ABSTRACT

Title of Thesis:	DYANMICS OF PHYTOPLANKTON POPULATIONS IN IRRIGATION PONDS
	Jaclyn Smith, Master of Science, 2022
Thesis Directed By:	Professor Robert Hill, Environmental Science and Technology Department

The dynamics of phytoplankton community structure in two agricultural irrigation ponds located in Maryland, USA were evaluated. Stable spatiotemporal patterns and zones of consistently higher and consistently lower phytoplankton functional group concentrations were established for both ponds. Moderate and strong correlations were found between the spatial patterns of several water quality parameters and phytoplankton concentrations. Additionally, zones of consistently higher and lower concentrations were found for the cyanobacteria pigment, phycocyanin. Chlorophyll, colored dissolved organic matter, and turbidity were the most influential predictors for phycocyanin concentrations. The prediction of phytoplankton community structure from water quality measurements with the random forest machine learning algorithm was possible and easily measured physicochemical parameter models offered the best model performance. Results of this work indicate that in-situ water quality measurements may be a cost-effective and faster alternative to time-intensive microscopy analysis of phytoplankton, allowing for more efficient water quality monitoring.

DYNAMICS OF PHYTOPLANKTON POPULATIONS IN IRRIGATION PONDS

by

Jaclyn Smith

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2022

Advisory Committee:
Professor Robert Hill, Chair
Dr. Patrick Kangas
Dr. Yakov Pachepsky
Jennifer Wolny

Dedication

I dedicate this thesis to my father, who unfortunately never got to see it in its completed form. I wish nothing more than to hear you say that you don't understand a single thing in this document, but nevertheless you're proud of me.

Acknowledgements

I would like to thank the following organizations and individuals for their assistance and support over the course of this degree. I thank the Environmental Microbial Food Safety Laboratory at the USDA-ARS along with the University of Maryland, College Park: Department of Environmental Science and Technology for providing the financial support for this research.

I would like to thank Dr. Robert Hill, who served as the Major Professor in my Master of Science program, who helped substantially in the preparation of this thesis, publications, as well as dedicated a considerable amount of time with me practicing for departmental presentations. I would also like to acknowledge and show my appreciation to my co-advisor Dr. Yakov Pachepsky for not only providing me a start in the environmental science career field, but also mentoring, training, and sharing pieces of his vast knowledge with me in every aspect of research from data collection to publishing.

I would also very much like to acknowledge Ms. Jennifer Wolny for educating and proving me with all the tools necessary to make the phytoplankton portion of this thesis possible. I look forward to everything else you can teach me as well as future collaborations. Thank you for always being my walking, talking, and breathing phytoplankton encyclopedia.

I would like to also thank my coworker at the USDA, Mr. Matthew Stocker for eagerly advising and training me in the laboratory along with your committed dedication to the assistance and execution of both laboratory and field experiments. Additionally, I would like to thank USDA coworkers: Ms. Billie Morgan and Ms. Nhu Le for their help with field and laboratory analyses and experiments.

Table of Contents

Structure and Organization

The organization of this thesis presents an introduction and literature review on phytoplankton importance and dynamics in freshwater sources (Chapter 1), and the results from three experiments elucidating the seasonal and spatial dynamics along with a machine learning application for prediction of phytoplankton community structure in agricultural irrigation ponds (Chapters 2, 3, and 4).

Chapter 1 is a review of our current knowledge surrounding phytoplankton in lentic ecosystems. The importance of phytoplankton in freshwater ecosystems and ecological issues associated with phytoplankton and various functional groups are discussed. Additionally, current knowledge on spatial and temporal variability of phytoplankton and the cyanobacteria pigment, phycocyanin, as well as current machine learning models for phytoplankton estimation are reviewed in depth.

Chapter 2 describes the first research study which focuses on spatiotemporal variability of phytoplankton functional groups and water quality parameters in two agricultural irrigation ponds. Mean relative differences were utilized to distinguish locations which were consistently higher, or consistently lower, than the ponds' average concentrations for each measurement. Correlations between phytoplankton functional group patterns and water quality parameter patterns are also reported. The materials within Chapter 2 were published in November 2021 in the Frontiers in Water special issue, Functional Diversity of Aquatic Microorganisms and Their Roles in Water Quality.

Chapter 3 describes the second research study which examines the most influential environmental covariates affecting the phycocyanin concentrations, an

accessory pigment belonging only to cyanobacteria, in two agricultural irrigation ponds. Phycocyanin is often used a proxy to detect and measure cyanobacteria, including harmful species, in aquatic systems. Regression tree analyses were conducted using machine learning techniques to determine the most influential water quality parameters. Additionally, average quartile ranks were used to elucidate locations within the ponds with consistently higher concentrations of phycocyanin pigment. The materials within Chapter 3 were published in the journal Environmental Monitoring and Assessment in October of 2020.

Chapter 4 focuses on the creation and evaluation of a random forest machine learning algorithm for the estimation of phytoplankton community structure, presented at a broad functional group level, from water quality measurements in two agricultural irrigation ponds. Overall model performance is described along with the performance of the model when spatially applied across the ponds. The most influential water quality predictors for each phytoplankton functional group model are also reported. The materials within Chapter 4 were submitted to the journal Phycology in March 2022 and are currently in the second phase of review, with publication expected within a few weeks.

A summary of the findings and conclusions for all three experimental chapters are presented in Chapter 5. Knowledge gaps and avenues of future research for each experimental chapter are also discussed as part of the summary.

Chapter 1 – Introduction and Literature Review

Agricultural Irrigation Ponds

Feeding the fast approaching eight billion people in the world wouldn't be possible without the increased food production from modern advancements in agriculture (United Nations et al., 2019). Soil, water, and solar energy are some of the necessities for successful agricultural practices. Water, specifically, is essential for the production of healthy and bountiful crops. Agricultural lands span across nearly every climate. Some climates generate sufficient rainfall to produce healthy crops, while other climates rely on other sources of water for irrigation. Irrigation water may be withdrawn from either ground water sources and/or surface waters such as lakes, reservoirs, streams, rivers, and ponds. It is estimated that there are 277,400,000 small (0.001-0.01 km^2) lakes and ponds globally, covering almost 700,000 km^2 of total surface area (Downing et al., 2006). According to the 2018 Irrigation and Water Management Survey, over half of all irrigation water in the United States is obtained from surface water sources (USDA, 2018). However, not all farms may have access to naturally occurring surface waters, and access to local ground water may be limited. In these instances, manmade ponds, embankments, or impoundments may be used to fulfill irrigation needs. Downing et al. (2006) estimates that the global total area of agricultural ponds to be 77,000 km^2 with approximately 22,000 km^2 of agricultural ponds in the contiguous United States. Irrigation ponds, whether natural occurring or man-made, provide a source of water to assist farmers in meeting crop irrigation demands (López-Felices et al., 2020).

Phytoplankton importance in lentic ecosystems

Phytoplankton are free-floating photosynthetic microorganisms which are found in various fresh and marine waters. Phytoplankton have been extensively used as a bio-indicator of water quality and ecosystem health (Patrick, 1973; Reynolds et al., 2012; Smith, 2003). The richness and uniformity of a phytoplankton community as well as the presence or absence of individual species may be indicative of the water quality. Phytoplankton communities have been extensively utilized as a bio-marker of the trophic level status of a waterbody (Adloff et al., 2018; Rimet & Druart, 2018; Xiao et al., 2013). They have also been used to confirm eutrophication (Kauppila et al., 1995; Ren et al., 2016; Varol, 2019), pollution (Feki-Sahnoun et al., 2014; Hu et al., 2012; Shi et al., 2012) and other anthropogenic effects (Feki-Sahnoun et al., 2018; Shi et al., 2015). Phytoplankton are a microbial community often studied for water quality purposes due to their diversity and high environmental sensitivity (Jakhar, 2013; Thakur et al., 2013).

Due to the high diversity of phytoplankton, populations are often binned into groups based on morphology, physiology, adaptations, and ecological attributes (Jin et al., 2020; Reynolds et al., 2002; Varol, 2019). These groups are often referred to as phytoplankton functional groups. Utilizing this method of phytoplankton enumeration, community data may be better used to indicate habitat conditions rather than using the presence or absence of individual species (Reynolds et al. 2002). Common phytoplankton functional groups are diatoms (Bacillariophyta), dinoflagellates (Dinophyta), flagellates, green algae (Chlorophyta), and cyanobacteria (Cyanophyta; also commonly referred to as blue-green algae), each of which possess

qualities that may influence and be indicative of different facets of water quality (Reynolds et al., 2002; Shi et al., 2012; Xiao et al., 2013). The organisms within phytoplankton functional groups possess many different morphological features but share ecological niches which may influence and be indicative of the water quality that they inhabit.

Diatoms have been utilized as water quality indicators since the mid 20th century with early studies in both the United States (Marshall, 1967; Patrick, 1948; Weber, 1971) and Europe (Round, 1961). Diatoms exhibit numerous characteristics which enable their use as bio-indicators for fresh and marine waters including high sensitive to water chemistry changes, abundance in many aquatic systems, and are thoroughly documented with regard to their taxonomy and ecology (Reid et al., 1995). The presence of diatoms alone does not indicate good water quality, but certain diatom species likely indicate water quality conditions from degraded to pristine. Bate et al. (2004) presented a water quality index indicating chemical constituents which would likely be present in river water where a specific diatom species was found. Schoeman (1979) previously devised a simple method of water quality assessment arranging diatom species into four indicator groups based on a more extensive method created by Lange-Bertalot (1979) that differentiates diatoms species for water quality assessment. The Chesapeake Bay Program assesses the water quality throughout Maryland and Virginia using plankton species and water quality data to create a plankton index of biotic index (P-IBI) (Lacouture et al., 2006). The phytoplankton index has also been applied to smaller, freshwater lacustrine systems, as demonstrated by Kane et al. (2009).

Dinoflagellates exist either as a vegetative cell (motile form) throughout the water column or as cysts often found within sediments or substrates (Carty & Parrow, 2015). Dinoflagellate cysts found within sediments are often used as bio-indicators of water quality. Pospelova et al. (2005) reported sites with the highest levels of toxic pollution and hypertrophic conditions often had the lowest richness and concentrations of dinoflagellate cysts. Whereas, Dale et al. (1999) measured double the concentrations of total dinoflagellate cysts and an increase in one species, *Lingulodinium polyedra*, with increased eutrophication. Certain species of dinoflagellates in their planktonic state are known to proliferate in eutrophic conditions (Baohong et al., 2021; Glibert & Burkholder, 2006). Some studies reported when anthropogenic eutrophication was present in coastal waters, diatom populations decreased and other algal groups, such as dinoflagellates, tended to persist (Casé et al., 2008; Kim et al., 2009; Yusoff et al., 2002).

Flagellates, and more specifically euglenoids, in freshwater ecosystems are a promising indicator organism for early detection of possible pollution and/or water degradation (Singh et al., 2013). Euglenoids as a pollution bio-indicator suggest likely organic matter contamination from an external source such as nearby livestock fecal contamination (Nweze, 2009). Species in the *Euglena* group may also pose as an ecotoxicological risk assessment tool due to the species' abilities to grow in wastewater and their sensitivity to various environmental stressors in sewage, which may also be present in water, such as heavy metals, excessive visible and UV radiation, salinity, herbicides, and toxic compounds (Krajčovič et al., 2015).

The presence of chlorophytes in water may indicate pollution. It has been reported that chlorophytes are often present when nutrient concentrations are relatively high, such as in eutrophic and hypertrophic waterbodies. Jensen et al. (1994) found that chlorophytes tended to dominate in hypertrophic, shallow lakes and attributed their growth to the continual input of nutrients from sediments and external sources, such as runoff. Similar findings were reported by Marshall et al. (2006) indicating that the least impaired tidal waters of the Chesapeake Bay tended to have lower chlorophyte biomass than impaired or degraded waters. Additionally, chlorophyte biomasses have been found to be the greatest within areas associated with intensive fish farming and its resulting eutrophication (Mäkinen & Aulio, 1986). Meyer (1971) stated that phosphorus and nitrogen ions present after a cyanobacteria bloom were easily consumed by chlorophyte populations indicating that a large green algae population in a waterbody may suggest water quality degradation. Although chlorophytes may indicate impaired waters, Ali et al. (2010) claims that chlorophytes are responsible for releasing oxygen which offset foul and/or septic conditions which may develop under eutrophic conditions.

Cyanobacteria have a longstanding and well-studied relationship with degraded aquatic environments and are effective indicators of water quality (Pinedo et al., 2007; Soltani et al., 2012; Teta et al., 2017). These microorganisms may quickly multiply and create large blooms on the water surface or along the benthos (Bouma-Gregson et al., 2017; Chorus & Bartram, 1999; Zanchett & Oliveira-Filho, 2013). Cyanobacteria blooms often arise when there is an increase in nutrients, typically nitrogen and phosphorus (Chellappa et al., 2009; Lee et al., 2015a), and

usually occur in low flow and/or stagnant waterbodies (Lee et al. 2015a; Steinberg & Hartmann, 1988). Heterocystous cyanobacteria contain specialized cells which allows for atmospheric nitrogen fixation, thus expanding their pool of nutrient resources. While some non-heterocystous cyanobacteria are capable of nitrogen fixation, the majority cannot, and rely on nutrients available within the water (Paerl, 1996). It has been established that occurrences of cyanobacterial blooms are becoming more frequent due to increases in eutrophication and pollution of waterbodies (O'Neil et al., 2012; Riedinger-Whitmore et al., 2005), but improvement in water quality may decrease the occurrences of these blooms (Gemza, 1997). Climate change poses additional concerns for the frequency, intensity, and duration of cyanobacteria blooms. Proposed effects of global climate change will result in conditions more conducive to cyanobacteria survival and bloom formation such as rising water temperatures, increased water stratification, and increases in precipitation bringing additional nutrients to the water column (Havens et al., 2019; Paerl, 2017, 2018).

While all phytoplankton blooms can cause detrimental effects such as hypoxia, some blooms possess the ability to produce toxins that cause additional harm. These harmful algal blooms (HABs) and cyanobacteria HABs (cyanoHABs) are often comprised of a monospecific population of phytoplankton which can produce a toxin that is detrimental and poses both environmental and human health risks. In marine and estuarine environments HABs are usually composed of dinoflagellates or diatoms (Fu et al., 2012; Hinder et al., 2012) and in freshwater environments HABs are usually composed of cyanobacteria (Paerl et al., 2011; Paerl & Scott, 2010). In the marine environment, diatoms species within the genera

Nitzschia and *Pseudo-nitzchia* are widely known for producing domoic acid, which is the biotoxin responsible for Amnesic Shellfish Poisoning and responsible for significant losses in fisheries revenue, animal mortality, and poising outbreaks (Anderson et al., 2021; Smith et al., 2018). Many marine and some freshwater species of dinoflagellates are known to form harmful algal blooms (i.e., red tides) which can produce biotoxins (Dale, 2000; Li et al., 2015; Rengefors & Legrand, 2001; Steidinger et al., 2008). For example, brevetoxins produced by *Karenia brevis*, have been well established to cause shellfish poisoning and marine fauna mortalities as well as respiratory and gastrointestinal issues in humans exposed to *K. brevis* blooms and toxins (Steidinger, 2009). CyanoHABs, found in freshwaters and the freshwater – estuarine interface are responsible for the production of several toxins including microcystins, nodularins, cylindrospermopsin, anatoxins, and saxitoxins (Huisman et al., 2018). Cyanotoxins are associated with fish kills and the contamination of drinking and recreational waters (Paerl, 2017; Paerl & Otten, 2013). Additionally, cyanotoxins can affect birds, mammals, and humans by causing respiratory issues, contact dermatitis, and when ingested liver, digestive, and neurological issues (Buratti et al., 2017; Huisman et al., 2018; Wood, 2016).

Issues arising from phytoplankton in agricultural irrigation ponds

Aside from being a useful bioindicator of water quality, phytoplankton can often be detrimental to the quality of freshwater pond waters. Phytoplankton blooms, most frequently cyanobacteria, cause a reduction in the dissolved oxygen content and light infiltration of a waterbody (Huisman et al., 2018; Paerl & Otten, 2013). These environmental conditions create uninhabitable hypoxic conditions which can hinder

aquatic vegetation growth and reduce the diversity to or cause mortalities within the aquatic community (Diaz & Breitburg, 2009; Tourville-Poirier et al., 2010). There are also implications from the use of hypoxic water in agricultural settings. Using hypoxic waters for irrigation purposes may lead to oxygen depletion of crop roots which can impede crop growth and decrease crop yields (Bhattarai et al., 2008; Maestre-Valero & Martínez-Alvarez, 2010). Thus, a phytoplankton bloom has the capacity to render a waterbody unusable for agricultural purposes. Phytoplankton may also trigger issues with agricultural equipment. Algal blooms in irrigation waters can lead to clogged pipes and/or filters, resulting in additional wear on irrigation equipment (i.e., irrigation pumps)(Bucks et al., 1979; Nakayama & Bucks, 1991). Boman (1995) estimated that nearly half (46%) of all irrigation clogging was attributed to algae. Similarly, phytoplankton can be detrimental to aquaculture ponds. Beyond inducing hypoxic conditions, cyanobacteria may produce metabolites which add undesirable flavors to the fish (Paerl & Tucker, 1995). It has also been noted that similar to agricultural irrigation practices, phytoplankton in aquaculture systems can lead to the clogging of pipes and filters (Hangwelani et al., 2021; Ramli et al., 2018). The additional stress from clogging may result in more frequent equipment replacements and cleanings.

Certain species of cyanobacteria pose an environmental and human health risk due to their ability to produce toxins, known as cyanotoxins. Exposure to cyanotoxins can occur through ingestion, inhalation, and/or direct contact and may affect the nervous system, skin, and/or the liver (Carmichael, 2001). Cyanotoxins can harm both humans and other animals such as wildlife and livestock. Currently there are no

known antidotes or cures for cyanotoxin poisoning and treatment of symptoms is based on palliative care (Pulido, 2016). To date, there are no regulations in place, only health advisories for cyanotoxins in drinking water, ground water, recreational water, and/or irrigation waters (US Environmental Protection Agency, 2015). Recently, it has been established that cyanotoxins present in irrigation waters can be transported to nearby crops and produce (Corbel et al., 2016; Saqrane & Oudra, 2009). It has also been found that when cyanotoxin-contaminated water is used for irrigation, these toxins can bioaccumulate in agricultural soils (Corbel et al., 2014; Lee et al., 2017) and even in the flesh of crops and produce (Bittencourt-Oliveira et al., 2016; Buratti et al., 2017; Corbel et al., 2016; Kittler et al., 2012). Thus, irrigation water that contains cyanobacteria and cyanotoxins can present food safety risks to both humans and animals (Svirčev et al., 2017).

Spatial and temporal variability of phytoplankton in lentic fresh waterbodies

Uniformity of the phytoplankton throughout a waterbody should not be assumed regardless of waterbody size. Phytoplankton distribution within a water source may be highly variability and may differ between groups and even species within a community (Lewis, 1978). Furthermore, external factors such as wind (Fragoso et al., 2008), morphology (Li et al., 2013), and water flow (Cloern et al., 1992) may impact phytoplankton distributions within a waterbody. Research on phytoplankton communities has previously been conducted to determine spatial and temporal population trends, to assess species composition, and community responses to changes in water quality. Spatial trends in phytoplankton populations have been studied in reservoirs (Ferral et al., 2017; Ren et al., 2016; Varol, 2019), lakes (Wang

et al., 2015; Wu et al., 2013; Xiao et al., 2013), rivers (Marshall, 2009; Shi et al., 2012, 2015), wetland ponds (Soininen et al., 2007), and estuaries (Marshall et al., 2005). Temporal trends of how the phytoplankton community changes in time within a waterbody have also been documented (Su et al., 2017; Wang et al., 2015; Zheng et al., 2015) and long-term phytoplankton datasets have aided in assessing longstanding temporal trends (Hernandez Cordero et al., 2020; Nishikawa et al., 2010). Several studies have evaluated the phytoplankton community composition within a waterbody (Adloff et al., 2018; Cellamare et al., 2010; Rimet & Druart, 2018) and, furthermore, how the phytoplankton communities respond in relation to water quality changes (Hu et al., 2012; Shi et al., 2012; Xiao et al., 2013).

At the regional level, temporal and/or spatial trends have been assessed along with water quality measures in the Chesapeake Bay and its tributaries and estuaries (Hernandez Cordero et al., 2020; Marshall, 2013, 2014; Marshall et al., 2005, 2006). These studies have indicated how species and/or functional groups of phytoplankton may exist in waterbodies on both spatial and temporal scales. Aside from a laboratory study (DeLorenzo et al., 2002), to our knowledge, no attempts have been made to reveal spatiotemporal aspects of the phytoplankton communities in agricultural irrigation ponds.

To determine spatial patterns in a water source, intensive sampling regimens are necessary. Table 1.1 shows a review of 25 publications on phytoplankton monitoring in various fresh waterbodies. This literature review indicates inconsistent methods for where to sample, how many samples to take, and how often to take samples for water quality assessment. Of 25 studies, 12 of these studies reported

spatiotemporal findings on fewer than 5 total sampling dates. Minimal sampling frequency may not allow for an accurate depiction of long-term spatial trends in phytoplankton populations. Furthermore, more than half (15 of 25) of the studies based their findings on 10 or fewer sampling locations within each waterbody. While less than 10 samples may provide a fine enough spatial scale to be able to assess for phytoplankton patterns on small waterbodies such as ponds, small lakes and rivers (Celewicz-Goldyn & Kuczynska-Kippen, 2008; Soininen et al., 2007; Touchart & Bouny, 2008), these small sample numbers may leave considerable water surface areas unanalyzed on larger waterbodies such as estuaries, reservoirs, large lakes, and river basins (Crossetti et al., 2013; Haque et al., 2021; Temponeras et al., 2000; Yang et al., 2020). Fine scale sampling is pertinent to establish spatiotemporal patterns within the phytoplankton community of a waterbody. Figure 1.1 is a graphical representation of the literature review conducted as a part of this study and highlights the number of samples and sampling frequency of the studies reported in table 1.1. The general trend of the data in the literature review is that the more samples that are taken the less often the waterbody is sampled. Alternatively, the less samples taken at each sampling the more often the pond is visited for sampling. The total number of samples taken from the two ponds analyzed in this study were on the high end of sampling frequency of the studies reported in Table 1.1 and Figure 1.1: Pond 2 being the highest of all the studies and Pond 1 having only two other studies with higher total samples. Furthermore, the number of samples per visit at each of the studied ponds was larger than all but four of the studies reported in Table 1.1. Of these four studies with more sampling locations, all studies reported fewer number of visits than

the ponds reported in this study. The design of this work provides some of the best coverages of both spatial and temporal compared with the designs found in the literature reviewed.

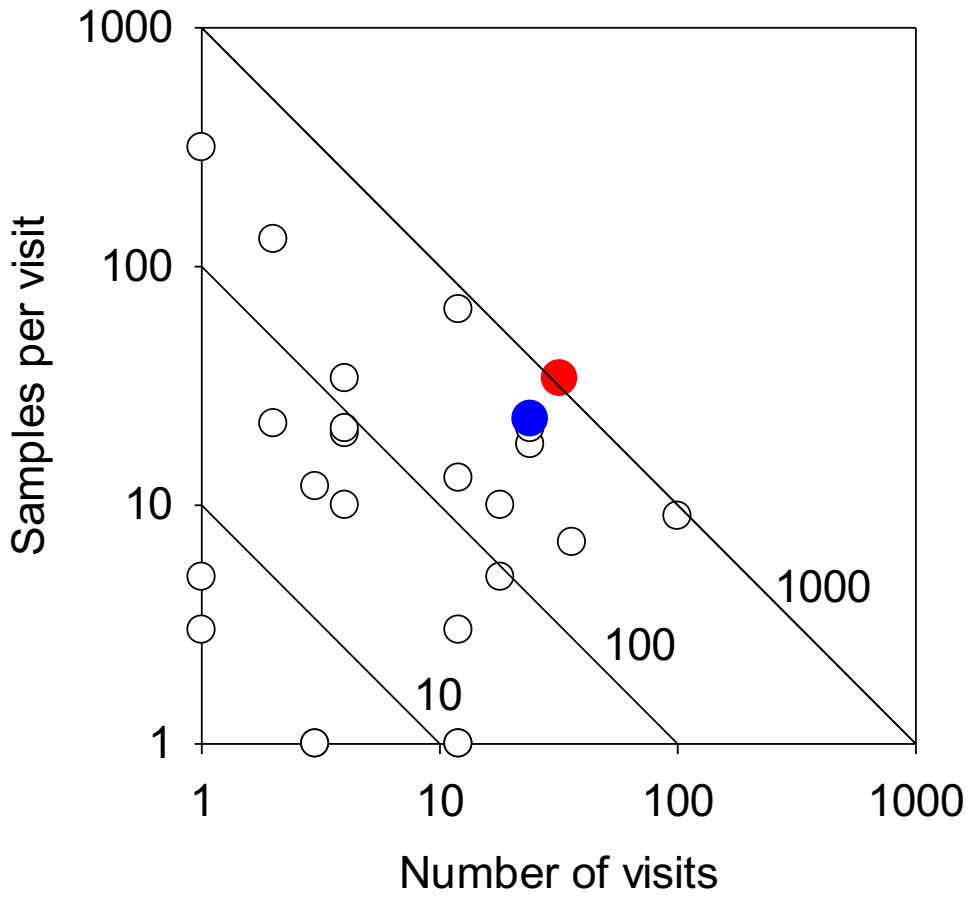


Figure 1.1 Graph depicting the number of samples taken and the frequency of sampling for each study reviewed as part of a literature search. Hollow circles represent studies that are reported in the Table 1 literature review. The blue (Pond 1) and red (Pond 2) circles represent the ponds analyzed in this study.

Determining spatial patterns among phytoplankton groups can have implications on water quality monitoring efforts. If stable spatial patterns throughout time can be established in a waterbody, improvements may be made to the monitoring of phytoplankton populations. Since a waterbody cannot be assumed to be

homogenous and if the spatial variability of phytoplankton populations is inherent across a waterbody, taking one or a limited number of samples may not provide an accurate representation of the phytoplankton community within the entire waterbody and sampling locations should not be chosen arbitrarily. Efforts should be placed on analyses to distinguish locations which are consistently higher, lower, and about the same concentrations as the waterbody's average phytoplankton concentrations. This characterization allows for more informed and efficient sampling to be performed for water quality assessments.

Table 1.1 – Summary of phytoplankton monitoring studies.											
	Target Microorganism/ Parameter	Location	Water Type	Sites	Time of Day	Sampling Frequency	Sample Depth	Samples per visit	Sampling Locations	Total Samples	Reference
1	Diatoms, Dinoflagellates, Cyanobacteria	Central Argentina	Reservoir	1	9:30am – 5:30pm	2 days	0.2 and 14m	22	11	42	Alexander & Imberger, 2009
2	Cyanobacteria, diatoms, chlorophytes, euglenoids	Bangladesh	River	1	9:00am – 12:00pm	Monthly (12)	Surface	1	4	48	Haque et al., 2021
3	Complete phytoplankton composition	Russia	Lake	1	n/a	1 date	0, 5, 10, 25 and 50m	315	63	n/a	Bondarenko et al., 1996
4	Complete phytoplankton composition	Poland	Pond	1	n/a	1 date	<1.5m	3	3	3	Celewicz-Goldyn & Kuczynska-Kippen, 2008
5	Cyanobacteria, diatoms, green algae, flagellates	China	Lake	1	n/a	1991-1999, seasonally 36 dates total	n/a	7	7	252	Chen et al., 2003
6	Complete phytoplankton composition	China	River	1	n/a	2 dates	n/a	130	130	260	Ding et al., 2021
7	Diatoms, green algae, euglenoids,	Vietnam	River	1	n/a	Bi-weekly (24)	0.5m	18	18	432	Nguyen et al., 2022

	cyanobacteria, dinoflagellates										
8	Complete phytoplankton composition	India	River	1	n/a	Seasonally (3)	Surface and "near bottom"	12	6	36	Chowdhury et al., 2017
9	Complete phytoplankton composition	Minnesota, USA	Lakes	2	n/a	Seasonally (4)	0.5M	34	15 and 19	136	Cloern et al., 1992
10	20 phytoplankton functional groups	Rio Grande do Sol	Lake	1	n/a	2001 - 2006, Twice a year	Surface	3	3	36	Crossetti et al., 2013
11	Complete phytoplankton composition	China	Lake	1	n/a	Monthly for 1 year	Surface, 1m, 2m	66	22	792	Deng et al., 2007
12	Cyanobacteria, green algae, diatoms, Euglenoids	China	Lake and River	11	n/a	Seasonally (4)	Surface	20	20	880	Jiang et al., 2014
13	Complete phytoplankton composition	China	Lake	1	n/a	Biweekly from autumn to spring and weekly in summer	0.5m	5	5	222	Li et al., 2013
14	Diatoms, chlorophytes, cyanobacteria	Spain	Reservoirs	4	n/a	2002-2003 Monthly (24)	1m	21	21	504	Moreno-Ostos et al., 2008

15	Complete phytoplankton composition	Finland	Ponds	25	n/a	Once	0.5m	5	5	125	Soininen et al., 2007
16	Complete phytoplankton composition	China	River	1	n/a	3 dates	Surface	1	10	30	Song et al., 2020
17	Complete phytoplankton composition	Greece	Lake	1	n/a	Monthly (9)	Surface and 1m intervals	Varies through season	2	n/a	Temponeras et al., 2000
18	Complete phytoplankton composition	China	Lake	1	n/a	Monthly (12)	0.5m	1	5	60	Tian et al., 2013
19	Complete phytoplankton composition	France	Pond	2	n/a	2 dates and 1 date	n/a	1	20 and 12	36	Touchart & Bouny, 2008
20	Cyanobacteria, chlorophytes, diatoms	China	Reservoir	1	n/a	Seasonal (4)	1m, 4-7m, and 10-20m	21	7	84	Yang et al., 2020
21	6 Phytoplankton functional groups	China	River Basin	1	n/a	2015-2016 Bi-annually (4)	10cm	10	10	40	Zhou et al., 2019
22	Cyanobacteria	China	Reservoirs	13	n/a	Spring, summer, and autumn of 2014-2015 (6)	0-2m	Varied on waterbody	90	480	Zhao et al., 2019
23	6 phytoplankton functional groups	India	Lagoon	1	n/a	Monthly (12)	Surface	13	13	156	Srichandan et al., 2015

24	Cyanobacteria, chlorophytes, diatoms	China	Estuary	1	n/a	Monthly (18)	n/a	10	10	180	Ren et al., 2016
25	Cyanobacteria, green algae, diatoms	China	River	1	n/a	Weekly during wet season (Jun-Oct) 2012-2017 (100)	n/a	9	9	900	Yang et al., 2020

In-situ sampling to relate cyanobacteria populations to water quality in lentic ecosystems

Cyanobacteria are the phytoplankton group that presents the most globally widespread risk to human and animal health due to the presence of cyanotoxins (Hilborn & Beasley, 2015; Wood, 2016). Traditional methods for identifying and enumerating cyanobacteria in a water source are time consuming and require highly trained specialists (Lawton et al., 1999). Measuring phycocyanin, a photosynthetic pigment found only in cyanobacteria, can be a quicker and easier alternative method for identifying and quantifying cyanobacteria in a water source. While several methodologies exist for extracting phycocyanin pigments (Horváth et al., 2013), fluorometry provides immediate results (Kasinak et al., 2015). Furthermore, deploying an in-situ phycocyanin sensor may provide instant real-time phycocyanin measurements allowing for improved bloom predictions and management (Bastien et al., 2011; Brient et al., 2008). Establishing the correlations and relationships between phycocyanin and other basic water quality parameters could allow for the use of inexpensive and more attainable in-situ probes to be utilized for cyanobacteria monitoring and management.

It should not be assumed that cyanobacteria and/or phycocyanin concentrations are homogeneous throughout a waterbody. Spatial and temporal trends and patterns in cyanobacteria populations have been reported in lakes (Otten et al., 2012; Tan et al., 2009), rivers (Genzoli & Kann, 2016), ponds (Andres et al., 2019; Rozina et al., 2018), and reservoirs (Briand et al., 2009). Spatiotemporal variability in

phycocyanin concentrations and cyanobacteria populations has been attributed to several factors such as wind (Foster et al., 2019), water flow (Paerl & Otten, 2013), and waterbody depth (Andres et al., 2019; Scheffer et al., 1997). Cyanobacteria spatial and temporal variability have also been explained by nutrient availability and the ability for cyanobacteria to proliferate in nutrient rich areas (Davis et al., 2009; Havens et al., 2003; Paerl et al., 2011). While studies on the spatial and temporal trends of cyanobacteria and its photosynthetic pigment phycocyanin have been completed in various freshwater sources, research is lacking on agricultural irrigation ponds where the risk of cyanotoxin transfer to livestock, crops, and produce exists.

It has been well documented that relationships between phycocyanin and water quality parameters exist in various sizes and types of waterbodies (Izydorczyk et al., 2005; McQuaid et al., 2011). Relationships between phycocyanin and water quality parameters have been established with temperature, turbidity, nutrients, chlorophyll, wave height, and meteorological measurements (Marion et al., 2012; Mchau et al., 2019; Song et al., 2013; Yang et al., 2021). Previously used methods to establish the relationships between phycocyanin concentrations and water quality parameters seemingly have been limited to correlations (Ahn et al., 2011; Izydorczyk et al., 2005) and linear regressions (Bastien et al., 2011; Thomson-Laing et al., 2020; Yang et al., 2021). Limited research has been performed using machine learning methodologies to model and reveal the complex relationships which may exist between the presence of phycocyanin and basic water quality measurements. To our knowledge the use of machine learning to discover relationships between phycocyanin and water quality parameters has only been attempted in two rivers

(Heddam et al., 2019) and a lake (Almuhtaram et al., 2021). An understanding of these complex relationships could improve water quality monitoring by allowing rapid in-situ measurements to be used as a surrogate measure for detecting and tracking cyanobacteria populations within agricultural ponds.

Monitoring and modeling environmental controls of phytoplankton populations in lentic ecosystems

Water quality monitoring as defined by Bartram and Ballance (2020) is the long-term standardized sample collection, measurement, and observation in order to define the status and trends of a waterbody. The purpose behind water quality monitoring is to ensure that the water being assessed can be utilized either for recreational, drinking, and/or irrigation purposes. Long-term datasets are useful for trend evaluations and ultimately can be used to help predict future conditions and guidelines for resource management plans. To predict or prevent phytoplankton growth is pertinent in building an understanding of the environmental controls which dictate the diversity and abundance of phytoplankton, and long-term monitoring data makes this prediction feasible (Pathak et al., 2021; Pinckney et al., 1997; Read et al., 2014). The identification and enumeration of phytoplankton is time intensive and requires specialized taxonomic knowledge (Lawton et al., 1999) or equipment that may be expensive, such as flow cytometers or imaging flow cytometers (Bergkemper & Weisse, 2018; Read et al., 2014). If relationships can be detected between phytoplankton populations and more easily measured water quality parameters, this association can allow for better prediction capabilities with data collected by more

moderate means. To our knowledge no long-term monitoring studies have been performed for agricultural irrigation ponds.

Various types of analyses have been performed to help understand the relationships between water quality parameters and phytoplankton groups. Correlations have been utilized extensively to establish relationships between water quality parameters and diatoms (Pan et al., 1996; Pourafrasyabi & Ramezanpour, 2014; Wu et al., 2013), green algae (de Figueiredo et al., 2006; Kane et al., 2009), and cyanobacteria (Davis et al., 2009; Sadegh et al., 2021; Smith et al., 1987). Regression analysis has also been applied to highlight relationships among water quality parameters and phytoplankton groups (Descy et al., 2016; Rao et al., 2021; Schönfelder et al., 2002). Statistical regressions have been successfully used to predict and distinguish phytoplankton community composition and dynamics in shallow freshwater lakes (Cheruvelil et al., 2008; Peng et al., 2021; Tian et al., 2013) and reservoirs (Zeng et al., 2017) but not within agricultural irrigation ponds.

Although regressions and correlations allow for linear relationships to be established, these models fail to recognize non-linear relationships present amongst a dataset (Jordan & Mitchell, 2015; Murphy, 2012) and, therefore, may potentially overlook influential relationships between water quality parameters and phytoplankton populations. Machine learning, a branch of artificial intelligence, is able to map non-linear and complex relationships within a dataset (Jordan & Mitchell, 2015). Neural networks have been utilized to extract non-linear and complex water quality relationships associated with cyanobacteria (Rousso et al., 2020; Yabunaka et al., 1997) and chlorophyll-a (Liu et al., 2015; Wu et al., 2020). The random forest

algorithm, another form of machine learning, has proven to be a very powerful algorithm due to its prediction capabilities (Breiman, 2001). A review of available literature on machine learning and its application to phytoplankton populations is depicted in Table 1.2. This literature review reveals that the application of random forests to predict and determine the environmental controls of phytoplankton have been studied in various freshwater sources such as reservoirs (Zeng et al., 2017), lakes (Mellios et al., 2020), and rivers (Shin et al., 2017), but research is lacking for small waterbodies such as agricultural irrigation ponds. Additionally, the review reveals that when multiple models were compared for accuracy with the random forest algorithm, in all but two instances the random forest algorithm was determined to be the superior model.

Phytoplankton is an extremely important component of lentic ecosystems, and this review manifests its importance in environmental, public health, and management decisions. Based on the review of the status of the knowledge on phytoplankton in lentic ecosystems including agricultural environments and importance of irrigation ponds in agricultural enterprises, the primary research goal for this study was to form a better understanding of the dynamics and relationships of phytoplankton populations in agricultural irrigation ponds. The specific objectives for this study are:

1. Reveal and quantify patterns in the spatial and temporal variability of phytoplankton in agricultural irrigation ponds.
2. Evaluate in-situ sensing of the cyanobacteria pigment phycocyanin as the means for characterization of spatial variability of cyanobacteria populations and its relation to other water quality parameters.

3. Model the effects of the environmental controls on phytoplankton in agricultural irrigation ponds using machine learning algorithms.

Table 1.2 – Summary of phytoplankton studies incorporating machine learning methodologies.								
Target Parameter/ Group	Water Type	Predictors	Models	Best Model	Metric	Ideal value for metric	Performance	Reference
Cyanobacteria, Diatoms, Green algae	Reservoir	Temp, Secchi, pH, Ca, Mg, DO, DIN, TP, Cl, SO4	ANN, RF, SVM, RT	RF	R	1	0.806	Zeng et al., 2017
Phytoplankton Functional Groups (21 total)	Lakes	Chl-a, SPC, DOC, NH4, NO3, NO2, DO, PO4, TP, Si, Temp, TSS, SEC, ALK	RF	RF	n/a	n/a	n/a	Derot et al., 2020
Chlorophyll-*a*	Fresh and estuarine reservoirs	PO4, NH4, NO3, Temp, Wind, Solar	ANN, SVM	SVM	R2	1	0.75, 0.45	Park et al., 2021
Cyanobacteria	Lakes	Chl-a, TN, TP, AT ELV, SA, WD	RT, KNN, SVM, RF	RF	Accuracy	100%	95.45%	Mellios et al., 2020
Cyanobacteria	Reservoir	DO, ELV, pH, SPC, Temp, NTU, Si, TN, NH4, NO3+NO2, PO4, TP, FC, Chl-a, Fe, SS, TN:TP	RF, SVM, BT	SVM	RMSE	0	n/a	Harris & Graham, 2017
Cyanobacteria	Reservoir	NH4, NO3, PO4, SEC, BOD, Temp, SPC, pH	ANN	n/a	RMSE	0	2.1594	Srisuksomwong & Pekkoh, 2020

Diatoms	river	NH4, Cl, K, Ca, Na, Mg, Si, PO4, SO4, TP, TSS, DD, BF, Runoff, Precip, WD	RF, GLM	GLM	R2	1	0.25	Sun et al., 2022
Cyanobacteria, Diatoms, Chlorophytes	Lake	SEC, WD, Temp, DO, SPC, ORP, TN, NH4, NO3, TP, PO4, Chl-a	GLM	n/a	R2	1	Cyano: 0.43 Dia:0.28 Chl:0.51	Zhu et al., 2018
Chlorophyll-*a*	Lake	Chl-a, DO, TP, PO4, TN, NO3, NO2, SEC, Temp, ALK, Cl, WC, NTU	RF, GAM	RF	R2	1	0.702	Zhang et al., 2021
Diatoms, Cyanobacteria, Chlorophytes	River	Temp, pH, DO, SPC, Na, TDS, TSS, TN, TP, Si, POC, DOC, Chl-a	ANN	n/a	r	1	0.678	Duong et al., 2019
Diatoms, Cyanobacteria, Chlorophytes	Reservoir, river	Temp, DO, pH, TN, TP, Ca, Mg, K, Na, Si, SO4, NO3, Cl	BPANN,	n/a	r	1	0.66-0.97	Wang et al., 2018
Chlorophyll-*a*	Lake	DO, PO4, NO3, NH4, Si, DON, DOP, POC, Temp	ANN	n/a	R2	1	0.83	Wang & Wang, 2021

Ca: Calcium ion, Mg: Magnesium ion, Cl: Chloride ion, SO4: Sulfate, NH4: Ammonia, DO: Dissolved oxygen, Si: Silica, Temp: Water temperature, SEC: Secchi disk, AT: Air temperature, TN: Total nitrogen, TP: Total phosphorus, Chl-a: Chlorophyll a, SPC: Conductivity, ORP, Oxidation reduction potential, DIN: Dissolved organic nitrogen, NO3: Nitrate, NO2: Nitrite, TSS: Total suspended solids, ALK: Alkalinity, PO4: phosphate, DOC: Dissolved organic carbon, Solar: Solar radiation, Wind: Wind speed, ELV:

Elevation, SA: Surface area, WD: Water depth, NTU: Turbidity, FC: Fecal Coliforms, SS: Suspended Solids, Fe: Iron, BOD: Biological oxygen demand, K: Potassium ion, Na: Salinity, DD: Daily discharge, BF: Baseflow, Precip: Precipitation, Runoff: Surface runoff, WC: Water color, POC: Particulate organic carbon, TDS: Total dissolved solids, DON: Dissolved organic nitrogen, DOP: Dissolved organic phosphorus. ANN: Artificial neural network, RF: Random Forest algorithm, RT: Regression Trees, SVM: Support vector machine, BT: Boosted tree, GLM: Generalized linear model, GAM: Generalized additive model, BPANN: Back propagation artificial neural networ

Chapter 2: Temporal stability of phytoplankton functional groups within two agricultural irrigation ponds in Maryland, USA

2.1. Introduction

Phytoplankton are commonly found members of microbial populations within many diverse waterbodies including agricultural irrigation ponds. These primary producers are an important component of the food web within aquatic ecosystems. Previous research has shown that phytoplankton may be an effective bio-indicator of water quality and also a reflection of ecosystem health (Adloff et al., 2018; Su et al., 2017; Wang et al., 2015).

Freshwater phytoplankton populations are typically divided into functional groups based on morphology, physiology, adaptations, and ecological attributes (Jin et al., 2020; Reynolds et al., 2002; Varol, 2019). Three major phytoplankton functional groups are diatoms (Bacillariophyta), green algae (Chlorophyta), and cyanobacteria (Cyanophyta; also commonly referred to as blue-green algae), each of which possess different qualities that may influence and be indicative of water quality (Shi et al., 2012, 2015; Xiao et al., 2013). The richness and uniformity of the phytoplankton community may also indicate different water properties and a range of water qualities from pristine to degraded water quality conditions. Phytoplankton communities have been utilized as an indication of the trophic state of a waterbody (Hu et al., 2012; Ren et al., 2016; Rimet & Druart, 2018), to confirm eutrophication (Ren et al., 2016; Varol, 2019), pollution and/or other anthropogenic effects (Feki-Sahnoun et al., 2018; Shi et al., 2015). The use of phytoplankton functional groups in more complex assessments, such as understanding biogeochemical models (Shimoda

& Arhonditsis, 2016) and in the development of remote sensing technologies (Vandermeulen et al., 2017; Wolanin et al., 2016; Xi et al., 2017) continues to be a growing research area in large waterbodies or on broad scales, but less is known about the temporal stability of these groups on smaller scale irrigation water systems (e.g., irrigation ponds, retention ponds, and aquaculture ponds).

Agricultural irrigation water has been shown to play a substantial role in the microbial contamination of fresh produce and foodborne illness outbreaks (Jongman & Korsten, 2018; Uyttendaele et al., 2015; World Health Organization, 2008). Certain groups of phytoplankton can form large proliferations or "blooms" and release toxins into the environment (Bouma-Gregson et al., 2017; Wood, 2016) which can be biotransported into the food supply (Bittencourt-Oliveira et al., 2016; Buratti et al., 2017). This presents both environmental and human health risks. Monitoring of irrigation water quality is important to avoid the transport of degraded and potentially contaminated waters to nearby crops.

Research on phytoplankton communities has previously been conducted across numerous waterbody types to determine spatial and temporal population trends, assess species composition, and community responses to changes in water quality. Within the Chesapeake Bay watershed, long-term phytoplankton data sets have been used to augment and support water quality guidelines in lakes, rivers, and estuaries (Hernandez Cordero et al., 2020; Marshall, 2013, 2014; Marshall et al., 2006, 2009), but not specifically for agricultural irrigation waters.

Although phytoplankton may be used as water quality bio-indicators, attempts to integrate phytoplankton community assessments to agricultural irrigation water

quality seemingly have been limited to laboratory studies (DeLorenzo et al., 2002). The objective of this study was to determine if temporally stable spatial patterns of phytoplankton functional groups exist within temperate agricultural irrigation ponds and if these groups could be correlated to easily measured water quality parameters which could lead to potential improvements in on-farm water quality monitoring and aid with the prediction and mitigation of food-safety issues.

2.2. Material and methods

2.2.1. Pond Monitoring

Sampling was conducted at two working farms for two consecutive growing seasons (2017-2018). These ponds were chosen because water was routinely drawn for irrigation of co-located crop fields. Each pond was sampled six times during the May through October growing season, with an exception to Pond 2 in 2017 with only five sampling dates. For 2017 sampling occurred from May to August and for 2018 from June to October. This resulted in a total of 276, and 242 phytoplankton samples collected for Pond 1 and Pond 2, respectively. Both ponds were located within a one-

hour drive from the USDA-ARS laboratory, so samples were maintained at ambient temperature and processed the same day as collection.

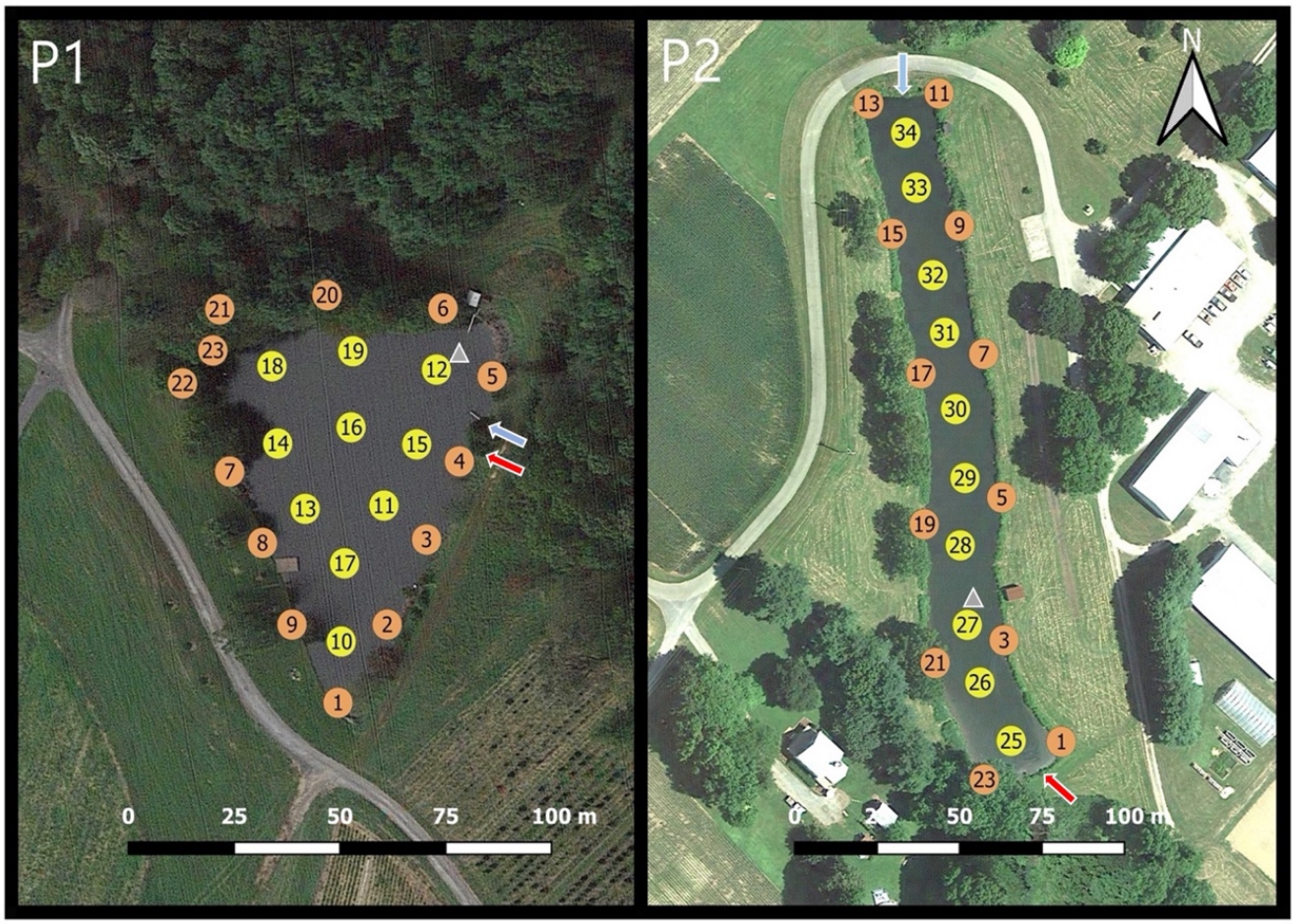


Figure 2.1. Sampling locations for both Pond 1 (P1) and Pond 2 (P2). Station location number is stated inside the circle. Yellow circles indicate interior water sampling locations and orange circles indicate nearshore sampling locations. Blue arrows represent inflow points and outflows are represented by red arrows. Irrigation intake is represented by a grey triangle.

Pond 1 is a 1.01-acre man-made embankment pond located in Germantown, MD, USA with an average depth of 2.7m (Figure 2.1- P1). Vegetation surrounding Pond 1 embankments consisted of deciduous trees and shrubs along the northern and eastern banks with the remaining embankments having a grass cover. The pond is surrounded by crop fields. When the water level in this pond gets low, the farm operators will occasionally pump water into Pond 1 from another pond which is

stream-fed. The inflow and outflows are both located near sampling location 15. The irrigation pump intake is located near location 12 and is approximately two-three feet below the water's surface. The photic zone in Pond 1, as determined by Secchi depth, averages 0.8 m. In 2017 and 2018, the algicide copper sulfate was commonly used to treat the water in Pond 1.

Pond 2 is located at the University of Maryland Wye Research Center in Wye Mills, MD, USA (Figure 2.1-P2). This pond is a 1.05-acre excavated pond with an average depth of 2.7m and most of the bank areas are covered with grass and dense shrubs. Large trees are also present along the perimeter but are approximately 20m from the water's edge. This pond is surrounded by crop fields, farm buildings, and one residential property. In March of each year, the surrounding crop fields receive chemical fertilizers, but no animal manures are applied. This pond is primarily fed through rainfall which typically enters through an ephemeral creek that leads into a culvert near location 12. This culvert tends to have a substantial inflow only when precipitation has recently occurred. On the south end of the pond, there is a water-level dependent outflow drain near location 24. The irrigation pump intake is near location 27 and is approximately two-three feet below the water's surface. The depth of the photic zone, determined by Secchi depth, for Pond 2 averages 0.5 m.

2.2.2. Sample collection, handling, and storage

Pond 1 had 23 sampling locations and Pond 2 had 22 sampling locations (Figure 2.1). Surface water samples were taken at a depth of 0-15cm. Nearshore samples were taken with a 500mL hand grab sampler at approximately 1.5m from the shoreline. Interior samples were taken from a boat with GPS tracking used to provide

consistency of sampling locations between different sampling dates. Sampling locations remained the same for every sampling date over both years. After collection, samples were immediately placed into a cooler without ice to help maintain the original ambient water temperature. Samples were then transported to the lab for analysis.

2.2.3. In-field measurements

In-situ water quality measurements were taken concurrently with sample collection using a YSI Exo-2 sonde (YSI Inc., Yellow Springs, OH). The YSI sonde was used to measure temperature (°C), dissolved oxygen (DO mg L^{-1}), pH, fluorescent dissolved organic matter (*f*DOM, RFU), chlorophyll-*a* (CHL YSI, RFU), phycocyanin (Phyco YSI, RFU), and turbidity (NTU). A Secchi disk was used to measure water transparency, approximating the photic zone depth (m). Precipitation data was obtained from weather stations located within three km of each pond.

2.2.4. Laboratory measurements

Water samples were measured for colored dissolved organic matter (CDOM, µg L^{-1}), in-vivo or whole-cell chlorophyll-*a* (CHL RFU, RFU), and phycocyanin (Phyco LAB, µg L^{-1}) using an Aquafluor fluorometer (Turner Designs, San Jose, CA). Samples were also processed and measured for extracted chlorophyll (CHL EXT, µg L^{-1}) following EPA method 445 (Environmental Protection Agency, 1997) using an Aquafluor fluorometer. For the extraction process, approximately 100mL of pond water was vacuum filtered using 0.7µm glass fiber filters (Whatman, Maidstone, United Kingdom) and steeped in a 90% acetone and 10% deionized water solution overnight at 4°C before being analyzed with the fluorometer. A subsample of

approximately 50mL was taken for phytoplankton identification and enumeration. This subsample was preserved with Lugol's iodine solution at a 1% final concentration. Subsamples were stored at 4°C and in the dark to prevent phytoplankton cell degradation until microscopic analysis could be completed.

2.2.5. Microscope analysis

During examination and enumeration of the preserved phytoplankton samples each phytoplankton was identified to the lowest taxon possible using John et al. (2011) and Bellinger & Sigee (2015). To assess the phytoplankton community at the group level species data was recorded as cell abundance (cells L^{-1}) and then classified into one of four major phytoplankton functional groups: diatoms, dinoflagellates, chlorophytes (including motile and non-motile species), and cyanobacteria as done for corresponding long-term, regional datasets (Lamlou, 1977; Marshall, 2013, 2014; Marshall et al., 2006). Because of the infrequent occurrence of dinoflagellate species in both ponds over the two years these data were not included in the final analysis but are available in Appendix A, Supplemental Figure 2.2. The cell abundance data for potentially toxic cyanobacteria species were compared with cell abundances presented in national and regional action guidelines (Environmental Protection Agency, 2019; Virginia Department of Health, 2015).

All phytoplankton samples were examined using a Nikon Ts2R inverted microscope (Nikon Instruments Inc., Melville, NY) and a modified Utermöhl method as described in Marshall & Alden (1990). A two- or three-mL Lugol's iodine preserved sample was pipetted into a chambered covered glass slide (Thermo Scientific, Rochester, NY), and allowed to settle for thirty minutes to one hour. After

settling, enumeration started in the upper left-hand corner of the chambered slide. After the first frame was counted, the next frame would be moved down and to the right to avoid frame overlap and possible double counting of algal cells. This movement of the field of view created a diagonal pattern across the cover glass slide. The frames were counted in this pattern until either a 200-cell minimum or 20 frames were examined.

2.2.6. Statistics and graphics

To assess spatio-temporal stability of phytoplankton functional groups, mean relative difference method (MRD) was applied. The mean relative difference indicates how an individual location compares to the pond average over multiple sampling dates and reveals areas that are consistently higher or lower than the pond's average for a measured parameter. This method applied here follows those reported in other spatial pattern studies (Pachepsky et al., 2017; Stocker et al., 2018). The relative difference RD_{ij} between the observation of variable x at location *i* at time *j* (x_{ij}), and the spatial average of x at the same time ($\langle x \rangle_j$), is defined as:

$$RD_{ij} = x_{ij} - \langle x_j \rangle$$

The MRD for location *i* then becomes

$$MRD_i = \frac{1}{N_t} \sum_{j=1}^{N_t} RD_{ij}$$

Where N_t is the number of sampling days, and $i = 1, 2, \ldots, N_i$, where N_i is the total number of locations.

The coefficient of variation (CV) was computed for each phytoplankton functional group for each date and pond. The calculation for CV is defined as:

$$CV = \sigma_{ij} / \mu_{ij}$$

Where σ_{ij} is the population standard deviation of phytoplankton functional group *i* on sampling date *j* and μ_{ij} is the population mean of phytoplankton functional group *i* on sampling date *j*.

Mean relative differences and Spearman rank correlations were computed in RStudio. Correlations were considered moderate if $r \geq 0.400$ (p values, P1=0.059 P2=0.065) and considered strong if $r \geq 0.600$ (p values, P1<0.001 P2<0.001). Sigmaplot v. 13 (SYSTAT, Chicago, IL, USA) and QGIS (OSGeo, Switzerland) were used to create visual representations of the data.

2.3. Results

2.3.1. Data summary

2.3.1.1. Weather data

Daily ambient air temperature and precipitation data for both ponds and years are displayed in Figure 2.2. In 2017, there was an increase in the air temperatures at both ponds from May to July. In 2018, the initial increase in temperature was less pronounced than in 2017. Pond 1 experienced more rainfall in 2018 compared to 2017. Information on the number of days following the last rainfall event from sampling dates and total rainfall accumulations may be seen in Appendix A, Supplemental Table 2.1. Over the two years, sampling at Pond 1 was performed six times with a rainfall event occurring the day before sampling, three times with a

rainfall event occurring one to three days before sampling, and three times when a rainfall event was four or more days before sampling. At Pond 2, sampling was done twice with a rainfall event occurring the day before sampling, six times with a rainfall event occurring one to three days before sampling, and three times with a rainfall event occurring four or more days before sampling. Major precipitation events (>6cm) at Pond 1 occurred on 7/28/17 and 7/21/18 with daily rainfall accumulations of 10.04 and 14.10 cm, respectively. Sampling near both major precipitation dates was avoided, and sampling was not conducted for three days following a major event. At Pond 2, major precipitation events occurred on 7/28/17, 7/29/17, 8/7/17, and 7/21/18 with daily rainfall accumulations of 8.28 cm, 8.40 cm, 16.76 cm, and 8.03 cm, respectively. Sampling was avoided within three days of these rainfall events with the exception of the 8/7/17 event. Sampling occurred on 8/8/17 which was one day following a major rainfall event.

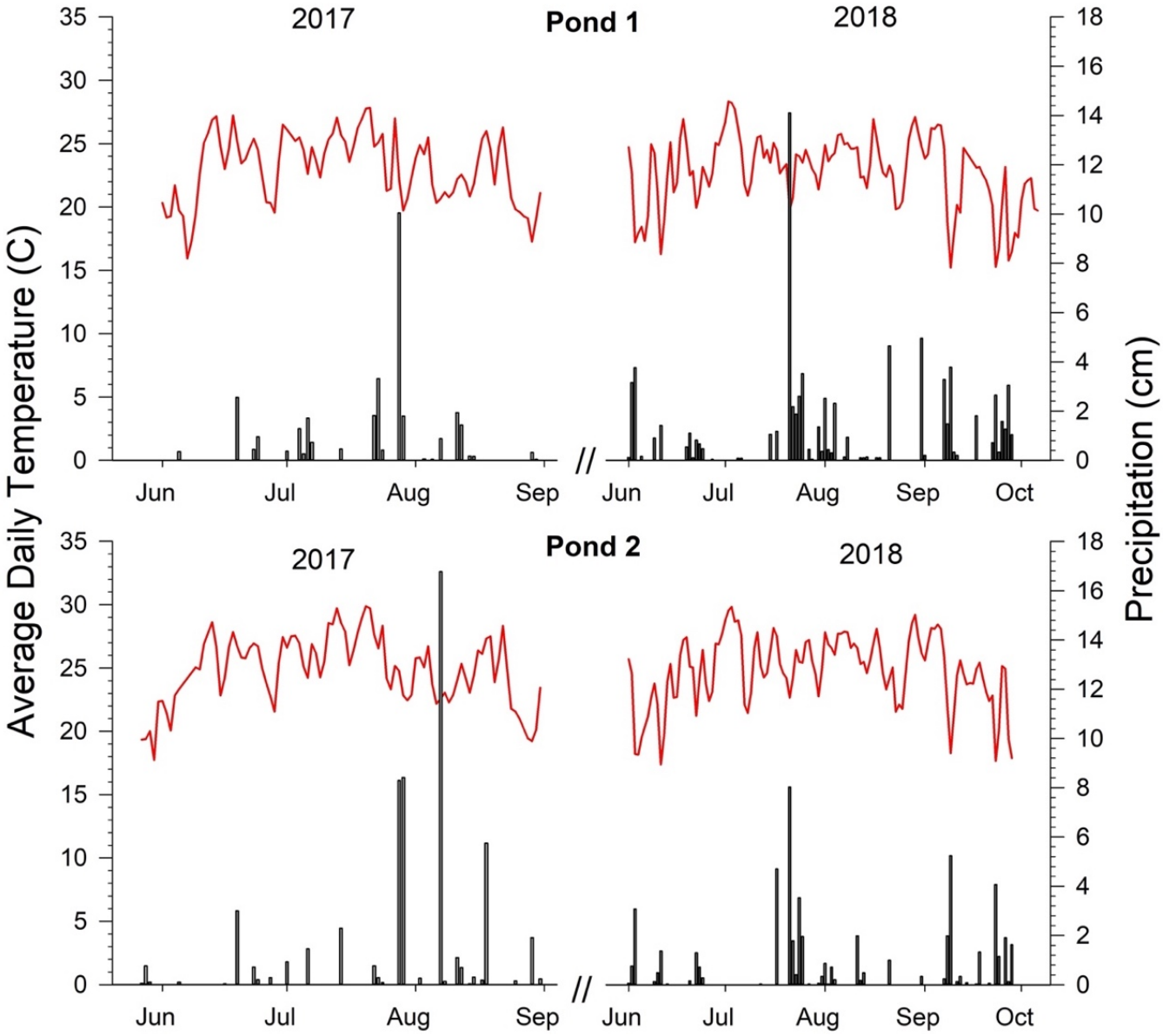


Figure 2.2. Precipitation and ambient air temperature data for both ponds. Precipitation data is represented in bars and temperature data are shown with red lines.

2.3.1.2. Water quality parameters

Time series data of water quality parameters measured for 2017 and 2018 are presented in Appendix A, Supplemental Tables 2.2 and 2.3. Mean values of all measurements related to phytoplankton pigments (Phyco YSI, CHL YSI, EXT CHL, LAB CHL, and Phyco LAB) were generally higher at Pond 2 than at Pond 1 for both

2017 and 2018. Specific conductance and pH measurements were lower for both ponds in 2018 compared to 2017. The positive relationship between higher pH and higher DO concentrations was more pronounced for Pond 2 compared to Pond 1. Algicide was applied to Pond 1 after the first sampling date on 7/1/18. Consequently, in Pond 1 all measurements related to phytoplankton pigments (Phyco YSI, CHL YSI, CHL EXT, LAB CHL, and Phyco LAB) displayed large decreases on sampling date 7/5/18. Phycocyanin measurements remained low for the remainder of the sampling season (Phyco YSI, Phyco LAB), while chlorophyll measurements recovered after two sampling dates (CHL YSI, CHL EXT, LAB CHL). A decrease in Phyco YSI, CHL YSI, and CHL EXT measurements was seen in Pond 2 in 2017 following a 16.8cm rainfall event. Phycocyanin measurements (Phyco YSI and Phyco Lab) for both ponds in 2018 indicated a cyanobacterial bloom was present during the first sampling dates. Phyco Lab measurements on the first sampling dates were 114 μg L^{-1} and 110 μg L^{-1} for Pond 1 and Pond 2, respectively. Furthermore, these blooms were also visually identified by the appearance of green surface scums and confirmed via microscopy analysis of phytoplankton samples. Phycocyanin measurements remained approximately the same in Pond 2 during the entire sampling season.

2.3.1.3. Phytoplankton functional groups

The time series data of log concentrations of green algae, diatoms, and cyanobacteria for both ponds and years are presented in the box plot graphs of Figure 2.3. Descriptive statistics for all phytoplankton groups, both ponds, and both sampling years are reported in Appendix A, Supplemental Table 2.4. Green algae displayed the lowest variability and cyanobacteria displayed the highest variability

among the phytoplankton functional groups for both ponds. The coefficients of variation (CV) values are presented in Appendix A, Supplemental Table 2.5. In 2017 the CVs for Pond 1 ranged from 0.024 to 0.066 for green algae, 0.064 to 0.124 for diatoms, and 0.074 to 0.214 for cyanobacteria. The CVs followed a similar pattern in Pond 1 during 2018 with green algae CVs ranging from 0.023 to 0.051, diatoms from 0.044 to 0.132, and cyanobacteria from 0.040 to 0.262. The green algae CVs were generally lower in Pond 2 than the values for diatoms and cyanobacteria. The CVs for diatoms and cyanobacteria did not follow the same pattern as found for Pond 1, the overall ranges of the diatom CVs were less than the cyanobacteria CVs for each respective sampling season. Green algae and diatoms had a similar intra-seasonal (May-August) trend during 2017 at both ponds wherein population growth occurred from May to June followed by a period of stabilization for the remainder of the sampling season. Diatoms and green algae in Pond 1 exhibited similar trends in 2018 (June-October) displaying a period of stabilization from June to July followed by a drop in concentrations for the remainder of the sampling season. Cyanobacteria trends were drastically different from 2017 to 2018 for both ponds. A cyanobacteria bloom was observed within both ponds during June 2018. During this study, copper sulfate was applied to Pond 1 on 7/1/18 and impacted the total phytoplankton concentrations, particularly decreasing the abundance of cyanobacteria species.

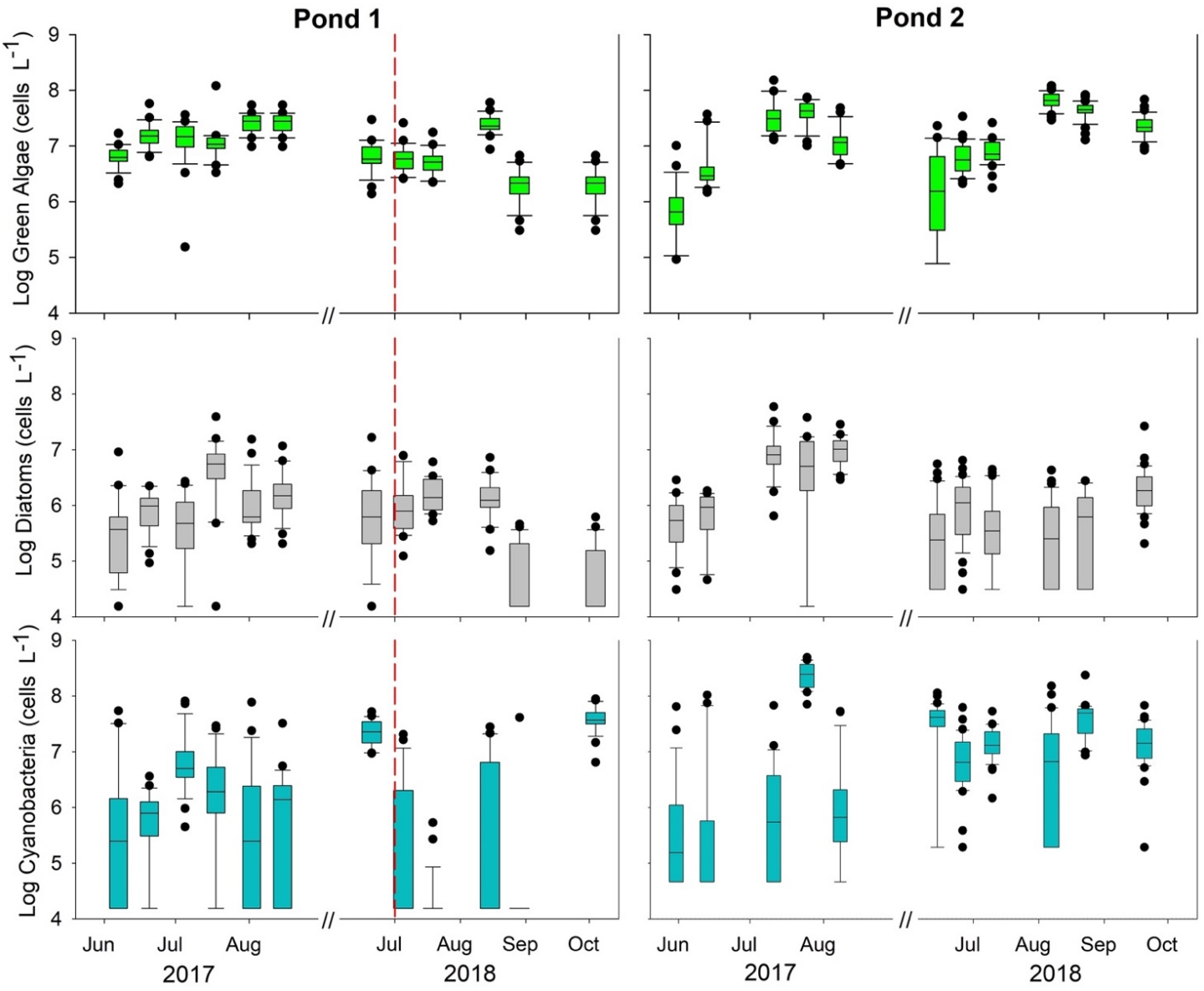


Figure 2.3: Time series data of log concentrations of the major phytoplankton functional groups (green algae, diatoms, and cyanobacteria) for both years. The mid-point line of the box plots represents the median for each date. Outliers (10^{th} and 90^{th} percentiles) are represented by black circles. The red dashed line represents an algicide application that took place on 7/1/18 at Pond 1.

2.3.2. Temporally stable patterns of phytoplankton functional groups

Temporal stability was assessed by considering the standard errors of the mean relative differences for each location. The mean relative differences along with standard error bars are displayed in Appendix A, Supplemental Figure 2.1. Small

standard errors indicate that a location has minimal phytoplankton variation between each sampling date and large standard errors indicate substantial phytoplankton variation between sampling dates. Green algae, diatoms, and cyanobacteria displayed temporally stable spatial patterns in both ponds and over the entire two-year study period.

2.3.2.1. Pond 1

The MRD values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the 2-years of observations at Pond 1 are shown in Appendix A, Supplemental Figure 2.1. Visual representations of the locations with consistently higher and consistently lower concentrations of each phytoplankton functional groups are displayed in Figures 2.4-2.6. The same patterns were observed for all three phytoplankton functional groups in Pond 1. The MRDs of each group tended to be lower for the interior sampling locations, and higher for the nearshore sampling locations. For green algae (Figure 2.4-P1), zones with consistently lower concentrations were all interior locations, except for location 6 where the irrigation pump is located. Zones of high concentrations of green algae were seen at the southern shoreline of the pond (locations 1, 8, 10), as well as locations 5 and 23. The southern shoreline of the pond is very shallow and located adjacent and downhill from crop fields. Location 5 is the site of an inflow pipe which pumps water from a nearby stream-fed pond. Location 23 is a very shallow area with aquatic vegetation and has an inflow from an ephemeral stream. Locations of consistently higher and lower concentrations of diatoms (Figure 2.5-P1) were very similar to those of green algae. Low concentrations of diatoms were exclusively observed at interior sampling

locations. High concentrations of diatoms were observed for locations 1, 5, 6, 23 (previously described), and 2. Locations with consistently higher and lower cyanobacteria concentrations are displayed in Figure 2.6-P1. Low cyanobacteria concentrations were found at all interior sampling locations except for location 20. High cyanobacteria concentrations were found at all nearshore sampling locations except for location 17. Consistently high concentrations of cyanobacteria were also seen close to the ephemeral stream inflow at locations 21 and 22; and at location 1 near the crop fields.

2.3.2.2. Pond 2

The MRD values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the 2-year period for Pond 2 are shown in Appendix A, Supplemental Figure 2.1. Visual representations of MRDs for green algae, diatoms, and cyanobacteria are displayed in Figures 2.4-2.6. Low MRDs for green algae (Figure 2.4-P2) were all nearshore sampling locations of the pond, although there was a somewhat dispersed distribution with similar values not observed within one specific area of the pond. Locations with consistently higher concentrations of green algae were mostly interior sampling locations except for sampling location 21, which is a shallow location with aquatic vegetation present. High concentrations of diatoms were found at nearshore sampling locations and within a small zone on the southeastern shoreline of the pond where there is a water level dependent outflow drain (Figure 2.5-P2). Cyanobacteria MRDs displayed a zonal pattern (Figure 2.6-P2) with higher cyanobacteria concentrations located in the northern portion of the pond apart from the observations at location 21 (as previously described). Consistently low

concentrations of cyanobacteria formed a zone in the middle of the pond containing both interior and nearshore sampling locations.

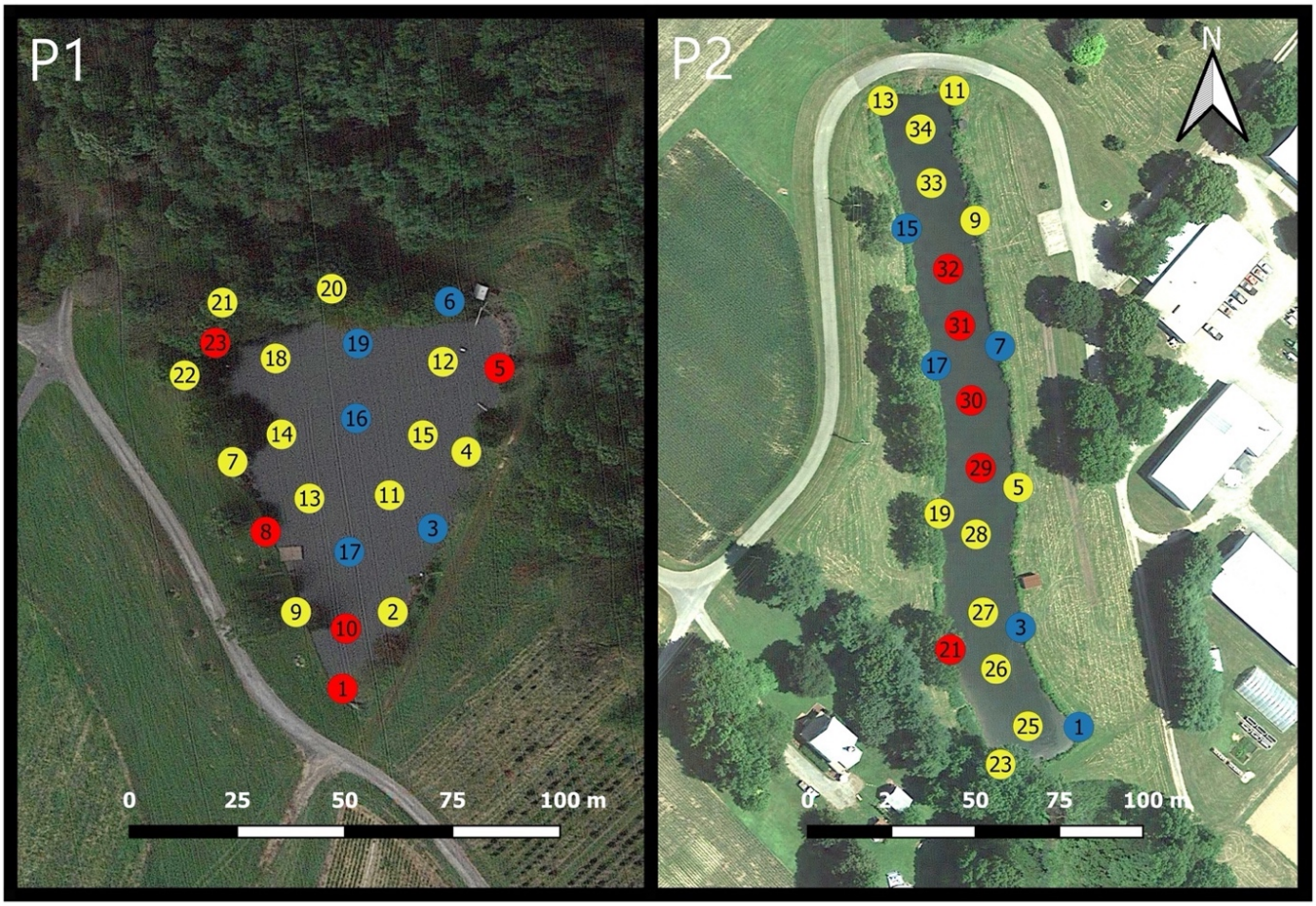


Figure 2.4: Mean relative differences (MRD) of green algae over the two-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red, and locations between the 25th and 75th percentile are displayed in yellow.

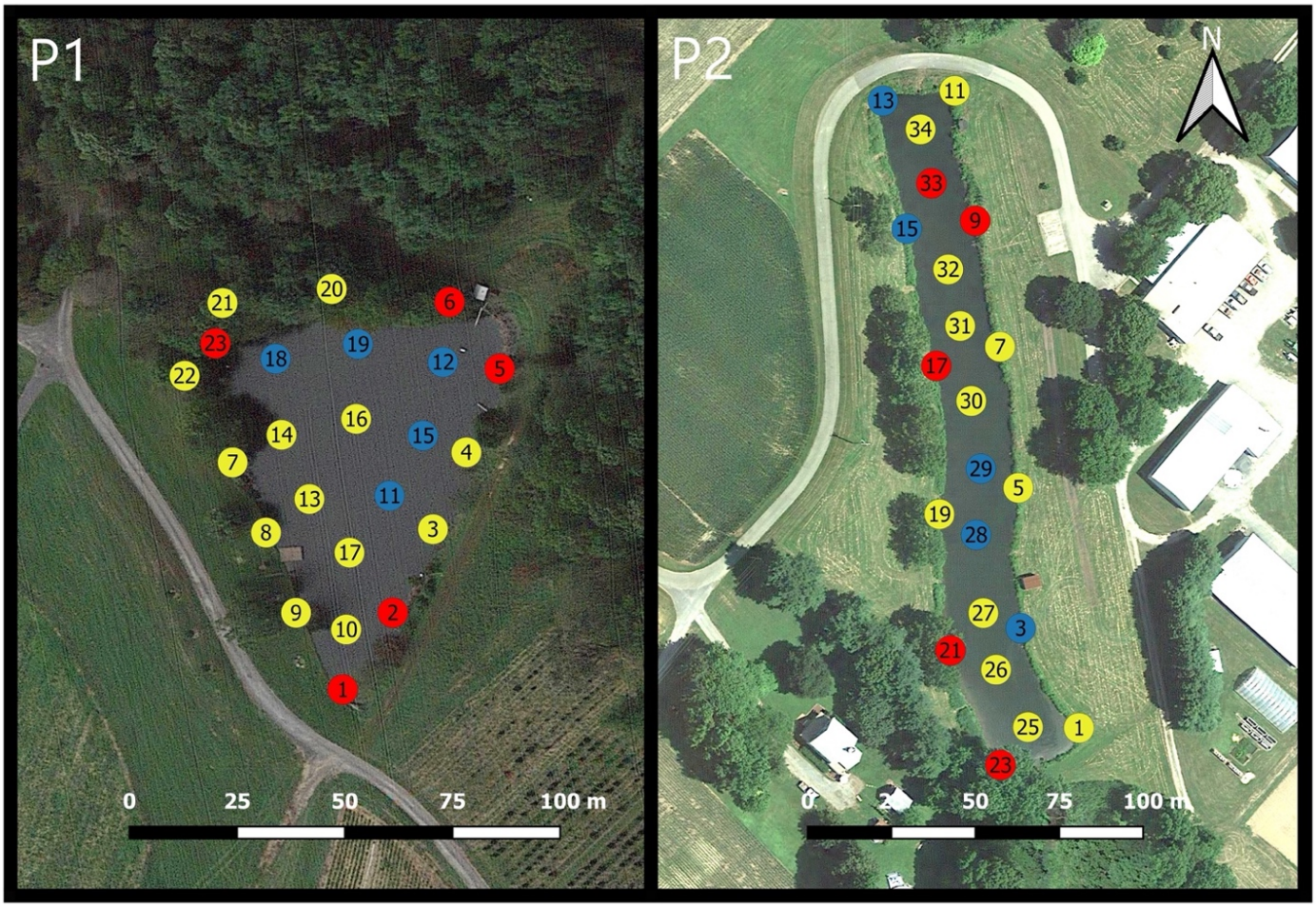


Figure 2.5: Mean relative differences (MRD) of diatoms over the two-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red, and locations between the 25th and 75th percentile are displayed in yellow.

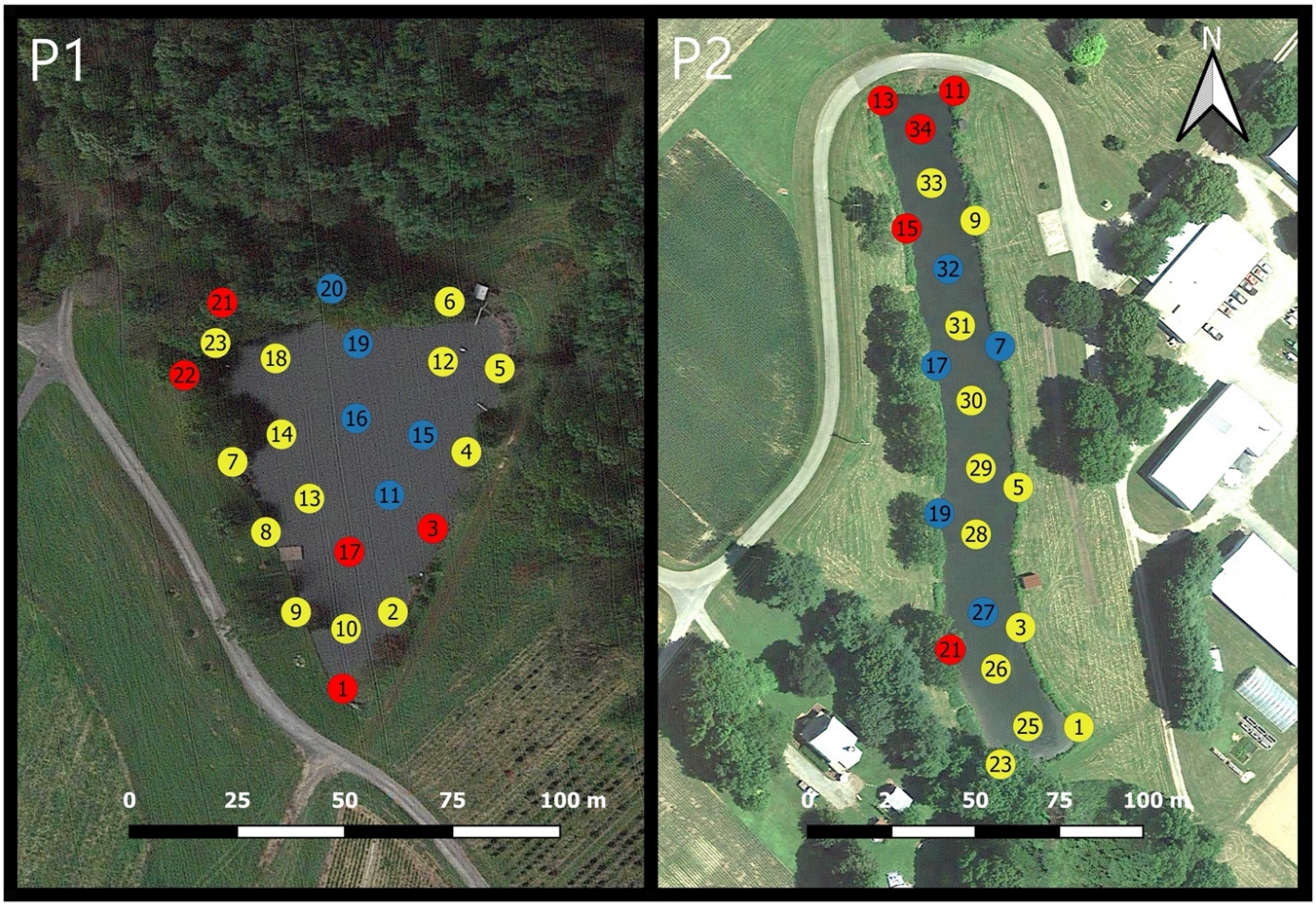


Figure 2.6: Mean relative differences (MRD) of cyanobacteria over the two-year period for Pond 1 (P1) and Pond 2 (P2). Locations with MRD values below the 25th percentile are displayed in blue, locations above the 75th percentile are displayed in red and locations between the 25th and 75th percentile are displayed in yellow.

2.3.3. Water quality patterns

The mean relative difference values of measured water quality parameters (Temp, DO, SPC, pH, NTU, Phyco YSI, CHL YSI, *f*DOM, and CHL EXT) for both ponds are shown in Appendix A, Supplemental Figures 2.3a-i. Within Pond 1, low MRDs were observed for temperature, DO, and pH for nearshore locations. For turbidity, Phyco YSI, CHL YSI, *f*DOM, and extracted chlorophyll, high MRDs were associated with nearshore locations and low MRDs were associated with interior locations within Pond 1. Within Pond 2, similar trends were observed with high MRD

values for temperature, DO, and pH being observed at interior locations and low MRD values found at the nearshore locations. An inverse distribution was seen for turbidity, Phyco YSI, CHL YSI, *f*DOM, and extracted chlorophyll within Pond 2. Low MRD values were typically observed for the interior locations and high MRD values were observed for the nearshore locations. Therefore, both ponds exhibited differences in the water quality parameters between the interior and the nearshore sampling locations.

2.3.4. Phytoplankton and water quality MRD correlations

The Spearman rank correlations between the mean relative differences of the water quality parameters and the mean relative differences of phytoplankton groups are displayed in Table 2.1. Moderate correlations were defined as $r \geq 0.400$ (p values, P1=0.059 P2=0.065) and are highlighted in yellow. Strong correlations were defined as $r \geq 0.600$ (p values, P1<0.001 P2<0.001) and are highlighted in blue. Moderate and strong correlations were observed within Pond 1 for the green algae MRDs and most of the water quality MRDs (DO, SPC, pH, NTU, Phyco YSI, CHL YSI, and CHL EXT). Lower correlations were observed for diatom MRDs and cyanobacteria MRDs within Pond 1. There were no moderate or strong correlations observed for diatom MRDs in Pond 1. The cyanobacteria MRDs were moderately correlated with the MRDs of the SPC, pH, and NTU parameters. Pond 2 differed from Pond 1 regarding MRD correlations. Within Pond 2, green algae MRDs were characterized with fewer moderate correlations than diatom MRDs and cyanobacteria MRDs. Green algae MRDs within Pond 2 had a strong correlation with the MRDs of extracted chlorophyll. Diatom MRDs correlated strongly with most water quality MRDs

(Temp, DO, SPC, pH, and NTU) and moderately with CHL EXT. Cyanobacteria MRDs correlated moderately with Phyco YSI and CHL YSI and strongly with most water quality MRDs (Temp, SPC, pH, and CHL EXT).

Table 2.1. Spearman rank correlations between the mean relative differences of water quality parameters and MRD values of phytoplankton functional groups. Moderate correlations were defined as R≥0.400 and are highlighted in yellow (p values, P1=0.059 P2=0.065). Strong correlations were defined as R≥0.600 and are highlighted in blue (p values, P1<0.001 P2<0.001).

Spearman Rank Correlations Between Water Quality MRDs and Phytoplankton MRDs						
2017 + 2018	**Pond 1**			**Pond 2**		
	Green Algae	**Diatoms**	**Cyanobacteria**	**Green Algae**	**Diatoms**	**Cyanobacteria**
Temp	0.186	<0.001	0.163	<0.001	**0.771**	**0.639**
DO	**0.498**	0.134	0.161	0.024	**0.616**	0.317
SPC	**0.906**	0.142	**0.703**	0.221	**0.713**	**0.842**
pH	**0.895**	0.173	**0.931**	0.004	**0.849**	**0.684**
NTU	0.213	0.104	**0.625**	0.233	**0.903**	0.196
Phyco YSI	0.174	0.003	0.103	0.005	0.387	**0.589**
CHL YSI	0.220	0.001	0.231	0.007	0.297	**0.467**
***f*DOM**	0.017	<0.001	0.029	0.125	0.043	0.390
CHL EXT	**0.668**	0.012	0.033	**0.692**	**0.471**	**0.875**

2.4. Discussion

The abundance and species distribution of phytoplankton taxa has been used as a bioindicator of water quality across freshwater and marine systems for decades (Patrick, 1973; Reynolds et al., 2012; Smith, 2003). Decadal phytoplankton data sets have proven to be useful when examining the seasonal periodicity and long-term trends in coastal water quality (Hernandez Cordero et al., 2020; Marshall et al., 2009; Nishikawa et al., 2010), however, similar longitudinal datasets are lacking for

agricultural irrigation waters despite the fact it has been reported that land-use and nutrient loading can impact phytoplankton biodiversity in agricultural waters (Zhang et al., 2020). Smith et al. (2020) demonstrated that within agricultural irrigation ponds there was a relationship between easily measured environmental co-variates, such as CDOM and NTU, and cyanobacteria (phycocyanin) concentrations. However, the temporal and spatial stability of the cyanobacteria, or other phytoplankton functional groups, in these ponds was not examined. Here, an assessment of the phytoplankton community present during the May to October growing season, when agricultural irrigation water is used most frequently and the risk due to cyanotoxins is greatest, is presented.

The agricultural irrigation ponds examined in this study, located on working farms in Maryland, did not exhibit drastically different phytoplankton populations during the growing seasons of 2017 and 2018. Diatom concentrations did not differ within the two ponds and were comparable with concentrations found in other temperate freshwater lakes and reservoirs (Gorokhova & Zinchenko, 2019; Jia et al., 2019; Rollwagen-Bollens et al., 2013) including lakes studied by Marshall (2013, 2014) in Virginia, located south of this study area. Mean concentrations of green algae were similar within the two ponds, but Pond 1 had a smaller overall range of concentrations than Pond 2. Concentrations of green algae were comparable to values reported for other freshwater systems (Dembowska et al., 2018; Gorokhova & Zinchenko, 2019; Khaliullina & Fazlieva, 2019). Pond 2 had slightly higher concentrations of green algae and cyanobacteria. These higher values may potentially be explained by the absence of an algicide application for Pond 2. Concentrations of

cyanobacteria within Pond 1 were similar to those previously reported within temperate lakes (Dembowska et al., 2018; Jia et al., 2019), including those studied locally by Marshall (2013, 2014). However, due to recurrent cyanobacteria blooms composed mainly of *Aphanizomenon* spp. and *Microcystis wesenbergii* in Pond 2, cell concentrations were comparable with concentrations reported within small temperate lakes in which cyanobacteria blooms frequently occur (Lee et al., 2015; Woodhouse et al., 2016), including other Maryland lakes (Tango & Butler, 2008; J. Wolny, unpublished data), but were greater than those recorded by Marshall (2013, 2014).

Both ponds displayed spatial and temporal variability of major phytoplankton functional groups during the two growing seasons. Spatio-temporal variations of phytoplankton communities have been documented within freshwater lakes (Naselli-Flores & Padisák, 2016; Wu et al., 2014; Xiao et al., 2018), wetland ponds (Soininen et al., 2007), reservoirs (Alexander & Imberger, 2009), rivers (Marshall et al., 2009), and estuaries (Marshall et al., 2006). While the phytoplankton community temporal trends noted in this study were similar to those reported by Marshall (2013, 2014) for Virginia lakes, comparisons between these earlier studies and spatial variation are not possible due to the limited spatial variance in the Virginia lakes dataset. The heterogeneity or homogeneity of phytoplankton communities should not be an assumed trait within a waterbody. As explained by Lewis (1978), not all species or groups of phytoplankton continuously exhibit heterogenous distributions, but rather homogenous and heterogenous distributions may synchronously exist within a waterbody. Additionally, exogenous forces, such as wind, water flow, and lake

morphology, have all been documented to attribute to the spatial variation of green algae, diatoms, and cyanobacteria (Li et al., 2013).

Two forms of spatial trends were observed within the two ponds during this study. The predominant spatio-temporal trends that were present within both ponds appeared to be a contrast between interior and nearshore sampling locations. Within Pond 1, this trend was displayed for all groups. Pond 1 had consistently higher concentrations of green algae, cyanobacteria, and diatoms at nearshore sampling locations; and consistently lower concentrations of green algae, cyanobacteria, and diatoms at interior locations. This pattern of higher concentrations at nearshore sampling locations versus interior sampling locations was reported for both ponds in a preceding study using average quartile ranks of phycocyanin concentrations (Smith et al., 2020). Higher concentrations of phytoplankton being closer to the shoreline of shallow waterbodies has been attributed to several different concepts. Bondarenko et al. (1996) stated that spatial distribution of phytoplankton was related to water depth, with shallow waters being richest in phytoplankton. Both ponds in this study were only 2.7 m deep, thus indicating that even in shallow environments depth-dependent gradients can be set up within the phytoplankton community. In other studies, greater abundances of phytoplankton were found in stands of *Phragmites australis* and other aquatic plants, due to the creation of favorable water quality conditions, including increased phosphorus concentrations (Celewicz-Goldyn & Kuczynska-Kippen, 2008, 2017) and in zones with elevated nutrient concentrations and water temperatures (Chen et al., 2003). Aquatic vegetation was noted at both ponds and future work will

look to correlate the spatial patterns of the phytoplankton community with the characteristics of the resident aquatic vegetation.

Similarly, within Pond 2 consistently higher concentrations of diatoms were found at nearshore locations. The opposite trend was observed for green algae within Pond 2 wherein consistently higher concentrations of green algae were observed for interior sampling locations, and consistently lower concentrations of green algae were observed for nearshore sampling. While this is a difference from co-located Pond 1, this trend has been previously documented by Celewicz-Goldyn & Kuczynska-Kippen (2008) who indicated that the greatest abundance of small chlorophytes was found in open waters where the potential threat of predation from zooplankton was less.

The major spatiotemporal patterns observed for Pond 2 were the formation of zones in which cyanobacteria were the dominant taxa. Consistently higher concentrations of cyanobacteria were found at the northern portion of the pond (sampling locations: 11, 13, 15, and 34) near a culvert, which following precipitation events provides inflow of potentially nutrient-rich waters to the pond. This high cyanobacteria biomass zone was also established in the preceding study on quartile ranks of phycocyanin concentrations (Smith et al., 2020). Other studies have documented spatial trends of cyanobacteria among other phytoplankton species due to either nutrient rich runoff or river inflow (Marshall et al., 2006; Powell et al., 1975; Woodhouse et al., 2016). Other potential explanations for the formation of these cyanobacteria-rich zones could be wind or wind driven water flow as noted by Cloern

et al. (1992) and Fragoso et al. (2008) or microhabitats set up through thermal stratification as noted by Vasas et al. (2013)

There was a zone of consistently low cyanobacteria concentrations within Pond 2 that was located near the middle of the pond (sampling locations: 17, 27, 32, 7, 19, 5). The pump house and water intake pipe for the farm irrigation system is in this area. The location of the irrigation intake pipe has important implications for food safety. It has been well-established that irrigation waters with toxigenic cyanobacteria can contaminate crops (Miller & Russell, 2017), remain in soils for extended periods of time (Machado et al., 2017), and may even be taken up by the root system of the produce (Lee et al., 2017). Thus, placing an irrigation intake system in a location with consistently higher concentrations of cyanobacteria may increase produce contamination risks. It appears that the pump and intake infrastructure in Pond 2 is located in a low-risk zone, as cell concentrations of potentially toxic cyanobacteria species never exceeded EPA or regional guidelines (Environmental Protection Agency, 2019; Virginia Department of Health, 2015). However, future research and monitoring efforts should focus on determining the prevalence of cyanotoxins in these irrigation waters.

Of note, a copper sulfate algicide was applied to Pond 1 midway through the study, on 7/1/18, to mitigate a bloom of *Microcystis*, a potentially toxigenic cyanobacteria species. While the concentration of copper sulfate used is unknown, all measured water quality parameters decreased significantly following this application. These reductions were comparable to values reported by Schrader et al. (2000) and Song et al. (2011) within other inland waters that were assessed during and after

treatments with copper sulfate. Average concentrations of DO, pH, CDOM, and *f*DOM returned to pre-application levels about one month after application. For the algal pigments (CHL RFU, CHL YSI, CHL EXT, and phycocyanin), all concentrations decreased after the copper sulfate application and slowly recovered to either pre-application levels or higher by the end of August. The return of chlorophyll-*a* readings to previous values was also reported by Dia (2016) and Effler et al. (1980) following low-level algicide treatments (8-14 µg L^{-1}) in freshwater lakes. Elder & Horne (1978) reported the recovery of pre-treatment algal populations in as little as five days after treatment and attributed this recovery to copper sulfate possibly being beneficial for biological activity if applied in very low concentrations (5-10 µg L^{-1}). The effect of algicide on cyanobacteria concentrations was more pronounced than the effect on green algae and diatoms concentrations. Similar responses were reported by Padovesi-Fonseca & Philomeno (2004) and XiaoLi et al. (2009) wherein cyanobacteria concentrations, including *Microcystis aeruginosa, Cylindrospermopsis raciborskii,* and *Anabaena flos-aquae,* decreased and green algae and diatoms became the dominant taxa after an algicide application, with the subsequent population changes attributed to cyanobacteria species sensitivity to copper. The rate of recovery of the phytoplankton community to the application of copper sulfate, or other algicides, should be monitored if the algicide application is meant to act as a safeguard to crops from cyanotoxin exposure via irrigation waters. It should be mentioned that the treatment of waterbodies with algicides can immediately release large quantities of cyanotoxins, if toxin-producing algal species are highly concentrated (Greenfield et al., 2014; Zhou et al., 2013). An assessment of

phytoplankton community composition should be performed prior to an algicide application if the water is to be used for crop irrigation or as drinking water for livestock to safeguard against the introduction of concentrated biotoxins.

Although not similar, both ponds expressed moderate and strong correlations between the spatial patterns of phytoplankton functional groups and water quality parameters. Pond 1 had strong water quality correlations with green algae, while Pond 2 had strong correlations with diatoms and cyanobacteria. The correlations between water quality and phytoplankton spatial trends provides helpful insights for irrigation pond monitoring. The examination of phytoplankton community structure using microscopy is an intensive analysis which requires extensive laboratory infrastructure and highly trained personnel (Lawton et al., 1999). However, if strong correlations exist among water quality parameters and optical properties associated with distinct phytoplankton functional groups, the option of using less specialized monitoring methods, such as in-situ sensors or drone-based imagery could be employed for routine resource management. These technologies would be efficient and cost-effective methods capable of being used by a broader group of personnel to safeguard against irrigating crops with degraded water drawn from agricultural irrigation waters. While identifying water quality covariates and the use of optical techniques to assess the phytoplankton functional groups present in water will not identify toxic versus non-toxic phytoplankton species it can provide the information necessary to make better informed decisions about when and where to conduct toxin risk assessments.

2.5. Conclusions

Using a mean relative difference analysis to assess spatiotemporal stability, it was determined that phytoplankton functional groups exhibited stable spatiotemporal trends in the two agricultural irrigation ponds evaluated in this study. Temporally stable spatial patterns of the three phytoplankton functional groups studied here were found within both ponds over the two sampling years. Both ponds had locations where phytoplankton group concentrations were consistently higher or lower than the pond's average concentrations. Typically, these patterns could be classified into two categories: nearshore or interior sampling locations or zones. These distributions indicate the importance of sampling locations for water quality monitoring purposes. If sampling is performed in areas of consistently higher or lower concentrations of phytoplankton, that sample may not be an accurate representation of the phytoplankton community within the entire waterbody. Because of the correlation between water quality parameters and certain phytoplankton functional groups it may be possible to employ broad-based technologies to routinely monitor irrigation waters for potentially harmful cyanobacteria instead of relying on labor intensive microscopy methods. However, it is important to note that there are other types of agricultural ponds, such as aquaculture ponds and retention ponds. While this study can provide a framework for assessing agricultural ponds no extrapolation should be made to these other water sources from the finding presented here.

Chapter 3: Intraseasonal variation of phycocyanin concentrations and environmental covariates in two agricultural irrigation ponds in Maryland, USA

3.1 Introduction

Cyanobacteria, also referred to as blue-green algae, are common in many types of ecosystems and are particularly prevalent in freshwaters. These microorganisms can multiply quickly and create large proliferations or "blooms" on the water surface or along the benthos (Bouma-Gregson et al., 2017; Chorus & Bartram, 1999; Zanchett & Oliveira-Filho, 2013). Cyanobacteria blooms tend to appear in low-flow or stagnant waterbodies that are nutrient enriched (Lee et al., 2015; O'Neil et al., 2012; Steinberg & Hartmann, 1988) and therefore, the presence of large cyanobacteria populations may function as an indicator of poor or degraded water quality (Douterelo et al., 2004; Perona et al., 1998; Teta et al., 2017). In response to excess nutrient concentrations, dense cyanobacteria blooms can form throughout the water column. These blooms can reduce dissolved oxygen concentrations, create hypoxic conditions, and reduce light penetration to depths below the surface layer, all which can inhibit the growth of submerged aquatic vegetation, impair critical fisheries habitat, and reduce aquatic community diversity (Aboal et al., 2000; Huisman et al., 2018; Paerl & Otten, 2013; Tourville-Poirier et al., 2010).

Cyanobacteria blooms are also called harmful algal blooms (cyanoHABs) because they may present risks to both environmental and human health. Certain species of cyanobacteria have the capacity to produce substances that are toxic and often these toxins are harmful to humans and livestock through acute exposure and

bioaccumulation (summarized in Hilborn & Beasley [2015] and Wood [2016]). Exposure to cyanotoxins can occur through ingestion, inhalation, or direct contact. The three most problematic cyanotoxins from a resource management standpoint are microcystin and cylindrospermopsin, which affect the liver, and anatoxin-a, which affects the nervous system (Carmichael, 1994, 2001; Environmental Protection Agency, 2019).

Recently, it has been determined that cyanotoxins present in waters used for irrigation can be transported to produce or other plants (Corbel et al., 2016; Saqrane & Oudra, 2009). Bittencourt-Oliveira et al. (2016), Buratti et al. (2017), Corbel et al. (2014), Kittler et al. (2012) and Lee et al. (2017) have documented cyanotoxin bioaccumulation in crops, produce, and agricultural soils when contaminated waters are used as irrigation source water. Consumption of foods irrigated with cyanotoxin-containing water may present severe risks to human and animal health (Saqrane & Oudra, 2009; Svirčev et al., 2017).

Traditionally, cyanobacteria have been identified and enumerated using microscope-based analyses, but this process can be time-consuming, and the accuracy is dependent on the skill level of analysts (Lawton et al., 1999). The concentration of phycocyanin, a pigment unique to cyanobacteria, may be used as a surrogate measurement for the presence and abundance of cyanobacteria, including toxigenic species. Phycocyanin can be measured with extraction methods that require intensive laboratory work (Horváth et al., 2013; Sarada et al., 1999; Silveira et al., 2007) or via fluorometry performed with either a benchtop fluorometer or an in-situ water quality probe, the latter providing near instantaneous results (Brient et al., 2007; Kasinak et

al., 2015). However, the detection and quantification of phycocyanin may be strongly affected by other water quality parameters. Determination of the relationship between phycocyanin concentrations and commonly or more readily evaluated water quality parameters may allow for the rapid assessment of cyanobacteria concentrations, and by extension, the potential for the presence of cyanotoxins in the water. Currently, it is not well-known which water quality parameters have the strongest effect on phycocyanin concentrations in agricultural irrigation ponds.

Regression trees are an artificial intelligence tool used to search for and describe complex relationships between variables within data sets (Breiman, 2017; Loh, 2011). The regression tree algorithm splits the dataset into subsets that are homogeneous regarding the output variable and are separated from each other as much as possible with respect to the output variable. The first split divides the dataset into two groups. After this, those two groups can be further divided into two more splits, and so forth. If a split cannot be completed, the split will end with a "leaf" which contains the average value of the output variable. The use of this technology to elucidate patterns in non-linear relationships in aquatic systems has been summarized by Chau (2006) and Quetglas et al. (2011). More recently, Harris & Graham, (2017) and Millie et al. (2014) have employed machine learning to monitor and predict harmful cyanobacteria blooms in large bodies of water used for recreational activities and as drinking water reservoirs. The use of similar technology in agricultural water systems, however, is untested. The objectives of this work were to (1) identify the most influential environmental covariates affecting the phycocyanin concentrations in two Maryland agricultural irrigation ponds by using the machine learning

methodology of decision trees and (2) to analyze the environmental covariates to create quartile ranks with respect to phycocyanin concentrations and therefore identify water quality parameters that are suitable for use in monitoring phycocyanin concentrations.

3.2 Materials & Methods

3.2.1 Pond Monitoring

Sampling was conducted at two irrigation ponds located on working farms in Maryland. These ponds, referred to as Pond 1 and Pond 2, were chosen because water was routinely drawn to irrigate fruit and vegetable crops. Sampling of these ponds was conducted during the growing season which was between June and October 2018. Each pond was sampled six times during this period and since both ponds were located within a 1-hour drive from the laboratory water quality samples were processed the same day as collection.

Pond 1 is a 1.01-acre man-made embankment pond with an average depth of 2.7 m located near Germantown, MD. There is vegetation consisting of deciduous trees and shrubs along the northern and eastern banks and the remaining banks have grass cover. The pond is surrounded by crop fields. When the water level in this pond gets low, farm operators pump in water from an adjacent creek-fed pond. The inflow and outflows are both located near sampling location 15 and the irrigation pump intake is located near sampling location 12 (Appendix B, Supplemental Figure 3.1A).

Pond 2 is located at the University of Maryland Wye Research Center in Wye Mills, MD. This pond is a 1.05-acre, excavated pond with an average depth of 2.7 m and most of the bank is covered with grass and dense shrubs. Large trees are also

present along the perimeter but are not close to the water. This pond is surrounded by crop fields, farm buildings, and one residential property. In March, the surrounding fields receive chemical fertilizer, but no animal manures are used. This pond is primarily fed through rainfall which enters through an ephemeral creek that leads into a culvert near sampling location 12 (Appendix B, Supplemental Figure 3.1B). This culvert tends to have a substantial inflow only when precipitation has recently occurred. On the south end of the pond there is a water-level dependent outflow drain located near sampling location 24 and the irrigation pump intake is located near sampling location 27 (Appendix B, Supplemental Figure 3.1B).

Sampling locations were the same for every sampling date. Pond 1 had 23 sampling locations and Pond 2 had 34 sampling locations (Appendix B, Supplemental Figures 3.1A and 3.1B, respectively). Water quality and phytoplankton community samples were always taken at a depth of 0 - 15 cm. Nearshore samples were taken with a 500-mL hand grabber (Dynalon, Rochester, NY, USA) at approximately 1.5 m from the shoreline. Interior samples were taken from a small boat and GPS was used to provide consistency of sampling locations between different sampling dates. After collection, samples were immediately placed into a cooler and transported to the lab for analysis. Water quality variables were measured in-situ with a YSI Exo-2 sonde (YSI Inc., Yellow Springs, OH, USA) at the same time and place the water samples were collected. The YSI sonde measured temperature (°C), dissolved oxygen (DO mg L^{-1}), pH, fluorescent dissolved organic matter (*f*DOM measured in relative fluorescent units, RFU), chlorophyll-*a* content (RFU), and turbidity (NTU). The YSI sonde was calibrated prior to each sampling date according to the manufacturer's

recommendations in the support manual. Precipitation data was collected at a weather station < 1 km away from each pond. Both weather stations are owned and operated by the University of Maryland. At the time of collection phytoplankton community samples were fixed with 5% unacidified Lugol's iodine solution in a ratio of 0.5 mL of fixative to 50 mL of sample. Phytoplankton samples were stored in the dark at 4 °C until analysis.

3.2.2 Laboratory Analyses

All water quality samples were analyzed using Aquaflor fluorometers (Turner Designs, San Jose, CA, USA) for colored dissolved organic matter (CDOM, µg L^{-1}), in-vivo or whole-cell chlorophyll-*a* (CHL RFU, RFU), and phycocyanin (Phyco, µg L^{-1}) prior to extraction. Samples were also processed for extracted chlorophyll-*a* (CHL EXT, µg L^{-1}) according to EPA method 445 (EPA, 1997) and measured using an Aquaflor fluorometer. For the extraction process, approximately 100 mL of pond water was vacuum filtered onto 0.7 µm glass fiber filters (Whatman, Maidstone, UK) and steeped in a 90% acetone and 10% deionized water solution overnight before being analyzed on the fluorometer. The CDOM and extracted chlorophyll channels were calibrated using standards provided by TurnerDesigns (San Jose, CA, USA). The phycocyanin channel was calibrated with PB-11 by PROzyme (Hayward, CA, USA).

Phytoplankton community samples were analyzed using a Nikon Eclipse T*s*2R inverted light microscope (Nikon Instruments, Melville, NY, USA) using a modified Utermöhl method as described in Marshall and Alden (1990). Phytoplankton taxa were identified morphologically to the lowest taxon possible

using John et al. (2002). For comparison to the chlorophyll and phycocyanin data, taxon data was summed into phytoplankton functional groups including cyanobacteria, chlorophytes, diatoms, dinoflagellates, and uncharacterized flagellates. Cyanobacteria were further categorized within the taxonomic orders Chroococcales, Oscillatoriales, and Nostocales as defined by Komárek et al. (2014).

3.2.3 Data Analysis with Regression Trees

Regression trees were created using RStudio (RStudio team, Boston, MA, USA) with the rpart.plot package. Each sampling date was analyzed independently with the regression tree analysis. A text file containing all parameters measured for that sampling date was uploaded into RStudio. This text file was then processed with the rpart.plot package to produce a corresponding phycocyanin regression tree for that date. Only the first three splits were taken into consideration to determine the primary predictors.

Each split was made according to a condition. If a condition was true for a dataset, this dataset was included in the left branch after the split. If a condition was not true for a specific dataset, then the dataset was included in the right branch. For example, if pH was < 7, then all datasets with values of $pH < 7$ would form the left branch and all data points with pH values ≥ 7 would form the right branch. The most influential inputs were the variables in the conditional statements in primary splits at the top of the regression tree.

3.2.4 Ranking sampling locations by the phycocyanin concentrations over the observation period

The phycocyanin concentrations per se were not suitable for ranking sampling locations over the observation period because at different days concentrations across the entire pond could be higher or lower than on other dates. Therefore, comparing concentrations at different days for the same location did not inform about the rank of location relative to other locations. To overcome this, we employed the ranking method demonstrated in Stocker et al. (2019) which allowed us to determine the cumulative probability distribution functions for phycocyanin concentrations on each sampling date and the quartile that each sampling location belonged to. This provided six quartile numbers for each sampling location, one for each observation date which were then averaged and ranked. For example, if a location contained phycocyanin concentrations mostly in the first and second quartile, its rank would be between 1 and 2. This location would have mostly low concentrations of phycocyanin. Conversely, if the phycocyanin concentrations for a location were found mostly in the third or fourth quartiles, then the rank of this location would be between 3 and 4 and would contain mostly high concentrations of phycocyanin.

3.2.5 Statistics and Graphics

Correlations were computed in RStudio. Correlations were considered significant if the associated p values were < 0.05. The Kolmogorov-Smirnov non-parametric test was used to compare the overall equal distribution of interior and nearshore samples and to find the probability of median concentrations being similar.

Sigmaplot v. 13 (SYSTAT, Chicago, IL, USA) and ArcGIS Pro v2.31 (ESRI, Redlands, CA, USA) were used to create visual representations of the data.

3.3 Results

3.3.1 Data Summary

Precipitation occurred prior to Pond 1 sampling dates 6/20/18 (7.6mm), 7/19/18 (18.5mm), and 8/15/19 (2.5mm) and Pond 2 sampling dates 6/26/18 (22.9), 8/23/18 (9.9mm) and 9/20/18 (13.5mm). All other sampling dates were preceded by a minimum of 48 hours of baseflow conditions. Information on the number of days from last rainfall and rain accumulations can be seen in Appendix B, Supplemental Table 3.1. Phycocyanin concentrations generally were between 20 and 50 μg L^{-1} at Pond 1, except for 6/20/18, when concentrations were 114 μg L^{-1} (Figure 3.1). Increased phycocyanin concentrations on this date were due to a cyanobacteria bloom that made up 72.04% of the phytoplankton community (Appendix B, Supplemental Table 3.2). In Pond 1, sampling dates after 6/20/18 indicated that cyanobacteria made up < 20% of the phytoplankton community composition, except on 10/4/18 when cyanobacteria comprised 94.70% of the phytoplankton community. Phycocyanin concentrations were higher throughout the summer at Pond 2 (70 to 170 μg L^{-1}) and displayed an increasing trend over the sampling season. The phytoplankton community in Pond 2 was dominated by cyanobacteria most of the sampling season with all but two dates having a community comprised of > 40% cyanobacteria (Appendix B, Supplemental Table 3.2). Time series data of phycocyanin concentrations and other water quality parameters are shown in Appendix B, Supplemental Table 3.3. Mean values and variability of all measured parameters were

generally higher at Pond 2 than at Pond 1. Pond 1 was treated with the algicide copper sulfate on 7/1/18 which resulted in a substantial decrease in mean phycocyanin concentrations when routine sampling occurring on 7/5/18. Following this treatment, the chlorophyll-*a*, turbidity, CDOM, *f*DOM, and phycocyanin concentrations drastically decreased (Appendix B, Supplemental Table 3.3). The microscopy analysis of the phytoplankton sample collected on 7/5/18 also indicated that phytoplankton cell concentrations had severely declined after the algicide application (data not shown). Measurements returned to pre-treated levels by 8/29/18. The last sampling date for Pond 1 was on 10/4/18 and by this date most chlorophyll-*a* concentrations dropped drastically with the change of season.

Consistent trends in water quality variables from date to date were generally not present in Pond 2. The only exceptions to this were turbidity and chlorophyll-*a* concentrations measured with the YSI sonde. At Pond 2, turbidity generally increased throughout the summer while the chlorophyll-*a* concentrations measured with the YSI sonde decreased throughout the 6 sampling dates. There was a substantial drop in *f*DOM, CDOM, and all three chlorophyll-*a* concentrations on 7/10/18 at Pond 2. Most of these concentrations returned to prior ranges by the next sampling date.

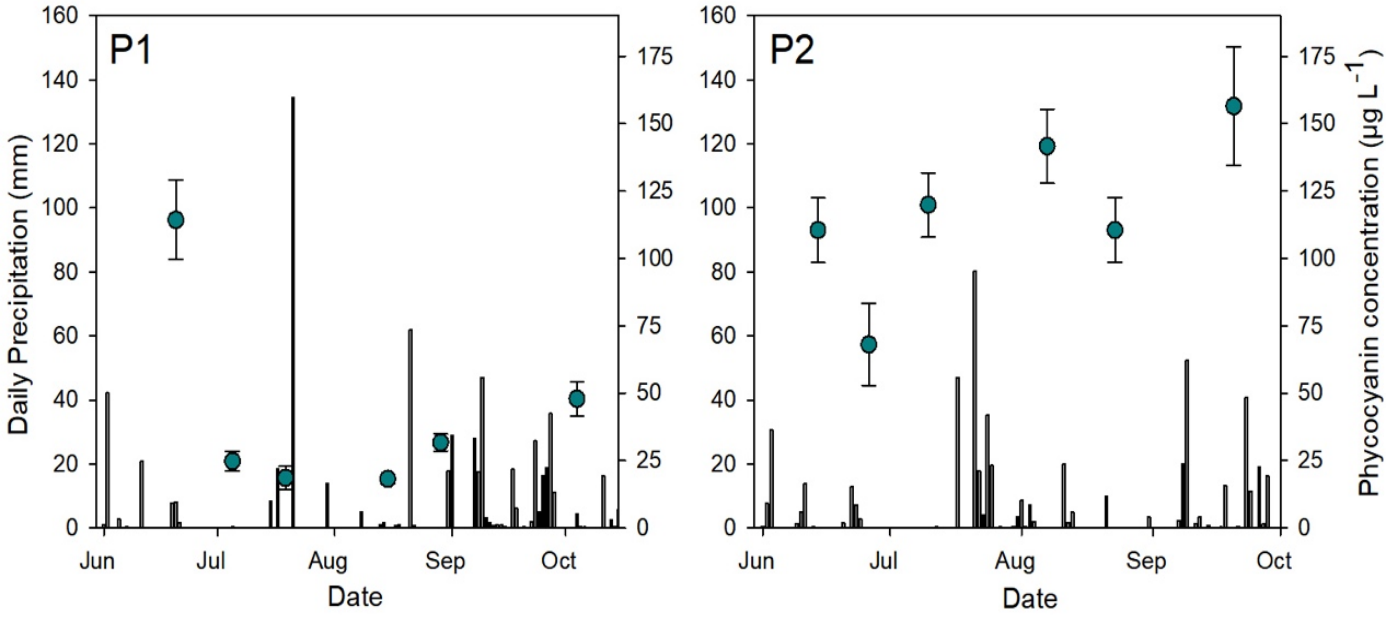


Figure 3.1. Precipitation data and phycocyanin concentrations for both ponds. Precipitation data is represented in bars and phycocyanin concentrations are shown with symbols.

3.3.2 Correlations

The strength of the relationships between water quality parameters was assessed by computing Pearson's correlation coefficients (r). Correlations were considered significant if the associated p values were < 0.05. The correlations determined for both ponds are reported in Appendix B, Supplemental Table 3.4. Dissolved oxygen concentrations and pH had significant correlations in both ponds with r values ranging from 0.448 to 0.811 in Pond 1 and from 0.868 to 0.970 in Pond 2. In Pond 1, phycocyanin concentrations had a moderate positive correlation with CDOM and CHL EXT concentrations with r values ranging from 0.519 to 0.761 and 0.525 to 0.900, respectively. Other significant correlations observed in Pond 1 were between concentrations of CHL RFU and CHL EXT (r values from 0.525 to 0.910), CDOM and CHL EXT (r values from 0.619 to 0.819), and CHL RFU and CDOM (r

values from 0.462 to 0.915). There was also a positive relationship between turbidity and the CHL YSI concentration (r = 0.761) on 8/15/2018. Moderate to strong negative correlations were seen in Pond 1 between the concentrations of *f*DOM and DO (r values from -0.473 to -0.625), pH (r values from -0.435 to -0.564), and NTU (r value of -0.701). Negative relationships were also observed between concentrations of DO and CDOM (r values from -0.568 to -0.755), CHL EXT (r values from -0.480 to -0.483), and phycocyanin (r value of -0.509).

In Pond 2, phycocyanin concentrations displayed both positive and negative correlations with concentrations of CHL YSI (r values of -0.381 to 0.916), CDOM (r values from -0.616 to 0.991), CHL RFU (r values from -0.598 to 0.993), and CHL EXT (r values from -0.493 to 0.962). The majority of correlations between phycocyanin concentrations and CHL YSI, CDOM, and CHL RFU concentrations were significant and positive in Pond 2; however, on the 8/29/18, there was a significant negative correlation. Extracted chlorophyll-*a* (CHL EXT) concentration had moderate to strong positive correlations with turbidity (r value from 0.389 to 0.836), CHL YSI (r values from 0.365 to 0.900), CDOM (r values from 0.575 to 0.963), and CHL RFU (r values from 0.435 to 0.968) concentrations. The concentrations of CHL RFU also had strong positive correlations with CHL YSI (r values from 0.460 to 0.969) and CDOM (r values from 0.636 to 0.998) concentrations, as well as moderate positive correlations with dissolved oxygen (r values from 0.341 to 0.528) and turbidity (r values from 0.414 to 0.789). Concentrations of CDOM displayed moderate correlations with concentrations of DO (r values from 0.355 to 0.569) and turbidity (r values from 0.427 to 0.808) and had a

strong correlation with CHL YSI (r values from 0.578 to 0.970) concentrations. The concentration of *f*DOM was negatively correlated with all parameters in Pond 2 and had significant r values ranging from -0.343 to -0.891.

3.3.3 Regression Tree Analysis

Regression trees relating phycocyanin concentrations to environmental covariates at each pond and each observation date are shown in Appendix B, Supplemental Figure 3.2. Table 3.1 presents the environmental covariates that provided the first two splits and therefore were the most influential predictors according to the regression tree algorithm. The root mean square error (RMSE) and r^2 values for the regression trees can be seen in Appendix B, Supplemental Table 3.5. The most frequent influential variables were concentrations of CDOM, CHL EXT, and turbidity (NTU). The concentration of CDOM was found to be a leading predictor 32% of the time, CHL EXT 28% of the time, and turbidity 24% of the time. Concentrations of DO, *f*DOM, and pH were found to be the most influential predictor in fewer cases (16%). Inspection of the regression trees (Appendix B, Supplemental Figure 3.2) shows that after a split, lower phycocyanin concentrations were found in a smaller range of values of the splitting environmental covariate for CDOM, extracted chlorophyll-*a* and turbidity, except on two occasions. The exceptions, 7/19/18 and 8/23/18, had larger turbidity values yielding smaller phycocyanin concentrations in Pond 1 and Pond 2, respectively. Dissolved oxygen concentrations only showed up as a predictor once in each pond and higher dissolved oxygen concentrations corresponded to lower phycocyanin concentrations. On 7/19/18 the phytoplankton community was comprised mostly of non-cyanobacteria algae with only 0.5% of the

community being cyanobacteria species. Similarly, on 8/23/18, less than 50 % of the phytoplankton community was comprised of cyanobacteria. Overall, CDOM, CHL EXT, and NTU concentrations were found as primary predictors and lower concentrations or readings of these predictors typically resulted in lower phycocyanin concentrations for both ponds.

Table 3.1: Most influential water quality variables, as identified by regression tree analyses, and effect on phycocyanin concentrations for Pond 1 and Pond 2.

Date:	First Split		Second split to the Left		Second split to the right	
	Split variable	Phycocyanin is smaller if	Split variable	Phycocyanin is smaller if	Split variable	Phycocyanin is smaller if
Pond 1						
06/20/18	CHL EXT	CHL EXT is smaller	CHL EXT	CHL EXT is smaller	NONE	NONE
07/05/18	CDOM	CDOM is smaller	CDOM	CDOM is smaller	NONE	NONE
07/19/18	NTU	NTU is larger	DO	DO is larger	NONE	NONE
08/15/18	CHL EXT	CHL EXT is smaller	NTU	NTU is smaller	NONE	NONE
08/29/18	CHL EXT	CHL EXT is smaller	CDOM	CDOM is smaller	NONE	NONE
10/04/18	CHL EXT	CHL EXT is smaller	NTU	NTU is smaller	NONE	NONE
Pond 2						
06/14/18	CDOM	CDOM is smaller	pH	pH is larger	NONE	NONE
06/26/18	CDOM	CDOM is smaller	CDOM	CDOM is smaller	NONE	NONE
07/10/18	CDOM	CDOM is smaller	NTU	NTU is smaller	NONE	NONE
08/07/18	CHL EXT	CHL EXT is smaller	CDOM	CDOM is smaller	*f*DOM	CDOM is smaller
08/23/18	NTU	NTU is larger	DO	DO is larger	NONE	NONE
09/20/18	NTU	NTU is smaller	CHL EXT	CHL EXT is smaller	NONE	NONE

DO - dissolved oxygen (mg L^{-1}), NTU - turbidity (NTU), CDOM - colored dissolved organic matter (µg L^{-1}), CHL EXT – extracted chlorophyll-*a* concentrations (µg L^{-1}).

3.3.4 Average Ranks

Locations with quartile ranks of 1-2 represented locations of generally lower phycocyanin concentrations throughout the summer and locations with ranks of 3-4 represented locations which harbored higher concentrations of phycocyanin concentrations throughout the summer. In Pond 1 (Figure 3.2A), sampling locations falling within the 3rd and 4th quartiles were generally shoreline or near shore locations. Conversely, interior sampling locations tended to fall into the 1st and 2nd quartiles. In Pond 1, the average quartile for interior sampling locations was 2.2 and the average quartile for shoreline sampling locations was 2.7. In Pond 2 (Figure 3.2B), higher quartiles were seen at the northern area of the pond where the inflow culvert is located and lower quartiles towards the southern end of the pond near the outflow. In Pond 2, almost all interior samples fell within the 1st and 2nd quartile, except for sampling points 33 and 34 which were near the northern portion of the pond. In Pond 2, the average quartile for interior locations was 2.0 and the average quartile for nearshore locations was 2.7.

In Pond 1, significant differences between phycocyanin concentrations in the interior and nearshore samples were only seen in two of the six sampling dates. For Pond 1, these significant dates were the last two sampling dates, 8/29/18 and 10/4/18. In Pond 2 significant differences between the nearshore and interior phycocyanin concentrations were seen in three of the six sampling dates. For Pond 2, these significant dates were the first sampling date, 6/14/18, and the last two sampling dates, 8/23/18 and 9/20/18.

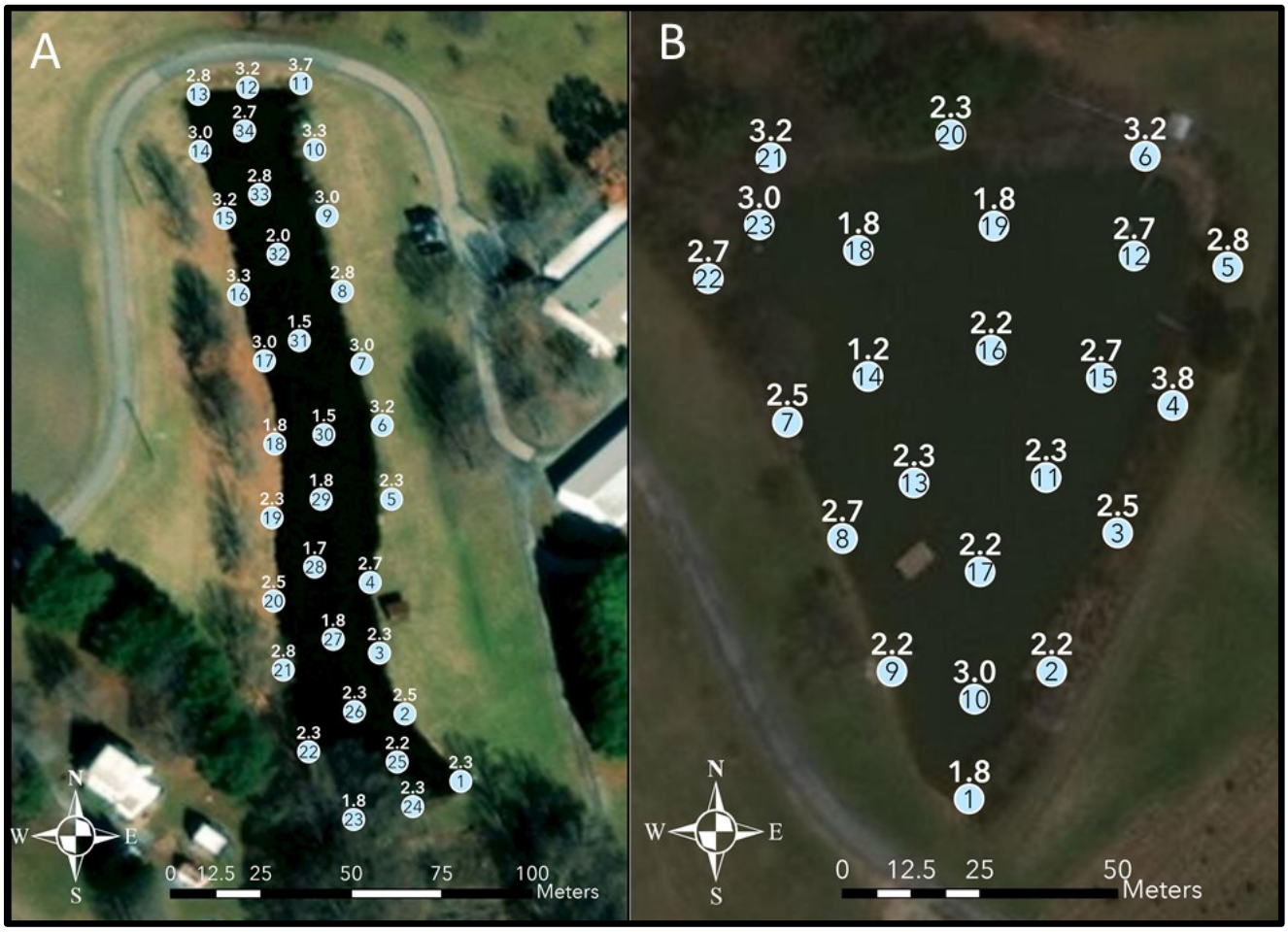


Figure 3.2: Average quartiles of phycocyanin concentrations measured in each location throughout the sampling season for Pond 1 (A) and Pond 2 (B). Number located in the circle is the sampling location designation. Number located above the circle denotes the average phycocyanin quartile rank for that location.

3.4 Discussion

Phycocyanin concentrations in Pond 2 were comparable with several other studies performed in eutrophic reservoirs (Kong et al., 2014; Li et al., 2015; Randolph et al., 2008), ponds (Kasinak et al., 2015) and bays (Mishra et al., 2013). Pond 1 phycocyanin concentrations were low and were comparable with concentrations reported in mesotrophic freshwater reservoirs (Li et al., 2010; Sengpiel, 2007; Song et al., 2013). Because phycocyanin concentrations were representative of other waterbodies and water quality parameters remained relatively stable throughout the sampling season (Appendix B, Supplemental Table 3.1), the study areas examined here provided good systems in which to assess which measurable environmental covariates could be used as predictors of phycocyanin concentrations.

Average concentrations of phycocyanin in Pond 1 were relatively constant after the algicide application in the beginning of July and only increased slightly on the last sampling date in October. In Pond 2, phycocyanin concentrations demonstrated a steady increase throughout the summer, which has been shown by other authors (McQuaid et al., 2011; Otsuki et al., 1994), and is also reflected in remote sensing analyses for cyanobacteria blooms (Sayers et al., 2016; Wynne & Stumpf, 2015) and ecological studies (summarized in Paerl & Otten, 2013). Because of climatic patterns in the Chesapeake Bay region, cyanobacteria populations start to develop in late-May to early-June and begin to decline in mid- to late- October (Marshall et al., 2005; Marshall & Alden, 1990; Wood et al., 2014; J. Wolny, unpublished data). Concentrations were noticeably different between ponds with Pond 2 having approximately 2.5 to 3 times greater phycocyanin concentrations from

July 1 onward. Precipitation seemed to have little effect on cyanobacterial communities with only slight increases in phycocyanin concentrations after rainfall events. After rain events there was no significant increase in *f*DOM concentrations which may indicate that rainfall did not transport runoff containing large amounts of organic matter. Future sampling strategies should be designed with the temporal scale needed to assess the immediate and long-term impacts of rainfall on agricultural pond water quality and phytoplankton community constituents. Temperature did not strongly correlate with phycocyanin increase in this work. One possible reason for that is the narrow range of temperatures measured at different locations on the same observation day. These ranges were on average 2.1°C. Another possible reason could be the relatively wide range of optimum temperatures for growth of the cyanobacteria populations found in these ponds. Konopka & Brock (1978) reported optimal growth occurred between 20 and 30 °C for many of the species we observed in our study sites. On all sampling days, water temperatures were in this interval, so the growth conditions were optimal for cyanobacteria during the entire observation period. The multicollinearity, i.e., correlation of temperature with other regression inputs, should also be considered. In the case of multicollinearity, the dependence on temperature could be masked by the dependence on another variable correlated with temperature. While possible, we did not observe a strong correlation of water quality parameters with temperature on any given day of the study. Although temperature is an important factor of cyanobacteria growth and metabolism, the observational setup in this work appeared to not be appropriate for the demonstration of the effects of temperature.

On 7/1/18, the algicide copper sulfate was applied to Pond 1. While the concentration of copper sulfate used in Pond 1 is unknown, all measured water quality parameters decreased significantly after this application. This is comparable to what Schrader & Kingsbury (2000) and Song et al. (2011) reported for other inland waters that were assessed during and after treatments with copper sulfate. Average concentrations of DO, pH, CDOM, and *f*DOM returned to pre-algicide application concentrations about one month after application. For the algal pigment measurements (CHL RFU, CHL YSI, CHL EXT and phycocyanin), all concentrations decreased after copper sulfate application and slowly recovered to either pre-application or greater levels by the last summer sampling date of 8/29/18. Return of chlorophyll-*a* readings to previous values after algicide application within days to weeks was also seen by Dia (2016) and Effler et al. (1980) with low-level treatments (8-14 $\mu g\ L^{-1}$) of algicide. Elder & Horne (1978) observed recovery of algal populations in as little as 5 days after treatment and attributed this to copper sulfate possibly being beneficial to biological activity if applied in very low concentrations (5-10$\mu g\ L^{-1}$). The rate of recovery of the phytoplankton community to the application of copper sulfate, or other algicides, should be monitored if the algicide application is meant to act as a safeguard to the exposure of crops to cyanotoxins via irrigation waters. It should also be noted that treatment of waterbodies with algicides can immediately release large quantities of biotoxins, if toxin-producing species are highly concentrated (Greenfield et al., 2014; Zhou et al., 2013), so assessment of algal community composition should be evaluated prior to algicide application if the water is to be used for crop irrigation or as drinking water for livestock.

In general, three variables consistently appeared to be primary predictors of phycocyanin concentrations and were the most influential variables from the regression tree analysis: the concentrations of extracted chlorophyll-*a* (CHL EXT), colored dissolved organic matter (CDOM), and turbidity (NTU). Chlorophyll-*a* can be regarded as a good primary predictor because cyanobacteria contain both phycocyanin and chlorophyll-*a* pigments. However, if phycocyanin and chlorophyll-a concentrations are divergent, this could indicate a phytoplankton community not dominated by cyanobacteria, though both Beutler et al. (2004) and Izydorczyk et al. (2009) have demonstrated some flaws with detecting cyanobacteria pigments via fluorometric analysis alone. Izydorczyk et al. (2009) and Zamyadi et al. (2016) summarized the water quality conditions and phytoplankton community composition that may cause discrepancies between pigment measurements taken with field-deployed probes and fluorometry. These scenarios should be investigated for agricultural ponds such as those studied here. It is also important to note that CDOM and turbidity can result in a decrease in the excitation of chlorophyll in water samples causing a dampening effect on the in-vivo chlorophyll fluorescence and potentially altering results (Gitelson et al., 2008; Schalles, 2006; Witte et al., 1982). Turbidity was frequently in the first two splits as a predictor of phycocyanin concentrations. This is likely because the presence of cyanobacteria in the water increases turbidity. The more cyanobacteria present in the water column the less transparent the water will be, having negative consequences on the underlying ecosystem (Havens, 2008; Simis et al., 2005). Additionally, some cyanobacteria, such as species within the Oscillatoriales, are more shade tolerant and can dominate in light-limited turbid,

eutrophic waters (Havens, 2008; Scheffer et al., 1997). Conversely, in environments where there is less submerged aquatic vegetation increased light availability throughout the water column allows cyanobacteria to proliferate (Hudon et al., 2014; J. Wolny, unpublished data) and become more toxic as the production of microcystin may be protective from photo-oxidation (Paerl & Otten, 2013; Zilliges et al., 2011). Colored dissolved organic matter was the other water quality parameter that presented as a primary predictor of phycocyanin concentrations in the regression tree analysis. Cyanobacteria can contribute to the total CDOM concentration during bloom maintenance and degradation phases; this CDOM can provide nutrients to fuel existing or new blooms as it is broken down by photodegradation and heterotrophic bacteria into simpler organics, including urea, and inorganic compounds (Shanmugam et al., 2016; Steinberg et al., 2004; Xie et al., 2012; Zhang et al., 2014). However, CDOM may also offer protection to algal cells from UV irradiation (Patidar et al., 2015) thus providing a more suitable habitat for bloom proliferation.

In both ponds, smaller predictor values resulted in smaller phycocyanin concentrations. However, there were two dates, 7/19/18 and 8/23/18, where turbidity was the primary split predictor but smaller phycocyanin concentrations were found for larger turbidity values. This may be due to a rainfall episode a few days prior to each of the sampling dates. On 7/19/18, there was a rainfall event with total accumulation of 18.5 mm of rain and on 8/21/18 a total accumulation of 9.9 mm of rain fell. Rainfall was above average for the study area during 2018, particularly for May through September (Winter et al., 2020). The saturated ground may have contributed to increased surface run-off that influenced the measured turbidity values.

The regression tree analysis in this work did not consider rainfall as an input variable since rainfall can be important for the average level of cyanobacteria on different days whereas the influential parameters in this work were determined from the differences across the pond on the same day. A sampling design with finer scale temporal monitoring may have allowed us to better assess how rain events affect agriculture ponds in the short- and long-term.

While in general the ponds had the same predictors, they differed in the primary influential predictors. In Pond 1, chlorophyll-*a* was the primary split predictor in 4 out of 6 sampling dates while CDOM and turbidity occurred more so in the secondary splits. This relates well with the phytoplankton community composition data based on cell concentrations obtained through light microscopy analysis. For Pond 1, four out of six dates had a phytoplankton community that was comprised of less than 20% cyanobacteria. This indicates that Pond 1 was not dominated by a cyanobacterial community for a majority of the summer allowing for other chlorophyll-*a* producing algae to proliferate. In Pond 2, which was dominated by cyanobacteria and experienced blooms of the nitrogen-fixing cyanobacterium, *Aphanizomenon*, CDOM was the primary split predictor for half of the sampling dates. Benavides et al. (2018) have shown that DOM concentrations play an important role in stimulating nitrogen-fixation in marine diazotrophic cyanobacteria. Similarly, Vähätalo et al. (2011) showed that terrestrial DOM could support coastal plankton communities through biotransfer mediated by picocyanobacteria. The presence of diazotrophs may enrich the pond water with nitrogen, using this enriched pond water for irrigation may have a fertigation effect, which would be interesting to

evaluate in the future. To date, the role of DOM in small eutrophic ponds, such as those in this study, has not been investigated.

Overall, the ranks of nearshore locations were slightly higher than that of the interior locations. Cyanobacteria tend to proliferate in low flow and stagnant environments (Paerl & Otten, 2013) such as those noted along shorelines of the two study sites. There was a difference in phycocyanin concentrations between interior and near shore samples in both ponds, although only significant in two of six dates and three of six dates for Pond 1 and Pond 2, respectively. Foster et al. (2019) stated that near shore or shoreline accumulations of cyanobacteria are driven mostly by wind. While wind direction and speed were not noted for this study this is a reasonable assumption to make given the open areas in which these ponds are situated and their shallow depths. Associations have been made between water quality, wind, and depth for other small waterbodies (Andres et al., 2019; Ji & Havens, 2019).

Interestingly, locations of higher quartile ranks seem to group together in certain zones of the ponds which shows a stronger pattern than simply whether the sampling location was located on the interior or bank of the pond. In Pond 1 there seems to be two zones of high concentrations of phycocyanin: one in the northwest corner near locations 21, 22, and 23, and another on the east side of the pond near locations 4, 5, 6, 12, and 15. These zones may be a product of important features of the pond. The northwest zone is located near an inflow point, which is believed to be ground water fed. This area is relatively shallow compared to the rest of the pond. Scheffer et al. (1997) found shallow lakes to be more conducive to cyanobacteria growth because there is less thermal stratification and groundwater discharge has

been hypothesized as a nutrient source for inland and coastal HABs (Hagerthey & Kerfoot, 2005; Hu et al., 2006). Kosten et al. (2011) have hypothesized that shallow waterbodies are closer to the sediment-water boundary in which nutrients can re-suspend into the water column and Rengefors et al. (2004) demonstrated that mixing action in the littoral zone can stimulate the growth of cyanobacteria, such as *Microcystis* and *Dolichospermum*. The eastern zone where the pond's outflow and inflow are located may have a higher nutrient concentration compared to other regions in the pond as a result of water transfer activities from a nearby creek. Nutrient allocation within the pond or in the feeder creek was not studied during this survey but should be considered in future efforts. Additionally, the northern end of Pond 2, where an ephemeral creek brings in water during rain events had, on average, higher phycocyanin concentrations. This inflow consists of runoff from adjacent, fertilized agricultural fields. Cyanobacteria are known to congregate where nutrients are accessible (Davis et al., 2009; Havens et al., 2003; Paerl et al., 2011) and can even migrate to nutrient sources (Butitta et al., 2017). The fine-scale spatial gradient elucidated in the 1-acre ponds used in this study demonstrates that sampling results may be substantially different depending on the locations within the pond which are sampled. Thus, the choice of a sampling location should not be an arbitrary decision and if feasible multiple locations within cyanobacteria-prone waters should be routinely monitored. Samples from one zone of the pond may contain greater populations of potentially toxic cyanobacteria than other zones which could alter the conclusion of water quality assessments. Cyanobacteria toxins were not routinely assayed during this study. However, the phytoplankton communities in both ponds

were dominated by potentially toxigenic cyanobacteria (i.e., *Microcystis*, *Aphanizomenon*, and several other species from the Order Nostocales) from June to October. Resource managers who oversee these ponds, and others like them that are routinely used for agricultural irrigation purposes, should implement a toxin testing plan when water quality parameters, such as elevated DO and pH measurements, coupled with elevated turbidity or CDOM concentrations, indicate the likelihood of a cyanobacteria bloom.

The most influential phycocyanin covariates, i.e., chlorophyll, CDOM, and turbidity, appear to be retrievable with remote sensing technologies and algorithms (Giardino et al., 2012; Kutser et al., 2020; Pyo et al., 2016). Therefore, drones with appropriate imaging equipment could be used to obtain data on the spatial distribution of influential covariates (Kislik et al., 2018) and help define zones for cyanobacteria monitoring. This opens an interesting research avenue to explore.

3.5 Conclusions

Using the machine learning process of decision trees, we determined that extracted chlorophyll-*a*, colored dissolved organic matter and turbidity were the most influential predictors of phycocyanin concentrations for both ponds on all but one of the sampling dates. Monitoring of the spatial patterns of these environmental covariates in ponds or lakes may be helpful to identify spatial patterns of cyanobacteria populations. Zones of consistently higher concentrations of phycocyanin were found in both of the agricultural irrigation ponds studied in this work. This demonstrates the importance of deciding where to sample when monitoring for water quality and harmful algae, even in small waterbodies like these

irrigation ponds. If sampling is conducted at zones with consistently high phycocyanin concentrations, it may not accurately represent the remaining area or potentially the majority of the waterbody. Conversely, measuring from zones with consistently low phycocyanin concentrations may lead to underestimating the potential risk from cyanotoxins in irrigation waters. Understanding the spatial patterns of cyanobacteria in irrigation ponds and the relationships between environmental covariates with easily measured water quality parameters may assist in better and more efficient pond management practices and the ability to better predict potentially toxic cyanobacteria blooms. There were impacts to both irrigation ponds following rain events, but finer scale temporal monitoring is needed to put these findings in context with cyanobacteria blooms.

Chapter 4: Examining the relationship between phytoplankton community structure and water quality measurements: a machine learning application

4.1 Introduction

Phytoplankton community composition and abundance is often used in assessments of recreational, aquaculture, and drinking water quality. Long-term monitoring studies conducted in marine and estuarine waters used for aquaculture activities (Marić et al., 2012; Marshall et al., 2009) and in freshwater lakes and reservoirs used to provide drinking water and recreational areas (Chen et al., 2003; Wynne & Stumpf, 2015; Znachor et al., 2020) have demonstrated distinctive relationships between certain phytoplankton community constituents and water temperature, salinity, and nutrient concentrations. However, long-term phytoplankton community composition studies in small-bodied agricultural irrigation waters to examine similar relationships, are lacking.

The examination of waters for phytoplankton community composition and abundance is a time-consuming activity that relies on the expertise of well-trained phytoplankton taxonomists or automated technologies, such as flow cytometry, that may be cost-prohibitive to many water quality management programs (Bergkemper & Weisse, 2018; Lawton et al., 1999). Satellite imagery has proven useful for monitoring phytoplankton community structure in large lakes (>24,000 acres, (Ho et al., 2019)) but does not yet have the spatial scale needed to remotely observe smaller waterbodies that are increasingly being used in agricultural irrigation applications (López-Felices et al., 2020). Hence, alternative techniques are being explored to examine the relationships between more easily measured water quality parameters

(i.e., temperature, chlorophyll-*a*, and specific conductance) and phytoplankton community composition and abundance. The presence of such relationships makes the use of regression analysis feasible for predicting phytoplankton concentrations using measured water quality parameters. Regression analyses have been used to predict the occurrence of bloom-forming cyanobacteria in shallow lakes (Descy et al., 2016; Rao et al., 2021), green algae in reservoirs (Fornarelli et al., 2013), and diatoms in estuaries, rivers, and lakes (Gayoso, 1998; Schönfelder et al., 2002).

Regression analyses were used to successfully predict the composition of phytoplankton communities in a drinking water reservoir near Beijing, China that was greater than 44,000 acres (Zeng et al., 2017). However, as noted by Cheruvelil et al. (2008), scale and regionalization are important factors to consider when conducting water quality assessments and applying water quality standards. Recently, machine learning provided several versatile techniques to establish models suitable to create 'phytoplankton – water quality' relationships. The work presented here assesses the use of the random forest analysis, similar to the regression analyses employed by Zeng et al. (2017), to estimate phytoplankton abundance in small scale (1 acre) agriculture ponds used for crop irrigation. The objective of this work was to evaluate the performance of the random forest algorithm in estimating the phytoplankton functional groups from in-situ water quality measurements of different complexity which were obtained during three years of spatially intensive observations at two agricultural irrigation ponds.

Phytoplankton community structure has long been used to assess trophic changes in aquatic systems (Reynolds, 1998) with shifts from green algae dominated

communities to cyanobacteria dominated communities indicating eutrophic conditions (Duarte et al., 1992; Watson et al., 1997). For this study three phytoplankton groups were considered critical to assess in relationship to water quality parameters due to their abundance within local freshwater phytoplankton populations. Previous studies by Parson and Parker (1989) and Marshall (2013, 2014) demonstrated that between 70-80% of regional freshwater lake phytoplankton community structure was composed of green algae (Chlorophytes), diatoms (Bacillariophytes), and cyanobacteria (Cyanophytes). Due to the harmful and potentially toxic effects of cyanobacteria blooms on human and environmental health, the detection, prediction, and modeling of these blooms has become a focus for resource managers (Rousso et al., 2020; Stauffer et al., 2019; Stumpf et al., 2016). Additionally, there is growing concern about the risk that cyanotoxins may pose to the agriculture industry through degraded water quality and the transfer of cyanotoxins from irrigation waters to crops and livestock, particularly as climate change increases the occurrence and toxicity of cyanobacteria blooms (Lee et al., 2017; Weralupitiya et al., 2022; Wood, 2016).

4.2 Methods

4.2.1 Data collection

Phytoplankton and water quality sampling was conducted every two weeks at two 1-acre ponds on working farms in Maryland during the 2017 and 2018 growing seasons (May – October). Pond 1 (Figure 4.1-P1) located in Germantown, Maryland is a man-made embankment pond with in-flow from a co-located pond; 23 stations were routinely sampled in this pond. Pond 2 (Figure 4.1-P2) located in Wye Mills,

Maryland (University of Maryland Wye Research Center) is an excavated pond with inflow from an ephemeral creek; 34 stations were routinely sampled in this pond. Phytoplankton samples and water quality measurements were made at all stations in Pond 1. Phytoplankton samples at Pond 2 were collected at fewer locations, consisting of odd numbered nearshore locations and all interior sampling locations (22 stations) whereas water quality measurements were made at all stations. Full site descriptions are provided in Smith et al. (2020). In-situ measurements were taken along with a water sample for laboratory processing at each sampling location. A YSI Exo-2 sonde (Yellow Springs Instruments, Yellow Spring, OH, USA) was used to measure temperature (TEMP), dissolved oxygen (DO), specific conductance (SPC), pH, fluorescent dissolved organic matter (FDOM), and turbidity (NTU). As a proxy for phytoplankton density, both chlorophyll-*a* (CHL) and phycocyanin (Phyco) were measured with the YSI Exo-2 sonde as demonstrated by Brient et al. (2008) and Song et al. (2013). Water samples were measured for colored dissolved organic matter (CDOM) using a Turner Designs AquaFluor fluorometer (San Jose, CA, USA). Identification and enumeration of phytoplankton was performed using a modified Utermöhl method described in Marshall & Alden (1990) with taxa identified according to John et al. (2011) and Bellinger & Sigee (2015). For full details of sampling methodologies see Smith et al. (2021). Water quality sampling methods and phytoplankton analyses for the 2019 sampling year were the same as those performed in 2017 and 2018 but occurred on a less routine schedule. In Pond 1 there were six sampling dates in 2017, six sampling dates in 2018, and three sampling dates in 2019.

In Pond 2 there were five sampling dates in 2017, six sampling dates in 2018, and two sampling dates in 2019.

Field work conducted in 2017 and 2018 yielded 518 phytoplankton samples, in-situ measurements, and laboratory-based water quality measurements. For the purpose of the random forest analysis phytoplankton data was examined at the functional group level (diatoms, pelagic green algae, and cyanobacteria). While other taxa (i.e., dinoflagellates) were observed with microscopy analyses, the low spatial and temporal occurrence and abundance of these taxa over the course of the study precluded examination with the random forest analysis. The data collected in 2019 were used as a blind dataset to test the random forest model.

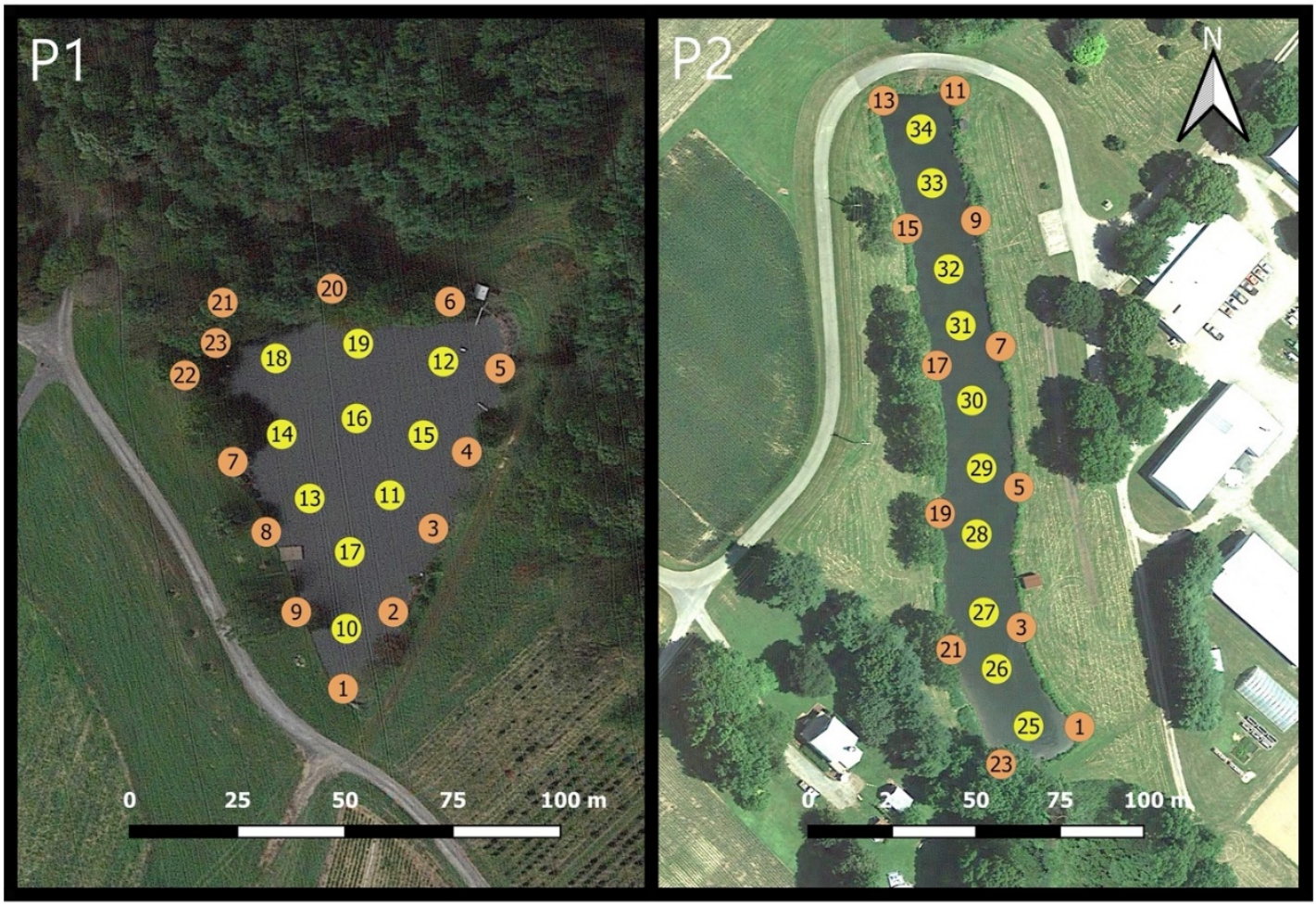


Figure 4.1. Sampling locations for both Pond 1 (P1) and Pond 2 (P2). Sampling location number is shown inside the circle. Yellow circles indicate interior water sampling locations and orange circles indicate nearshore sampling locations.

4.2.2 Modeling with the random forest algorithm

The machine learning random forest algorithm was used to predict phytoplankton functional group concentrations and the most influential parameters for each group. The random forest algorithm is an extension of decision tree algorithms. Each decision tree splits a dataset into multiple subgroups. At each split the data is divided into two groups: one group contains the most similar values within the dataset and another group containing the most dissimilar values within the dataset. The splitting process ends with subgroups called nodes, averages over which is the

sought prediction. While regression trees alone are very informative about influential parameters of the data set, a single tree is often not the best at prediction. The random forest algorithm builds many decision trees and averages their outputs. The result is more accurate outputs which are better suited for prediction models.

Random forest models with various inputs and outputs were developed in this study. Three input datasets (A, B, and C) were used for each of the three output datasets of phytoplankton functional groups (diatoms, green algae, and cyanobacteria). The input set A included physio-chemical parameters, i.e., TEMP, pH, DO, NTU, and SPC. In 2018 photosynthetic active radiation (PAR) was included in input set A. The input set B included parameters related to organic constituents, i.e., CHL, Phyco, CDOM, FDOM. Input set C included nutrients and macro elements, i.e., potassium, calcium, magnesium, ammonium, nitrate, and phosphate. For 2017 and 2017+2018 data sets, the random forest model was developed with input set A, input set B, and combined input sets A and B (AB). For the 2018 dataset random forest models were developed with input set A, input set B, input set C, and combinations of these input sets (AB, AC, BC, ABC). All random forest computations were completed in Rstudio (Rstudio Team, Boston, MA, USA) using the 'randomForest' package.

4.2.3 Model performance metrics

To evaluate the model's prediction capabilities, the root-mean-squared errors (RMSEs) were computed with the predicted and measured values as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\log C_{i,meas} - \log C_{i,predict}\right)^2}$$

where $log\ C(i,meas)$ and $log\ C(i,predict)$ are measured and predicted concentrations for the i^{th} dataset and N is the total number of datasets. The RMSE values were computed for testing datasets for individual regression trees of the random forest models if the independent testing data were not available. If the independent data was available, the RMSE vales were computed from the testing datasets and predictions of the random forest models. T-tests were used to determine significant differences in accuracy metric results, and a p value of 0.05 was selected.

The Williams-Kloot test (Williams & Kloot, 1953) was utilized to compare performance of pairs of random forest models obtained with different inputs for estimating phytoplankton concentrations. The test consists of computing the slope of the inward regression using the following equation:

$$Y - 1(2)\ \ Y(1) + \ Y(2)(2) = \lambda\ \ Y(2) - \ Y(1)(2)$$

where Y is the measured concentration, $Y(1)$ is the predicted concentration from model 1 and $Y(2)$ is the predicted concentration from model 2. If λ is positive and significantly different from zero, then the performance of model 2 is better than performance of model 1. If λ is negative and significantly different from zero, then the performance of model 1 is the better of the two models. A p value of 0.05 was selected to determine significance in the Williams-Kloot test applications.

The ratio of coefficients of variance (CVs) were also calculated to compare the variation of interior locations with the variation of nearshore sampling locations for phytoplankton functional groups and water quality parameters. The equation for calculating the ratio of CVs for each parameter is as follows:

$$Ratio\ of\ CV = \quad \sigma(n)(2)\ \mu(n)(2)\ \ \sigma(i)(2)\ \mu(i)(2)$$

where σ is the standard deviation and μ is the mean of the interior *(i)* parameters or nearshore (*n*) parameters.

The input variable importance was quantified by the Mean Decrease Accuracy (%IncMSE) as implemented in the Rstudio randomForest package. The %IncMSE reflects the loss of model accuracy when a variable is scrambled, i.e., its values are randomly rearranged. The model decreases of accuracy are computed for each tree in the forest and the percentage of decrease of accuracy is averaged over all trees to get the mean value.

4.3 Results

4.3.1 Data summary

Average daily temperatures, precipitation, and sampling dates for 2017 and 2018 are displayed in Appendix C, Supplemental Figure 4.1. The total number of phytoplankton data points was 4,144 for both years and ponds. There were 1,554 datapoints after phytoplankton data were combined into functional groups. Summary statistics for phytoplankton functional groups and water quality data over the two growing seasons are presented in Table 4.1 and Supplemental Table 4.1 (Appendix C). The most dominant and commonly occurring phytoplankton taxa were all representatives of eutrophic, shallow, small water bodies per the functional group classifications of Reynolds et al. (Reynolds et al., 2002). Diatom concentrations for both years ranged from 4.19 to 7.59 log cells L^{-1} and from 4.19 to 7.77 log cells L^{-1} in Pond 1 (*Aulacoseira* spp.) and Pond 2 (*Aulacoseira* spp. and *Cyclotella* spp.), respectively. In Pond 1, cyanobacteria (*Microcystis* spp.) ranged from 4.19 to 7.95 log cells L^{-1} and green algae (*Coelastrum* spp. and *Scenedesmus* spp.) ranged from 5.49

to 8.08 log cells L^{-1} for both years. In Pond 2 cyanobacteria (*Aphanizomenon* spp., *Dolichospermum* spp. and *Microcystis* spp.) and green algae (*Closterium* spp. and *Scenedesmus* spp.) ranged from 4.67 to 8.69 log cells L^{-1} and from 4.89 to 8.18 log cells L^{-1}, respectively. In 2017 and 2018 green algae had the highest average cell concentrations, followed by cyanobacteria and then diatoms in Pond 1. At Pond 2, cyanobacteria had the lowest average concentrations of the phytoplankton groups in 2017, whereas in 2018 diatoms had the lowest average concentrations. For both ponds and years, the median and the means of each phytoplankton group were similar in value indicating a symmetrical dataset. The only exception to this was in 2018 at Pond 1 where the cyanobacteria cell concentration mean was higher than the median, indicating a skewed dataset.

In 2017 a total of eight physio-chemical parameters and organic constituents commonly used to assess water quality were measured in the field and used in set A and set B for training the random forest algorithm. An additional 11 physio-chemical, organic constituents, and nutrient/macro element parameters were added in 2018 and applied across set A, set B, and set C training datasets. Average TEMP, DO, SPC, pH, NTU, and Phyco did not differ from 2017 to 2018 in both Pond 1 and Pond 2. CHL averages doubled from 2017 to 2018 at Pond 2. While CHL concentrations were low at Pond 1 for both 2017 and 2018, there was a decrease in 2018. This may be the result of routine algicide application to Pond 1 during the study period.

Table 1. Average values of measured parameters for 2017 and 2018

Variable	Units	Pond 1	Pond 2
2017			
Diatoms	Log cells/L	5.94	6.32
Green Algae	Log cells/L	7.10	6.92
Cyanobacteria	Log cells/L	5.98	6.14

TEMP	°C	26.31	27.37
DO	mg/L	10.39	11.82
SPC	uS/cm	162.52	161.41
pH		8.82	8.18
NTU		5.16	10.82
Phyco	RFU	1.14	3.49
CHL	RFU	3.83	10.74
FDOM	ppb	13.06	30.34
2018			
Diatoms	Log cells/L	5.61	5.63
Green algae	Log cells/L	6.70	7.10
Cyanobacteria	Log cells/L	5.64	7.08
TEMP	°C	26.99	27.54
DO	mg/L	10.07	14.91
SPC	uS/cm	148.83	142.59
pH		7.65	8.17
NTU		4.20	13.86
Phyco	RFU	0.95	4.86
CHL	RFU	2.80	19.97
FDOM	ppb	20.72	34.90
CDOM	ug/L	76.80	177.00
Sub. light 0cm		712.72	1113.02
Sub. light 7.5cm		624.42	810.60
Sub. light 15cm		540.88	578.13
PAR		950.46	1312.86

4.3.2 Model accuracy

The RMSEs that characterized the random forest model performance are shown in Figure 4.2. The differences in RMSE between ponds were relatively small. However, in almost all instances RMSE values for Pond 2 were larger than those for Pond 1 for all three phytoplankton functional groups (green algae, diatoms, and cyanobacteria) and all three time periods (2017, 2018, and 2017+2018). It is also worth noting that the ranges of log phytoplankton concentrations, computed from minimum and maximum values in Appendix C, Supplemental Table 4.1, were also slightly greater in Pond 2 than in Pond 1 for all three phytoplankton functional groups and all three time periods. The smallest and the largest RMSEs for the combined year

data were found for green algae and cyanobacteria, respectively. RMSE values for diatoms were in most cases an intermediate value. An exception to this was in 2018 in Pond 2; here diatoms RMSEs were larger than the cyanobacteria values. RMSEs of the 2018 model were lower than RMSE of 2017 model. Mean values of measured parameters in the 2018 dataset were also lower than those from 2017 (Appendix C, Supplemental Table 4.1). RMSE values of the combined dataset 2017+2018 were smaller than the RMSE values of 2018. Creating the combined year dataset improved the robustness of the random forest models.

The small differences in RMSE between random forest models using input set A and input set AB implied that there may be not a significant difference between model performance. All Williams-Kloot tests yielded positive λ values indicating that modeling with set AB as the input may be superior to model created with set A as the input. The Williams-Kloot test showed that there was a significant difference between models for green algae in Pond 1 ($p<0.001$) and cyanobacteria in Pond 2 ($p=0.010$), but not for green algae in Pond 2, cyanobacteria in Pond 1, nor diatoms in either pond.

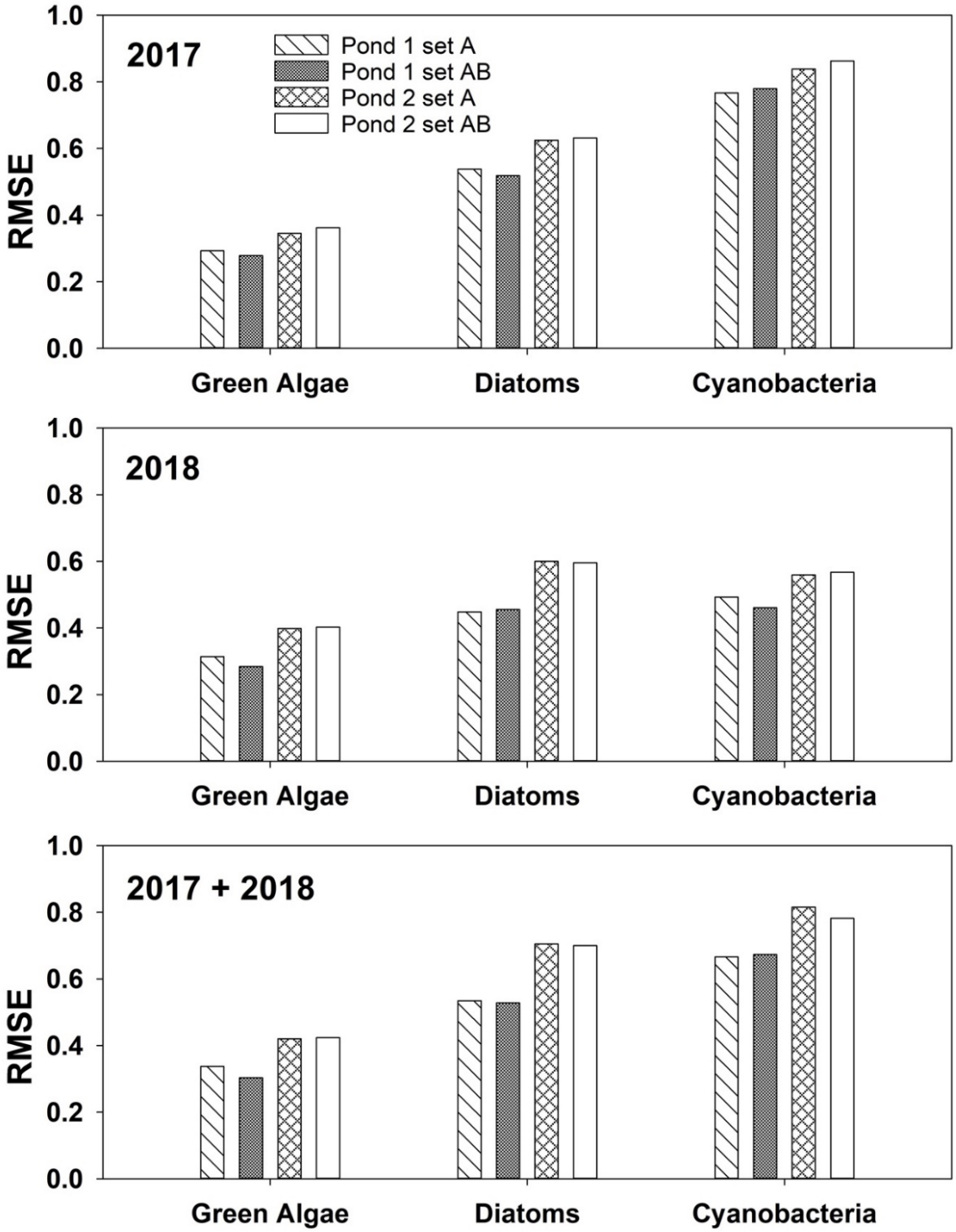


Figure 4.2: Root-mean-squared errors of the random forest models for green algae, diatoms, and cyanobacteria for Ponds 1 and 2, with data from 2017, 2018, and 2017+2018.

4.3.3 Model validation

Models developed with the 2017 and 2018 datasets were tested using data collected in 2019. The results are shown in Figure 4.3. When using the random forest model on blind 2019 data, the RMSE results did not mirror what was predicted during model development using the 2017 and 2018 data. The RMSE values for green algae 2019 predictions were larger than the values for the 2017 and 2018 datasets using any combination of input sets A and B. For 2019, cyanobacteria continued to produce the higher RMSE values, whereas diatoms presented the lowest RMSE values. Pond 2 continued to have higher RMSEs for diatoms than Pond 1. Cyanobacteria in Pond 1 displayed higher RMSE values in 2019 whereas for training years (2017 and 2018), Pond 2 typically had higher cyanobacteria RMSE values. Overall, green algae RMSEs were much higher for the 2019 validation data compared to the training dataset. In all instances RMSE values were lower when the model was run with set AB parameters. The Williams-Kloot test determined that the AB model was superior to the A model. The AB model performance was significantly different ($p<0.05$) for all groups and both ponds, except for Pond 2 diatoms ($p=0.84$).

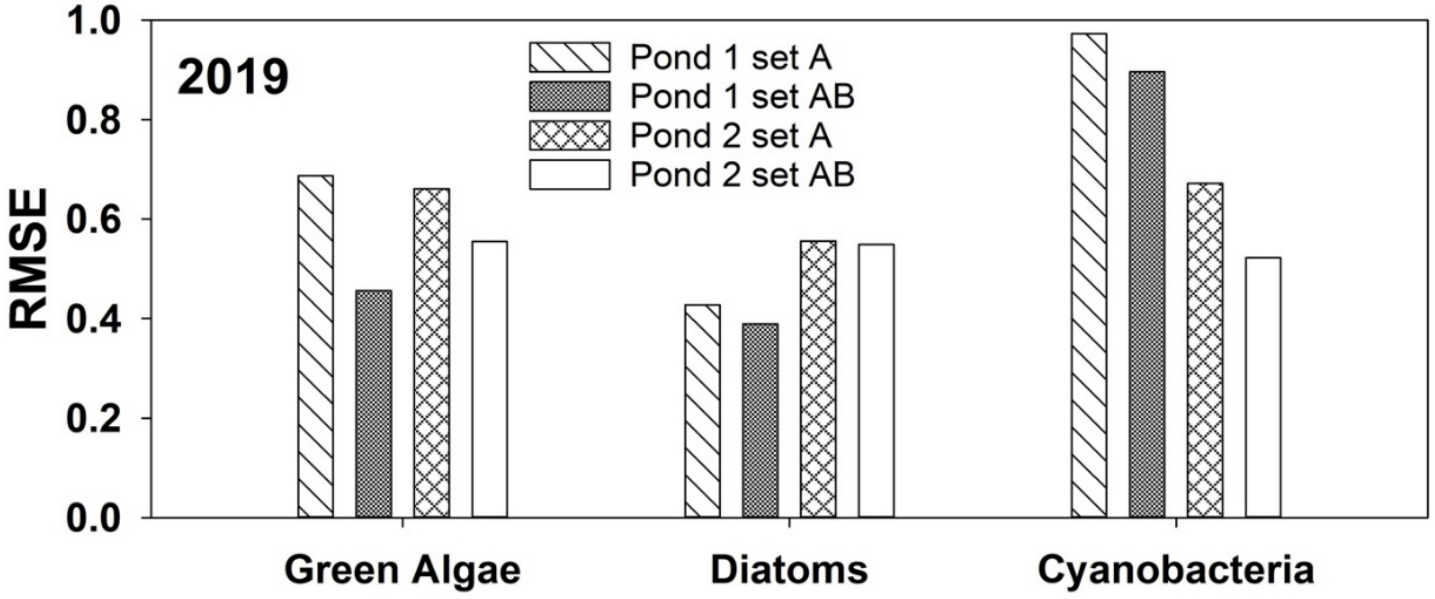


Figure 4.3. Root-mean-squared errors of the random forest models for green algae, diatoms, and cyanobacteria for Ponds 1 and 2, with blind data from 2019.

4.3.4 Spatial patterns of random forest model performances

Spatial distribution of the individual location errors with data from 2017+1018 and set A parameters is shown in Figures 4.4-4.6. There was a pattern of lower RMSE values for interior locations compared with nearshore locations in each pond. For all three groups of phytoplankton within both ponds, the lowest RMSE values were found in the interior of the ponds, except the outflow area of Pond 2 (location 23). The average RMSE values were larger, and the performance of the models was reduced at nearshore locations compared to interior locations for all phytoplankton groups at Pond 2. Separation of nearshore locations from interior locations revealed that, in Pond 2, the probability (t-test) of the average RMSE being the same over nearshore and interior locations was very low ($p < 0.01$) for green algae and cyanobacteria. The probability of RMSE values being the same for nearshore and

interior locations for diatoms in Pond 2 was greater but still small ($p<0.1$). In Pond 1, no substantial differences in average RMSE for nearshore and interior locations were found for green algae ($p>0.5$), and only moderate differences were found for diatoms and cyanobacteria ($p<0.1$). The percentage of sampling dates in which the CV was larger for nearshore locations compared to interior locations can be found in Table 4.2. For diatom and cyanobacteria, more than 54.6% of the sampling dates had higher CVs for nearshore samples compared to interior samples for both ponds. For green algae, Pond 2 (63.6%) had a higher percentage of dates with nearshore variability being higher than Pond 1 (41.7%). Overwhelmingly most water quality measurements had high CVs at nearshore samples with most being greater than 75% of the sampling dates. The exception to this is both SPC (63.6% of dates) and Phyco (72.7% of dates) in Pond 2.

Spatial distribution of the individual location errors with data from 2017+2018 and set AB parameters is shown in Appendix C, Supplemental Figures 4.2-4.4. Similar to the spatial distributions of errors of the model using set A parameters, set AB parameters shows a similar pattern of interior locations mostly containing the lowest RMSE values. This was true for cyanobacteria and diatoms at Pond 1 and green algae and cyanobacteria at Pond 2. A t-test of set AB showed that no differences were found in the average RMSEs for interior and nearshore locations for green algae ($p>0.05$) at Pond 1. Diatom RMSEs at both ponds exhibited moderate ($p<0.1$) differences between interior and nearshore locations. Significant ($p<0.05$) differences between nearshore and interior RMSEs were found for cyanobacteria at both ponds and green algae at Pond 2.

Table 4.2. Percent of dates in which the coefficient of variation (CV) was larger for nearshore locations.

Table 4.2		
Parameter	**Pond 1 %**	**Pond 2 %**
Green algae	41.7	63.6
Diatoms	66.7	63.6
Cyanobacteria	66.7	54.6
TEMP	75.0	81.8
DO	100.0	90.9
SPC	75.0	63.6
pH	100.0	100.0
NTU	91.7	81.8
Phyco	91.7	72.7
CHL	75.0	81.8
FDOM	75.0	90.9

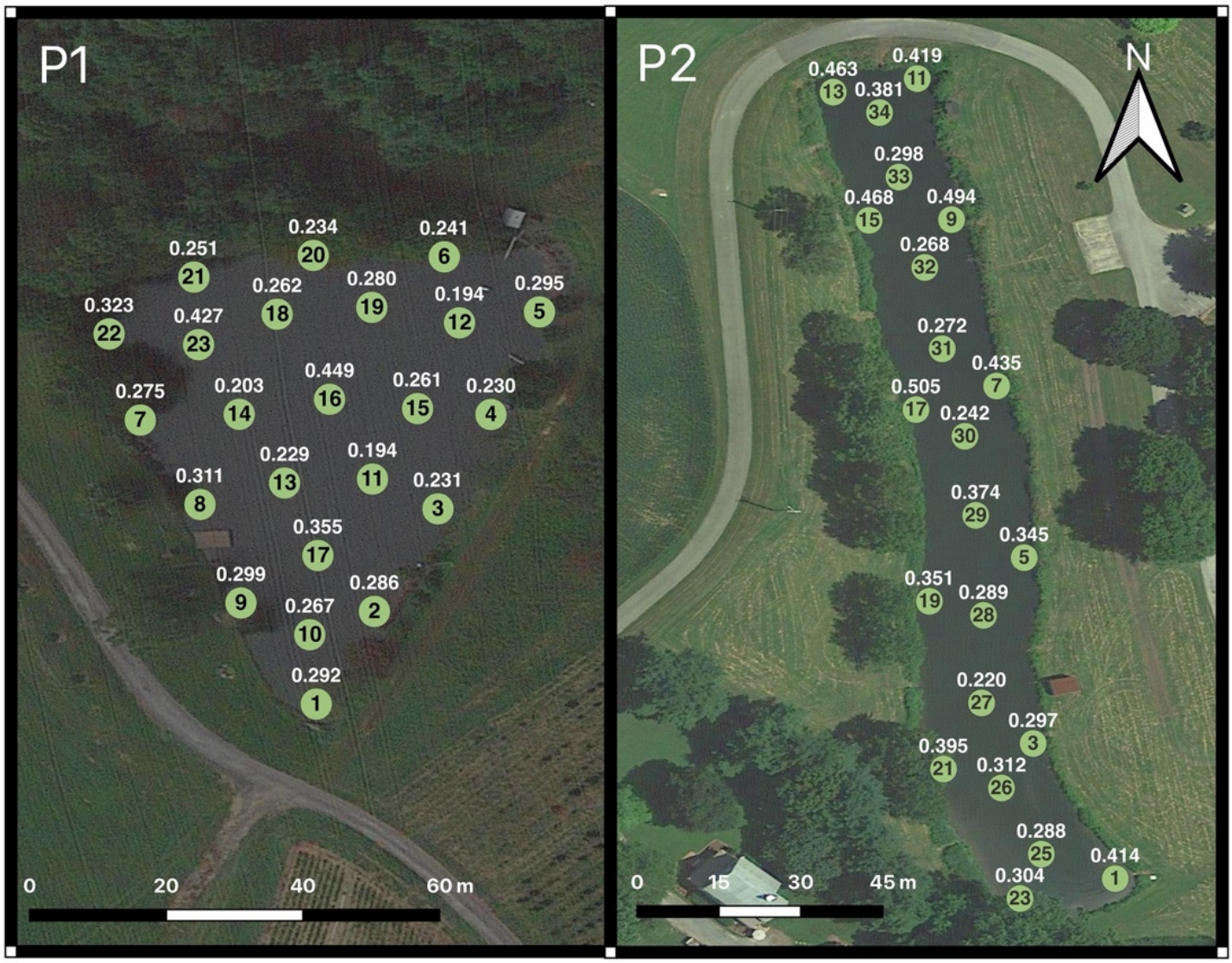


Figure 4.4: Spatial pattern of root-mean-squared errors calculated for green algae using set A parameters errors for 2017+2018 combined. Number located inside of symbol indicates sampling location number. Number above location indicates the RMSE for green algae at that location.

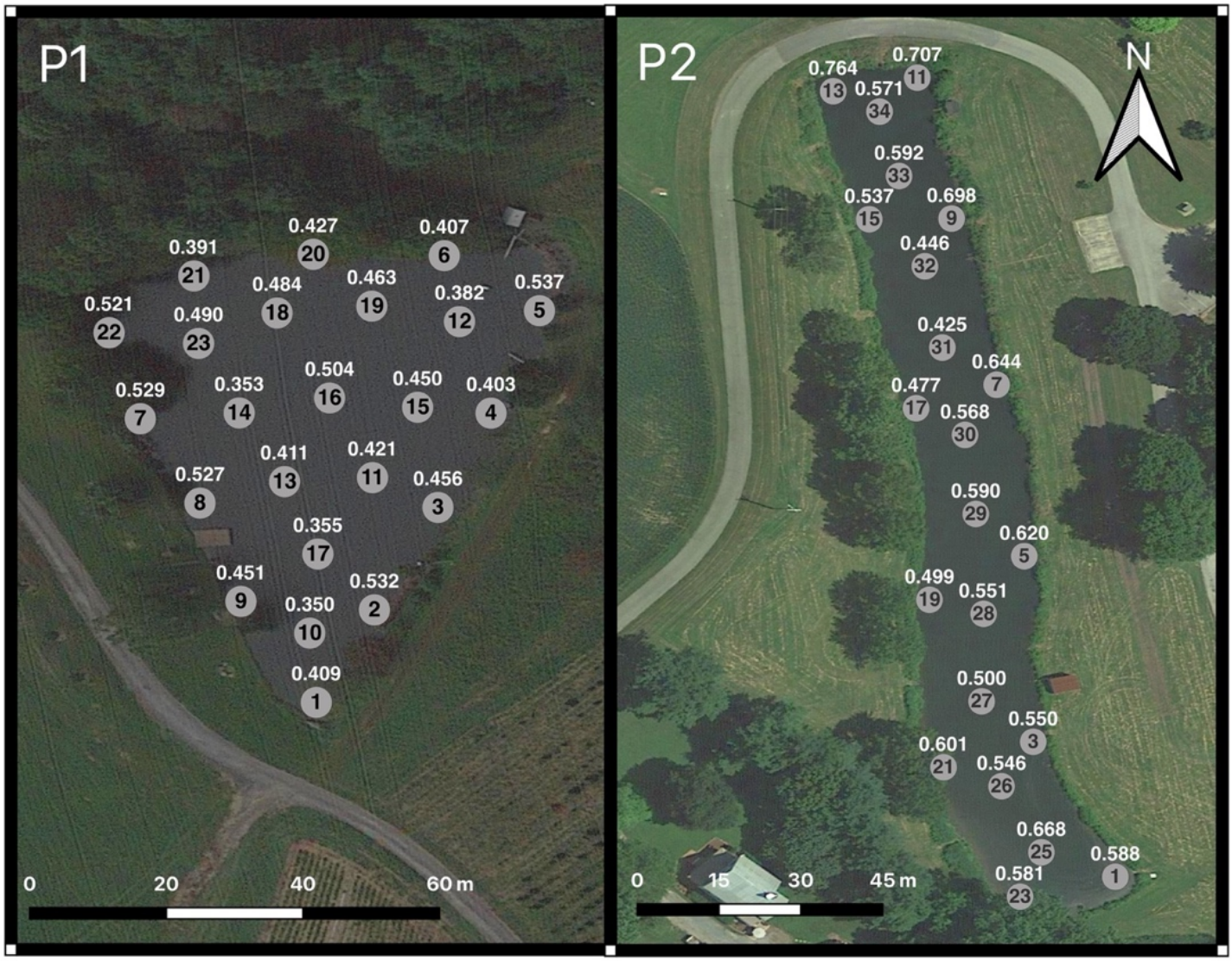


Figure 4.5: Spatial pattern of root-mean-squared errors calculated for diatoms using set A parameters errors for 2017+2018 combined. Number located inside of symbol indicates sampling location number. Number above location indicates the calculated RMSE for cyanobacteria at that location.

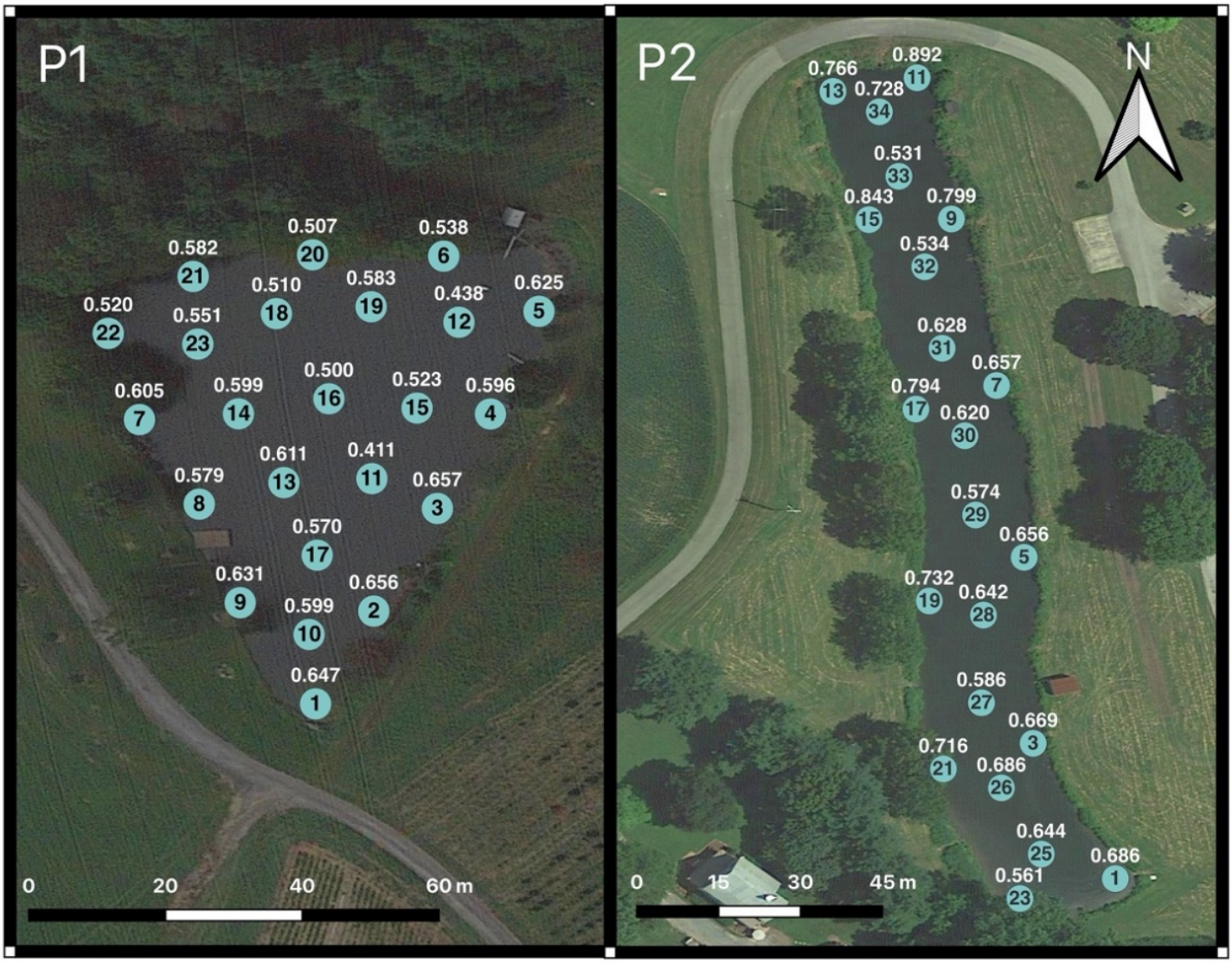


Figure 4.6: Spatial pattern of root-mean-squared calculated for cyanobacteria using set A parameters errors for 2017 and 2018 combined. Number located inside of symbol indicates sampling location number. Number above location indicates the calculated RMSE for cyanobacteria at that location.

4.3.5 Importance of variables-predictors

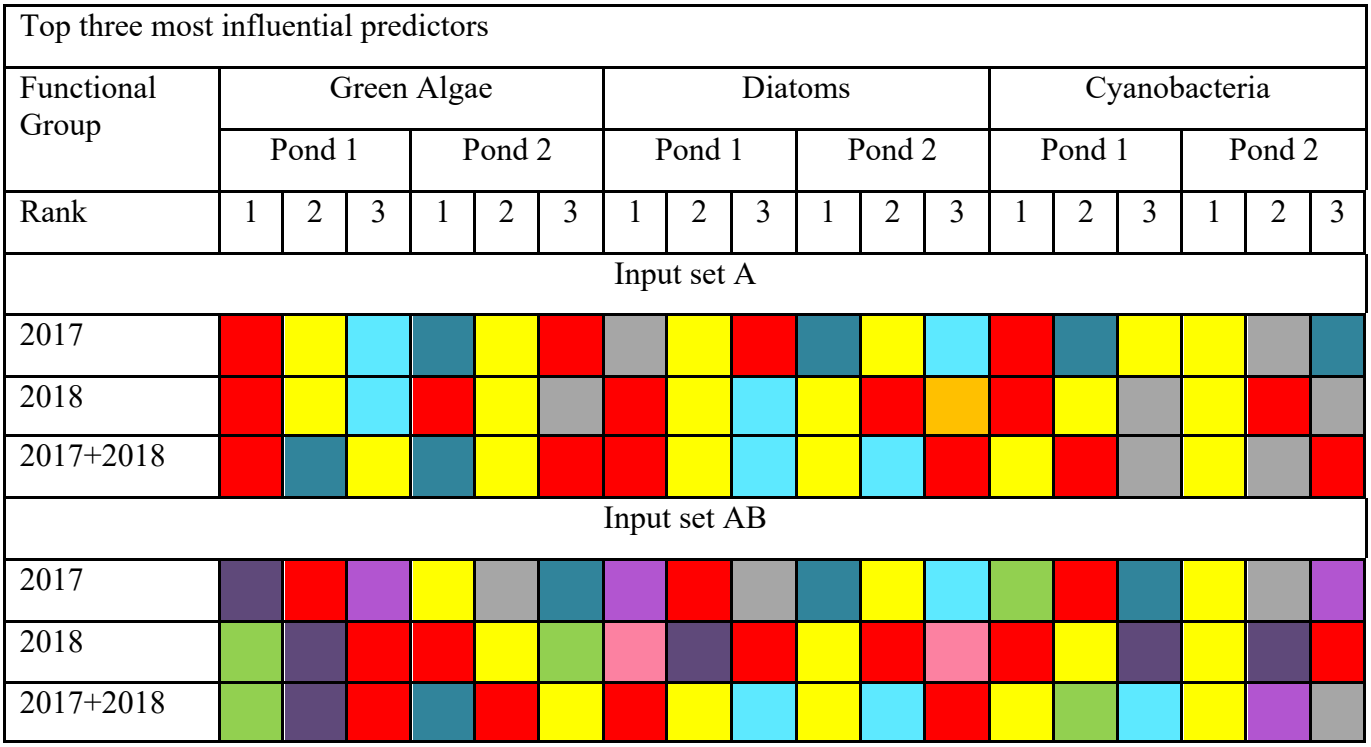
The top three most important predictors for each dataset and model for are shown in Tables 4.3 and 4.4, and in Figure 4.7. SPC and TEMP were the most influential predictors; found in 63% of all cases when using input set A and in 46% of

all cases when using input set AB. NTU was seen in 7% and FDOM was seen in 11% of all cases where input set A and input set AB were used, respectively. Predictors from set A continued to have high importance (total of 61%) when the input set AB was used.

There was no significant difference between the ponds when considering the top three most influential predictors when input set A was used. Using input set A, SPC was the most influential predictor, with nine occurrences for each pond. TEMP was in the top three most influential predictors nine times for Pond 1 and seven times in Pond 2. There was a greater difference between the ponds when input set AB was used. SPC was the most influential predictor three times for Pond 1 and nine times for Pond 2. TEMP was among the most influential predictor eight times for Pond 1 and five times for Pond 2. The influence of CHL was more prominent for Pond 1 (4 times) than for Pond 2 (once). Similarly, the fDOM was more prominent for Pond 1 (five times) than for the Pond 2 (once). Overall, with the AB input set, predictors of the input set A were less influential in Pond 1 (52% of all occurrences) than in Pond 2 (78% of all occurrences).

There were clear differences among the phytoplankton groups. NTU was the most influential predictor when assessing cyanobacteria, yet CHL was not. For green algae, DO was the most frequent influential predictor with the input set A, but no influence of DO was found when the organic matter-related inputs were included as part of set AB. CDOM was found as an influential predictor only for diatoms. Diatoms in Pond 2 had the same most influential predictors with inputs sets A and AB. The same was true for diatoms with combined 2017+2018 data in Pond 1.

Top three most influential predictors were different in 2017 and 2018 in most cases, with green algae in Pond 1 being an exception. The 2017+2018 dataset, in some cases, led to the influential predictors being the same as individual years modeled separately (e.g., green algae in Pond 2 with the input set A, diatoms in Pond 1 with the input set A, green algae with the input set AB in Pond 1, cyanobacteria in Pond 2 with input sets A and AB). The nutrient-related variables available in 2018 are grouped in input set C. These proved to be most important when all available input variables (input set ABC) were used as input for green algae and diatoms (Tables 4.3 and 4.4), but not cyanobacteria. Because nutrient data was only collected in 2018 it was excluded from the 2017+2018 dataset to avoid unequal weighting across all parameters.



	Temperature		PAR
	Specific Conductance		*f*DOM

	pH		Phycocyanin
	Turbidity		Chlorophyll-*a*
	Dissolved Oxygen		CDOM

Figure 4.7. Top three most important predictors in random forest models of green algae, diatoms, and cyanobacteria populations in pond waters with datasets from 2017, 2018 and combined 2017+2018.

4.3.6 Sensitivity to inputs

The mean decrease of accuracy (%IncMSE) are shown below each variable in Tables 4.3 and 4.4. For Pond 1 in 2017 and 2018, %IncMSE values were low for green algae and diatoms (<15), and slightly higher for cyanobacteria (>15). The combination of years (2017+2018) produced higher increases of mean square error indicating that multiyear data allowed for predictors to be more influential. For Pond 2, the sensitivity to the important variables tended to be higher than in Pond 1. The values of %IncMSE in Pond 2 was less than 30 for green algae and diatoms for both years. Cyanobacteria had larger (>30) increase in mean square error values with the highest value being 143, indicating cyanobacteria predictions were more sensitive to the influential predictors than predictions for green algae and diatoms. Multiyear data tended to increase the %IncMSE values causing greater sensitivity to influential predictors.

Table 4.3. Most influential predictors and the increase in accuracy for that variable for Pond 1 as determined using the random forest algorithm.

Pond 1 Random forest most influential predictors									
Input Group	Green Algae			Diatoms			Cyanobacteria		
	Imp Var 1	Imp Var 2	Imp Var 3	Imp Var 1	Imp Var 2	Imp Var 3	Imp Var 1	Imp Var 2	Imp Var 3
2017									
A	TEMP	SPC	DO	NTU	SPC	TEMP	TEMP	pH	SPC
	4.3	2.7	2.6	14.3	12.6	12.4	21.0	15.5	14.9
B	FDOM	Phyco	Chl	Phyco	FDOM	Chl	Chl	Phyco	FDOM
	6.7	3.8	3.2	22.0	15.4	10.9	37.3	22.2	17.3
AB	FDOM	TEMP	Phyco	Phyco	TEMP	NTU	Chl	TEMP	pH
	2.8	2.8	1.9	9.7	8.0	7.7	14.2	12.2	11.0

2018									
A	TEMP	SPC	DO	TEMP	SPC	DO	TEMP	SPC	NTU
	9.5	6.1	6.1	10.0	8.0	7.	32.3	25.9	16.1
B	Chl	FDOM	CDOM	FDOM	CDOM	Chl	FDOM	Chl	CDOM
	20.6	8.9	4.2	17.5	16.0	9.1	44.9	30.1	23.3
C	K	NO3	Ca2	NO3	H2PO4	Ca2	NO3	K	H2PO4
	13.1	4.8	4.8	12.9	9.7	8.6	29.1	28.7	18.9
AB	Chl	FDOM	TEMP	CDOM	FDOM	TEMP	TEMP	SPC	FDOM
	10.3	5.3	4.5	6.6	6.0	5.5	18.5	17.7	15.6
AC	K	Ca2	NO3	TEMP	NO3	H2PO4	TEMP	SPC	K
	6.8	3.9	3.7	6.6	5.9	5.1	18.0	15.8	15.2
BC	K	Chl	Ca2	FDOM	NO3	CDOM	FDOM	K	NO3
	8.1	8.0	3.7	7.6	7.4	6.8	21.7	17.5	14.3
ABC	Chl	K	Mg2	FDOM	H2PO4	NO3	TEMP	SPC	K
	5.3	5.1	2.9	4.7	4.4	4.1	12.0	11.8	11.5
2017 + 2018									
A	TEMP	pH	SPC	TEMP	SPC	DO	SPC	TEMP	NTU
	14.8	13.3	11.3	33.4	31.8	23.5	69.8	43.5	41.3
B	Chl	FDOM	Phyco	FDOM	Phyco	Chl	Chl	FDOM	Phyco
	33.5	14.0	6.6	42.4	40.6	38.9	81.3	79.8	58.4
AB	Chl	FDOM	TEMP	TEMP	SPC	DO	SPC	Chl	DO
	16.9	8.1	6.8	23.1	21.0	16.5	47.7	34.5	28.1

Table 4.4. Most influential predictors and increase in mean square error for that variable for Pond 2 as determined using the random forest algorithm.

Pond 2 Random forest most influential predictors									
Input Group	Green Algae			Diatoms			Cyanobacteria		
	Imp Var 1	Imp Var 2	Imp Var 3	Imp Var 1	Imp Var 2	Imp Var 3	Imp Var 1	Imp Var 2	Imp Var 3
2017									
A	pH	SPC	TEMP	pH	SPC	DO	SPC	NTU	pH
	22.5	19.9	6.4	24.1	13.3	11.6	88.6	49.9	16.0
B	FDOM	Phyco	Chl	FDOM	Phyco	Chl	Phyco	Chl	FDOM
	26.3	20.4	8.3	29.9	19.1	12.3	118.2	32.4	24.3
AB	SPC	pH	TEMP	pH	SPC	DO	SPC	NTU	Phyco
	16.7	16.5	6.8	15.7	11.2	9.8	60.6	36.6	33.4
2018									
A	TEMP	SPC	NTU	SPC	TEMP	PAR	SPC	TEMP	NTU
	21.9	19.2	13.6	10.6	10.2	10.1	16.7	10.9	7.9
B	Chl	CDOM	FDOM	CDOM	FDOM	Chl	FDOM	CDOM	Chl
	31.6	22.5	16.3	23.5	18.1	16.7	29.0	16.0	15.1
C	K	Mg2	NH4	Mg2	Ca2	H2PO4	Mg2	K	H2PO4
	26.0	18.3	12.3	15.9	13.5	13.0	15.3	13.9	12.7
AB	TEMP	SPC	Chl	SPC	TEMP	CDOM	SPC	FDOM	TEMP
	15.4	12.7	10.3	8.1	7.5	7.3	11.3	10.0	7.8
AC	K	TEMP	Mg2	Mg2	H2PO4	Ca2	SPC	Mg2	TEMP

	13.5	10.2	9.9	7.8	6.2	5.9	10.3	7.3	6.6
BC	K	Mg2	NH4	Mg2	H2PO4	Ca2	FDOM	Mg2	Chl
	19.8	13.9	11.4	11.5	8.8	8.6	11.7	11.5	7.8
ABC	K	NH4	NO3	Mg2	Ca2	H2PO4	SPC	FDOM	Mg2
	10.2	9.2	8.8	7.1	5.3	4.8	8.1	6.9	6.2
2017 + 2018									
A	pH	SPC	TEMP	SPC	DO	TEMP	SPC	NTU	TEMP
	30.7	25.7	25.6	52.6	38.8	26.6	104.5	59.3	45.3
B	Phyco	Chl	FDOM	FDOM	Phyco	Chl	Phyco	Chl	FDOM
	41.9	38.9	26.8	63.0	60.1	40.0	142.7	78.3	63.0
AB	pH	TEMP	SPC	SPC	DO	TEMP	SPC	Phyco	NTU
	22.9	19.9	19.1	39.5	26.6	21.4	65.9	59.6	35.6

4.4 Discussion

Earlier work by Smith et al. (2020, 2021) demonstrated the correlation between several basic water quality parameters and cyanobacteria populations, as well as the temporal stability of phytoplankton populations within these ponds. Here the relationship between more complex water quality parameters and phytoplankton groups were examined with machine learning. Phytoplankton functional group concentrations in the two agricultural irrigation ponds in this study did not vary greatly nor were the community compositions significantly different, both representing communities of eutrophic, shallow, small-bodied waters. Average diatom and green algae concentrations were similar between years and the two ponds. Despite the routine application of the algicide copper sulfate during the study, phytoplankton concentrations in Pond 1 were comparable to those reported in regional (Marshall, 2013, 2014) and global lakes (Dembowska et al., 2018; Jia et al., 2019). Pond 2 had recurrent cyanobacteria blooms during the study making the phytoplankton concentrations more comparable to those reported in small lakes by Lee et al. (2015) and in local waters by Tango & Butler (2008). Pond 2 phytoplankton

concentrations were slightly higher than Pond 1 concentrations and can potentially be explained by routine algicide use in Pond 1. All three phytoplankton populations in Pond 1 were greater in 2017 than 2018, whereas the opposite was true for Pond 2, except for diatom concentrations which were slightly higher in 2017 than 2018.

Root mean square errors (RMSEs), a metric used to evaluate model performance, for the 2017, 2018, and 2017+2018 models (sets A and AB) varied depending on phytoplankton group. Green algae models tended to have the best performance, followed by diatoms, and then cyanobacteria. In a review by Shimoda & Arhonditsis (2016), green algae were found to have the least error of the three phytoplankton groups similar to the results in this study. Cyanobacteria models had higher RMSEs than green algae models in both our findings and those reviewed by Shimoda & Arhonditsis (2016). This could be explained by the natural spatial and temporal variability of cyanobacteria blooms making accurate population predictions more challenging (Beversdorf et al., 2017; Smith et al., 2021). While various types of models were used in the review by Shimoda & Arhonditsis (2016), the RMSEs from this work indicate that the random forest model is a superior model for predicting green algae when compared to the diatom and cyanobacteria models. Additionally, in the work of Di Maggio et al. (2016) where the same three functional groups were studied, cyanobacteria were found to have the least accurate model performance during peak biomass periods. However, Thomas et al. (2017) noted that cyanobacteria were more predictable than diatoms and green algae across many time scales in an alpine lake. Both ecosystem type and available input variables appear to affect the comparative performance of the random forest algorithm in predictions of

phytoplankton functional groups. The robustness of the model during the growing seasons is characterized by the RMSE values presented in this paper since these RMSEs are averages over the datasets used for testing by the random forest algorithm. Since this study only focused on assessing the accuracy of the prediction model in agricultural irrigations ponds during the growing season (May – October), when waters were used for irrigation purposes and when cyanobacteria biomass, and subsequently risks from cyanotoxins, was expected to be greatest in this region (Marshall et al., 2005; Tango & Butler, 2008), to better assess this model's performance in comparison to similar models additional training and validation needs to be done using data collected outside of the growing season and in varying waterbody types. In this study, sampling was conducted during periods of time between rainfall events, when irrigation is more likely to take place due to crop production demands, elevated temperatures, and reduced soil moisture (Paul et al., 2021). To better equip this model for prediction during all weather conditions and all seasons, additional sampling and training of the model would be necessary.

Model performance did not differ drastically between years. The exception to this is for cyanobacteria predictions wherein RMSE values decreased substantially from 2017 to 2018, indicating better performance of the 2018 models. In Pond 1, models predicting diatoms and cyanobacteria performed better in 2018 compared to 2017. Similarly, in Pond 2, better model performance in 2018 was seen for cyanobacteria predictions and to a lesser extent, diatom predictions. Furthermore, the combined 2017+2018 datasets had higher RMSE values than when using just the 2018 dataset, but lower than when only the 2017 dataset was used. For all three

groups and both sets of parameters (A and AB), 2018 had the best model performance as indicated by the lowest RMSEs. Thomas et al. (2017) found that multiyear datasets were able to produce reasonable performance and attributed it to the model having more data points to train the machine learning algorithm with. Our individual years had fewer data points than the combined year models. While 2018 had the lowest RMSE values of the three data sets, the use of 2017+2018 caused a decrease in RMSE values for 2017. Furthermore, it was determined that the prediction of 2019 data was not as accurate as the prediction of the 2017 and 2018 years. Additional monitoring would help to determine if the model performance of future years is comparable to the accuracy represented in the 2017 and 2018 evaluations.

The addition of organic constituent-related input parameters did not improve model performance overall. While some aspects of the model saw a small increase in performance, others saw a small decrease and no general pattern could be defined. This follows many other studies that showed the use of inputs, similar to this study's set A parameters (DO, pH, NTU, and TEMP), tended to be most important and produced the best prediction results (Fragoso et al., 2008; Huang et al., 2014; Liu et al., 2019). According to Rigosi et al. (2010), a model based on water quality physical parameters often has superior performance and this was attributed to the high level of complexity found in biological processes. Likewise, while the nutrient and macro element parameters in input set C were highly influential when evaluating the 2018 data, the difference in model performance across phytoplankton groups may be due to the complex and interrelated way each phytoplankton group utilizes different nutrients and macro elements (Bradshaw et al., 2012; Finkel et al., 2010), which was

not captured with just one year of training data. The presence of short blooms of both nitrogen-fixing and non-nitrogen-fixing cyanobacteria in the study area (Smith et al., 2021), which can utilize different forms of nitrogen and impact the overall nitrogen budget (Agawin et al., 2007; Newell et al., 2019) also may not have been equitably represented in this dataset. However, the ability to use the random forest algorithm to predict phytoplankton functional groups using only set A inputs is beneficial for a wide range of resource monitoring and research applications since the set A input parameters are often the least expensive and easiest parameters to collect, thus predictions can be quickly and easily done.

Overall, spatial distributions of RMSE values differed based on phytoplankton functional group. Green algae had the lowest spatial average RMSEs (P1=0.278, P2=0.356), cyanobacteria had the highest spatial average RMSEs (P1=0.567, P2=0.679), and average RMSEs for diatoms were in between (P1=0.446, P2=0.578) for both ponds and models. This indicates that the set A and AB models were the most accurate in predicting the spatial green algae concentrations for the 2017+2018 dataset. In general, interior waters tended to exhibit the lowest RMSE values in both ponds and for models with both inputs sets A and AB showing that the random forest algorithm predicted interior concentrations of green algae best, followed by diatoms and cyanobacteria. In a prior study, on the temporal and spatial variability of phytoplankton functional groups within these two agricultural irrigation ponds, it was established that interior waters tended to be less variable than nearshore waters (Smith et al., 2021). This stability allows the model to better predict the phytoplankton community structure in those locations. Variations in phytoplankton

concentrations tended to be greater in nearshore samples when compared to interior waters using an assessment of CV. In over 50% of the sampling dates, CVs were higher for nearshore samples except for green algae in Pond 1. Similarly, water quality CVs in both ponds were almost always higher for nearshore locations, with most nearshore variability being higher in 75% or more of the sampling dates. This pattern was also observed in the study by Awada et al. (2021), for marine waters; the model developed by these authors performed best in open water locations of the Gulf of Sirte and had poorer agreement between measured and simulated concentrations of chlorophyll-*a* along the shoreline. In Lake Taihu, locations closer to the shoreline tended to have higher simulation errors than central lake locations (Huang et al., 2014). However, in a study on Lake Okeechobee the random forest algorithm had better model results at nearshore locations as opposed to pelagic locations and Zhang et al. (2021) attributed this to poor phytoplankton growth in the pelagic zones caused by wind-driven sediment resuspension.

For all three phytoplankton groups, there was almost no change in RMSE values from models run using set A parameters to models run using set AB parameters indicating that the additional parameters did not impact the predictive abilities of the random forests. The ability of the random forest model to predict phytoplankton community structure or chlorophyll-*a* concentrations accurately on set A parameters (TEMP, pH, NTU, and SPC) alone has been noted in several other studies (Derot et al., 2020; X. Liu et al., 2019; Zeng et al., 2017). Whereas other studies (Cheng et al., 2021; Yajima & Derot, 2017) found that biological parameters (biological oxygen demand, chlorophyll-*a* concentrations) were more important for

phytoplankton prediction models. Biological oxygen demand was not measured in this study but should be considered for future modeling efforts as it is known to be spatially and temporally variable in lake waters (Carpenter et al., 1979; Wang et al., 2007) and can be positively correlated with potentially toxigenic cyanobacteria species (Karadžić et al., 2013) and overall phytoplankton biomass (Wang et al., 2007), both of which are of concern to agricultural resource managers.

Overall, this study found that the most important variables tended to be set A parameters (TEMP, pH, NTU, and SPC) for both ponds. TEMP was determined to be the most recurrent parameter in the top three most influential parameters for all groups and both ponds. This is comparable to numerous other random forest models used for phytoplankton prediction (Derot et al., 2020; Kehoe et al., 2015; Liu et al., 2019; Rousso et al., 2020; Yajima & Derot, 2017). Other set A parameters which were also reported in the top three most influential parameters, but to a lesser degree than TEMP in this study were SPC, NTU, pH, and DO. SPC appeared to be the most influential predictor in input set A. A possible reason for this could be the correlation between SPC and nutrient ion concentrations in agricultural waters (Taboada-Castro et al., 2004) and intercoupled relationship between specific nutrient forms and concentrations and phytoplankton groups (Varol, 2019).

The only instance when set A parameters were not the most influential parameters was in 2018 when nutrients (input set C) were measured and used as inputs. Nutrients being the most influential or important parameters is in line with numerous assessments of phytoplankton community structure using random forest algorithms. Dunker et al. (2016) found a strong relationship between orthophosphate,

nitrogen, and chlorophytes. Total nitrogen, total phosphate, nitrate, and nitrite were identified as the most important predictors in phytoplankton models used in Lake Okeechobee (Zhang et al., 2021) and Lake Taihu (Wang et al. 2007). However, these studies took place in lakes considerably larger than the ponds studied here. Small waterbodies (<12 acres), which are increasing used in agricultural practices in the Mid-Atlantic region, often have a greater biodiversity than larger bodies of water and can experience more climatic stress (Brönmark & Hansson, 2002; Chopyk et al., 2018), highlighting the need to refine models to local conditions. Since nutrients were only measured in 2018, these parameters were not included in the 2017 nor combined year models. The modeling robustness of nutrient parameters, when compared to set A parameters, has yet to be determined for ponded agricultural waters. It should be noted that the collection and laboratory processing of nutrient samples can be a labor-intensive and costly process that has successfully been augmented with modeling for riverine systems (Harrison et al., 2021; Leigh et al., 2019). However, assessments similar to those outlined by Harrison et al. (2021) and Leigh et al. (2019) in agricultural irrigation waters have not been conducted and should be the focus of future model developing and training efforts.

In a review of predictive and forecasting models for cyanobacteria by Rousso et al. (2020) it was found that parameters similar to this study's input set A (TEMP, DO, pH) were reported as the most influential predictors in 38.5% of publications surveyed. Nutrients were reported as most influential parameters in 30.5% of the total publications surveyed. One of the least influential predictors reported (6% of publications) were similar to the parameters included input set B (FDOM, CHL,

Phyco) which is comparable to the findings in this study. As noted in a cyanobacteria research forecast by Burford et al. (2020) future modeling efforts should incorporate CO_2 dynamics that will reflect future climate scenarios, temporally relevant weather patterns, as well as the intricate relationship cyanobacteria have with the food web, all factors which ultimately will influence agricultural irrigation water quality.

Physiochemical parameters being the most important predictors for the three major phytoplankton groups is beneficial for water quality management. Enumeration of phytoplankton is time intensive, requires highly trained staff, and/or expensive infrastructure (Lawton et al., 1999; Stauffer et al., 2019; Zeng et al., 2017), whereas parameters such as temperature, dissolved oxygen, pH, conductivity, and turbidity can be easily and affordably measured in real time with an in-situ sensor. The quick acquisition and input of these parameters into a modeling application allows for the prediction of major phytoplankton groups by machine learning algorithms to be performed by a broader group of individuals that could lead to more timely alerts of potentially harmful phytoplankton species.

4.5 Conclusions

The prediction and estimation of phytoplankton functional groups in two working agricultural irrigation ponds was feasible with machine learning methodology and the random forest algorithm. Random forest models predicting green algae were found to be superior when compared to diatom and cyanobacteria predictions. The RMSE values of the model obtained with two years of data were in between the RMSE values obtained with data from individual years. When model performance was mapped, interior sampling locations tended to have lower model

error than nearshore sampling locations indicating that phytoplankton predictions are more accurate for interior waters compared to nearshore waters. Furthermore, minimal differences in model performance were seen when additional input set B parameters were added. Models using physical parameters (input set A) tended to be the models with the best performance. Physical parameters (TEMP, pH, DO, SPC, and NTU) were the most frequent influential predictors for the random forest algorithm allowing water quality managers to potentially bypass the use of time intensive and expensive monitoring procedures for those which can be obtained easily, affordably, and in real time. Development of machine learning models to provide site-specific estimates of phytoplankton groups from more easily obtainable water quality parameters presents a promising research avenue to explore.

Chapter 5: Conclusions

Phytoplankton form a ubiquitous microbial community that is often studied to assess ecological functioning, water quality, and human health risks associated with drinking, recreational, and aquaculture waters. Prior research on phytoplankton has been performed on various fresh and marine waterbodies, but very little work has been performed on agricultural irrigation ponds. A better understanding of a phytoplankton community's dynamics and environmental relationships allows for more efficient and feasible management practices and prediction capabilities to be implemented.

This work was performed to (a) reveal and quantify the dynamics of spatiotemporal phytoplankton in agricultural irrigation ponds, and (b) relate those dynamics to the spatiotemporal variability of water quality covariates. An initial goal was to determine if there were stable spatial patterns in phytoplankton concentrations across ponds over time. We then researched spatial variability of the cyanobacteria-related pigment phycocyanin and its relation to other measured water quality parameters. Finally, machine learning was employed to model the influence of environmental controls on the phytoplankton communities and attempted to predict or estimate phytoplankton community structure.

Chapter 2 presents the results on temporally stable spatial patterns of phytoplankton functional groups within two agricultural irrigation ponds in Maryland, USA. Conclusions are as follows.

Stable spatial patterns were established for green algae, diatoms, and cyanobacteria in both agricultural irrigation ponds. Zones which were consistently

higher or lower than the pond's average phytoplankton concentrations were found for all phytoplankton groups and in both ponds. Two main patterns were detected, establishing zones as either a cluster of locations which had consistently higher or lower concentrations of phytoplankton, or locations having consistently higher or lower concentrations at nearshore or interior sampling locations.

Spatiotemporal patterns had implications on water quality monitoring design and implementation. An agricultural irrigation pond cannot be assumed to contain a homogenous mixture of phytoplankton. A water sample may drastically underestimate or overestimate the concentrations of phytoplankton if it is collected only in one location of the pond or in locations where the phytoplankton concentrations are consistently lower or higher than the average across the pond. This spatial heterogeneity also indicates that the location of the irrigation pump intakes should not be arbitrarily chosen. The intake location may affect the nearby fields and crops by providing water that may contain toxins produced by cyanoHAB species.

Moderate and strong correlations were found between the spatial patterns of phytoplankton group concentrations and measured water quality parameters in both studied irrigation ponds. Existence of such correlations suggests that in-situ water quality sensing technologies may be used to monitor irrigation ponds providing a timesaving, cost-effective alternative to intensive microscopy analysis. This alternative can deliver near-instantaneous results to monitoring irrigation water sources for their quality in terms of phytoplankton-related parameters.

The findings in this study represent a specific region in the Mid-Atlantic United States. No extrapolations should be made to other irrigation ponds and other

agricultural ponds such as retention and aquaculture ponds. There is a need to investigate the spatiotemporal variations in the phytoplankton community in other agriculture-related ponds. It also remains to be seen if the diurnal dynamics in phytoplankton communities in such ponds can be strong enough to affect phytoplankton concentrations and community structure and correlations between them, water quality parameters, and phytoplankton attributes. The existence of phytoplankton-related toxicity should be determined at the diurnal scale, as it may have implications for irrigation scheduling.

Chapter 3 reports data and analyses on intraseasonal variations of phycocyanin – a pigment which is often used to indicate cyanobacteria presence in waters. Concentrations of phycocyanin and water quality covariates were studied in two agricultural irrigation ponds in Maryland, USA. The research led to the following conclusions.

The machine-learning algorithm of decision trees assisted in determining water quality variables that were important predictors for the observed concentrations of phycocyanin. These variables were: extracted chlorophyll, colored dissolved organic matter, and turbidity. Average quartile ranks were instrumental in highlighting locations within the ponds that had consistently higher or consistently lower phycocyanin concentrations. Zones of predominantly low and predominantly high phycocyanin concentrations were found in both agricultural irrigation ponds studied in this work.

Monitoring of agricultural irrigation ponds for phycocyanin may assist in early detection of cyanoHABs thus reducing food safety risks associated with cyanotoxins. Zones of consistently higher or lower phycocyanin concentrations must be sampled to represent the entire waterbody. Determining spatial patterns and zones of phycocyanin can assist in informed and efficient water quality monitoring for cyanobacteria.

Understanding the relationships between phycocyanin and easily measured water quality parameters may assist in developing better and more efficient pond management practices along with the ability to better predict potentially toxic cyanobacteria blooms. Measuring phycocyanin might be a more economical option for water quality management. Further improvements in understanding which water quality parameters are the most influential variables for phycocyanin and cyanobacteria may give rise to even more cost-effective and faster alternatives to microscopy and/or toxin analysis for cyanobacteria and cyanotoxin early detection and monitoring.

Future research should consider that detecting and quantifying phycocyanin cannot differentiate between toxic and non-toxic cyanobacteria species. In this study, there were impacts to both irrigation ponds following rain events, but finer scale temporal monitoring is needed to put these findings in the context of cyanobacteria blooms. Determining the intraseasonal variations and spatial patterns of toxin-producing cyanobacteria species in irrigation ponds is essential to food safety and water quality monitoring. Research of the spatiotemporal variation of phycocyanin

may help contribute to the proper assessment of the human health and food safety risks associated with cyanobacteria populations in agricultural irrigation waters.

Chapter 4 contains an example of the application of machine learning to estimate phytoplankton concentrations in agricultural irrigation ponds from water quality measurements. Conclusions are as follows.

The machine learning algorithm 'random forest' was capable of predicting and estimating phytoplankton functional groups within irrigation ponds. The green algae models performed the best compared to the diatoms and cyanobacteria models. Modeling with multiyear in-situ data had a degree of accuracy comparable with individual year models. Sampling locations in the interior of the ponds tended to have the least model errors compared to estimations near the shorelines.

Model accuracy was compared with two input data sets, one data set with physicochemical parameters (pH, dissolved oxygen, temperature, conductivity, and turbidity) and the other data set with physicochemical and organic constituent-related measurements (chlorophyll-*a*, fluorescent dissolved organic matter, and colored dissolved organic matter). Models with the physicochemical water quality inputs tended to produce superior results to those models with both physicochemical and organic constituent-related water quality input datasets. The random forest algorithm showed that physicochemical parameters were the most influential predictors when both input parameter groups were used together.

Physicochemical parameters can be obtained easily and quickly in-situ from water quality sensors. Physicochemical data inputs were the most influential

predictors and offered the best model performance indicating that these parameters are useful to predict and estimate the phytoplankton groups in agricultural irrigation ponds. This procedure could allow water quality managers to avoid costly and time-consuming monitoring procedures, such as imaging flow cytometry or microscopy analysis, for the identification and enumeration of the phytoplankton communities that are needed to make risk assessments regarding cyanoHABs and cyanotoxins in irrigation waters. Indications are that physicochemical parameters examined with the use of random forest models allow for a quick, affordable, and straightforward process for the prediction and estimation of phytoplankton functional groups within agricultural irrigation ponds.

The distribution of phytoplankton within these agricultural irrigation ponds was not random. Consistent temporal patterns were present, and the variations in phytoplankton concentrations, or their pigments, was reflected water quality variations. Future research should look at other in-situ sensed variables as potential inputs to improve phytoplankton estimation models. In addition, other machine learning algorithms should be investigated and compared with the currently used models. It should be noted that this study represents the use of two farm irrigation ponds within the Mid-Atlantic region of the United States. Future longitudinal research is needed before applying the machine learning methodologies described here to other irrigation ponds located in different areas of the United States. Developing a more general model, applicable to a wider range of irrigation water sources, is a necessary avenue of research as climate change alters the timing and available water resources for farming practices.

Appendices

Appendix A: Supplementary material for Chapter 2

Supplemental Table 1. Sampling dates from 2017 and 2018 with corresponding rainfall event information for Pond 1 and Pond 2.

Dates	# of days from last rainfall to sampling date	Total days of rainfall	Total accumulation (cm)
Pond 1			
6/7/17	0*	2	.25
6/20/17	0*	1	.89
7/5/17	0*	1	.94
7/18/17	0*	1	5.08
8/2/17	3	2	10.85
8/15/17	0*	4	4.70
6/20/18	0*	1	0.76
7/5/18	14	3	1.75
7/19/18	2	1	1.85
8/15/18	1	2	.25
8/29/18	7	2	6.27
10/4/18	6	7	11.63
Pond 2			
5/31/17	1	3	0.91
6/13/17	7	1	0.10
7/11/17	4	1	1.45
7/25/17	0*	3	1.12
8/8/17	0*	1	16.76
6/14/18	3	3	1.98
6/26/18	2	3	2.29
7/10/18	17	3	2.29
8/7/18	3	6	2.18
8/23/18	2	1	0.99
9/20/18	2	2	1.35

* there was rain the day before sampling occurred.

Supplemental Table 2: Time series data of 2017 water quality parameters for Pond 1 and Pond 2.

2017	Pond 1						Pond 2				
	6/7/17	6/21/17	7/5/17	7/18/17	8/2/17	8/16/17	5/31/17	6/13/17	7/11/17	7/25/17	8/8/18
Temp	22.73	27.14	28.00	29.14	25.44	25.43	24.52	30.73	29.75	29.61	22.24
DO	10.52	9.12	11.92	11.02	10.16	9.58	16.94	14.49	11.98	10.33	5.35
SPC	156.40	168.85	170.20	175.82	150.74	153.10	151.49	160.10	174.77	179.11	141.57
pH	8.51	8.70	9.39	8.86	8.64	8.79	9.35	8.81	8.20	8.10	6.48
NTU	2.88	4.82	4.06	7.38	6.28	5.54	4.83	8.38	3.85	32.23	4.84
Phyco YSI	0.22	0.32	0.88	2.34	1.17	1.94	0.94	2.86	1.58	11.74	0.33
CHL YSI	1.88	2.34	6.25	7.56	1.68	3.29	8.65	15.10	12.27	15.27	2.39
***f*DOM**	3.42	2.19	4.82	6.24	36.65	25.02	32.57	24.99	23.62	28.26	42.24
CHL EXT	4.69	4.19	15.77	15.52	10.58	20.25	64.26	10.65	25.59	170.44	9.71
Secchi	0.89	1.02	0.80	0.58	0.54	0.51	0.43	N/A	0.70	0.31	N/A

Supplemental Table 3: Time series data of 2018 water quality parameters for Pond 1 and Pond 2.

2018	Pond 1						Pond 2					
	6/20/18	7/5/18	7/19/18	8/15/18	8/29/18	10/4/18	6/14/18	6/26/18	7/10/18	8/7/18	8/23/18	9/20/18
Temp	26.59	29.57	28.25	26.83	27.73	22.97	25.45	27.12	28.64	31.93	25.92	26.21
DO	11.59	7.89	8.96	10.50	11.23	10.23	20.73	16.78	11.11	14.05	11.46	15.33
SPC	141.37	156.44	166.90	146.49	132.82	148.95	136.37	146.25	162.08	124.84	139.56	146.44
pH	8.46	7.30	7.25	7.46	8.06	7.36	8.82	8.49	7.51	8.50	7.37	8.32
NTU	8.80	3.76	2.57	2.47	3.97	3.60	9.30	6.04	10.17	18.68	18.52	20.47
Phyco YSI	2.51	0.60	0.28	0.64	0.91	0.76	6.07	4.16	4.49	4.72	5.03	4.69
CHL YSI	2.42	1.12	1.03	4.55	6.93	0.73	36.64	26.28	12.90	17.01	13.69	13.31
***f*DOM**	21.35	7.11	1.92	29.94	25.58	38.39	39.56	30.01	27.51	40.84	38.75	32.76
CHL EXT	36.75	6.25	5.47	28.88	42.80	12.46	272.44	233.39	82.27	163.02	266.06	157.99
CDOM	85.01	43.37	32.32	95.33	102.56	102.19	211.71	166.18	133.34	203.32	188.46	158.97
LAB CHL	223.80	87.49	94.74	399.87	638.30	90.39	1182.11	1393.71	652.03	1183.12	1286.87	1053.45
Phyco LAB	114.28	24.84	18.58	18.23	31.67	47.92	110.49	68.07	119.90	141.60	110.49	156.53
Secchi	1.07	0.95	1.42	0.98	0.89	1.15	0.56	0.50	0.58	0.41	0.45	0.52

Pond 1 was treated with copper sulfate on 7/1/18.

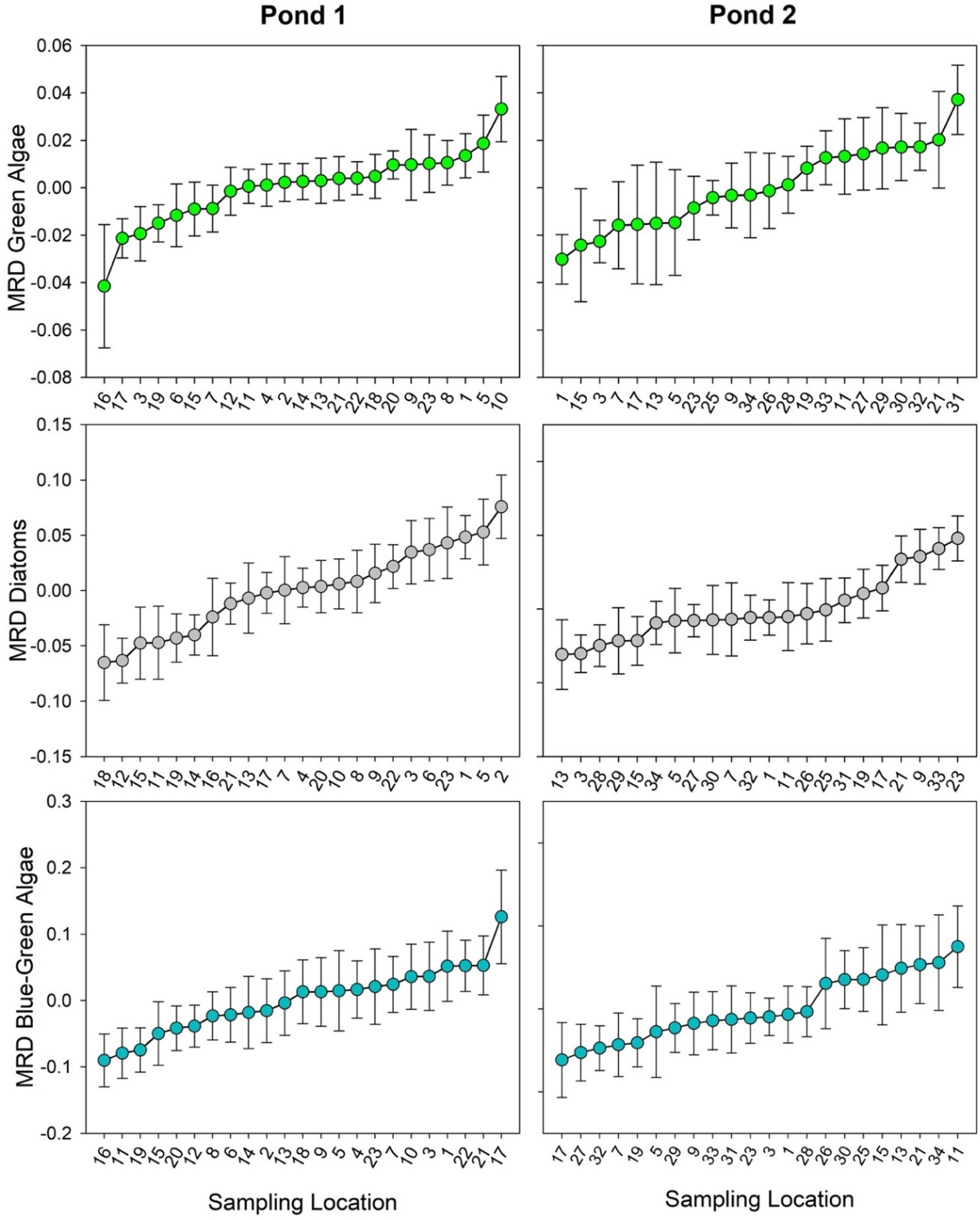
Supplemental Table 4: Descriptive statistics of phytoplankton functional groups.

Time Series Data – Mean, standard deviation, and number of samples taken									
Group	**Green Algae**			**Diatoms**			**Cyanobacteria**		
Dates	**Mean**	**Std Dev**	**# Samples**	**Mean**	**Std Dev**	**# Samples**	**Mean**	**Std Dev**	**# Samples**
Pond 1									
6/7/17	7.042	0.353	23	5.469	0.679	23	5.508	1.138	23
6/20/17	6.812	0.192	23	5.887	0.376	23	5.644	0.739	23
7/5/17	7.187	0.206	23	5.588	0.686	23	6.768	0.498	23
7/18/17	7.091	0.469	23	6.564	0.674	23	6.141	0.948	23
8/2/17	7.044	0.286	23	5.969	0.459	23	5.490	1.175	23
8/15/17	7.405	0.179	23	6.181	0.419	23	5.601	1.106	23
6/20/18	6.807	0.276	23	5.798	0.722	23	7.352	0.228	23
7/5/18	6.760	0.234	23	5.919	0.463	23	5.020	1.203	23
7/19/18	6.719	0.214	23	6.172	0.275	23	4.309	0.403	23
8/15/18	7.391	0.173	23	6.110	0.368	23	5.223	1.352	23
8/29/18	6.257	0.319	23	4.662	0.614	23	4.337	0.714	23
10/4/18	6.257	0.319	23	4.667	0.594	23	7.585	0.249	23
Pond 2									
5/31/17	5.836	0.469	22	5.625	0.487	22	5.411	0.917	22
6/13/17	6.598	0.380	22	5.788	0.483	22	5.382	1.233	22
7/11/17	7.513	0.289	22	6.883	0.399	22	5.696	1.003	22
7/25/17	7.590	0.218	22	6.347	1.107	22	8.366	0.229	22
8/8/17	7.051	0.275	22	6.980	0.254	22	5.862	0.869	22
6/14/18	6.243	0.800	22	5.213	0.949	22	7.351	0.885	22
6/26/18	6.810	0.271	22	5.940	0.502	22	6.871	0.489	22
7/10/18	6.875	0.223	22	5.487	0.716	22	7.124	0.328	22
8/7/18	7.818	0.164	22	5.111	0.831	22	6.240	1.392	22
8/23/18	7.587	0.172	22	5.015	0.959	22	7.609	0.338	22
9/20/18	7.323	0.214	22	6.250	0.367	22	7.058	0.629	22

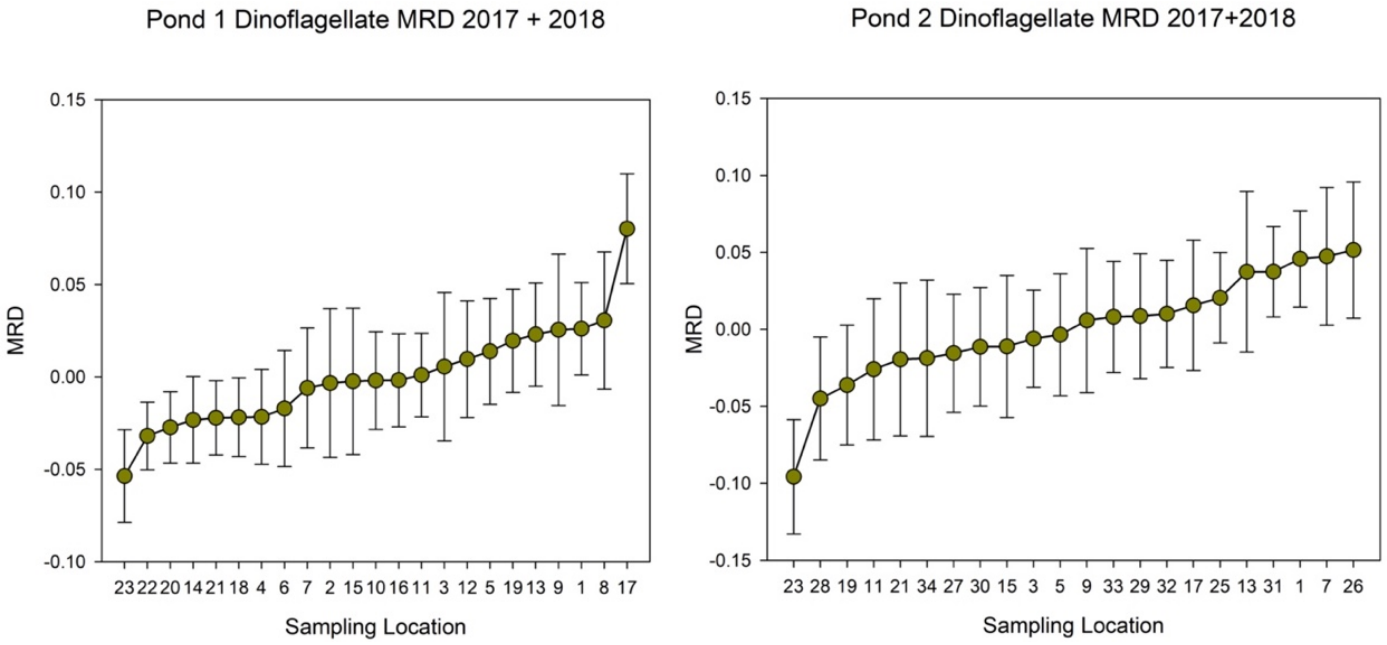
Supplemental Table 5: Coefficient of variation values for Pond 1 and Pond 2

Coefficient of Variation			
Dates	**Green Algae**	**Diatoms**	**Cyanobacteria**
Pond 1			
6/7/17	0.050	0.124	0.207
6/20/17	0.028	0.064	0.131
7/5/17	0.029	0.123	0.074
7/18/17	0.066	0.103	0.154
8/2/17	0.041	0.077	0.214
8/15/17	0.024	0.068	0.197
6/20/18	0.041	0.124	0.040
7/5/18	0.035	0.078	0.249
7/19/18	0.032	0.044	0.217
8/15/18	0.023	0.060	0.262
8/29/18	0.051	0.132	0.212
10/4/18	0.051	0.127	0.066
Pond 2			
5/31/17	0.080	0.087	0.169
6/13/17	0.058	0.084	0.229
7/11/17	0.039	0.058	0.176
7/25/17	0.029	0.174	0.027
8/8/17	0.039	0.036	0.148
6/14/18	0.129	0.139	0.107
6/26/18	0.040	0.092	0.075
7/10/18	0.032	0.111	0.042
8/7/18	0.019	0.140	0.161
8/23/18	0.021	0.143	0.041
9/20/18	0.025	0.060	0.062

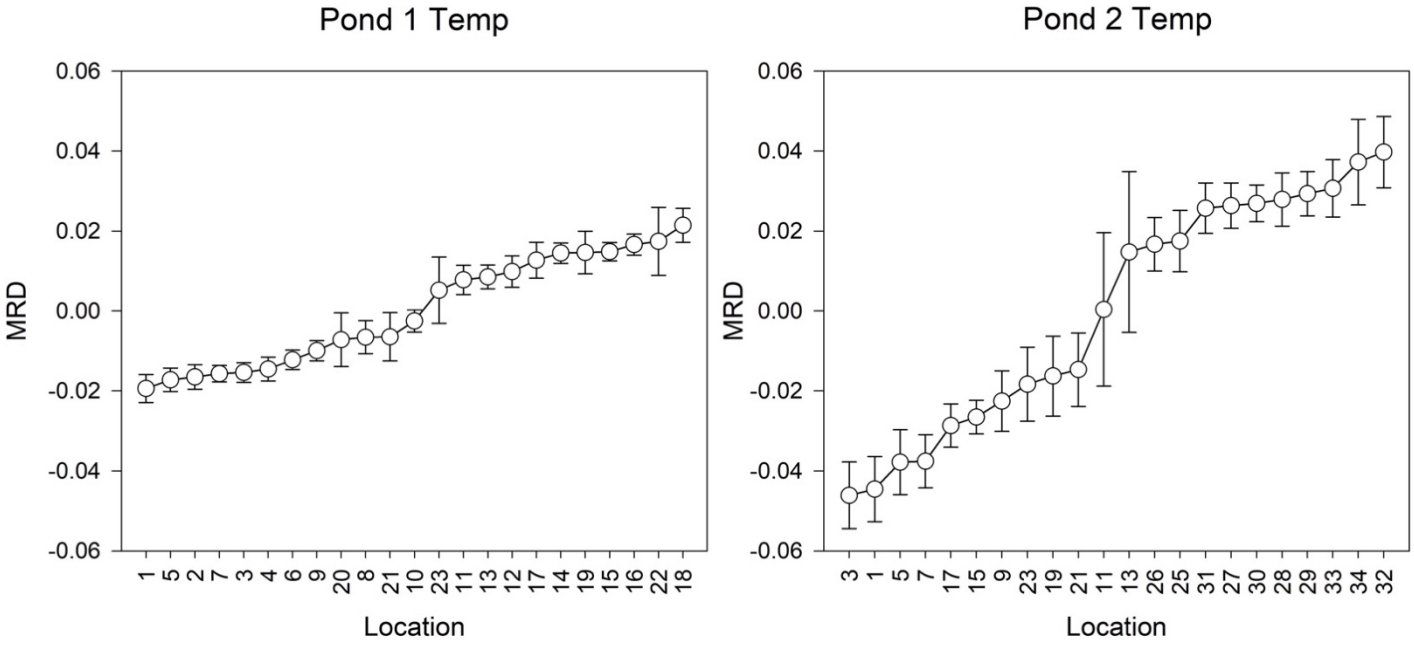
Supplemental Figure 2.1: The mean relative difference values of the logarithms of green algae, diatoms, and cyanobacteria concentrations computed over the two-year period for Pond 1 and Pond 2.



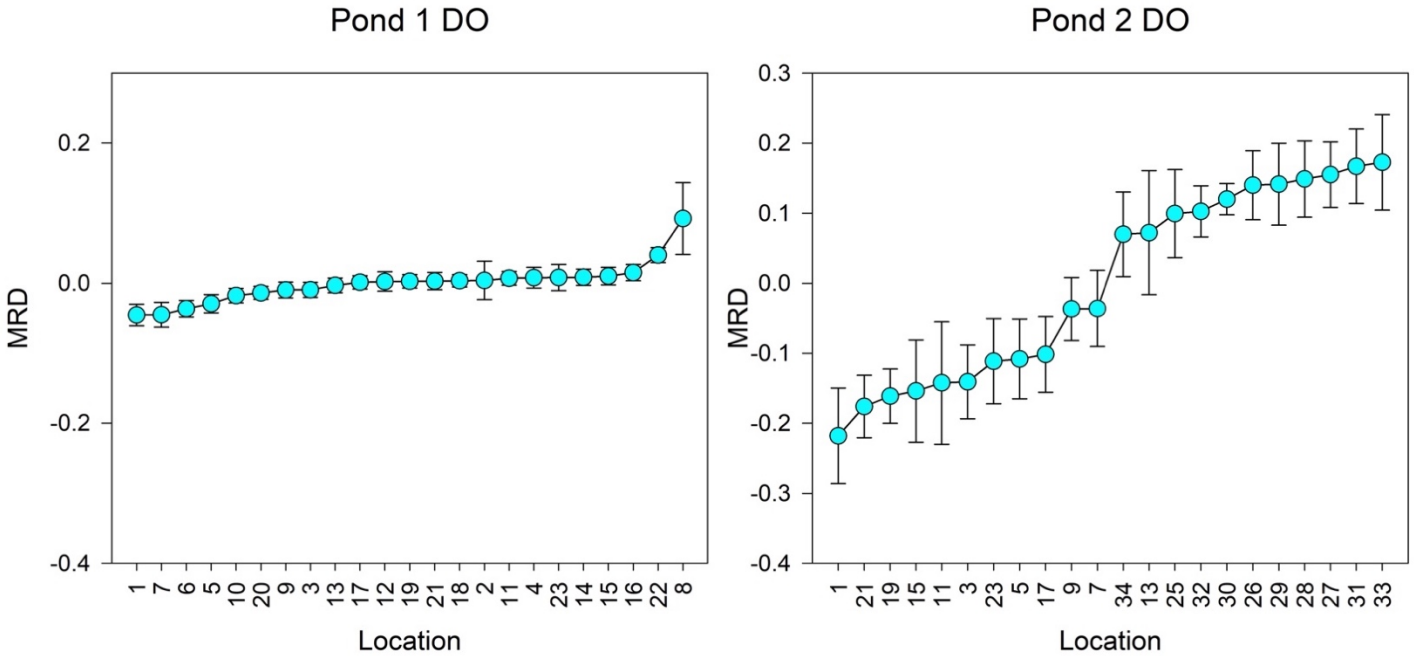
Supplemental Figure 2.2: The mean relative difference values of the logarithms of dinoflagellate concentrations computed over the two-year period for both Pond 1 and Pond 2.



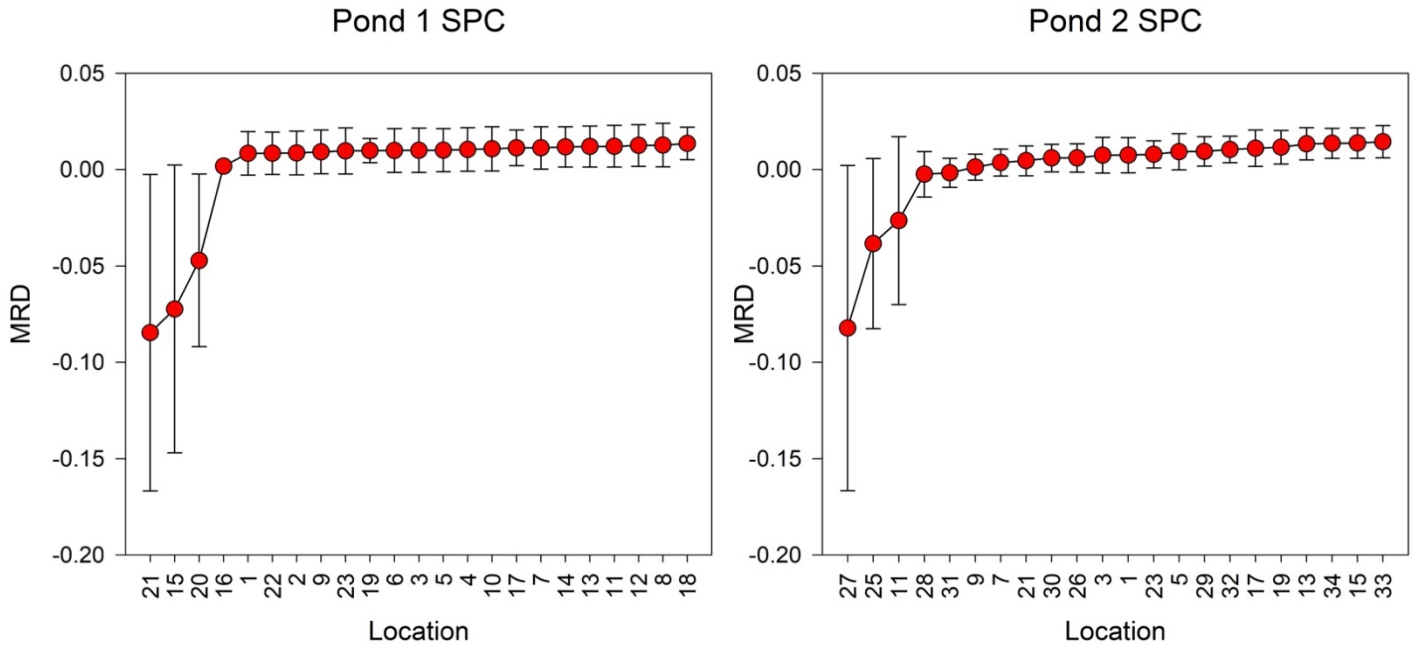
Supplemental Figure 2.3a: The mean relative difference values of temperature over the two-year period for Pond 1 and Pond 2.



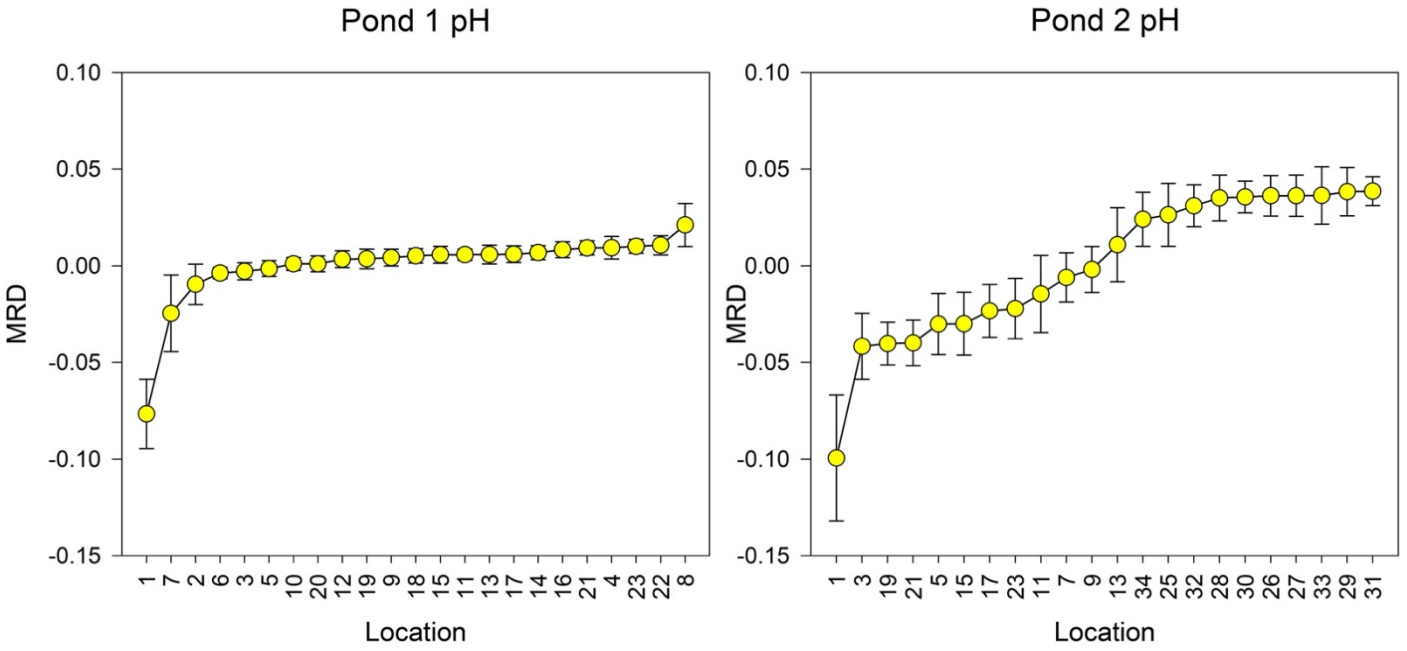
Supplemental Figure 2.3b: The mean relative difference values of dissolved oxygen over the two-year period for Pond 1 and Pond 2.



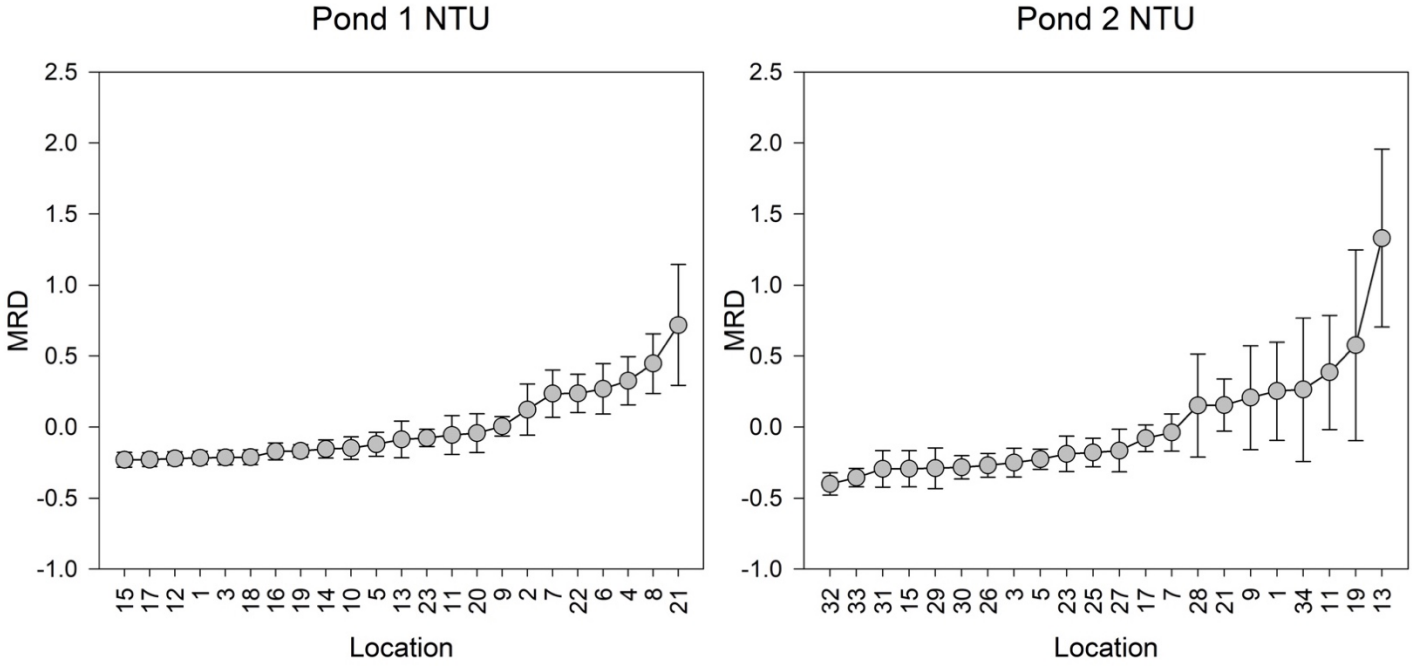
Supplemental Figure 2.3c: The mean relative difference values of specific conductance over the two-year period for Pond 1 and Pond 2.



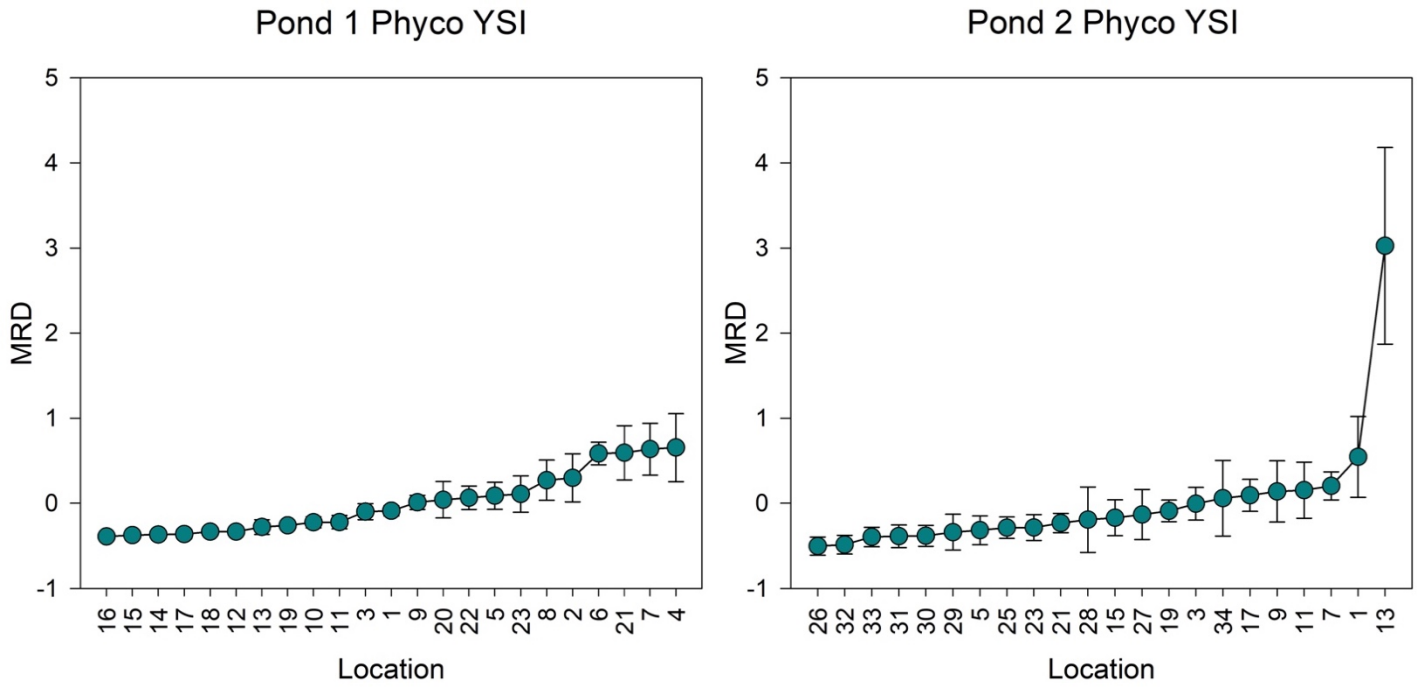
Supplemental Figure 2.3d: The mean relative difference values of pH over the two-year period for Pond 1 and Pond 2.



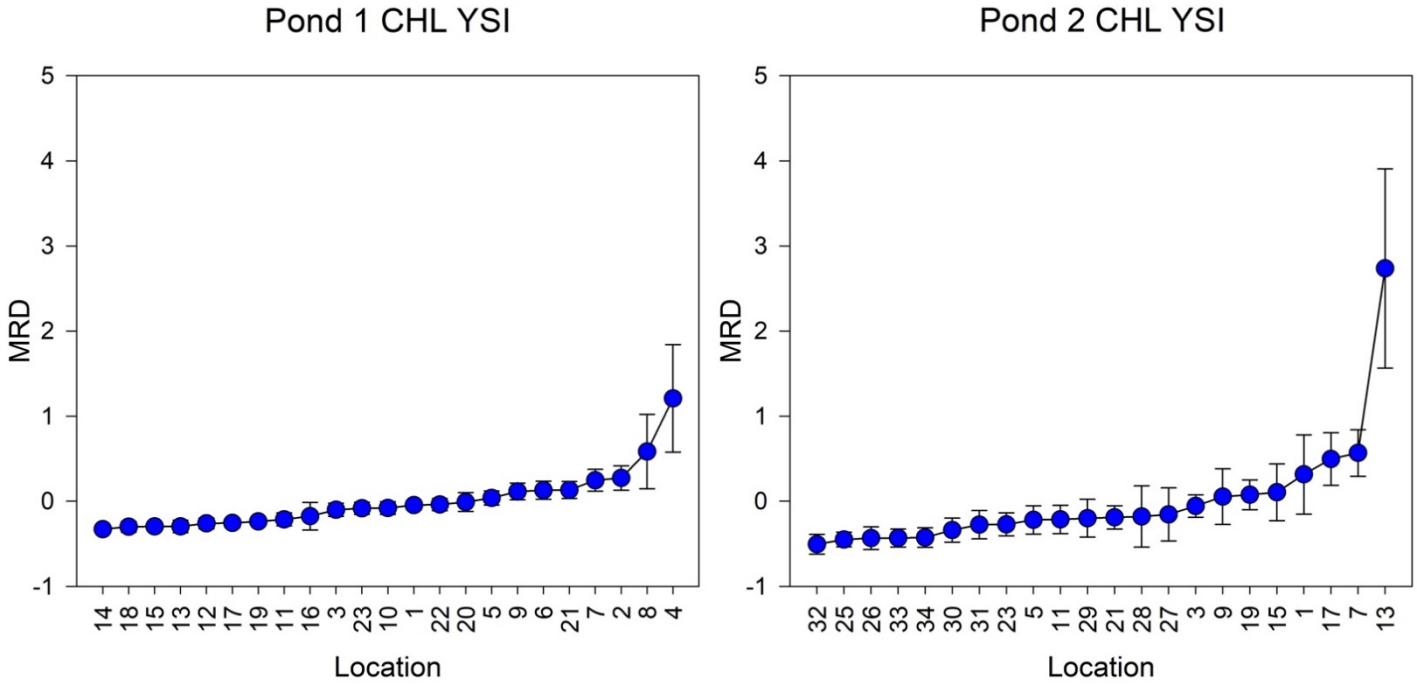
Supplemental Figure 2.3e: The mean relative difference values of turbidity over the two-year period for Pond 1 and Pond 2.



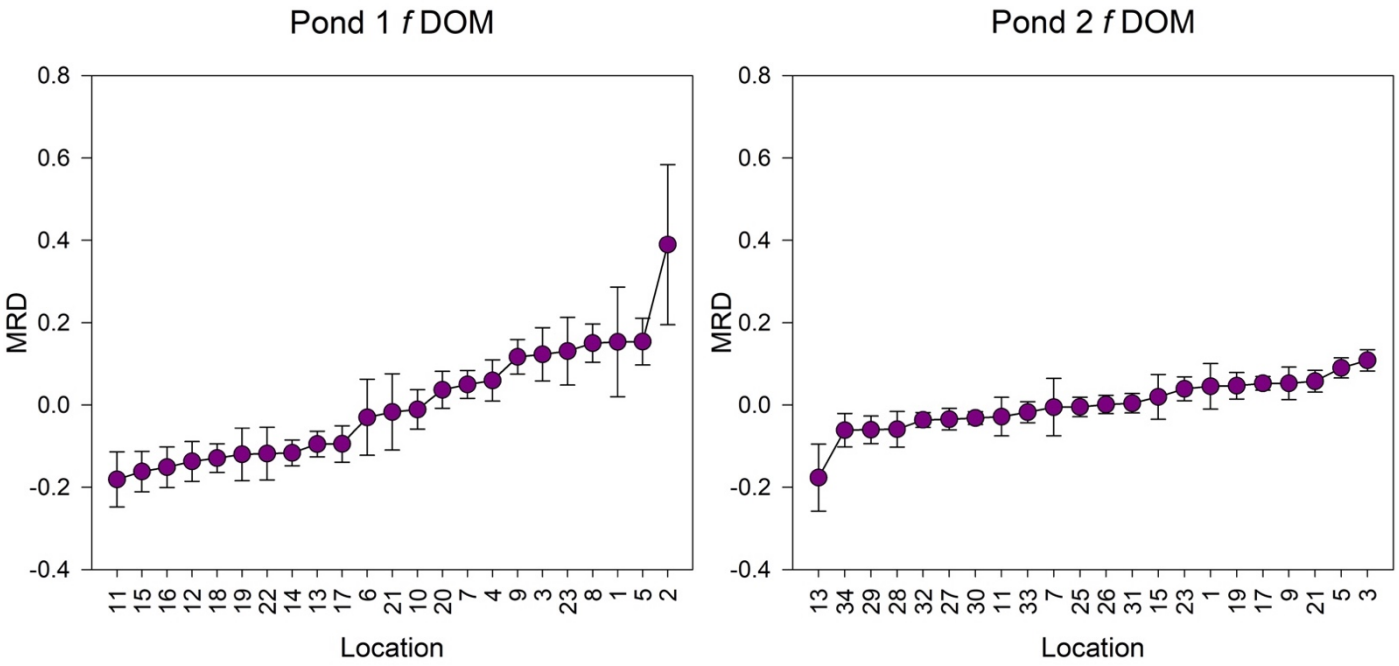
Supplemental Figure 2.3f: The mean relative difference values of phycocyanin from the YSI sonde over the two-year period for Pond 1 and Pond 2.



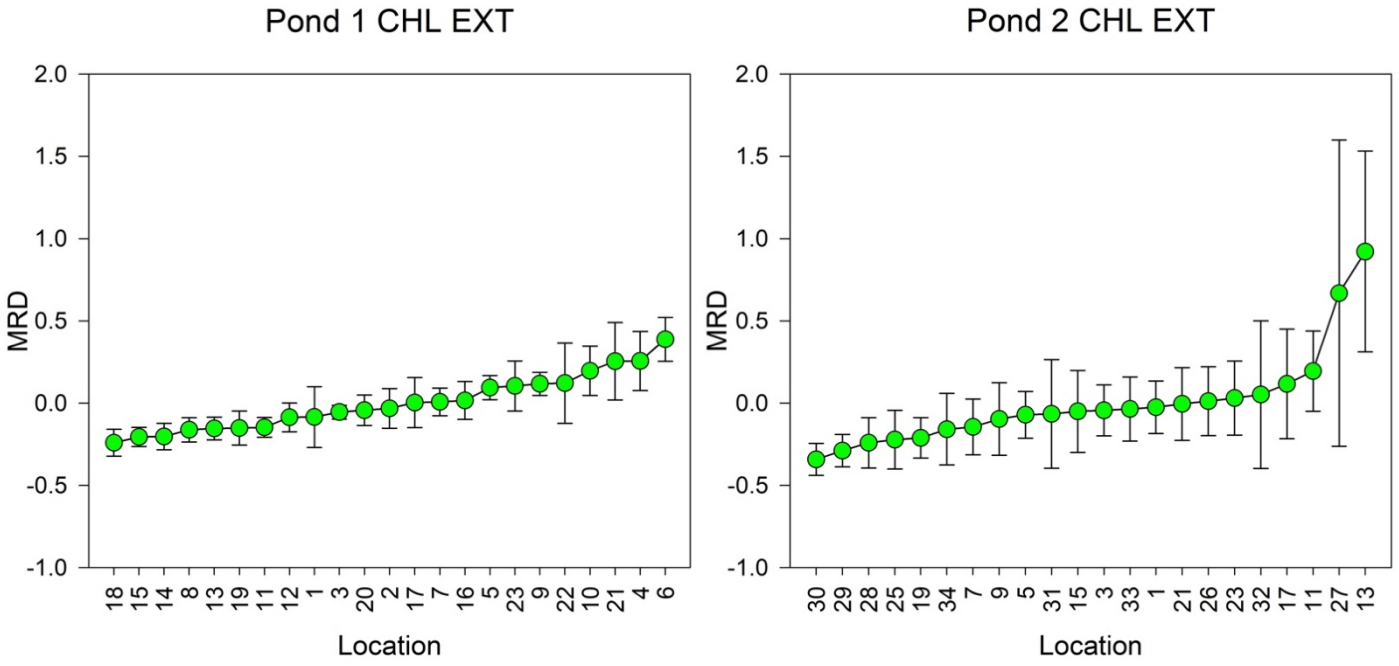
Supplemental Figure 2.3g: The mean relative difference values of chlorophyll pigment from the YSI sonde over the two-year period for Pond 1 and Pond 2.



Supplemental Figure 2.3h: The mean relative difference values of fluorescent dissolved organic matter over the two-year period for Pond 1 and Pond 2.

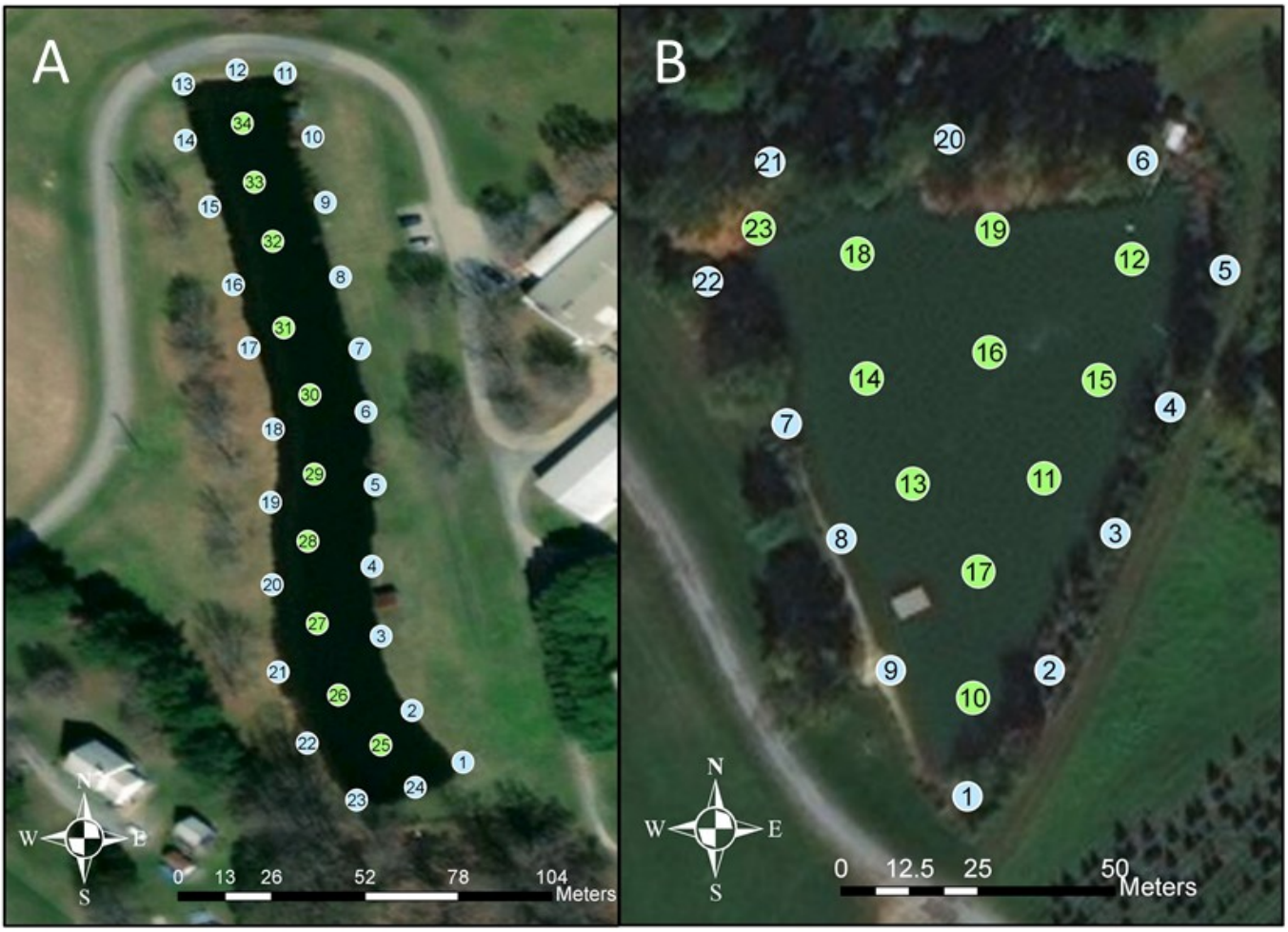


Supplemental Figure 2.3i: The mean relative difference values of extracted chlorophyll over the two-year period for Pond 1 and Pond 2.

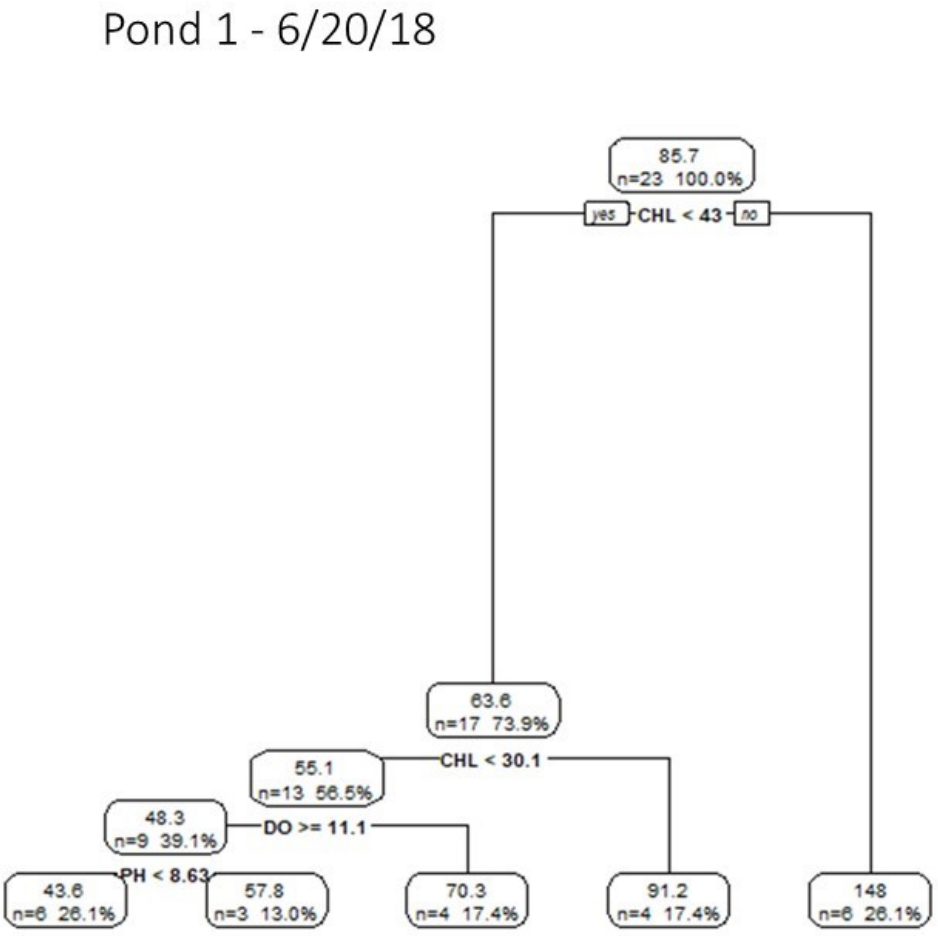


Appendix B: Supplementary material for Chapter 3

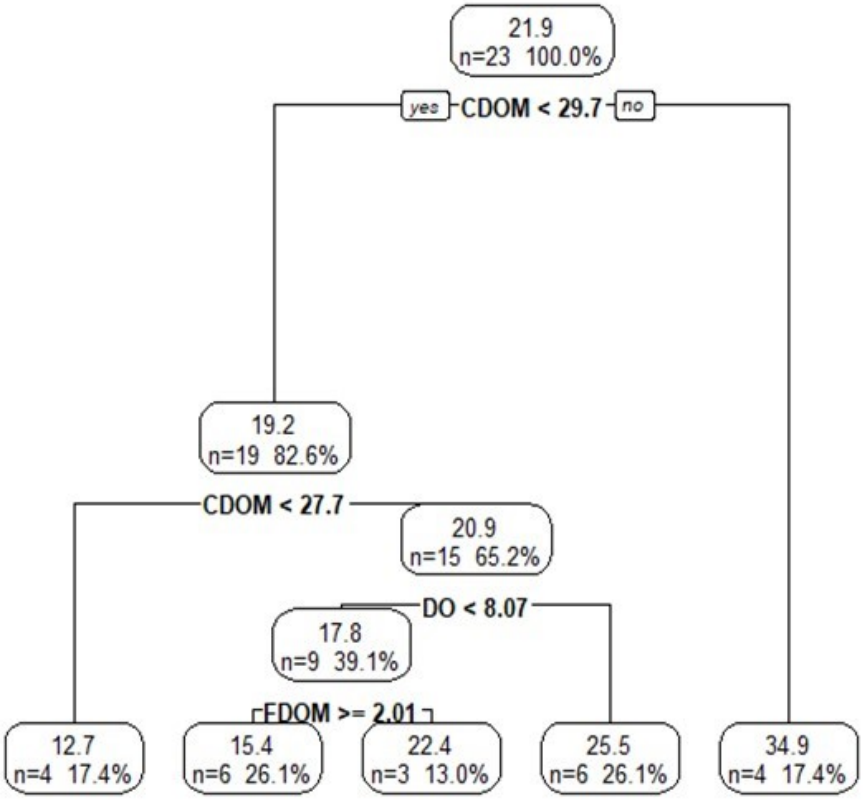
Supplemental Figure 3.1. Sampling locations for both Pond 1 (A) and Pond 2 (B). Location number is located inside the circle. Green circles indicate interior water sampling locations and blue circles indicate near-shore sampling locations.



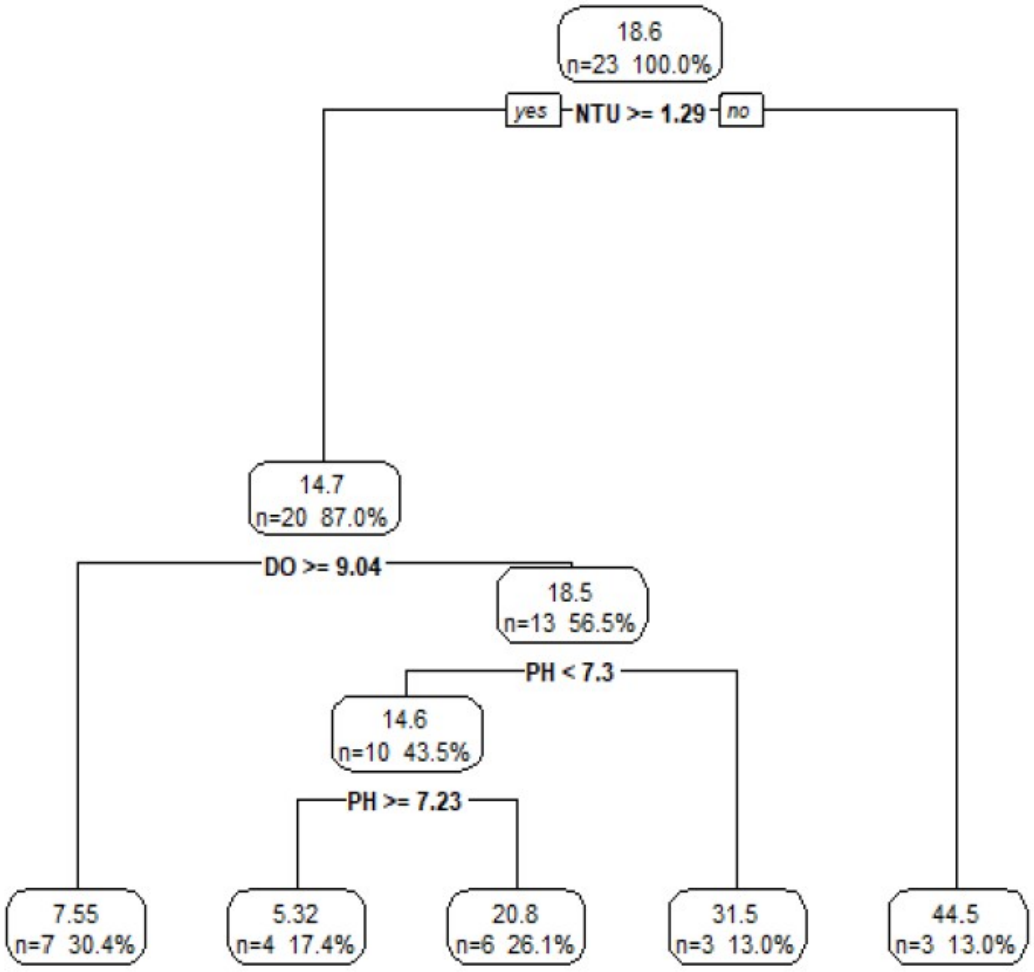
Supplemental Figure 3.2. Regression trees for both Pond 1 and Pond 2 for each sampling date. Pond number and sampling dates are located in upper left-hand corner of each tree. Dates are in chronological order for each pond.



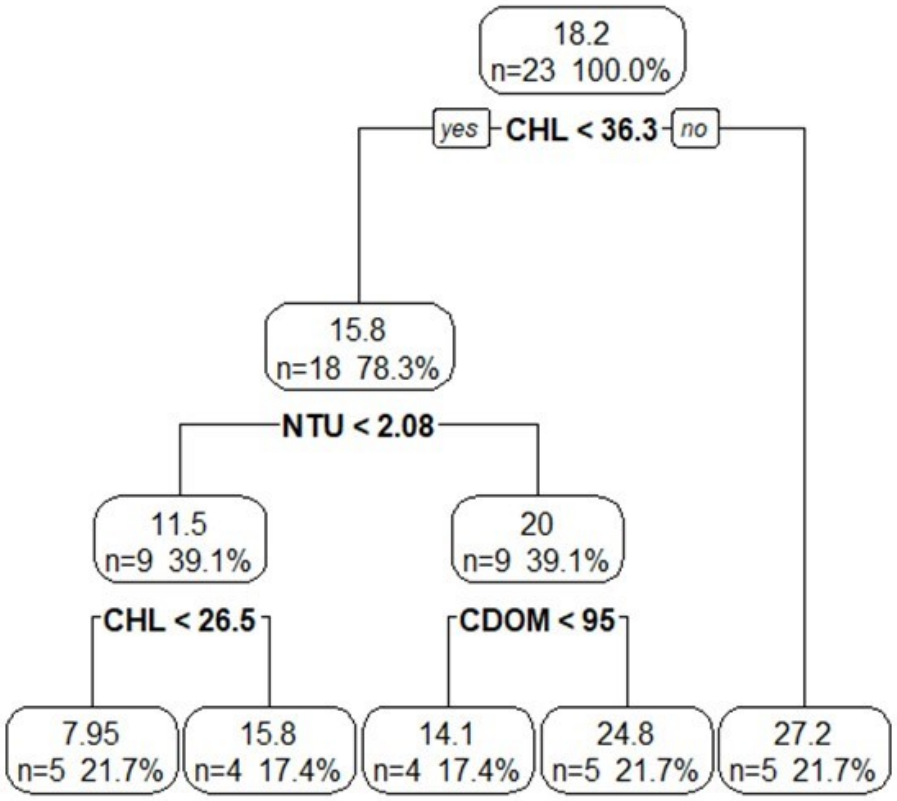
Pond 1 - 7/5/18


21.9
n=23 100.0%
yes
CDOM < 29.7
no
19.2
n=19 82.6%
CDOM < 27.7
20.9
n=15 65.2%
DO < 8.07
17.8
n=9 39.1%
FDOM >= 2.01
12.7
n=4 17.4%
15.4
n=6 26.1%
22.4
n=3 13.0%
25.5
n=6 26.1%
34.9
n=4 17.4%

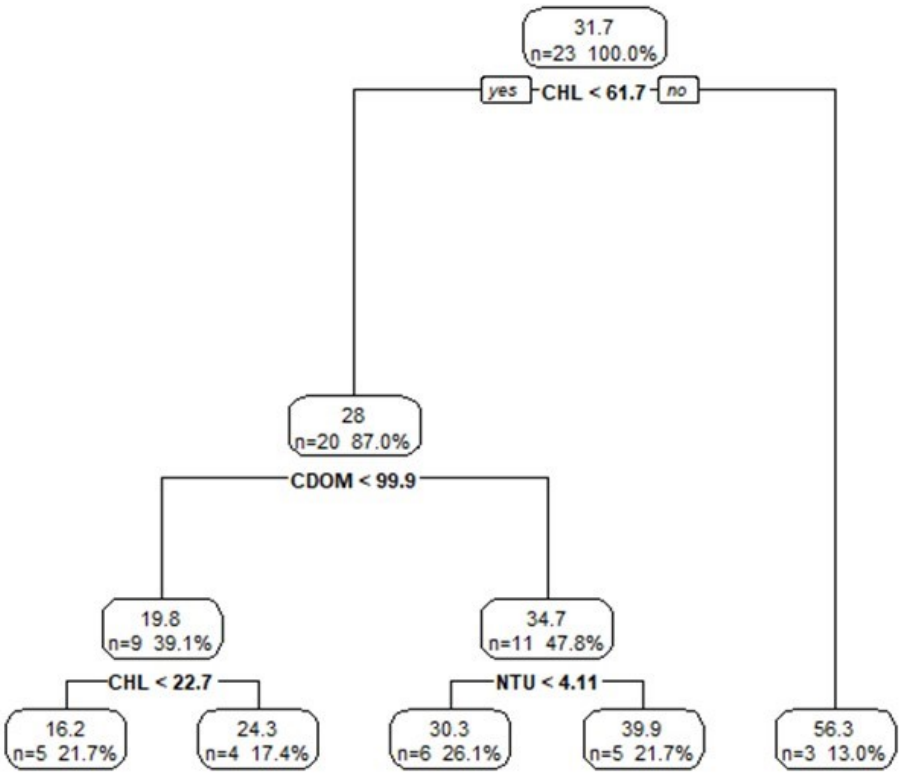
Pond 1 - 7/19/18


18.6
n=23 100.0%
yes
NTU >= 1.29
no
14.7
n=20 87.0%
DO >= 9.04
18.5
n=13 56.5%
PH < 7.3
14.6
n=10 43.5%
PH >= 7.23
7.55
n=7 30.4%
5.32
n=4 17.4%
20.8
n=6 26.1%
31.5
n=3 13.0%
44.5
n=3 13.0%

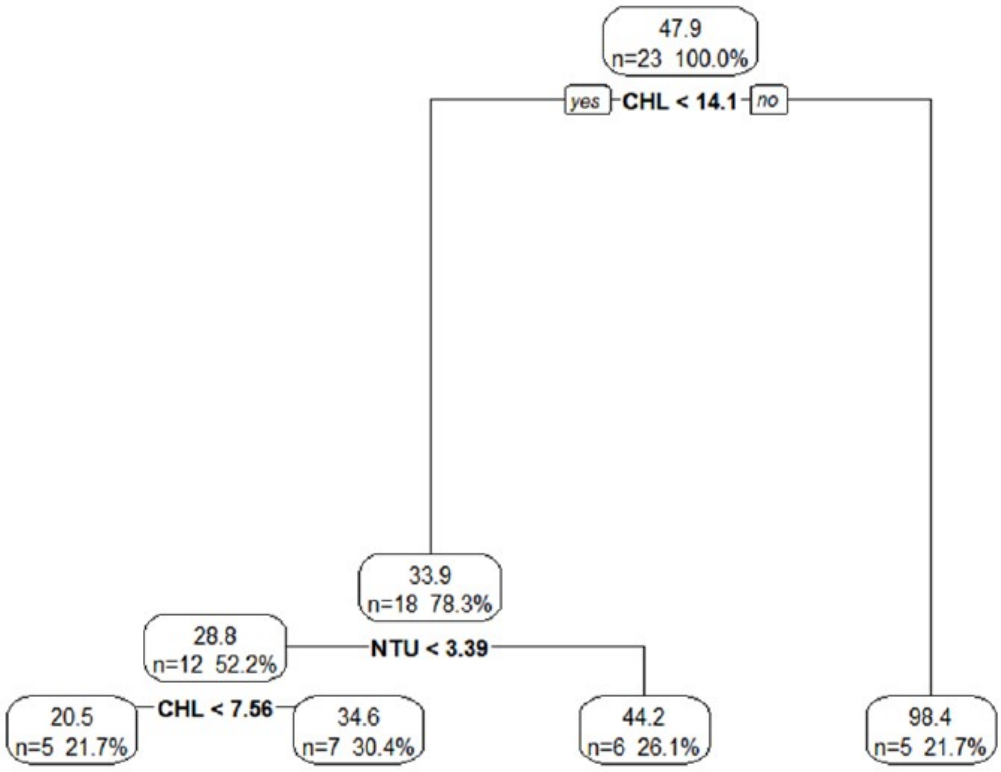
Pond 1 - 8/15/18


18.2
n=23 100.0%
yes
CHL < 36.3
no
15.8
n=18 78.3%
NTU < 2.08
11.5
n=9 39.1%
20
n=9 39.1%
CHL < 26.5
CDOM < 95
7.95
n=5 21.7%
15.8
n=4 17.4%
14.1
n=4 17.4%
24.8
n=5 21.7%
27.2
n=5 21.7%

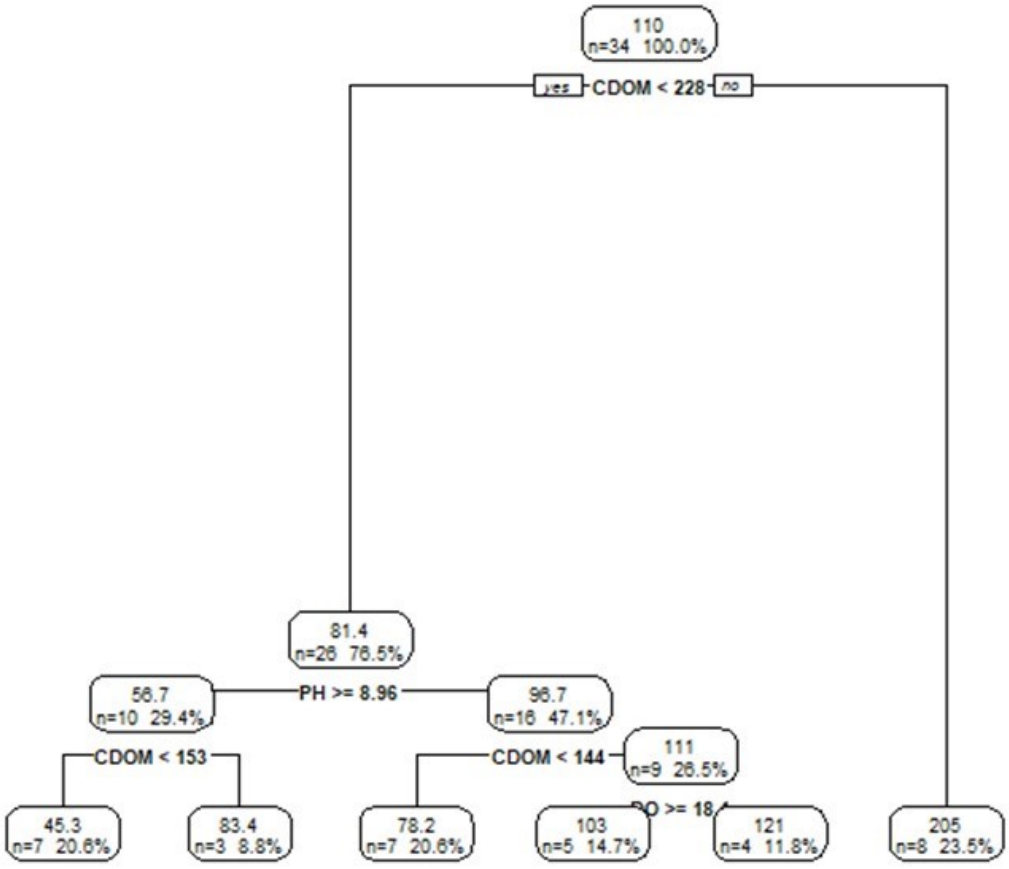
Pond 1 - 8/29/18


31.7
n=23 100.0%
yes
CHL < 61.7
no
28
n=20 87.0%
CDOM < 99.9
19.8
n=9 39.1%
34.7
n=11 47.8%
CHL < 22.7
NTU < 4.11
16.2
n=5 21.7%
24.3
n=4 17.4%
30.3
n=6 26.1%
39.9
n=5 21.7%
56.3
n=3 13.0%

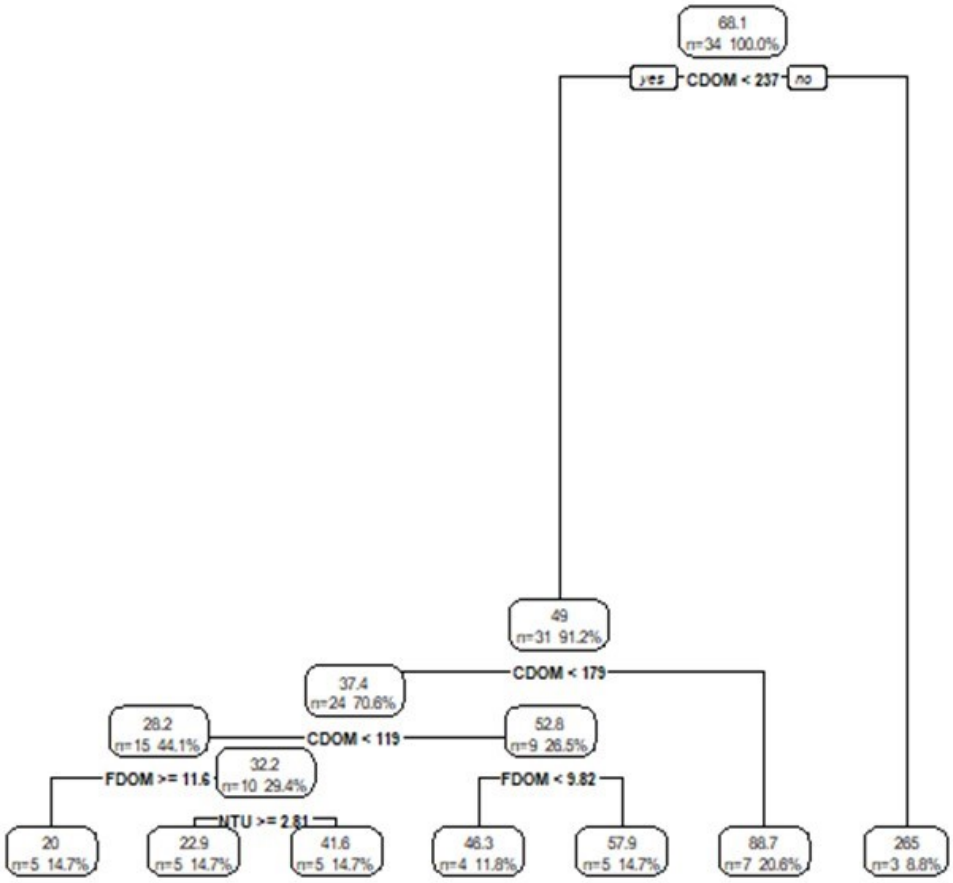
Pond 1 - 10/04/18



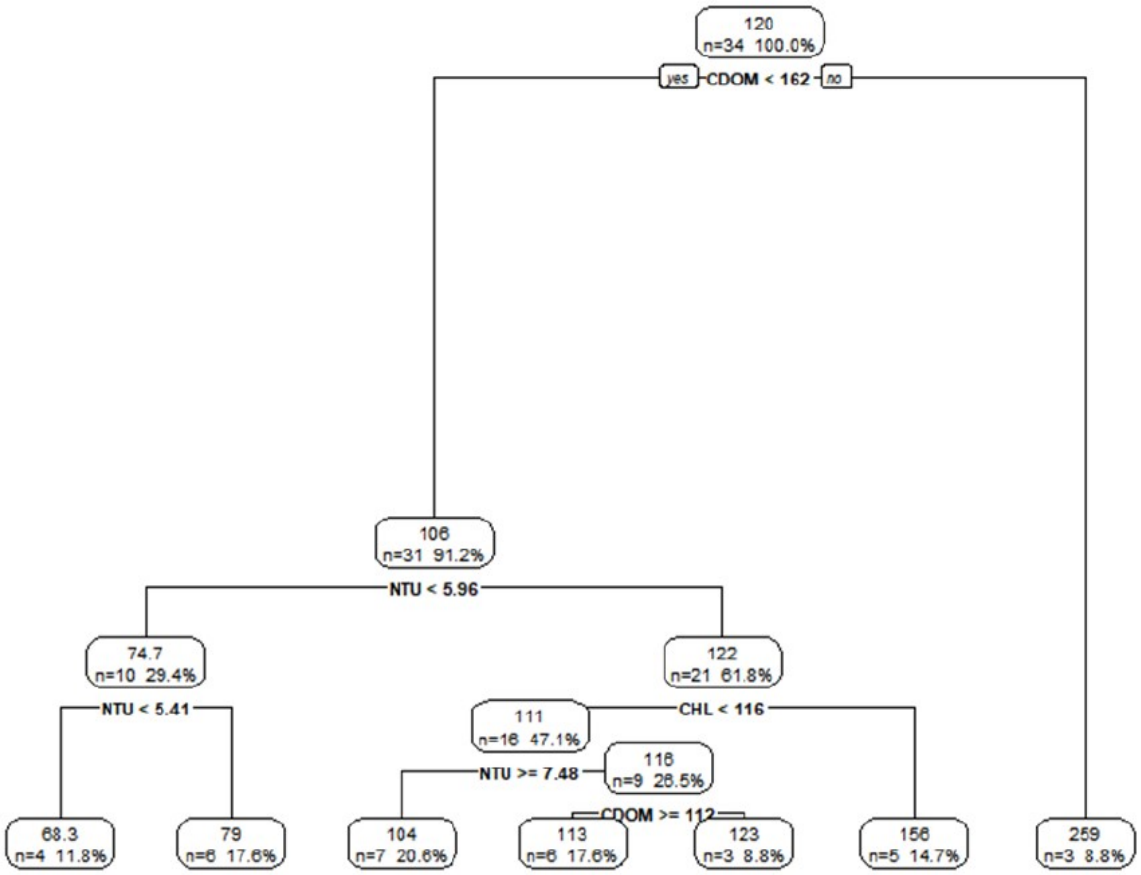
Pond 2 - 6/14/18


110
n=34 100.0%
yes
CDOM < 228
no
81.4
n=26 76.5%
56.7
n=10 29.4%
PH >= 8.96
96.7
n=16 47.1%
CDOM < 153
CDOM < 144
111
n=9 26.5%
45.3
n=7 20.6%
83.4
n=3 8.8%
78.2
n=7 20.6%
103
n=5 14.7%
DO >= 18
121
n=4 11.8%
205
n=8 23.5%

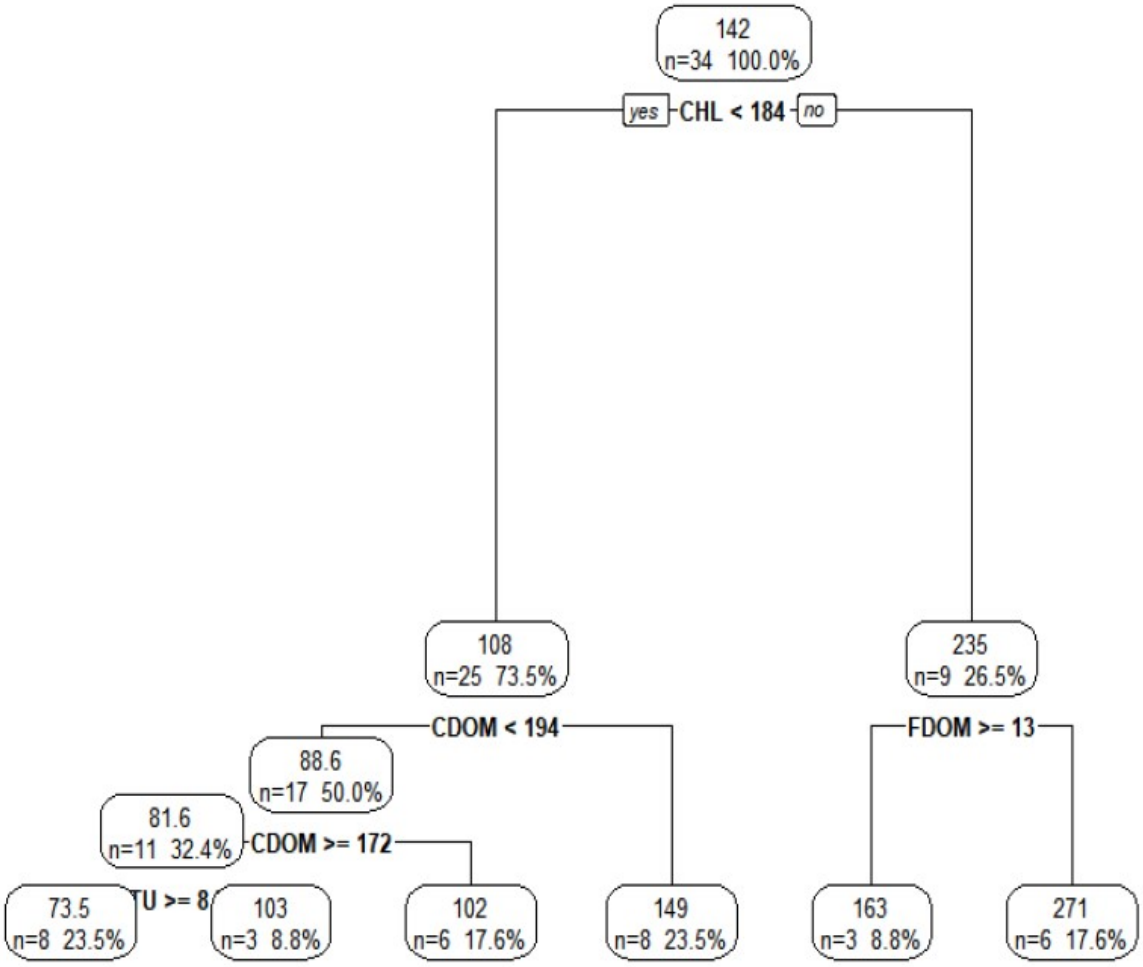
Pond 2 - 6/26/18


68.1
n=34 100.0%
yes
CDOM < 237
no
49
n=31 91.2%
37.4
n=24 70.6%
CDOM < 179
28.2
n=15 44.1%
CDOM < 119
52.8
n=9 26.5%
32.2
n=10 29.4%
FDOM >= 11.6
FDOM < 9.82
NTU >= 2.81
20
n=5 14.7%
22.9
n=5 14.7%
41.6
n=5 14.7%
46.3
n=4 11.8%
57.9
n=5 14.7%
88.7
n=7 20.6%
265
n=3 8.8%

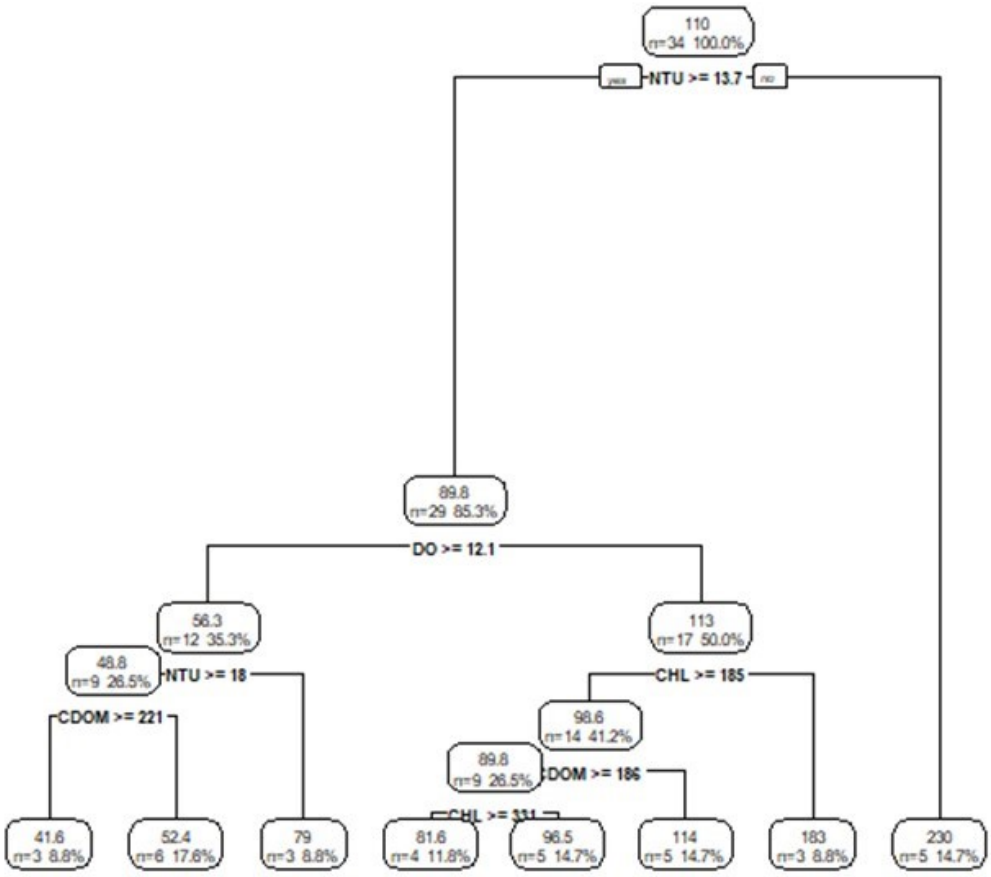
Pond 2 - 7/10/18


120
n=34 100.0%
yes
CDOM < 162
no
106
n=31 91.2%
NTU < 5.96
74.7
n=10 29.4%
122
n=21 61.8%
NTU < 5.41
111
n=16 47.1%
CHL < 116
116
n=9 26.5%
NTU >= 7.48
CDOM >= 112
68.3
n=4 11.8%
79
n=6 17.6%
104
n=7 20.6%
113
n=6 17.6%
123
n=3 8.8%
156
n=5 14.7%
259
n=3 8.8%

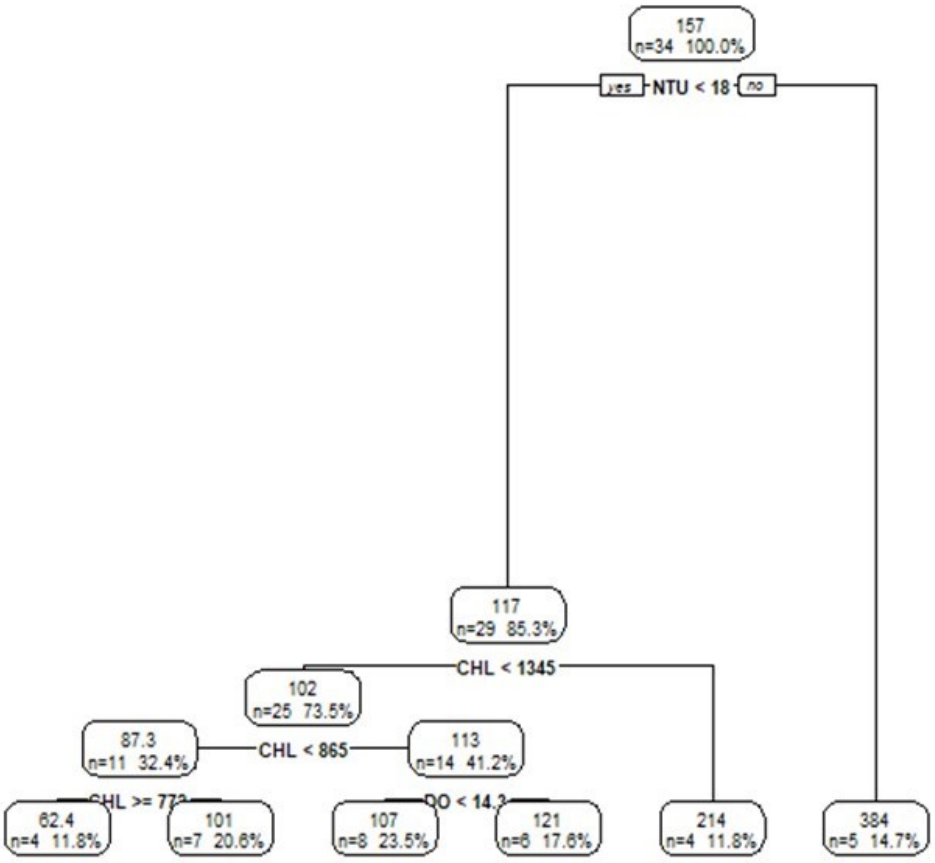
Pond 2 - 8/7/18


142
n=34 100.0%
yes
CHL < 184
no
108
n=25 73.5%
235
n=9 26.5%
CDOM < 194
FDOM >= 13
88.6
n=17 50.0%
81.6
n=11 32.4%
CDOM >= 172
TU >= 8
73.5
n=8 23.5%
103
n=3 8.8%
102
n=6 17.6%
149
n=8 23.5%
163
n=3 8.8%
271
n=6 17.6%

Pond 2 - 8/23/18


110
n=34 100.0%
yes
NTU >= 13.7
no
89.8
n=29 85.3%
DO >= 12.1
56.3
n=12 35.3%
48.8
n=9 26.5%
NTU >= 18
CDOM >= 221
41.6
n=3 8.8%
52.4
n=6 17.6%
79
n=3 8.8%
113
n=17 50.0%
CHL >= 185
98.6
n=14 41.2%
89.8
n=9 26.5%
CDOM >= 186
CHL >= 331
81.6
n=4 11.8%
96.5
n=5 14.7%
114
n=5 14.7%
183
n=3 8.8%
230
n=5 14.7%

Pond 2 - 9/20/18


157
n=34 100.0%
yes
NTU < 18
no
117
n=29 85.3%
CHL < 1345
102
n=25 73.5%
87.3
n=11 32.4%
CHL < 865
113
n=14 41.2%
62.4
n=4 11.8%
101
n=7 20.6%
107
n=8 23.5%
121
n=6 17.6%
214
n=4 11.8%
384
n=5 14.7%

Supplemental Table 3.1: Sampling dates with corresponding rainfall information for Pond 1 and Pond 2.

Dates	# of days from last rainfall to sampling date	Total days of rainfall	Total rainfall accumulation (mm)
Pond 1			
6/20/18	0*	1	7.6
7/5/18	14	3	17.5
7/19/18	2	1	18.5
8/15/18	1	2	2.5
8/29/18	7	2	62.7
10/4/18	6	7	116.3
Pond 2			
6/14/18	3	3	19.8
6/26/18	2	3	22.9
7/10/18	17	3	22.9
8/7/18	3	6	21.8
8/23/18	2	1	9.9
9/20/18	2	2	13.5

*Date with zero means there was rain the day before sampling occurred.

Supplemental Table 3.2: Percentage of the algal community that was classified as cyanobacteria using light microscopy analysis on each sampling date for Pond1 and Pond 2.

Date	% Cyanobacteria	% Nostocales	% Chroococcales	% Oscillatoriales
Pond 1				
06/20/2018	72.04	0.56	97.47	1.97
07/05/2018	18.90	35.59	26.49	37.92
07/19/2018	0.50	100.00	0.00	0.00
08/15/2018	12.57	49.27	50.73	0.00
08/29/2018	3.67	0.00	100.00	0.00
10/04/2018	94.70	100.00	0.00	0.00
Pond 2				
06/14/2018	85.72	98.65	1.35	0.00
06/26/2018	49.45	90.06	9.94	0.00
07/10/2018	62.41	76.01	23.99	0.00
08/07/2018	22.07	30.13	69.87	0.00
08/23/2018	48.81	39.86	60.02	0.12
09/20/2018	39.22	12.33	87.67	0.00

Supplemental Table 3.3: Time series of water quality parameters for Pond 1 and Pond 2.

Date	DO	pH	NTU	CDOM	*f*DOM	CHL RFU	CHL EXT	CHL YSI	Phyco
Pond 1									
6/20	11.6 ± 0.8	8.5 ± 0.2	8.8 ± 4.8	85.0 ± 6.9	7.0 ± 0.9	223.8 ± 56.1	36.7 ± 18.4	2.4 ± 0.7	114.3 ± 70.5
7/5	7.9 ± 0.4	7.3 ± 0.1	3.8 ± 2.2	43.4 ± 2.1	2.4 ± 0.5	87.5 ± 30.8	6.3 ± 2.3	1.1 ± 0.5	24.8 ± 17.5
7/19*	9.0 ± 0.3	7.3 ± 0.2	2.6 ± 2.3	32.3 ± 1.3	0.8 ± 0.3	94.7 ± 26.5	5.5 ± 2.3	1.0 ± 0.8	18.6 ± 20.6
8/15	10.5 ± 0.3	7.5 ± 0.1	2.5 ± 1.3	95.3 ± 3.6	10.0 ± 0.6	399.9 ± 100.5	28.9 ± 12.8	4.6 ± 1.4	18.2 ± 7.9
8/29	11.2 ± 0.5	8.1 ± 0.4	4.0 ± 1.9	102.6 ± 7.6	8.6 ± 0.8	497.6 ± 114.9	42.8 ± 19.6	6.9 ± 2.3	31.7 ± 15.2
10/04	10.2 ± 0.5	7.4 ± 0.2	3.6 ± 0.7	102.2 ± 4.5	12.6 ± 0.6	90.4 ± 19.2	12.5 ± 8.4	0.7 ± 0.1	47.9 ± 30.4
Pond 2									
6/14	20.7 ± 3.0	8.8 ± 0.3	9.3 ± 11.8	211.7 ± 140.2	13.1 ± 1.7	1182.1 ± 1388.5	272.4 ± 385.0	36.6 ± 49.4	110.5 ± 70.0
6/26	16.8 ± 4.1	8.5 ± 0.5	6.0 ± 5.6	166.2 ± 118.9	10.2 ± 1.6	1393.7 ± 2634.9	233.4 ± 475.2	26.3 ± 34.1	68.1 ± 88.9
7/10	11.1 ± 2.7	7.5 ± 0.3	10.2 ± 8.8	133.3 ± 92.4	9.1 ± 1.3	407.7 ± 362.0	82.3 ± 35.3	12.9 ± 29.2	119.9 ± 69.4
8/7	14.1 ± 3.1	8.5 ± 0.5	18.7 ± 21.6	203.3 ± 40.6	13.6 ± 2.3	1183.1 ± 697.5	163.0 ± 117.0	17.0 ± 9.8	141.6 ± 80.5
8/23	11.5 ± 3.8	7.4 ± 0.6	18.5 ± 4.5	188.5 ± 22.9	12.9 ± 0.4	1286.9 ± 427.2	266.1 ± 103.0	13.7 ± 4.6	110.5 ± 70.0
9/20	15.3 ± 2.7	8.3 ± 0.6	20.5 ± 34.3	159.0 ± 26.2	13.3 ± 9.1	1053.5 ± 442.6	158.0 ± 84.5	4.7 ± 3.5	156.5 ± 128.4

*Pond 1 was treated with copper sulfate on 7/1/18.
The "±" sign separates average and standard deviations across both ponds.
DO - dissolved oxygen (mg L^{-1}), NTU - turbidity (NTU), CDOM - colored dissolved organic matter (µg L^{-1}), *f*DOM - fluorescent dissolved organic matter (relative fluorescent units), CHL RFU – chlorophyll-*a* fluorescence (relative fluorescent units), CHL EXT – extracted chlorophyll-*a* concentrations (µg L^{-1}), CHL YSI – chlorophyll-*a* measured via YSI sonde (relative fluorescent units), Phyco – Phycocyanin concentrations (µg L^{-1}).

Supplemental Table 3.4: Pearson correlation coefficients for Pond 1 and Pond 2. The significant critical value for Pond 1 was r =0.413. The significant value for Pond 2 was r=0.339. Only significant correlations are displayed (P < 0.05).

Pearson Correlations	Pond 1										Pond 2									
		DO	pH	NTU	CHL YSI	*f*DOM	CDOM	CHL RFU	CHL EXT	Phyco		DO	pH	NTU	CHL YSI	*f*DOM	CDOM	CHL RFU	CHL EXT	Phyco
DO	6/20		0.586					0.422			6/14		0.868			-0.343				
DO	7/5					-0.473	-0.755	-0.737	-0.483		6/26		0.935	0.465	0.431	-0.810	0.533	0.528	0.461	0.517
DO	7/19		0.448				0.419				7/10		0.909				0.375	0.373		0.402
DO	8/15		0.811		-0.682	-0.598					8/7		0.920		0.453	-0.672	0.355	0.341		
DO	8/29		0.759	-0.464	-0.659	-0.625	-0.568		-0.480	-0.509	8/23		0.970	0.366		-0.702	0.569	0.532		-0.452
DO	10/4		0.764								9/20		0.931	0.373	0.421					0.451
pH	6/20					-0.449					6/14									
pH	7/5										6/26					-0.706	0.380	0.380		0.360
pH	7/19								-0.670		7/10				0.340		0.393	0.392		0.420
pH	8/15				-0.417	-0.564					8/7			0.373		-0.779				0.350
pH	8/29				-0.427	-0.549	-0.422				8/23					-0.653	0.548	0.562		-0.476
pH	10/4					-0.435					9/20				0.349					0.376
NTU	6/20										6/14				0.844	-0.823	0.808	0.789	0.836	0.705
NTU	7/5					-0.701					6/26				0.839	-0.728	0.427	0.414	0.389	0.424
NTU	7/19										7/10				0.573	-0.795	0.568	0.555		0.549
NTU	8/15				0.761						8/7					-0.520				0.351
NTU	8/29										8/23				0.626		0.565	0.479	0.541	-0.562
NTU	10/4						0.508	0.512	0.480	0.486	9/20				0.631				0.531	0.701
CHL YSI	6/20										6/14					-0.862	0.903	0.8662	0.900	0.749
CHL YSI	7/5										6/26					-0.696	0.578	0.559	0.577	0.554
CHL YSI	7/19						0.507	0.795			7/10					-0.608	0.970	0.969		0.845
CHL YSI	8/15										8/7						0.732	0.776	0.665	0.454
CHL YSI	8/29					0.762				0.433	8/23						0.638	0.599	0.365	-0.381
CHL YSI	10/4					0.792	0.613				9/20					0.540	0.713	0.460	0.836	0.916
*f*DOM	6/20										6/14						-0.891	-0.695	-0.855	-0.533
*f*DOM	7/5							0.485			6/26						-0.486	-0.476	-0.430	-0.465
*f*DOM	7/19						0.465		0.463		7/10						-0.581	-0.580		-0.575
*f*DOM	8/15										8/7									-0.451
*f*DOM	8/29										8/23									
*f*DOM	10/4						0.494				9/20						0.753	0.779	0.568	
CDOM	6/20							0.594	0.707	0.668	6/14							0.636	0.954	0.489
CDOM	7/5							0.606	0.711	0.519	6/26							0.998	0.963	0.991
CDOM	7/19							0.462			7/10							0.998		0.899
CDOM	8/15							0.783	0.619	0.548	8/7							0.958	0.808	0.667
CDOM	8/29								0.819	0.672	8/23							0.947	0.575	-0.616
CDOM	10/4							0.915	0.775	0.761	9/20							0.908	0.842	0.668
CHL RFU	6/20										6/14								0.680	0.915
CHL RFU	7/5								0.525		6/26								0.968	0.993
CHL RFU	7/19										7/10									0.889

CHL RFU	8/15								0.690	0.563	8/7								0.850	0.614
CHL RFU	8/29										8/23								0.435	-0.598
CHL RFU	10/4								0.910	0.864	9/20								0.664	0.406
CHL EXT	6/20									0.873	6/14									0.522
CHL EXT	7/5									0.525	6/26									0.962
CHL EXT	7/19										7/10									
CHL EXT	8/15									0.623	8/7									0.593
CHL EXT	8/29									0.727	8/23									-0.493
CHL EXT	10/4									0.900	9/20									0.801

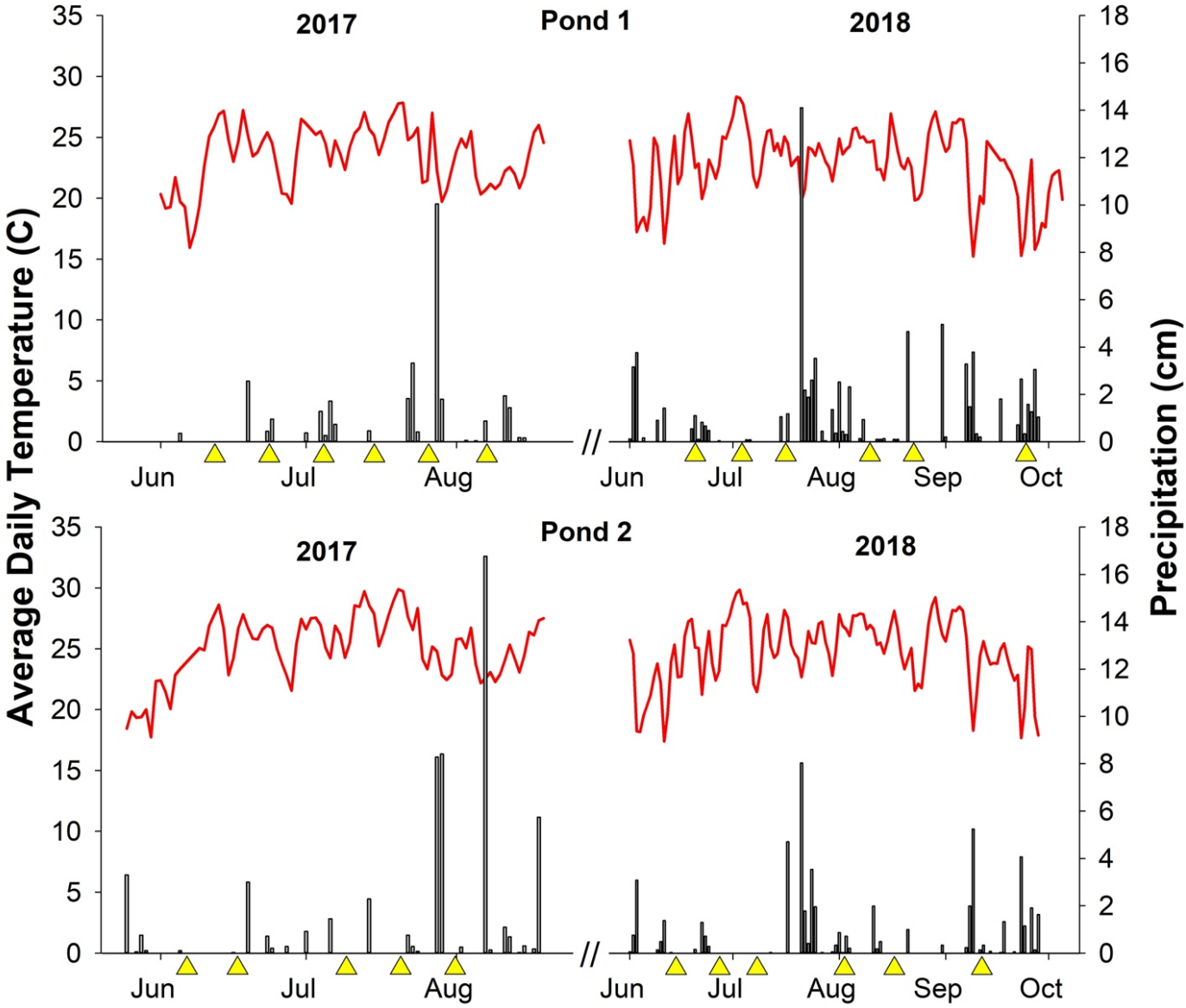
Supplemental Table 3.5: Root mean square error (RMSE) and r^2 values for regression trees. Each column displays the analyses run with different chlorophyll-*a* values measured using different analytical techniques.

Date	CHL YSI		CHL EXT		CHL RFU	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Pond 1						
06/20/18	0.699	37.778	0.675	39.281	0.700	37.714
07/05/18	0.393	13.313	0.393	13.313	0.393	13.313
07/19/18	0.441	15.036	0.441	15.036	0.441	15.036
08/15/18	0.789	3.564	0.892	2.544	0.649	4.590
08/29/18	0.696	8.175	0.708	8.009	0.658	8.676
10/04/18	0.686	16.649	0.877	10.426	0.871	10.682
Pond 2						
06/14/18	0.690	38.374	0.690	38.374	0.839	27.638
06/26/18	0.565	57.749	0.565	57.749	0.565	57.749
07/10/18	0.544	46.162	0.548	45.931	0.528	46.930
08/07/18	0.712	42.588	0.716	42.299	0.713	42.463
08/23/18	0.782	32.198	0.801	30.765	0.801	30.778
09/20/18	0.896	40.732	0.685	71.006	0.680	71.508

CHL RFU – chlorophyll-*a* fluorescence (relative fluorescent units), CHL EXT – extracted chlorophyll-*a* concentrations ($\mu g\ L^{-1}$), CHL YSI – chlorophyll-*a* measured via YSI sonde (relative fluorescent units).

Appendix C: Supplementary material for Chapter 4

Supplemental Figure 4.1. Weather data for Pond 1 and Pond 2 for 2017 and 2018. Average daily temperature readings are represented by the red line and daily precipitation is represented by black bars. Sampling dates are indicated with a yellow triangle at the x axis.



Supplemental Table 4.1. Summary statistics for all measured parameters and phytoplankton functional groups for 2017 and 2018.

Minimum, maximum, mean, and median values of measured phytoplankton functional groups and water quality data for 2017 and 2018									
Variable	**Units**	**Pond 1**				**Pond 2**			
2017		**Min**	**Max**	**Mean**	**Median**	**Min**	**Max**	**Mean**	**Median**
Diatoms grouped	Log cells/L	4.19	7.59	5.94	6.00	4.19	7.77	6.32	6.46
Green Algae grouped	Log cells/L	6.03	8.08	7.10	7.14	4.97	8.18	6.92	7.08
Cyanobacteria grouped	Log cells/L	4.74	7.91	5.98	6.11	4.67	8.69	6.14	5.81
TEMP	°C	22.33	29.84	26.31	26.34	21.81	33.74	27.37	29.16
DO	mg/L	8.14	15.44	10.39	10.24	2.74	24.87	11.82	12.01
SPC	µS/cm	149.50	178.30	162.52	162.15	138.90	183.20	161.41	160.15
pH		6.90	9.56	8.82	8.84	6.18	9.90	8.18	8.49
NTU		1.81	23.78	5.16	5.05	1.52	68.61	10.82	4.73
Phyco	RFU	0.12	8.43	1.14	0.93	-0.27	34.18	3.49	0.83
CHL	RFU	0.87	46.54	3.83	2.21	0.84	123.31	10.74	5.83
FDOM	ppb	-0.05	39.53	13.06	5.29	9.39	47.63	30.34	29.11
2018									
Diatoms grouped	Log cells/L	4.45	7.22	5.61	5.75	4.49	7.42	5.63	5.79
Green algae grouped	Log cells/L	5.49	7.78	6.70	6.72	4.89	8.08	7.10	7.15
Cyanobacteria grouped	Log cells/L	4.19	7.95	5.64	4.19	5.29	8.38	7.08	7.25
TEMP	°C	22.34	30.10	26.99	27.06	24.00	36.28	27.54	26.95
DO	mg/L	7.04	12.68	10.07	10.41	5.83	27.18	14.91	14.94
SPC	µS/cm	1.70	168.10	148.83	149.05	10.10	164.10	142.59	145.70
pH		6.47	8.67	7.65	7.46	6.46	9.31	8.17	8.37
NTU		1.20	24.00	4.20	3.28	1.00	178.46	13.86	9.85
Phyco	RFU	0.10	10.65	0.95	0.61	0.45	35.11	4.86	3.32
CHL	RFU	0.51	11.12	2.80	1.79	1.37	166.42	19.97	10.43
FDOM	ppb	0.49	41.25	20.72	23.60	9.09	50.64	34.90	35.58
CDOM	µg/L	29.94	113.15	76.80	91.38	93.54	826.40	177.00	161.97
Submerged light -0cm		7.00	1809.00	712.72	793.00	55.00	2128.00	1113.02	1195.50
Submerged light – 7.5cm		7.00	1578.00	624.42	653.50	7.00	1946.00	810.60	848.50
Submerged light – 15cm		6.00	1479.00	540.88	546.50	5.00	1933.00	578.13	607.00
PAR		4.00	1278.00	950.46	928.00	48.00	2200.00	1312.86	1373.00

Figure 4.2. Spatial RMSE for 2017+2018 calculated for green algae using set AB parameters. Number located inside of symbol indicates sampling location number. Number above location indicates the calculated RMSE for green algae at that location.

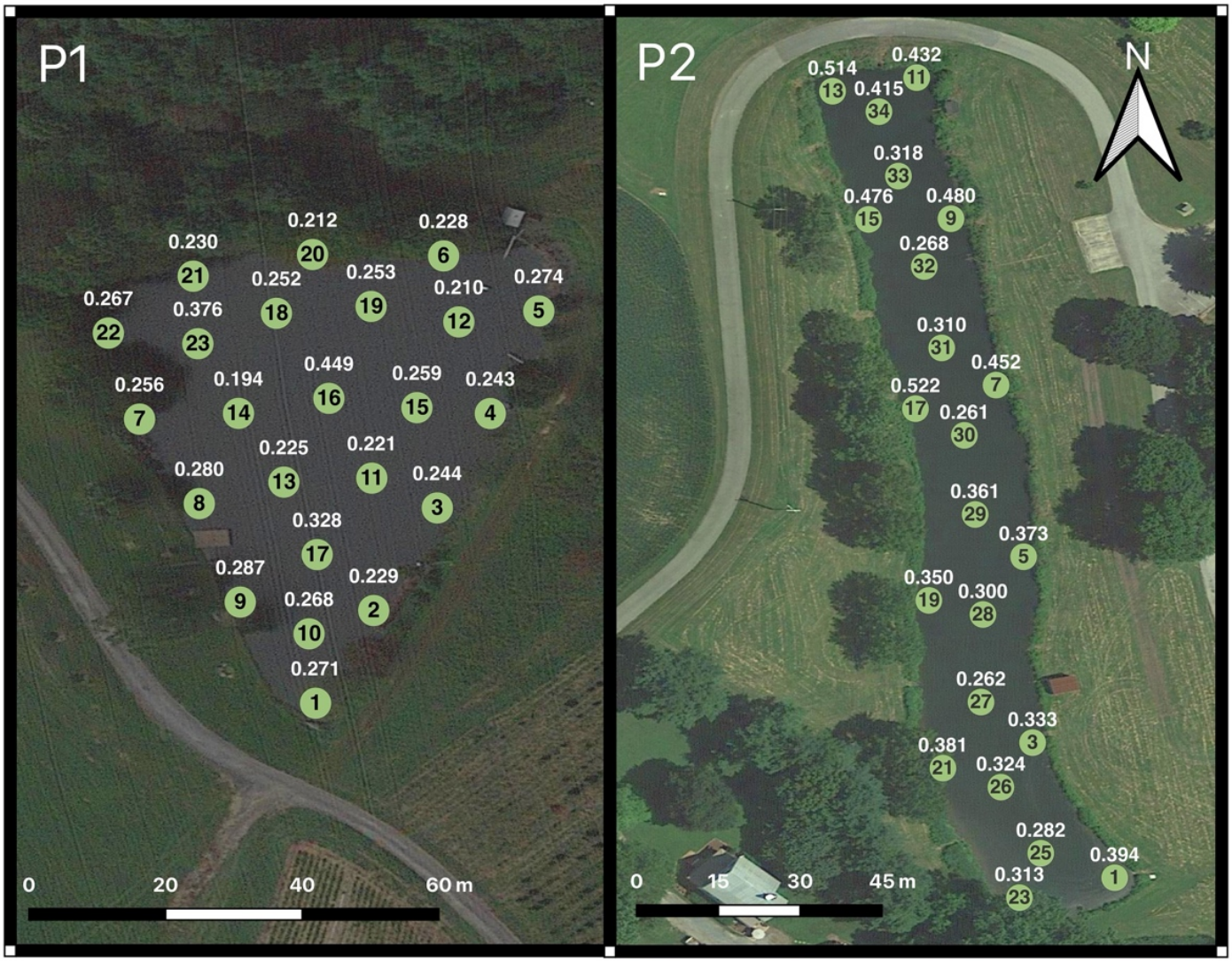


Figure 4.3. Spatial RMSE for 2017+2018 calculated for diatoms using set AB parameters. Number located inside of symbol indicates sampling location number. Number above location indicates the calculated RMSE for diatoms at that location.

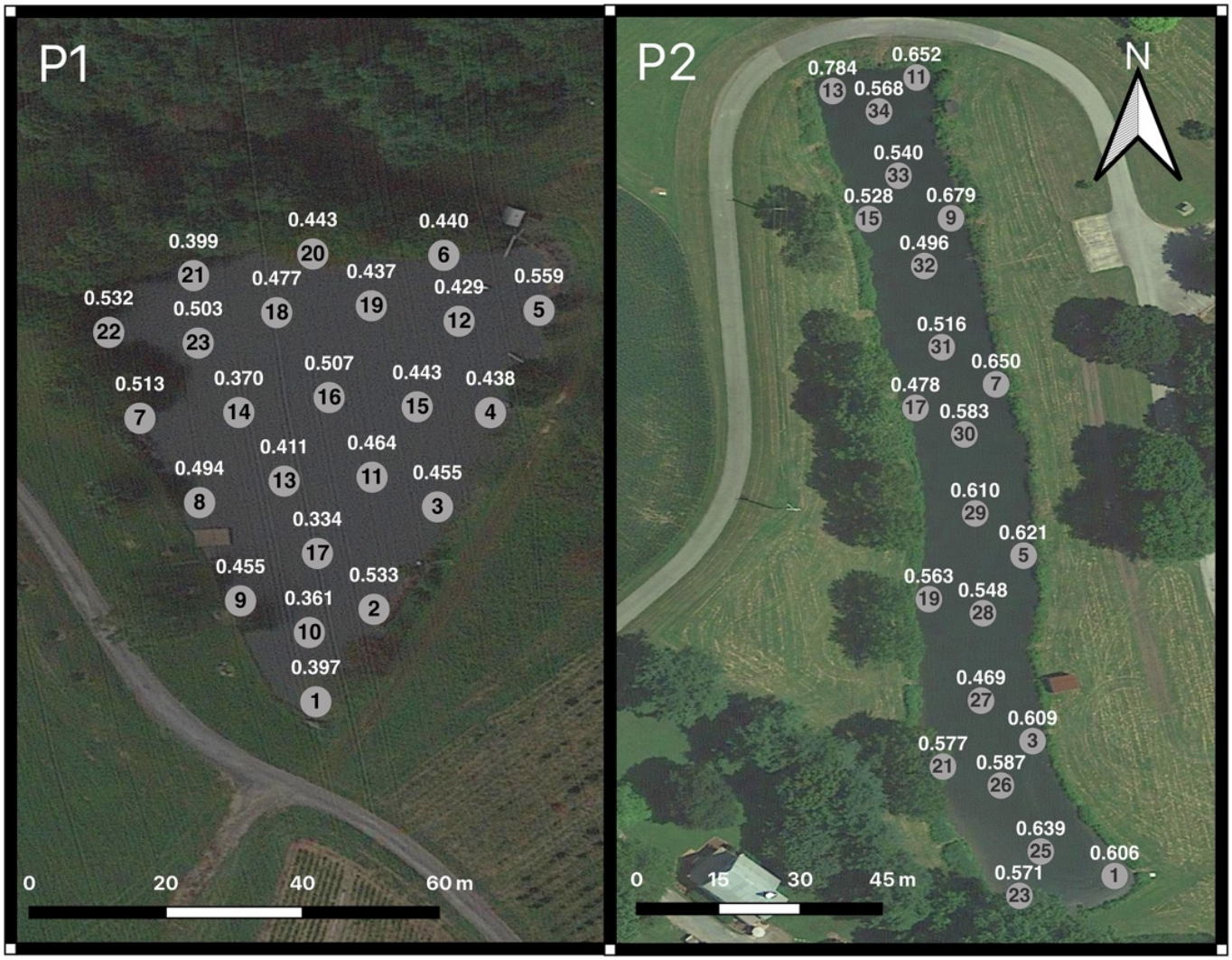
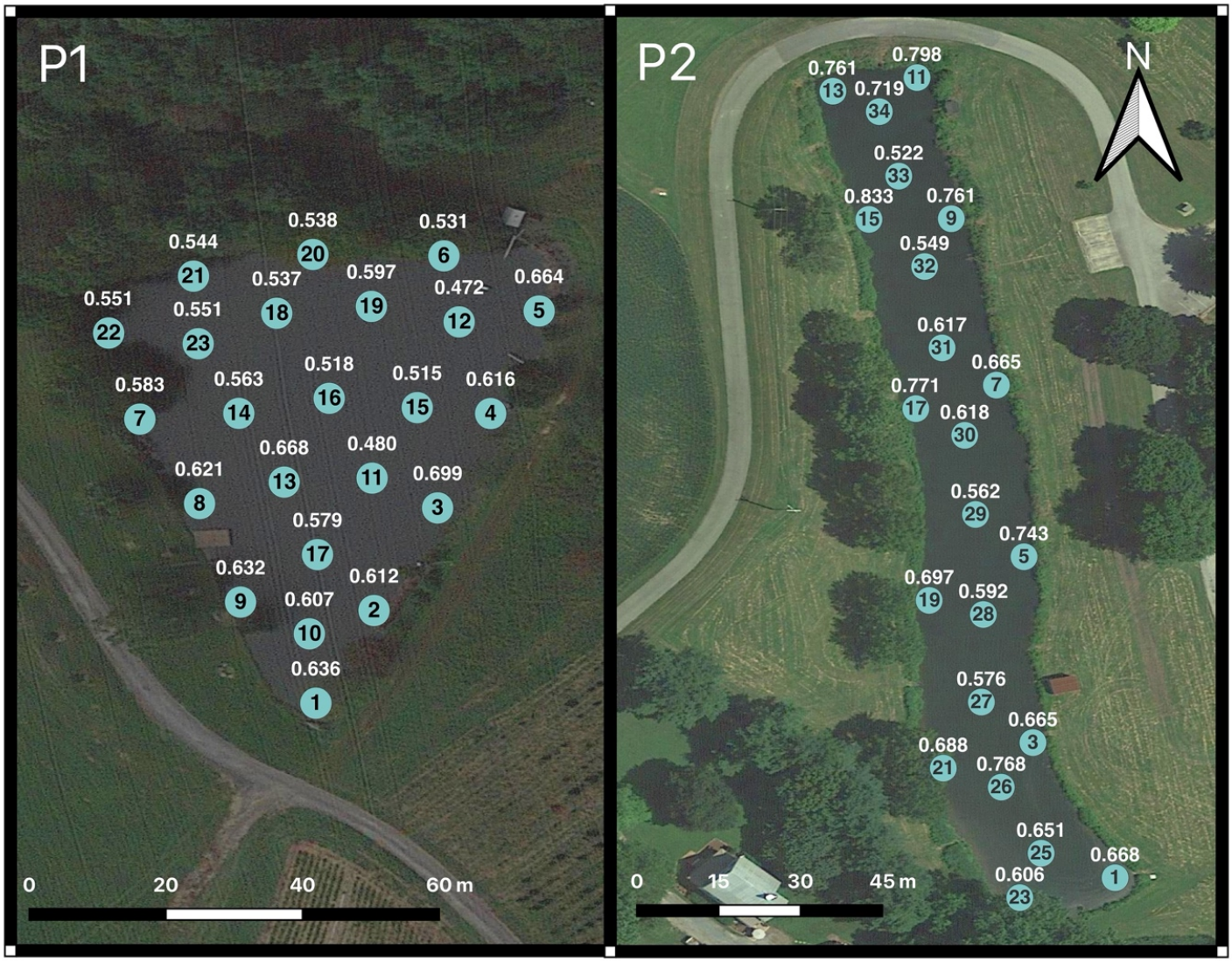


Figure 4.4. Spatial RMSE for 2017+2018 calculated for cyanobacteria using set AB parameters. Number located inside of symbol indicates sampling location number. Number above location indicates the calculated RMSE for diatoms at that location.



Bibliography

Aboal, M., Puig, M. A., Ríos, H., & López-Jiménez, E. (2000). Relationship between macroinvertebrate diversity and toxicity of Cyanophyceae (Cyanobacteria) in some streams from Eastern Spain. *SIL Proceedings, 1922-2010*, *27*(1), 555–559. https://doi.org/10.1080/03680770.1998.11901297

Adloff, C. T., Bem, C. C., Reichert, G., & Azevedo, J. C. R. de. (2018). Analysis of the phytoplankton community emphasizing cyanobacteria in four cascade reservoirs system of the Iguazu River, Paraná, Brazil. *Brazilian Journal of Water Resources*, *23*. https://doi.org/10.1590/2318-0331.0318170050

Agawin, N. S. R., Rabouille, S., Veldhuis, M. J. W., Servatius, L., Hol, S., van Overzee, H. M. J., & Huisman, J. (2007). Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non—Nitrogen-fixing phytoplankton species. *Limnology and Oceanography*, *52*(5), 2233–2248. https://doi.org/10.4319/lo.2007.52.5.2233

Ahn, C.-Y., Oh, H.-M., & Park, Y.-S. (2011). Evaluation of environmental factors on cyanobacterial bloom in eutrophic reservoir using artificial neural networks: Modeling of cyanobacterial bloom. *Journal of Phycology*, *47*(3), 495–504. https://doi.org/10.1111/j.1529-8817.2011.00990.x

Alexander, R., & Imberger, J. (2009). Spatial distribution of motile phytoplankton in a stratified reservoir: The physical controls on patch formation. *Journal of Plankton Research*, *31*(1), 101–118. https://doi.org/10.1093/plankt/fbn101

Ali, P. A. W. F. P., Khan, Z., Shinwari, Fazli, A., & Sarim, M. (2010). Contribution to the algal flora (Chlorophyta) of fresh waters of District Swat. N.W.F.P., Pakistan. *Pakistan Journal of Botany*, *42*(5), 3457–3462.

Almuhtaram, H., Zamyadi, A., & Hofmann, R. (2021). Machine learning for anomaly detection in cyanobacterial fluorescence signals. *Water Research*, *197*, 117073. https://doi.org/10.1016/j.watres.2021.117073

Anderson, D. M., Fensin, E., Gobler, C. J., Hoeglund, A. E., Hubbard, K. A., Kulis, D. M., Landsberg, J. H., Lefebvre, K. A., Provoost, P., Richlen, M. L., Smith, J. L., Solow, A. R., & Trainer, V. L. (2021). Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae*, *102*, 101975. https://doi.org/10.1016/j.hal.2021.101975

Andres, A. S., Main, C. R., Pettay, D. T., & Ullman, W. J. (2019). Hydrophysical and hydrochemical controls of cyanobacterial blooms in Coursey Pond, Delaware (USA). *Journal of Environmental Quality*, *48*(1), 73–82. https://doi.org/10.2134/jeq2018.03.0108

Awada, H., Aronica, S., Bonanno, A., Basilone, G., Zgozi, S. W., Giacalone, G., Fontana, I., Genovese, S., Ferreri, R., Mazzola, S., Spagnolo, B., Valenti, D., & Denaro, G. (2021). A novel method to simulate the 3D chlorophyll distribution in marine oligotrophic waters. *Communications in Nonlinear Science and Numerical Simulation*, *103*, 106000. https://doi.org/10.1016/j.cnsns.2021.106000

Baohong, C., Kang, W., Xu, D., & Hui, L. (2021). Long-term changes in red tide outbreaks in Xiamen Bay in China from 1986 to 2017. *Estuarine, Coastal and Shelf Science*, *249*, 107095. https://doi.org/10.1016/j.ecss.2020.107095

Bartram, J., & Ballance, R. (Eds.). (2020). *Water Quality Monitoring: A practical guide to the design and implementation of freshwater quality studies and monitoring programmes*. CRC Press. https://doi.org/10.1201/9781003062110

Bastien, C., Cardin, R., Veilleux, É., Deblois, C., Warren, A., & Laurion, I. (2011). Performance evaluation of phycocyanin probes for the monitoring of cyanobacteria. *Journal of Environmental Monitoring*, *13*(1), 110–118. https://doi.org/10.1039/C0EM00366B

Bate, G., Smailes, P., & Adams, J. (2004). A water quality index for use with diatoms in the assessment of rivers. *Water SA*, *30*(4), 493–498.

Bellinger, E. D., & Sigee, D. C. (2015). *Freshwater Algae* (1st ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118917152

Benavides, M., Martias, C., Elifantz, H., Berman-Frank, I., Dupouy, C., & Bonnet, S. (2018). Dissolved organic matter influences n_2 fixation in the New Caledonian Lagoon (Western Tropical South Pacific). *Frontiers in Marine Science*, *5*. https://www.frontiersin.org/article/10.3389/fmars.2018.00089

Bergkemper, V., & Weisse, T. (2018). Do current European lake monitoring programmes reliably estimate phytoplankton community changes? *Hydrobiologia*, *824*(1), 143–162. https://doi.org/10.1007/s10750-017-3426-6

Beutler, M., Wiltshire, K., Reineke, C., & Hansen, U. (2004). Algorithms and practical fluorescence models of the photosynthetic apparatus of red cyanobacteria and Cryptophyta designed for the fluorescence detection of red cyanobacteria and cryptophytes. *Aquatic Microbial Ecology*, *35*, 115–129. https://doi.org/10.3354/ame035115

Beversdorf, L. J., Weirich, C. A., Bartlett, S. L., & Miller, T. R. (2017). Variable cyanobacterial toxin and metabolite profiles across six eutrophic lakes of differing physiochemical characteristics. *Toxins*, *9*(2), 62. https://doi.org/10.3390/toxins9020062

Bhattarai, S. P., Midmore, D. J., & Pendergast, L. (2008). Yield, water-use efficiencies and root distribution of soybean, chickpea and pumpkin under different subsurface drip irrigation depths and oxygation treatments in vertisols. *Irrigation Science*, *26*(5), 439. https://doi.org/10.1007/s00271-008-0112-5

Bittencourt-Oliveira, M. do C., Cordeiro-Araújo, M. K., Chia, M. A., Arruda-Neto, J. D. de T., Oliveira, Ê. T. de, & Santos, F. dos. (2016). Lettuce irrigated with contaminated water: Photosynthetic effects, antioxidative response and bioaccumulation of microcystin congeners. *Ecotoxicology and Environmental Safety*, *128*, 83–90. https://doi.org/10.1016/j.ecoenv.2016.02.014

Boman, B. J. (1995). Effects of orifice size on microsprinkler clogging rates. *Applied Engineering in Agriculture*, *11*(6), 839–843.

Bondarenko, N. A., Guselnikova, N. E., Logacheva, N. F., & Pomazkina, G. V. (1996). Spatial distribution of phytoplankton in Lake Baikal, Spring 1991. *Freshwater Biology*, *35*(3), 517–523. https://doi.org/10.1111/j.1365-2427.1996.tb01765.x

Bouma-Gregson, K., Power, M. E., & Bormans, M. (2017). Rise and fall of toxic benthic freshwater cyanobacteria (*Anabaena* spp.) in the Eel River: Buoyancy

and dispersal. *Harmful Algae*, *66*, 79–87. https://doi.org/10.1016/j.hal.2017.05.007

Bradshaw, C., Kautsky, U., & Kumblad, L. (2012). Ecological stoichiometry and multi-element transfer in a coastal ecosystem. *Ecosystems*, *15*(4), 591–603. https://doi.org/10.1007/s10021-012-9531-5

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2017). *Classification and regression trees*. Routledge. https://doi.org/10.1201/9781315138470.

Briand, E., Escoffier, N., Straub, C., Sabart, M., Quiblier, C., & Humbert, J.-F. (2009). Spatiotemporal changes in the genetic diversity of a bloom-forming *Microcystis* aeruginosa (Cyanobacteria) population. *The ISME Journal*, *3*(4), 419–429. https://doi.org/10.1038/ismej.2008.121

Brient, L., Lengronne, M., Bertrand, E., Rolland, D., Sipel, A., Steinmann, D., Baudin, I., Legeas, M., Rouzic, B. L., & Bormans, M. (2008). A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies. *Journal of Environmental Monitoring*, *10*(2), 248–255. https://doi.org/10.1039/B714238B

Brönmark, C., & Hansson, L.-A. (2002). Environmental issues in lakes and ponds: Current state and perspectives. *Environmental Conservation*, *29*(3), 290–307. https://doi.org/10.1017/S0376892902000218

Bucks, D. A., Nakayama, F. S., & Gilbert, R. G. (1979). Trickle irrigation water quality and preventive maintenance. *Agricultural Water Management*, *2*(2), 149–162. https://doi.org/10.1016/0378-3774(79)90028-3

Buratti, F., Manganelli, M., Vichi, S., Stefanelli, M., Scardala, S., Testai, E., & Funari, E. (2017). Cyanotoxins: Producing organisms, occurrence, toxicity, mechanism of action and human health toxicological risk evaluation. *Archives of Toxicology*, *91*(3), 1049–1130. https://doi.org/10.1007/s00204-016-1913-6

Burford, M. A., Carey, C. C., Hamilton, D. P., Huisman, J., Paerl, H. W., Wood, S. A., & Wulff, A. (2020). Perspective: Advancing the research agenda for improving understanding of cyanobacteria in a future of global change. *Harmful Algae*, *91*, 101601. https://doi.org/10.1016/j.hal.2019.04.004

Butitta, V. L., Carpenter, S. R., Loken, L. C., Pace, M. L., & Stanley, E. H. (2017). Spatial early warning signals in a lake manipulation. *Ecosphere*, *8*(10), 1–11.

Carmichael, W. W. (1994). The toxins of cyanobacteria. *Scientific American*, *270*(1), 78–86.

Carmichael, W. W. (2001). Health Effects of Toxin-Producing Cyanobacteria: "The CyanoHABs." *Human and Ecological Risk Assessment: An International Journal*, *7*(5), 1393–1407. https://doi.org/10.1080/20018091095087

Carpenter, S. R., Gurevitch, A., & Adams, M. S. (1979). Factors causing elevated biological oxygen demand in the littoral zone of Lake Wingra, Wisconsin. *Hydrobiologia*, *67*(1), 3–9. https://doi.org/10.1007/BF00020871

Carty, S., & Parrow, M. W. (2015). Chapter 17—Dinoflagellates. In J. D. Wehr, R. G. Sheath, & J. P. Kociolek (Eds.), *Freshwater Algae of North America*

(Second Edition) (pp. 773–807). Academic Press. https://doi.org/10.1016/B978-0-12-385876-4.00017-7

Casé, M., Leça, E. E., Leitão, S. N., SantAnna, E. E., Schwamborn, R., & Moraes Junior, A. T. (2008). Plankton community as an indicator of water quality in tropical shrimp culture ponds. *Marine Pollution Bulletin*, *56*(7), 1343–1352.

Celewicz-Goldyn, S., & Kuczynska-Kippen, N. (2008). Spatial distribution of phytoplankton communities in small water body. *Roczniki Akademii Rolniczej w Poznaniu. Botanika-Steciana*, *12*. http://agro.icm.edu.pl/agro/element/bwmeta1.element.agro-09ba7268-dd81-45ba-993f-5bff48b31977

Celewicz-Gołdyn, S., & Kuczyńska-Kippen, N. (2017). Ecological value of macrophyte cover in creating habitat for microalgae (diatoms) and zooplankton (rotifers and crustaceans) in small field and forest water bodies. *PLOS ONE*, *12*(5), e0177317. https://doi.org/10.1371/journal.pone.0177317

Cellamare, M., Rolland, A., & Jacquet, S. (2010). Flow cytometry sorting of freshwater phytoplankton. *Journal of Applied Phycology*, *22*(1), 87–100. https://doi.org/10.1007/s10811-009-9439-4

Chau, K. (2006). A review on integration of artificial intelligence into water quality monitoring. *Marine Pollution Bulletin*, *52*, 726–733.

Chellappa, N., Câmara, F., & Rocha, O. (2009). Phytoplankton community: Indicator of water quality in the Armando Ribeiro Gonçalves Reservoir and Pataxó Channel, Rio Grande do Norte, Brazil. *Brazilian Journal of Biology*, *69*(2), 241–251.

Chen, Y., Qin, B., Teubner, K., & Dokulil, M. T. (2003). Long-term dynamics of phytoplankton assemblages: *Microcystis*-domination in Lake Taihu, a large shallow lake in China. *Journal of Plankton Research*, *25*(4), 445–453. https://doi.org/10.1093/plankt/25.4.445

Cheng, Y., Bhoot, V. N., Kumbier, K., Sison-Mangus, M. P., Brown, J. B., Kudela, R., & Newcomer, M. E. (2021). A novel random forest approach to revealing interactions and controls on chlorophyll concentration and bacterial communities during coastal phytoplankton blooms. *Scientific Reports*, *11*(1), 19944. https://doi.org/10.1038/s41598-021-98110-9

Cheruvelil, K. S., Soranno, P. A., Bremigan, M. T., Wagner, T., & Martin, S. L. (2008). Grouping Lakes for Water Quality Assessment and Monitoring: The Roles of Regionalization and Spatial Scale. *Environmental Management*, *41*(3), 425–440. https://doi.org/10.1007/s00267-007-9045-7

Chopyk, J., Allard, S., Nasko, D. J., Bui, A., Mongodin, E. F., & Sapkota, A. R. (2018). Agricultural Freshwater Pond Supports Diverse and Dynamic Bacterial and Viral Populations. *Frontiers in Microbiology*, *0*. https://doi.org/10.3389/fmicb.2018.00792

Chorus, I., & Bartram, J. (Eds.). (1999). *Toxic cyanobacteria in water: A guide to their public health consequences, monitoring, and management*. E & FN Spon.

Chowdhury, M., Hardikar, R., Chanjaplackal Kesavan, H., Thomas, J., Mitra, A., Rokade, M. A., Naidu, V. S., & Sukumaran, S. (2017). Nutrient stoichiometry and freshwater flow in shaping of phytoplankton population in a tropical

monsoonal estuary (Kundalika Estuary). *Estuarine, Coastal and Shelf Science*, *198*, 73–91. https://doi.org/10.1016/j.ecss.2017.08.019

Cloern, J. E., Alpine, A. E., Cole, B. E., & Heller, T. (1992). Seasonal changes in the spatial distribution of phytoplankton in small, temperate-zone lakes. *Journal of Plankton Research*, *14*(7), 1017–1024. https://doi.org/10.1093/plankt/14.7.1017

Corbel, S., Mougin, C., & Bouaïcha, N. (2014). Cyanobacterial toxins: Modes of actions, fate in aquatic and soil ecosystems, phytotoxicity and bioaccumulation in agricultural crops. *Chemosphere*, *96*, 1–15. https://doi.org/10.1016/j.chemosphere.2013.07.056

Corbel, S., Mougin, C., Nélieu, S., Delarue, G., & Bouaïcha, N. (2016). Evaluation of the transfer and the accumulation of microcystins in tomato (*Solanum lycopersicum* cultivar MicroTom) tissues using a cyanobacterial extract containing microcystins and the radiolabeled microcystin-LR ((14)C-MC-LR). *The Science of the Total Environment*, *541*, 1052–1058. https://doi.org/10.1016/j.scitotenv.2015.10.004

Crossetti, L. O., Becker, V., Cardoso, L. de S., Rodrigues, L. R., da Costa, L. S., & da Motta-Marques, D. (2013). Is phytoplankton functional classification a suitable tool to investigate spatial heterogeneity in a subtropical shallow lake? *Limnologica*, *43*(3), 157–163. https://doi.org/10.1016/j.limno.2012.08.010

Dale, B. (2000). Dinoflagellate cysts as indicators of cultural eutrophication and industrial pollution in coastal sediments. In R. E. Martin (Ed.), *Environmental Micropaleontology: The Application of Microfossils to Environmental*

Geology (pp. 305–321). Springer US. https://doi.org/10.1007/978-1-4615-4167-7_14

Dale, B., Thorsen, T. A., & Fjellså, A. (1999). Dinoflagellate cysts as indicators of cultural eutrophication in the oslofjord, norway. *Estuarine, Coastal and Shelf Science*, *48*(3), 371–382.

Davis, T. W., Berry, D. L., Boyer, G. L., & Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms. *Harmful Algae*, *8*(5), 715–725. https://doi.org/10.1016/j.hal.2009.02.004

de Figueiredo, D. R., Reboleira, A. S. S. P., Antunes, S. C., Abrantes, N., Azeiteiro, U., Gonçalves, F., & Pereira, M. J. (2006). The effect of environmental parameters and cyanobacterial blooms on phytoplankton dynamics of a Portuguese temperate lake. *Hydrobiologia*, *568*(1), 145–157. https://doi.org/10.1007/s10750-006-0196-y

DeLorenzo, M. E., M., Taylor, L. A., Lund, S. A., Pennington, E. D., Strozier, E. D., & Fulton, M. H. (2002). Toxicity and bioconcentration potential of the agricultural pesticide endosulfan in phytoplankton and zooplankton. *Archives of Environmental Contamination and Toxicology*, *42*(2), 173–181. https://doi.org/10.1007/s00244-001-0008-3

Dembowska, E. A., Mieszczankin, T., & Napiórkowski, P. (2018). Changes of the phytoplankton community as symptoms of deterioration of water quality in a shallow lake. *Environmental Monitoring and Assessment*, *190*(2), 95. https://doi.org/10.1007/s10661-018-6465-1

Deng, D.-G., Xie, P., Zhou, Q., Yang, H., & Guo, L.-G. (2007). Studies on Temporal and Spatial Variations of Phytoplankton in Lake Chaohu. *Journal of Integrative Plant Biology*, *49*(4), 409–418. https://doi.org/10.1111/j.1744-7909.2007.00390.x

Derot, J., Yajima, H., & Schmitt, F. G. (2020). Benefits of machine learning and sampling frequency on phytoplankton bloom forecasts in coastal areas. *Ecological Informatics*, *60*, 101–174. https://doi.org/10.1016/j.ecoinf.2020.101174

Descy, J.-P., Leprieur, F., Pirlot, S., Leporcq, B., Van Wichelen, J., Peretyatko, A., Teissier, S., Codd, G. A., Triest, L., Vyverman, W., & Wilmotte, A. (2016). Identifying the factors determining blooms of cyanobacteria in a set of shallow lakes. *Ecological Informatics*, *34*, 129–138. https://doi.org/10.1016/j.ecoinf.2016.05.003

Di Maggio, J., Fernández, C., Parodi, E. R., Diaz, M. S., & Estrada, V. (2016). Modeling phytoplankton community in reservoirs. A comparison between taxonomic and functional groups-based models. *Journal of Environmental Management*, *165*, 31–52. https://doi.org/10.1016/j.jenvman.2015.08.027

Dia, S. A. (2016). *Efficacy of three algaecides on algal blooms in hypereutrophic lakes.* [Master's Thesis]. American University of Beirut.

Diaz, R. J., & Breitburg, D. L. (2009). Chapter 1 The Hypoxic Environment. In J. G. Richards, A. P. Farrell, & C. J. Brauner (Eds.), *Fish Physiology* (Vol. 27, pp. 1–23). Academic Press. https://doi.org/10.1016/S1546-5098(08)00001-0

Ding, Y., Pan, B., Zhao, G., Sun, C., Han, X., & Li, M. (2021). Geo-climatic factors weaken the effectiveness of phytoplankton diversity as a water quality indicator in a large sediment-laden river. *Science of The Total Environment*, *792*, 148346. https://doi.org/10.1016/j.scitotenv.2021.148346

Douterelo, I., Perona, E., & Mateo, P. (2004). Use of cyanobacteria to assess water quality in running waters. *Environmental Pollution*, *127*(3), 377–384.

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., McDowell, W. H., Kortelainen, P., Caraco, N. F., Melack, J. M., & Middelburg, J. J. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, *51*(5), 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Duarte, C. M., Agustí, S., & Canjield Jr., D. E. (1992). Patterns in phytoplankton community structure in Florida lakes. *Limnology and Oceanography*, *37*(1), 155–161. https://doi.org/10.4319/lo.1992.37.1.0155

Dunker, S., Nadrowski, K., Jakob, T., Kasprzak, P., Becker, A., Langner, U., Kunath, C., Harpole, S., & Wilhelm, C. (2016). Assessing in situ dominance pattern of phytoplankton classes by dominance analysis as a proxy for realized niches. *Harmful Algae*, *58*, 74–84. https://doi.org/10.1016/j.hal.2016.08.004

Duong, T. T., Hoang, T. T. H., Nguyen, T. K., Le, T. P. Q., Le, N. D., Dang, D. K., Lu, X., Bui, M. H., Trinh, Q. H., Dinh, T. H. V., Pham, T. D., & Rochelle-newall, E. (2019). Factors structuring phytoplankton community in a large tropical river: Case study in the Red River (Vietnam). *Limnologica*, *76*, 82–93. https://doi.org/10.1016/j.limno.2019.04.003

Effler, S. W., Litten, S., Field, S. D., Tong-Ngork, T., Hale, F., Meyer, M., & Quirk, M. (1980). Whole lake responses to low level copper sulfate treatment. *Water Research*, *14*(10), 1489–1499. https://doi.org/10.1016/0043-1354(80)90015-9

Elder, J. F., & Horne, A. J. (1978). Copper cycles and $CuSO_4$ algicidal capacity in two California lakes. *Environmental Management*, *2*(1), 17–30. https://doi.org/10.1007/BF01866443

Environmental Protection Agency. (1997). *In Vitro Determination of Chlorophyll a and Pheophytin a in Marine and Freshwater Algae by Fluorescence*. *EPA-600R15006*, 22.

Environmental Protection Agency. (2019). *Recommendations for Cyanobacteria and Cyanotoxin Monitoring in Recreational Waters*. *EPA-823-R-19-001*, 16.

Feki-Sahnoun, W., Hamza, A., Béjaoui, B., Mahfoudi, M., Rebai, A., & Bel Hassen, M. (2018). Multi-table approach to assess the biogeography of phytoplankton: Ecological and management implications. *Hydrobiologia*, *815*(1), 229–251. https://doi.org/10.1007/s10750-018-3566-3

Feki-Sahnoun, W., Hamza, A., Mahfoudi, M., Rebai, A., & Hassen, M. B. (2014). Long-term microphytoplankton variability patterns using multivariate analyses: Ecological and management implications. *Environmental Science and Pollution Research*, *21*(19), 11481–11499. https://doi.org/10.1007/s11356-014-3009-2

Ferral, A., Solis, V., Frery, A., Orueta, A., Bernasconi, I., Bresciano, J., & Scavuzzo, C. M. (2017). Spatio-temporal changes in water quality in an eutrophic lake

with artificial aeration. *Journal of Water and Land Development*, *35*(1), 27–40. https://doi.org/10.1515/jwld-2017-0065

Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., & Raven, J. A. (2010). Phytoplankton in a changing world: Cell size and elemental stoichiometry. *Journal of Plankton Research*, *32*(1), 119–137. https://doi.org/10.1093/plankt/fbp098

Fornarelli, R., Galelli, S., Castelletti, A., Antenucci, J. P., & Marti, C. L. (2013). An empirical modeling approach to predict and understand phytoplankton dynamics in a reservoir affected by interbasin water transfers. *Water Resources Research*, *49*(6), 3626–3641. https://doi.org/10.1002/wrcr.20268

Foster, G. M., Graham, J. L., & King, L. R. (2019). Spatial and temporal variability of harmful algal blooms in Milford Lake, Kansas, May through November 2016: U.S. *Geological Survey Scientific Investigations Report*, 1–46.

Fragoso, C. R., Marques, D. M. L. M., Collischonn, W., Tucci, C. E. M., & van Nes, E. H. (2008). Modelling spatial heterogeneity of phytoplankton in Lake Mangueira, a large shallow subtropical lake in South Brazil. *Ecological Modelling*, *219*(1), 125–137. https://doi.org/10.1016/j.ecolmodel.2008.08.004

Fu, F., Tatters, A., & Hutchins, D. (2012). Global change and the future of harmful algal blooms in the ocean. *Marine Ecology Progress Series*, *470*, 207–233. https://doi.org/10.3354/meps10047

Gayoso, A. (1998). Long-term phytoplankton studies in the Bahía Blanca estuary, Argentina. *ICES Journal of Marine Science*, *55*(4), 655–660. https://doi.org/10.1006/jmsc.1998.0375

Gemza, A. F. (1997). Water Quality Improvements during Hypolimnetic Oxygenation in Two Ontario Lakes. *Water Quality Research Journal of Canada*, *32*(2), 365–390.

Genzoli, L., & Kann, J. (2016). *Evaluation of phycocyanin probes as a monitoring tool for toxigenic cyanobacteria in the Klamath River below Iron Gate Dam*. Klamath Tribal Water Consortium. https://doi.org/10.13140/RG.2.2.23897.31841

Giardino, C., Candiani, G., Bresciani, M., Lee, Z., Gagliano, S., & Pepe, M. (2012). BOMBER: A tool for estimating water quality and bottom properties from remote sensing images. *Computers & Geosciences*, *45*, 313–318.

Gitelson, A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., & Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, *112*, 3582–3593.

Glibert, P. M., & Burkholder, J. M. (2006). The Complex Relationships Between Increases in Fertilization of the Earth, Coastal Eutrophication and Proliferation of Harmful Algal Blooms. In E. Granéli & J. T. Turner (Eds.), *Ecology of Harmful Algae* (pp. 341–354). Springer. https://doi.org/10.1007/978-3-540-32210-8_26

Gorokhova, O. G., & Zinchenko, T. D. (2019). Phytoplankton of the Usa River and its tributaries (Kuibyshev Reservoir Basin). *Biology Bulletin*, *46*(10), 1382–1389. https://doi.org/10.1134/S1062359019100121

Greenfield, D. I., Duquette, A., Goodson, A., Keppler, C. J., Williams, S. H., Brock, L. M., Stackley, K. D., White, D., & Wilde, S. B. (2014). The effects of three chemical algaecides on cell numbers and toxin content of the cyanobacteria *Microcystis aeruginosa* and *Anabaenopsis* sp. *Environmental Management*, *54*(5), 1110–1120. https://doi.org/10.1007/s00267-014-0339-2

Hagerthey, S. E., & Kerfoot, W. C. (2005). Spatial variation in groundwater-related resource supply influences freshwater benthic algal assemblage composition. *Journal of the North American Benthological Society*, *24*, 807–819.

Hangwelani, S., Matjutla John Matjutla, M., & Ray, G. J. (2021). Impact of pigeon pea fish feed formula on the limnology of small-holder aquaculture systems during tilapia fish feeding trials, Vhembe District, Limpopo Province. *International Journal of Environmental Impacts: Management, Mitigation and Recovery*, *4*(4), 388–399. https://doi.org/10.2495/EI-V4-N4-388-399

Haque, Md. A., Jewel, Md. A. S., Akhi, Most. M., Atique, U., Paul, A. K., Iqbal, S., Islam, M. S., Das, S. K., & Alam, M. M. (2021). Seasonal dynamics of phytoplankton community and functional groups in a tropical river. *Environmental Monitoring and Assessment*, *193*(11), 704. https://doi.org/10.1007/s10661-021-09500-5

Harris, T. D., & Graham, J. L. (2017). Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking water reservoir using a 14-year dataset. *Lake and Reservoir Management*, *33*, 32–48.

Harrison, J. W., Lucius, M. A., Farrell, J. L., Eichler, L. W., & Relyea, R. A. (2021). Prediction of stream nitrogen and phosphorus concentrations from high-

frequency sensors using Random Forests Regression. *Science of The Total Environment*, *763*, 143005. https://doi.org/10.1016/j.scitotenv.2020.143005

Havens, K. E. (2008). Cyanobacteria blooms: Effects on aquatic ecosystems. *Advances in Experimental Medicine and Biology*, *619*, 733–747.

Havens, K. E., James, R. T., East, T. L., & Smith, V. H. (2003). N:P ratios, light limitation, and cyanobacterial dominance in a subtropical lake impacted by non-point source nutrient pollution. *Environmental Pollution*, *122*(3), 379–390.

Havens, K. E., Ji, G., Beaver, J. R., Fulton, R. S., & Teacher, C. E. (2019). Dynamics of cyanobacteria blooms are linked to the hydrology of shallow Florida lakes and provide insight into possible impacts of climate change. *Hydrobiologia*, *829*(1), 43–59. https://doi.org/10.1007/s10750-017-3425-7

Heddam, S., Sanikhani, H., & Kisi, O. (2019). Application of artificial intelligence to estimate phycocyanin pigment concentration using water quality data: A comparative study. *Applied Water Science*, *9*(7), 164. https://doi.org/10.1007/s13201-019-1044-3

Hernandez Cordero, A. L., Tango, P. J., & Batiuk, R. A. (2020). Development of a multimetric water quality indicator for tracking progress towards the achievement of Chesapeake Bay water quality standards. *Environmental Monitoring and Assessment*, *192*(2), 94. https://doi.org/10.1007/s10661-019-7969-z

Hilborn, E. D., & Beasley, V. R. (2015). One Health and cyanobacteria in freshwater systems: Animal illnesses and deaths are sentinel events for human health risks. *Toxins*, *7*, 1374–1395.

Hinder, S. L., Hays, G. C., Edwards, M., Roberts, E. C., Walne, A. W., & Gravenor, M. B. (2012). Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*, *2*(4), 271–275. https://doi.org/10.1038/nclimate1388

Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, *574*(7780), 667–670. https://doi.org/10.1038/s41586-019-1648-7

Horváth, H., Kovács, A. W., Riddick, C., & Présing, M. (2013). Extraction methods for phycocyanin determination in freshwater filamentous cyanobacteria and their application in a shallow lake. *European Journal of Phycology*, *48*(3), 278–286.

Hu, C., Muller-Karger, F. E., & Swarzenski, P. W. (2006). Hurricanes, submarine groundwater discharge, and Florida's red tides. *Geophys. Res. Lett*, *33*, 11601,. https://doi.org/10.1029/2005GL025449.

Hu, C., Ou, Y., Zhang, D., Zhang, H., Yan, C., Zhao, Y., & Zheng, Z. (2012). Phytoremediation of the polluted Waigang River and general survey on variation of phytoplankton population. *Environmental Science and Pollution Research*, *19*(9), 4168–4175. https://doi.org/10.1007/s11356-012-0931-z

Huang, J., Gao, J., Hörmann, G., & Fohrer, N. (2014). Modeling the effects of environmental variables on short-term spatial changes in phytoplankton

biomass in a large shallow lake, Lake Taihu. *Environmental Earth Sciences*, *72*(9), 3609–3621. https://doi.org/10.1007/s12665-014-3272-z

Hudon, C., De Sève, M., & Cattaneo, A. (2014). Increasing occurrence of the benthic filamentous cyanobacterium Lyngbya wollei: A symptom of freshwater ecosystem degradation. *Freshwater Science*, *33*, 606–618.

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483.

Izydorczyk, K., Carpentier, C., Mrówczyński, J., Wagenvoort, A., Jurczak, T., & Tarczyńska, M. (2009). Establishment of an Alert Level Framework for cyanobacteria in drinking water resources by using the Algae Online Analyser for monitoring cyanobacterial chlorophyll a. *Water Research*, *43*(4), 989–996. https://doi.org/10.1016/j.watres.2008.11.048

Izydorczyk, K., Tarczynska, M., Jurczak, T., Mrowczynski, J., & Zalewski, M. (2005). Measurement of phycocyanin fluorescenceas an online early warning system for cyanobacteria in reservoir intake water. *Environmental Toxicology*, *20*(4), 425–430. https://doi.org/10.1002/tox.20128

Jakhar, P. (2013). Role of phytoplankton and zooplankton as health indicators of aquatic ecosystem: A review. *International Journal of Innovative Research and Studies*, *2*, 489–500.

Jensen, J., Jeppesen, E., Olrik, K., & Kristensen, P. (1994). Impact of nutrients and physical factors on the shift from cyanobacteria to chlorophyte dominance in

shallow Danish Lakes. *Canadian Journal of Fishing and Aquatic Science*, *51*, 1692–1699.

Ji, G., & Havens, K. (2019). Periods of extreme shallow depth hinder but do not stop long-term improvements of water quality in Lake Apopka, Florida (USA. *Water*, *11*, 538,. https://doi.org/10.3390/w11030538.

Jia, J., Gao, Y., Song, X., & Chen, S. (2019). Characteristics of phytoplankton community and water net primary productivity response to the nutrient status of the Poyang Lake and Gan River, China. *Ecohydrology*, *12*(7), e2136. https://doi.org/10.1002/eco.2136

Jiang, Y.-J., He, W., Liu, W.-X., Qin, N., Ouyang, H.-L., Wang, Q.-M., Kong, X.-Z., He, Q.-S., Yang, C., Yang, B., & Xu, F.-L. (2014). The seasonal and spatial variations of phytoplankton community and their correlation with environmental factors in a large eutrophic Chinese lake (Lake Chaohu). *Ecological Indicators*, *40*, 58–67. https://doi.org/10.1016/j.ecolind.2014.01.006

Jin, Y., Yu, R., Zhang, Z., Zhang, Q., Li, M., Cao, Z., Wu, L., & Hao, Y. (2020). Spatiotemporal variability of phytoplankton functional groups in a shallow eutrophic lake from cold, arid regions. *Environmental Monitoring and Assessment*, *192*(6), 371. https://doi.org/10.1007/s10661-020-08349-4

John, D. M., Whitton, B. A., & Brook, A. J. (2011). *The freshwater algal flora of the British Isles: An identification guide to freshwater and terrestrial algae*. Cambridge University Press. London. 724 p.

Jongman, M., & Korsten, L. (2018). Irrigation water quality and microbial safety of leafy greens in different vegetable production systems: A review. *Food Reviews International*, *34*(4), 308–328. https://doi.org/10.1080/87559129.2017.1289385

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kane, D. D., Gordon, S. I., Munawar, M., Charlton, M. N., & Culver, D. A. (2009). The Planktonic Index of Biotic Integrity (P-IBI): An approach for assessing lake ecosystem health. *Ecological Indicators*, *9*(6), 1234–1247. https://doi.org/10.1016/j.ecolind.2009.03.014

Karadžić, V., Simić, G. S., Natić, D., Ržaničanin, A., Ćirić, M., & Gačić, Z. (2013). Changes in the phytoplankton community and dominance of *Cylindrospermopsis raciborskii* (Wolosz.) Subba Raju in a temperate lowland river (Ponjavica, Serbia). *Hydrobiologia*, *711*(1), 43–60. https://doi.org/10.1007/s10750-013-1460-6

Kasinak, J. M. E., Holt, B. M., Chislock, M. F., & Wilson, A. E. (2015). Benchtop fluorometry of phycocyanin as a rapid approach for estimating cyanobacterial biovolume. *Journal of Plankton Research*, *37*(1), 248–257.

Kauppila, P., Hällfors, G., Kangas, P., Kokkonen, P., & Basova, S. (1995). Late summer phytoplankton species composition and biomasses in the Eastern Gulf of Finland. *Ophelia*, *42*(1), 179–191. https://doi.org/10.1080/00785326.1995.10431503

Kehoe, M. J., Chun, K. P., & Baulch, H. M. (2015). Who smells? Forecasting taste and odor in a drinking water reservoir. *Environmental Science & Technology*, *49*(18), 10984–10992. https://doi.org/10.1021/acs.est.5b00979

Khaliullina, L. Y., & Fazlieva, A. R. (2019). The composition and seasonal dynamics of planktonic green algae in the mouth of the Kazanka River (Republic of Tatarstan). *Journal of Environmental Treatment Techniques*, *5*, 993–998.

Kim, S. Y., Moon, C. H., Cho, H. J., & Lim, D. I. (2009). Dinoflagellate cysts in coastal sediments as indicators of eutrophication: A case of Gwangyang Bay, South sea of Korea. *Estuaries and Coasts*, *32*(6), 1225–1233.

Kislik, C., Dronova, I., & Kelly, M. (2018). UAVs in support of algal bloom research: A review of current applications and future opportunities. *Drones*, *2*, 35.

Kittler, K., Schreiner, M., Krumbein, A., Manzei, S., Koch, M., Rohn, S., & Maul, R. (2012). Uptake of the cyanobacterial toxin cylindrospermopsin in Brassica vegetables. *Food Chemistry*, *133*, 875–879.

Komárek, J., Kaštovský, J., Mareš, J., & Johansen, J. R. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia*, *86*, 295–335.

Kong, Y., Lou, I., Zhang, Y., Lou, C. U., & Mok, K. M. (2014). Using an online phycocyanin fluorescence probe for rapid monitoring of cyanobacteria in Macau freshwater reservoir. *Hydrobiologia*, *741*(1), 33–49.

Konopka, A., & Brock, T. D. (1978). Effect of temperature on blue-green algae (Cyanobacteria) in Lake Mendota. *Applied and Environmental Microbiology*, *36*(4), 572–576.

Kosten, S., Huszar, V. L. M., Bécares, E., Costa, L. S., Donk, E., Hansson, L. A., Jeppesen, E., Kruk, C., Lacerat, G., Mazzeo, N., Meester, L. D., Moss, B., Lürling, M., Nõges, T., Romo, S., & Scheffer, M. (2011). Warmer climates boost cyanobacterial dominance in shallow lakes. *Global Change Biology*, *18*(1), 118–126.

Krajčovič, J., Vesteg, M., & Schwartzbach, S. D. (2015). Euglenoid flagellates: A multifaceted biotechnology platform. *Journal of Biotechnology*, *202*, 135–145.

Kutser, T., Hedley, J., Giardino, C., Roelfsema, C., & Brando, V. E. (2020). Remote sensing of shallow waters – a 50-year retrospective and future directions. *Remote Sensing of Environment*, *240*, 111619.

Lacouture, R. V., Johnson, J. M., Buchanan, C., & Marshall, H. G. (2006). Phytoplankton Index of Biotic Integrity for Chesapeake Bay and its tidal tributaries. *Estuaries and Coasts*, *29*(4), 598–616.

Lamlou, V. W. (1977). *Distribution of phytoplankton in Maryland lakes*. Environmental Protection Agency, Office of Research and Development, Environmental Monitoring and Support Laboratory, Monitoring Operations Division.

Lange-Bertalot, H. (1979). Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwigia*, *64*, 285–304.

Lawton, L., Marsalek, B., Padisak, J., & Chorus, I. (1999). Determination of cyanobacteria in the laboratory. In I. Chorus & J. Bartram (Eds.), *Toxic Cyanobacteria in Water: A guide to their public health consequences, monitoring and management*. E & FN Spon.

Lee, S., Jiang, X., Manubolu, M., Riedl, K., Ludsin, S. A., Martin, J. F., & Lee, J. (2017). Fresh produce and their soils accumulate cyanotoxins from irrigation water: Implications for public health and food security. *Food Research International*, *102*, 234–245. https://doi.org/10.1016/j.foodres.2017.09.079

Lee, T. A., Rollwagen-Bollens, G., & Bollens, S. M. (2015a). The influence of water quality variables on cyanobacterial blooms and phytoplankton community composition in a shallow temperate lake. *Environmental Monitoring and Assessment*, *187*(6), 315. https://doi.org/10.1007/s10661-015-4550-2.

Lee, T. A., Rollwagen-Bollens, G., Bollens, S. M., & Faber-Hammond, J. J. (2015). Environmental influence on cyanobacteria abundance and microcystin toxin production in a shallow temperate lake. *Ecotoxicology and Environmental Safety*, *114*, 318–325. https://doi.org/10.1016/j.ecoenv.2014.05.004

Leigh, C., Kandanaarachchi, S., McGree, J. M., Hyndman, R. J., Alsibai, O., Mengersen, K., & Peterson, E. E. (2019). Predicting sediment and nutrient concentrations from high-frequency water-quality data. *PLOS ONE*, *14*(8), e0215503. https://doi.org/10.1371/journal.pone.0215503

Lewis, W. M. (1978). Spatial distribution of the phytoplankton in a tropical lake (Lake Lanao, Philippines). *Internationale Revue Der Gesamten Hydrobiologie*

Und Hydrographie, *63*(5), 619–635. https://doi.org/10.1002/iroh.19780630508

Li, F., Zhang, H., Zhu, Y., Xiao, Y., & Chen, L. (2013). Effect of flow velocity on phytoplankton biomass and composition in a freshwater lake. *Science of The Total Environment*, *447*, 64–71. https://doi.org/10.1016/j.scitotenv.2012.12.066

Li, J., Glibert, P. M., & Gao, Y. (2015). Temporal and spatial changes in Chesapeake Bay water quality and relationships to *Prorocentrum minimum, Karlodinium veneficum*, and CyanoHAB events, 1991–2008. *Harmful Algae*, *42*, 1–14. https://doi.org/10.1016/j.hal.2014.11.003

Li, L., Li, L., & Song, K. (2015). Remote sensing of freshwater cyanobacteria: An extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. *Remote Sensing of Environment*, *157*, 9–23.

Li, L., Sengpiel, R. E., Pascual, D. L., Tedesco, L. P., Wilson, J. S., & Soyeux, A. (2010). Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. *International Journal of Remote Sensing*, *31*(15), 4147–4162.

Liu, X., Feng, J., & Wang, Y. (2019). Chlorophyll a predictability and relative importance of factors governing lake phytoplankton at different timescales. *Science of The Total Environment*, *648*, 472–480. https://doi.org/10.1016/j.scitotenv.2018.08.146

Liu, Y., Xi, D.-G., & Li, Z.-L. (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), e0119082. https://doi.org/10.1371/journal.pone.0119082

Loh, W. (2011). Classification and regression trees. *Wires Data Mining and Knowledge Discovery*, *1*(1), 14–23.

López-Felices, B., Aznar-Sánchez, J. A., Velasco-Muñoz, J. F., & Piquer-Rodríguez, M. (2020). Contribution of irrigation ponds to the sustainability of agriculture. A review of worldwide research. *Sustainability*, *12*(13), 5425. https://doi.org/10.3390/su12135425

Machado, J., Campos, A., Vasconcelos, V., & Freitas, M. (2017). Effects of microcystin-LR and cylindrospermopsin on plant-soil systems: A review of their relevance for agricultural plant quality and public health. *Environmental Research*, *153*, 191–204. https://doi.org/10.1016/j.envres.2016.09.015

Maestre-Valero, J. F., & Martínez-Alvarez, V. (2010). Effects of drip irrigation systems on the recovery of dissolved oxygen from hypoxic water. *Agricultural Water Management*, *97*(11), 1806–1812.

Mäkinen, A., & Aulio, K. (1986). The use of *Cladophora glomeratae* (Chlorophyta) as an indicator of coastal eutrophication. *Publ. Water Res. Inst. Finland*, *68*, 160–163.

Marić, D., Kraus, R., Godrijan, J., Supić, N., Djakovac, T., & Precali, R. (2012). Phytoplankton response to climatic and anthropogenic influences in the north-

eastern Adriatic during the last four decades. *Estuarine, Coastal and Shelf Science*, *115*, 98–112. https://doi.org/10.1016/j.ecss.2012.02.003

Marion, J. W., Lee, J., Wilkins, J. R., Lemeshow, S., Lee, C., Waletzko, E. J., & Buckley, T. J. (2012). In vivo phycocyanin flourometry as a potential rapid screening tool for predicting elevated microcystin concentrations at eutrophic lakes. *Environmental Science & Technology*, *46*(8), 4523–4531. https://doi.org/10.1021/es203962u

Marshall, H. G. (1967). Plantkon in James River Estuary, Virginia. I. Phytoplankton in Willoughby Bay and Hampton Roads. *Chesapeake Science*, *8*(2), 90–101. https://doi.org/10.2307/1351153

Marshall, H. G. (2009). Phytoplankton of the York River. *Journal of Coastal Research*, *S1*, 59–65. https://doi.org/10.2112/1551-5036-57.sp1.59

Marshall, H. G. (2013). Phytoplankton in Virginia lakes and reservoirs. *Virginia Journal of Science*, *64*(1–2), 3–15. https://doi.org/doi:10.25778/F2RH-ZD44

Marshall, H. G. (2014). Phytoplankton in Virginia lakes and reservoirs: Part II. *Virginia Journal of Science*, *65*(1), 3–8. https://doi.org/10.25778/fq7v-mt85

Marshall, H. G., & Alden, R. W. (1990). A comparison of phytoplankton assemblages and environmental relationships in three estuarine rivers of the Lower Chesapeake Bay. *Estuaries*, *13*, 287–300.

Marshall, H. G., Burchardt, L., & Lacouture, R. (2005). A review of phytoplankton composition within Chesapeake Bay and its tidal estuaries. *Journal of Plankton Research*, *27*(11), 1083–1102. https://doi.org/10.1093/plankt/fbi079

Marshall, H. G., Lacouture, R. V., Buchanan, C., & Johnson, J. M. (2006). Phytoplankton assemblages associated with water quality and salinity regions in Chesapeake Bay, USA. *Estuarine, Coastal and Shelf Science*, *69*(1), 10–18. https://doi.org/10.1016/j.ecss.2006.03.019

Marshall, H. G., Lane, M. F., Nesius, K. K., & Burchardt, L. (2009). Assessment and significance of phytoplankton species composition within Chesapeake Bay and Virginia tributaries through a long-term monitoring program. *Environmental Monitoring and Assessment*, *150*(1), 143–155. https://doi.org/10.1007/s10661-008-0680-0

Mchau, G. J., Makule, E., Machunda, R., Gong, Y. Y., & Kimanya, M. (2019). Phycocyanin as a proxy for algal blooms in surface waters: Case study of Ukerewe Island, Tanzania. *Water Practice and Technology*, *14*(1), 229–239. https://doi.org/10.2166/wpt.2019.005

McQuaid, N., Zamyadi, A., Prévost, M., F. Bird, D., & Dorner, S. (2011). Use of in vivo phycocyanin fluorescence to monitor potential microcystin -producing cyanobacterial biovolume in a drinking water source. *Journal of Environmental Monitoring*, *13*(2), 455–463. https://doi.org/10.1039/C0EM00163E

Mellios, N., Moe, S. J., & Laspidou, C. (2020). Machine learning approaches for predicting health risk of cyanobacterial blooms in northern European lakes. *Water*, *12*(4), 1191. https://doi.org/10.3390/w12041191

Meyer, R. L. (1971). *A Study of Phytoplankton Dynamics in Lake Fayetteville As A Means of Assessing Water Quality* (Technical Report No. PUB010A; Vol. 10, pp. 1–68).

Miller, A., & Russell, C. (2017). Food crops irrigated with cyanobacteria-contaminated water: An emerging public health issue in Canada. *Environmental Health Review*, *60*(3), 58–63. https://doi.org/10.5864/d2017-021

Millie, D. F., Fahnenstiel, G. L., Carrick, H. J., Ardjmand, E., Young II, W. A., Sayers, M. J., & Shuchman, R. A. (2014). Using artificial intelligence for CyanoHAB niche modeling: Discovery and visualization of *Microcystis*–environmental associations within western Lake Erie. *Canadian Journal of Fisheries & Aquatic Sciences*, *71*, 1642–1654.

Mishra, S., Mishra, D. R., Lee, Z., & Tucker, C. S. (2013). Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: A quasi-analytical approach. *Remote Sensing of Environment*, *133*, 141–151.

Moreno-Ostos, E., Cruz-Pizarro, L., Basanta, A., & Glen George, D. (2008). The spatial distribution of different phytoplankton functional groups in a Mediterranean reservoir. *Aquatic Ecology*, *42*(1), 115–128. https://doi.org/10.1007/s10452-007-9087-1

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Nakayama, F. S., & Bucks, D. A. (1991). Water quality in drip/trickle irrigation: A review. *Irrigation Science*, *12*(4), 187–192. https://doi.org/10.1007/BF00190522

Naselli-Flores, L., & Padisák, J. (2016). Blowing in the wind: How many roads can a phytoplankton walk down? A synthesis on phytoplankton biogeography and spatial processes. *Hydrobiologia*, *764*(1), 303–313. https://doi.org/10.1007/s10750-015-2519-3

Newell, S. E., Davis, T. W., Johengen, T. H., Gossiaux, D., Burtner, A., Palladino, D., & McCarthy, M. J. (2019). Reduced forms of nitrogen are a driver of non-nitrogen-fixing harmful cyanobacterial blooms and toxicity in Lake Erie. *Harmful Algae*, *81*, 86–93. https://doi.org/10.1016/j.hal.2018.11.003

Nguyen, A. T., Dao, T.-S., Strady, E., Nguyen, T. T. N., Aimé, J., Gratiot, N., & Némery, J. (2022). Phytoplankton characterization in a tropical tidal river impacted by a megacity: The case of the Saigon River (Southern Vietnam). *Environmental Science and Pollution Research*, *29*(3), 4076–4092. https://doi.org/10.1007/s11356-021-15850-x

Nishikawa, T., Hori, Y., Nagai, S., Miyahara, K., Nakamura, Y., Harada, K., Tanda, M., Manabe, T., & Tada, K. (2010). Nutrient and phytoplankton dynamics in Harima-Nada, Eastern Seto Inland Sea, Japan during a 35-year period from 1973 to 2007. *Estuaries and Coasts*, *33*(2), 417–427. https://doi.org/10.1007/s12237-009-9198-0

Nweze, N. (2009). Algal diversity of Adada River, Nigeria. I. Chlorophyta (green algae) and Euglenophyta (euglenoids). *Plant Products Research Journal*, *13*(1).

O'Neil, J. M., Davis, T. W., Burford, M. A., & Gobler, C. J. (2012). The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae*, *14*, 313–334.

Otsuki, A., Omi, T., Hashimoto, S., Aizaki, M., & Takamura, N. (1994). HPLC fluorometric determination of natural phytoplankton phycocyanin and its usefulness as cyanobacterial biomass in highly eutrophic shallow lake. *Water, Air, and Soil Pollution*, *76*, 383–396.

Otten, T. G., Xu, H., Qin, B., Zhu, G., & Paerl, H. W. (2012). Spatiotemporal patterns and ecophysiology of toxigenic *Microcystis* blooms in Lake Taihu, China: Implications for water quality management. *Environmental Science & Technology*, *46*(6), 3480–3488. https://doi.org/10.1021/es2041288

Pachepsky, Y., Kierzewski, R., Stocker, M., Sellner, K., Mulbry, W., Lee, H., & Kim, M. (2017). Temporal stability of *Escherichia coli* concentrations in waters of two irrigation ponds in Maryland. *Applied and Environmental Microbiology*, *84*(3), e01876-17, /aem/84/3/e01876-17.atom. https://doi.org/10.1128/AEM.01876-17

Padovesi-Fonseca, C., & Philomeno, M. G. (2004). Effects of algicide (copper sulfate) application on short-term fluctuations of phytoplankton in Lake Paranoá, Central Brazil. *Brazilian Journal of Biology*, *64*(4), 819–826. https://doi.org/10.1590/S1519-69842004000500011

Paerl, H. W. (1996). A comparison of cyanobacterial bloom dynamics in freshwater, estuarine and marine environments. *Phycologia*, *35*(sup6), 25–35. https://doi.org/10.2216/i0031-8884-35-6S-25.1

Paerl, H. W. (2017). Controlling harmful cyanobacterial blooms in a climatically more extreme world: Management options and research needs. *Journal of Plankton Research*, *39*(5), 763–771. https://doi.org/10.1093/plankt/fbx042

Paerl, H. W. (2018). Mitigating toxic planktonic cyanobacterial blooms in aquatic ecosystems facing increasing anthropogenic and climatic pressures. *Toxins*, *10*(2), 76. https://doi.org/10.3390/toxins10020076

Paerl, H. W., & Otten, T. G. (2013). Harmful cyanobacterial blooms: Causes, consequences, and control. *Microbial Ecology*, *65*, 995–1010.

Paerl, H. W., & Scott, J. T. (2010). Throwing fuel on the fire: Synergistic effects of excessive nitrogen inputs and global warming on harmful algal blooms. *Environmental Science & Technology*, *44*(20), 7756–7758. https://doi.org/10.1021/es102665e

Paerl, H. W., & Tucker, C. S. (1995). Ecology of blue-green algae in aquaculture ponds. *Journal of the World Aquaculture Society*, *26*(2), 109–131. https://doi.org/10.1111/j.1749-7345.1995.tb00235.x

Paerl, H. W., Xu, H., McCarthy, M. J., Zhu, G., Qin, B., Li, Y., & Gardner, W. S. (2011). Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Research*, *45*(5), 1973–1983.

Pan, Y., Stevenson, R. J., Hill, B. H., Herlihy, A. T., & Collins, G. B. (1996). Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. *Journal of the North American Benthological Society*, *15*(4), 481–495. https://doi.org/10.2307/1467800

Park, Y., Lee, H. K., Shin, J.-K., Chon, K., Kim, S., Cho, K. H., Kim, J. H., & Baek, S.-S. (2021). A machine learning approach for early warning of cyanobacterial bloom outbreaks in a freshwater reservoir. *Journal of Environmental Management*, *288*, 112415. https://doi.org/10.1016/j.jenvman.2021.112415

Parson, M. J., & Parker, B. C. (1989). Algal Flora in Mountain Lake, Virginia: Past and Present. *Castanea*, *54*(2), 79–86.

Pathak, D., Hutchins, M., Brown, L., Loewenthal, M., Scarlett, P., Armstrong, L., Nicholls, D., Bowes, M., & Edwards, F. (2021). Hourly prediction of phytoplankton biomass and its environmental controls in lowland rivers. *Water Resources Research*, *57*(3). https://doi.org/10.1029/2020WR028773

Patidar, S. K., Chokshi, K., George, B., Bhattacharya, S., & Mishra, S. (2015). Dominance of cyanobacterial and cryptophytic assemblage correlated to CDOM at heavy metal contamination sites of Gujarat, India. *Environmental Monitoring and Assessment*, *187*(1), 1–9.

Patrick, R. (1948). Factors effecting the distribution of diatoms. *The Botanical Review*, *14*(8), 473–524. https://doi.org/10.1007/BF02861575

Patrick, R. (1973). Use of Algae, Especially Diatoms, in the Assessment of Water Quality. *Biological Methods for the Assessment of Water Quality*, 76–95. https://doi.org/10.1520/STP34718S

Paul, M., Negahban-Azar, M., Shirmohammadi, A., & Montas, H. (2021). Developing a Multicriteria Decision Analysis Framework to Evaluate

Reclaimed Wastewater Use for Agricultural Irrigation: The Case Study of Maryland. *Hydrology*, *8*(1), 4. https://doi.org/10.3390/hydrology8010004

Peng, X., Zhang, L., Li, Y., Lin, Q., He, C., Huang, S., Li, H., Zhang, X., Liu, B., Ge, F., Zhou, Q., Zhang, Y., & Wu, Z. (2021). The changing characteristics of phytoplankton community and biomass in subtropical shallow lakes: Coupling effects of land use patterns and lake morphology. *Water Research*, *200*, 117235. https://doi.org/10.1016/j.watres.2021.117235

Perona, E., Bonilla, I., & Mateo, P. (1998). Epilithic cyanobacterial communities and water quality: An alternative tool for monitoring eutrophication in the Alberche River (Spain. *Journal of Applied Phycology*, *10*(2), 183–191.

Pinckney, J. L., Millie, D. F., Vinyard, B. T., & Paerl, H. W. (1997). *Environmental controls of phytoplankton bloom dynamics in the Neuse River Estuary, North Carolina, U.S.A. 54*, 11.

Pinedo, S., García, M., Satta, M. P., Torres, M. de, & Ballesteros, E. (2007). Rocky-shore communities as indicators of water quality: A case study in the Northwestern Mediterranean. *Marine Pollution Bulletin*, *55*(1–6), 126–135.

Pospelova, V., Chmura, G. L., Boothman, W. S., & Latimer, J. S. (2005). Spatial distribution of modern dinoflagellate cysts in polluted estuarine sediments from Buzzards Bay. *Marine Ecology Progress Series*, *292*, 23–40.

Pourafrasyabi, M., & Ramezanpour, Z. (2014). Phytoplankton as bio-indicator of water quality in Sefid Rud River, Iran (South of Caspian Sea). *Caspian J. Env. Sci.*, *12*(1), 31–40.

Powell, T. M., Richerson, P. J., Dillon, T. M., Agee, B. A., Dozier, B. J., Godden, D. A., & Myrup, L. O. (1975). Spatial scales of current speed and phytoplankton biomass fluctuations in Lake Tahoe. *Science*, *189*(4208), 1088–1090. https://doi.org/10.1126/science.189.4208.1088

Pulido, O. (2016). Phycotoxins by harmful algal blooms (HABS) and human poisoning: An overview. *International Clinical Pathology Journal*, *2*. https://doi.org/10.15406/icpjl.2016.02.00062

Pyo, J., Ha, S., Pachepsky, Y. A., Lee, H., Ha, R., Nam, G., Kim, M. S., Im, J., & Cho, K. H. (2016). Chlorophyll-a concentration estimation using three difference bio-optical algorithms, including a correction for the low-concentration range: The case of the Yiam Reservoir, Korea. *Remote Sensing Letters*, *7*, 407–416.

Quetglas, A., Ordines, F., & Guijarro, B. (2011). The use of artificial neural networks (ANNs) in aquatic ecology. In C. L. P. Hui (Ed.), *Artificial neural networks—Application* (pp. 576–586). InTech Publishers.

Ramli, N. M., Giatsis, C., Yusoff, F. M., Verreth, J., & Verdegem, M. (2018). Resistance and resilience of small-scale recirculating aquaculture systems (RAS) with or without algae to pH perturbation. *PLOS ONE*, *13*(4), e0195862. https://doi.org/10.1371/journal.pone.0195862

Randolph, K., Wilson, J., Tedesco, L., Li, L., Pascual, D. L., & Soyeux, E. (2008). Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. *Remote Sensing of Environment*, *112*(11), 4009–4019.

Rao, K., Zhang, X., Wang, M., Liu, J., Guo, W., Huang, G., & Xu, J. (2021). The relative importance of environmental factors in predicting phytoplankton shifting and cyanobacteria abundance in regulated shallow lakes. *Environmental Pollution*, *286*, 117555. https://doi.org/10.1016/j.envpol.2021.117555

Read, D. S., Bowes, M. J., Newbold, L. K., & Whiteley, A. S. (2014). Weekly flow cytometric analysis of riverine phytoplankton to determine seasonal bloom dynamics. *Environmental Science: Processes & Impacts*, *16*(3), 594–603. https://doi.org/10.1039/C3EM00657C

Reid, M. A., Tibby, J. C., Penny, D., & Gell, P. A. (1995). The use of diatoms to assess past and present water quality. *Australian Journal of Ecology*, *20*(1), 57–64.

Ren, X., Yang, K., Che, Y., Wang, M., Zhou, L., & Chen, L. (2016). Spatial and temporal assessment of the initial pattern of phytoplankton population in a newly built coastal reservoir. *Frontiers of Earth Science*, *10*(3), 546–559. https://doi.org/10.1007/s11707-015-0543-2

Rengefors, K., Gustafsson, S., & Ståhl-Delbanco, A. (2004). Factors regulating the recruitment of cyanobacterial and eukaryotic phytoplankton from littoral and profundal sediments. *Aquatic Microbial Ecology*, *36*, 213–226.

Rengefors, K., & Legrand, C. (2001). Toxicity in Peridinium aciculiferum—An adaptive strategy to outcompete other winter phytoplankton? *Limnology and Oceanography*, *46*(8), 1990–1997. https://doi.org/10.4319/lo.2001.46.8.1990

Reynolds, C. S. (1998). What factors influence the species composition of phytoplankton in lakes of different trophic status? *Hydrobiologia*, *369*(0), 11–26. https://doi.org/10.1023/A:1017062213207

Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L., & Melo, S. (2002). Towards a functional classification of the freshwater phytoplankton. *Journal of Plankton Research*, *24*(5), 417–428. https://doi.org/10.1093/plankt/24.5.417

Reynolds, C. S., Maberly, S. C., Parker, J. E., & Ville, M. M. D. (2012). Forty years of monitoring water quality in Grasmere (English Lake District): Separating the effects of enrichment by treated sewage and hydraulic flushing on phytoplankton ecology. *Freshwater Biology*, *57*(2), 384–399. https://doi.org/10.1111/j.1365-2427.2011.02687.x

Riedinger-Whitmore, M. A., Whitmore, T. J., Smoak, J. M., Brenner, M., Moore, A., Curtis, J., & Schelske, C. L. (2005). Cyanobacterial proliferation is a recent response to eutrophication in many Florida lakes: A paleolimnological assessment. *Lake and Reservoir Management*, *21*(4), 423–435.

Rigosi, A., Fleenor, W., & Rueda, F. (2010). State-of-the-art and recent progress in phytoplankton succession modelling. *Environmental Reviews*, *18*(NA), 423–440. https://doi.org/10.1139/A10-021

Rimet, F., & Druart, J.-C. (2018). A trait database for phytoplankton of temperate lakes. *Annales de Limnologie - International Journal of Limnology*, *54*, 18. https://doi.org/10.1051/limn/2018009

Rollwagen-Bollens, G., Bollens, S. M., Gonzalez, A., Zimmerman, J., Lee, T., & Emerson, J. (2013). Feeding dynamics of the copepod *Diacyclops thomasi*

before, during and following filamentous cyanobacteria blooms in a large, shallow temperate lake. *Hydrobiologia*, *705*(1), 101–118. https://doi.org/10.1007/s10750-012-1385-5

Round, F. (1961). The diatoms of a core Esthwaite Water. *New Phytologist*, *60*(1), 43–59.

Rousso, B. Z., Bertone, E., Stewart, R., & Hamilton, D. P. (2020). A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*, *182*, 115959. https://doi.org/10.1016/j.watres.2020.115959

Rozina, T., Eleršek, T., Justin, M. Z., & Meglič, A. (2018). Combined use of chlorophyll a and phycocyanin fluorescence sensors for quantification and differentiation of phytoplankton: A useful approach for small surface water bodies. *Acta Biologica Slovenica*, *61*(2), 29–37.

Sadegh, A. S., Sidoumou, Z., Dia, M., Pinchetti, J. L. G., & Bouaïcha, N. (2021). Impacts of phosphorus loads on the water quality and the proliferation of harmful cyanobacteria in Foum-Gleita Reservoir (Mauritania). *Annales de Limnologie - International Journal of Limnology*, *57*, 1. https://doi.org/10.1051/limn/2020029

Saqrane, S., & Oudra, B. (2009). CyanoHAB occurrence and water irrigation cyanotoxin contamination: Ecological impacts and potential health risks. *Toxins*, *1*(2), 113–122.

Sarada, R., Pillai, M. G., & Ravishankar, G. A. (1999). Phycocyanin from *Spirulina* sp.: Influence of processing of biomass on phycocyanin yield, analysis of

efficacy of extraction methods and stability studies on phycocyanin. *Process Biochemistry*, *34*(8), 795–801.

Sayers, M., Fahnenstiel, G. L., R.A., S., & Whitley, M. (2016). Cyanobacteria blooms in three eutrophic basins of the Great Lakes: A comparative analysis using satellite remote sensing. *International Journal of Remote Sensing*, *37*, 4148–4171.

Schalles, J. F. (2006). Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and CDOM concentrations. *Remote Sensing and Digital Image Processing*, *9*, 27–79.

Scheffer, M., Rinaldi, S., Gragnani, A., Mur, L. R., & Nes, E. (1997). On the dominance of filamentous cyanobacteria in shallow, turbid lakes. *Ecology*, *78*(1), 272–282.

Schoeman, F. R. (1979). Diatoms as indicators of water quality in the upper Hennops River Transvaal, South Africa. *Journal of the Limnological Society of Southern Africa*, *5*(2), 73–78.

Schönfelder, I., Gelbrecht, J., Schönfelder, J., & Steinberg, C. E. W. (2002). Relationships between littoral diatoms and their chemical environment in northeastern German lakes and rivers. *Journal of Phycology*, *38*(1), 66–89. https://doi.org/10.1046/j.1529-8817.2002.01056.x

Schrader, K. K., Duke, S. O., Kingsbury, S. K., Tucker, C. S., Duke, M. V., Dionigi, C. P., Millie, D. F., & Zimba, P. V. (2000). Evaluation of ferulic acid for controlling the musty-odor cyanobacterium, *Oscillatoria perornata*, in

aquaculture ponds. *Journal of Applied Aquaculture*, *10*(1), 1–16. https://doi.org/10.1300/J028v10n01_01

Schrader, K., & Kingsbury, K. (2000). Evaluation of limnocoral for studying the effects of phytotoxic compounds on plankton and water chemistry in aquaculture ponds. *Journal of the World Aquaculture Society*, *31*(3), 403–415.

Sengpiel, E. (2007). *Using airborne hyperspectral imagery to estimate chlorophyll a and phycocyanin in three central Indiana mesotrophic to eutrophic reservoirs*. Indiana University – Purdue University.

Shanmugam, P., Varunan, T., Nagendra Jaiganesh, S. N., Sahay, A., & Chauhan, P. (2016). Optical assessment of colored dissolved organic matter and its related parameters in dynamic coastal water systems. *Estuarine, Coastal and Shelf Science*, *175*, 126–145.

Shi, X., Liu, X., Liu, G., Sun, Z., & Xu, H. (2012). Application of phytoplankton communities for monitoring water quality in the Hangzhou section of Jing-Hang Grand Canal, southern China. *Fundamental and Applied Limnology*, 1–11. https://doi.org/10.1127/1863-9135/2012/0238

Shi, X., Sun, Z., Liu, G., & Xu, H. (2015). Insights into community-based discrimination of water quality status using an annual pool of phytoplankton in mid-subtropical canal systems. *Environmental Science and Pollution Research*, *22*(2), 1199–1206. https://doi.org/10.1007/s11356-014-3391-9

Shimoda, Y., & Arhonditsis, G. B. (2016). Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of

knowledge. *Ecological Modelling*, *320*, 29–43. https://doi.org/10.1016/j.ecolmodel.2015.08.029

Shin, J., Yoon, S., & Cha, Y. (2017). Prediction of cyanobacteria blooms in the lower Han River (South Korea) using ensemble learning algorithms. *DESALINATION AND WATER TREATMENT*, *84*, 31–39. https://doi.org/10.5004/dwt.2017.20986

Silveira, S. T., Burkert, J. F. M., Costa, J. A. V., Burkert, C. A. V., & Kalil, S. J. (2007). Optimization of phycocyanin extraction from *Spirulina platensis* using factorial design. *Bioresource Technology*, *98*(8), 1629–1634.

Simis, S. G. H., Peters, S. W. M., & Gons, H. J. (2005). Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland waters. *Limnology and Oceanography*, *50*(1), 237–245.

Singh, U. B., Ahluwalia, A. S., Sharma, C., Jindal, R., & Thakur, R. K. (2013). Planktonic indicators: A promising tool for monitoring water quality (early-warning signals). *Ecology, Environment and Conservation*, *19*(3), 793–800.

Smith, J., Connell, P., Evans, R. H., Gellene, A. G., Howard, M. D. A., Jones, B. H., Kaveggia, S., Palmer, L., Schnetzer, A., Seegers, B. N., Seubert, E. L., Tatters, A. O., & Caron, D. A. (2018). A decade and a half of *Pseudo-nitzschia* spp. And domoic acid along the coast of southern California. *Harmful Algae*, *79*, 87–104. https://doi.org/10.1016/j.hal.2018.07.007

Smith, J. E., Stocker, M. D., Wolny, J. L., Hill, R. L., & Pachepsky, Y. A. (2020). Intraseasonal variation of phycocyanin concentrations and environmental covariates in two agricultural irrigation ponds in Maryland, USA.

Environmental Monitoring and Assessment, *192*(11), 706. https://doi.org/10.1007/s10661-020-08664-w

Smith, J. E., Wolny, J. L., Stocker, M. D., Hill, R. L., & Pachepsky, Yakov A. (2021). Temporal stability of phytoplankton functional groups within two agricultural irrigation ponds in Maryland, USA. *Frontiers in Water*, *3*, 14. https://doi.org/10.3389/frwa.2021.724025

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research*, *10*(2), 126–139. https://doi.org/10.1065/espr2002.12.142

Smith, V. H., Willén, E., & Karlsson, B. (1987). Predicting the summer peak biomass of four species of blue-green algae (Cyanophyta/Cyanobacteria) in Swedish lakes. *JAWRA Journal of the American Water Resources Association*, *23*(3), 397–402. https://doi.org/10.1111/j.1752-1688.1987.tb00818.x

Soininen, J., Kokocinski, M., Estlander, S., Kotanen, J., & Heino, J. (2007). Neutrality, niches, and determinants of plankton metacommunity structure across boreal wetland ponds. *Écoscience*, *14*(2), 146–154. https://doi.org/10.2980/1195-6860(2007)14[146:NNADOP]2.0.CO;2

Soltani, N., Khodaei, K., Alnajar, N., Shahsavari, A., & Ashja Ardalan, A. (2012). Cyanobacterial community patterns as water quality bioindicators. *Iranian Journal of Fisheries Sciences*, *11*(4), 876–891.

Song, J., Hou, C., Liu, Q., Wu, X., Wang, Y., & Yi, Y. (2020). Spatial and temporal variations in the plankton community because of water and sediment

regulation in the lower reaches of Yellow River. *Journal of Cleaner Production*, *261*, 120972. https://doi.org/10.1016/j.jclepro.2020.120972

Song, K., Li, L., Tedesco, L., Clercin, N., Hall, B., Li, S., Shi, K., Liu, D., & Sun, Y. (2013). Remote estimation of phycocyanin (PC) for inland waters coupled with YSI PC fluorescence probe. *Environmental Science and Pollution Research*, *20*(8), 5330–5340. https://doi.org/10.1007/s11356-013-1527-y

Song, L., Marsh, T. L., Voice, T. C., & Long, D. T. (2011). Loss of seasonal variability in a lake resulting from copper sulfate algaecide treatment. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(9), 430–435. https://doi.org/10.1016/j.pce.2010.04.016

Srichandan, S., Kim, J. Y., Bhadury, P., Barik, S. K., Muduli, P. R., Samal, R. N., Pattnaik, A. K., & Rastogi, G. (2015). Spatiotemporal distribution and composition of phytoplankton assemblages in a coastal tropical lagoon: Chilika, India. *Environmental Monitoring and Assessment*, *187*(2), 47. https://doi.org/10.1007/s10661-014-4212-9

Srisuksomwong, P., & Pekkoh, J. (2020). Artificial Neural Network Model to Prediction of Eutrophication and Microcystis Aeruginosa Bloom. *Emerging Science Journal*, *4*(2), 129–135. https://doi.org/10.28991/esj-2020-01217

Stauffer, B. A., Bowers, H. A., Buckley, E., Davis, T. W., Johengen, T. H., Kudela, R., McManus, M. A., Purcell, H., Smith, G. J., Vander Woude, A., & Tamburri, M. N. (2019). Considerations in harmful algal bloom research and monitoring: Perspectives from a consensus-building workshop and technology

testing. *Frontiers in Marine Science*, *6*, 399. https://doi.org/10.3389/fmars.2019.00399

Steidinger, K. A. (2009). Historical perspective on *Karenia brevis* red tide research in the Gulf of Mexico. *Harmful Algae*, *8*(4), 549–561. https://doi.org/10.1016/j.hal.2008.11.009

Steidinger, K. A., Landsberg, J. H., Flewelling, J. L., Kirkpatrick, B., Walsh, P. J., Smith, S. L., Fleming, L. E., Solo-Gabriele, H. M., & Gerwick, W. H. (2008). Toxic dinoflagellates. In *Oceans and Human Health: Risks and Remedies from the Seas*. Elsevier Science Publishers.

Steinberg, C. E. W., & Hartmann, H. M. (1988). Planktonic bloom-forming cyanobacteria and the eutrophication of lakes and rivers. *Freshwater Biology*, *20*(2), 279–287.

Steinberg, D. K., Nelson, N. B., Carlson, C. A., & Prusak, A. C. (2004). Production of chromophoric dissolved organic matter (CDOM) in the open ocean by zooplankton and the colonial cyanobacterium *Trichodesmium* spp. *Marine Ecology Progress Series*, *267*, 45–56.

Stocker, M. D., Jeon, D. J., Sokolova, E., Lee, H., Kim, M. S., & Pachepsky, Y. A. (2020). Accounting for the three-dimensional distribution of *Escherichia coli* concentrations in pond water in simulations of the microbial quality of water withdrawn for irrigation. *Water*, *12*(6), 1708. https://doi.org/10.3390/w12061708

Stocker, M. D., Pachepsky, Y. A., Hill, R. L., Sellner, K. G., Macarisin, D., & Staver, K. W. (2019). Intraseasonal variation of *E. coli* and environmental covariates

in two irrigation ponds in Maryland, USA. *Science of the Total Environment*, *670*, 732–740.

Stocker, M. D., Penrose, M., & Pachepsky, Y. A. (2018). Spatial patterns of *Escherichia coli* concentrations in sediment before and after high-flow events in a first-order creek. *Journal of Environmental Quality*, *47*(5), 958–966. https://doi.org/10.2134/jeq2017.11.0451

Stumpf, R. P., Johnson, L. T., Wynne, T. T., & Baker, D. B. (2016). Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. *Journal of Great Lakes Research*, *42*(6), 1174–1183. https://doi.org/10.1016/j.jglr.2016.08.006

Su, X., Steinman, A. D., Xue, Q., Zhao, Y., Tang, X., & Xie, L. (2017). Temporal patterns of phyto- and bacterioplankton and their relationships with environmental factors in Lake Taihu, China. *Chemosphere*, *184*, 299–308. https://doi.org/10.1016/j.chemosphere.2017.06.003

Sun, X., Wu, N., Hörmann, G., Faber, C., Messyasz, B., Qu, Y., & Fohrer, N. (2022). Using integrated models to analyze and predict the variance of diatom community composition in an agricultural area. *Science of The Total Environment*, *803*, 149894. https://doi.org/10.1016/j.scitotenv.2021.149894

Svirčev, Z., Drobac, D., Tokodi, N., Mijovic, B., Codd, G., & Meriluoto, J. (2017). Toxicology of microcystins with reference to cases of human intoxications and epidemiological investigations of exposures to cyanobacteria and cyanotoxins. *Archives of Toxicology*, *91*, 621–650.

Taboada-Castro, M. M., Diéguez-Villar, A., & Taboada-Castro, M. T. (2004). Transfer of Nutrients and Major Ions of an Agricultural Catchment to Runoff Waters: Analysis of Their Spatial Distribution. *Conserving Soil and Water for Society: Sharing Solutions*, 1–4.

Tan, X., Kong, F., Yu, Y., & Zhang, M. (2009). Spatio-temporal variations of phytoplankton community composition assayed by morphological observation and photosynthetic pigment analyses in Lake Taihu (China). *African Journal of Biotechnology*, *8*(19), 4160–4168. https://doi.org/10.4314/ajb.v8i19.65202

Tango, P. J., & Butler, W. (2008). Cyanotoxins in tidal waters of Chesapeake Bay. *Northeastern Naturalist*, *15*(3), 403–416.

Temponeras, M., Kristiansen, J., & Moustaka-Gouni, M. (2000). Seasonal variation in phytoplankton composition and physical-chemical features of the shallow Lake Doïrani, Macedonia, Greece. In C. S. Reynolds, M. Dokulil, & J. Padisák (Eds.), *The Trophic Spectrum Revisited* (pp. 109–122). Springer Netherlands. https://doi.org/10.1007/978-94-017-3488-2_10

Teta, R., Romano, V., Della Sala, G., Picchio, S., De Sterlich, C., Mangoni, A., Di Tullio, G., Costantino, V., & Lega, M. (2017). Cyanobacteria as indicators of water quality in Campania coasts, Italy: A monitoring strategy combining remote/proximal sensing and in situ data. *Environmental Research Letters*, *12*(2), 024001.

Thakur, R. K., Jindal, R., Singh, U. B., & Ahluwalia, A. S. (2013). Plankton diversity and water quality assessment of three freshwater lakes of Mandi (Himachal Pradesh, India) with special reference to planktonic indicators. *Environmental*

Monitoring and Assessment, *185*(10), 8355–8373. https://doi.org/10.1007/s10661-013-3178-3

Thomas, M. K., Fontana, S., Reyes, M., Kehoe, M., & Pomati, F. (2017). The predictability of a lake phytoplankton community, from hours to years. *Ecology Letters*, *21*, 619–628. https://doi.org/10.1101/230722

Thomson-Laing, G., Puddick, J., & Wood, S. A. (2020). Predicting cyanobacterial biovolumes from phycocyanin fluorescence using a handheld fluorometer in the field. *Harmful Algae*, *97*, 101869. https://doi.org/10.1016/j.hal.2020.101869

Tian, C., Lu, X., Pei, H., Hu, W., & Xie, J. (2013). Seasonal dynamics of phytoplankton and its relationship with the environmental factors in Dongping Lake, China. *Environmental Monitoring and Assessment*, *185*(3), 2627–2645. https://doi.org/10.1007/s10661-012-2736-4

Tian, C., Pei, H., Hu, W., & Xie, J. (2013). Phytoplankton variation and its relationship with the environmental factors in Nansi Lake, China. *Environmental Monitoring and Assessment*, *185*(1), 295–310. https://doi.org/10.1007/s10661-012-2554-8

Touchart, L., & Bouny, J. (2008). Phytoplankton geographic spatialization in two ponds in Limousin (France). *Applied Geography*, *28*(4), 295–310. https://doi.org/10.1016/j.apgeog.2008.02.002

Tourville-Poirier, A. M., Cattaneo, A., & Hudon, C. (2010). Benthic cyanobacteria and filamentous chlorophytes affect macroinvertebrate assemblages in a large

fluvial lake. *Journal of the North American Benthological Society*, *29*, 737–749.

United Nations, Department of Economic and Social Affairs, & Population Division. (2019). *World population prospects Highlights, 2019 revision Highlights, 2019 revision*.

US Environmental Protection Agency. (2015). *Drinking Water Health Advisories for Two Cyanobacterial Toxins* (Health Advisory No. 820F15003; p. 3). EPA.

USDA. (2018). *2018 Irrigation and Water Management Survey* (Census of Agriculture AC17SS1). USDA. https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Farm_and_Ranch_Irrigation_Survey/index.php

Uyttendaele, M., Jaykus, L.-A., Amoah, P., Chiodini, A., Cunliffe, D., Jacxsens, L., Holvoet, K., Korsten, L., Lau, M., McClure, P., Medema, G., Sampers, I., & Jasti, P. R. (2015). Microbial hazards in irrigation water: Standards, norms, and testing to manage use of water in fresh produce primary production. *Comprehensive Reviews in Food Science and Food Safety*, *14*(4), 336–356. https://doi.org/10.1111/1541-4337.12133

Vähätalo, A. V., Aarnos, H., Hoikkala, L., & Lignell, R. (2011). Photochemical transformation of terrestrial dissolved organic matter supports hetero- and autotrophic production in coastal waters. *Marine Ecology Progress Series*, *423*, 1–14.

Vandermeulen, R. A., Mannino, A., Neeley, A., Werdell, J., & Arnone, R. (2017). Determining the optimal spectral sampling frequency and uncertainty

thresholds for hyperspectral remote sensing of ocean color. *Optics Express*, *25*(16), A785–A797. https://doi.org/10.1364/OE.25.00A785

Varol, M. (2019). Phytoplankton functional groups in a monomictic reservoir: Seasonal succession, ecological preferences, and relationships with environmental variables. *Environmental Science and Pollution Research*, *26*(20), 20439–20453. https://doi.org/10.1007/s11356-019-05354-0

Vasas, G., Farkas, O., Borics, G., Felföldi, T., Sramkó, G., Batta, G., Bacsi, I., & Gonda, S. (2013). Appearance of *Planktothrix rubescens* bloom with [D-Asp3, Mdha7]MC-RR in gravel pit pond of a shallow lake-dominated area. *Toxins*, *5*(12), 2434–2455. http://dx.doi.org/10.3390/toxins5122434

Virginia Department of Health. (2015). *Virginia Recreational Water Guidance for Microcystin and Microcystis Blooms Provisional Guidance*. https://townhall.virginia.gov/L/ViewGDoc.cfm?gdid=5479

Wang, B., Qiu, X.-L., Peng, X., & Wang, F. (2018). Phytoplankton community structure and succession in karst cascade reservoirs, SW China. *Inland Waters*, *8*(2), 229–238. https://doi.org/10.1080/20442041.2018.1443550

Wang, L., Wang, C., Deng, D., Zhao, X., & Zhou, Z. (2015). Temporal and spatial variations in phytoplankton: Correlations with environmental factors in Shengjin Lake, China. *Environmental Science and Pollution Research*, *22*(18), 14144–14156. https://doi.org/10.1007/s11356-015-4640-2

Wang, Q., & Wang, S. (2021). A predictive model of chlorophyll a in western Lake Erie based on artificial neural network. *Applied Sciences*, *11*(14), 6529. https://doi.org/10.3390/app11146529

Wang, X., Lu, Y., He, G., Han, J., & Wang, T. (2007). Exploration of relationships between phytoplankton biomass and related environmental variables using multivariate statistic analysis in a eutrophic shallow lake: A 5-year study. *Journal of Environmental Sciences*, *19*(8), 920–927. https://doi.org/10.1016/S1001-0742(07)60152-1

Watson, S. B., McCauley, E., & Downing, J. A. (1997). Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. *Limnology and Oceanography*, *42*(3), 487–495. https://doi.org/10.4319/lo.1997.42.3.0487

Weber, C. I. (1971). *A Guide to the Common Diatoms at Water Pollution Surveillance System Stations*. U.S. Environmental Protection Agency, National Environmental Research Center, Analytical Quality Control Laboratory.

Weralupitiya, C., Wanigatunge, R. P., Gunawardana, D., Vithanage, M., & Magana-Arachchi, D. (2022). Cyanotoxins uptake and accumulation in crops: Phytotoxicity and implications on human health. *Toxicon*, *211*, 21–35. https://doi.org/10.1016/j.toxicon.2022.03.003

Williams, E. J., & Kloot, N. (1953). Interpolation in a series of correlated observations. *Australian Journal of Applied Science*, *4*(1), 1–17.

Winter, J. M., Huang, H., Osterberg, E. O., & Mankin, J. S. (2020). Anthropogenic impacts on the exceptional precipitation of 2018 in the Mid-Atlantic United States. *Bulletin of the American Meteorological Society*, *101*(1), S5–S10.

Witte, W. G., Whitlock, C. H., Harriss, R. C., Usry, J. W., Poole, L. R., Houghton, W. M., Morris, W. D., & Gurganus, E. A. (1982). Influence of dissolved organic materials on turbid water optical properties and remote-sensing reflectance. *Journal of Geophysical Research: Oceans*, *87*(C1), 441–446. https://doi.org/10.1029/JC087iC01p00441

Wolanin, A., Soppa, M. A., & Bracher, A. (2016). Investigation of spectral band requirements for improving retrievals of phytoplankton functional types. *Remote Sensing*, *8*(10), 871. https://doi.org/10.3390/rs8100871

Wood, J. D., Franklin, R. B., Garman, G., McIninch, S., Porter, A. J., & Bukaveckas, P. A. (2014). Exposure to the cyanotoxin microcystin arising from interspecific differences in feeding habits among fish and shellfish in the James River Estuary, Virginia. *Environmental Science & Technology*, *48*(9), 5194–5202. https://doi.org/10.1021/es403491k

Wood, R. (2016). Acute animal and human poisonings from cyanotoxin exposure—A review of the literature. *Environment International*, *91*, 276–282. https://doi.org/10.1016/j.envint.2016.02.026

Woodhouse, J. N., Kinsela, A. S., Collins, R. N., Bowling, L. C., Honeyman, G. L., Holliday, J. K., & Neilan, B. A. (2016). Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. *The ISME Journal*, *10*(6), 1337–1351. https://doi.org/10.1038/ismej.2015.218

World Health Organization. (2008). *Microbiological Hazards in Fresh Leafy Vegetables and Herbs: Meeting Report* (Vol. 14). World Health Organization.

Wu, S., Du, Z., Wang, Y., Lin, T., Zhang, F., & Liu, R. (2020). Modeling spatially anisotropic nonstationary processes in coastal environments based on a directional geographically neural network weighted regression. *Science of The Total Environment*, *709*, 136097. https://doi.org/10.1016/j.scitotenv.2019.136097

Wu, Z., Cai, Y., Liu, X., Xu, C. P., Chen, Y., & Zhang, L. (2013). Temporal and spatial variability of phytoplankton in Lake Poyang: The largest freshwater lake in China. *Journal of Great Lakes Research*, *39*(3), 476–483. https://doi.org/10.1016/j.jglr.2013.06.008

Wu, Z., He, H., Cai, Y., Zhang, L., & Chen, Y. (2014). Spatial distribution of chlorophyll a and its relationship with the environment during summer in Lake Poyang: A Yangtze-connected lake. *Hydrobiologia*, *732*(1), 61–70. https://doi.org/10.1007/s10750-014-1844-2

Wynne, T. T., & Stumpf, R. P. (2015). Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western Lake Erie from 2002–2014. *Toxins*, *7*(5), 1649–1663. https://doi.org/10.3390/toxins7051649

Xi, H., Hieronymi, M., Krasemann, H., & Röttgers, R. (2017). Phytoplankton group identification using simulated and in situ hyperspectral remote sensing reflectance. *Frontiers in Marine Science*. http://dx.doi.org/10.3389/fmars.2017.00272

Xiao, L.-J., Zhu, Y., Yang, Y., Lin, Q., Han, B.-P., & Padisák, J. (2018). Species-based classification reveals spatial processes of phytoplankton meta-communities better than functional group approaches: A case study from three

freshwater lake regions in China. *Hydrobiologia*, *811*(1), 313–324. https://doi.org/10.1007/s10750-017-3502-y

Xiao, R., Su, S., Ghadouani, A., & Wu, J. (2013). Spatial analysis of phytoplankton patterns in relation to environmental factors across the southern Taihu basin, China. *Stochastic Environmental Research and Risk Assessment*, *27*(6), 1347–1357. https://doi.org/10.1007/s00477-012-0670-1

XiaoLi, Z., LiRong, S., & XiaoMing, Z. (2009). Effects of copper sulfate treatment on eutrophic urban lake phytoplankton communities. *Acta Hydrobiologica Sinica*, *33*(4), 596–602.

Xie, H., Bélanger, S., Song, G., Benner, R., Taalba, A., Blais, M., Tremblay, J.-É., & Babin, M. (2012). Photoproduction of ammonium in the southeastern Beaufort Sea and its biogeochemical implications. *Biogeosciences*, *9*, 3047–3061.

Yabunaka, K., Hosomi, M., & Murakami, A. (1997). Novel application of a back-propagation artificial neural network model formulated to predict algal bloom. *Water Science and Technology*, *36*(5), 89–97. https://doi.org/10.1016/S0273-1223(97)00464-2

Yajima, H., & Derot, J. (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), 206–220. https://doi.org/10.2166/hydro.2017.010

Yang, J., Holbach, A., Stewardson, M. J., Wilhelms, A., Qin, Y., Zheng, B., Zou, H., Qin, B., Zhu, G., Moldaenke, C., & Norra, S. (2021). Simulating chlorophyll-a fluorescence changing rate and phycocyanin fluorescence by using a multi-

sensor system in Lake Taihu, China. *Chemosphere*, *264*, 128482. https://doi.org/10.1016/j.chemosphere.2020.128482

Yang, M., Xia, J., Cai, W., Zhou, Z., Yang, L., Zhu, X., & Li, C. (2020). Seasonal and spatial distributions of morpho-functional phytoplankton groups and the role of environmental factors in a subtropical river-type reservoir. *Water Science and Technology*, *82*(11), 2316–2330. https://doi.org/10.2166/wst.2020.489

Yusoff, F. M., Zubaidah, M. S., Matias, H. B., & Kwan, T. S. (2002). Phytoplankton succession in intensive marine shrimp culture ponds treated with a commercial bacterial product. *Aquaculture Research*, *33*(4), 269–278.

Zamyadi, A., Choo, F., Newcomb, G., Stuetz, R., & Henderson, R. K. (2016). A review of monitoring technologies for real-time management of cyanobacteria: Recent advances and future direction. *Trends in Analytical Chemistry*, *85*.

Zanchett, G., & Oliveira-Filho, E. C. (2013). Cyanobacteria and cyanotoxins: From impacts on aquatic ecosystems and human health to anticarcinogenic effects. *Toxins*, *5*(10), 1896–1917.

Zeng, Q., Liu, Y., Zhao, H., Sun, M., & Li, X. (2017). Comparison of models for predicting the changes in phytoplankton community composition in the receiving water system of an inter-basin water transfer project. *Environmental Pollution*, *223*, 676–684. https://doi.org/10.1016/j.envpol.2017.02.001

Zhang, F., Harir, M., Moritz, F., Zhang, J., Witting, M., Wu, Y., Schmitt-Kopplin, P., Fekete, A., & Hertkorn, N. (2014). Molecular and structural characterization

of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. *Water Research*, *57*, 280–294.

Zhang, J., Zhi, M., & Zhang, Y. (2021). Combined generalized additive model and random forest to evaluate the influence of environmental factors on phytoplankton biomass in a large eutrophic lake. *Ecological Indicators*, *130*, 108082. https://doi.org/10.1016/j.ecolind.2021.108082

Zhang, Z., Gao, J., & Cai, Y. (2020). The direct and indirect effects of land use and water quality on phytoplankton communities in an agriculture-dominated basin. *Environmental Monitoring and Assessment*, *192*(12), 760. https://doi.org/10.1007/s10661-020-08728-x

Zhao, C. S., Shao, N. F., Yang, S. T., Ren, H., Ge, Y. R., Feng, P., Dong, B. E., & Zhao, Y. (2019). Predicting cyanobacteria bloom occurrence in lakes and reservoirs before blooms occur. *Science of The Total Environment*, *670*, 837–848. https://doi.org/10.1016/j.scitotenv.2019.03.161

Zheng, W., Li, F., Shi, H., Huo, Y., Li, Y., Chi, Y., Guo, Z., Wang, Y., Shen, C., Liu, J., & Qiao, M. (2016). Spatiotemporal heterogeneity of phytoplankton diversity and its relation to water environmental factors in the southern waters of Miaodao Archipelago, China. *Acta Oceanologica Sinica*, *35*(2), 46–55. https://doi.org/10.1007/s13131-016-0811-3

Zhou, S., Shao, Y., Gao, N., Deng, Y., Qiao, J., Ou, H., & Deng, J. (2013). Effects of different algaecides on the photosynthetic capacity, cell integrity and microcystin-LR release of *Microcystis aeruginosa*. *Science of The Total*

Environment, *463–464*, 111–119. https://doi.org/10.1016/j.scitotenv.2013.05.064

Zhou, Y., Zhang, Y., Liang, T., & Wang, L. (2019). Shifting of phytoplankton assemblages in a regulated Chinese river basin after streamflow and water quality changes. *Science of The Total Environment*, *654*, 948–959. https://doi.org/10.1016/j.scitotenv.2018.10.348

Zhu, R., Wang, H., Chen, J., Shen, H., & Deng, X. (2018). Use the predictive models to explore the key factors affecting phytoplankton succession in Lake Erhai, China. *Environmental Science and Pollution Research*, *25*(2), 1283–1293. https://doi.org/10.1007/s11356-017-0512-2

Zilliges, Y., Kehr, J., Meissner, S., Ishida, K., Mikkat, S., Hagemann, M., Kaplan, A. B., T., & Dittmann, E. (2011). The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of Microcystis under oxidative stress conditions. *PLoS ONE*, *6*, 17615.

Znachor, P., Nedoma, J., Hejzlar, J., Seďa, J., Komárková, J., Kolář, V., Mrkvička, T., & Boukal, D. S. (2020). Changing environmental conditions underpin long-term patterns of phytoplankton in a freshwater reservoir. *Science of The Total Environment*, *710*, 135626. https://doi.org/10.1016/j.scitotenv.2019.135626