyll grab samples. Data is from May 15th, 2018. Adjusted R^2 = 0.171, P = 0.00961. Upper right: scale location plot to test for homoscedasticity; Lower Left: Q-Q plot to check for normality; Lower Right: Cook's distance (test for outliers).

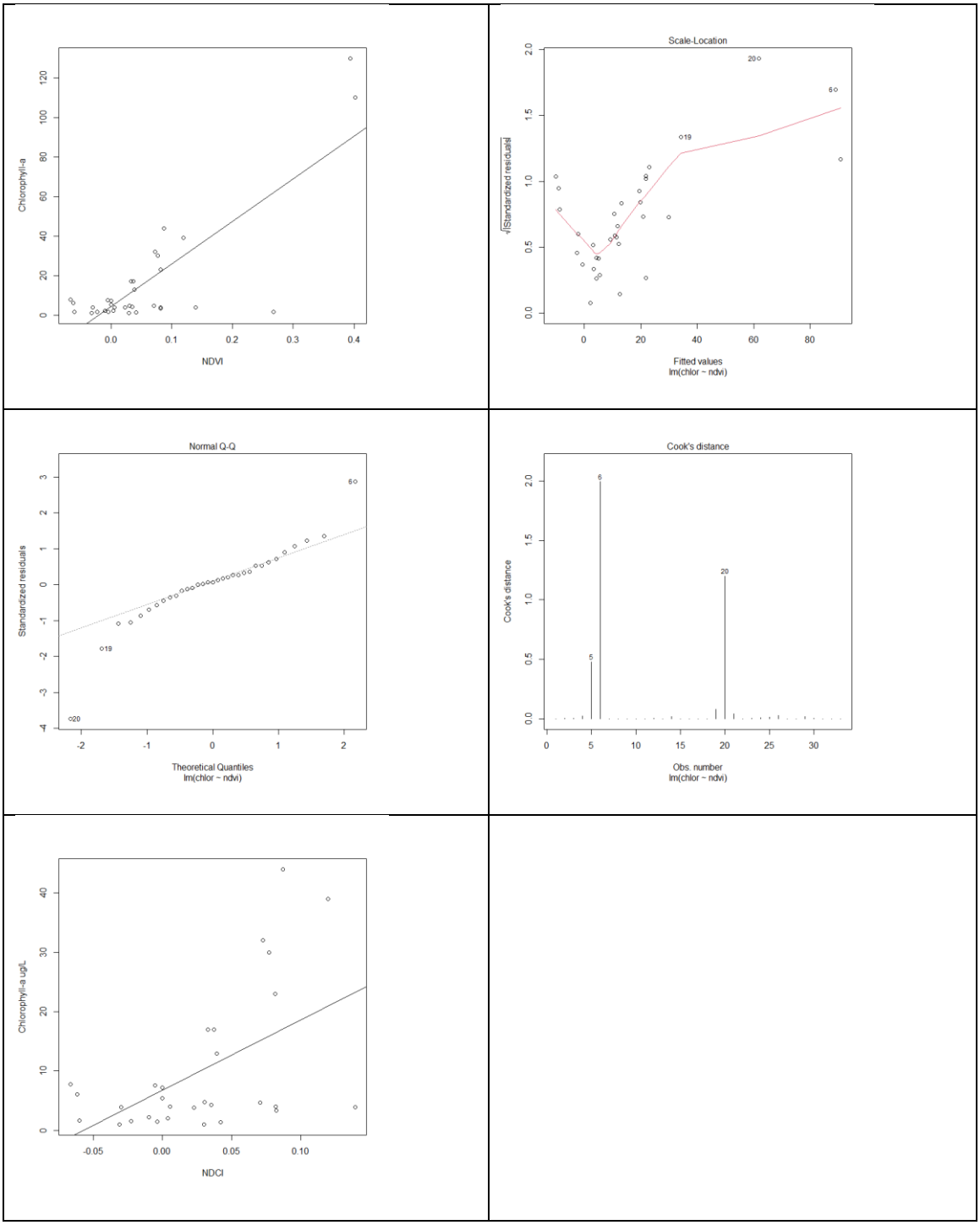

Figure 9. NDCI vs. chlorophyll for samples from over 30 lakes in the Twin Cities region on 5-15-2018 (Upper Left); Met Council chlorophyll-a grab samples. NDVI was positively correlated with chlorophyll (P = 1.193e-08, adjusted R^2 = 0.644). Upper right and Middle: regression diagnostic plots. Lower left: three potential outliers removed based on Cook's distance (R^2 = 0.248, p = 0.00301).

4.1.3 UAV results

The first flight on 9/5/2021 at Little Rock Lake, Benton County, MN, showed a weak, positive correlation between both VWRI, NDVI, and chlorophyll-a measured using a YSI sonde, however the relationship between VWRI and chlorophyll-a was not significant, unlike for NDVI and chlorophyll-a (VWRI: $R_{adj}^2 = 0.17$, $p = 0.06$, Fig. 10; NDVI: $R_{adj}^2 = 0.323$, $p = 0.00527$, Fig. 11). VWRI was log transformed, since the p value improved after the log transform, and given that logarithmic relationships between chlorophyll-a and band ratio algorithms have been report in the past [87]. Both regressions showed substantial departures from the assumption of homogeneity of variance.

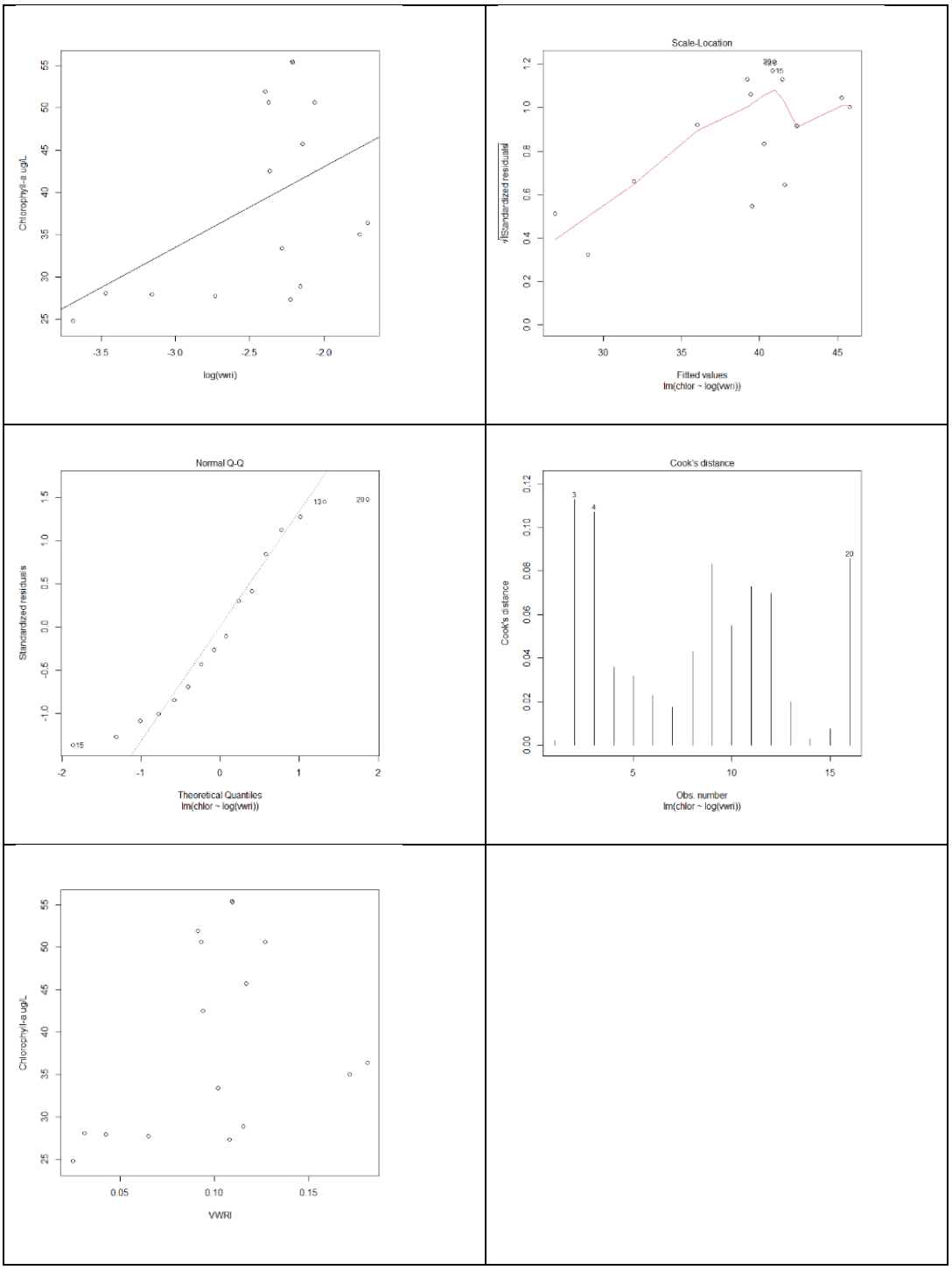

Figure 10. Top Left: Chlorophyll as a function of VWRI on 9/5/2021 at Little Rock Lake, Benton County, MN with logarithmic best fit; Adjusted $R^2 = 0.17$. Upper right: scale location plot to test for homoscedasticity, Middle left: Q-Q to check for normality, Middle Right: Cook's distance (test for outliers). Lower left: non-transformed data.

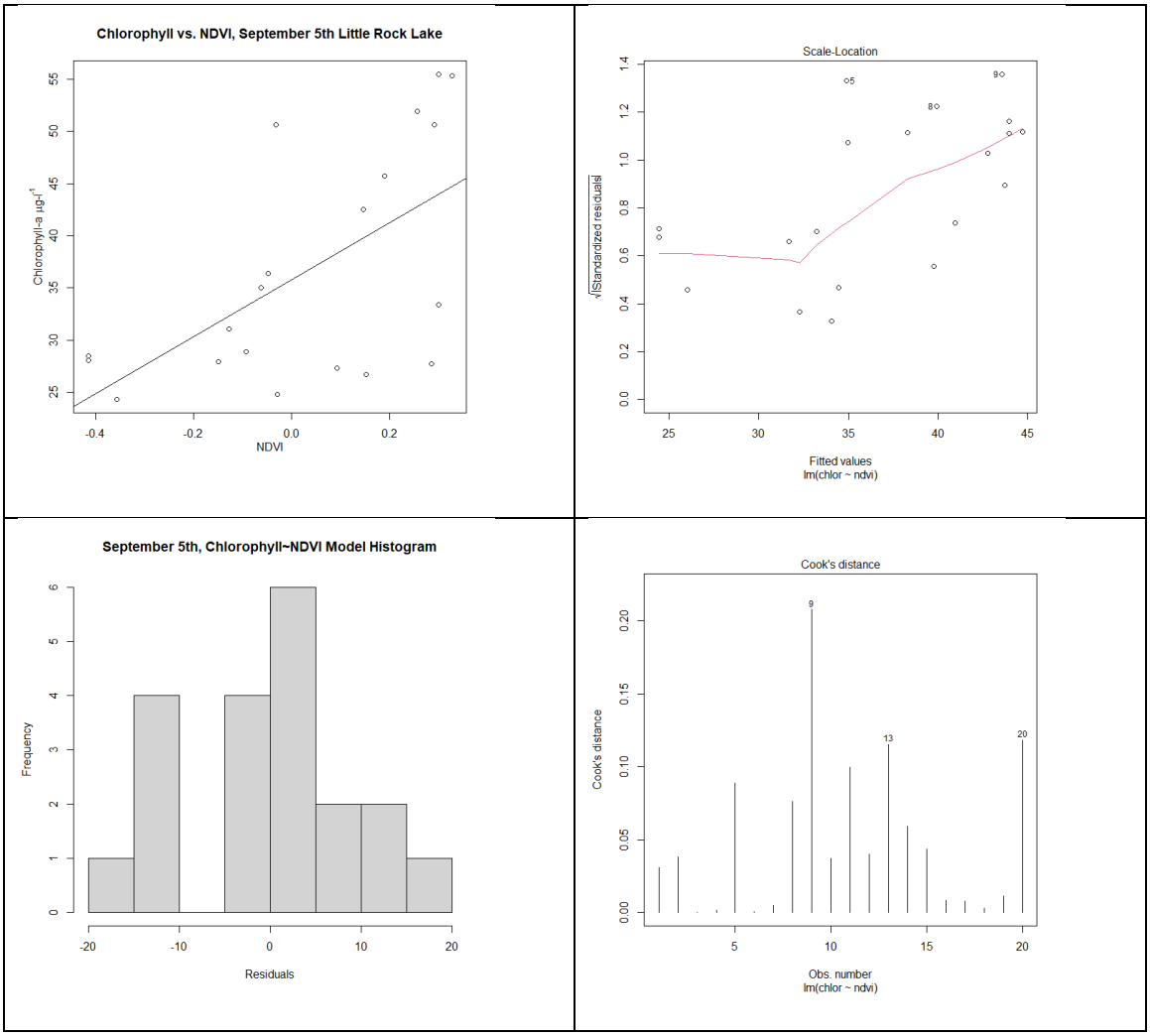

Figure 11. Top Left: Chlorophyll as a function of NDVI on 9/5/2021 at Little Rock Lake, Benton County, MN with linear best fit; $R^2 = 0.323$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

On 9/14/2020 for Bass Lake, both NDVI and VWRI had strong, positive relationships with chlorophyll-a. The relationship between chlorophyll-a and NDVI appeared to follow a polynomial fit ($R_{adj}^2 = 0.986$, P = 4.12e-11, Fig. 12), while VWRI appeared closer to a linear fit ($R_{adj}^2 = 0.866$, p = 3.07e-07, Fig. 13). While NDVI ranged from about -0.5 to 0.6, there was a -0.08 to -0.01, there was a large section missing samples from about 0.2-0.4 NDVI and -0.07 to -0.03 VWRI (100-200 µg/L chlorophyll-

a). VWRI appeared to almost have a polynomial fit, however, a few of the high VWRI samples show a decline in chlorophyll-a with VWRI, which made a polynomial fit challenging. Both regressions showed substantial departures from the assumption of homogeneity of variance. Although both datasets passed a Shapiro-Wilks test, the low sample size is low, and histograms show some evidence of skew/tails.

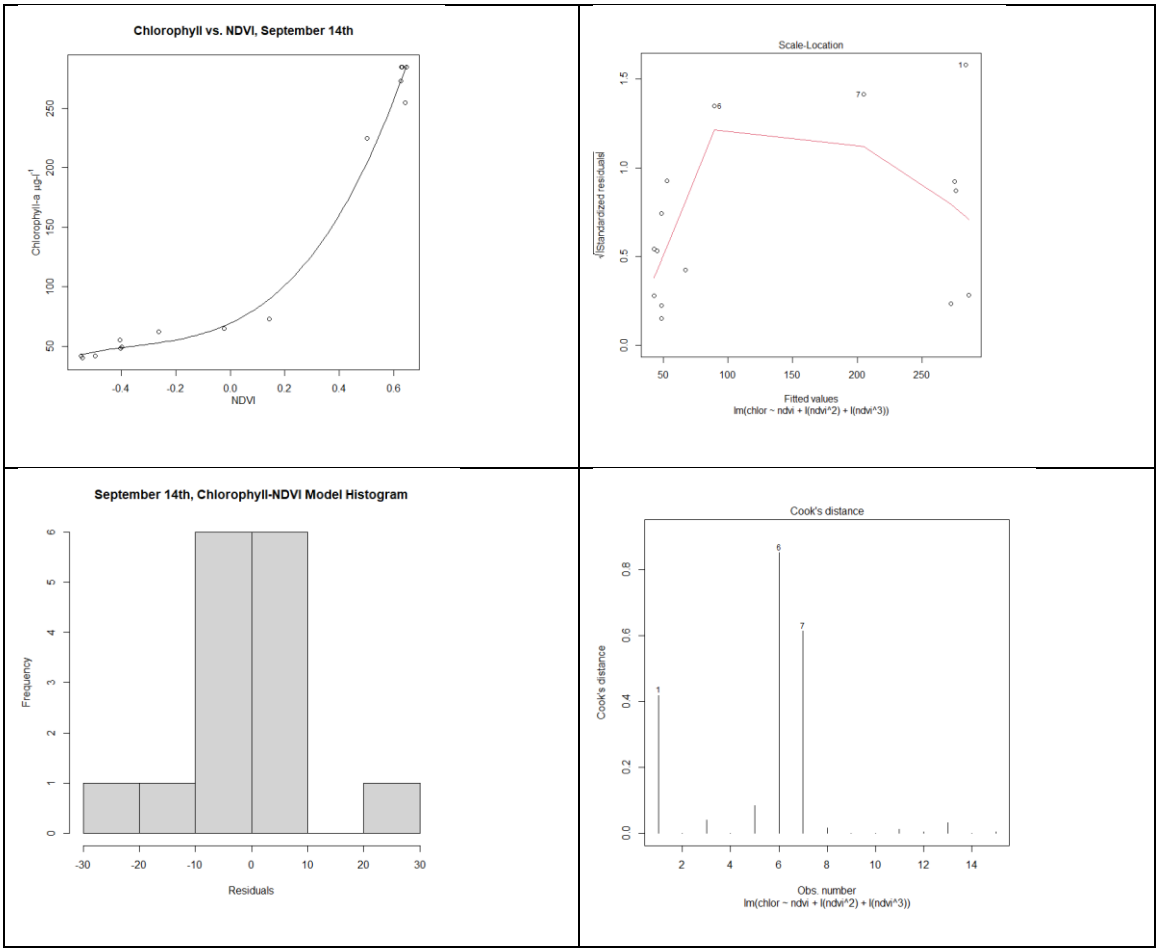

Figure 12. Top Left: Chlorophyll as a function of NDVI on 9/14/2021 at Bass Lake, Faribault County, MN with 3rd order polynomial best fit. $R^2 = 0.986$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

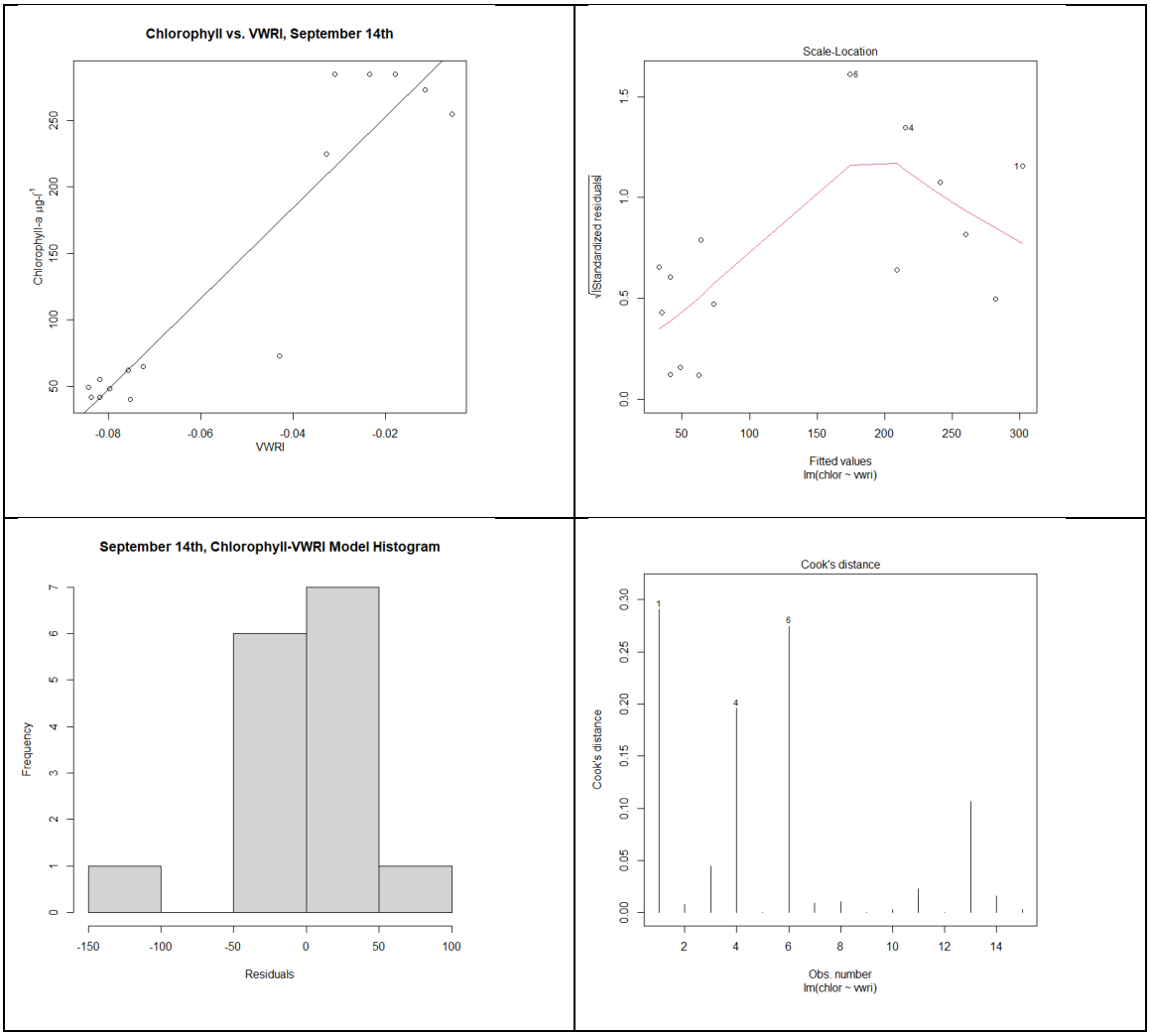

Figure 13. Top Left: Chlorophyll as a function of VWRI on 9/14/2021 at Bass Lake, Faribault County, MN with linear best fit. $R^2 = 0.866$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

On 9/18/2020 for Bass Lake, both NDVI and VWRI showed polynomial, strong, relationships with chlorophyll-a, however, the relationship for VWRI ($R_{adj}^2 = 0.799$, $p = 0.000296$, Fig. 15) was stronger than NDVI ($R_{adj}^2 = 0.665$, $p = 0.00774$, Fig. 14). VWRI was fit with a concave 2nd order polynomial, while NDVI was fit with a 3rd order polynomial. Besides showing an improved fit with a 2nd order polynomial for VWRI

compared to a linear fit, the linear relationship would have had a negative slope, leading to extremely high chlorophyll-a concentration estimates for pixels with the lowest VWRI values. Both datasets showed trends in the scale-location plots, and although they passed the Shapiro-Wilks test ($p > 0.05$), right skews were present in the histogram.

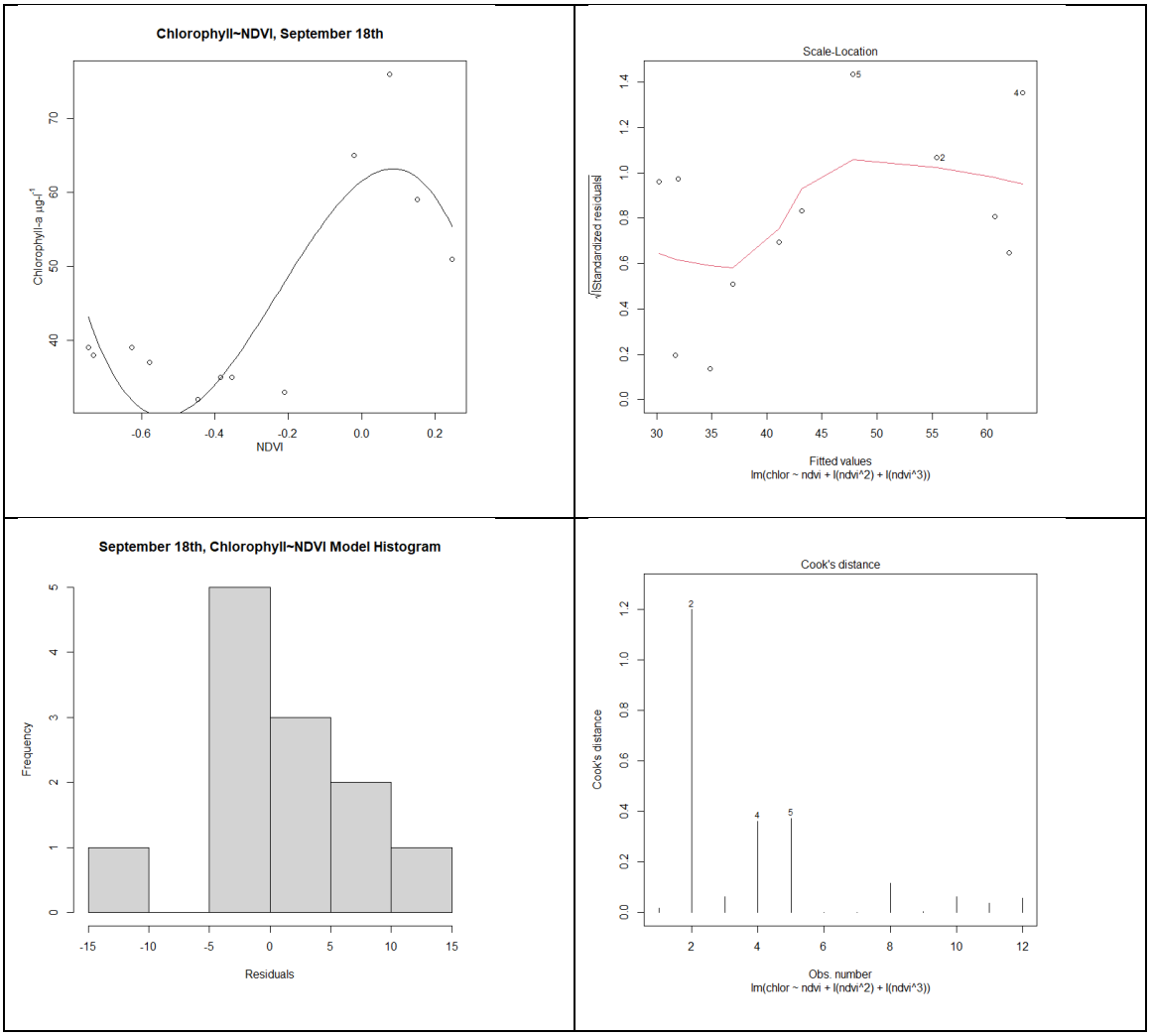

Figure 14. Top Left: Chlorophyll as a function of NDVI on 9/18/2021 at Bass Lake, Faribault County, MN with polynomial best fit; $R^2 = 0.665$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

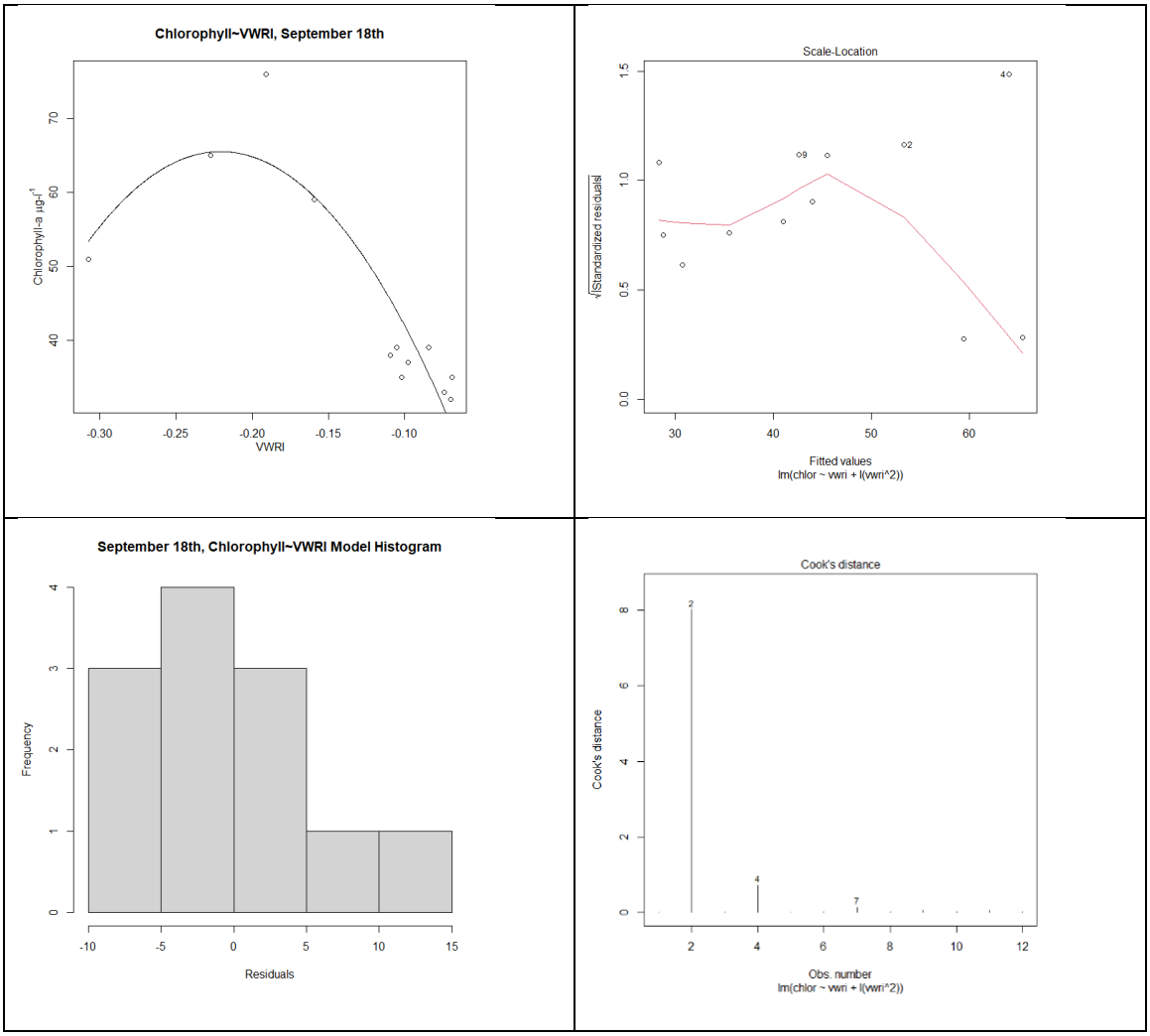

Figure 15. Top Left: Chlorophyll as a function of VWRI on 9/18/2021 at Bass Lake, Faribault County, MN with polynomial best fit; $R^2 = 0.799$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

On 9/26/2020 for Bass Lake, NDVI was positively correlated to chlorophyll-a ($R_{adj}^2 = 0.577$, p = 2.45e-04, Fig. 16), while the VWRI relationship was not significant ($R_{adj}^2 = 0.0785$, p = .145, Fig. 17). However, outliers appeared to be present in both plots, and were confirmed by the Cook's Distance plot. After removing 1 outlier from each dataset the fits improved significantly, with NDVI ($R_{adj}^2 = 0.785$, p = 3.06e-06, Fig. 16) still showing a stronger relationship than VWRI ($R_{adj}^2 = 0.467$, p = 0.00212, Fig. 17).

Both datasets showed potential violations of regression assumptions, with a right skew for NDVI and left skew for VWRI; while the scale-location plots suggested lack of homogeneity of variance, NDVI seemed to show a trend of higher variance in the center, while the high variance on the left for VWRI appeared to be only due to a single sampling point.

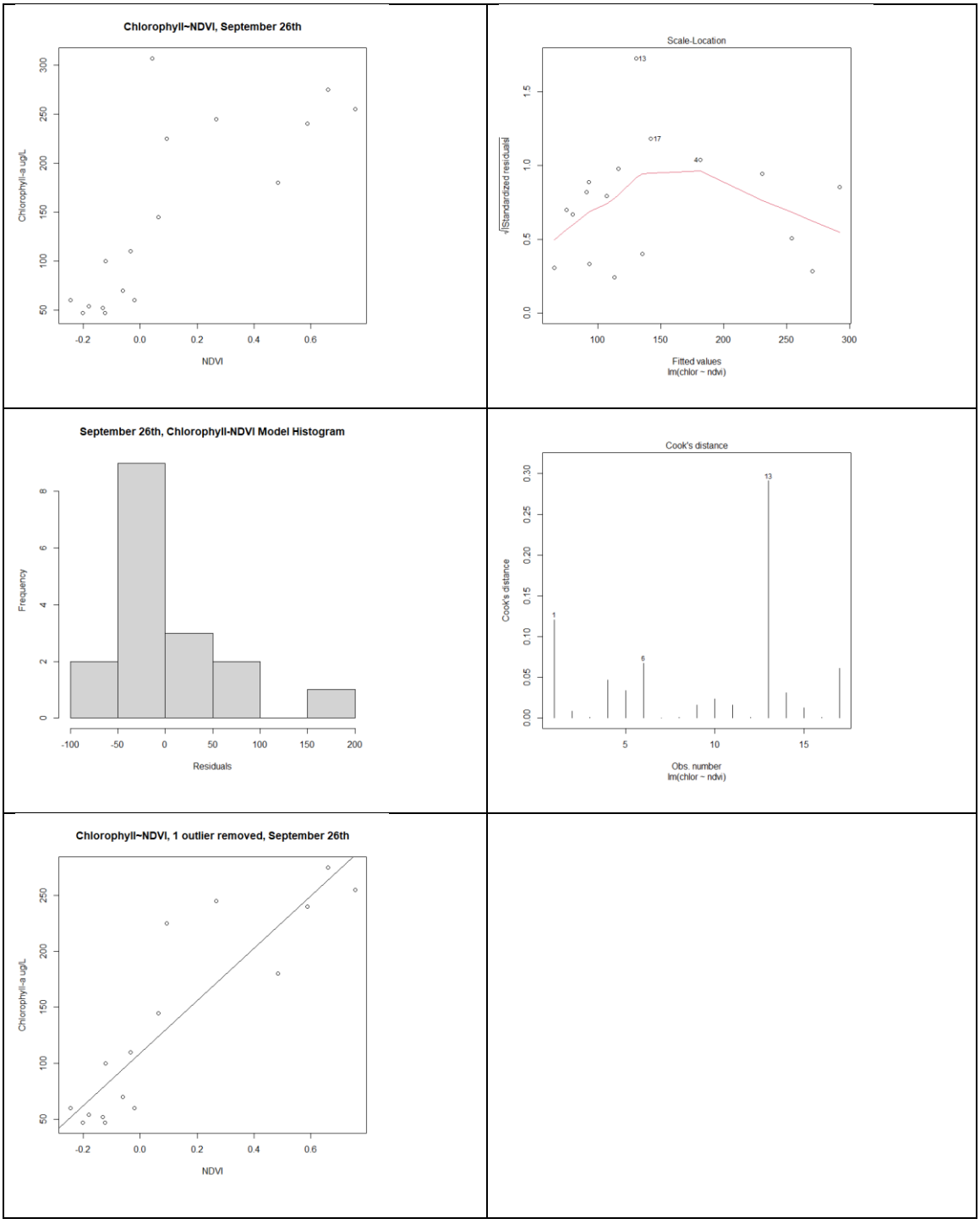

Figure 16. Top Left: Chlorophyll as a function of NDVI on 9/26/2021 at Bass Lake, Faribault County, MN with linear best fit; $R^2 = 0.577$. Upper right: scale location plot to test for homoscedasticity, Middle left: histogram to check for normality, Middle Right: Cook's distance (test for outliers). Lower left: 1 potential outlier removed: $R^2 = 0.785$.

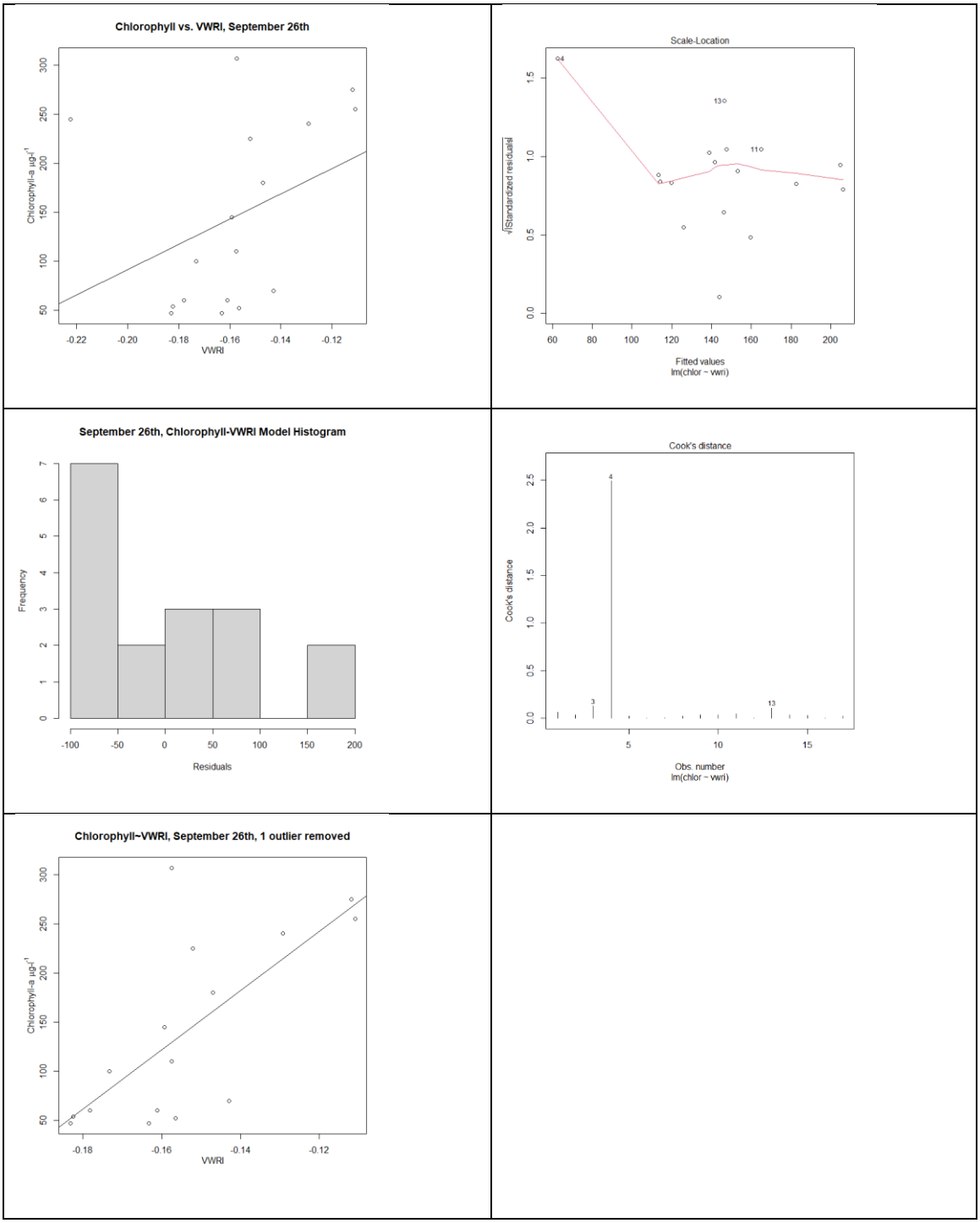

Figure 17. Top Left: Chlorophyll as a function of VWRI on 9/26/2021 at Bass Lake, Faribault County, MN with linear best fit; $R_{adj}^2 = 0.0785$. Upper right: scale location plot to test for homoscedasticity, Middle left: histogram to check for normality, Middle Right: Cook's distance (test for outliers). Lower left: 1 potential outlier removed, $R_{adj}^2 = 0.467$.

On 10/8/2020 for Bass Lake, both NDVI and VWRI were positively correlated to chlorophyll-a, with VWRI ($R_{adj}^2 = 0.797$, $p = 0.00427$, Fig. 19) showing a stronger relationship than NDVI ($R_{adj}^2 = 0.631$, $p = 0.0203$, Fig. 18). The NDVI plot suggests the possibility of a polynomial relationship, and while a second order polynomial relationship had a stronger fit than the linear relationship ($R_{adj}^2 = 0.8112$, $p = 0.01584$), this would have resulted in increasing chlorophyll-a with decreasing NDVI, so the linear fit was chosen for NDVI. One limitation of the results from October 8th was the lowest sample size of all flights ($n = 7$). However, all samples were collected off of the dock or from the shore, which may have reduced the error due to drift in the kayak.

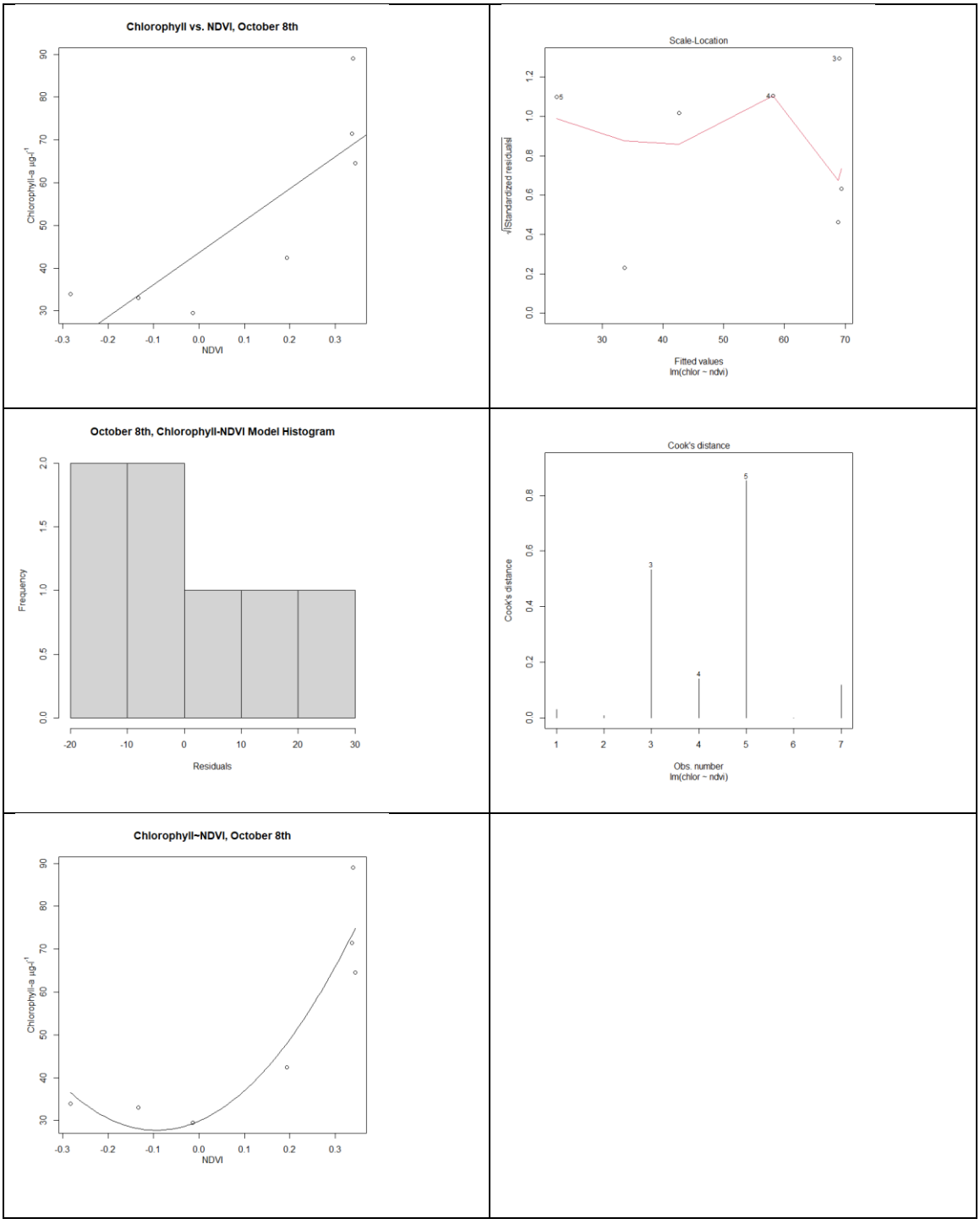

Figure 18. Top Left: Chlorophyll as a function of NDVI on 10/8/2021 at Bass Lake, Faribault County, MN with linear best fit. $R^2 = 0.631$. Upper right: scale location plot to test for homoscedasticity, Middle left: histogram to check for normality, Middle Right: Cook's distance (test for outliers). Lower Right: second polynomial relationship ($R_{adj}^2 = 0.8112$).

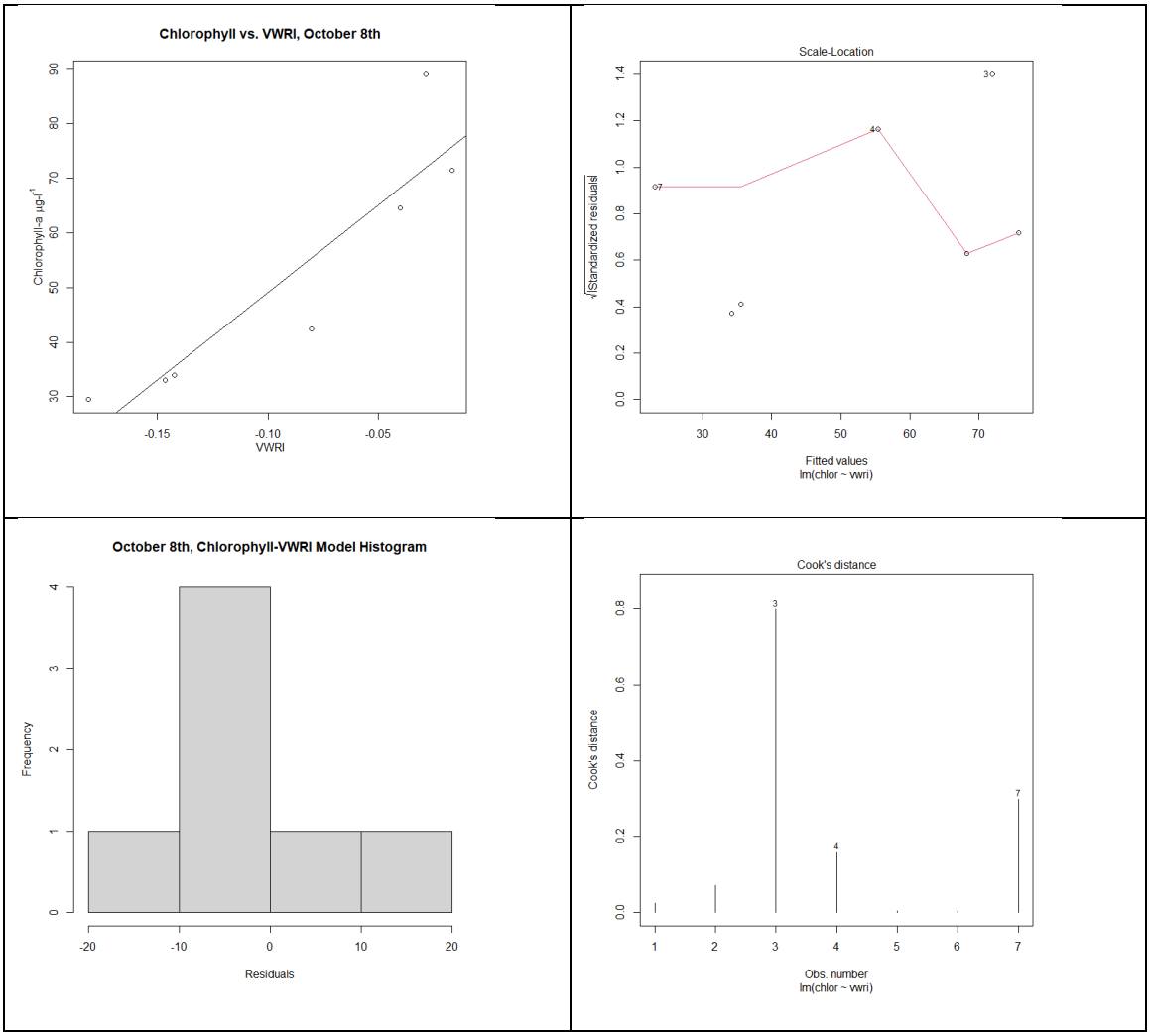

Figure 19. Top Left: Chlorophyll as a function of VWRI on 10/8/2021 at Bass Lake, Faribault County, MN with linear best fit. $R^2 = 0.797$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

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, Fig. 20) 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, Fig. 21). Both datasets showed violations of regression assumptions which could not be improved with transformation: NDVI showed an increasing variance in the center, while VWRI showed a decrease in variance at lower

chlorophyll-a, and both datasets had a right skew. While potential outliers were present in the NDVI plot, they generally fell among other samples with significant variation from the best fit line, so seemed to represent the variation in the dataset well, rather than acting as an outlier, so no outliers were removed.

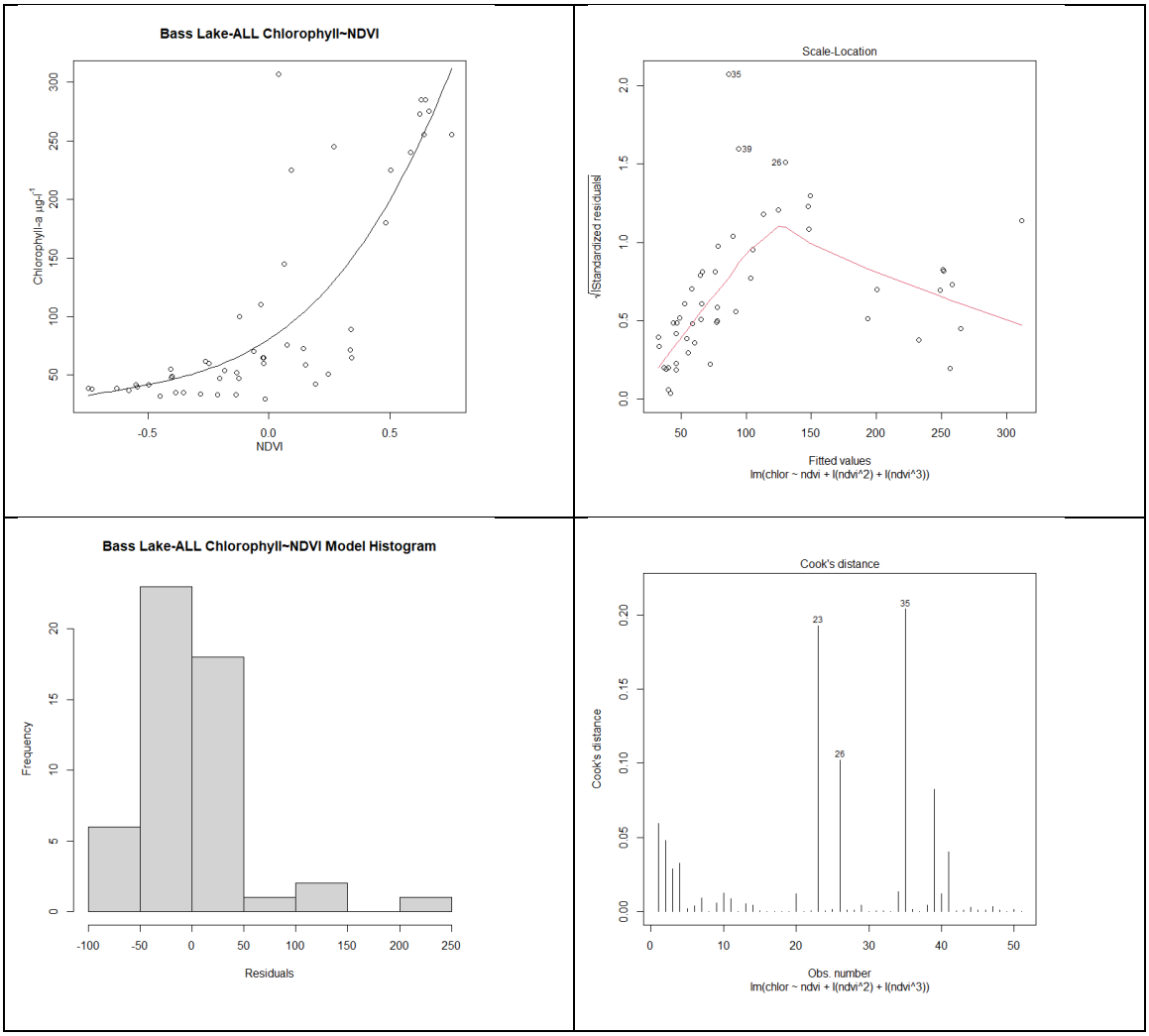

Figure 20. Top Left: Chlorophyll as a function of NDVI at Bass Lake combining all data from four flights in the fall of 2020, with polynomial best fit; $R^2 = 0.678$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).

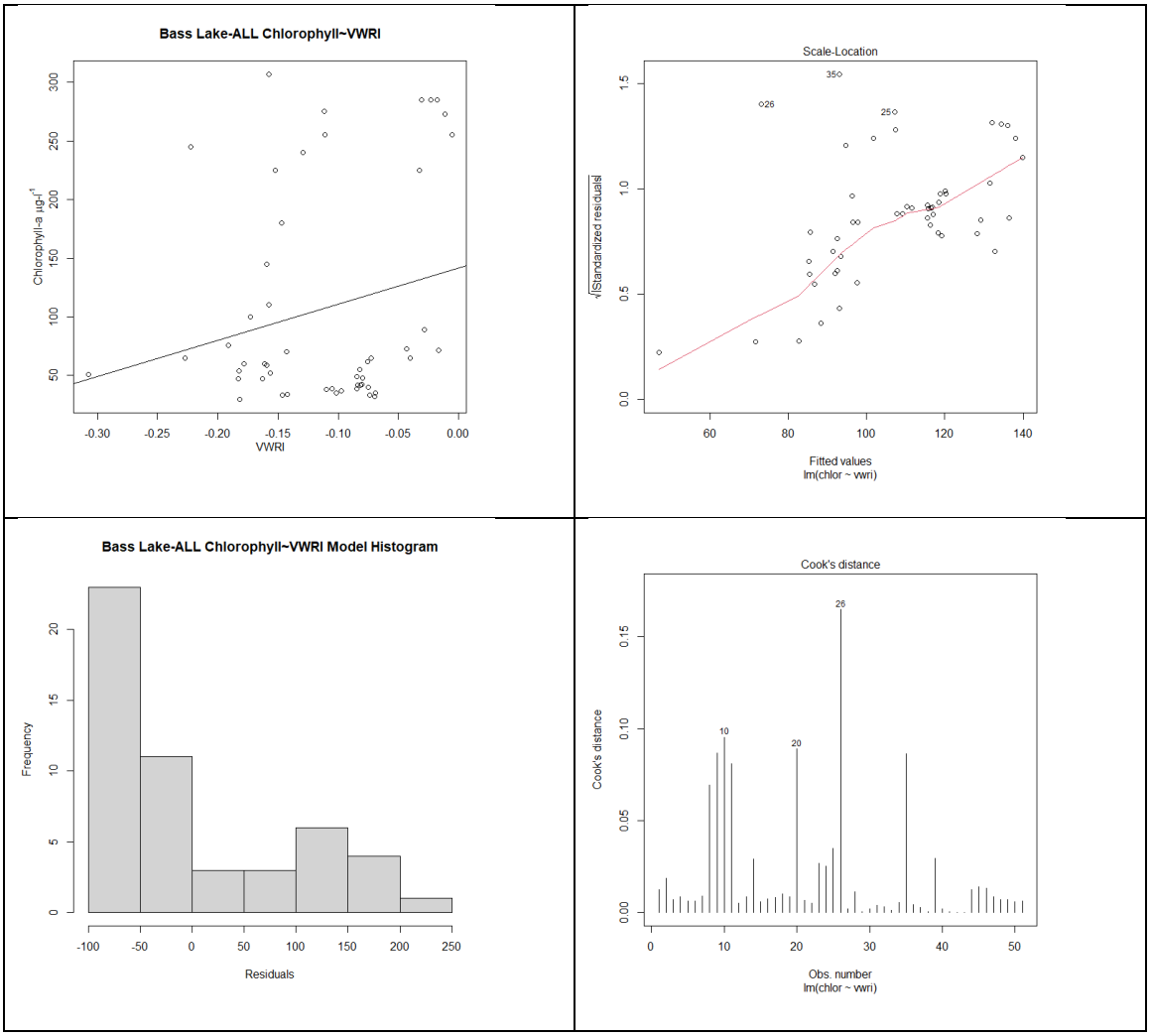

Figure 21. Top Left: Chlorophyll as a function of VWRI combining all 4 flights in the fall of 2020 at Bass Lake, Faribault County, MN with linear best fit; $R^2 = 0.130$. Upper right: scale location plot to test for homoscedasticity, Lower left: histogram to check for normality, Lower Right: Cook's distance (test for outliers).Top Right: and Lower: regression diagnostic plots.

All of the regression results from the 5 flights, including R_{adj}^2, p values, model equations, term estimates for regression, and Shapiro-Wilks normality test results are shown in Table 9 - Appendix. While both VWRI and NDVI generally had positive correlations with chlorophyll-a, NDVI had a stronger relationship than VWRI in three out

of the five flights and the combined dataset, and VWRI was the only of the two metrics to contain flights with non-significant p values (> 0.05) or near zero R^2 (in the combined dataset). Of the 5 flights, Little Rock Lake had a worse performing model than the Bass Lake flights.

4.1.3.1 Spatial Patterns and Stitched Imagery

Samples were collected in Little Rock Lake using Kayak only (Fig. 22). While the chlorophyll-a concentrations estimated by the NDVI (Fig. 25) and VWRI (Fig. 26) models at Little Rock Lake was low (ranging from 0-63 µg/L), a visible green scum was present on much of the northern shore of the lake, especially the center and eastern side of the flight area. VWRI (Fig. 23) ranged from -0.45 to 0.67, and NDVI from -3.68-1.02 (Fig. 24). VWRI, NDVI, and chlorophyll-a maps suggest that algae was not as abundant on the northwestern portion of the monitored area. The area south, and west of the boat ramp but near the shore had higher algal abundance, while near the shore east of the boat ramp had lower chlorophyll-a according to the NDVI model, and areas of higher chlorophyll-a off-shore to the east.

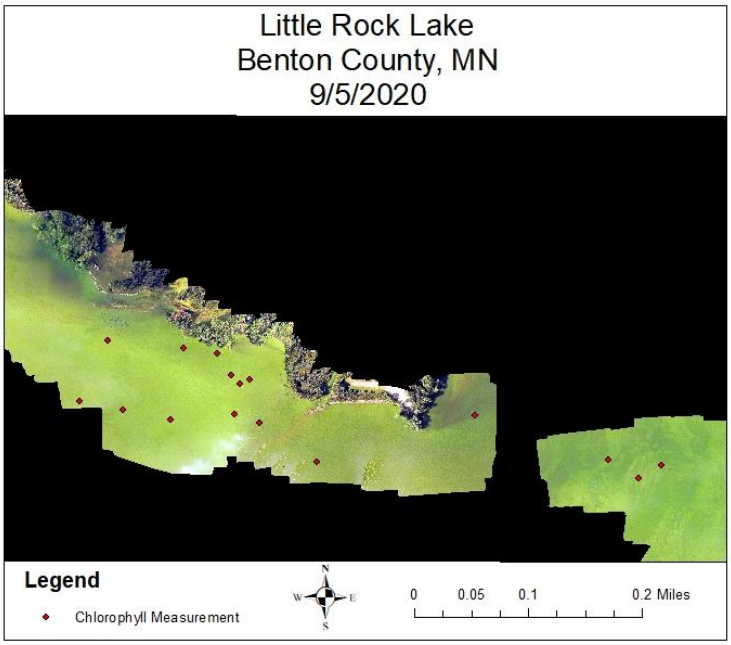

Figure 22. Little Rock Lake YSI sonde chlorophyll measurement points. A phantom 4 was used to fly over the northern shore of Little Rock Lake on 9/5/2021, and Drone Deploy was used to stitch imagery together.

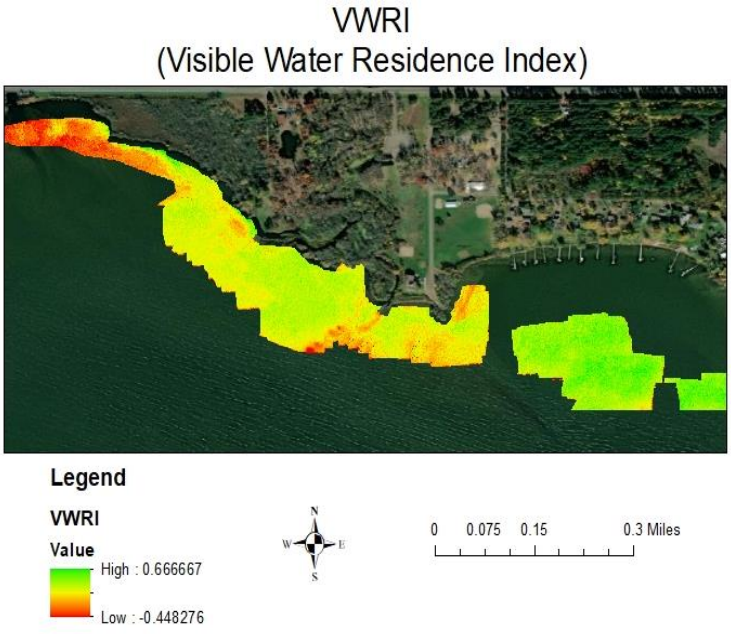

Figure 23. Little Rock Lake VWRI on 9/5/2021.

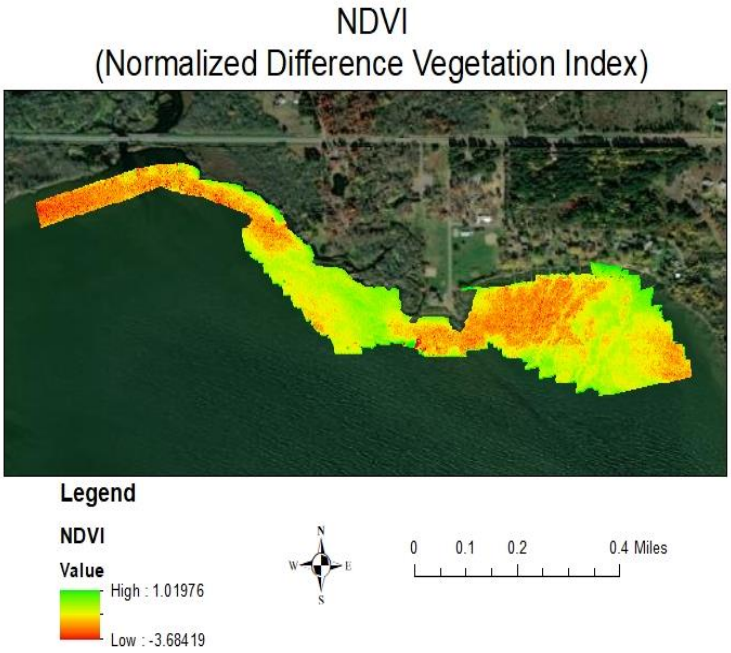

Figure 24. Little Rock Lake NDVI on 9/5/2021.

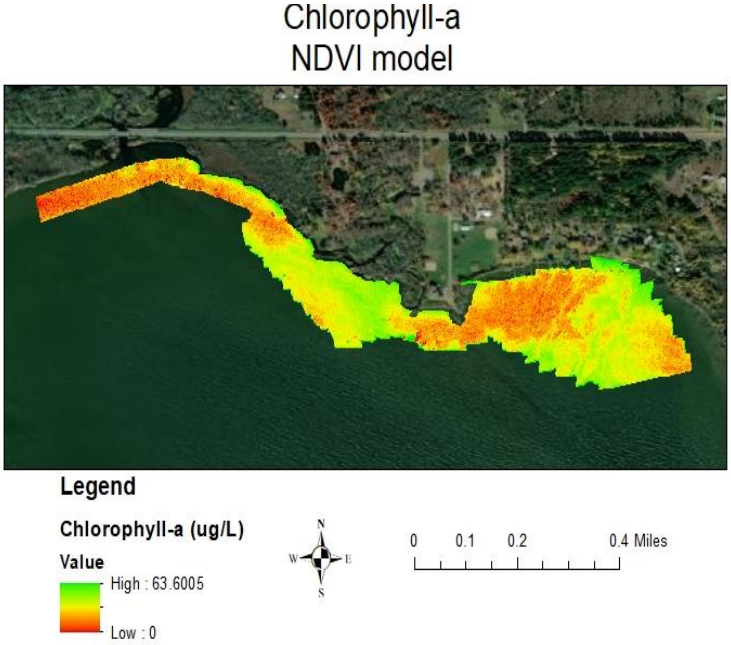

Figure 25. Little Rock Lake, Chlorophyll-a (ug/L) modeled from NDVI, 9/5/2021.

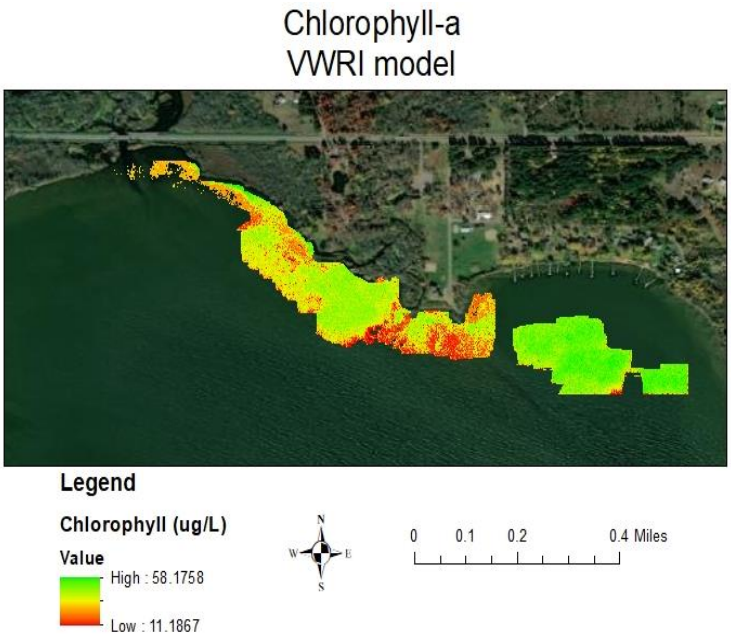

Figure 26. Little Rock Lake, Chlorophyll-a (ug/L) modeled from VWRI, 9/5/2021.

On 9/14/20, the UAV only flew over a small portion of the lake (the northeastern corner (Fig. 27). Samples were collected using a combination of sampling rod and kayak. NDVI ranged from -4.27 to 1.025 (Fig. 28), while VWRI ranged from -0.6 to 1.0 (Fig. 29). Both models of chlorophyll showed a hypereutrophic area near the shore, with maximum chlorophyll-a concentrations estimated above 300 ug/L. However, the NDVI maps and models (Fig. 28 and Fig. 30) showed a clear band of high but lower concentrations (around 100 ug/L) further offshore, and finally low concentrations further offshore (0-50 ug/L), with a patch of dense algae further east. The VWRI map and chlorophyll-a model map (Fig. 29 and Fig. 31) showed more variability in modeled chlorophyll-a further offshore, with more patches of high concentration of chlorophyll-a offshore compared to the NDVI map.

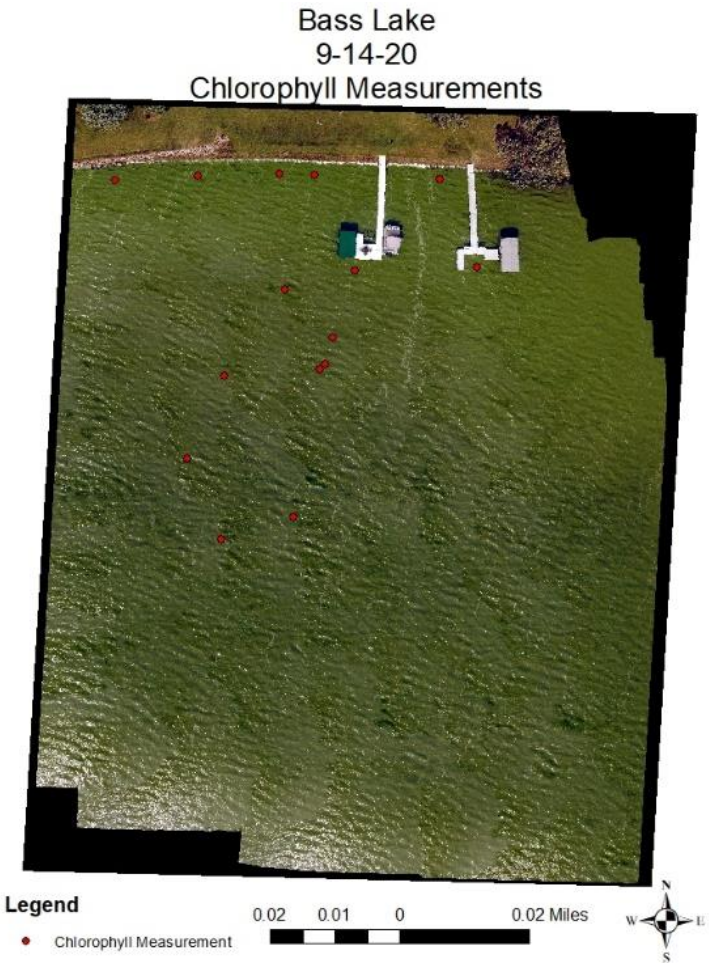

Figure 27. Bass Lake, 9/14/2020, chlorophyll-a sampling point locations.

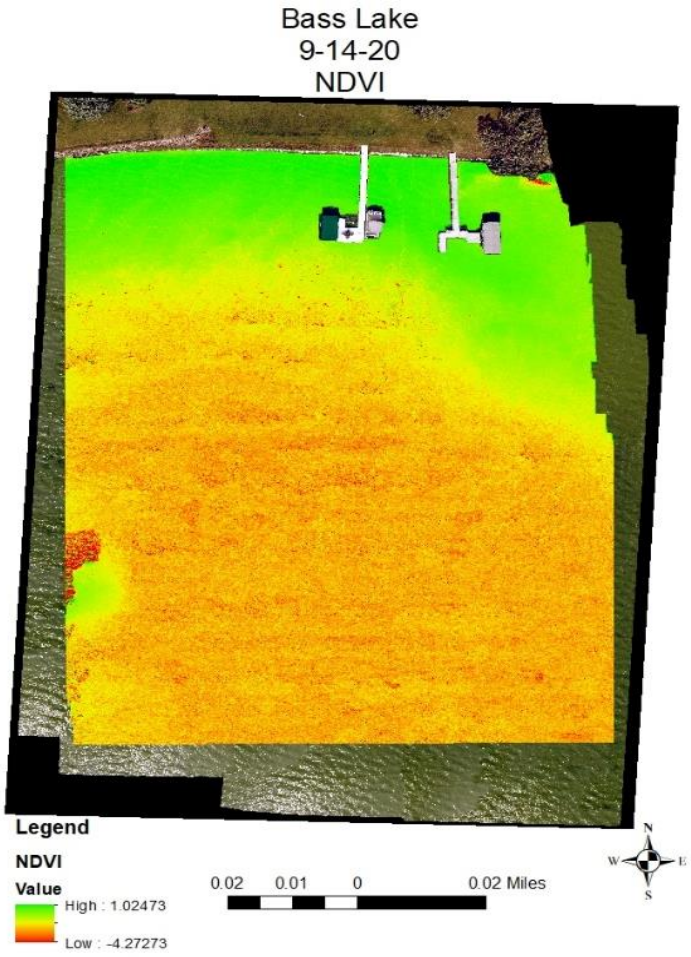

Figure 28. Bass Lake, 9/14/2020, NDVI.

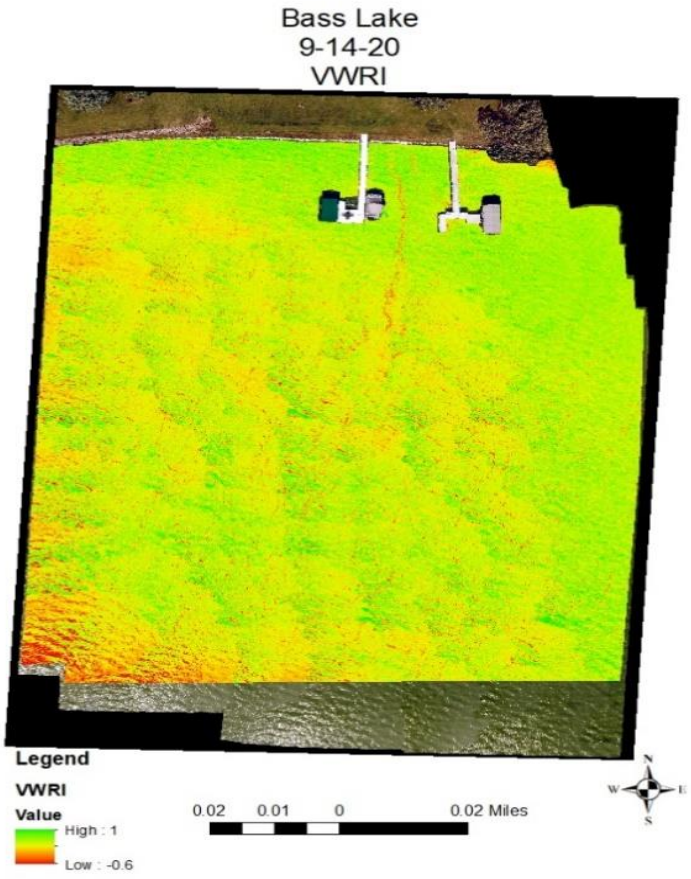

Figure 29. Bass Lake, 9/14/2020, VWRI.

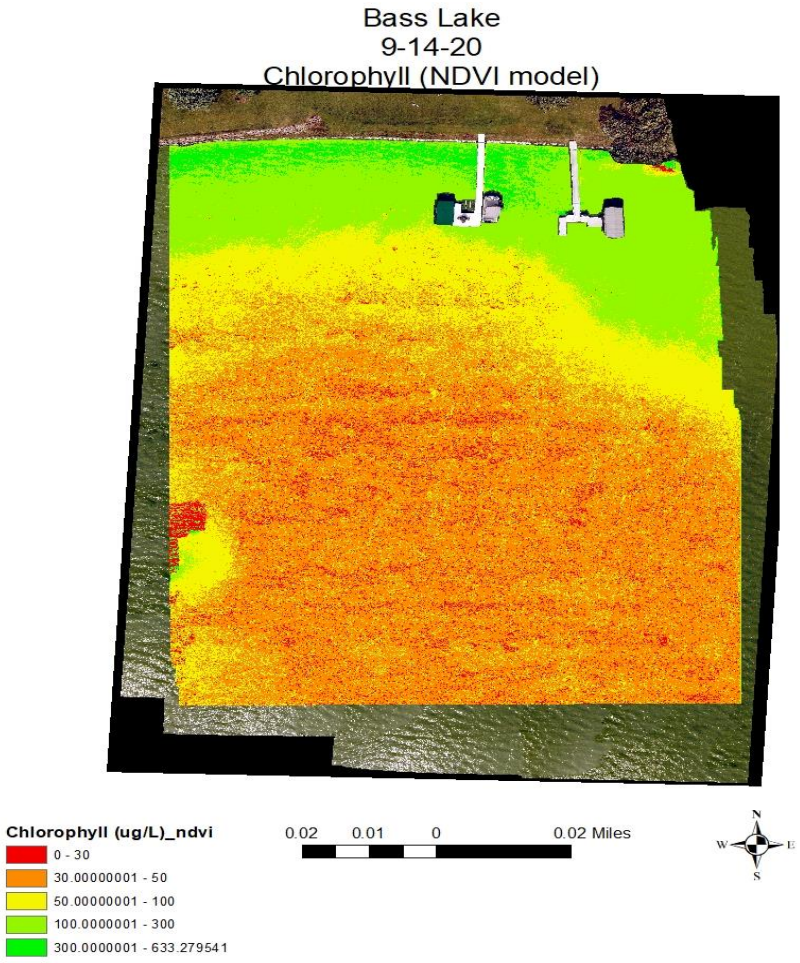

Figure 30. Bass Lake, 9/14/2020, chlorophyll-a estimated using best fit NDVI model.

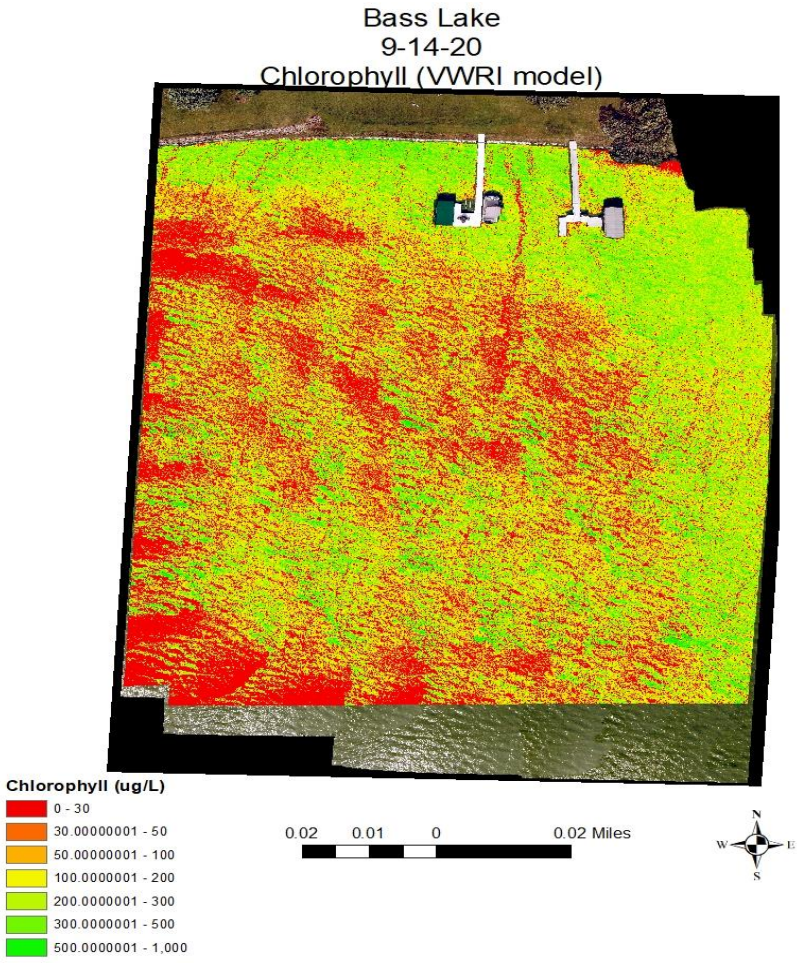

Figure 31. Bass Lake, 9/14/2020, chlorophyll-a estimated using best fit VWRI model.

Sampling on 9/18/2020 was again performed using a combination of sampling rod and kayak (Fig. 32). While the 2 previous flights showed high concentrations of chlorophyll-a, especially along the shore, patches of dense algae were less common. NDVI ranged from -4.27 to 0.89 (Fig. 33), and VWRI ranged from -0.83 to 0.67 (Fig. 34). The range of chlorophyll-a modeled by both NDVI (Fig. 35) and VWRI (Fig. 36) was comparable, with maximum chlorophyll-a of 63 ug/L for NDVI, and 65 ug/L for

VWRI. The NDVI chlorophyll-a model showed higher algal abundance on the northeastern portion of the lake near the shore than the monitored patch on the southwest side, and lower chlorophyll-a offshore. However, the VWRI model showed relatively high concentrations of chlorophyll-a offshore and throughout the monitored area compared to the NDVI model.

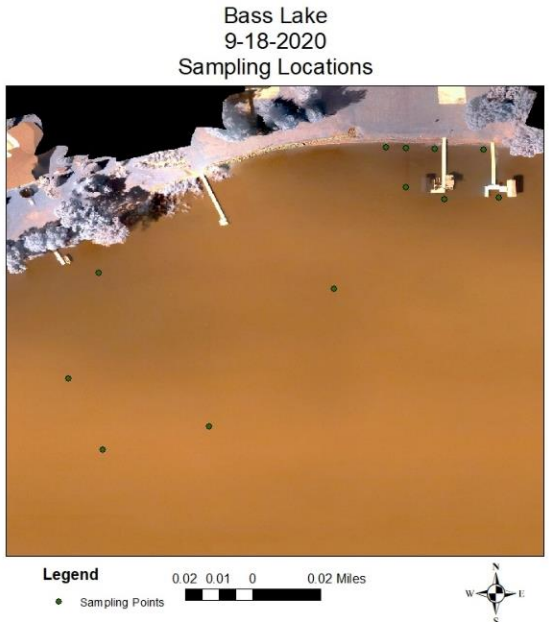

Figure 32. False color image, with YSI sonde chlorophyll measurement sampling locations at Bass Lake, on 9/18/2020.

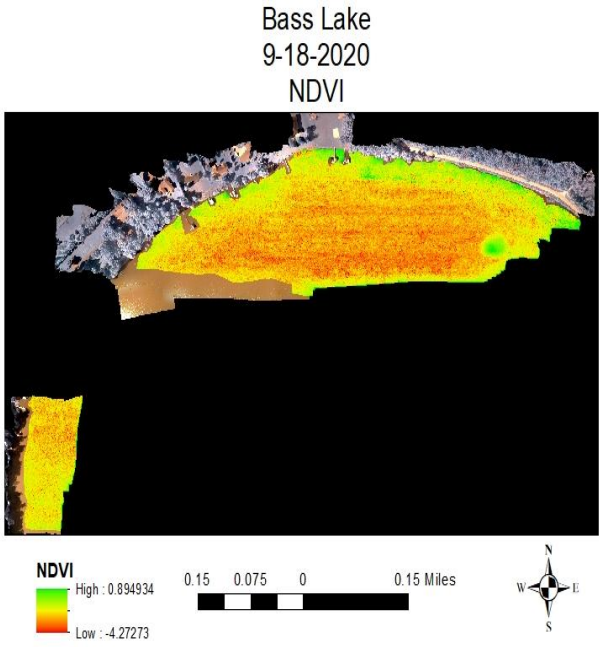

Figure 33. Bass Lake, 9/18/2020 NDVI.

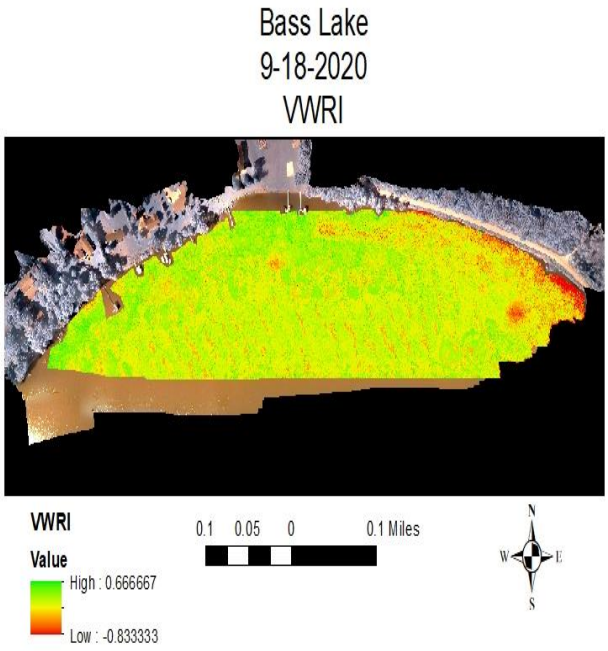

Figure 34. Bass Lake, 9/18/2020 VWRI.

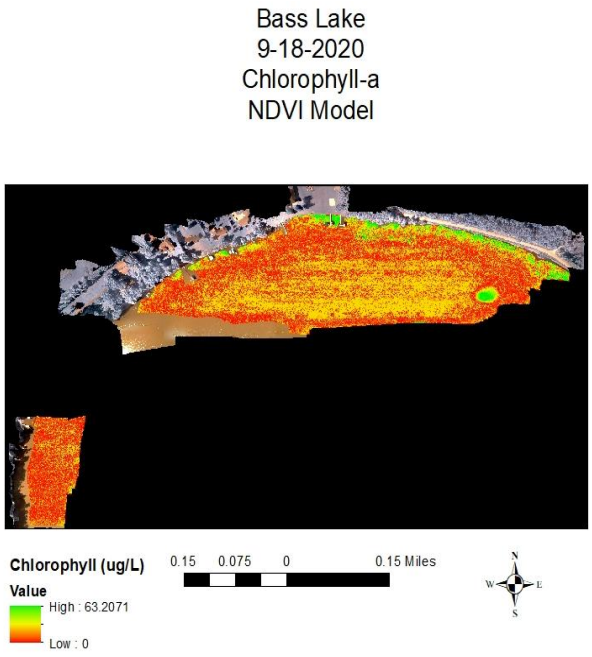

Figure 35. Bass Lake, 9/18/2020, Chlorophyll-a (ug/L) estimated from NDVI model.

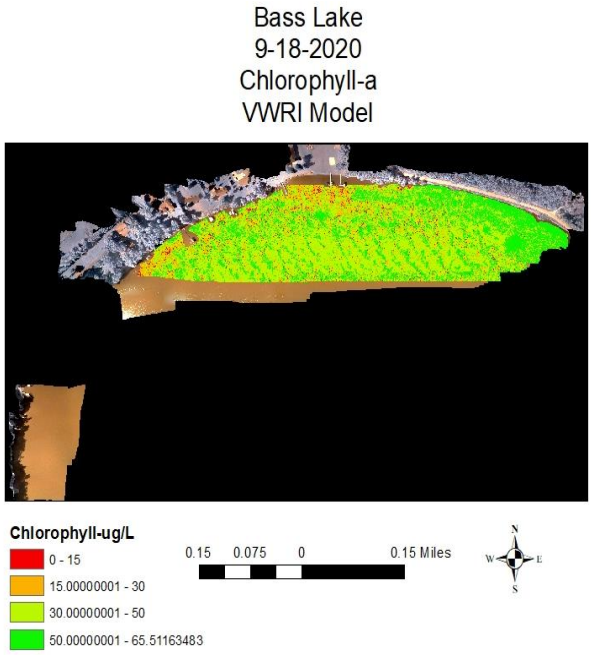

Figure 36. Bass Lake, 9/18/2020, Chlorophyll-a (ug/L) estimated from VWRI model.

Sampling on 9/26/20 also was carried out using a combination of sampling rod and kayaking. A visible surface scum was present along the shoreline in the Northeastern corner of the lake where sampling occurred (Fig. 37). NDVI ranged from -3.44 to 0.92 (Fig. 38) and VWRI from -0.91 to 0.79 (Fig. 39). Chlorophyll-a concentrations modeled from NDVI ranged from 0-325 ug/L (Fig. 40). The NDVI modeled chlorophyll-map highlighted the algae along the shore. The range of chlorophyll-a modeled by VWRI was from 0-3,000 ug/L. Although the concentrations modeled by VWRI were generally higher, they also highlighted the algae on the northern shore. However, the VWRI map appeared to show much more variability in algal abundance than the NDVI map.

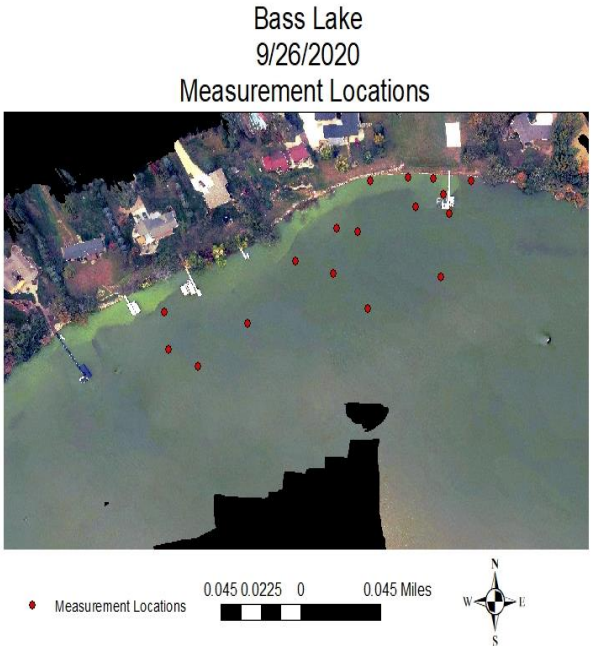

Figure 37. 9/26/2020, YSI sonde chlorophyll-a measurement locations in Bass Lake, Faribault County, MN.

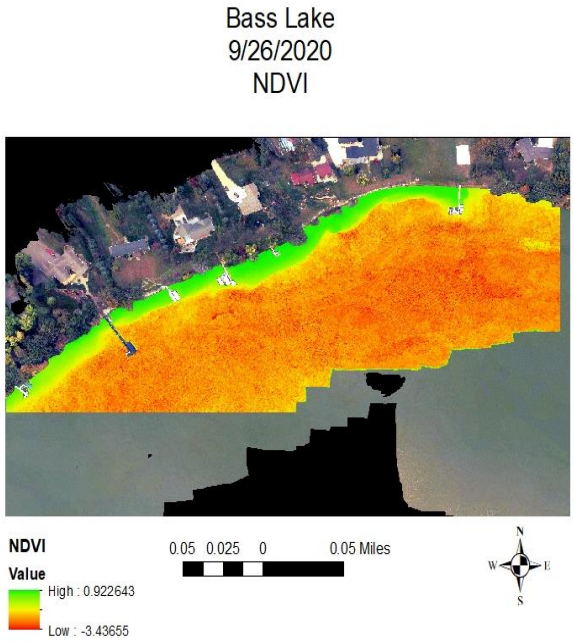

Figure 38. Bass Lake, 9/26/2020, NDVI.

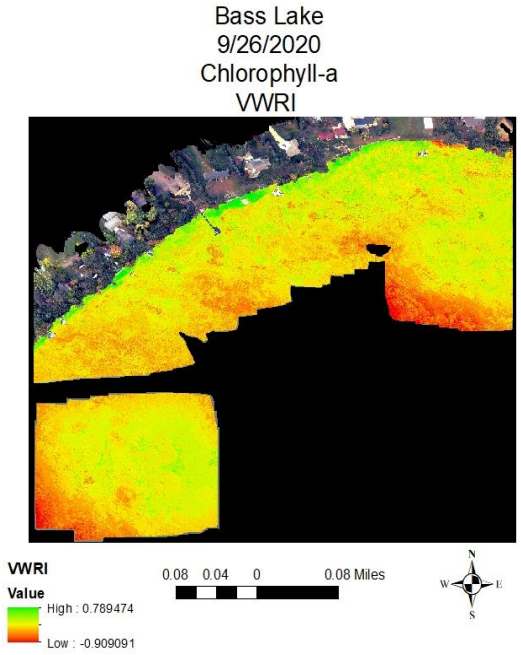

Figure 39. Bass Lake, 9/26/2020, VWRI.

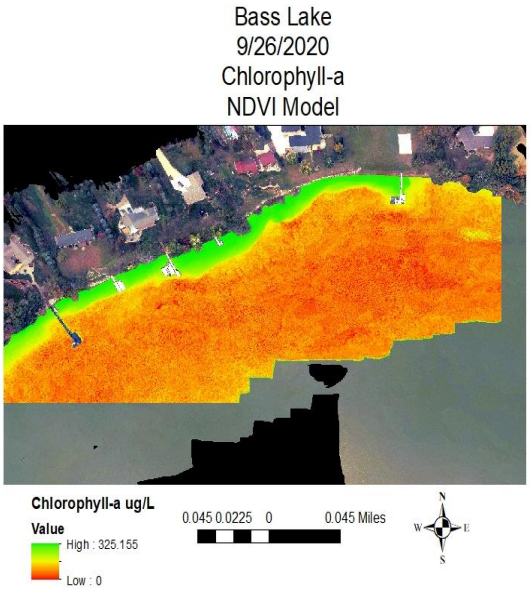

Figure 40. Bass Lake, 9/26/2020, Chlorophyll-a estimated from NDVI model. Lower area coverage than VWRI flight due to failure of drone deploy to stitch all imagery.

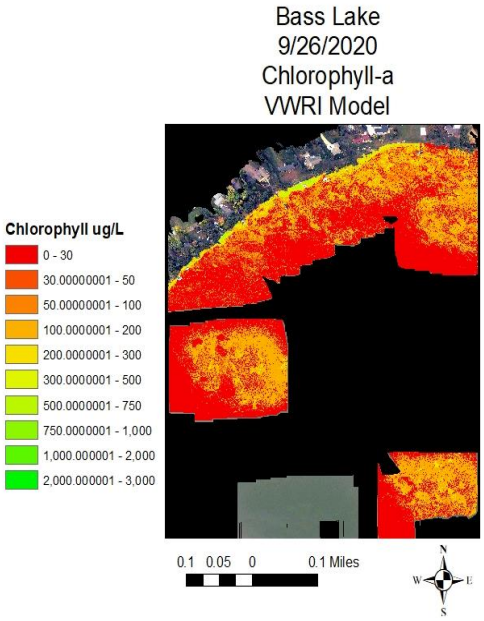

Figure 41. Bass Lake, 9/26/2020, Chlorophyll-a (µg/L) estimated using VWRI model. Although maximum chlorophyll-a was near 3,000 µg/L, mean chlorophyll-a was 63.2 µg/L with a standard deviation of 84.0 µg/L, showing that the extreme values only occurred for a small portion of the pixels.

On 10/8/2021, samples were only collected with a sampling rod, and a thick algal scum was again visible on the northern shore (Fig. 42). By flying at the maximum flight height of 121.9 meters, and replacing the drone batteries between flights, the Phantom 4 drone was able to map most of the lake within a period of roughly 1 hour, using 4 flights. NDVI ranged from -3.66 to 1.20 (Fig. 43), and highlighted higher algal concentrations along the shore of most of the lake. VWRI was again, more patchy than NDVI (Fig. 44), and ranged from -.85 to 0.63. The NDVI model estimated a range of chlorophyll-a from 0-133 ug/L (Fig. 45), while VWRI showed a range from 0-283 ug/L (Fig. 46). However, majority of the highest VWRI concentrations estimated were between 50-100 ug/L, and higher values were extremely rare and near impossible to observe on the map, making the concentrations relatively comparable between the 2 models. The VWRI model appeared to show thinner strips of algae along the shore than NDVI, in addition to more variability.

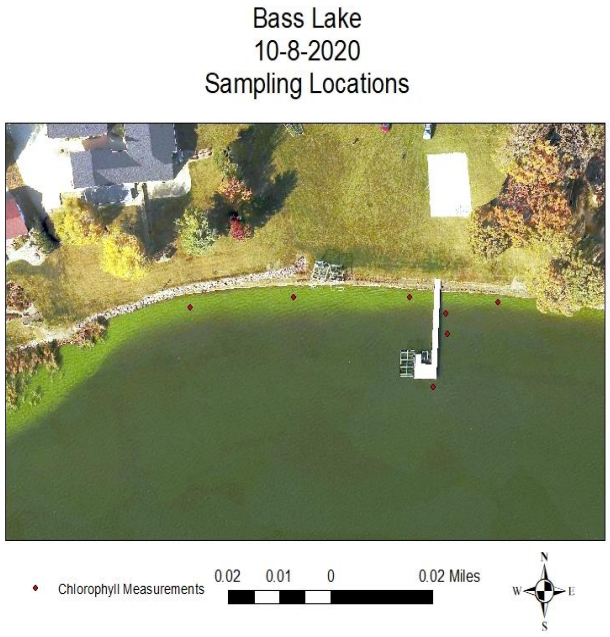

Figure 42. Bass Lake, 10/8/2020, sampling locations. Samples were collected using a 10-foot sampling rod from either the shore or dock.

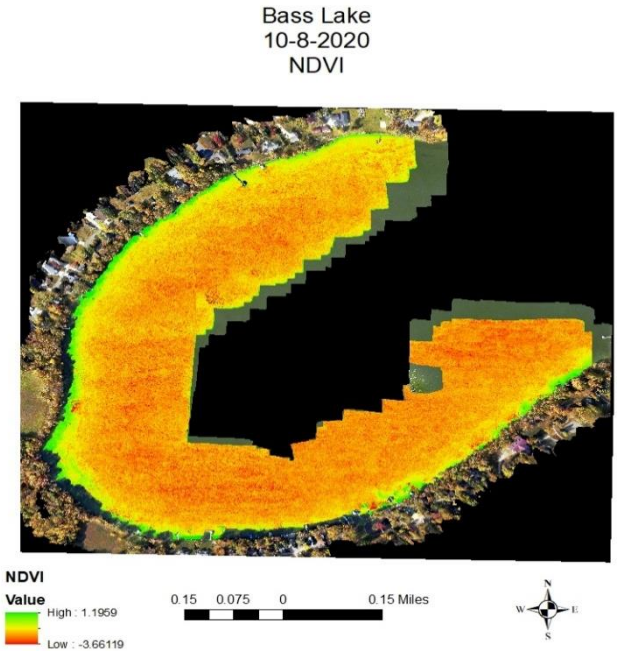

Figure 43. Bass Lake, 10/8/2020, NDVI.

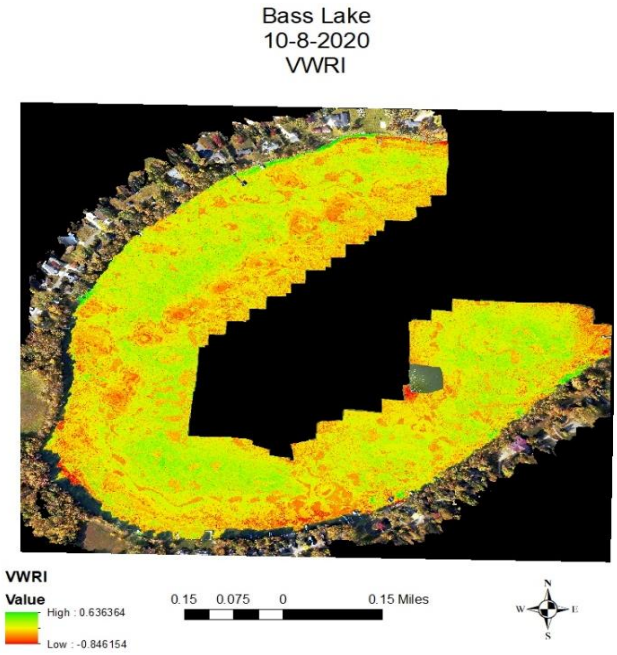

Figure 44. Bass Lake, 10/8/2020, VWRI.

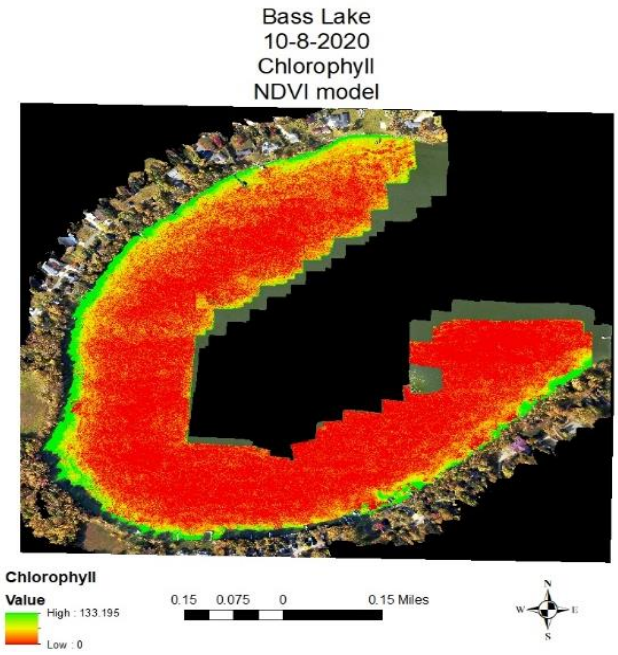

Figure 45. Bass Lake, 10/8/2020, Chlorophyll-a (ug/L) modeled using NDVI model.

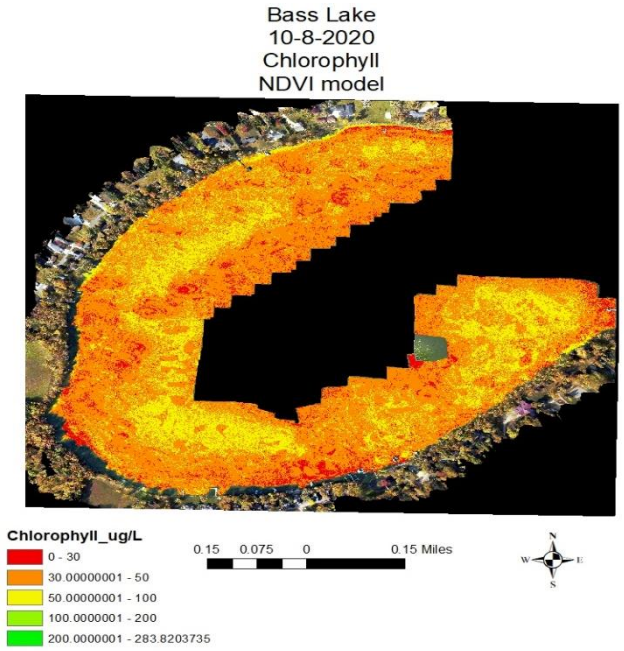

Figure 46. Bass Lake, 10/8/2020, Chlorophyll-a (ug/L) modeled using NDVI model.

4.2 Time Series Forecasting Results

4.2.1 Multivariate Parameter Selection

Stepwise deletion using multiple regression of the physical, chemical, and meteorological parameters found that the simplest model was predicting chlorophyll-a using turbidity ($p = 0.000143$), specific conductivity ($p = 2.90e\text{-}06$), water temperature ($p = 0.000756$), the past 60 days (2 months) of precipitation ($p = 6.02e\text{-}05$), with a R_{adj}^2 of 0.343 ($p = 1.36e\text{-}12$, table 1). Although the multiple regression failed the assumption of normality, due to a left skew, transformation was unable to improve the distribution. These 4 predictor variables were chosen as input variables along with past chlorophyll-a for forecasting chlorophyll-a using a multivariate forecast.

Table 1. Multiple regression results from stepwise regression of weather station variables (wind speed, direction, precipitation), and water quality variables (conductivity, water temperature, turbidity, pH) on chlorophyll-a measured at high frequency. Non-significant parameters were removed stepwise until the simplest model remained (chlorophyll ~ turbidity + conductivity + water temperature + 2months-precip).

Variable	Estimate	P value	Adjusted R^2
Turbidity	-0.00396	0.000143	
Specific Conductivity	-0.00399	2.90e-06	
Water Temperature	0.00481	0.000756	
Past 2 months precipitation	0.0267	6.02e-05	
Intercept	3.859	<2e-16	
Overall		1.36e-12	0.343

4.2.2 Hyperparameterization

The number of input days used in the multivariate LSTM model had an extremely variable effect on RMSE, with no clear trend (Fig. 47). 22 input days had the highest RMSE, and 5 input days and 30 input days had the lowest RMSE. Based on these results, input days was assumed to be a relatively unimportant hyperparameter. 10 days was selected as one of the troughs in the plot, to be held constant during the coarse grid scale searching. In contrast, the number of network nodes had a relatively dramatic effect with a clear trend, with the RMSE around 200 until 35 network nodes, when it increase and peaked at 45 nodes with a RMSE of approximately 1,400, and then declined but was still

high a RMSE of 600 at 50 network nodes (Fig. 48). The number of epochs showed small effect on RMSE similar to the number of input days, ranging from about 40-140 (Fig. 49). However, there was a more obvious trend of higher RMSE after 200 epochs, so the lower epoch values were examined during the coarse grid scale search. The number of batches showed a general trend of increasing RMSE as more batches were added (Fig. 50).

Based on these results, a coarse grid scale search using a range of 10-30 nodes with 1 or 2 hidden layers, 50 or 150 epochs, and 1 or 40 batches was chosen for coarse scale grid searching (see Table 12 and 13 - Appendix for full list of results). The top three models for the 2019 and 2020 datasets with either 1 or 2 hidden layers are shown in table 2. The top model for 2019 was 10, 20, 150, 40, while the top model for 2020 was 30, 20, 150, 40. Because the top model for 2019 was also the 4th ranked model for 2020 (a RMSE of 29.3 for 2020), this model was selected for both years for simplicity. 2 hidden layers had a greater performance (lower RMSE) than 1 hidden layer in general. The multivariate LSTM model was found to have a spike in RMSE in 2019 followed by a decline towards the 10th forecasting day (Fig. 51). However, in 2020, the trend followed the opposite pattern, with an increase in RMSE near the 10th forecasting day (Fig. 52).

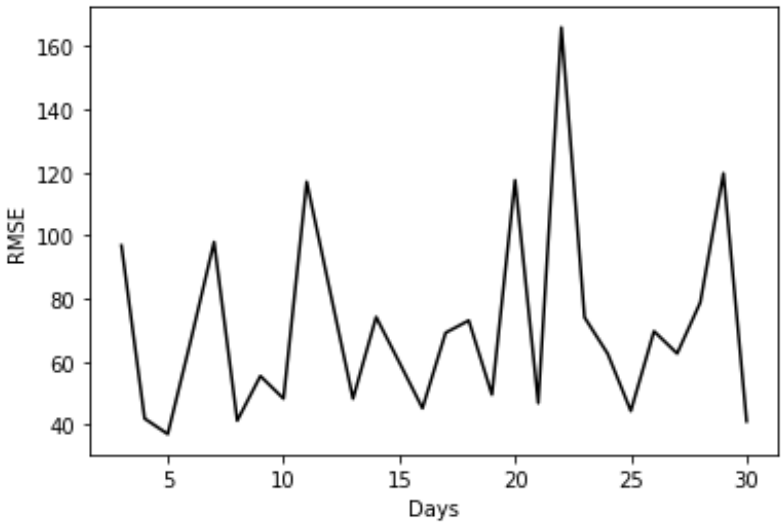

Figure 47. Impact of number of input days on RMSE (keeping other hyper parameters fixed) for 2019 chlorophyll-a forecast (tested on a multivariate LSTM). 10 repeats per configuration. RMSE in an average across all models (separate model for days 1 – 10).

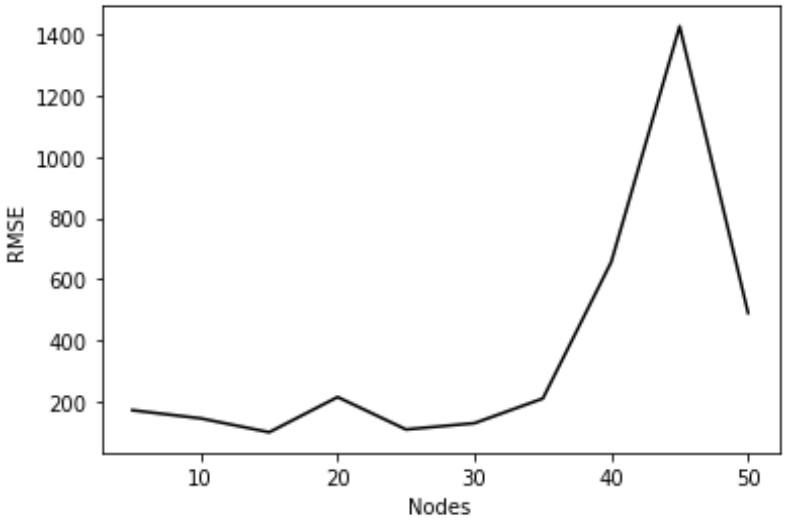

Figure 48. Impact of number of network nodes on RMSE (keeping other hyper parameters fixed) for 2019 chlorophyll-a forecast (tested on a multivariate LSTM forecast). 10 repeats per configuration. RMSE in an average across all models (separate model for days 1 – 10).

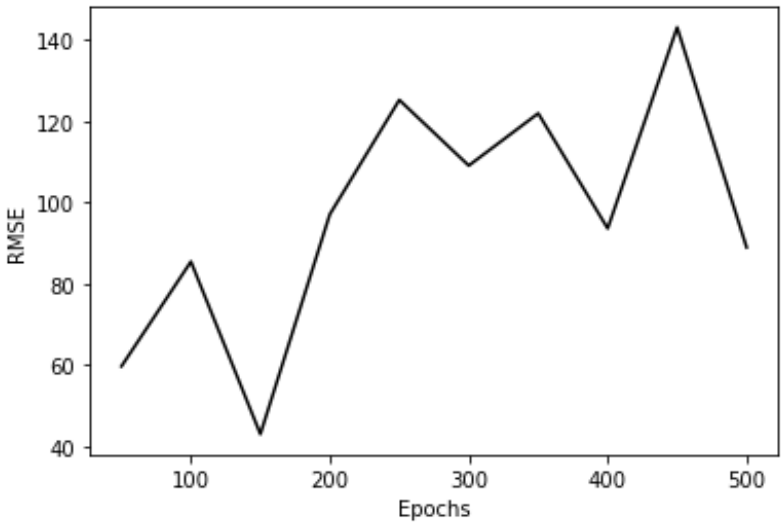

Figure 49. Impact of number of epochs on RMSE (keeping other hyper parameters fixed) for 2019 chlorophyll-a forecast (tested on a multivariate LSTM forecast). 10 repeats per configuration. RMSE in an average across all models (separate model for days 1 – 10).

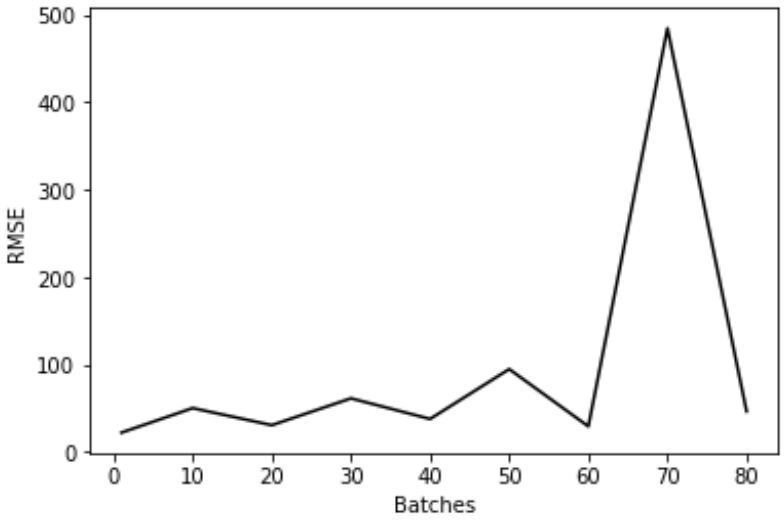

Figure 50. Impact of number of Batches on RMSE (keeping other hyper parameters fixed) for 2019 chlorophyll-a forecast (tested on a multivariate LSTM forecast). 10 repeats per configuration. RMSE in an average across all models (separate model for days 1 – 10).

Table 2. Top three models (lowest RMSE) for coarse scale grid search testing impact of hidden layers, epochs, batches on average model RMSE for multivariate chlorophyll-a forecast using LSTM. 2019 and 2020 data were tested separately. 12 or 36 configurations were tested for 1 hidden layer and 2 hidden layers, respectively, with three repeats per configuration.

Hidden Layers	Year	Nodes-1	Nodes-2	Epochs	Batches	RMSE
2	2019	10	20	150	40	40.45046
2	2019	20	20	150	1	40.50747
2	2019	10	20	50	1	42.43067
2	2020	30	20	150	40	13.21203
2	2020	30	30	50	1	22.5401
2	2020	10	10	50	40	27.72061
1	2019	20	-	50	1	78.28478
1	2019	10	-	50	40	114.074
1	2019	30	-	50	40	131.0668
1	2020	10	-	50	40	42.9673
1	2020	30	-	50	40	78.41266
1	2020	20	-	150	40	107.2038

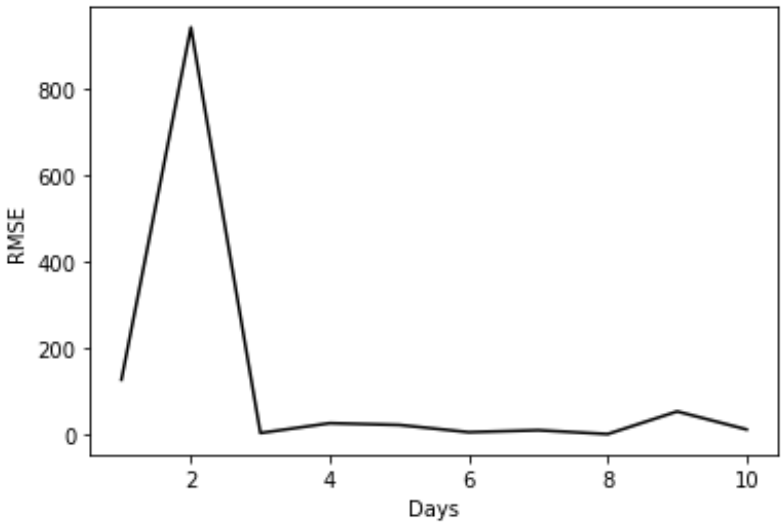

Figure 51. RMSE with top model configuration for 2019 chlorophyll-a data for forecasting days 1-10, using multivariate LSTM forecast.

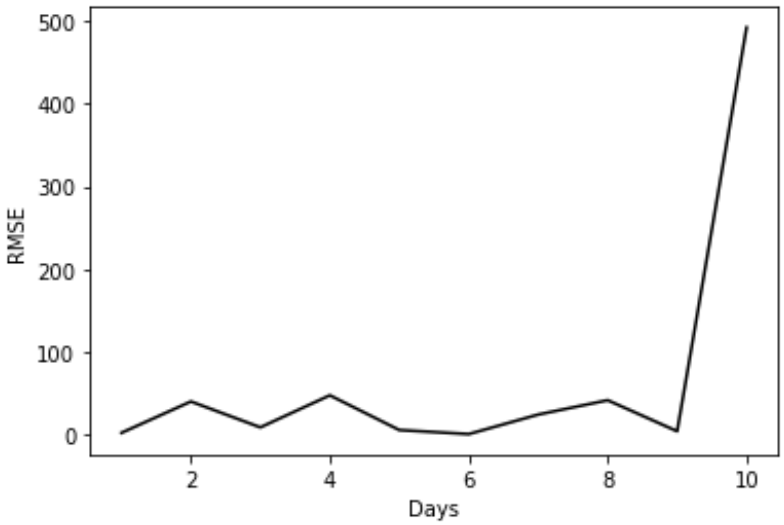

Figure 52. RMSE with top model configuration for 2020 chlorophyll-a data for forecasting days 1-10, using multivariate LSTM forecast.

4.2.3 Model Comparisons

The 10-day multivariate forecast showed large spikes and troughs in the chlorophyll-a forecast in both 2019 and 2020 (Fig. 53, Fig. 57). The RMSE for the 2019 and 2020 multivariate forecasts were 9.702 and 2.791, respectively (table 3). The univariate LSTM forecast showed a smoother forecast than the multivariate forecast (Fig. 54, Fig. 58), and had a lower RMSE in both 2019 and 2020: 1.197 and 1.767, respectively (table 3). While the RMSE was higher for 2019 than 2020 for the multivariate forecast, the univariate forecast had its lowest RMSE in 2019. Wavelet-LSTM had the highest RSME of all the models (Fig.55, Fig. 59). ARIMA had the lowest RMSE of the three model approaches, with a RMSE of 1.160 in 2019 (Fig. 56), and 0.936 in 2020 (Fig. 60) (table 3). However, ARIMA predicted a flat line, due to auto.arima selecting the optimum autoregressive parameter as 0 (in 2020), and therefore there was little variability in the ARIMA point forecast besides the error range around the mean point forecast (Fig. 56, Fig. 60).

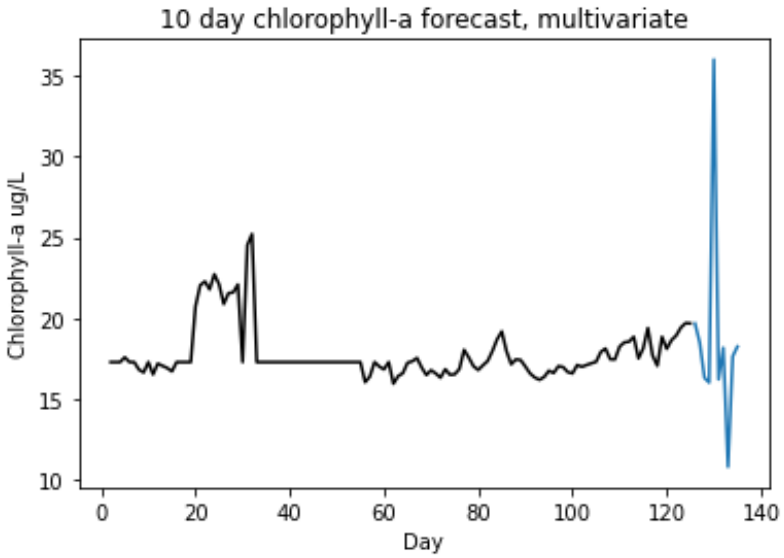

Figure 53. 10 day forecast for multivariate LSTM, for 2019 chlorophyll-a data. Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median.

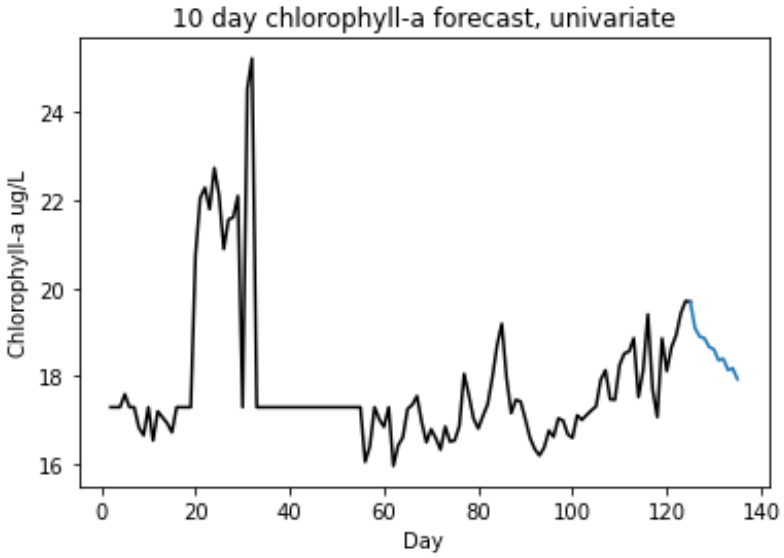

Figure 54. 10-day chlorophyll-a forecast using univariate LSTM, 2019 data from Bass Lake. Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median.

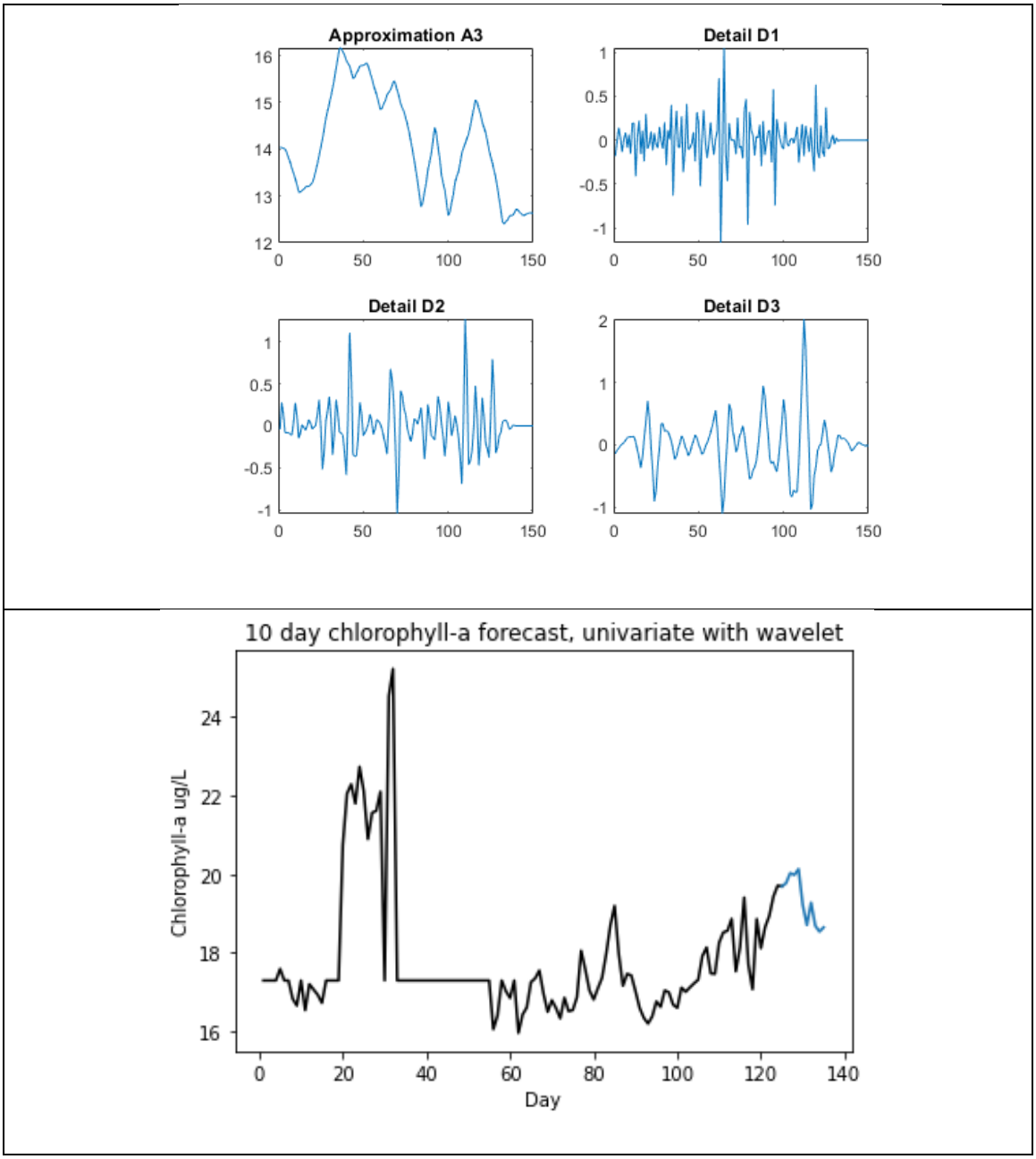

Figure 55. Wavelet–LSTM transformation and forecast. Upper: DB3 transformation, A3 approximation series, and D1, D2, D3 detail series for chlorophyll-a from Bass Lake. Lower: 2019, and 10-day forecast; black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median).

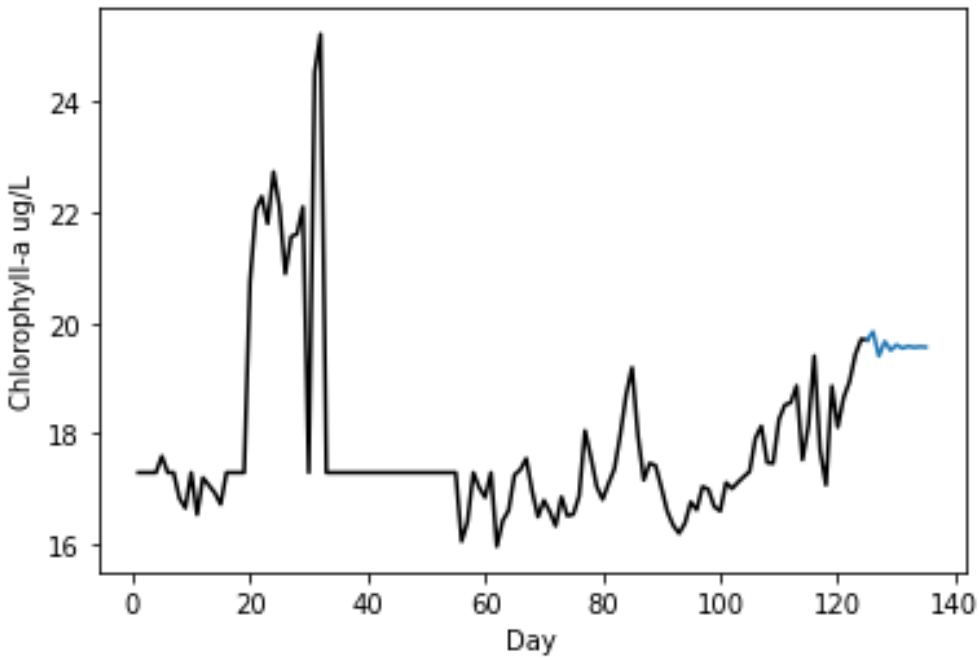

Figure 56. 10-day chlorophyll-a forecast using ARIMA, 2019 data from Bass Lake. ARIMA order 1,1,2. Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median).

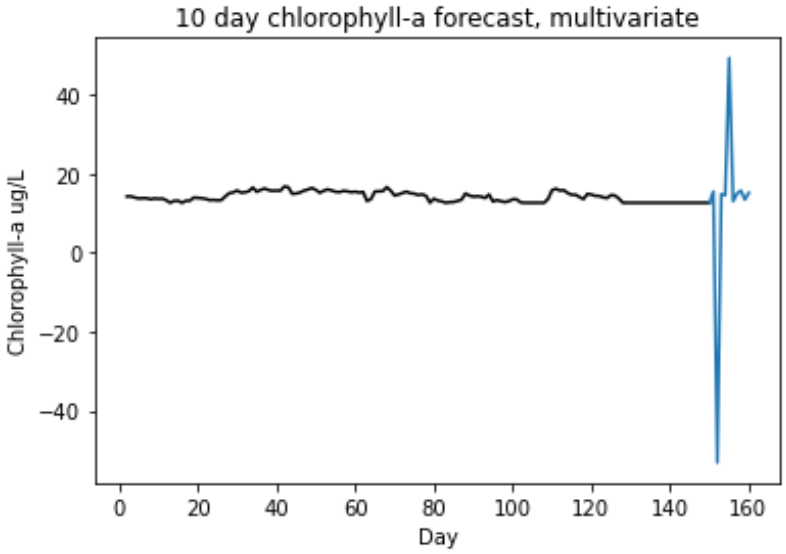

Figure 57. 10 day forecast for multivariate LSTM, for 2020 chlorophyll-a data. Black – actual data, blue – forecast. Flat regions show periods with missing data (imputed to the median).

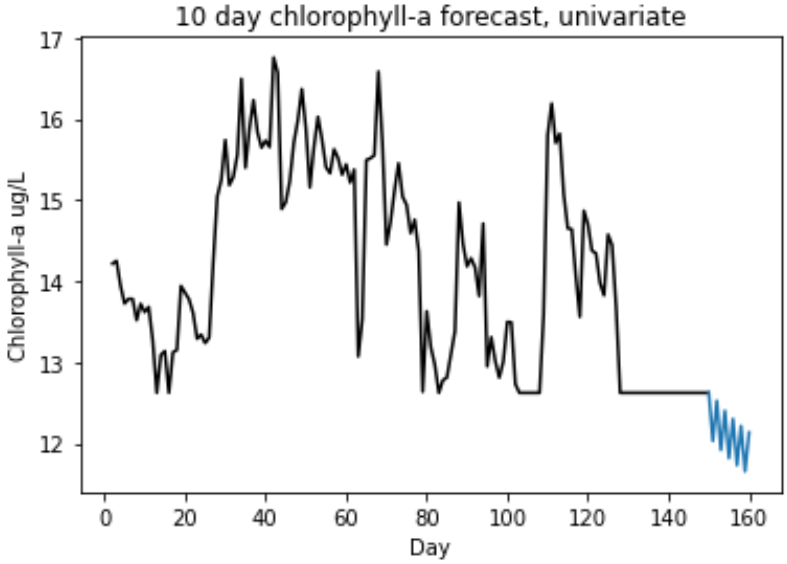

Figure 58. 10-day chlorophyll-a forecast, univariate, 2020. Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median).

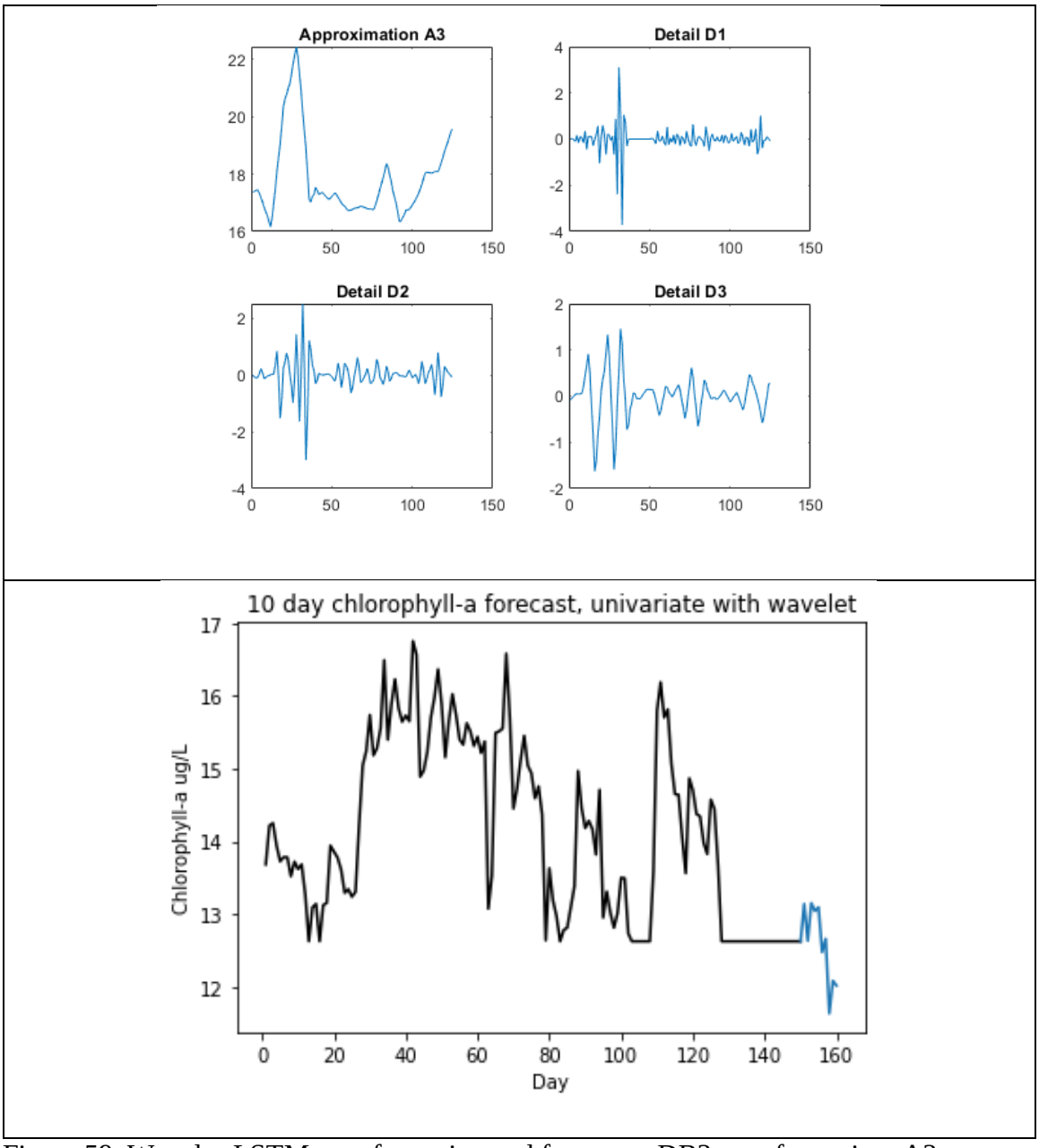

Figure 59. Wavelet-LSTM transformation and forecasts, DB3 transformation, A3 approximation series, and D1, D2, D3 detail series from chlorophyll-a From Bass Lake, 2020, and 10-day forecast (blue). Flat regions show periods with missing data (imputed to the median).

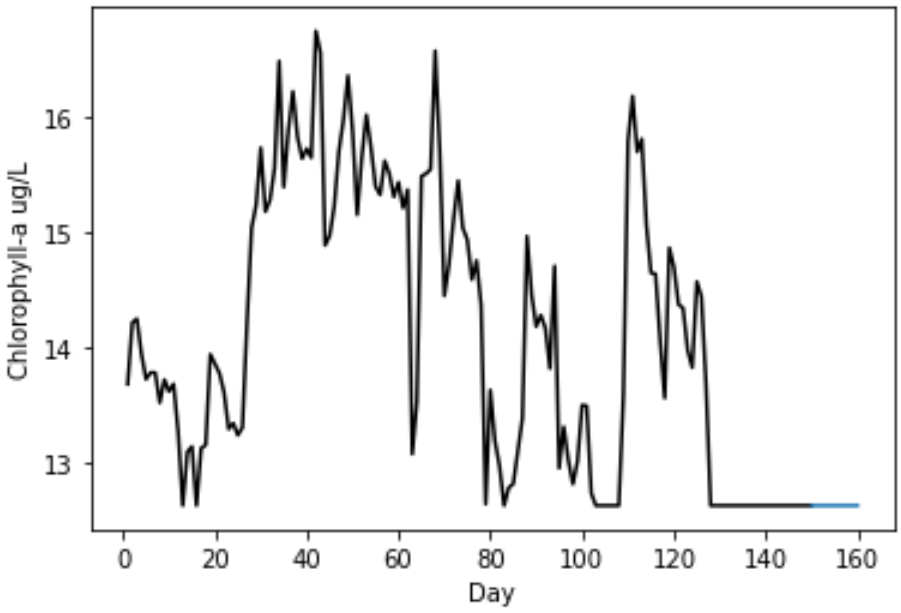

Figure 60. 10-day chlorophyll forecast, ARIMA, 2020. ARIMA Order 0,1,2. Top: forecast from area with missing data (imputed to the median). Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median).

Table 3. Comparison of RMSE between LSTM (multivariate and univariate), and ARIMA.

Method	RSME		Cost	Benefits
	2020	2019		
LSTM, multivariate	9.702	2.791	Poor RMSE Spikes in data	Takes correlated variables into account
LSTM, univariate	1.197	1.767		Good RMSE Reasonable long term forecast

ARIMA	1.160	0.936	Long term forecast seems unlikely (flat)	Good RMSE
Wavelet-LSTM (univariate), DB3 wavelet	44.0	8.46	Worst performance	May improve with better hyperparameters or alternative wavelet form

4.3 Impact of Interannual Variability of Meteorological Factors at Bass Lake

A multiple regression of maximum temperature, and 1-week, 2-week, 30-day, and 60-day precipitation totals for predicting median August chlorophyll-a concentrations from Sentinel-2 satellite imagery for Bass Lake showed 2 variables to be significant in the simplest model following stepwise deletion: 2 week precipitation totals, and 60 day precipitation totals (Fig. 61, table 4). The past 2 weeks precipitation total was the stronger of the 2 relationships, with a positive correlation and the lower of the p values ($p = 0.0158$). The 60 day precipitation total was negatively correlated to median chlorophyll-a (0.0301). The overall model had a R_{adj}^2 of 0.363 ($p = 0.042$). The relationship between chlorophyll-a and past 2 weeks of precipitation appears linear in the pair plot (Fig. 61), and while a negative correlation does appear possible in the 60-day precipitation plot, there appears to be more variability in the graph with 2 possible outlier with chlorophyll-

a greater than 40 ug/L potentially driving the negative correlation (Fig. 61, upper left plot).

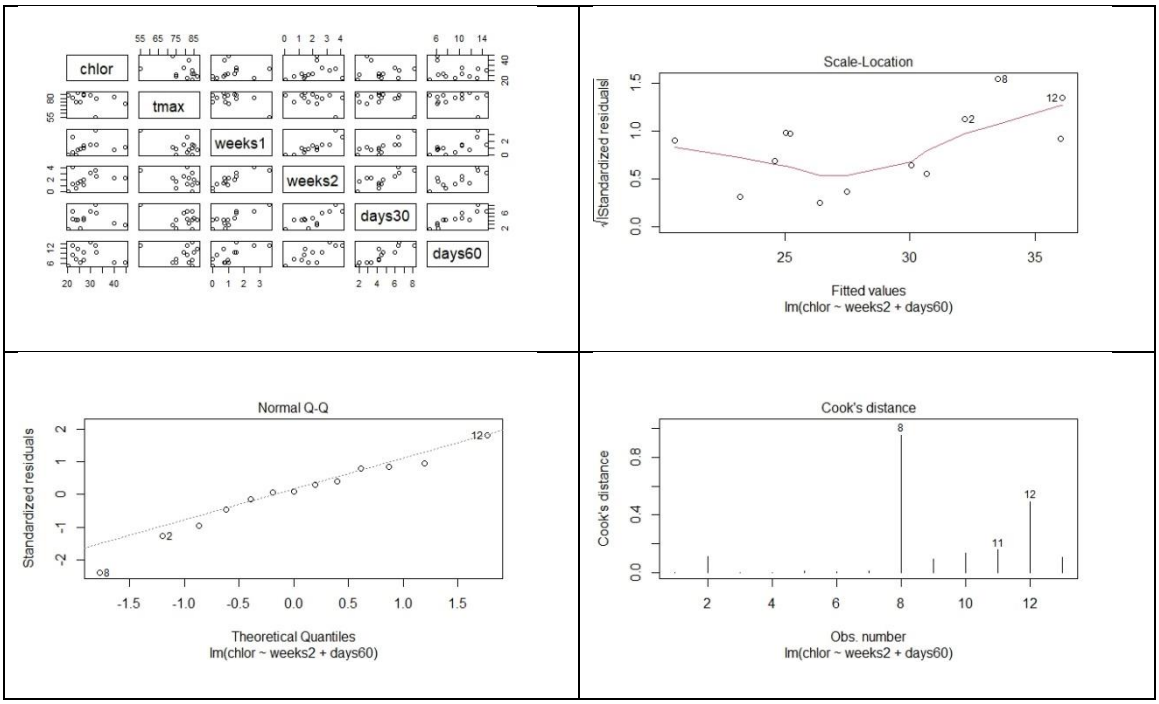

Figure 61. Multiple Regression of chlorophyll-a vs. meteorological factors at Bass Lake, and regression diagnostics. Average chlorophyll is from Sentinel 2; predictors include maximum temperature (tmax), and the past 7, 14, 30, and 60 days of precipitation.

Table 4. Multiple regression results for impact of precipitation on average chlorophyll-a at Bass Lake from 2017-2020. Stepwise removal of least significant terms was carried out until the simplest model (including 2-weeks and 60-days precipitation totals as the factors) remained.

	Weeks2 precipitation	Days60 precipitation	Overall	Intercept
P value	0.0158	0.0301	0.042	6.51*10e-05
Estimate	5.9455	-1.8908		
Adjusted R^2			$R^2 = 0.363$	

4.3.1 Meteorological Impacts on Chlorophyll-a in Western Corn Belt Plains

After stepwise deletion of non-significant variables, multiple regression of precipitation, temperature, and snowfall variables in predicting median chlorophyll-a during August in 160 Minnesota Lakes in the Western Corn Belt Plains showed that the 2019 model was left with only 2 variables, AN19 and JJA19, both of which become non-significant after stepwise deletion was complete ($p > 0.05$ for each variable). The overall p value for the 2019 model was also not significant ($p = 0.0530$, $R_{adj}^2 = 0.025$, table 5, Fig. 62). However, for the 2020 model, MAM20 ($p = 0.0192$, table 6) and JJA20 were both significant ($p = 0.0014$), 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), suggesting that the model failed to explain much of the

variability in chlorophyll-a (Fig. 62). While a slight negative trend is visible especially in the 2020 data, multiple lakes used the same meteorological data due to proximity to the data source, but contained a high variability in chlorophyll-a concentrations.

Table 5. Multiple regression results from 160 lakes in 2019, impact of precipitation and temperature on chlorophyll-a. Only 2 predictor variables were significant when all predictors were included (annual precipitation in 2019 AN19, and June-July-August precipitation in 2019 JJA19 ($p < 0.05$), but the correlations were not significant after removing the other non-significant predictors). $R^2 = 0.025$, $P = 0.0530$.

Variable	P value	Estimate	Adjusted R^2
Intercept	0.00497	188.732	
AN19	0.319	-3.441	
JJA19	0.713	-1.441	
Overall	0.0530		$R^2 = 0.025$

Table 6. Multiple regression results from 160 lakes in 2020, impact of precipitation and temperature on chlorophyll-a. After stepwise deletion of non-significant variables, the simplest model of chlorophyll-a in 2020 included March-April-May precipitation in 2020 MAM20, and June-July-August precipitation in 2020 JJA20 ($p < 0.05$). $R^2 = 0.0519$, $P = 0.00596$.

Variable	P value	Estimate	Adjusted R^2
Intercept	0.0012	415.888	
MAM20	0.0192	-26.905	
JJA20	0.0014	-6.014	
Overall	0.00596		0.0519

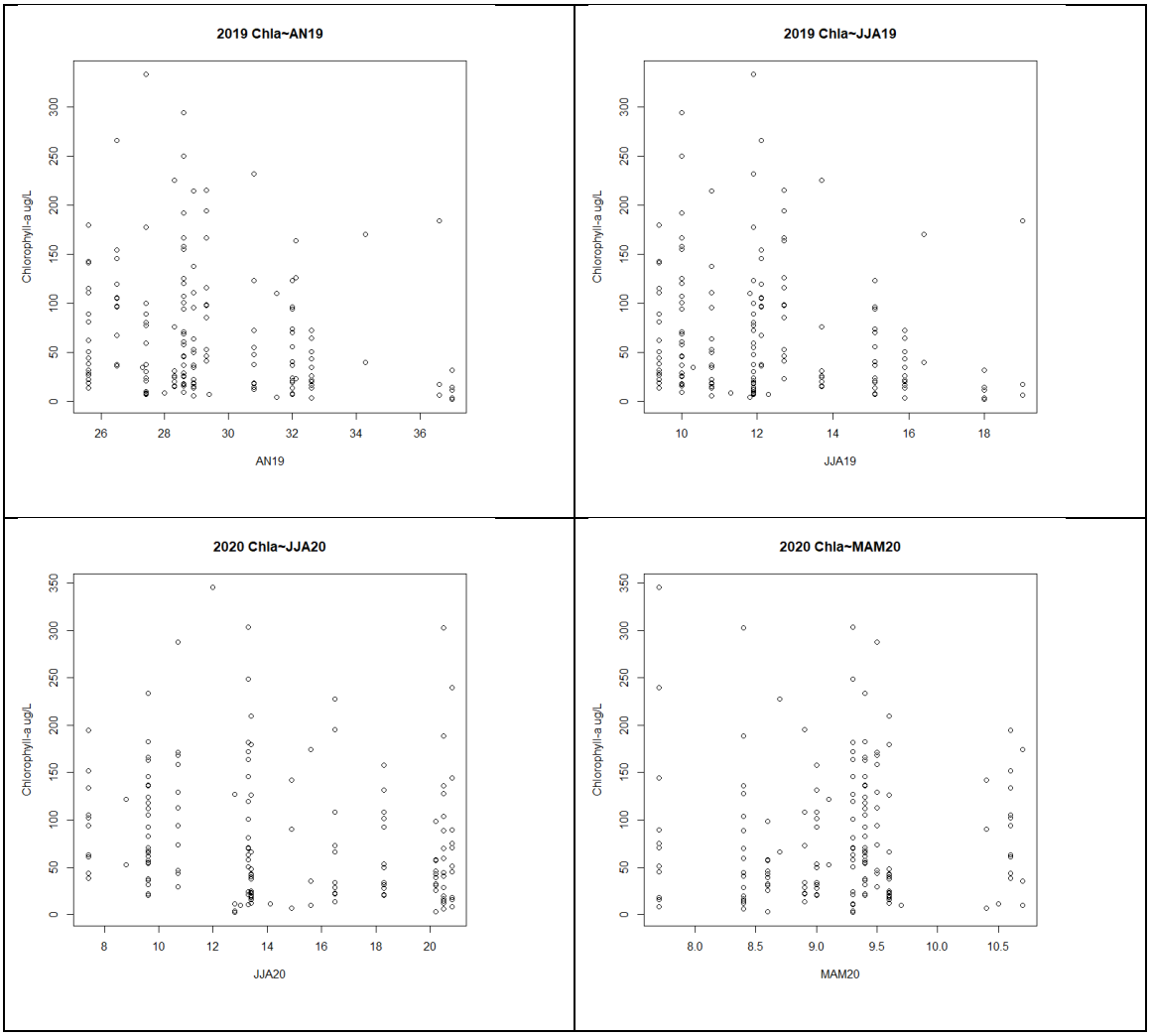

Figure 62. Precipitation impacts on chlorophyll-a in 160 WCBP lakes. Chlorophyll-a in 2019 and 2020 vs. annual precipitation in 2019 (AN19), June-July-August precipitation in 2019 (JJA19), June-July-August precipitation in 2020 (JJA20), and March-April-May precipitation in 2020 (MAM20). Other precipitation and temperature variables were removed via stepwise deletion. Only the 2020 model was significant (R^2= 0.0519, P = 0.00596).

5 Discussion

5.1 Overview of Major Findings

This study attempts to bring two tools to managing the problem of a lack of routine HAB monitoring programs in Minnesota combined with a high cost of monitoring: 1) remote sensing (trail cameras and UAVs), where the use of a potentially lower cost but understudied and recently developed remote sensing metric (VWRI) was compared to the more commonly used band ratio algorithm (NDVI) using UAVs, and 2) time series forecasting, exploring the potential of using univariate chlorophyll-a forecasts to reduce monitoring costs, and comparing the effectiveness of machine learning and classical time series approaches. The impact of meteorological factors (temperature, precipitation) on chlorophyll-a was examined in Bass Lake (interannual variability), Faribault County, MN, to understand factors controlling HABs at the lake, and across other Western Corn Belt Plains lakes; understanding these factors is critical to predicting whether a particular year will likely have HAB problems.

5.2 Remote Sensing

5.2.1 Trail Cameras and Sentinel-2

This study found a poor relationship between VWRI calculated from trail camera images and chlorophyll-a using both a large set of images and a subset with better image quality. These results suggest that VWRI is not a good predictor of chlorophyll-a in Bass Lake using trail cameras. Poor image quality could be a factor in the lack of relationship,

and could have resulted from waves, shadows, and reflection. However, since there was no relationship between VWRI and chlorophyll-a using a smaller subset of the best quality images, changes in the relationship between VWRI and chlorophyll-a over time may be a problem for this method. If this is the case, VWRI analyses could still be useful for mapping blooms using UAVs [42] or other imagery collected during a single day (satellites, smart phones), but it is not recommended to combine data from multiple days for VWRI. The trail cameras are recommended for use to keep an eye on water quality conditions remotely using visual inspection, but are not recommended for quantitative algal biomass estimation.

Although the relationship between VWRI and chlorophyll-a was stronger for Sentinel-2 data than the trail cameras, the results suggest that VWRI may not always be an effective index for modeling chlorophyll-a in Minnesota lakes. However, the satellite analysis involved combining data from multiple lakes across the twin cities area. Although the Met Council dataset was extensive and could be used for Sentinel-2 analyses, the results from this study suggest state agencies should consider sampling with higher spatial coverage at times (coinciding with Sentinel-2 overpasses) to allow the examination of VWRI using a dataset from a single lake. Although the VWRI analyses using UAVs provides support to using VWRI to model chlorophyll, more extensive sampling data coinciding with Sentinel-2 imagery would assist in determining the impact of using samples from different lakes in our VWRI analyses with Sentinel-2. Also, the poor NDCI-chlorophyll relationship, and separate clusters of data which were present after removing a few potential outliers suggested that significant differences may have

existed in the NDCI-chlorophyll-a relationship between lakes, and that intensive sampling on a single lake would be a superior approach. Zhang et al. [98] notes that the relationship between NDVI and chlorophyll-a is affected by turbidity, so variability in sediment concentrations between flights could be one reason NDVI or VWRI is inconsistent across lakes on different dates.

The Northern Kentucky University classifier had serious problems using the trail camera images from Bass Lake (Fig. 7), since it suggested that the majority of days between May 2019 and October 2019 had at least some images with 100% probability of blue-green algae presence during most days, which was not realistic based on visual inspection of the images. Methodological differences between the classifier developed by The Northern Kentucky University and our setup could be one factor: Northern Kentucky University deployed cameras in the center of lakes, while ours were on the shore, which could have reduced our image quality relative to what the classifier was trained on. However, the classifier is intended to deal with shadows and reflection. It seems possible that differences in optical properties between lakes (such as sediment concentration) could be one factor, and thereby the HAB APP classifier, which was trained using data from other lakes, may have been a poor fit for Bass Lake.

5.2.2 Comparison of VWRI and NDVI using UAVs

There was usually a positive relationships between both VWRI, and NDVI to chlorophyll-a concentrations. Previous laboratory research has suggested that VWRI is strongly correlated to chlorophyll-a [75]. NDVI or comparable algorithms have been

commonly tested using UAVs (BNDVI – [87]), although this is the first published study to our knowledge to test Sentera's NDVI camera for use in mapping algal blooms. Unlike the laboratory results in Sharaishi et al. 2018 [75], our VWRI values tended to be negative. NDVI values also ranged lower than expected. Since NDVI technically should be between -1 and 1, these values below -1 or above 1 are likely because NDVI was calculated from raw digital numbers, instead of converting the digital numbers to spectral reflectance first [46]. Digital numbers can be converted to spectral reflectance (ratio if the incident light to upwelling irradiance), however, in this study incident light was not measured. In the future, using an incident light sensor (ILS) to measure incident light, or using a calibration target would allow for digital numbers to be converted to reflectance, and likely would yield values between -1 to 1.

Although this is the first study to our knowledge which tested VWRI in the field, other RGB band ratio algorithms have been explored in other studies or for other applications. For example, Li et al. [47] monitored leaf area index using a variety of band ratio algorithms similar to VWRI (VARI - Visible Atmospherically Resistant Index, Normalized Green-Red, Red-Green-Blue Vegetation Index, and others). Our results suggest that NDVI outperformed VWRI, as expected, 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.

5.3 Time Series Forecasting of Chlorophyll-a

The results from this study confirm previous suggestions from Xiao et al. [92] that univariate forecasts of chlorophyll-a can be performed successfully at a much lower cost than a multivariate forecasts. Like Du et al. [24], this study found high model performance of chlorophyll-a forecasting using a multi-step rolling forecast (NAR in their study, LSTM, this study). While ARIMA showed low forecasting error (lower than the neural network, in contrast to Xiao et al. [92]), our results suggest that producing a realistic long term forecast is equally important to having a low error. 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. Examples of a low cost setup could be a Turner Cyclops-7F probe with a commercial grade data logger (~$4,000), or by combining the Cyclops-7F with a DIY data logger such as EnviroDIY Mayfly (< $100, total cost ~ $2,100). A univariate chlorophyll-a forecast using either ARIMA or LSTM would be valuable if the results were presented on

a webpage and the forecast was updated daily; this would assist by providing an early warning to lake users who need to make informed recreation decisions and to water quality managers who could prepare to sample or monitor forecasted blooms.

5.4 Impact of Meteorological Factors on HABs

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. Page et al. [59] found that higher spring precipitation followed by dry summers contributed enough nutrients early in the year, but a stable water column during the summer to support algal biomass. Our negative correlation with 60-day precipitation matches the results from Page et al. [59], but it is unclear why 2-week precipitation could spur algae growth. If the algae are nutrient starved later in the summer as the temperature warms up, a short pulse of nutrients could potentially allow a bloom to continue and develop further [38], reaching a higher maximum chlorophyll concentration. The lack of strong trends in the relationship between chlorophyll-a and meteorological patterns reflects similar results to the nationwide study of meteorological impacts on chlorophyll-a [38]. It is likely that other factors which control the chemical, biological, and physical conditions of a lake such as watershed area, land use, trophic state [98], slope, presence of point sources [96] create so much variability in the relationship between summer chlorophyll-a and meteorological factors that a broad scale analysis is not feasible. Zhang

[98] notes that as a lake becomes hypereutrophic it eventually becomes released from nutrient limitation, and meteorological factors begin to dominate controlling blooms.

5.5 Future research, Limitations, and Challenges

Although our study shows promise for using UAVs to detect and map algal blooms, there are some limitations of the technique to be aware of. First, the processing extent of a multi-rotor drone such as the Phantom 4 is limited to relatively small areas [87]. By flying at the maximum possible flight (121.9 meters), a high degree of image overlap is possible while covering a relatively large area, yet mapping the majority of a small lake like Bass Lake still takes ~ 1 hour and 4 flights. If there are areas of concern such as a swimming beach, drinking water intake, in a smaller lake, or to monitor shoreline areas where algae often accumulates with the wind (such as the northern shore of Bass Lake), UAVs may be appropriate. Many challenges of using UAVs to map chlorophyll-a were apparent in this study. Low sample sizes caused problems in our flights due to the potential for overfitting a regression relationship; for example, failing to sample across the entire range of chlorophyll-a / NDVI leads to making major assumptions about trends in the regression in un-sampled ranges. Future studies using UAVs to map chlorophyll-a in Minnesota Lakes should attempt to use large samples sizes; a sample size ranging from a minimum of 10 to 40 samples is recommended, since sampling is time consuming and must coincide with the UAV flight. Image stitching presents another challenge, as Drone Deploy can fail if the UAV is flown too low. Therefore, it is critical to fly as high as possible, and to overfly the shoreline and

study area, and to use at least 80% overlap between images. Future UAV work in MN should explore low costs methods for image stitching; Drone deploy requires a $150/month subscription, while Sentera's Field Agent software requires very specific and high power hardware (16 Gb ram, 4 core CPU with 8 processing threads). When a high-grade GPS is not available for capturing GPS coordinates, using either a sampling rod or an anchored boat would assist in limiting error in GPS coordinates; sampling as close as possible to the drone flight and noting any movement of algal patches between the flight and sampling would also assist in reducing error. Finally, flights should be conducted on low-wind days to reduce the presence of waves, and conducted in mid-morning to mid-day to improve image quality [87].

In the future, hyperparameters should be tested in more detail (using a larger grid search with more repeats, for all models including the univariate model and wavelet model). While the wavelet transformation reduced performance dramatically of the univariate LSTM forecast, it may that each series (A3, D1, D2, D3) are sensitive to hyperparameters, and future research should test the performance of wavelet-LSTM using different hyperparameters and potentially wavelet forms (DB-3 used in this study). One challenge in developing a forecasting model is that determining the appropriate hyperparameters can be extremely time consuming. Testing each possible hyperparameter setup in a grid search yields a multi-dimensional space of possibilities, which can yield extremely long computation times. The approach used in this study, was to examine each hyperparameter by itself, and to use these results to reduce the size of the grid search to the range where the hyperparameters perform successfully. Using this

approach may help in increasing the speed at which a forecasting model can be parameterized, which is important because a forecasting model needs to be produced quickly as new data is collected. Future studies should examine the impact of fitting a new model for each forecast (more computationally intensive but could increase accuracy and impact results). If data from a sonde with nutrient probes or high frequency nutrient measurements are available, it would be valuable to see if the variability which a multiple regression can explain would be increased, since nutrient measurements are one of the main parameters missing from this study [19].

One major limitation of our examination of factors controlling interannual variability in chlorophyll-a is a relatively short term data set of only 4 years. Zhang et al. [98] notes that short term datasets are limited in their ability to understand long term trends and factors controlling chlorophyll-a over large scales. Since Sentinel-2 only has data since 2016, the available dataset will continue to grow over time, so it an important area of future research. Obtaining meteorological data closer to each lake would increase the detail in the data for individual lakes and increase variability in the regressions. Controlling for more factors such as watershed area:surface ratio, trophic state (select lakes with similar nutrient and average chlorophyll-a concentrations), would likely be necessary to draw broad conclusions about the patterns of precipitation that can lead to years in Minnesota Lakes with problematic algal blooms. Although wet springs and drier summers may contribute to spurring blooms in some lakes at some times, overall the patterns of meteorological factors leading to a bad bloom season may be extremely

complex and variable. Our results at Bass Lake suggest that in this lakes algal blooms may be induced due to nutrient input from recent precipitation events.

6 Conclusion

To conclude, this study examined remote sensing methodology for chlorophyll-a estimation, and tested the effectiveness of VWRI. VWRI was compared to NDCI using satellites (Sentinel-2, and VWRI was compared to NDVI using a UAV. (NDVI and NDCI) to RGB (VWRI) BRs. While VWRI was unsuccessful at estimating chlorophyll-a in trail cameras, the results of this study suggests that VWRI can successfully be used to model chlorophyll-a in Minnesota Lakes using UAVs. However, since the VWRI-chlorophyll-a relationship varied over time, while NDVI had a more consistent relationship, NDVI mapping using UAVs requires fewer chlorophyll-a sampling trips, while VWRI requires chlorophyll-a sampling during each flight. In addition, in this study, NDVI provided more reliable algal maps, outlining blooms better than VWRI.

Univariate time series forecasting using LSTM performed better in general than multivariate LSTM forecasting, and ARIMA also showed good performance. These results suggest that chlorophyll-a forecasting can be performed successfully in Minnesota Lakes using a univariate approach, allowing for forecasting at lower costs than a multivariate setup.

The results from this study also suggest that algal blooms at Bass Lake are correlated to precipitation. A negative correlation with 60-day precipitation and positive correlation with 2-week precipitation suggest that both long term dry weather and

nutrient loading from recent rain events may lead to algal blooms. This study also suggests a slight negative correlation between precipitation and chlorophyll-a is present at a broader scale across Minnesota Lakes, which should be explored using longer term data at a local scale for individual lakes.

6.1 Acknowledgements

Thanks to The Hoppie Lab (Bryce Hoppie and Owen Lott) for setting up the trail camera, sonde, and weather station, use of YSI sondes, and collaborating on algal sampling, the Water Resources Center at Minnesota State University, Mankato, for use of UAV equipment, Leif Olmanson and Ben Page at the University of Minnesota for providing Sentinel-2 chlorophyll-a maps, and Mike Waters at Northern Kentucky University for sending trail camera imagery through their classifier. 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.

7 References

[1] Ram Avtar, Pankaj Kumar, Hitesh Supe, Dou Jie, Netranada Sahu, Binaya Kumar Mishra, and Ali P. Yunus. 2020. Did the COVID-19 Lockdown-Induced Hydrological Residence Time Intensify the Primary Productivity in Lakes? Observational Results Based on Satellite Remote Sensing. *Water* 12, 9 (September 2020), 2573.

[2] Mohamad Javad Alizadeh, Kavianpour, Mohamad Reza Kavianpour, Ozgur Kisi, and Vahid Nourani. 2017. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of hydrology* 548 (May 2017), 588-597.

[3] Lorraine C. Backer, Jan H. Landsberg, Melissa Miller, Kevin Keel, and Tegwin K. Taylor. 2013. Canine cyanotoxin poisonings in the United States (1920s–2012): Review of suspected and confirmed cases from three data sources. *Toxins 5*, 9 (September 2013), 1597-1628.

[4] Shams Ali Baig, Linglin Huang, Tiantian Sheng, Xiaoshu Lv, Zhe Yang, Muhammad Qasim, and Xinhua Xu. 2017. Impact of climate factors on cyanobacterial dynamics and their interactions with water quality in South Taihu Lake, China. *Chemistry and Ecology 33*, 1 (January 2017), 76-87.

[5] Rahim Barzegar, Mohammad Taghi Aalami, and Jan Adamowski. 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment* (February 2020), 1-19.

[6] Richard Beck, Shengan Zhan, Hongxing Liu, Susanna Tong, Bo Yang, Min Xu, Zhaoxia Ye, Yan Huang, Song Shu, Qiusheng Wu, Shujie Wang, Kevin Berling, Andrew Murray, Erich Emery, Molly Reif, Joseph Harwood, Jade Young, Christopher Nietch, Dana Macke, Mark Martin, Garrett Stillings, Richard Stump, and Haibin Su. 2016. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral

aircraft imagery and dense coincident surface observations. *Remote Sensing of Environment 178* (June 2016), 15-30.

[7] Richard Beck, Min Xu, Shengan Zhan, Hongxing Liu, Richard A. Johansen, Susanna Tong, Bo Yang, Song Shu, Quisheng Wu, Shujie Wang, Kevin Berling, Andrew Murray, Erich Emery, Molly Reif, Joseph Harwood, Jade Young, Mark Martin, Garrett Stillings, Richard Stump, and Haibin Su, Zhaoxia Ye, and Yan Huang. 2017. Comparison of satellite reflectance algorithms for estimating phycocyanin values and cyanobacterial total biovolume in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. *Remote Sensing* 9, 6 (May 2017), 538.

[8] Walter G. Bradley, R. X. Miller, T. D. Levine, E. W. Stommel, and P. A. Cox. 2018. Studies of environmental risk factors in amyotrophic lateral sclerosis (ALS) and a phase I clinical trial of L-serine. *Neurotoxicity research 33,* 1 (January 2018), 192-198.

[9] Bryan W. Brooks, James M. Lazorchak, Meredith DA Howard, Mari-Vaughn V. Johnson, Steve L. Morton, Dawn AK Perkins, Euan D. Reavie, Geoffrey I. Scott, Stephanie A. Smith, and Jeffery A. Steevens. 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?. *Environmental toxicology and chemistry* 35, 1 (January 2016), 6-13.

[10] Vinay Boddula, Lakshmish Ramaswamy, and Deepak Mishra. 2017. CyanoSense: A Wireless Remote Sensor System using Raspberry-Pi and Arduino with

Application to Algal Bloom. *2017 IEEE International Conference on AI & Mobile Services (AIMS)* (June 2017), 85-88.

[11] Bryan Brooks, James M. Lazorchak, Meredith DA Howard, Mari-Vaughn V. Johnson, Steve L. Morton, Dawn AK Perkins, Euan D. Reavie, Geoffrey I. Scott, Stephanie A. Smith, and Jeffery A. Steevens. 2017. In Some Places, in Some Cases and at Some Times, Harmful Algal Blooms are the Greatest Threat to Inland Water Quality. *Environmental toxicology and chemistry*, *36,* 5 (April 2017), 1125.

[12] Jason Brownlee. 2018. Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python. *Machine Learning Mastery*.

[13] Isabel Caballero, Raúl Fernández, Oscar Moreno Escalante, Luz Mamán, and Gabriel Navarro. 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 (May 2020), 1-14.

[14] Wayne W. Carmichael and Gregory L. Boyer. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful algae* 54 (April 2016), 194-212.

[15] Stephen R. Carpenter, Babak MS Arani, Paul C. Hanson, Marten Scheffer, Emily H. Stanley, and Egbert Van Nes. 2020. Stochastic dynamics of Cyanobacteria in long-term high-frequency observations of a eutrophic lake. *Limnology and Oceanography Letters* 5, 5 (October 2020), 331-336.

[16] Qiuwen Chen, Tiesheng Guan, Liu Yun, Ruonan Li, and Friedrich Recknagel. 2015. Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials. *Harmful Algae 43* (March 2015), 58-65.

[17] Hyungmin Cho, U-Jin Choi, Heekyung Park. 2018. Deep learning application to time-series prediction of daily chlorophyll-a concentration. *WIT Transactions on Ecology and the Environment* 215, 157-163.

[18] H Cho and H Park. 2019. Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast. *IOP Conference Series: Earth and Environmental Science,* 351, 1 (July 2019). IOP Publishing.

[19] Rory Coffey, Michael J. Paul, Jen Stamp, Anna Hamilton, and Thomas Johnson. A Review of Water Quality Responses to Air Temperature and Precipitation Changes 2: Nutrients, Algal Blooms, Sediment, Pathogens. *JAWRA Journal of the American Water Resources Association,* 55, 4 (August 2019), 844-868.

[20] Victoria G. Christensen, Ryan P. Maki, Erin A. Stelzer, Jack E. Norland, and Eakalak Khan. 2019. Phytoplankton community and algal toxicity at a recurring bloom in Sullivan Bay, Kabetogama Lake, Minnesota, USA. *Scientific reports*, *9,* 1 (November 2019), 1-11.

[21] Timothy W. Davis, Richard Stumpf, George S. Bullerjahn, Robert Michael L. McKay, Justin D. Chaffin, Thomas B. Bridgeman, and Christopher Winslow. 2019. Science meets policy: a framework for determining impairment designation

criteria for large waterbodies affected by cyanobacterial harmful algal blooms. *Harmful algae*, 81 (January 2019), 59-64.

[22] Eugen Diaconescu. 2008. The use of NARX neural networks to predict chaotic time series. *Wseas Transactions on computer research*, 3, 3 (March 2008), 182-191.

[23] Mike Dickman. 1969. Some Effects of Lake Renewal on Phytoplankton Productivity and Species Composition 1. *Limnology and Oceanography*, 14, 5 (September 1969), 660-666.

[24] Zhenhong Du., Mengjiao Qin, Feng Zhang, and Renyi Liu. 2018. Multistep-ahead forecasting of chlorophyll a using a wavelet nonlinear autoregressive network. *Knowledge-Based Systems*, 160 (November 2018), 61-70.

[25] Azhari A. Elhag and Hanaa Abu-Zinadah. 2020. Forecasting under applying machine learning and statistical models. *Thermal Science* 24, Suppl. 1, 131-137.

[26] Maria M. Parisio, Roberto Parisio, Tommaso Filippini, Valerio Mantione, Armando Platania, Anna Odone, Carlo Signorelli, Vladamiro Pietrini, Jessica Mandrioli, Sergio Teggi, Sofia Costanzini, Cristaldi Antonio, Pietro Zuccarello, Geo Oliveri Conti, Alessandra Nicoletti, Mario Zappia, Marco Vincenti, and Margherita Ferrante. Living near waterbodies as a proxy of cyanobacteria exposure and risk of amyotrophic lateral sclerosis: A population based case-control study. *Environmental research* 186 (July 2020), 109530.

[27] Joonseong Gim and Kwan-dong Park. Comparison of positioning accuracy using the pseudorange from android GPS raw measurements. *The Journal of Advanced Navigation Technology* 21, 5 (October 2017), 514-519.

[28] L.C. Gomes and L.E. Miranda. Hydrologic and climatic regimes limit phytoplankton biomass in reservoirs of the Upper Paraná River Basin, Brazil. *Hydrobiologia* 457, 1 (August 2001), 205-214.

[29] , Kathiresan Gopal, and Mahendran Shitan. Development of a Web Portal to Forecast the Monthly Mean Chlorophyll Concentration of the Waters off Peninsular Malaysia's West Coast. *Malaysian Journal of Mathematical Sciences* 12,1 (January 2018), 99-119.

[30] Alexander Grossmann and Jean Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis* 15, 4 (July 1984), 723-736.

[31] Xuexiang He, Yen-Ling Liu, Amanda Conklin, Judy Westrick, Linda K. Weavers, Dionysios D. Dionysiou, John J. Lenhart, Paula J. Mouser, David Szlag, and Harold W. Walker. 2016. Toxic cyanobacteria and drinking water: Impacts, detection, and treatment. *Harmful algae* 54 (April 2016), 174-193.

[32] Xiaoyu He, Suixiang Shi, Xiulin Geng, Lingyu Xu, and Xiaolin Zhang. 2021. Spatial-temporal attention network for multistep-ahead forecasting of chlorophyll. *Applied Intelligence* (January 2021), 1-13.

[33] R.E. Hecky and P. Kilham. 1988. Nutrient limitation of phytoplankton in freshwater and marine environments: a review of recent evidence on the effects of enrichment. *Limnology and oceanography 33*, 4part2 (July 1988), 796-822.

[34] T. Hein, C. Baranyi, G. Heiler, C. Holarek, P. Riedler, and F. Schiemer. 1999. Hydrology as a major factor determining plankton development in two floodplain

segments and the River Danube, Austria. *Large Rivers* (December 1999), 439-452.

[35] Steve Heiskary, Matthew Lindon, and Jesse Anderson. 2014. Summary of microcystin concentrations in Minnesota lakes. *Lake and Reservoir Management*, 30, 3 (July 2014), 268-272.

[36] Elizabeth D. Virginia A. Roberts, Lorraine Backer, Erin DeConno, Jessica S. Egan, James B. Hyde, David C. Nicholas, Eric J. Wiegert, Laurie M Billing, Mary DiOrio, Marika C. Mohr, F. Joan Hardy, Timothy J. Wade, Jonathan S. Yoder, and Michele C. Hlavsa. 2014. Algal bloom–associated disease outbreaks among users of freshwater lakes—United States, 2009–2010. *MMWR. Morbidity and mortality weekly report* 63, 1 (January 2014), 11.

[37] Walter R. Hill and Allen W. Knight. Nutrient and Light Limitation of Algae in Two Northern California Streams. *Journal of Phycology* 24, 2 (June 1988), 125-132.

[38] Jeff C. Ho and Anna M. Michalak. Exploring temperature and precipitation impacts on harmful algal blooms across continental US lakes. *Limnology and Oceanography* 65,5 (May 2020), 992-1009.

[39] Rob J. Hyndman and George Athanasopoulos. 2018. *Forecasting: principles and practice*. *OTexts*.

[40] Chippie Kislik, Iryna Dronova, and Maggi Kelly. 2018. UAVs in support of algal bloom research: a review of current applications and future opportunities. *Drones* 2, 4 (December 2018), 35.

[41] Sarah Kocher. 2020. Dry summer primes lakes for algae growth; Little Rock is no exception. *St. Cloud Times* (July 2020).

[42] Tiit Kutser. 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *International Journal of Remote Sensing* 30, 17 (August 2009), 4401-4425.

[43] Annette B.G. Janssen, Jan H. Janse, Arthur HW Beusen, Manqi Chang, John A. Harrison, Inese Huttunen, Xiangzhen Kong, Jasmin Rost, Sven Teurlincx, Tineke Troost, Dianneke van Wijk, Wolf M. Mooij. 2019. How to model algal blooms in any lake on earth. *Current opinion in environmental sustainability*, 36 (February 2019), 1-10.

[44] Kwang-Seuk Jeong, Dong-Kyun Kim, Jong-Mun Jung, Myoung-Chul Kim, and Gea-Jae Joo. 2008. Non-linear autoregressive modelling by Temporal Recurrent Neural Networks for the prediction of freshwater phytoplankton dynamics. E*cological modelling* 211, (3-4) (March 2008), 292-300.

[45] Gooyong Lee, Faridah Othman, Shaliza Ibrahim, and Min Jang. 2016. Determination of the forecasting-model parameters by statistical analysis for development of algae warning system. *Desalination and Water Treatment*, 57, 55 (November 2016), 26773-26782.

[46] Kyung-Do Lee, Lee Ye-Eun, Chan-Won Park, Suk-Young Hong, and Sang-I Na. 2016. Study on Reflectance and NDVI of Aerial Images using a Fixed-Wing UAV. *Korean Journal of Soil Science and Fertilizer* 49, 6 (December 2016), 731-742.

[47] Songyang Li, Fei Yuan, Syed Tahir Ata-UI-Karim, Hengbiao Zheng, Tao Cheng, Xiaojun Liu, Yongchao Tian, Yan Zhu, Weixing Cao, and Qiang Cao. "Combining

color indices and textures of UAV-based digital imagery for rice LAI estimation." *Remote Sensing* 11, 15 (January 2019), 1763.

[48] Xue Li, Jian Sha, and Zhong-Liang Wang. 2018. Application of feature selection and regression models for chlorophyll-a prediction in a shallow lake. *Environmental Science and Pollution Research*, 25, 20 (July 2018), 19488-19498.

[49] Matthew Lindon, Steven Heiskary. 2009. Blue-green algal toxin (microcystin) levels in Minnesota lakes. *Lake and Reservoir Management* 25, 3 (September 2009), 240-252.

[50] Jianping Liu, Yuchao Zhang, and Xin Qian. 2009. Modeling chlorophyll-a in Taihu Lake with machine learning models. *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, (June 2009), 1-6.

[51] Yu Liu, Du-Gang Xi, and Zhao-Liang Li. 2015. Determination of the optimal training principle and input variables in artificial neural network model for the biweekly chlorophyll-a prediction: A case study of the Yuqiao Reservoir, China. *PloS one* 10, 3 (March 2015), e0119082.

[52] Xia Liu, Jianfeng Feng, and Yuqiu Wang. 2019. Chlorophyll a predictability and relative importance of factors governing lake phytoplankton at different timescales. *Science of the Total Environment* 648 (January 2019), 472-480.

[53] Fang Lu, Zhi Chen, Wenquan Liu, and Hongbo Shao. 2016. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecological engineering* 95 (October 2016), 422-429.

[54] Wenguang Luo, Senlin Zhu, Shiqiang Wu, and Jiangyu Dai. 2019. Comparing artificial intelligence techniques for chlorophyll-a prediction in US lakes. *Environmental Science and Pollution Research* 26, 29 (October 2019), 30524-30532.

[55] Spyros Makridakis Evangelos Spiliotis, and Vassilios Assimakopoulos. 2018. Statistical and Machine Learning forecasting methods: Concerns and ways forward." *PloS one* 13, 3 (March 2018), e0194889.

[56] Ina S. Markham, Terry R. Rakes. 1998. The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. *Computers & operations research*, 25, 4 (April 1998), 251-263.

[57] Anna M. Michalak, V. Balaji, Dario Del Giudice, Eva Sinha, Yuntao Zhou, and Jeff C. Ho. 2017. Are extreme hydro-meteorological events a prerequisite for extreme water quality impacts? Exploring climate impacts on inland and coastal waters. *AGU Fall Meeting Abstracts* Vol. 2017, H43L-1797.

[58] Sachidananda, Mishra and Deepak R. Mishra. 2012. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment* 117 (February 2012), 394-406.

[59] Benjamin P. Page, Abhishek Kumar, and Deepak R. Mishra. 2018. A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial

harmful algal bloom. *International journal of applied earth observation and geoinformation* 66 (April 2018), 69-81.

[60] Benjamin P. Page, Leif G. Olmanson, and Deepak R. Mishra. 2019. A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment* 231 (September 2019), 111284.

[61] Yongeun Park, Kyung Hwa Cho, Jihwan Park, Sung Min Cha, and Joon Ha Kim. 2015. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Science of the Total Environment* 502 (January 2015), 31-41.

[62] Hans W. Paerl and Rolland S. Fulton, Pia H. Moisander, and Julianne Dyble. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *TheScientificWorldJournal* 1, (April 2001), 76-113.

[63] Hans W. Paerl and Jef Huisman. 2008. Blooms like it hot. *Science*, 320, 5872 (April 2008), 57-58.

[64] Hans W. Paerl, and J. Thad Scott. 2010. Throwing fuel on the fire: synergistic effects of excessive nitrogen inputs and global warming on harmful algal blooms. *Environ. Sci. Technol.* 44, 20 (August 2010), 7756-7758.

[65] Hans W. Paerl and Malcom A. Barnard. 2020. Mitigating the global expansion of harmful cyanobacterial blooms: Moving targets in a human-and climatically-altered world. *Harmful algae*, 96 (June 2020), 101845.

[66] Monica Pinardi, Mariano Bresciani, Paolo Villa, Ilaria Cazzaniga, Alex Laini, Viktor Tóth, Ali Fadel, Martina Austoni, Andrea Lami, and Claudia Giardino. 2018. Spatial and temporal dynamics of primary producers in shallow lakes as seen from space: Intra-annual observations from Sentinel-2A. *Limnologica* 72 (June 2018), 32-43.

[67] Taher Rajaee and Amir Boroumand. 2015. Forecasting of chlorophyll-a concentrations in South San Francisco Bay using five different models. *Applied Ocean Research* 53 (October 2015), 208-217.

[68] Friedrich F. Recknagel, Ilia Ostrovsky, Hongqing Cao, Tamar Zohary, and Xiaoqing Zhang. 2013. Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets. *Ecological Modelling* 255 (April 2013), 70-86.

[69] Elke S. Reichwaldt, and Anas Ghadouani. 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. *Water research* 46, 5 (April 2012), 1372-1393.

[70] Virginia A. Roberts, Marissa Vigar, Lorraine Backer, Gabriella E. Veytsel, Elizabeth D. Hilborn, Elizabeth I. Hamelin, Kayla L. Vanden Esschert, Joana Y. Lively, Jennifer R. Cope, Michele C. Hlavsa, and Jonathan S. Yoder. 2020. Surveillance for harmful algal bloom events and associated human and animal illnesses—One health harmful algal bloom system, United States, 2016–2018. *Morbidity and Mortality Weekly Report*, 69, 50 (December 2020), 1889.

[71] Cecile S Rousseaux and Watson W. Gregg. 2017. Forecasting ocean chlorophyll in the Equatorial Pacific. *Frontiers in Marine Science* 4 (July 2017), 236.

[72] Sentera. 2017. False Color to NDVI Conversion Precision NDVI Single Sensor.

[73] Shaoling Shang, Zhongping Lee, Gong Lin, Chuanmin Hu, Lianghai Shi, Yongnian Zhang, Xueding Li, Jingyu Wu, and Jing Yan. 2017. Sensing an intense phytoplankton bloom in the western Taiwan Strait from radiometric measurements on a UAV. *Remote Sensing of Environment* 198 (September 2017), 85-94.

[74] Li Shen, Huiping Xu, and Xulin Guo. 2012. Satellite remote sensing of harmful algal blooms (HABs) and a potential synthesized framework. *Sensors*, 12, 6 (June 2012), 7778-7803.

[75] 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.

[76] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. 2018. A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (December 2018).

[77] Talita Silva, Brigitte Vinçon-Leite, Bruno Lemaire, Briac Le Vu, Catherine Quiblier, François Prévot, Catherine Freissinet, Michel Calzas, Yves Degres, and Bruno Tassin. 2011. Water Quality in urban lakes: from continuous monitoring to

forecasting. Application to cyanobacteria dynamics in Lake Enghien (France). *European Geosciences Union General Assembly* (April 2011).

[78] Eva Sinha and Anna M. Michalak. 2016. Precipitation dominates interannual variability of riverine nitrogen loading across the continental United States. *Environmental science & technology* 50, 23 (December 2016), 12874-12884.

[79] Ankita Srivastava, Shweta Singh, Chi-Yong Ahn, Hee-Mock Oh, and Ravi Kumar Asthana. 2013. Monitoring approaches for a toxic cyanobacterial bloom. *Environmental science & technology* 47, 16 (August 2013), 8999-9013.

[80] Shahaboddin Shamshirband, Ehsan Jafari Nodoushan, Jason E. Adolf, Azizah Abdul Manaf, Amir Mosavi, and Kwok-wing Chau. 2019. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Engineering Applications of Computational Fluid Mechanics* 13, 1 (January 2019), 91-101.

[81] Val H. Smith. 1998. Cultural eutrophication of inland, estuarine, and coastal waters. In *Successes, limitations, and frontiers in ecosystem science*, Springer, New York, NY, 7-49.

[82] Robert W. Sterner. 2008. On the phosphorus limitation paradigm for lakes. *International Review of Hydrobiology* 93, (4-5) (October 2008), 433-445.

[83] Richard P. Stumpf, Timothy W. Davis, Timothy T. Wynne, Jennifer L. Graham, Keith A. Loftin, Thomas H. Johengen, Duane Gossiaux, Danna Palladino, and

Ashley Burtner. 2016. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful algae* 54 (April 2016), 160-173.

[84] Elizabeth B. Stumpner, Brian A. Bergamaschi, Tamara EC Kraus, Alexander E. Parker, Frances P. Wilkerson, Bryan D. Downing, Richard C. Dugdale, Michael C. Murrell, Kurt D. Carpenter, James L. Orlando, and Carol Kendall. 2020. Spatial variability of phytoplankton in a shallow tidal freshwater system reveals complex controls on abundance and community structure. *Science of the Total Environment*, 700 (January 2020), 134392.

[85] G. Tanber, G. 2014. Toxin leaves 500,000 in northwest Ohio without drinking water. *Reuters News Service Online*. < https://www.reuters.com/article/us-usa-water-ohio/toxin-leaves-500000-in-northwest-ohio-without-drinking-water-idUSKBN0G20L120140802 > (Accessed July 2021).

[86] Wenchong Tian, Zhenliang Liao, and Jin Zhang. 2017. An optimization of artificial neural network model for predicting chlorophyll dynamics. *Ecological modelling* 364 (November 2017), 42-52.

[87] Deon Van der Merwe and Kevin Price. 2015. Harmful algal bloom characterization at ultra-high spatial and temporal resolution using small unmanned aircraft systems. *Toxins*, 7, 4 (April 2015), 1065-1078.

[88] He Yi Wang and Xu Chang Yang. 2013. Elman's Recurrent Neural Network Applied to Forecasting Algal Dynamic Variation in Gonghu Bay. In *Advanced Materials Research* 779, 1352-1358). Trans Tech Publications.

[89] Li Wang, Tianrui Zhang, Xiaoyi Wang, Xuebo Jin, Jiping Xu, Jiabin Yu, Huiyan Zhang, and Zhiyao Zhao. 2019. An approach of improved Multivariate Timing-Random Deep Belief Net modelling for algal bloom prediction. *Biosystems engineering*, 177 (January 2019), 130-138.

[90] Mike Waters, M. Kannan, Jim Lazorchak, and Joel Allen. 2016. Harmful algal bloom smart device application: using image analysis and machine learning techniques for classification of harmful algal blooms. *SETAC NA Annual Meeting, Minneapolis, MN*.

[91] Di Wu, Ruopu Li, Feiyang Zhang, and Jia Liu. 2019. A review on drone-based harmful algae blooms monitoring. *Environmental monitoring and assessment* 191, 4 (April 2019), 1-11.

[92] 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.

[93] Hiroshi Yajima and Jonathan Derot. 2017. Application of the Random Forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases. *Journal of Hydroinformatics* 20, 1 (January 2017), 206-220.

[94] Lin Ye, Qinghua Cai, Min Zhang, and Lu Tan. 2014. Real-time observation, early warning and forecasting phytoplankton blooms by integrating in situ automated

online sondes and hybrid evolutionary algorithms. *Ecological informatics* 22 (July 2014), 44-51.

[95] Hye-Suk Yi, Sangyoung Park, Kwang-Guk An, and Keun-Chang Kwak. 2018. Algal bloom prediction using extreme learning machine models at artificial weirs in the Nakdong River, Korea. *International journal of environmental research and public health* 15,10 (October 2018), 2078.

[96] , Zhenyu Yu, Kun Yang, Yi Luo, and Chunxue Shang. 2020. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. *Journal of Hydrology* 582 (March 2020), 124488.

[97] G. Peter Zhang. 2012. Neural networks for time-series forecasting. *Handbook of natural computing* (November 2012), 461-477.

[98] Tiantian Zhang, Hong Hu, Xiaoshuang Ma, and Yaobo Zhang. 2020. Long-Term Spatiotemporal Variation and Environmental Driving Forces Analyses of Algal Blooms in Taihu Lake Based on Multi-Source Satellite and Land Observations. *Water* 12, 4 (April 220), 1035.

8 Appendix

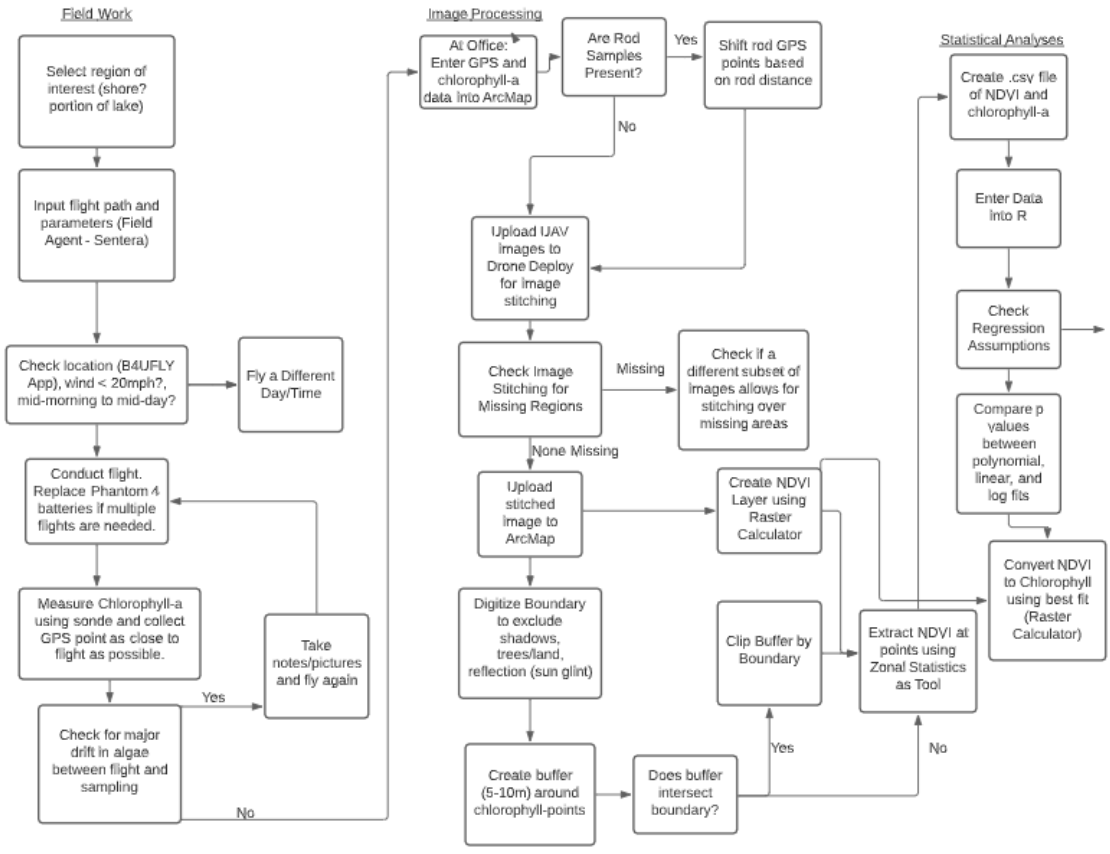

Figure 63 - Appendix. Flow Chart of UAV mapping (NDVI using Phantom 4 with Sentera camera).

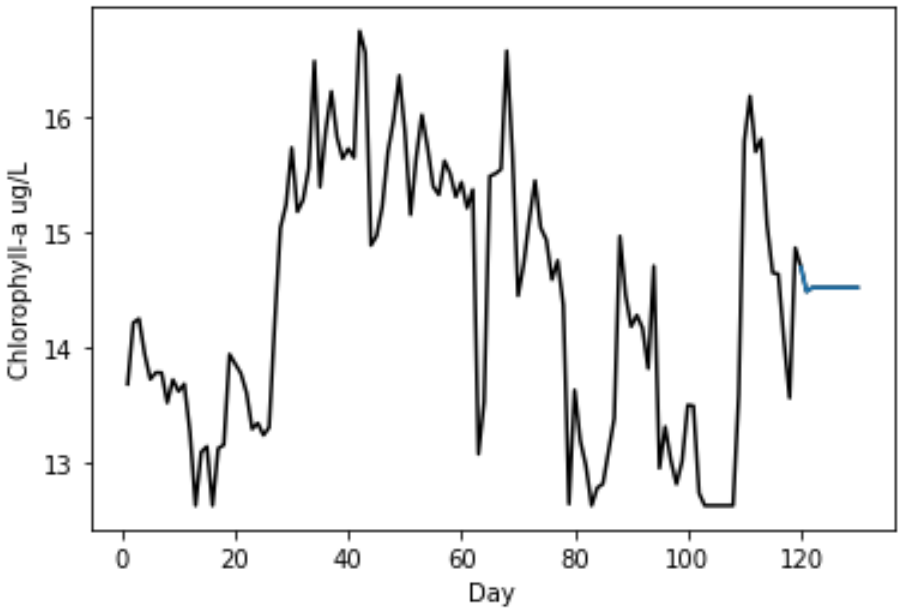

Figure 64 - Appendix. Earlier ARIMA forecast (2020) with discussion of future work related to missing data and re-estimation. ARIMA order 0,1,2. Top: forecast from area with missing data (imputed to the median). Black – actual data, blue – 10 day forecast. Flat regions show periods with missing data (imputed to the median).

Further Discussion of Figure 64

In order to obtain time series forecasts, missing data (periods while the sonde broke down and required repair) had to be imputed (to the median). However, during 2020 the time series forecasts came from a period near the end of the year using imputed data. Due to a lack of variability and inaccuracy of the imputed data, the reliability of the 10-day forecasts from an imputed period should be viewed with caution. The RMSE for each forecast method was calculated by averaging the error from multiple 10-day forecasts over 33% of the test set; selecting an area without missing data may provide a more reliable RMSE, although the impact of a small period of missing data over the entire test set is limited. The ARIMA, 2020 results were repeated by performing the forecast from earlier in the year while the sonde was functioning. This shows that there is variability in the ARIMA forecast initially, followed by a flat forecast even with this earlier forecast. The results suggest that ARIMA forecasts may be more useful for the first few days of the forecast, and that a rolling forecast where the ARIMA model is retrained (re-estimation) could provide a longer term forecast which can show expected variability in chlorophyll-a (non-flat), while still having good performance. Given time constraints it was not possible to re-run the forecasts using periods with non-missing data, or to repeat the forecast with re-estimation, so this is an area which is suggested for future work.

Table 7 - Appendix. Literature Review of Strengths and Weaknesses of Chlorophyll Time Series Forecasting Approaches.

Reference	Type	Method	Location	Time Period	Strengths	Weaknesses
[24]	Machine Learning	Wavelet-NAR, multi-step	Wenzhou coast	2014-2015	low accumulated error, maintains dependencies, low computational cost, good prediction	still potential for accumulated error
[80]	Machine Learning (Ensemble)	Bates Granger or least squares with Wavelet-ANN	Hilo Bay, Hawaii	2012-2016	increased accuracy and uncertainty compared to single model, especially for longer horizons, good for nonlinear problems	model accuracy fell dramatically after 3 days
[92]	Machine Learning	Wavelet-ANN	Siling Reservoir China, Lake Winnebago, USA	2011-2012, 2013	wavelet-transform reduces error and improves accuracy to regular ANN, requires shorter data set that classical forecasting, good	model has only been tested on eutrophic lakes, may be best for short term

					performance for a single parameter model	
[18]	Machine Learning	LSTM	Geum River, South Korea	2013-2017	good deep learning approach for time series, good performance for longer horizons, LSTM learns long term dependencies	potential for accumulated error
[93]	Machine Learning	Random Forest	Urayama Reservoir and Lake Shinji, Japan	1981-2015	good for multivariate forecasting problems with large datasets	prone to overfitting (like ANN)
[67]	Machine Learning	Support Vector Machine Regression (hybrid with GA)	San Fransisco Bay	1993-2013	adequate performance	optimal tuning of hyperparameters is critical (GA used here to identify), lower performance than ANN here
[67]	Classical	WMLR (Multiple Linear Regression	San Fransisco Bay	1993-2013	adequate performance, wavelet transform	lower performance than ANN here, not strong for

		with wavelet transform)			improved performance compared to MLR	nonlinear processes like chlorophyll
[16]	Classical	ARIMA	Lake Taihu, China	2010-2011	better performance than multivariate method (MLR) here, only requires one input variable	
[25]	Classical	SARIMA	inland waters in Saudia Arabia	2017-2018	worked well to forecast using intermittent data at month scale (Sentinel-2)	inappropriate when seasonal parameters are not present
[5]	Machine Learning	CNN, LSTM and CNN-LSTM hybrid	Small Prespa Lake, Greece	2012-2013	Hybrid CNN-LSTM outperforms both LSTM and CNN, CNN is good at extracting time-invariant features	Standalone CNN is poor at predicting high values, while standalone LSTM is poor at predicting low values
[29]	Classical	Holt-Winters seasonal (exponential smoothing)	West Coast of Malaysia	2002-2017	includes seasonal component, additive model had good performance for	poor choice if data does not include seasonality

					monthly mean data	
[61]	Machine Learning	SVM	Juam and Yeongsan Reservoirs, Korea		good performance in relatively long horizon (7 days)	prediction accuracy may decline after 7 days
[71]	Mathematical Modeling	NOBM	Equatorial Pacific Ocean	2012-2015	Good performance for 3 month forecast	Uncertainties associated with model forcing and satellite data

Table 8 - Appendix. Chlorophyll-a concentrations (measured in the laboratory) by the Metropolitan Council). Samples were collected from 34 twin cities area lakes on 5/15/2018.

ID	Lake Name	Chlorophyll-a (µg/L)
82001900-01	South Twin Lake	7.8
82004900-01	Big Carnelian Lake	6.1
10000500-01	Courthouse Lake	1
82004600-01	Square Lake	1.7
82002602-01	Mud Lake	32
10022500-01	Brickyard Clayhole	1.5
82003600-01	Turtle Lake	4.7
10022600-01	Firemens Clayhole	1.6
82005204-01	Big Marine Lake	3.9
10001400-01	Hazeltine Lake	30
82007600-01	Barker Lake	4
82005400-01	Bone Lake	13
10024900-01	Big Woods Lake	39
10021600-01	McKnight Lake	17
13005300-01	Big Comfort Lake	7.2
02013000-01	Pickerel Lake	3.8
70007800-01	Haas Lake	4.3
10021700-01	Jonathan Lake	23
10021800-01	Grace Lake	17
82015900-01	Forest Lake	3
27010700-01	Parkers Lake	1
10006900-01	Benton Lake	110

82015900-02	Forest Lake	4
10007000-01	Meuwissen Lake	130
19002400-01	Wood Lake	3.4
82015900-03	Forest Lake	5.4
82001800-01	North Twin Lake	1.7
82001600-01	Silver Lake	3.9
10002900-01	Miller Lake	44
82002500-01	Louise Lake	2.1
82001502-01	Loon Lake	7.6
82008000-01	Keewahtin Lake	1.4
19002500-01	Keller Lake	4.8
82010300-01	Olson Lake	2.2
82002500-01	Louise Lake	2.1
82001502-01	Loon Lake	7.6
82008000-01	Keewahtin Lake	1.4
19002500-01	Keller Lake	4.8
82010300-01	Olson Lake	2.2

Table 9 - Appendix. Regression R^2 and p values from UAV flights in Bass Lake and Little Rock Lake.

Date	Model	Adjusted R^2	Type	Overall P value	Term-1 P-value	Term-2 P-value	Term-3 P-value	Shapiro Test
9/14/2020 Bass	Chlorophyll ~NDVI	0.986	Polynomial	4.12e-11	.00777	2.13e-05	0.0726	0.0944
Chlorophyll = $209.139*NDVI^3 + 219.541*NDVI^2 + 105.835*NDVI + 69.254$								
9/14/2020 Bass	Chlorophyll ~VWRI	0.866	Linear	3.07e-07				0.173
Chlorophyll = 3422.22*VWRI + 321.38								
10/8/2020 Bass	Chlorophyll ~NDVI	0.631	Linear	0.0203				0.848
Chlorophyll = 74.881*NDVI + 43.645								
10/8/2020 Bass	Chlorophyll ~VWRI	0.797	Linear	0.00427				0.525

Chlorophyll = 318.812*VWRI + 80.940								
9/26/2020 Bass	Chlorophyll ~NDVI	0.577	Linear	2.45e-04				0.00154
	1 outlier removed	0.785	Linear	3.06e-06				
Full Model: Chlorophyll = 226.68*NDVI + 120.97 1 Outlier Removed: Chlorophyll=234.18*NDVI+109.09								
9/26/2020 Bass	Chlorophyll ~VWRI	0.0785	Linear	0.145				0.0703
	1 outlier removed	0.467	Linear	0.00212				
Full Model: Chlorophyll = 226.68*VWRI +120.97 Outliers Removed: Chlorophyll = 3018.3*VWRI+604.3								
9/18/2020 Bass	Chlorophyll ~NDVI	0.665	Polynomial	0.00774	0.0510	0.0737	0.0325	0.684
Chlorophyll = 37.877*NDVI -186.852*NDVI^2-274.367*NDVI^3 + 61.506								
9/18/2020 Bass	Chlorophyll ~VWRI	0.799	Polynomial	0.000296	0.0006 41	0.0023 7		0.769
Chlorophyll = -1613.74*VWRI^2 -712.39 *VWRI -13.11								
9/5/2020 Little Rock	Chlorophyll ~NDVI	0.323	Linear	0.00527				0.751
Chlorophyll = 27.295*NDVI + 35.766								
9/5/2020	Chlorophyll ~VWRI	0.17	Natural Log	0.0632				0.169

Little Rock								
Chlorophyll = 9.518*ln(vwri) + 62.035								
Fall 2020 Bass Lake	Chlorophyll ~NDVI	0.678	Polynomial	2.98e-12	0.003920	0.000552	0.4686 55	2.23*10e-6
Chlorophyll = 135.53*NDVI+160.26*NDVI^2+86.91*NDVI^3 + 80.43								
Fall 2020 Bass Lake	Chlorophyll ~VWRI	0.027	Linear	0.130				8.67*1-e-07
Chlorophyll = 307.72*VWRI + 141.55								

Table 10 - Appendix. From 160 lakes in the Western Cornbelt Plains; 2019 Average monthly chlorophyll, annual precipitation (an), snowfall (snow), March-April-May precipitation (MAM), June-July-August precipitation (JJA), March-April May average temperature(MAMT), June-July-August average temperature (JJAT).

	Year: 2020						
Lake	chla	an	snow	mam	jja	mamt	jjat
Bamber	2.1	24.3	37.6	9.3	12.8	54.1	81.5
George	2.2	24.3	37.6	9.3	12.8	54.1	81.5
Hiniker Pond	3.2	31.4	34.5	8.6	20.2	54.6	83.2
Bamber	3.5	24.3	37.6	9.3	12.8	54.1	81.5
Juni	5.8	31.3	39.3	8.4	20.5	42.5	83.9
Kohlmeier	7.1	27.4	31.3	10.4	14.9	53.9	82.8

Madison	8.1	30.7	29.3	7.7	20.8	54.6	83.2
Zumbro	9.7	24.1	25.5	9.7	13.0	54.1	81.5
Beaver	9.8	29.4	36.5	10.7	15.6	53.9	82.4
Eberhart	10.2	25.1	35.1	9.3	13.3	54.9	82.9
Silver	11.4	24.3	37.6	9.3	12.8	54.1	81.5
Mill Pond	11.5	26.7	27.8	10.5	14.1	53.4	81.5
Bullhead	11.7	26.2	34.5	9.6	13.4	55.8	84.3
Sleepy Eye	12.0	31.3	39.3	8.4	20.5	42.5	83.9
Goose	13.4	27.5	24.2	8.9	16.5	54.9	82.1
Erickson	14.2	31.3	39.3	8.4	20.5	42.5	83.9
Middle	15.8	30.7	29.3	7.7	20.8	54.6	83.2
Fedji	15.9	26.2	34.5	9.6	13.4	55.8	84.3
Peterson	16.8	31.3	39.3	8.4	20.5	42.5	83.9
Oak Leaf	17.8	30.7	29.3	7.7	20.8	54.6	83.2
School	18.0	26.2	34.5	9.6	13.4	55.8	84.3
Swan	19.3	31.3	39.3	8.4	20.5	42.5	83.9
St. James	19.4	26.2	34.5	9.6	13.4	55.8	84.3
Seymour	20.0	20.0	22.0	9.4	9.6	54.8	82.8
Long	20.3	26.2	34.5	9.6	13.4	55.8	84.3
Perch	20.4	30.1	31.7	9.0	18.3	53.9	82.4
Upper Twin	20.8	25.1	35.1	9.3	13.3	54.9	82.9
Silver	21.3	30.1	31.7	9.0	18.3	53.9	82.4

Cedar	21.8	20.0	22.0	9.4	9.6	54.8	82.8
Freeborn	22.2	27.5	24.2	8.9	16.5	54.9	82.1
Kansas	22.5	26.2	34.5	9.6	13.4	55.8	84.3
South Walnut	22.8	27.5	24.2	8.9	16.5	54.9	82.1
Sulem	23.2	26.2	34.5	9.6	13.4	55.8	84.3
Goose	24.0	25.1	35.1	9.3	13.3	54.9	82.9
Wood	24.5	26.2	34.5	9.6	13.4	55.8	84.3
Mills	25.8	31.4	34.5	8.6	20.2	54.6	83.2
Alice	27.6	30.1	31.7	9.0	18.3	53.9	82.4
Altermatt	28.4	31.3	39.3	8.4	20.5	42.5	83.9
Freeborn	28.9	27.5	24.2	8.9	16.5	54.9	82.1
Bass	29.6	22.2	32.3	9.5	10.7	55.1	83.5
Eagle (South)	31.2	31.4	34.5	8.6	20.2	54.6	83.2
Perch	31.4	30.1	31.7	9.0	18.3	53.9	82.4
Knights	31.9	30.1	31.7	9.0	18.3	53.9	82.4
Big Twin	31.9	20.0	22.0	9.4	9.6	54.8	82.8
Lily	32.3	31.4	34.5	8.6	20.2	54.6	83.2
Clear	33.5	30.1	31.7	9.0	18.3	53.9	82.4
Minnesota	34.0	27.5	24.2	8.9	16.5	54.9	82.1
Rice	34.2	30.1	31.7	9.0	18.3	53.9	82.4
St. Olaf	35.7	29.4	36.5	10.7	15.6	53.9	82.4

Creek	35.9	20.0	22.0	9.4	9.6	54.8	82.8
Cedar	37.5	20.0	22.0	9.4	9.6	54.8	82.8
Hanska	37.5	26.2	34.5	9.6	13.4	55.8	84.3
Hall	38.0	19.5	32.2	10.6	7.4	54.8	81.8
Spring	39.0	31.4	34.5	8.6	20.2	54.6	83.2
Omsrud	40.1	26.2	34.5	9.6	13.4	55.8	84.3
Armstrong	40.9	31.3	39.3	8.4	20.5	42.5	83.9
Irish	41.8	26.2	34.5	9.6	13.4	55.8	84.3
Mary	42.9	26.2	34.5	9.6	13.4	55.8	84.3
Eagle (North)	42.9	31.4	34.5	8.6	20.2	54.6	83.2
George	44.0	19.5	32.2	10.6	7.4	54.8	81.8
Mountain	44.0	22.2	32.3	9.5	10.7	56.5	84.7
Lieberg	44.3	31.3	39.3	8.4	20.5	42.5	83.9
Madison	44.9	30.7	29.3	7.7	20.8	54.6	83.2
Crystal	45.8	31.4	34.5	8.6	20.2	54.6	83.2
Lura	46.7	22.2	32.3	9.5	10.7	55.1	83.5
Willow Creek	48.5	26.2	34.5	9.6	13.4	55.8	84.3
Buffalo	49.7	30.1	31.7	9.0	18.3	53.9	82.4
Geneva	50.3	25.1	35.1	9.3	13.3	54.9	82.9
George	50.8	30.7	29.3	7.7	20.8	54.6	83.2
Budd	52.7	19.9	30.9	9.1	8.8	55.1	83.5

Amber	52.9	19.9	30.9	9.1	8.8	55.1	83.5
Silver	53.7	30.1	31.7	9.0	18.3	53.9	82.4
Susan	53.9	20.0	22.0	9.4	9.6	54.8	82.8
Severson	54.7	20.0	22.0	9.4	9.6	54.8	82.8
Loon	57.4	31.4	34.5	8.6	20.2	54.6	83.2
North Silver	57.5	20.0	22.0	9.4	9.6	54.8	82.8
Eagle (North)	57.9	31.4	34.5	8.6	20.2	54.6	83.2
East Side	58.0	25.1	35.1	9.3	13.3	54.9	82.9
Butterfield	59.7	31.3	39.3	8.4	20.5	42.5	83.9
Swag	60.6	19.5	32.2	10.6	7.4	54.8	81.8
South Silver	61.1	19.5	32.2	10.6	7.4	54.8	81.8
Pierce	61.8	20.0	22.0	9.4	9.6	54.8	82.8
Rose	62.9	19.5	32.2	10.6	7.4	54.8	81.8
Lower Twin	63.4	25.1	35.1	9.3	13.3	54.9	82.9
Eagle	65.7	20.0	22.0	9.4	9.6	54.8	82.8
School	65.8	26.2	34.5	9.6	13.4	55.8	84.3
Faribault	66.5	27.1	25.0	8.7	16.5	53.9	82.8
Temperance	67.6	20.0	22.0	9.4	9.6	54.8	82.8
Linden	69.7	31.3	39.3	8.4	20.5	42.5	83.9
Geneva	69.9	25.1	35.1	9.3	13.3	54.9	82.9

Fountain (West Bay)	70.3	25.1	35.1	9.3	13.3	54.9	82.9
Middle	70.5	30.7	29.3	7.7	20.8	54.6	83.2
Little Tuttle	70.9	20.0	22.0	9.4	9.6	54.8	82.8
Walnut	72.8	27.5	24.2	8.9	16.5	54.9	82.1
Charlotte	73.4	22.2	32.3	9.5	10.7	55.1	83.5
Madison	75.2	30.7	29.3	7.7	20.8	54.6	83.2
Bear	81.4	25.1	35.1	9.3	13.3	54.9	82.9
Little Twin	82.6	20.0	22.0	9.4	9.6	54.8	82.8
Clear	88.9	31.3	39.3	8.4	20.5	42.5	83.9
Wita	89.2	30.7	29.3	7.7	20.8	54.6	83.2
Oak Glen (Main Bay)	90.0	27.4	31.3	10.4	14.9	53.9	82.8
Willmert (Main Bay)	92.3	20.0	22.0	9.4	9.6	54.8	82.8
Goose	92.4	30.1	31.7	9.0	18.3	53.9	82.4
Clayton	93.6	19.5	32.2	10.6	7.4	54.8	81.8
Ida	94.2	22.2	32.3	9.5	10.7	55.1	83.5
Born	98.4	31.4	34.5	8.6	20.2	54.6	83.2
Penny	100.4	25.1	35.1	9.3	13.3	54.9	82.9
Mott	101.2	30.1	31.7	9.0	18.3	53.9	82.4

East Chain	101.9	19.5	32.2	10.6	7.4	54.8	81.8
Gilman	104.0	31.3	39.3	8.4	20.5	42.5	83.9
Okamanpeedan	105.0	19.5	32.2	10.6	7.4	54.8	81.8
Clam	105.4	20.0	22.0	9.4	9.6	54.8	82.8
Perch	108.2	30.1	31.7	9.0	18.3	53.9	82.4
Morin	108.3	27.5	24.2	8.9	16.5	54.9	82.1
Clear	111.8	20.0	22.0	9.4	9.6	54.8	82.8
Kiester	112.4	22.2	32.3	9.5	10.7	55.1	83.5
Perry	118.1	20.0	22.0	9.4	9.6	54.8	82.8
Albert Lea	119.2	25.1	35.1	9.3	13.3	54.9	82.9
Sisseton	121.6	19.9	30.9	9.1	8.8	55.1	83.5
Fish	124.0	20.0	22.0	9.4	9.6	54.8	82.8
Ewy	126.6	26.2	34.5	9.6	13.4	55.8	84.3
Willow Reservoir 6A	126.8	24.3	37.6	9.3	12.8	54.1	81.5
Domeier	127.5	31.3	39.3	8.4	20.5	42.5	83.9
Martin	129.0	22.2	32.3	9.5	10.7	55.1	83.5
Loon	131.8	30.1	31.7	9.0	18.3	53.9	82.4
Imogene	133.8	19.5	32.2	10.6	7.4	54.8	81.8
Fish	135.7	20.0	22.0	9.4	9.6	54.8	82.8
Strom	136.2	31.3	39.3	8.4	20.5	42.5	83.9

Buffalo	137.0	20.0	22.0	9.4	9.6	54.8	82.8
Rice	141.7	27.4	31.3	10.4	14.9	53.9	82.8
Albert	144.1	30.7	29.3	7.7	20.8	54.6	83.2
White	145.5	25.1	35.1	9.3	13.3	54.9	82.9
Bright	146.1	20.0	22.0	9.4	9.6	54.8	82.8
Sager	151.5	19.5	32.2	10.6	7.4	54.8	81.8
Indian	157.8	30.1	31.7	9.0	18.3	53.9	82.4
Murphy	158.9	22.2	32.3	9.5	10.7	55.1	83.5
Fox	162.9	20.0	22.0	9.4	9.6	54.8	82.8
Fountain (East Bay)	163.9	25.1	35.1	9.3	13.3	54.9	82.9
Willmert (South Bay)	166.1	20.0	22.0	9.4	9.6	54.8	82.8
Buffalo	168.3	22.2	32.3	9.5	10.7	55.1	83.5
High	171.2	22.2	32.3	9.5	10.7	55.1	83.5
School Section	172.0	25.1	35.1	9.3	13.3	54.9	82.9
Lonergan	174.5	29.4	36.5	10.7	15.6	53.9	82.4
Round	179.9	26.2	34.5	9.6	13.4	55.8	84.3
Albert Lea	182.2	25.1	35.1	9.3	13.3	54.9	82.9
Iowa	182.8	20.0	22.0	9.4	9.6	54.8	82.8
Somsen	188.6	31.3	39.3	8.4	20.5	42.5	83.9

Halls	194.3	19.5	32.2	10.6	7.4	54.8	81.8
Rice	195.7	27.5	24.2	8.9	16.5	54.9	82.1
Case	209.5	26.2	34.5	9.6	13.4	55.8	84.3
Crystal	227.8	27.1	25.0	8.7	16.5	53.9	82.8
Iowa	233.5	20.0	22.0	9.4	9.6	54.8	82.8
Long	239.6	30.7	29.3	7.7	20.8	54.6	83.2
Pickeral	248.7	25.1	35.1	9.3	13.3	54.9	82.9
Canright	287.9	22.2	32.3	9.5	10.7	55.1	83.5
Zanders	302.9	31.3	39.3	8.4	20.5	42.5	83.9
Albert Lea	303.5	25.1	35.1	9.3	13.3	54.9	82.9
Boise	346.0	22.3	32.6	7.7	12.0	42.5	83.9

Table 11 - Appendix. From 160 lakes in the Western Cornbelt Plains; 2020 Average monthly chlorophyll, annual precipitation (an), snowfall (snow), March-April-May precipitation (MAM), June-July-August precipitation (JJA), March-April May average temperature(MAMT), June-July-August average temperature (JJAT).

	Year: 2019						
Lake	chla	an	snow	mam	jja	mamt	jjat
Bamber	2.3	37.0	71.5	14.8	18.0	50.6	78.2
George	14.5	37.0	71.5	14.8	18.0	50.6	78.2
Hiniker Pond	4.0	32.6	57.6	12.8	15.9	50.3	78.5
Bamber	3.6	37.0	71.5	14.8	18.0	50.6	78.2

Juni	7.7	27.4	65.1	12.6	11.9	50.4	77.8
Kohlmeier	7.1	36.6	65.4	13.8	19.0	50.2	79.5
Madison	12.3	30.8	57.2	16.0	11.9	50.3	78.5
Zumbro	21.9	32.6	59.5	12.3	15.9	50.6	78.6
Beaver	5.0	31.5	69.1	14.9	11.8	50.0	78.9
Eberhart	14.1	25.6	56.5	12.8	9.4	50.5	79.5
Silver	12.1	37.0	71.5	14.8	18.0	50.6	78.2
Mill Pond	8.7	28.0	58.9	11.6	11.3	50.6	78.5
Bullhead	5.8	28.9	55.7	14.5	10.8	50.9	80.0
Sleepy Eye	8.8	27.4	65.1	12.6	11.9	50.4	77.8
Goose	26.4	28.3	42.4	11.6	13.7	51.9	79.5
Erickson	21.1	27.4	65.1	12.6	11.9	50.4	77.8
Middle	14.7	30.8	57.2	16.0	11.9	50.3	78.5
Fedji	37.2	28.9	55.7	14.5	10.8	50.9	80.0
Peterson	10.8	27.4	65.1	12.6	11.9	50.4	77.8
Oak Leaf	37.8	30.8	57.2	16.0	11.9	50.3	78.5
School	19.2	28.9	55.7	14.5	10.8	50.9	80.0
Swan	8.3	27.4	65.1	12.6	11.9	50.4	77.8
St. James	50.1	28.9	55.7	14.5	10.8	50.9	80.0
Seymour	25.8	28.6	54.0	14.9	10.0	50.8	79.5
Long	35.1	28.9	55.7	14.5	10.8	50.9	80.0
Perch	19.8	32.0	59.4	12.6	15.1	50.0	78.9

Upper Twin	29.1	25.6	56.5	12.8	9.4	50.5	79.5
Silver	123.1	32.0	59.4	12.6	15.1	50.0	78.9
Cedar	16.1	28.6	54.0	14.9	10.0	50.8	79.5
Freeborn	20.8	28.3	42.4	11.6	13.7	51.9	79.5
Kansas	15.8	28.9	55.7	14.5	10.8	50.9	80.0
South Walnut	16.5	28.3	42.4	11.6	13.7	51.9	79.5
Sulem	18.1	28.9	55.7	14.5	10.8	50.9	80.0
Goose	27.3	25.6	56.5	12.8	9.4	50.5	79.5
Wood	13.9	28.9	55.7	14.5	10.8	50.9	80.0
Mills	65.1	32.6	57.6	12.8	15.9	50.3	78.5
Alice	7.6	32.0	59.4	12.6	15.1	50.0	78.9
Altermatt	37.9	27.4	65.1	12.6	11.9	50.4	77.8
Freeborn	31.6	28.3	42.4	11.6	13.7	51.9	79.5
Bass	41.5	29.3	61.7	13.7	12.7	51.3	79.8
Eagle (South)	14.1	32.6	57.6	12.8	15.9	50.3	78.5
Perch	21.2	32.0	59.4	12.6	15.1	50.0	78.9
Knights	24.5	32.0	59.4	12.6	15.1	50.0	78.9
Big Twin	58.4	28.6	54.0	14.9	10.0	50.8	79.5
Lily	17.2	32.6	57.6	12.8	15.9	50.3	78.5
Clear	70.3	32.0	59.4	12.6	15.1	50.0	78.9
Minnesota	24.7	28.3	42.4	11.6	13.7	51.9	79.5

Rice	41.1	32.0	59.4	12.6	15.1	50.0	78.9
St. Olaf	4.8	31.5	69.1	14.9	11.8	50.0	78.9
Creek	17.5	28.6	54.0	14.9	10.0	50.8	79.5
Cedar	45.7	28.6	54.0	14.9	10.0	50.8	79.5
Hanska	64.0	28.9	55.7	14.5	10.8	50.9	80.0
Hall	67.9	26.5	75.4	12.0	12.1	50.4	78.3
Spring	51.1	32.6	57.6	12.8	15.9	50.3	78.5
Omsrud	53.1	28.9	55.7	14.5	10.8	50.9	80.0
Armstrong	24.1	27.4	65.1	12.6	11.9	50.4	77.8
Irish	18.7	28.9	55.7	14.5	10.8	50.9	80.0
Mary	96.2	28.9	55.7	14.5	10.8	50.9	80.0
Eagle (North)	72.6	32.6	57.6	12.8	15.9	50.3	78.5
George	154.5	26.5	75.4	12.0	12.1	50.4	78.3
Mountain	34.8	27.3	55.4	13.9	10.3	52.1	81.3
Lieberg	78.1	27.4	65.1	12.6	11.9	50.4	77.8
Madison	48.1	30.8	57.2	16.0	11.9	50.3	78.5
Crystal	43.7	32.6	57.6	12.8	15.9	50.3	78.5
Lura	46.9	29.3	61.7	13.7	12.7	51.3	79.8
Willow Creek	111.4	28.9	55.7	14.5	10.8	50.9	80.0
Buffalo	8.6	32.0	59.4	12.6	15.1	50.0	78.9
Geneva	44.6	25.6	56.5	12.8	9.4	50.5	79.5

George	55.0	30.8	57.2	16.0	11.9	50.3	78.5
Budd	125.9	32.1	57.5	16.4	12.7	51.3	79.8
Amber	23.3	32.1	57.5	16.4	12.7	51.3	79.8
Silver	94.2	32.0	59.4	12.6	15.1	50.0	78.9
Susan	107.2	28.6	54.0	14.9	10.0	50.8	79.5
Severson	18.7	28.6	54.0	14.9	10.0	50.8	79.5
Loon	35.3	32.6	57.6	12.8	15.9	50.3	78.5
North Silver	69.4	28.6	54.0	14.9	10.0	50.8	79.5
Eagle (North)	26.1	32.6	57.6	12.8	15.9	50.3	78.5
East Side	18.8	25.6	56.5	12.8	9.4	50.5	79.5
Butterfield	59.9	27.4	65.1	12.6	11.9	50.4	77.8
Swag	145.7	26.5	75.4	12.0	12.1	50.4	78.3
South Silver	38.1	26.5	75.4	12.0	12.1	50.4	78.3
Pierce	9.4	28.6	54.0	14.9	10.0	50.8	79.5
Rose	105.7	26.5	75.4	12.0	12.1	50.4	78.3
Lower Twin	22.9	25.6	56.5	12.8	9.4	50.5	79.5
Eagle	29.5	28.6	54.0	14.9	10.0	50.8	79.5
School	18.3	28.9	55.7	14.5	10.8	50.9	80.0
Faribault	39.9	34.3	63.2	14.4	16.4	50.2	79.5
Temperance	100.8	28.6	54.0	14.9	10.0	50.8	79.5
Linden	100.1	27.4	65.1	12.6	11.9	50.4	77.8

Geneva	89.2	25.6	56.5	12.8	9.4	50.5	79.5
Fountain (West Bay)	62.8	25.6	56.5	12.8	9.4	50.5	79.5
Middle	18.4	30.8	57.2	16.0	11.9	50.3	78.5
Little Tuttle	46.8	28.6	54.0	14.9	10.0	50.8	79.5
Walnut	15.3	28.3	42.4	11.6	13.7	51.9	79.5
Charlotte	194.4	29.3	61.7	13.7	12.7	51.3	79.8
Madison	232.2	30.8	57.2	16.0	11.9	50.3	78.5
Bear	38.9	25.6	56.5	12.8	9.4	50.5	79.5
Little Twin	61.1	28.6	54.0	14.9	10.0	50.8	79.5
Clear	177.7	27.4	65.1	12.6	11.9	50.4	77.8
Wita	72.8	30.8	57.2	16.0	11.9	50.3	78.5
Oak Glen (Main Bay)	17.5	36.6	65.4	13.8	19.0	50.2	79.5
Willmert (Main Bay)	26.5	28.6	54.0	14.9	10.0	50.8	79.5
Goose	37.1	32.0	59.4	12.6	15.1	50.0	78.9
Clayton	105.4	26.5	75.4	12.0	12.1	50.4	78.3
Ida	166.8	29.3	61.7	13.7	12.7	51.3	79.8
Born	20.7	32.6	57.6	12.8	15.9	50.3	78.5
Penny	111.3	25.6	56.5	12.8	9.4	50.5	79.5

Mott	56.4	32.0	59.4	12.6	15.1	50.0	78.9
East Chain	96.7	26.5	75.4	12.0	12.1	50.4	78.3
Gilman	89.4	27.4	65.1	12.6	11.9	50.4	77.8
Okamanpeedan	36.5	26.5	75.4	12.0	12.1	50.4	78.3
Clam	192.4	28.6	54.0	14.9	10.0	50.8	79.5
Perch	74.5	32.0	59.4	12.6	15.1	50.0	78.9
Morin	76.5	28.3	42.4	11.6	13.7	51.9	79.5
Clear	125.5	28.6	54.0	14.9	10.0	50.8	79.5
Kiester	215.0	29.3	61.7	13.7	12.7	51.3	79.8
Perry	120.4	28.6	54.0	14.9	10.0	50.8	79.5
Albert Lea	81.2	25.6	56.5	12.8	9.4	50.5	79.5
Sisseton	164.0	32.1	57.5	16.4	12.7	51.3	79.8
Fish	155.1	28.6	54.0	14.9	10.0	50.8	79.5
Ewy	22.7	28.9	55.7	14.5	10.8	50.9	80.0
Willow Reservoir 6A	32.3	37.0	71.5	14.8	18.0	50.6	78.2
Domeier	333.6	27.4	65.1	12.6	11.9	50.4	77.8
Martin	85.5	29.3	61.7	13.7	12.7	51.3	79.8
Loon	96.5	32.0	59.4	12.6	15.1	50.0	78.9
Imogene	97.4	26.5	75.4	12.0	12.1	50.4	78.3
Fish	37.5	28.6	54.0	14.9	10.0	50.8	79.5

Strom	30.5	27.4	65.1	12.6	11.9	50.4	77.8
Buffalo	71.5	28.6	54.0	14.9	10.0	50.8	79.5
Rice	184.2	36.6	65.4	13.8	19.0	50.2	79.5
Albert	19.0	30.8	57.2	16.0	11.9	50.3	78.5
White	32.5	25.6	56.5	12.8	9.4	50.5	79.5
Bright	158.2	28.6	54.0	14.9	10.0	50.8	79.5
Sager	119.6	26.5	75.4	12.0	12.1	50.4	78.3
Indian	14.0	32.0	59.4	12.6	15.1	50.0	78.9
Murphy	53.4	29.3	61.7	13.7	12.7	51.3	79.8
Fox	166.5	28.6	54.0	14.9	10.0	50.8	79.5
Fountain (East Bay)	51.0	25.6	56.5	12.8	9.4	50.5	79.5
Willmert (South Bay)	94.2	28.6	54.0	14.9	10.0	50.8	79.5
Buffalo	116.4	29.3	61.7	13.7	12.7	51.3	79.8
High	99.0	29.3	61.7	13.7	12.7	51.3	79.8
School Section	143.2	25.6	56.5	12.8	9.4	50.5	79.5
Lonergan	110.2	31.5	69.1	14.9	11.8	50.0	78.9
Round	214.7	28.9	55.7	14.5	10.8	50.9	80.0
Albert Lea	179.9	25.6	56.5	12.8	9.4	50.5	79.5
Iowa	249.9	28.6	54.0	14.9	10.0	50.8	79.5

Somsen	80.8	27.4	65.1	12.6	11.9	50.4	77.8
Halls	266.3	26.5	75.4	12.0	12.1	50.4	78.3
Rice	225.6	28.3	42.4	11.6	13.7	51.9	79.5
Case	137.6	28.9	55.7	14.5	10.8	50.9	80.0
Crystal	170.6	34.3	63.2	14.4	16.4	50.2	79.5
Iowa	294.0	28.6	54.0	14.9	10.0	50.8	79.5
Long	123.4	30.8	57.2	16.0	11.9	50.3	78.5
Pickeral	141.6	25.6	56.5	12.8	9.4	50.5	79.5
Canright	98.1	29.3	61.7	13.7	12.7	51.3	79.8
Zanders	80.4	27.4	65.1	12.6	11.9	50.4	77.8
Albert Lea	115.1	25.6	56.5	12.8	9.4	50.5	79.5
Boise	7.8	29.4	50.4	13.6	12.3	50.4	77.8

Table 12 - Appendix. Full list of ranking of hyperparameter setups based on RMSE for 2 hidden layers, tested on multivariate forecast of 2019 data.

Nodes-1	Nodes-2	Epochs	Batches	RMSE	Rank
10	20	150	40	40.5	1
20	20	150	1	40.5	2
10	20	50	1	42.4	3
10	20	50	40	46.1	4
20	30	50	1	46.4	5
30	10	50	1	52.6	6

10	30	50	40	58.2	7
30	20	50	1	65.8	8
20	30	150	40	68.9	9
10	30	150	1	84.7	10
30	10	150	40	85.1	11
10	30	50	1	92.4	12
30	20	50	40	102.2	13
20	10	50	1	117.3	14
20	10	150	1	128.3	15
10	20	150	1	174.1	16
20	30	150	1	180.4	17
10	10	50	1	216.0	18
30	20	150	40	243.1	19
30	10	50	40	268.5	20
10	10	150	40	270.8	21
20	10	50	40	337.3	22
10	10	50	40	372.2	23
20	30	50	40	523.8	24
30	30	50	1	623.8	25
30	10	150	1	650.8	26
20	20	50	40	800.5	27
20	10	150	40	1147.5	28

30	30	50	40	1257.1	29
20	20	150	40	2206.2	30
20	20	50	1	4336.1	31
30	20	150	1	4343.5	32
10	30	150	40	4506.4	33
10	10	150	1	7768.2	34
30	30	150	1	18642.7	35
30	30	150	40	1017599.8	36

Table 13 - Appendix. Full list of ranking of hyperparameter setups based on RMSE for 2 hidden layers, tested on multivariate forecast of 2020 data.

Nodes-1	Nodes-2	Epochs	Batches	RMSE	Rank
30	20	150	40	13.2	1
30	30	50	1	22.5	2
10	10	50	40	27.7	3
10	20	150	40	29.3	4
30	10	150	1	34.4	5
10	20	150	1	36.8	6
10	10	150	1	36.8	7
30	20	150	1	37.7	8
10	30	150	1	43.5	9
20	30	50	1	47.1	10

10	20	50	1	51.6	11
10	20	50	40	57.6	12
30	30	150	1	61.2	13
20	20	50	40	73.8	14
10	30	50	40	74.2	15
20	10	50	1	84.7	16
20	30	50	40	88.5	17
30	10	150	40	98.8	18
20	30	150	1	114.2	19
20	20	50	1	122.9	20
30	20	50	1	130.3	21
30	10	50	1	165.5	22
30	30	150	40	186.5	23
10	30	50	1	241.1	24
20	30	150	40	290.7	25
20	10	150	1	296.4	26
20	20	150	1	353.1	27
20	10	50	40	448.4	28
30	20	50	40	448.9	29
10	10	50	1	830.0	30
10	10	150	40	1048.2	31
20	20	150	40	1268.0	32

20	10	150	40	1338.4	33
30	30	50	40	2390.4	34
10	30	150	40	2583.0	35
30	10	50	40	8728.2	36

Table 14 - Appendix. Raw data for multivariate LSTM forecast of chlorophyll-a, from Bass Lake in 2019.

Date	Chlorophyll-a ug/L	Turbidity (NTU)	Conductivity (us/cm)	Temp (C)	Precip. (inches)
6/23/2019	17.3	5.5	351.6	21.4	3.1
6/24/2019	17.3	5.7	352.8	21.5	3.2
6/25/2019	17.3	5.7	350.6	21.8	3.5
6/26/2019	17.3	5.6	348.8	22.9	3.6
6/27/2019	17.6	5.0	349.7	23.3	2.7
6/28/2019	17.3	5.4	348.2	23.5	2.5
6/29/2019	17.3	5.5	347.5	25.4	2.7
6/30/2019	16.8	5.0	347.0	26.6	2.7
7/1/2019	16.7	5.2	347.0	25.8	2.7
7/2/2019	17.3	4.9	347.2	25.7	2.9
7/3/2019	16.5	5.4	340.4	26.5	3.9
7/4/2019	17.2	5.4	339.5	26.7	3.9
7/5/2019	17.1	11.4	339.8	27.1	3.9
7/6/2019	16.9	11.4	340.6	26.6	3.8
7/7/2019	16.7	11.4	338.5	27.1	4.2
7/8/2019	17.3	11.4	338.9	27.0	4.2
7/9/2019	17.3	11.4	332.4	23.1	4.2
7/10/2019	17.3	11.4	332.4	23.1	4.4
7/11/2019	17.3	9.2	362.8	26.7	4.5
7/12/2019	20.7	9.0	366.5	25.9	4.4
7/13/2019	22.0	9.0	362.8	26.2	4.1

7/14/2019	22.3	10.8	361.0	26.7	4.2
7/15/2019	21.8	8.9	363.9	27.4	4.2
7/16/2019	22.7	9.5	365.4	26.6	4.0
7/17/2019	22.1	9.8	366.5	26.0	4.0
7/18/2019	20.9	10.4	365.4	27.0	4.2
7/19/2019	21.6	13.4	367.9	28.1	4.6
7/20/2019	21.6	11.3	364.7	27.3	4.7
7/21/2019	22.1	13.2	365.7	26.1	4.9
7/22/2019	17.3	12.1	361.8	25.8	5.0
7/23/2019	24.5	13.8	361.7	26.2	5.0
7/24/2019	25.2	13.3	361.7	26.1	5.0
7/25/2019	17.3	13.4	361.3	25.1	4.7
7/26/2019	17.3	12.7	366.7	24.1	4.7
7/27/2019	17.3	15.0	353.0	27.5	4.9
7/28/2019	17.3	5.9	357.3	26.1	4.6
7/29/2019	17.3	7.8	357.0	25.5	4.4
7/30/2019	17.3	10.3	359.4	26.0	4.6
7/31/2019	17.3	11.1	355.1	25.8	4.6
8/1/2019	17.3	11.4	352.5	25.7	4.2
8/2/2019	17.3	13.8	345.7	26.1	3.3
8/3/2019	17.3	11.4	341.6	27.1	3.3
8/4/2019	17.3	9.1	332.6	27.1	3.3
8/5/2019	17.3	8.5	330.3	26.4	3.1
8/6/2019	17.3	10.1	338.9	26.7	2.8
8/7/2019	17.3	10.5	339.2	27.0	3.0
8/8/2019	17.3	12.0	337.8	26.8	3.0
8/9/2019	17.3	18.2	338.5	26.8	2.8
8/10/2019	17.3	13.4	335.1	26.3	2.7
8/11/2019	17.3	13.8	339.3	26.0	2.9
8/12/2019	17.3	15.7	341.3	26.0	3.1
8/13/2019	17.3	12.4	341.8	25.7	3.0
8/14/2019	17.3	13.1	344.5	25.3	3.0
8/15/2019	17.3	12.1	347.9	25.0	3.0

8/16/2019	17.3	12.6	340.7	24.8	2.5
8/17/2019	16.1	13.2	327.7	25.3	2.4
8/18/2019	16.4	10.4	329.9	25.0	2.0
8/19/2019	17.3	13.2	315.8	24.9	2.2
8/20/2019	17.0	13.1	328.8	25.5	2.0
8/21/2019	16.9	9.1	329.5	25.2	2.0
8/22/2019	17.3	8.8	329.9	24.9	2.0
8/23/2019	16.0	8.8	329.9	24.8	2.0
8/24/2019	16.4	9.7	330.2	24.1	2.0
8/25/2019	16.6	9.6	331.0	23.4	1.9
8/26/2019	17.3	8.1	331.7	23.1	1.6
8/27/2019	17.4	9.4	331.1	22.8	1.7
8/28/2019	17.6	9.6	330.7	22.1	1.8
8/29/2019	17.0	10.6	330.1	22.2	1.7
8/30/2019	16.5	8.6	331.0	22.4	1.7
8/31/2019	16.8	10.5	331.2	21.7	1.7
9/1/2019	16.6	10.6	332.4	21.3	1.7
9/2/2019	16.3	9.4	332.6	22.1	1.8
9/3/2019	16.9	9.4	334.4	22.3	1.8
9/4/2019	16.5	7.3	333.5	22.5	1.7
9/5/2019	16.5	9.7	333.9	22.1	1.6
9/6/2019	16.9	10.2	334.9	22.9	1.4
9/7/2019	18.1	9.1	335.0	22.1	1.4
9/8/2019	17.6	13.9	334.9	20.9	1.4
9/9/2019	17.0	13.4	337.5	20.2	1.6
9/10/2019	16.8	11.5	337.0	21.4	1.4
9/11/2019	17.1	14.4	330.9	22.1	1.4
9/12/2019	17.4	20.1	324.8	21.5	2.4
9/13/2019	18.0	15.1	328.1	20.7	4.0
9/14/2019	18.7	15.7	329.3	20.3	4.4
9/15/2019	19.2	26.1	329.6	22.2	4.4
9/16/2019	18.0	40.8	327.8	22.3	4.2
9/17/2019	17.2	20.2	328.4	22.7	4.2

9/18/2019	17.5	18.9	325.8	22.9	3.9
9/19/2019	17.4	15.0	322.7	23.5	4.8
9/20/2019	17.0	18.5	318.1	24.4	5.1
9/21/2019	16.6	16.9	320.3	23.8	5.1
9/22/2019	16.3	11.5	319.0	23.0	5.1
9/23/2019	16.2	13.6	321.6	22.2	5.1
9/24/2019	16.4	13.6	320.8	22.1	5.1
9/25/2019	16.8	13.8	321.5	21.6	5.3
9/26/2019	16.6	14.6	322.0	20.6	5.8
9/27/2019	17.0	15.4	323.4	20.1	5.7
9/28/2019	17.0	18.9	323.3	19.4	5.7
9/29/2019	16.7	18.3	324.3	18.3	5.7
9/30/2019	16.6	17.4	327.0	19.4	5.7
10/1/2019	17.1	19.3	322.8	18.9	5.7
10/2/2019	17.0	21.0	324.6	17.5	6.6
10/3/2019	17.1	15.3	324.6	16.9	7.4
10/4/2019	17.2	17.9	323.9	15.7	7.4
10/5/2019	17.3	17.5	323.1	15.0	7.4
10/6/2019	17.9	18.2	323.0	14.7	8.2
10/7/2019	18.1	15.3	322.7	14.5	8.4
10/8/2019	17.5	18.2	324.3	14.3	8.4
10/9/2019	17.5	18.0	325.4	14.3	8.3
10/10/2019	18.3	17.0	325.2	14.4	8.1
10/11/2019	18.5	17.7	325.9	12.9	8.1
10/12/2019	18.6	19.0	324.4	10.6	7.1
10/13/2019	18.9	13.1	325.8	9.5	5.5
10/14/2019	17.5	11.7	325.2	9.3	5.1
10/15/2019	18.1	10.8	326.3	9.2	5.1
10/16/2019	19.4	7.0	327.3	9.4	5.1
10/17/2019	17.7	7.4	327.4	9.5	5.1
10/18/2019	17.1	9.3	329.3	9.7	5.1
10/19/2019	18.9	4.5	329.1	10.1	4.2
10/20/2019	18.1	7.1	330.0	10.7	3.7

10/21/2019	18.7	8.7	329.6	10.6	3.7
10/22/2019	18.9	6.0	328.4	9.8	4.4
10/23/2019	19.4	3.6	328.8	9.0	5.0
10/24/2019	19.7	3.7	329.6	8.6	5.0
10/25/2019	19.7	3.2	329.3	8.1	4.8

Table 15 – Appendix. Raw data for multivariate LSTM forecast of chlorophyll-a, from Bass Lake in 2019.

Date	Chlorophyll-a (µg/L)	Turbidity (NTU)	Conductiity (µs/cm)	Temperature ©	Precip. (inches)
6/10/2020	13.7	3.5	360.2	23.2	4.8
6/11/2020	14.2	4.3	361.1	22.3	4.8
6/12/2020	14.3	5.0	363.0	22.7	4.8
6/13/2020	13.9	8.1	362.5	23.2	4.8
6/14/2020	13.7	7.9	363.0	22.1	4.7
6/15/2020	13.8	6.2	364.1	21.7	2.9
6/16/2020	13.8	6.0	362.6	22.4	2.8
6/17/2020	13.5	6.7	361.8	23.0	2.8
6/18/2020	13.7	6.2	363.4	23.3	3.0
6/19/2020	13.6	8.8	359.8	23.6	3.8
6/20/2020	13.7	8.1	360.6	24.0	3.8
6/21/2020	13.3	7.8	359.0	23.8	4.0
6/22/2020	12.6	8.1	350.0	23.0	4.7
6/23/2020	13.1	6.9	351.9	23.4	4.5
6/24/2020	13.1	7.2	353.9	23.1	4.4
6/25/2020	12.6	8.1	350.0	23.0	4.1
6/26/2020	13.1	6.7	357.0	24.2	4.2
6/27/2020	13.2	5.2	363.9	26.6	4.2
6/28/2020	13.9	7.2	364.3	26.4	4.2
6/29/2020	13.9	6.6	364.2	25.4	4.2
6/30/2020	13.8	7.8	365.3	25.7	3.0
7/1/2020	13.6	7.0	364.6	26.4	2.8
7/2/2020	13.3	5.8	362.9	28.0	2.6
7/3/2020	13.3	6.6	357.2	28.7	2.6
7/4/2020	13.2	8.4	364.7	29.6	2.5
7/5/2020	13.3	6.9	364.7	30.1	2.5
7/6/2020	14.2	4.8	365.4	29.8	2.5
7/7/2020	15.0	4.3	366.0	30.0	2.5
7/8/2020	15.2	4.7	365.6	29.8	2.4

7/9/2020	15.7	5.6	366.1	28.7	2.5
7/10/2020	15.2	5.6	365.9	28.5	2.7
7/11/2020	15.3	6.5	364.7	28.7	2.8
7/12/2020	15.6	7.1	364.9	28.4	3.1
7/13/2020	16.5	7.0	365.7	27.9	3.1
7/14/2020	15.4	6.6	366.8	27.3	3.1
7/15/2020	15.9	6.5	365.3	26.8	3.1
7/16/2020	16.2	5.9	366.1	26.5	3.1
7/17/2020	15.8	6.6	369.6	26.7	2.9
7/18/2020	15.6	7.2	370.2	27.4	2.6
7/19/2020	15.7	6.8	370.4	27.7	2.7
7/20/2020	15.7	6.6	369.1	27.9	2.4
7/21/2020	16.8	7.0	369.1	27.4	1.6
7/22/2020	16.6	6.2	367.6	27.2	1.6
7/23/2020	14.9	7.3	366.0	27.1	1.6
7/24/2020	15.0	7.8	365.7	27.3	1.6
7/25/2020	15.2	7.8	367.4	27.4	1.5
7/26/2020	15.7	8.1	358.6	26.9	3.7
7/27/2020	16.0	6.7	359.6	27.0	4.2
7/28/2020	16.4	6.7	356.6	27.0	4.1
7/29/2020	15.9	8.1	355.2	27.8	4.1
7/30/2020	15.2	7.1	351.1	27.4	4.0
7/31/2020	15.7	6.3	352.9	27.3	4.0
8/1/2020	16.0	5.7	352.6	27.3	4.0
8/2/2020	15.7	6.3	347.3	26.8	4.0
8/3/2020	15.4	5.6	352.8	26.2	4.0
8/4/2020	15.3	6.2	353.5	26.1	4.0
8/5/2020	15.6	6.3	354.3	25.1	4.0
8/6/2020	15.5	6.4	356.6	24.8	4.0
8/7/2020	15.3	7.1	356.5	24.9	3.8
8/8/2020	15.4	7.9	356.5	25.2	3.7
8/9/2020	15.2	10.1	355.9	25.5	3.9
8/10/2020	15.4	21.0	356.4	25.5	3.7
8/11/2020	13.1	40.9	353.8	24.5	3.8
8/12/2020	13.5	9.0	349.6	23.6	4.3
8/13/2020	15.5	12.7	352.3	25.2	4.7
8/14/2020	15.5	11.4	350.6	25.8	4.8
8/15/2020	15.6	12.5	346.2	26.0	5.1
8/16/2020	16.6	13.6	344.7	26.4	4.7
8/17/2020	15.7	12.6	340.0	26.2	4.6

8/18/2020	14.5	11.3	341.4	26.6	4.6
8/19/2020	14.7	11.2	341.4	25.6	4.5
8/20/2020	15.1	10.6	344.2	25.0	4.5
8/21/2020	15.5	12.1	344.8	25.2	4.5
8/22/2020	15.0	12.4	341.4	25.4	5.8
8/23/2020	14.9	12.8	341.5	26.3	6.0
8/24/2020	14.6	12.5	339.3	27.1	3.8
8/25/2020	14.8	13.9	336.2	27.2	3.3
8/26/2020	14.4	15.0	334.0	26.9	3.3
8/27/2020	12.6	17.8	334.6	28.1	3.3
8/28/2020	13.6	13.6	333.5	27.4	3.8
8/29/2020	13.2	15.3	332.0	26.4	4.0
8/30/2020	13.0	20.4	331.6	25.5	4.0
8/31/2020	12.6	36.6	332.5	24.8	4.7
9/1/2020	12.8	26.7	332.8	23.8	4.8
9/2/2020	12.8	15.6	332.6	23.4	4.8
9/3/2020	13.1	14.9	333.6	22.7	4.8
9/4/2020	13.4	17.1	333.1	22.4	4.8
9/5/2020	15.0	44.3	332.9	23.4	4.8
9/6/2020	14.5	39.7	331.4	22.5	5.1
9/7/2020	14.2	21.0	332.3	21.3	4.9
9/8/2020	14.3	17.8	333.2	19.8	4.8
9/9/2020	14.2	16.9	333.4	18.5	4.9
9/10/2020	13.8	19.0	333.5	17.8	4.4
9/11/2020	14.7	30.6	332.2	17.2	3.9
9/12/2020	12.9	16.5	332.8	17.0	3.9
9/13/2020	13.3	21.3	331.5	17.6	3.6
9/14/2020	13.0	24.8	332.7	18.5	3.6
9/15/2020	12.8	18.2	334.9	18.3	3.6
9/16/2020	13.0	15.0	334.2	18.8	3.6
9/17/2020	13.5	18.0	333.6	18.6	3.6
9/18/2020	13.5	20.8	334.4	18.5	3.6
9/19/2020	12.7	23.2	335.4	17.6	3.6
9/20/2020	12.6	18.9	334.2	17.0	2.4
9/21/2020	12.6	17.7	337.3	17.2	2.2
9/22/2020	12.6	16.1	338.3	17.9	2.2
9/23/2020	12.6	21.7	339.4	18.4	2.2
9/24/2020	12.6	23.2	339.7	19.7	2.2
9/25/2020	12.6	24.4	341.5	19.7	2.2
9/26/2020	13.6	30.6	339.5	19.4	1.7

9/27/2020	15.8	14.0	340.4	19.0	1.6
9/28/2020	16.2	18.1	340.2	17.9	1.6
9/29/2020	15.7	18.1	342.3	17.1	0.9
9/30/2020	15.8	18.3	341.7	16.5	0.8
10/1/2020	15.1	15.4	342.9	15.7	0.8
10/2/2020	14.7	20.3	342.2	14.8	0.8
10/3/2020	14.6	25.6	342.1	14.2	0.8
10/4/2020	14.1	15.4	343.8	14.0	0.8
10/5/2020	13.6	19.9	345.7	13.3	0.4
10/6/2020	14.9	17.5	346.6	13.9	0.4
10/7/2020	14.7	14.2	347.5	14.6	0.3
10/8/2020	14.4	20.1	348.6	14.9	0.2
10/9/2020	14.3	16.6	350.6	14.6	0.2
10/10/2020	14.0	16.5	349.5	15.7	0.1
10/11/2020	13.8	19.8	350.5	15.5	0.1
10/12/2020	14.6	15.3	348.1	15.3	2.0
10/13/2020	14.4	15.7	347.4	14.7	2.0
10/14/2020	13.7	15.9	348.4	13.9	2.0
10/15/2020	12.6	12.1	348.8	12.8	2.0
10/16/2020	12.6	12.6	349.3	11.7	2.0
10/17/2020	12.6	32.0	349.5	10.9	2.0
10/18/2020	12.6	42.2	349.0	10.4	2.0
10/19/2020	12.6	46.4	349.9	9.2	2.0
10/20/2020	12.6	36.1	349.5	8.0	2.1
10/21/2020	12.6	28.8	349.8	16.5	2.2
10/22/2020	12.6	4.7	349.5	15.4	2.4
10/23/2020	12.6	1.6	349.3	6.2	2.5
10/24/2020	12.6	4.3	349.5	5.4	2.5
10/25/2020	12.6	0.7	348.8	4.7	2.5
10/26/2020	12.6	0.7	349.7	4.4	2.5
10/27/2020	12.6	0.4	349.9	4.2	2.5
10/28/2020	12.6	0.1	350.3	3.9	2.4
10/29/2020	12.6	0.0	351.2	3.9	2.4
10/30/2020	12.6	3.0	351.3	10.9	2.4
10/31/2020	12.6	8.1	350.0	23.0	2.4
11/1/2020	12.6	8.1	350.0	23.0	2.4
11/2/2020	12.6	8.1	350.0	23.0	2.4
11/3/2020	12.6	8.1	350.0	23.0	2.4
11/4/2020	12.6	8.1	350.0	23.0	2.4
11/5/2020	12.6	8.1	350.0	23.0	2.4

11/6/2020	12.6	8.1	350.0	23.0	2.4