January 2023

Remote Sensing Of Chlorophyll-A In Small And Medium Lakes In The Minnesota Northern Forest Ecoregion

Jacob Alan Moll

REMOTE SENSING OF CHLOROPHYLL-A IN SMALL AND MEDIUM LAKES IN THE MINNESOTA NORTHERN FOREST ECOREGION

By

Jacob Alan Moll
Bachelor of Science Environmental Science and Mathematics, Bowling Green State University
2021

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science in Geography

Grand Forks, North Dakota
May
2023

PERMISSION

Title Remote Sensing of Chlorophyll-A in Small and Medium Lakes in the Minnesota Northern Forest Ecoregion

Department Geography

Degree Master of Science in Geography

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Jacob Alan Moll
May 2023

This thesis submitted by Jacob Alan Moll in partial fulfillment of the requirements for the Degree of Master of Science in Geography from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____

Gregory S. Vandeberg

_____

Jeffery A. VanLooy

_____

Mbongowo Mbuh

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

_____

Chris Nelson
Dean of the School of Graduate Studies

_____

Date

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank the University of Minnesota Itasca Biological Station, specifically Dr. Jim Cotner who provided inside knowledge into the lakes within the park and shared water quality data with me that helped further my understanding of Lake Itasca. I would also like to thank the station scientist, Dan Brumm, for all his help in coordinating the use of station canoes for data collection, and for his help in transporting the canoe around the park.

Thanks to JamesGuy Gierisch for keeping me company on the long days of collecting samples at Itasca State Park and helping paddle the canoe across the lakes. I would also like to acknowledge America View and North Dakota View who provided funding for my study through USGS grant agreement G18AP00077.Thank you to my committee members Dr. Jeff VanLooy and Dr. Mbongowo Mbuh for your review and comments on my thesis. Finally, I would like to thank my advisor Dr. Greg Vandeberg for his words of encouragement and guidance in my two years at UND.

ABSTRACT

Algae blooms have become increasingly common in many bodies of water and are occurring in more and more new locations. Past methods have relied on in-situ sampling to determine Chl-a concentrations and can be time and cost intensive. With the use of NASA's Harmonized Landsat imagery existing methods of index calculation can be implemented, with the added benefit of an increase in temporal resolution. This increase in temporal resolution allows for more detailed spatial and temporal analysis such as hot spot analysis. Harmonized Landsat imagery was proven sufficient in spatial and temporal resolution for detection of Chl-a in small and medium lakes using indices and hot spot analysis. However, index accuracy has been proven to be an issue in this study compared to other studies using Sentinel 2 or Landsat imagery.

CHAPTER 1

INTRODUCTION

1.1 Purpose

An algae bloom is defined as a visible growth of algae on the water and becomes a HAB (Harmful Algal Bloom) when it is harmful to humans, pets, livestock, and other organisms. Other negative effects of HABs include impacts on drinking and irrigation water, recreational activities, and fisheries (Wehr et al. 2015). Algal blooms occur in water bodies throughout the world ranging from the Great Lakes of North America to Hulan, China (Lekki et al. 2019, Cao et al. 2021). HABs have gained media attention when blooms affect the safety of drinking water. One common example of HABs that has been studied extensively occur in the western basin of Lake Erie near Toledo, Ohio in the United States (Lekki et al. 2019, Vincent et al. 2004). In August of 2014 a HAB left the City of Toledo without clean drinking water for three days making national news and bringing the issue of HABs to the eyes of public (Jetoo, Grover and Krantzberg 2015). This 'Toledo Water Crisis' and others drive the need and interest to further study algal blooms not only in the Great Lakes Region but globally.

Remote sensing of water and algal blooms has seen an increase in applications and has proven useful providing data when field collection or verification is difficult or costly. Many spectral indices have been created to aid in the detection of algal blooms. These indices use combinations of visible and near infrared bands as well as ratios to take advantage of the spectral properties of algae. The most common method of detecting algae involves the detection of chlorophyll-a (Chl-a) which has a green color. Because of this many indices take advantage of Chl-a's high reflectance in the near infrared (NIR) wavelengths and high absorption in the red, red-edge wavelengths. Some of the most popular indices include, normalized difference

chlorophyll index (NDCI), two band algorithms (2BDA) such as simple ratios, and three band algorithms (3BDA) (Mishra and Mishra 2012,Tzortziou et al. 2007, Moses et al. 2009).

1.2 Research Needs

Algal blooms are a common occurrence in the Great Lakes region with events like the one that caused the Toledo water crisis. However algal blooms also occur in other regions. One region that has naturally occurring algal blooms that are understudied is the Northern Lakes and Forests region of the United States. Specifically, the small and medium lakes within this region. One unique challenge that this area faces is the high number of small shallow lakes that are not easily studied using remote sensing. Many studies have examined the effectiveness of different satellites with varying spatial and temporal resolution for Chl-a monitoring on the Great Lakes and in the oceans. However, few if any studies have examined how moderate-resolution imagery (20-30m) from satellites such as the European Space Agency's (ESA) Sentinel 2 or the United States Geological Survey's (USGS) Landsat program can be used for the detection of Chl-a in small water bodies in the Northern Lakes and Forests ecoregion. Furthermore, NASA's Harmonized Landsat (HLS) Imagery provides unique advantages implementing moderate spatial resolution imagery at a higher temporal resolution than previously available.

Much of the attention on algal blooms in the Midwest United States has been focused on the Great Lakes. However, recently, in October of 2022, algal blooms in pristine northern Minnesota lakes have received media attention. In the Boundary Waters Canoe Area Wilderness two lakes that are protected and surrounded by wilderness were found to have algal blooms covering the entire water surface (Kraker 2022). While algae blooms occur naturally, they do not usually occur at this magnitude in lakes without extrinsic forces such as agricultural runoff or

changes in surrounding land cover. Few studies have used remote sensing techniques to analyze hotspots for how these blooms develop and grow within lakes.

1.3 Research Questions

This research addresses the lack of use of HLS data for Chl-a detection in small and medium lakes and applies indices derived from the HLS imagery to perform hotspot analysis. Two research questions were addressed in this study. The first question of this study is, if moderate resolution satellite imagery specifically NASA's Harmonized Landsat surface reflectance imagery is sufficient in both spatial and temporal resolution to be used as a proxy for in-situ data regarding algal blooms and their growth in small and medium sized lakes within the Northern Lakes and Forests ecoregion? The second research question that was addressed is, what spectral indices are most accurate in detecting Chl-a in lakes and ponds of the Northern Lakes and Forests ecoregion?

CHAPTER 2

LITERATURE REVIEW

2.1 Chlorophyll-a

This study and many others have used chlorophyll-a as an indicator of algae because it (Chl-a) "is the primary photosynthetic pigment of all oxygen-evolving photosynthetic organisms and is present in all algae and cyanobacteria" (Wetzel 2001). This means that for algae to be present on the lakes, Chl-a must also be present. However, the detection of Chl-a does not always indicate algae presence. This specifically is relevant to detection along coastal and shallow water areas where aquatic vegetation is present, but methods exist to limit the detection of this aquatic vegetation (Zabaleta et al. 2021). Chl-a has two absorption ranges on the electromagnetic spectrum, the first is at 430 nm (blue wavelengths of light) and the second is between 660-665 nm (red wavelengths of light) (Wetzel 2001). While Chl-a absorbs blue and red wavelengths it highly reflects light in the green (about 500-600nm) and has a greater reflectance near infrared wavelengths (about 700-800nm) (Jensen 2014).

2.2 Remote Sensing Satellites and Tools

In the early years of remote sensing of water some of the most popular sensors used were on board polar orbiting operation satellites such as the SeaWiFS Sensor onboard Orbview-2, the Moderate Resolution Imaging Spectrometer (MERIS), and Aqua MODIS. Early on much of the research was focused on more oceanographic applications and as a result sensors like these with spatial resolutions ranging from hundreds of meters to a few kilometers were sufficient. However, with the development of satellites such as Landsat 7 more studies were able to be conducted on smaller study areas than the oceans. One such study was conducted on Lake Erie in 2004 studying the Phycocyanin content in the water utilizing Landsat 5 and Landsat 7. One of

the findings of this study included that the spatial resolution of the Landsat series of satellites, about 30m by 30m was sufficient to study not only Lake Erie but the tributaries of the lake as well (Vincent et al. 2004). Due to the availability of sensors with larger spatial resolution many of the indices mentioned in the previous section were developed and studied using these sensors. Mishra and Mishra developed the NDCI index using MERIS data and before them Moses used MERIS data to evaluate red and near infrared two band and three band algorithm indices. Other studies were completed using Aqua MODIS data. Copado utilized MODIS FLH data products to verify media reported algal blooms in Mexico. Copado also utilized hot spot analysis with the data to discover algal bloom events that were not reported by the media (Copado-Rivera et al. 2020). Another study that was a precursor to Mishra and Mishra (2012) evaluated current MODIS algorithms over the Chesapeake Bay, they found that a red-green ratio performed better than the MODIS algorithms which at the time mostly involved blue and green wavelengths (Tzortziou et al. 2007).

Over time there has been a shift in the research from these more coastal ocean environments towards inland lakes. Recent studies have utilized satellites with a higher spatial resolution such as the European Space Agency's Sentinel 2 or the USGS's Landsat 8. Some of the studies that have utilized Sentinel 2 or Landsat 8 imagery have done so by applying indices such as NDCI, and the two and three band algorithms to the data. They have been successful in utilizing these indices on the data. One study utilized NDCI to map Chl-a concentrations in a lake in South America to find how releasing water from upstream reservoirs affected water quality (Aubriot et al. 2020). Aubriot et al combined Sentinel 2 imagery with hydrologic, and meteorologic data to predict Chl-a models. Other work with Sentinel 2 and Landsat 8 imagery has been completed in Brazil utilizing different band ratios to find which work best for

estimating Chl-a. The findings of this study reported that a near infrared-red ratio with Sentinel 2 bands was the best option (Watanabe et al. 2018). A third study evaluated the 2BDA, 3BDA, FLH, and NDCI indices utilizing a combination of Sentinel 2 and Landsat 8 imagery for Lake Chad in Africa (Buma and Lee 2020).

Landsat 8 and Sentinel 2 data have also been used in coincidence with one another to form a single data set for water clarity analysis. It was found that Modified Atmospheric correction for inland waters (MAIN) was successful for the use of Landsat 8 and Sentinel 2 data when observing water clarity in Minnesota lakes (Paige 2019). Paige found that between the MAIN and ACOLITE atmospheric corrections that bands 2, 3, and 4 of Sentinel 2 and Landsat 8 were very similar to one another ($R^2 > 0.88$). With the start of NASA's Harmonized Landsat and Sentinel-2 project (HLS) users are no longer required to perform atmospheric corrections or even worry about differences in spatial or radiometric resolution between the two satellites themselves. HLS data is processed so that users can treat the data as if it came from one sensor (Masek et al. 2022). This allows users to perform Earth observation analyses at a moderate spatial resolution (30 meters) with a high temporal resolution (2-3 days) using quality data provided by NASA.

Looking at the research that has been done for each satellite, there is substantially less research evaluating Sentinel 2 data than MERIS, MODIS, or even the Landsat series. It appears that recent research is trending towards using Sentinel 2 or Landsat 8 imagery. Out of the first wave of satellites presented in this section MERIS and SeaWiFS are both no longer operational. Meanwhile Sentinel 2 was launched in the past six years, Landsat 8 is about halfway through its lifespan and in 2021 the USGS launched Landsat 9. It appears that the preferred tools, and the

ones more readily available are satellites with sub kilometer spatial resolution. This is one of the reasons that these satellites have been selected for this study.

2.3 Chlorophyll-a Indices

There have been numerous indices proposed and utilized to aid in the detection of algal blooms. Studies observing marine environments and large freshwater lakes have utilized MODIS Fluorescence Line Height (FLH) data products. While other studies have utilized the NDCI (Mishra and Mishra 2012). Other common indices include simple ratios involving visible and near infrared wavelengths (Buma and Lee 2020). When studying algal blooms of inland lakes much of the water has more suspended sediment than marine environments. Due to this many indices that only involve visible wavelengths can be inaccurate for inland freshwater bodies (Moses et al. 2009).

Mishra and Mishra (2012) have shown NDCI to be more accurate for Chl-a estimation than red-green and NIR-red band ratios for the Mississippi River delta, Mobile Bay, and Chesapeake Bay. NDCI is defined as:

$$\frac{R_{rs}(708\ nm) - R_{rs}(665\ nm)}{R_{rs}(708\ nm) + R_{rs}(665\ nm)} \quad (1)$$

Where R_{rs} is defined as remote sensing reflectance at the specified wavelength. NDCI's success is mostly attributed due to its normalization. The reason this combination works so well is the same reason that the Normalized Difference Vegetation Index works well. Chl-a has peak reflectance in near infrared wavelengths and peak absorption in red wavelengths. Buma and Lee (2020) found that NDCI was one of the best indices for estimating Chl-a in Lake Chad, Africa using Landsat 8 Operational Land Imager (OLI) and Sentinel 2 Multi-Spectral Insturment (MSI) imagery. NDCI performed significantly better than 2BDA however it was only marginally better than 3BDA (Buma and Lee 2020).

Simple band ratio indices have also shown success in remote sensing applications. Tzortziou et al. (2007) demonstrated that the red-green ratio defined as:

$$\frac{R_{rs}(667\ nm)}{R_{rs}(554\ nm)} \tag{2}$$

Was utilized within Chesapeake Bay and could be used because non-algal particulate absorption is minimal at these wavelengths (Tzortziou et al. 2007). Another simple two band ratio that has been proven successful is the NIR-red ratio (Moses et al. 2009):

$$\frac{R_{rs}(708\ nm)}{R_{rs}(665\ nm)} \tag{3}$$

This ratio is well suited for detecting algae for similar reasons that NDCI is, at 708 nm Chl-a exhibits high reflectance while at 665 nm it exhibits high abosorbtion. Another reason that some of the blue-green ratios are not utilized for turbid waters is that they have differing optical properties than open water. Turbid waters have more blue wavelength absorbing particles which can complicate the use of these ratios (Moses et al. 2009, Tzortziou et al. 2007). Another index that has had success in the detection of Chl-a is FLH:

$$R_{rs}(560\ nm) - [R_{rs}(665\ nm) + (R_{rs}(480\ nm) - R_{rs}(665\ nm))] \tag{4}$$

FLH has been used along the coast of Veracruz, Mexico for the detection of algae bloom hotspots (Copado-Rivera et al. 2020).

While many of these indices have been developed for optically complex waters, they have been predominantly tested on large water bodies such as Chesapeake Bay, Mobile Bay, Great Lakes, etc. This leaves a lack of research in utilizing these different indices for smaller bodies of water. This study hopes to provide insight into how these different indices perform when applied to smaller lakes in the Northern Lakes and Forests region.

2.4 Hotspot Analysis

The use of hotspot analysis for studying Chl-a distributions has been an area of growth in recent years. Copado-Rivera (2020) used FLH derived hotspot analysis in conjunction with existing field data and news reports to verify the occurrence of HABs in Veracruz, Mexico. In 2021 another study, used hotspot analysis to understand the spatial distribution of Chl-a in artificial suburban lakes in South America (Zabaleta et al. 2021). Zabaleta found that hotspot analysis of remote sensing data allowed for a better understanding of pollutants in the lakes and could be used when field data collection is too expensive or not possible.

One common hotspot analysis method is G_i^*. G_i^* is calculated by comparing surrounding values with one another. G_i^* was created and developed by Getis and Ord (1995). It can be calculated using equation 5.

$$G_i^*(d) = \sum_j \frac{w_{ij}(d)x_j}{\sum_j x_j} \tag{5}$$

Where d is a distance-based neighborhood determined by the user defining the range of which adjacent values affect the target location. W is a symmetrical binary matrix with 1's in locations within the neighborhood distance, d, and 0's in all other locations including the location at x (Getis & Ord 1992).

While G_i^* is helpful in understanding the data, only using equation 5 does not provide the user with any level of significance for their results. To obtain significance, a Z-score must be calculated.

$$E(G_i^*) = \sum_j \frac{w_{ij}(d)}{n-1} \tag{6}$$

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{\sqrt{Var(G_i^*)}} \quad (7)$$

Expected G_i^* is calculated using equation 6, using this a Z-score can be calculated for the G_i^* to allow for a much better understanding of the data and the significance of the results. This Z-score value is what is used in ArcGIS Pro to create hot and cold spot maps showing the distribution of data. (Mitchel & Griffen 2021). There are a few factors that can alter the results of this analysis. First, features near the edge of the study area can have skewed values because they have less neighbors, increasing the weight that their existing neighbors have on the calculations. Second, when less than 30 features are present, results may be inaccurate because that outliers can have on the data (Mitchel & Griffen 2021).

CHAPTER 3

DATA AND METHODS

3.1 Study Area

The study area is Itasca State Park located in northern Minnesota (Figure 1). Itasca State Park is in the Laurentian Mixed Forest ecological province and is over 131.5 km^2 of old-growth red and white pine forest (Minnesota Department of Natural Resources 2022). Itasca is unique because it is also Minnesota's oldest state park, established in 1891, has over 100 lakes, and is home to the headwaters of the Mississippi River.

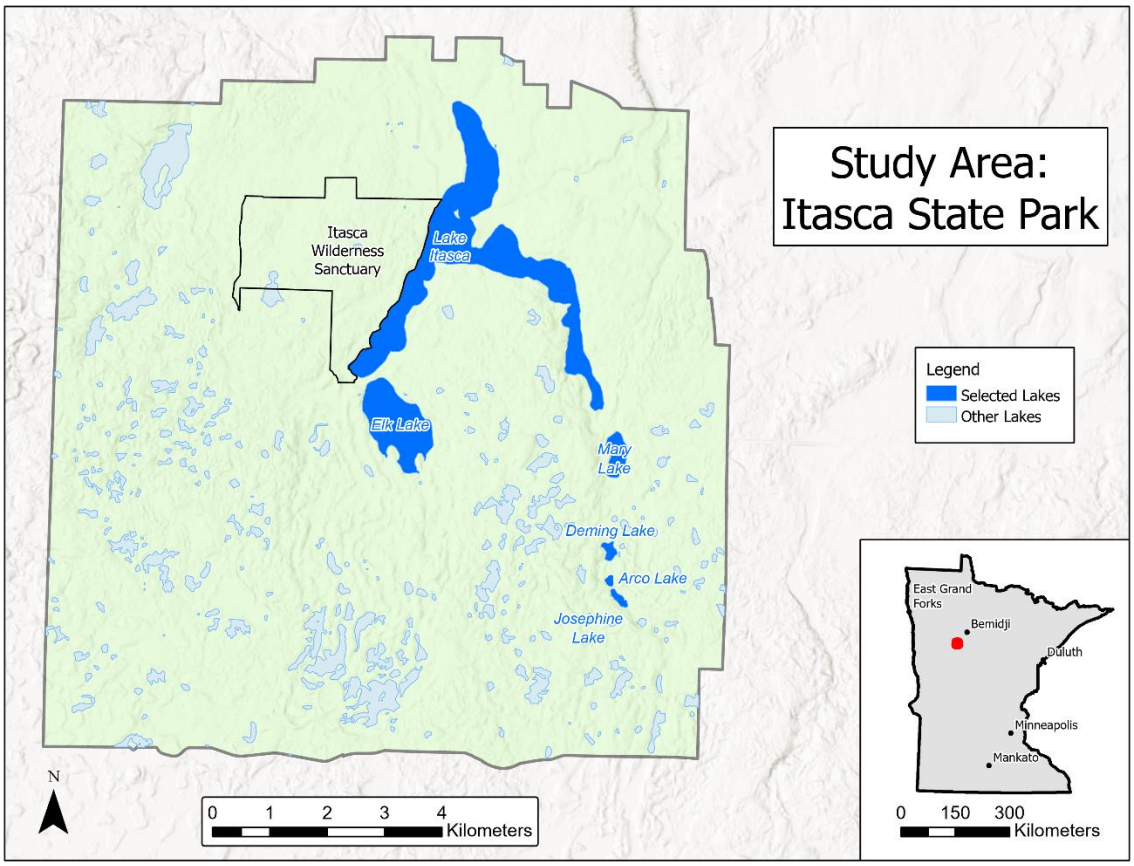


Figure 1 Study Area: Itasca State Park.

Past glacial activity has heavily impacted the landscape today. The park is situated on the Itasca Moraine, specifically on outwash and end moraine areas with mostly sandy glacial drift (Hobbs & Goebel 1982). Total average precipitation for the area ranges from 58-69 centimeters, most of which falling outside of winter months (Minnesota Department of Natural Resources 2023, February 16). What separates the lakes in the park from many other lakes is how little developed land is around them and how minimal human impacts are on these lakes. A study by the Minnesota Department of Natural Resources in 2006 shows that the most common landcover in the Itasca Moraine is forest, followed by wetland, then agriculture (Figure 2). From mid-June through October 2022 the wind direction in the area was predominantly northwest and south-southeast (Figure 3).

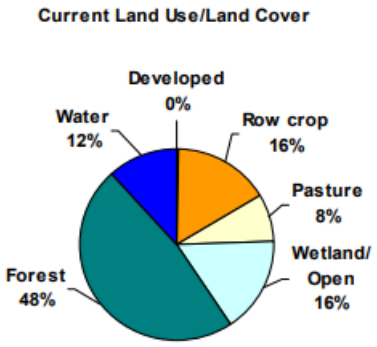


Figure 2 Itasca Moraine Landcover. Land use/land cover of Itasca Moraine region shows 24% use for agriculture (Row crop & Pasture) and 76% natural (Forest, Wetland, Water). *(Minnesota DNR 2006)*

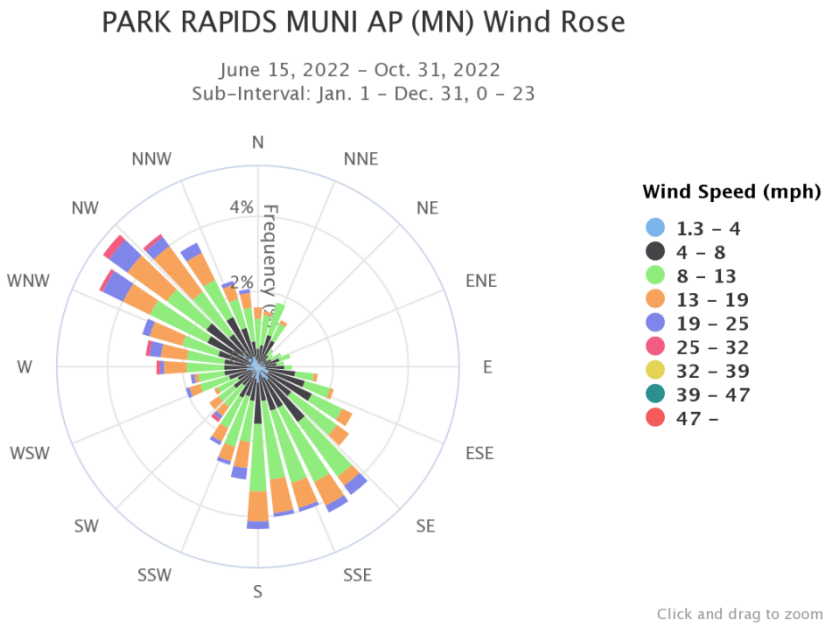


Figure 3 Study Area Wind Rose. Predominant wind direction for study area is northwest and south, southeast. (Midwest Regional Climate Center)

Because of the abundance of lakes in the Northern Lakes and Forests ecoregion not every lake can be studied. However, the Minnesota Department of Natural Resources and Minnesota Pollution Control Agency have the Sentinel Lakes program. The primary goal of this program is to "Identify important biological, physical and chemical trends in Minnesota lakes and monitor these trends over time" (Minnesota Department of Natural Resources 2023, February 6). Elk Lake in Itasca State Park is one of these Sentinel Lakes and has historic data dating back to 1985.

Another source of data for lakes in this region is the University of Minnesota's Itasca Biological Station and Laboratory. The research station has years of previous biological research, including limnology. Previous studies have determined that Arco Lake and Deming Lake are both meromictic, meaning that the layers of water do not intermix with one another (Frane & Walberg 1997, Reiter et al. 1998). Other research from the Minnesota Pollution Control Agency shows that Elk Lake is dimictic (Minnesota Pollution Control Agency 2009).

3.2 Water Sample Collection and Analysis

Chl-a samples were collected for Lakes Itasca, Elk, Mary, Deming, Arco, and Josephine during seven sampling events between July 2022 and October 2022. Samples were collected from a location within the middle of each lake, accessed by canoe. Field measurements of temperature, specific conductance, and pH were measured with a YSI 63. Dissolved oxygen was measured with a YSI Dissolved Oxygen meter, and secchi disk depth were recorded. The pH meter was calibrated the day of each sampling event using standards with measures of 3,7, and 10 pH. Chl-a samples were obtained using a 1-meter vertical composite sample of water at the sample locations and sample locations were recorded with GPS. Once on shore the composite sample of water was pumped through a Whatman Glass Microfiber filter (0.7 µm pore size), with the amount of water pumped recorded and the sample put in a cooler in the field, then a freezer until laboratory analysis was complete. During each sampling trip one duplicate sample was collected to test for variation within the laboratory analysis.

The methodology used for data collection and laboratory analysis was modified from the Environmental Protection Agency's Method 445.0 (Arar & Collins 1997). Filtered samples were placed in a grinding tube with 4 ml of 90% acetone and ground into a slurry. The grinding tube was then rinsed with another ml of 90% acetone, and the contents were placed in a 14 mL centrifuge tube, shaken, and placed in the fridge to sit for 18-20 hours. The next day each tube was shaken again and then placed in the centrifuge at 4,300 RPM for five minutes. The samples rested at room temperature with minimal light exposure for 30 minutes. Following this 5 ml of the sample were micro pipetted into a test tube and insert into the Turner Designs TD-700 laboratory fluorometer and the fluorescence was recorded. Then the sample was acidified to 0.002 N HCL by adding 0.15 ml of 0.1 N HCL. The sample set for 90 seconds and then a second

fluorescence reading was recorded. This process was then repeated for all samples with sample containers and the tissue grinder being cleaned and then rinsed with 90% acetone before use again.

The TD-700 Laboratory Fluorometer was calibrated using a direct concentration calibration method as described in the instrument's manual (Turner Designs 2002). The fluorometer was calibrated on August 18, 2022 and used for five analysis events spanning from calibration to October 27, 2022. The day of calibration a reading of two standards, 228 µg/L and 22.7 µg/L, was recorded and found to be within an acceptable range of the known concentrations. Carlson's Trophic State Index (TSI) (equation 8) was calculated to estimate the trophic state of each of the lakes (Wetzel 2001).

$$TSI(Chla) = 9.81 \ln(Chla) + 30.6 \quad (8)$$

Another measure taken to ensure the data were precise was the inclusion of laboratory reagent blanks (LRB). LRBs were created following each day of laboratory analysis by performing the EPA method 445 with an empty filter paper and the 90% acetone solution .LRBs were analyzed for each date laboratory analysis was completed, for a total of five LRBs. The LRBs showed how much error is being introduced from improper cleaning of laboratory materials or from errors performing the methodology. The acceptable concentration for the LRB is determined by the minimum measured concentration of Chl-a for that analysis batch. EPA Method 445 sets acceptable LRB levels as less than ten percent of the minimum from that batch (Arar & Collins 1997).

3.3 Data

Two datasets were used in the measurement and estimation of Chl-a. The first dataset used was the Harmonized Landsat and Sentinel-2 (HLS) surface reflectance imagery acquired

from NASA, at https://search.earthdata.nasa.gov/search?q=hls (Table 1). The HLS data was selected for this study because it has a moderate spatial resolution of 30 meters and a temporal resolution of two to three days. HLS can achieve this because it takes data from the European Space Agency's Sentinel-2 MSI and the USGS's Landsat 8-OLI. The data from each sensor goes through numerous processing steps until they are seamlessly integrated with one another and can be treated as the same data set (Figure 4). The biggest benefit of using HLS imagery compared to solely Landsat or Sentinel 2 is an increase in available images. HLS data is also convenient because NASA provides a quality assurance layer that identifies pixels that have cloud, cloud shadow, high aerosol and multiple other variables. To supplement the Harmonized Landsat and Sentinel imagery a lake database of water body boundaries provided by the Minnesota Geospatial Commons was used to provide lake boundaries (https://gisdata.mn.gov/).

Table 1 HLS Spectral Bands Nomenclature. Band wavelength ,number, and name of Harmonized Landsat bands and the corresponding Sentinel 2 and Landsat bands. (Masek et al. 2022)

Band name	OLI band number	MSI band number	HLS band code name L30	HLS band code name S30	Wavelength (micrometers)
Coastal Aerosol	1	1	B01	B01	0.43 – 0.45*
Blue	2	2	B02	B02	0.45 – 0.51*
Green	3	3	B03	B03	0.53 – 0.59*
Red	4	4	B04	B04	0.64 – 0.67*
Red-Edge 1	–	5	–	B05	0.69 – 0.71**
Red-Edge 2	–	6	–	B06	0.73 – 0.75**
Red-Edge 3	–	7	–	B07	0.77 – 0.79**
NIR Broad	–	8	–	B08	0.78 –0.88**
NIR Narrow	5	8A	B05	B8A	0.85 – 0.88*
SWIR 1	6	11	B06	B11	1.57 – 1.65*
SWIR 2	7	12	B07	B12	2.11 – 2.29*
Water vapor	–	9	–	B09	0.93 – 0.95**
Cirrus	9	10	B09	B10	1.36 – 1.38*

Thermal Infrared 1	10	–	B10	–	10.60 – 11.19*
Thermal Infrared 2	11	–	B11	–	11.50 – 12.51*

* from OLI specifications, ** from MSI specifications

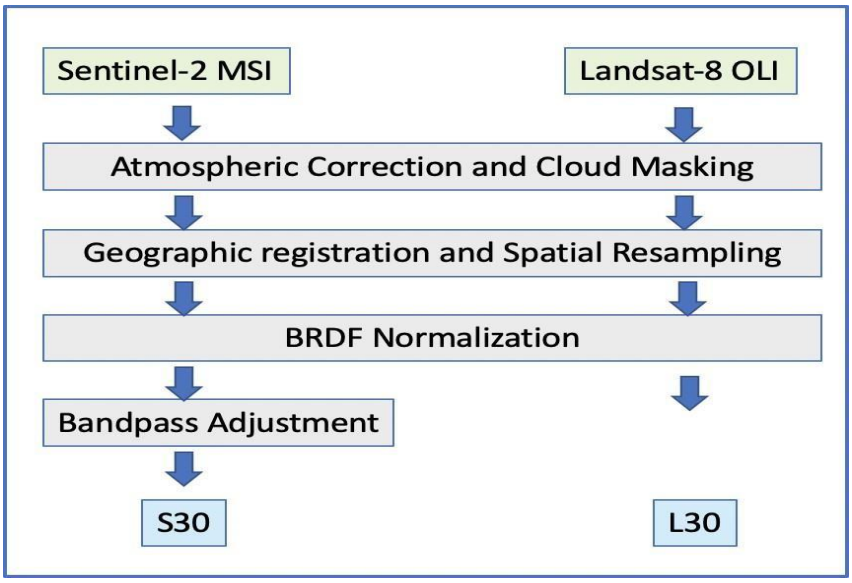


Figure 4 HLS Science Algorithm Processing Flow. (Masek et al. 2022)

3.4 Data Formatting and Processing

The Harmonized imagery required a substantial amount of cleaning up before analysis. Figure 5 shows an outline of how the data were processed. All the processing was performed using a combination of Python 3.9.11 and ArcGIS Pro 3.0.1. For further detail on how data were processed see appendix A. First, a cloud, cloud shadow, and aerosol mask from the quality assurance layer provided by NASA was created. Clouds and cloud shadows were masked to obtain more accurate index values. Due to a known issue with the atmospheric correction process NASA recommends masking out pixels that have high aerosol values (Masek et al. 2022). Next, the images were clipped to a 10-meter inner buffer of the lakes to help mitigate edge effect from shoreline and aquatic vegetation (Zabaleta et al. 2021). The data was masked using the clipped images and the masks created previously. Following this some of the reflectance values on the

lakes were negative, to fix these the negative reflectance values were set to no data and scaled to percent reflectance. Then the negative reflectance values were masked out to obtain clean images that can be used to calculate the indices of interest. The indices used in this study are GRB, Green-Blue Normalized Difference, FLH Blue, Red-Green ratio, Blue-Green ratio, NIR-Red ratio, and NDCI as defined in

Table ***2***. The index calculations were completed using python scripting with the arcpy module.

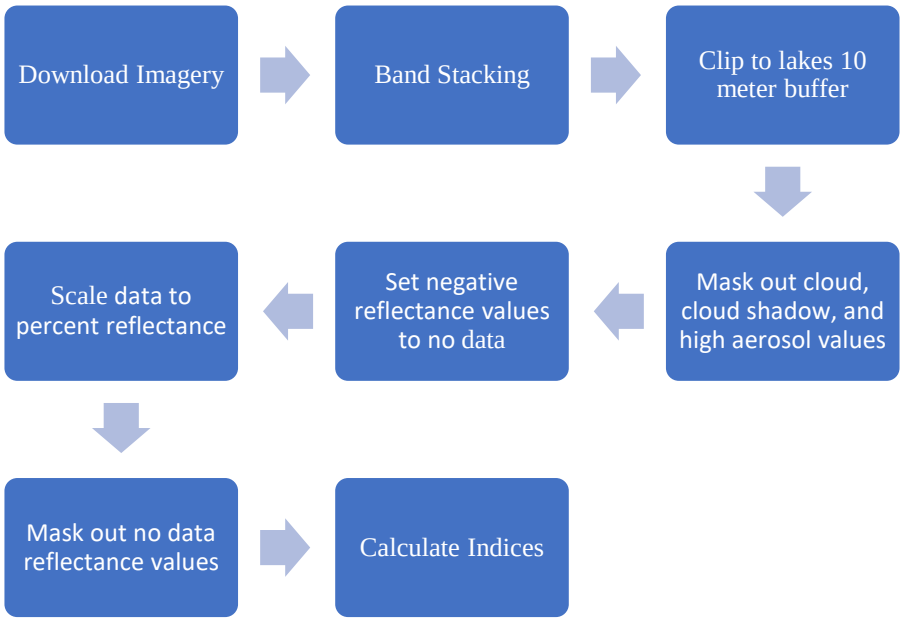


Figure 5 Imagery Processing Workflow.

Table 2 Index Equations.

Index	HLS L30 Equation	HLS S30 Equation
GRB	$\frac{B3 + B4}{B2}$	$\frac{B3 + B4}{B2}$
Gr-Bl Norm	$\frac{B3 - B2}{B3 + B2}$	$\frac{B3 - B2}{B3 + B2}$
FLH Blue	$float(B3) - [float(B4) + (float(B2) - float(B4))]$	$float(B3) - [float(B4) + (float(B2) - float(B4))]$
Red/Green Ratio	$\frac{B4}{B3}$	$\frac{B4}{B3}$
Green/Blue Ratio	$\frac{B3}{B2}$	$\frac{B3}{B2}$
NIR/Red Ratio	$\frac{B5}{B4}$	$\frac{B8A}{B4}$
NDCI	$\frac{B5 - B4}{B5 + B4}$	$\frac{B8A - B4}{B8A + B4}$

3.5 Sample Collection Timing and Hot Spot Analysis

The collection days were planned to be as close to a clear Sentinel 2 overpass as possible. The methods of this study are modified from Zabaleta et al 2021. The imagery was used in conjunction with a lake database of water body boundaries provided by the Minnesota Geospatial Commons to mask imagery to the water bodies then calculate different chlorophyll indices. Due to errors common with smaller water bodies, and interference with aquatic vegetation and lake bottoms, the mask created for the lakes will limit the lakes so that the area between the shoreline and up to 10 meters in will not be included (Zabaleta et al. 2021). Because algae blooms often occur near the shoreline, masking out near the shoreline can cause an under detection of blooms using this method. However, if Chl-a is detected within this area it is near impossible to differentiate whether the Chl-a is due to algae or vegetation presence. Using a linear regression these indices were tested against the in-situ data to determine which index is most accurate in the estimation of Chl-a. The most accurate index was used to perform a hotspot

analysis of the selected lakes with a Getis Ord Gi* hotspot analysis using the Optimized Hotspot Analysis tool in ArcGIS Pro 3.0 to explore the distribution of Chl-a within each lake (Figure 6).

Figure 6 Hotspot Analysis Workflow.

CHAPTER 4

RESULTS

4.1 Field Data and Chl-a Measurements

The field data collected as well as the laboratory analysis results for Chl-a concentrations in the lake, recorded as Lake Chl-a (ug/L) are shown in appendix B. Note that on 8/12/22 volume pumped for Lake Arco and Lake Josephine reads 'NA' this is because samples were not collected on those days because of an error. Also note that on 9/20/22 there are no pH readings, this is because of an equipment error on that day. Secchi Disk values ranged from 1-5m (Figure 7) , dissolved oxygen was between 7-13.5 mg/L (Figure 8), temperature from 7-25˚C (Figure 9), pH from 5.5-9.5 (Figure 10), specific conductance was between 50-350 µs (Figure 11), and Chl-a concentrations were between 0.1-16 µg/L (Figure 12).

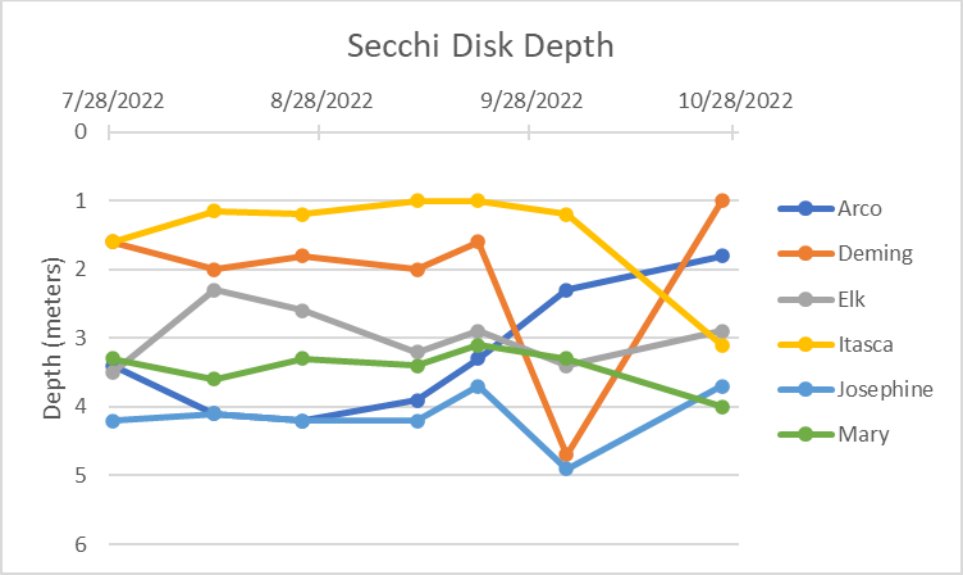


Figure 7 Secchi Disk Depth. Secchi disk depth was consistent from the end of July into the middle of September. With Arco, Elk, and Mary Lake seeing a decrease in depth and Deming, Itasca, and Josephine all increasing in late September.

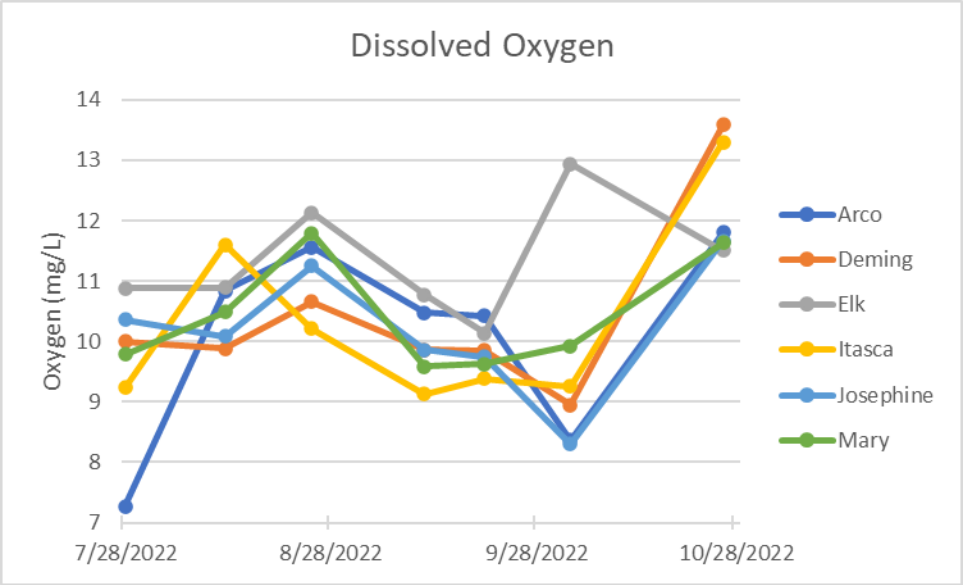


Figure 8 Dissolved Oxygen. Starting in late July Deming, Elk, Itasca, Josephine, and Mary Lake all had similar dissolved oxygen levels. In early August Arco Lake's dissolved oxygen levels reached similar levels to the other five lakes. The dissolved oxygen levels of these lakes followed similar rates of decrease and increase from early August to mid-September.

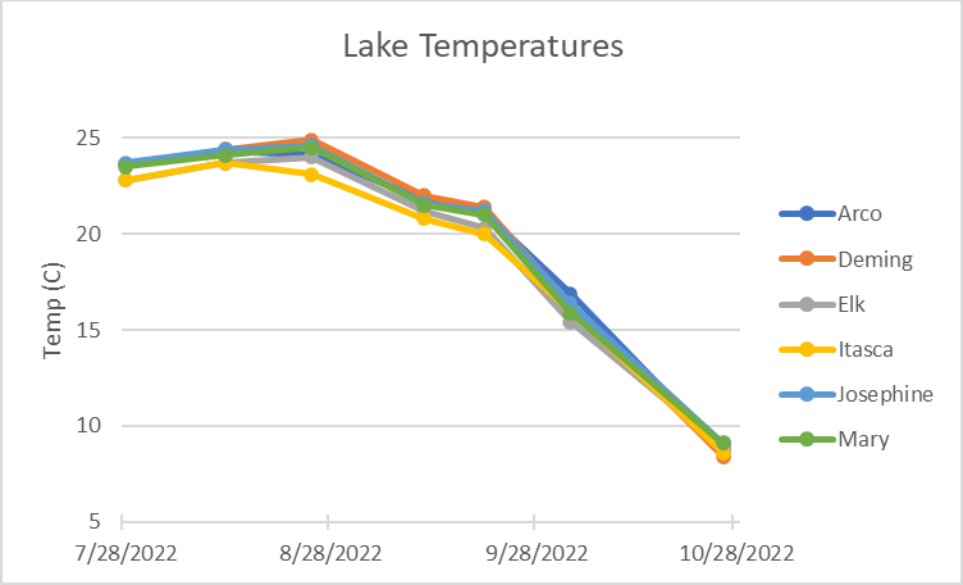


Figure 9 Lake Temperatures. All six lakes had similar water temperatures, starting near 23˚C and gradually declining to 8˚C in late October.

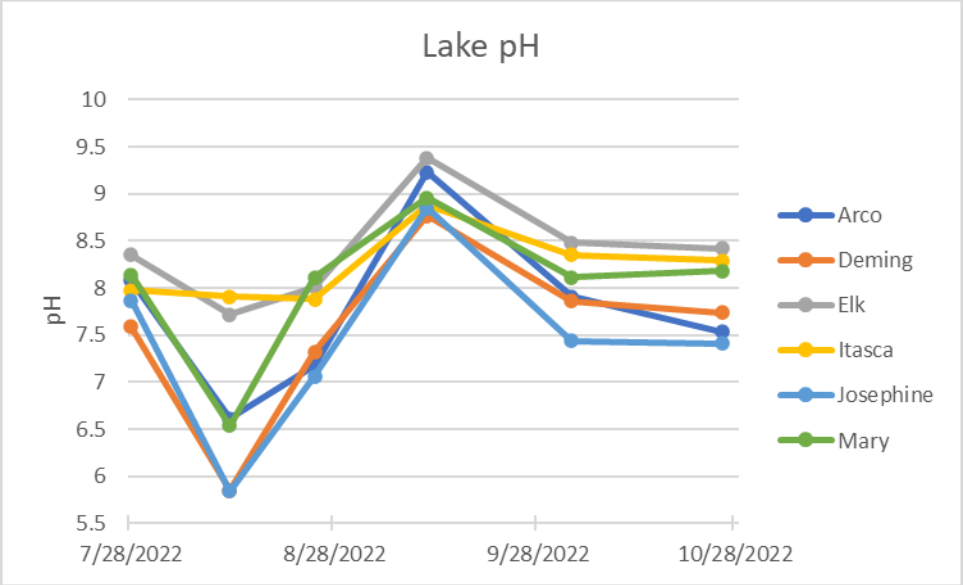


Figure 10 Lake pH. The smaller lakes, Mary, Arco, Deming, and Josephine, all had a range of pH values between 5.8-9.1. With Lake Itasca and Elk Lake ranging between 7.9-9.4.

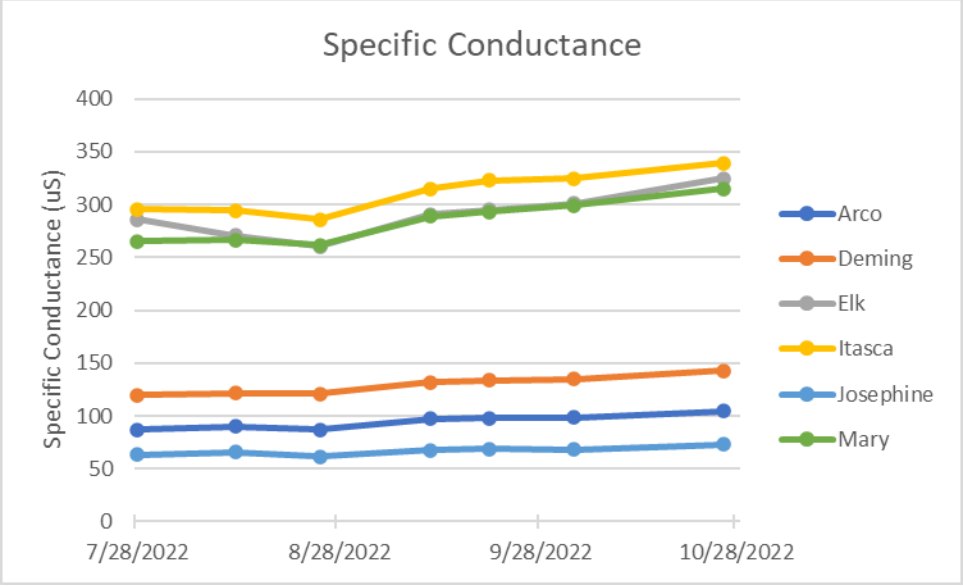


Figure 11 Specific Conductance. The lakes were split into two groups, one with a higher specific conductance, Lake Itasca, Elk Lake, and Mary Lake. And lakes with lower specific conductance, Deming Lake, Josephine Lake, and Arco Lake. All six lakes showed a gradual increase from late July to late October and were consistent across the study.

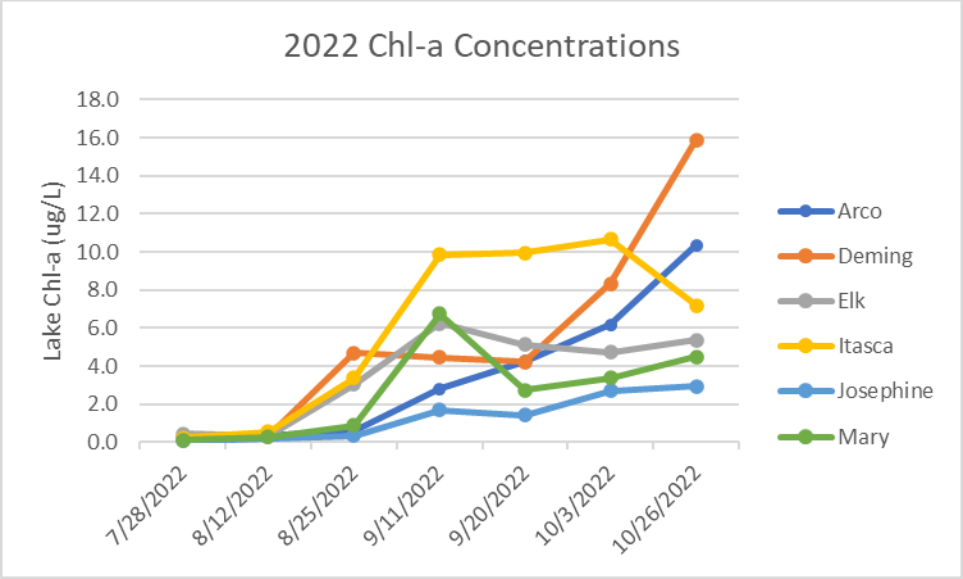


Figure 12 Chl-a Time Series. In late July Chl-a concentrations for each lake were all low. Following late August Chl-a concentrations for each lake started to diverge from one another. Lake Itasca rose to the highest concentration and remained steady around 10 µg/L from mid-September to early October and then had a slight decrease during the last sampling event. Arco and Deming Lake both showed rapid increases in concentration from early October to the last sampling event.

Table 3 shows a summary of the lakes including surface area, measured depth at sample locations, Maximum Carlson's Trophic State Index calculated using Chl-a concentrations and the corresponding trophic state classification.

Table 3 Lake Summaries and Mixing State.

Lake	Area (km^2)	Max Measured Depth (m)	TSI Max	Trophic State
Arco	0.022	10.7	53.5	Eutrophic
Deming	0.056	10.5	57.7	Eutrophic
Elk	1.181	13.4	47.1	Mesotrophic
Itasca	4.283	10.8	53.8	Eutrophic
Josephine	0.045	8.9	41.2	Mesotrophic
Mary	0.219	10.0	49.4	Mesotrophic

4.2 Quality Assurance

Multiple quality assurance methods were put into place to evaluate the accuracy and precision of the Chl-a detection in the laboratory. Field duplicates were collected for five of the

sampling events at the same location and with the same methods as the other samples. The percent relative difference between the samples and the corresponding duplicates was calculated (Table 4). The percent difference for the Chl-a measurements were all acceptable under EPA Method 445 (<15%) while the phy-a percent differences were higher.

Table 4 Duplicate Percent Difference.

Samples	% Difference Chl-a	% Difference Phy-a
ITA0825-1 & ITA0825-1D	2.94	**-89**
MARY0911-1 & MARY0911-1D	4.4	**-23**
ELK920-1 & ELK920-1D	5.1	**-42**
DEM1003-1 & DEM1003-1D	-2.4	2.0
ARC1026-1 & ARC1026-1D	-4.8	-1.6

Table 5 shows the LRB concentrations as well as the ten percent of minimum Chl-a concentrations for each analysis batch. All five LRBs were less than the ten percent mark for their respective minimums.

Table 5 Lab Blanks.

Date	Solution	Measured Chl-a (µg/L)	10% of Minimum Chl-a (µg/L)
8/19/22	Lab Blank	2.0	5.4
8/30/22	Lab Blank	7.6	18
9/30/22	Lab Blank	3.0	20
10/12/22	Lab Blank	2.4	20
10/27/22	Lab Blank	2.5	14

Table 6 shows the percent difference values of the fluorometer for Chl-a measurements using known standards and blanks. For the first two analysis events a high standard reading was recorded before the field samples were analyzed, and before every analysis a blank consisting of 90% acetone was recorded.

Table 6 Fluorometer Calibration. Fluorometer calibration dates, concentrations, and percent differences with percent differences greater than 15% in bold.

Date	Solution	Measured Chl-a (µg/L)	Actual Chl-a (µg/L)	Percent Difference
8/18/22	High Standard	227.2	228	-0.0035
8/18/22	Low Standard	22.4	22.7	-0.01
8/19/22	High Standard	229	228	0.4
8/19/22	Blank	0	0	0
8/30/22	High Standard	247.1	228	8.4
8/30/22	High Standard*	243.5	228	6.8
8/30/22	4.54 µg/L Dilution	4.1	4.54	-9.7
8/30/22	3.78 µg/L Dilution	2.5	3.78	**-33**
8/30/22	Blank	0	0	0
9/30/22	Blank	0	0	0
10/12/22	Blank	0	0	0
10/27/22	Blank	0	0	0

4.3 Index Evaluation

The results of the linear regression between the indices and the field measured Chl-a concentrations are shown in Table 7. I found that the GRB index was the most accurate with the largest R, R^2, adjusted R^2, and lowest root mean square error. Following that the green blue normalized difference was the next best across all measures and the FLH was third best.

Table 7 Ln Transformed Index Correlations. Index correlation results, with significance, size of sample, and RMSE. NDCI was not transformed due to negative values within the natural log.

Index	R	R^2	Adj. R^2	P-value	N	RMSE	RMSE (ug/L)
GRB	0.432	0.187	0.162	0.011	34	1.35	3.87
Gr/Bl Norm	0.366	0.134	0.109	0.026	37	2.71	15.08
FLH	0.340	0.116	0.089	0.045	35	3.47	32.14
Red/Green Ratio	0.324	0.105	0.079	0.051	37	3.55	34.87
Gr/Bl Ratio	0.300	0.088	0.062	0.075	37	3.48	32.35
NIR/Red Ratio	0.235	0.055	0.028	0.16	37	3.65	38.44
NDCI*	0.225	0.051	0.024	0.18	37	3.66	38.76

4.4 Hotspot Analysis

The hotspot analysis had differing results depending on the size of the lake. The smallest lake, Arco, was so small that at the resolution of the imagery (30m) there was not enough data to perform a hotspot analysis. The next two smallest lakes, Deming (Figure 13) and Josephine (Figure 14) successfully executed the hotspot analysis. However, neither has very many hotspots or cold spots.

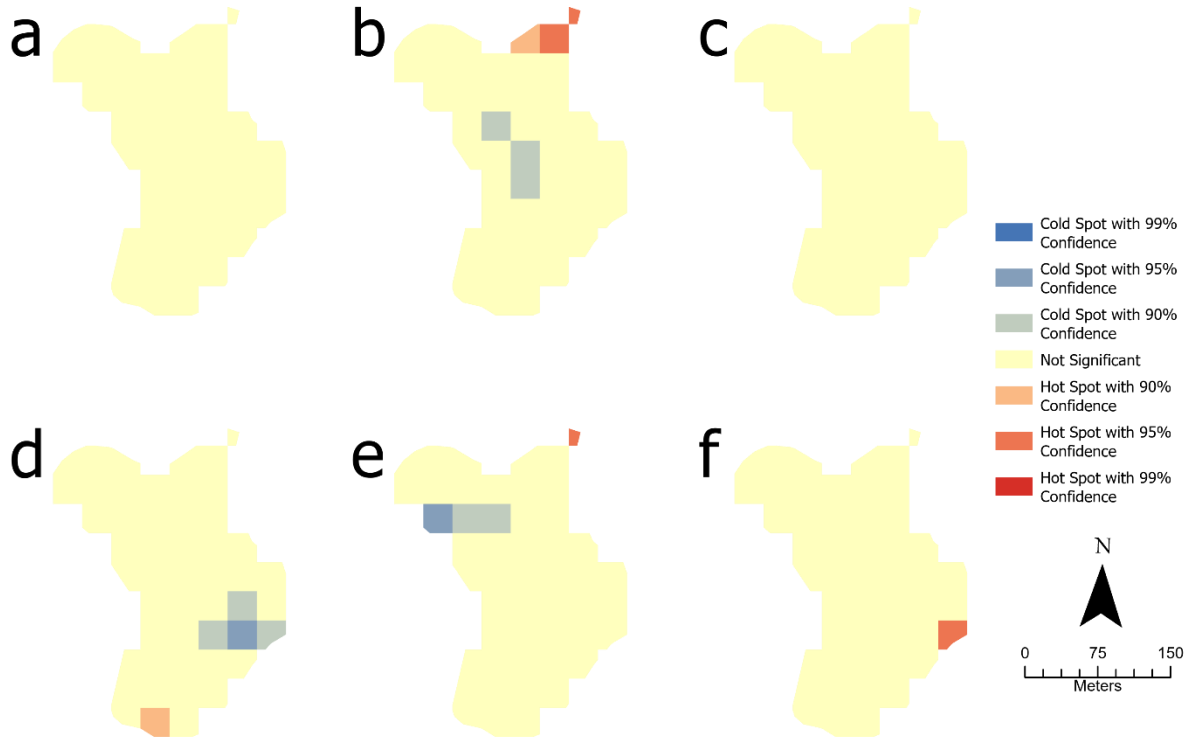


Figure 13 Deming Lake Chl-a Hotspot. Deming Lake did not have very many hot or cold spots. However, the few hot spots that were present were all located along the shoreline, and the cold spots were located more central or extended into the central portion of the lake.
Map **a** 28 June 2022, **b** 17 August 2022, **c** 25 August 2022, **d** 11 September 2022, **e** 3 October 2022, and **f** 19 October 2022

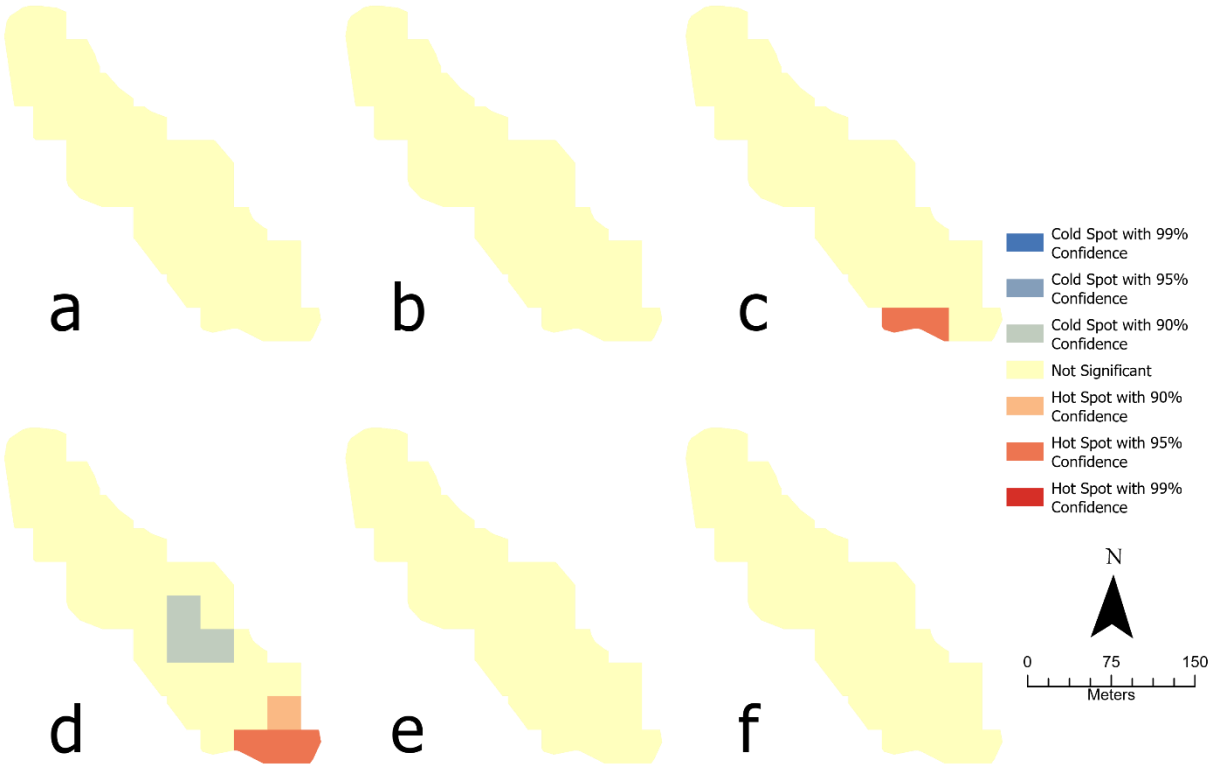


Figure 14 Josephine Lake Chl-a Hotspots. Josephine Lake had few hotspots with only two dates showing any. Both hot spots were located in the southern portion of the lake, and on September 11th there was also a cold spot in the central portion of the lake.
Map **a** 28 June 2022, **b** 12 August 2022, **c** 25 August 2022, **d** 11 September 2022, **e** 3 October 2022 and **f** 26 October 2022

Unlike the smaller lakes the larger three, Mary, Elk, and Itasca, have more hot and cold spots (Figures 15,16,17). All three of these lakes show variability in the Chl-a distribution throughout the summer season. There are, however, certain areas within each lake that appear more than often to be a hotspot. In Mary Lake (Figure 15) the southern portion of the lake shows a hotspot in four of the six images. While in Elk Lake (Figure 16) the northwest portion of the lake shows a similar pattern with four of the six images showing hotspots. In Itasca (Figure 17), a different pattern emerges. Rather than having consistent hotspots near the shoreline Itasca has

its hotspots around Schoolcraft Island and extending North. The west arm of Itasca may be a likely place for cold spots, with three of the six images showing cold spots throughout the arm.

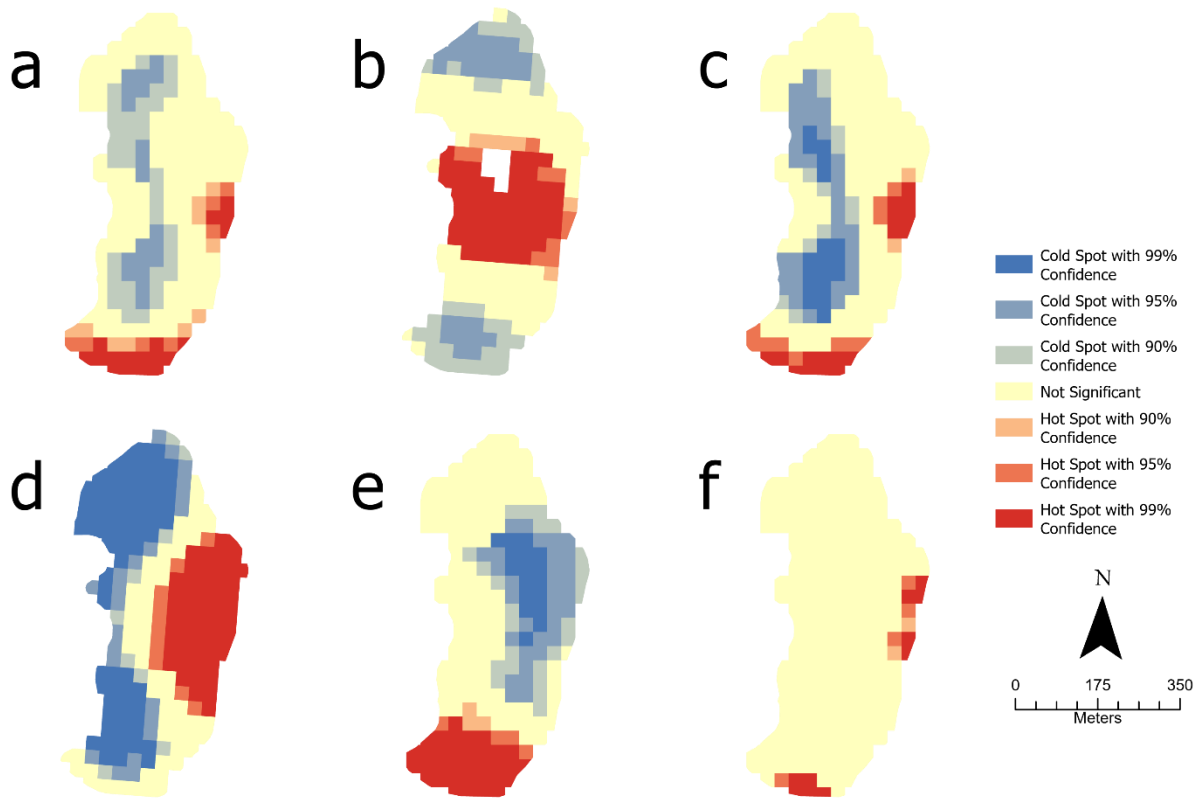


Figure 15 Mary Lake Chl-a Hotspots. Mary Lake had a common hot spot near the south end of the lake in maps a, c, d, and f. Similarly in maps a, c, and d there was a cold spot on the west side of the lake.
Map **a** 28 June 2022, **b** 17 August 2022, **c** 25 August 2022, **d** 11 September 2022, **e** 3 October 2022, and **f** 19 October 2022

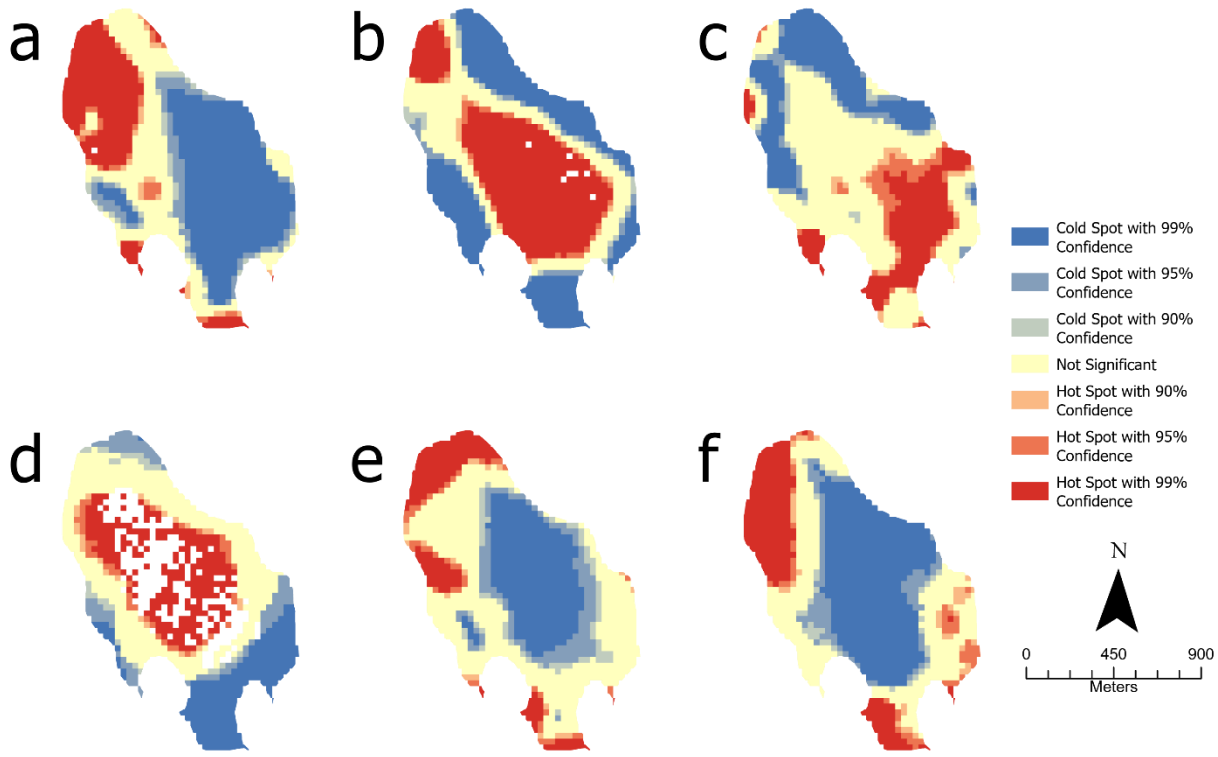


Figure 16 Elk Lake Chl-a Hotspots. Elk Lake consistently had some degree of a hot spot in the northwest portion of the lake, with map a, b, c, e, and f showing hot spots. Another location with less hot spot activity was in the southern part of the lake in map a, e, and f.
Map **a** 28 June 2022, **b** 17 August 2022, **c** 25 August 2022, **d** 11 September 2022, **e** 3 October 2022, and **f** 19 October 2022

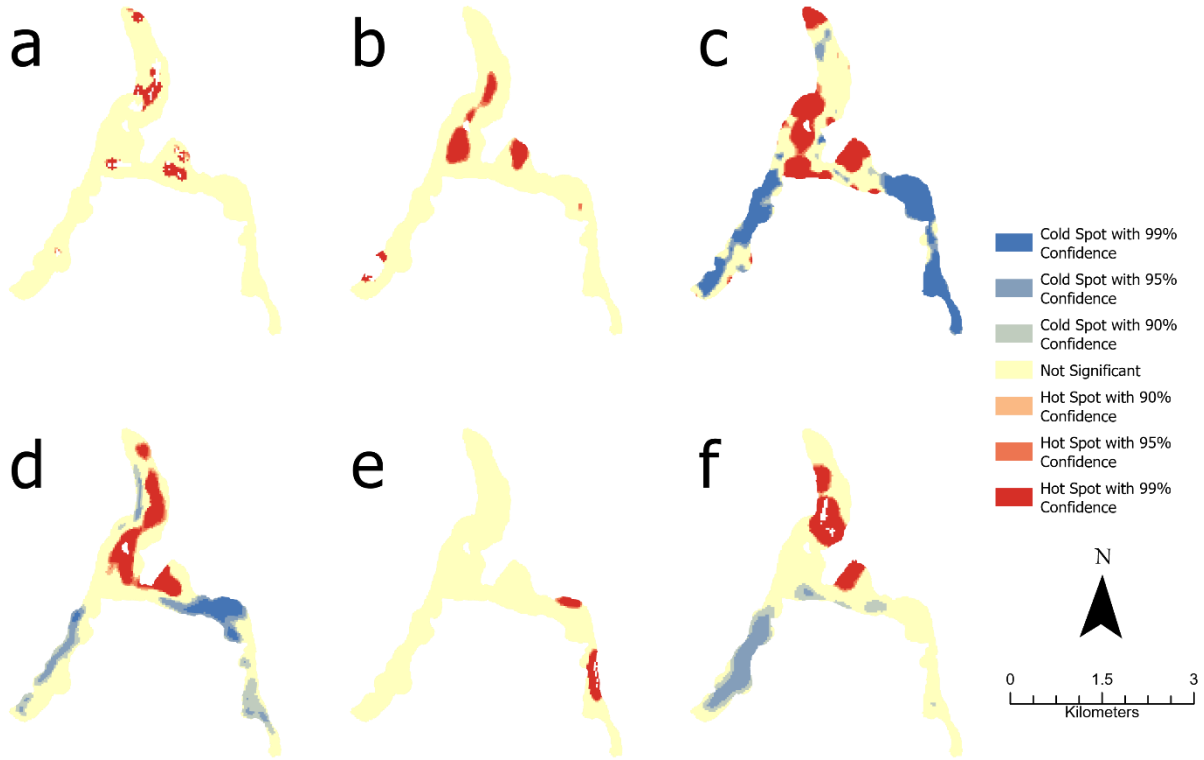


Figure 17 Lake Itasca Chl-a Hotspots. Schoolcraft Island is in the southern portion of the north arm of Lake Itasca and is the white spot in the hotspot of map d. Lake Itasca had sporadic hot spots in the north arm and in the north portion of the east arm. The cold spots were located in the south and east arms of the lake in map c, d, and f.
Map **a** 28 June 2022, **b** 17 August 2022, **c** 25 August 2022, **d** 11 September 2022, **e** 3 October 2022, and **f** 19 October 2022

CHAPTER 5

DISCUSSION

5.1 Water Quality Measures

Secchi Disk depth varied across the different lakes with Itasca having the lowest values and Josephine having the highest (Figure 7). Dissolved oxygen saw fluctuations from June to September and then a sharp rise in October in all lakes except for Elk Lake which saw a decrease in dissolved oxygen in October (Figure 8). All six lakes showed similar patterns for water temperature with a peak in late August and then decreasing throughout the season (Figure 9). The lakes also all had similar pH values with a peak in mid-September (Figure 10). Specific conductance was steady across all the lakes with two distinct groups forming. The first group of Elk, Itasca, and Mary had higher values. While the other group of Deming, Josephine, and Arco had lower values (Figure 11).

Looking at the field measurements, there were few moderate to strong correlations present. Also, many of the moderate to strong correlations were negative correlations. Temperature and Secchi disk depth showed the most correlations with other measurements (Table 8). Secchi disk depth had a moderate strength correlation with lake Chl-a, -0.4046, and a strong correlation with Lake Pheophytin-a, -0.6124. While temperature had a moderate correlation with Dissolved Oxygen, -0.3853, and a strong correlation with lake Chl-a, -0.6232. Specific conductance and pH also showed a moderate correlation, 0.3542.

Table 8 Pearson Correlation Coefficients. Pearson correlation coefficient values showing the strength of correlations between each variable with one another. Significance values are designated by *. ** Significant at 0.01 level. *Significant at 0.05 level.

	Secchi Disk (m)	Temp (C)	DO (mg/L)	pH	Chl-a (ug/L)	Specific Conductance (uS)
Secchi Disk (m)	1					
Temp (C)	-0.019	1				
DO (mg/L)	-0.032	-0.385*	1			
pH	0.024	-0.115	-0.42	1		
Chl-a (ug/L)	-0.406**	-0.625**	0.173	0.285	1	
Specific Conductance (uS)	-0.324*	-0.118	0.195	0.354*	0.196	1

**** Significant at 0.01 Level**
*** Significant at 0.05 Level**

Looking at the measured Chl-a for the six lakes they all appear to be rising as the months progress. From June until early October Itasca was the lake with the highest concentrations of Chl-a (Figure 12). However, from October 6th to October 26th Itasca was the only lake to see a decrease in Chl-a.

5.2 Lake Optical Properties

The best index, shown in Table 8, was GRB. This was not the expected result of this study. One of the reasons that the expected indices did not perform well for the lakes could be because of how limited an effect agriculture has on the watershed, with only 24% of the surrounding area being used for crop growth or animal pastures (Figure 2). Looking at a sample spectral signature from August 30, 2022 (Figure 18), a peak can be seen in the green band, Band 3. And two absorption windows in blue, band 2, and red, band 4, wavelengths. There is a peak in the NIR wavelengths, band 5. However, for most of the lakes this peak is much smaller than the peak in the green wavelengths. This differs from typical responses on healthy land vegetation, where the NIR reflectance is typically much greater than the green reflectance. This difference

could be due to the optical properties of these six specific lakes or have to do with the atmospheric corrections of the imagery.

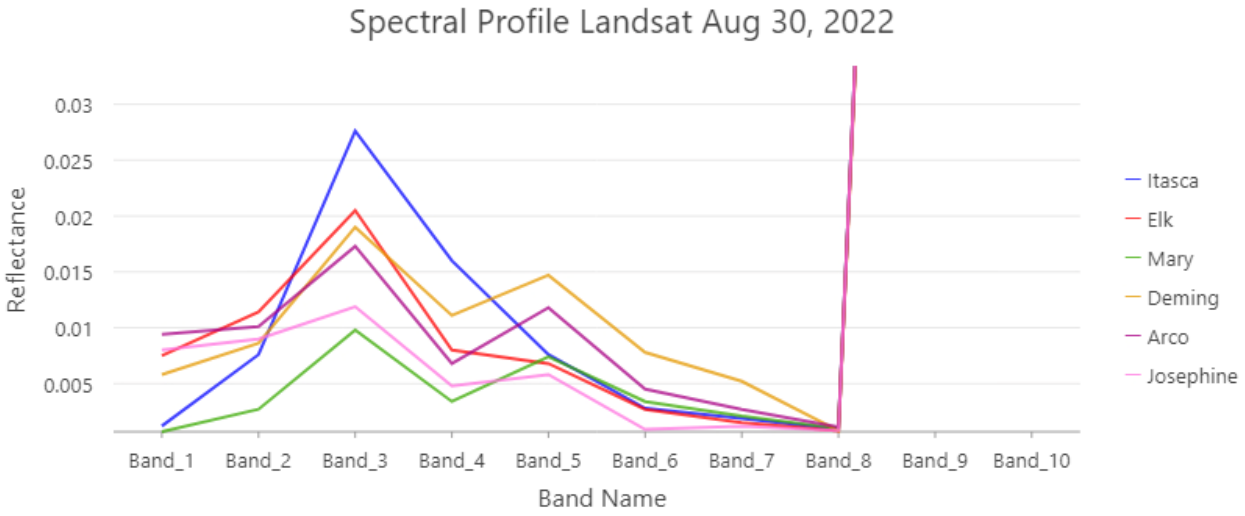


Figure 18 Spectral Signature 8/30/22. The spectral signature for a pixel in the center of each lake with Lake Itasca in blue, Elk Lake in red, Mary Lake in green, Deming Lake in orange, Arco Lake in purple, and Josephine Lake in pink. All the lakes show a peak in the green band (band 3) and all the lakes except Itasca peak again in the near infrared band (band 5).

5.3 HLS Imagery Performance

The HLS imagery has worked well for this study. While other studies have used both Landsat and Sentinel 2 imagery together, few have used the available and geometrically and atmospherically corrected HLS data. The time and energy spent on atmospherically correcting and resampling the data to match one another for analysis can now be used on the analysis itself. However, it may be advantageous to perform atmospheric corrections on a use-by-use case depending on the study area. The HLS data is processed using the Land Surface Reflectance Code (LaSRC) (Vernote el al. 2018). But when comparing different atmospheric corrections for Landsat and Sentinel 2 imagery Doxani (2018) found that the ACOLITE method performed better over water. Future studies could compare the HLS data with Landsat and Sentinel 2 imagery that has been processed with the ACOLITE method.

The spatial resolution of HLS imagery although improved from sensors such as MODIS still has its limits, particularly when it comes to hotspot analysis. The optimized hotspot analysis tool requires at minimum 30 features. With the imagery at a resolution of 30 meters, this leaves HLS data insufficient for hotspot analysis on lakes that are much smaller than 27,000 m^2. Because of this, I was unable to perform a hotspot analysis on Arco Lake which has a surface area of 22,000 m^2. The tradeoff of the limitations to spatial resolution is the higher temporal resolution available. This study spanned 123 days and during that time 43 images were collected. A minimum of 34 images were usable with the GRB index, providing on average a usable image about every three and a half days.

The HLS imagery overall performed well in the hotspot analysis. Except for Arco Lake mentioned above all the selected lakes were able to use the HLS data to perform a hotspot analysis. After applying a buffer along the edge of the lakes there did not appear to be a noticeable edge effect from aquatic vegetation. The hotspots produced from the imagery were able to provide a season-long summary from June to October and showed a great deal of variability between lakes and as the summer progressed. A few trends appeared within the hotspot images. First in Figure 16, Elk Lake showed a hotspot in five of the six images in the northwest corner of the lake. This can likely be attributed to the wind which commonly blows to the northwest and in that area of the lake there is a creek outlet that flows north into Lake Itasca. These two factors seem to funnel the Chl-a concentrations into that corner of the lake as the water moves in that direction. Another pattern on Lake Itasca (Figure 17) is the increase in hotspots near Schoolcraft Island in the central part of the lake. This is likely due to aquatic vegetation in the area showing more clearly on the imagery as the lake gets shallower leading up to the shoreline of the island. Also, on Lake Itasca it appears that the two southern arms of the

lake are more prone to cold spots. Again, this is likely due to the commonly northwestern wind direction and due to the overall flow of the lake which flows north becoming the headwaters of the Mississippi River.

Other studies that have used hotspot analysis to study the distribution of Chl-a in lakes have had similar findings. Zabaleta et al. (2021) found that the central regions of the lakes were more prone to cold spots while the areas near the lakeshore were more susceptible to hot spots. When regarding hotspot location with respect to inlets and outlets, Copado et al. (2020) found that in the Gulf of Mexico HAB hotspots were concentrated at the mouths of rivers entering the gulf. They (Copado) concluded that these hotspots were a result of the nutrient rich rivers emptying into the golf fueling the blooms. In the lakes at Itasca rather than seeing an increase in Chl-a at inlets, such as in the southern portion of the west arm of Lake Itasca, instead there was higher concentrations at the outlets of the lakes, for example at the Mississippi River headwater in the north arm of Lake Itasca and in the northwest corner of Elk Lake.

CHAPTER 6

CONCLUSION

The first research question that was addressed in this study focused on evaluating the spatial and temporal resolutions of the HLS data and how they affected hotspot analysis on the lakes. Due to limiting factors with the calculations of the hotspot analysis, not all the lakes were able to have hotspot maps. The hotspot calculations also showed less hot and cold spots within Deming and Josephine (the smaller lakes) than Itasca, Mary, and Elk (the larger lakes). This does not, however, imply that the HLS data performs better for the larger lakes. It could be due to several reasons including lake depth, water circulation, or trophic state. More detailed research would need to be completed to differentiate among these factors. Looking at the temporal resolution of the HLS data it provided several benefits. The first benefit is crucial when validating the imagery with field measurements and that is the about three-day revisit period of the imagery. This allowed for trips to the site to be planned on days that were likely to have minimal cloud cover allowing maximum data to form a relationship between measure Chl-a and the indices derived from the imagery. This also provides the most possible data for analysis of Chl-a distribution on the lakes and how it changes over time.

The second goal of this study was to determine the most suitable index for detecting Chl-a in the lakes of Itasca State Park. The index results were unexpected. NDCI was anticipated to be one of the top indices but was instead the worst index evaluated. In fact, indices that focused more on the green and blue bands generally performed better within the study area. While the GRB index was found to be the best it was still not a great option with only an adjusted R^2 of 0.162 it was far from the values near 0.8 of other studies. This difference in index performance could be because of the use of the HLS data. The HLS imagery is atmospherically corrected and

calibrated for land use, not water use. This is a potential problem when using this dataset. Future studies could correct the HLS imagery for water remote sensing and compare index results with this study. This could determine if the low index fit found in this study was due to the way the imagery was atmospherically corrected or if the selected indices were not adequate and the best index has yet to be tested.

By modifying for correction of water remote sensing, this methodology could be adopted to drinking water reservoirs, and lakes that have been negatively affected by algae blooms. With the HLS medium spatial resolution of 30 meters and temporal resolution of three to four days policy makers can have weekly data on what is happening in their water, and where algae are gathering. This methodology could also be used solely on Landsat imagery, to gather a historic record of data to determine if there are any long-term patterns.

References

Aubriot, L., B. Zabaleta, F. Bordet, D. Sienra, J. Risso, M. Achkar & A. Somma (2020) Assessing the origin of a massive cyanobacterial bloom in the Rio de la Plata (2019): Towards an early warning system. Water Research, 181.

Arar, E. J. & G. B. Collins. Method 445.0 In Vitro Determination of Chlorophyll a and Pheophytin ain Marine and Freshwater Algae by Fluorescence. U.S. Environmental Protection Agency, Washington, DC, 1997.

Buma, W. G. & S. I. Lee (2020) Evaluation of Sentinel-2 and Landsat 8 Images for Estimating Chlorophyll-a Concentrations in Lake Chad, Africa. Remote Sensing, 12.

Cao, M., S. Qing, E. Jin, Y. Hao & W. Zhao (2021) A spectral index for the detection of algal blooms using Sentinel-2 Multispectral Instrument (MSI) imagery: a case study of Hulun Lake, China. International Journal of Remote Sensing, 42, 4514-4535.

Copado-Rivera, A. G., J. Bello-Pineda, J. A. Ake-Castillo & P. Arceo (2020) Spatial modeling to detect potential incidence zones of harmful algae blooms in Veracruz, Mexico. Estuarine Coastal and Shelf Science, 243, 8.

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., et al. (2018). Atmospheric Correction Inter-Comparison Exercise. Remote Sensing, 10(3), 352. MDPI AG. Retrieved from http://dx.doi.org/10.3390/rs10020352.

ESRI (2022). ArcGIS Pro: Release 3.0. Redlands, CA: Environmental Systems Research Institute.

Frane, M. A., & Walberg, T. J. (1997). Arco and Demming: Two Worlds Divided Demming: When the Levee Breaks. IBS # 2260.

Getis, A. Ord, J.K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3).

Hobbs, H.C., Goebel, J.E.. (1982). S-01 Geologic map of Minnesota, Quaternary geology. Minnesota Geological Survey. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/60085.

Jensen, J. R., (2014). *Remote Sensing of the Environment an Earth Resource Perspective*. Pearson.

Jetoo, S., V. I. Grover & G. Krantzberg (2015) The Toledo Drinking Water Advisory: Suggested Application of the Water Safety Planning Approach. Sustainability, 7, 9787-9808.

Kraker, Dan. (2022, October 17). Boundary Waters algae blooms spark questions, concern. *MPRnews*. https://www.mprnews.org/story/2022/10/17/boundary-waters-algae-blooms-spark-questions-concern

Lekki, J., E. Deutsch, M. Sayers, K. Bosse, R. Anderson, R. Tokars & R. Sawtell (2019) Determining remote sensing spatial resolution requirements for the monitoring of harmful algal blooms in the Great Lakes. Journal of Great Lakes Research, 45, 434-443.

Masek, J.G, J. Ju, M. Claverie, S. Skakun, J.-C. Roger, E. Vermote, B. Franch, Z. Yin, J.L. Dungan. (2022). Harmonized Landsat Sentinel-2 (HLS) Product User Guide. Product Version 2.0. National Aeronautics and Space Administration.

Minnesota Department of Natural Resources (2022). Itasca State Park Summer Map. https://files.dnr.state.mn.us/maps/state_parks/spk00181_summer.pdf.

Minnesota Department of Natural Resources (2006). Tomorrow's Habitat for the Wild and Rare: An Action Plan for Minnesota Wildlife, Comprehensive Wildlife Conservation Strategy. Division of Ecological Services, Minnesota Department of Natural Resources

Minnesota Department of Natural Resources (2023, February 7). Monitoring Minnesota's Changing Lakes. https://www.dnr.state.mn.us/fisheries/slice/index.html

Minnesota Department of Natural Resources. (2023, February 16). Pine Moraines & Outwash Plains Subsection. https://www.dnr.state.mn.us/ecs/212Nc/index.html

Minnesota Pollution Control Agency. (2009). Elk Lake, Clearwater County Sentinel Lake Report. https://files.dnr.state.mn.us/fish_wildlife/fisheries/slice/fact-sheets/wq-slice15-0010.pdf

Mishra, S. & D. 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, 394-406.

Mitchell, A., & Griffin, L. S. (2021). *The Esri Guide to GIS analysis* (Second, Vol. 2). Esri Press.

Ord, J. K., & Getis, A. (1995). "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application." *Geographical Analysis*, 27(4), 287-306.

Page, B., Olmanson, L., Mishra, D.R. (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, Volume 231, 111284, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2019.111284

Reiter, C., Schmitz, J., & Helfman, S. (1998). A Limnological Study of Deming Lake. IBS # 2318.

Turner Designs. (2002). TD-700 Laboratory Fluorometer Operating Manual.

Tzortziou, M., A. Subramaniam, J. R. Herman, C. L. Gallegos, P. J. Neale & L. W. Harding (2007) Remote sensing reflectance and inherent optical properties in the mid Chesapeake Bay. Estuarine Coastal and Shelf Science, 72, 16-32.

Vernote, E. Roger, J.C., Franch, B. Skakun, S. (2018). LaSRC (Land Surface Reflectance Code): Overview, application and validation. International Geoscience and Remote Sensing Symposium.

Vincent, R. K., X. M. Qin, R. M. L. McKay, J. Miner, K. Czajkowski, J. Savino & T. Bridgeman (2004) Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. Remote Sensing of Environment, 89, 381-392.

Watanabe, Fernanda et al. Remote sensing of the chlorophyll-a based on OLI/Landsat-8 and MSI/Sentinel-2A (Barra Bonita reservoir, Brazil). Anais da Academia Brasileira de Ciências [online]. 2018, v. 90, n. 2 suppl 1 [Accessed 14 December 2021] , pp. 1987-2000. Available from: <https://doi.org/10.1590/0001-3765201720170125>. Epub 31 Aug 2017. ISSN 1678-2690. https://doi.org/10.1590/0001-3765201720170125.

Wehr, J. D., Sheath, R. G., & Kociolek, J. P. (2015). Freshwater Algae of North America: Ecology and classification / edited by John Wehr, Robert Sheath, J. Patrick Kociolek. Academic Press.

Wetzel, R. G. (2001). *Limnology: Lake and Rive Ecosystem*. Academic Press.

Zabaleta, B., Achkar, M., & Aubriot, L. (2021). Hotspot analysis of spatial distribution of algae blooms in small and medium water bodies. *Environmental Monitoring and Assessment, 193*(4). https://doi.org/ 10.1007/s10661-021-08944-z

APPENDIX A
Image Proccesing Workflow

First the HLS data were downloaded from NASA. The imagery was downloaded as individual bands and needed to be stacked. The bands were stacked using the python language in A.2. The order of the bands after stacking is shown in A.1. The next step was to clip the imagery to the 10 meter inner buffer of the lakes, this was done using A.3. Following that the imagery was masked using an ArcGIS Pro Model Builder A.4. Next, the negative values in the imagery were set to nodata using A.5. The next step was to convert to percent reflectance (A.6) then mask out the null values (A.7). Finally the Indices were calculated using Python scripting.

A.1 Band Stacking Band List

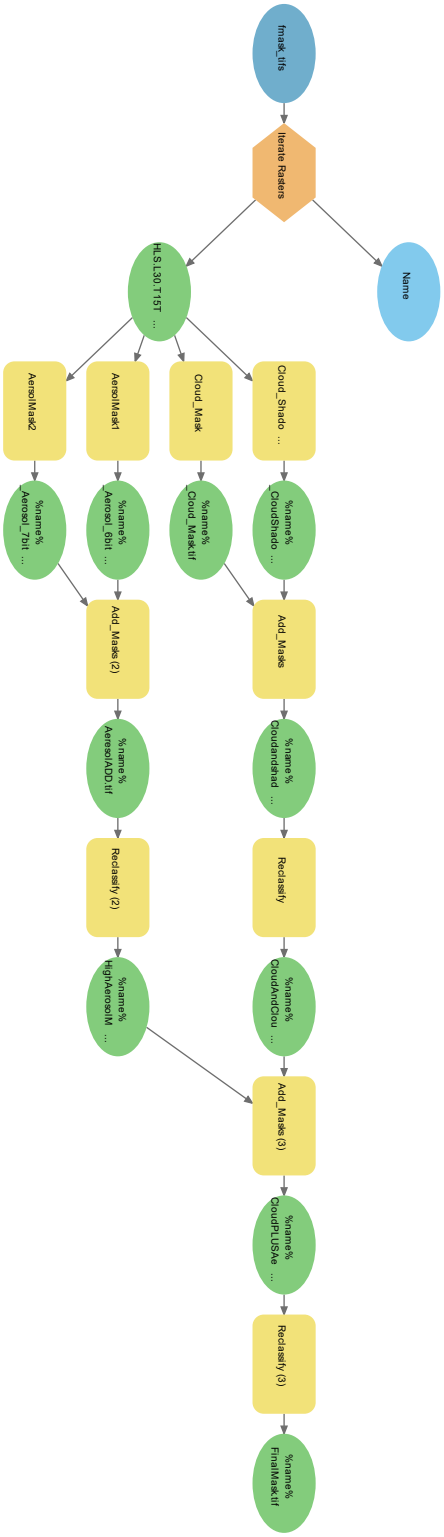
Sentinel 2 Band	Landsat Band	New Band Number
B1	B1	1
B2	B2	2
B3	B3	3
B4	B4	4
B5	B5	5
B6	B6	6
B7	B7	7
B8	B9	8
B9	B10	9
B10	B11	10
B11		11
B12		12
B8A		13

A.2 Band Stacking Script

```
import arcpy
import glob
arcpy.env.workspace = r"C:\cygwin64\home\jmoll"
arcpy.env.overwriteOutput = True

#Stacks all bands of Landsat tifs
rasters=arcpy.ListRasters("HLS_L30*","TIF")
print(rasters)

x=0
while x < len(rasters):

arcpy.management.CompositeBands("{};{};{};{};{};{};{};{};{};{};{}".format(rasters[x],rasters[x+1],raster
s[x+2],rasters[x+3],rasters[x+4],rasters[x+5],rasters[x+6],rasters[x+7],rasters[x+8],rasters[x+9]),
                        r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLSDownload\{}.tif".format(rasters[x][0:29]))
                                                    x+=10
##############################################

#Sentinel 2 Section
sentrasters=arcpy.ListRasters("HLS_S30*","TIF")
print(sentrasters)

x=0
while x < len(sentrasters):

arcpy.management.CompositeBands("{};{};{};{};{};{};{};{};{};{};{};{};{};{}".format(sentrasters[x],sentrast
ers[x+1],sentrasters[x+2],sentrasters[x+3],sentrasters[x+4],sentrasters[x+5],sentrasters[x+6],sentr
asters[x+7],sentrasters[x+8],sentrasters[x+9],sentrasters[x+10],sentrasters[x+11],sentrasters[x+12
]),
                        r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLSDownload\{}.tif".format(sentrasters[x][0:29]))
   x+=13
```

A.3 Imagery Clip and Lake Buffer Script

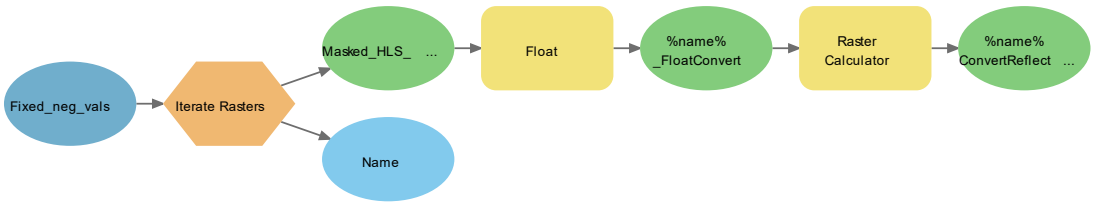
```
import arcpy

arcpy.env.workspace = r"C:\Users\jmoll\OneDrive - North Dakota University System\HLSDownload"
arcpy.env.overwriteOutput = True

rasters=arcpy.ListRasters("HLS_L30*","TIF")

x=0
while x<len(rasters):
   arcpy.management.Clip(r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLSDownload\{}".format(rasters[x]),
                "331047.984292942 5225393.44427143 336129.390492698 5234233.11731608",
                r"C:\Users\jmoll\OneDrive - North Dakota University
System\Clip_Images\{}.tif".format(rasters[x]),
                r"C:\Users\jmoll\OneDrive - North Dakota University
System\Lakes_10M_Buffer_Project.shp", "-9999", "ClippingGeometry", "MAINTAIN_EXTENT")
   x+=1


sentrasters=arcpy.ListRasters("HLS_S30*","TIF")

x=0
while x<len(sentrasters):
   arcpy.management.Clip(r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLSDownload\{}".format(sentrasters[x]),
                "331047.984292942 5225393.44427143 336129.390492698 5234233.11731608",
                r"C:\Users\jmoll\OneDrive - North Dakota University
System\Clip_Images\{}".format(sentrasters[x]),
                r"C:\Users\jmoll\OneDrive - North Dakota University
```



A.4 Mask Model

A.5 Set Negative Values to No Data Script

```
import arcpy
arcpy.env.workspace = r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLS_Masked_Data"
arcpy.env.overwriteOutput = True

rasters=arcpy.ListRasters("*","TIF")

x=0
while x < len(rasters):
   out_raster = arcpy.ia.SetNull(r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLS_Masked_Data\{}".format(rasters[x]),
                        r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLS_Masked_Data\{}".format(rasters[x]),
                        "VALUE < 0"); out_raster.save(r"C:\Users\jmoll\OneDrive - North Dakota
University System\Fixed_neg_vals\{}".format(rasters[x]))
```



A.6 Convert to Reflectance Model

A.7 Mask Out Null Values Script

```
import arcpy
arcpy.env.workspace = r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLS_Reflectance"
arcpy.env.overwriteOutput = True

rasters=arcpy.ListRasters("*","TIF")


x=0
while x < len(rasters):
   out_raster = arcpy.ia.Mask("{}".format(rasters[x]),
                        no_data_values = "NODATA",
                        no_data_interpretation = 1)
   out_raster.save(r"C:\Users\jmoll\OneDrive - North Dakota University
System\HLS_Clean_Data\HLS_Clean_{}.tif".format(rasters[x][7:21]+"_"+rasters[x][26:29]))

   x+=1
```

APPENDIX B

B.1 Sample Locations and Times

SiteID	Lake	Date_Collected	Time	Long	Lat
ARC1026-1	Arco	10/26/2022	3:37	-95.168	47.16582
DEM1026-1	Demming	10/26/2022	3:01	-95.168	47.17037
ELK1026-1	Elk	10/26/2022	12:02	-95.2211	47.19294
ITA1026-1	Itasca	10/26/2022	5:06	-95.2014	47.23115
JOS1026-1	Josephine	10/26/2022	4:10	-95.1669	47.16353
MARY1026-1	Mary	10/26/2022	2:00	-95.167	47.18663
ARC1003-1	Arco	10/3/2022	3:55	-95.168	47.16582
DEM1003-1	Demming	10/3/2022	3:25	-95.168	47.17037
ELK1003-1	Elk	10/3/2022	1:50	-95.2211	47.19294
JOS1003-1	Josephine	10/3/2022	4:20	-95.1669	47.16353
MARY1003-1	Mary	10/3/2022	2:34	-95.167	47.18663
ARC0920-1	Arco	9/20/2022	4:01	-95.168	47.16582
DEM0920-1	Demming	9/20/2022	3:23	-95.168	47.17037
ELK0920-1	Elk	9/20/2022	1:34	-95.2211	47.19294
ITA0920-1	Itasca	9/20/2022	5:27	-95.2014	47.23115
JOS0920-1	Josephine	9/20/2022	4:32	-95.1669	47.16353
MARY0920-1	Mary	9/20/2022	2:37	-95.167	47.18663
ARC0911-1	Arco	9/11/2022	5:50	-95.168	47.16582
DEM0911-1	Demming	9/11/2022	5:03	-95.168	47.17037
ELK0911-1	Elk	9/11/2022	3:02	-95.2211	47.19294
ITA0911-1	Itasca	9/11/2022	7:33	-95.2014	47.23115
JOS0911-1	Josephine	9/11/2022	6:32	-95.1669	47.16353
MARY0911-1	Mary	9/11/2022	4:00	-95.167	47.18663
ARC0825-1	Arco	8/25/2022	4:52	-95.168	47.16582
DEM0825-1	Demming	8/25/2022	4:08	-95.168	47.17037
ELK0825-1	Elk	8/25/2022	1:55	-95.2211	47.19294
ITA0825-1	Itasca	8/25/2022	6:46	-95.2014	47.23115
JOS0825-1	Josephine	8/25/2022	5:25	-95.1669	47.16353
MARY0825-1	Mary	8/25/2022	3:00	-95.167	47.18663
ARC0812-1	Arco	8/12/2022	6:00	-95.168	47.16582
DEM0812-1	Demming	8/12/2022	5:24	-95.168	47.17037
ELK0812-1	Elk	8/12/2022	3:40	-95.2211	47.19294
ITA0812-1	Itasca	8/12/2022	7:04	-95.2014	47.23115
JOS0812-1	Josephine	8/12/2022	6:15	-95.1669	47.16353
MARY0812-1	Mary	8/12/2022	4:35	-95.167	47.18663
ARC0728-1	Arco	7/28/2022	5:18	-95.1679	47.1653
DEM0728-1	Demming	7/28/2022	6:33	-95.1682	47.17053
ELK0728-1	Elk	7/28/2022	3:00	-95.2182	47.19096

ITA0728-1	Itasca	7/28/2022	7:30	-95.2014	47.23116
JOS0728-1	Josephine	7/28/2022	5:53	-95.1662	47.16337
MARY0728-1	Mary	7/28/2022	4:24	-95.1666	47.18676

B.2 Field Measurements and Laboratory Results

SiteID	Lake	Date Collect	Time	Secchi Disk(m)	Temp (C)	DO (mg/L)	pH	Spec. Conduct	Vol pumped (ml)	Lab Chla (ug/l)	Lake Chla (ug/l)	Lab Pheophytin a (ug/l)	Lake Pheophytin a (ug/l)	Reading before acid	reading after acid	r	Date analyzed
MARY0728-1	Mary	7/28/2022	4:24	3.3	23.5	9.79	8.14	265	500	5.4	0.1	97.3	1.9	56.2	53.6		8/19/2022
DEM0728-1	Demming	7/28/2022	6:33	1.6	23.6	10	7.59	120	500	9.8	0.2	147.9	3.0	87	82.3		8/19/2022
ARC0728-1	Arco	7/28/2022	5:18	3.4	23.6	7.27	8.08	87.1	500	6.3	0.1	91.8	1.8	54.2	51.2		8/19/2022
JOS0728-1	Josephine	7/28/2022	5:53	4.2	23.7	10.36	7.87	63.3	1000	8.6	0.1	96.8	1.0	59.1	55		8/19/2022
ELK0728-1	Elk	7/28/2022	3:00	3.5	22.8	10.88	8.35	286	600	27.8	0.5	240.6	4.0	153.4	140.1		8/19/2022
ITA0728-1	Itasca	7/28/2022	7:30	1.6	22.8	9.24	7.98	295.5	500	12.3	0.3	238.7	6.0	136.9	131		8/19/2022
DEM0812-1	Demming	8/12/2022	5:24	2	24.4	9.88	5.85	121.6	500	17.6	0.4	131.7	2.6	86.3	77.9		8/19/2022
ELK0812-1	ELk	8/12/2022	3:40	2.3	23.7	10.89	7.72	270.5	1000	31.0	0.3	254.1	2.5	163.6	148.8		8/19/2022
MARY0812-1	Mary	8/12/2022	4:35	3.6	24.1	10.49	6.54	266.5	500	14.4	0.3	206.5	4.1	122.2	115.3		8/19/2022
ITA0812-1	Itasca	8/12/2022	7:04	1.15	23.7	11.6	7.91	294.2	500	27.8	0.6	627.5	12.5	355.3	342		8/19/2022
ARC0812-1	Arco	8/12/2022	6:00	4.1	24.3	10.84	6.61	89.8									8/19/2022
JOS0812-1	Josephine	8/12/2022	6:15	4.1	24.4	10.08	5.84	65.8									8/19/2022
High std	NA	8/19/2022								229.0				229	119.5	1.916	8/19/2022
ARC0825-1	Arco	8/25/2022	4:52	4.2	24.1	11.56	7.18	86.9	1000	57.7	0.6	253.6	2.5	190.1	162.5		8/30/2022
ITA0825-1	Itasca	8/25/2022	6:46	1.2	23.1	10.21	7.88	285.8	500	168.2	3.4	273.3	5.5	310.8	230.4		8/30/2022
ITA0825-1D	Itasca	8/25/2022	6:46					NA	500	165.2	3.3	519.7	10.4	436.5	357.5		8/30/2022
JOS0825-1	Josephine	8/25/2022	5:25	4.2	24.6	11.25	7.06	61.5	1000	35.6	0.4	292.7	2.9	188.3	171.3		8/30/2022
DEM0825-1	Demming	8/25/2022	4:08	1.8	24.9	10.66	7.32	120.7	500	233.9	4.7	176.4	3.5	325.9	214.1		8/30/2022
MARY0825-1	Mary	8/25/2022	3:00	3.3	24.5	11.79	8.11	261.5	1000	88.1	0.9	214.1	2.1	199.8	157.7		8/30/2022
ELK0825-1	Elk	8/25/2022	1:55	2.6	24	12.13	8.01	260.3	1000	304.1	3.0	106.1	1.1	359.5	214.1		8/30/2022
ARC0911-1	Arco	9/11/2022	5:50	3.9	21.8	10.48	9.23	97.3	700	197.0	2.8	113.5	1.6	256.3	162.1		9/30/2022
DEM0911-1	Demming	9/11/2022	5:03	2	22	9.86	8.77	132	500	223.0	4.5	59.1	1.2	253.8	147.2		9/30/2022
ELK0911-1	Elk	9/11/2022	3:02	3.2	21.2	10.77	9.38	291	500	312.1	6.2	61.2	1.2	344	194.8		9/30/2022
ITA0911-1	Itasca	9/11/2022	7:33	1	20.8	9.13	8.88	315	500	327.8	9.8	230.4	6.9	448	291.3		9/30/2022
JOS0911-1	Josephine	9/11/2022	6:32	4.2	21.6	9.86	8.85	67.4	1000	170.1	1.7	93.6	0.9	218.9	137.6		9/30/2022
MARY0911-1	Mary	9/11/2022	4:00	3.4	21.5	9.58	8.96	289	500	338.2	6.8	131.6	2.6	406.9	245.2		9/30/2022
ARC0920-1	Arco	9/20/2022	4:01	3.3	21.1	10.42		98	600	256.0	4.3	167.0	2.8	343.2	220.8		9/30/2022
DEM0920-1	Demming	9/20/2022	3:23	1.6	21.4	9.85		133.4	500	211.9	4.2	219.8	4.4	326.6	225.3		9/30/2022
ELK0920-1	Elk	9/20/2022	1:34	2.9	20.3	10.14		295	500	257.3	5.1	92.8	1.9	305.7	182.7		9/30/2022
ITA0920-1	Itasca	9/20/2022	5:27	1	20	9.38		323	500	311.2	10.0	116.0	3.7	371.8	223		9/30/2022
JOS0920-1	Josephine	9/20/2022	4:32	3.7	21.2	9.74		68.6	1000	142.0	1.4	115.1	1.2	202.1	134.2		9/30/2022
MARY0920-1	Mary	9/20/2022	2:37	3.1	21	9.63		293	500	136.4	2.7	124.2	2.5	201.2	136		9/30/2022
ARC1003-1	Arco	10/3/2022	3:55	2.3	16.9	8.36	7.91	98.7	500	309.4	6.2	56.4	3.8	338.8	190.9		10/12/2022
DEM1003-1	Demming	10/3/2022	3:25	4.7	16.3	8.95	7.86	134.8	500	347.9	8.3	40.1	4.9	368.8	202.5		10/12/2022
DEM1003-1D	Demming	10/3/2022							500	425.5	8.5	35.3	4.8	443.9	240.5		10/12/2022
ELK1003-1	Elk	10/3/2022	1:50	3.4	15.4	12.94	8.48	301	800	378.8	4.7	46.7	2.8	403.2	222.1		10/12/2022
ITA10003-1	Itasca	10/3/2022		1.2	16	9.26	8.35	325	400	426.7	10.7	28.7	5.9	441.7	237.7		10/12/2022
JOS1003-1	Josephine	10/3/2022	4:20	4.9	16.4	8.3	7.44	68.2	700	188.3	2.7	33.2	1.7	205.6	115.6		10/12/2022
MARY1003-1	Mary	10/3/2022	2:34	3.3	15.9	9.93	8.11	299	600	202.7	3.4	28.0	2.0	217.3	120.4		10/12/2022
ELK1026-1	Elk	10/26/2022	12:02	2.9	8.9	11.51	8.42	325	500	268.4	5.4	49.7	3.3	294.3	166		10/28/2022
MARY1026-1	Mary	10/26/2022	2:00	4	9.1	11.63	8.18	315	500	224.0	4.5	35.0	2.7	242.3	135.2		10/28/2022
DEM1026-1	Demming	10/26/2022	3:01	1	8.4	13.59	7.74	142.7	400	453.3	15.9	53.3	9.3	481.1	264.4		10/28/2022
ARC1026-1	Arco	10/26/2022	3:37	1.8	8.7	11.8	7.53	104.3	400	414.2	10.4	56.4	6.1	443.6	245.6		10/28/2022
ARC1026-1D	Arco	10/26/2022							400	434.7	10.9	41.7	6.2	456.4	248.6		10/28/2022
JOS1026-1	Josephine	10/26/2022	4:10	3.7	9.1	11.67	7.41	73.3	400	118.2	3.0	38.0	2.0	138	81.5		10/28/2022
ITA1026-1	Itasca	10/26/2022	5:06	3.1	8.6	13.3	8.29	339	400	287.2	7.2	49.1	4.4	312.8	175.5		10/28/2022
LabBlank819										0.2		3.4		2	1.9		8/19/2022
LabBlank830										8.4		-1.5		7.6	3.6		8/30/2022
LabBlank930										1.3		3.3		3	2.4		9/30/2022
LabBlank1012										1.9		1.0		2.4	1.5		10/12/2002
LabBlank1027										1.7		1.6		2.5	1.7		
Mary911-1D	Mary	9/11/2023							500	254.8	5.1	50.8	3.2	281.3	159.5		9/30/2022
Elk920-1D	Elk	9/20/2022							500	200.8	4.0	59.6	2.7	231.9	135.9		