Cyanobacterial Blooms in the Southeast: Understanding Their Distribution and Management

by

RajReni Beckett Kaul

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 6th, 2011

Keywords: phytoplankton models, cyanobacteria, *Microcystis aeruginosa*, microcystin

Approved by

Alan E. Wilson, Chair, Assistant Professor of Fisheries and Allied Aquacultures
Covadonga Arias, Associate Professor of Fisheries and Allied Aquacultures
Scott R. Santos, Associate Professor of Biological Sciences

Abstract

Water is vital for life. The increased demand for freshwater resources dictates that current water practices must ensure access to and availability of high-quality water for future generations. Phytoplankton community structure is indicative of, and can influence, water quality. In freshwater systems, bloom-forming cyanobacteria are the primary group of phytoplankton that dominate nutrient-rich (eutrophic), aquatic habitats. Cyanobacteria can cause noxious blooms and have the potential to produce toxic secondary metabolites. Microcystin, a hepatotoxin associated with many cyanobacterial species, has been linked to the deaths of livestock, fishes and humans.

Over 70% of the 89 sites sampled in Alabama during the 2008, 2009, and 2010 summers were classified as "eutrophic" based on chlorophyll concentration, and 90% of sites had detectable levels of the cyanotoxin, microcystin. Given the prevalence of cyanobacteria and their related toxin, microcystin predictive correlation and regression tree (CART) and multiple linear regression models were created for algal, cyanobacterial, and cyanotoxin abundances as a function of chlorophyll (µg/L), phycocyanin (µg/L), and microcystin (µg/L) respectively. The CART models created have the potential to become a powerful tool for both resource managers and citizen scientists.

In an effort to understand the processes favoring toxic cyanobacterial blooms, scientists often examine the intraspecific variation of blooms through the use of molecular markers. In the case presented, detection sensitivity for unique isolates was compared using the well-established

phycocyanin spacer and a newly described multilocus approach using housekeeping genes. In the population tested, the new approach was able to differentiate all isolates as unique strains.

Acknowledgments

Many people helped me along the winding road leading to this thesis. I am forever indebted to Orlando "Ace" Sarnelle for the introduction to the world of limnology. My continued collaboration with Ace led to the data for the molecular chapter. My formative experiences with Ace gave me the motivation and perseverance to purse this thesis. I would like to thank Alan Wilson for recruiting and supporting me for the past three years. Alan was also instrumental in helping me find my passion for teaching.

I would not have been able to create the predictive models without the help of numerous people. The large dataset was created through the collaboration of the multiple state agencies. I don't think we would have been as ambitious without the support and successful collaboration with Gina Curvin at ADEM. Additionally, sampling the state in the hot summers required passion and dedication by many people in the Wilson lab, including Michael Chislock and Kristin Durham, for which I am grateful. The transformation of the models from a simple regression to what is presented here was greatly influenced by Tomas Höök. Tomas' gentle pushes and constant questioning helped fill gaps in my knowledge and shaped the framework of the models.

My committee acted as a much needed sounding board. Scott Santos' clarity of thought and focus was much needed given the amorphous nature of the projects. Cova Arias forced me to frame my work not only within a limnological perspective but also within a microbiological perspective. This broader perspective has benefited my work and my scientific self.

Finally, I must acknowledge the unconditional support of my friends and family. My friends were there to help celebrate my successes and pick me up after setbacks. I am grateful and astounded by my mother's patients and willingness to listen to countless hours of rambling about phytoplankton and cyanobacteria. I know the journey would have been unbearable without these people in my life. Thank you.

Table of Contents

Abstract ii

Acknowledgments iv

List of Tables viii

List of Figures ix

Chapter One: Cyanobacteria and Their Distribution 1

Cyanobacterial Distribution 1

Harmful Algal Blooms 4

Microcystis aeruginosa 6

Microcystin 7

Cyanobacteria in Alabama Waterbodies 8

Objectives 10

References 12

Chapter Two: Predicting Phytoplankton, Cyanobacterial and Microcystin Concentrations throughout the Southeastern U.S. 18

Introduction 18

Methods 20

Results 24

Discussion 25

References 33

Chapter Three: A multilocus alternative to the phycocyanin operon reveals hidden genetic diversity within a *Microcystis* population. 49

Introduction 49

Methods 50

Results and Discussion 51

References 53

Appendix: Use of Secchi models by citizen scientists 56

List of Tables

Chapter One: Predicting Phytoplankton, Cyanobacterial, and Microcystin Concentrations throughout the Southeastern U.S.

Table 1 Geographic distribution of waterbodies and sites sampled by year 37

Table 2 Summary of data set used to build models 37

Table 3 Percent bias in models 38

Table 4 Example Secchi models 38

Table 5 Example full models 39

Table 6 Final Secchi models 39

Table 7 Final full models 40

Chapter Two: MLST

Table 1 Genetic diversity by method (PC-igs and MLST) and individual genes for MLST..... 55

List of Figures

Chapter One: Predicting Phytoplankton, Cyanobacterial, and Microcystin Concentrations throughout the Southeastern U.S.

Figure 1 Distribution of waterbodies sampled. 41

Figure 2 Trends between response and predictor variables 42

Figure 3 Secchi CART models for response variables 43

Figure 4 Full CART models for response variables 46

Chapter One: Cyanobacteria and Their Distribution

Cyanobacterial Distribution

The cyanobacterial phylum is defined by the evolution of photosynthesis (Knoll 2003). These gram-negative, prokaryotes belong to the negibacteria subkingdom and are significantly older than eukaryotes (Cavalier-Smith 2010). Modern chloroplasts found in eukaryotes developed from ancient cyanobacteria phagocytized by eukaryotes ~600 Myr ago (Cavalier-Smith 2010). Ancestral cyanobacteria played a crucial role in the development of the modern environment. Throughout the Proterozoic Era (2500-543 Ma), cyanobacteria were the dominate primary producers in the oceans playing a critical roles in the carbon and nitrogen cycles while oxygenating the atmosphere (Knoll 2003). The essential role of this early phylum has been maintained to present day and is reflected by their sheer abundance and diversity, in addition to chloroplasts' maintained functionality in higher plant species. Current research shows that cyanobacteria are still a crucial component in the ocean-atmosphere feedback cycles through their association with carbon and nitrogen fixation (Zehr 2011).

In addition to the open ocean, symbiotic cyanobacteria fix nitrogen in coral reef systems on the continental shelf (Lesser 2004). However, cyanobacteria are more often negatively associated with coral reef systems and coral black band disease (Frias-Lopez et al. 2003). Cyanobacteria have a global distribution and can be found in terrestrial, freshwater, and marine systems in addition to latitudinal and elevational extremes. For example, in high Arctic microbial mats and high Chilean saline wetlands, freshwater and terrestrial cyanobacterial species co-exist to dominate phototrophic communities (Dorador et al. 2008, Jungblut et al. 2009). The cyanobacteria phylum also includes thermophiles, forming mats in hot spring beds (Papke et al. 2003). In slow moving or standing freshwaters, some buoyancy-regulating

cyanobacterial genera (*Microcystis, Anabaena, Aphanizomenon,* etc.) form surface scums instead of benthic mats. Cyanobacterial surface scum formations typically occurs in standing or low-flow freshwater systems during warm times of the year when the water column stratifies based on temperature-mediated density differences.

In the early 1970s, the United States government attempted to remediate anthropogenic eutrophic ecosystems through the abatement of phosphorus via the Clean Water Act of 1972. This law sparked an ongoing debate about the importance of nitrogen (N) and phosphorus (P) in controlling algal abundance. Early evidence supported the sole importance of P, however these early studies relied on unreplicated observational data collected during the reduction of point source pollution in Lake Washington (Edmondson 1970) and whole lake experiments in the Canadian Shield area (Schindler 1978). Supporting the role of P in stimulating phytoplankton growth, a significant positive correlation between algal biomass via chlorophyll and P has been observed on a global scale (Schindler 1978). However, this correlation eventually reaches an asymptote suggesting another limiting factor, such as light or N (McCauley et al. 1989). Trends become less complex when narrowed to only include only cyanobacterial species. Since some species of cyanobacteria are capable of fixing atmospheric nitrogen, many argue and have demonstrated the sole importance of P (Downing et al. 2001, Watson et al. 1997). Nitrogen fixing species are thought to have a competitive advantage at low concentrations of N (Paerl 1988). When cyanobacterial biomass is divided into those capable of fixing nitrogen and non-nitrogen fixers, clearer patterns of dominance emerge across an N:P ratio. However it is unclear if the energetically costly act of nitrogen fixing is the mediating mechanism instead of a more efficient phosphorus uptake system (Jensen et al. 1994). Conflicting results from lab and field-

based studies sustain the two competing theories (P vs. N & P) and resulting debate (reviewed by Smith and Schindler 2009).

The expected increased prevalence and persistence of cyanobacterial blooms under global climate change is far less controversial. Increased atmospheric temperatures leading to fewer and more severe precipitation events will increase nutrient loading and retention times (Paerl and Huisman 2008). Lakes will also stratify sooner and longer with increased epilimnetic temperatures, creating an environment conducive for photosynthetic cyanobacteria to outcompete eukaryotic algal species by a variety of mechanisms (Paerl and Huisman 2009). First, earlier and longer stratification periods will increase phytoplankton death rates for algal species unable to regulate their buoyancy (Paerl and Huisman 2009). Increased sinking rates are compounded by the decreased viscosity of warmer water. Second, cyanobacteria are tolerant of and may require high temperatures for optimal growth. As temperatures increase > 25°C, eukaryotic algal growth may decline while cyanobacterial growth rates approach their maxima (Coles and Jones 2000, Robarts and Zohary 1987). Third, the number of man-made reservoirs is expected to increase to meet an increasing global demand for reliable freshwater supplies (Schindler 2009). The alteration of flowing systems to reservoirs will create more suitable habitat leading to cyanobacterial range expansion. Finally, in addition to the previously mentioned competitive advantages, some bloom-forming cyanobacterial species are more salt tolerant, if not halophilic, relative to other eukaryotic algal groups. This tolerance will be advantageous as expected salinity increases in reservoirs and coastal areas that are infiltrated by rising sea levels (Paerl and Huisman 2008).

Harmful Algal Blooms

The increased frequency of cyanobacterial blooms has meaningful social and economic impacts. Scum forming cyanobacterial species are included in a group of organisms including dinoflagellates and diatoms known to form harmful algal blooms (HABs). These blooms are harmful due to their sheer density and/or through the production of toxic compounds. In coastal waters, shellfish filter-feed on planktonic dinoflagellates and diatoms, concentrating the toxins that lead to a variety of shellfish poisoning, such as amnesic, diarrhetic and neutotoxic shellfish poisoning. Toxins produced by harmful dinoflagellate species forming HABs are the most diverse and include five unique classes of toxins (Backer and McGillicuddy 2006). For example, *Karenia brevis*, is a brevetoxin-producing dinoflagellate (Backer et al. 2005). Victims of brevetoxin poisoning have shown symptoms of neurotoxic shellfish poisoning after ingesting oysters, clams, and other filter feeders exposed to *K. brevis* blooms (Steidinger 1993). These symptoms include (but are not limited to) diarrhea, headache, bradycardia, temperature sensation reversal, and vertigo (Baden 1983, Hughes and Merson 1976, McFarren et al. 1965, Sakamoto et al. 1987). Marine HABs also affect the consumption risk associated with reef fish. The lipid soluble gambiertoxin produced by the dinoflagellate, *Gambierdiscus toxicus,* can be passed through multiple trophic levels in the coral reef food chain and bioaccumulate in large predatory fish (Lange 1987). The victim consumes the seemingly healthy fish before displaying symptoms of ciguatera fish poisoning, which is similar to neurotoxic shellfish poisoning (Backer and McGillicuddy 2006).

Freshwater HABs, specifically cyanoHABs, are of greater health concern since humans directly consume freshwater. CyanoHABs are thick surface scums often composed of *Anabaena, Aphanizomenon, Cylindrospermopsis, Nodularia, Planktothrix*, and/or *Microcystis*

genera. These blooms negatively impact potable water quality and food web dynamics through the production of a wide range of toxins that can be classified as either 1) neurotoxic, 2) hepatotoxic, or 3) cytotoxic (Aráoz et al. 2010). Anatoxin-a, a potent neurotoxin, produced by *Anabaena* spp. blocks cholinergic synapses and sodium channels (Spivak et al. 1980, Thomas et al. 1993). CyanoHABs producing anatoxin-a have contaminated drinking water for wildlife and domestic animals causing the deaths of Lesser flamingos (Krienitz et al. 2003), cows (Carmichael and Gorham 1978), and dogs (Edwards et al. 1992, Gugger et al. 2005). In addition to deaths, anatoxin-a produced by *Anabaena* spp. can be concentrated by shellfish creating a consumption risk similar to marine shellfish and paralytic shellfish poisoning (Negri and Jones 1995). Microcystin, a hepatotoxin, is not as acute as cyanobacterial neurotoxins. However, microcystin-contaminated medical water has caused human deaths (Carmichael et al. 2001). Sub-lethal chronic exposure of microcystin in contaminated water sources also negatively affects human health (Zhang et al. 2009). Like other cyanotoxins, microcystin can be concentrated in primary and secondary consumer tissues, typically the liver or hepatopancreas (Garcia et al. 2010).

CyanoHABs also negatively impact energy flow in food webs. Organisms relying on cyanoHABs as a food source show signs of stressed metabolisms. Zooplankton and rotifers fed diets of toxic cyanoHABs species have reduced feeding rates, fecundity and increased mortality (Demott et al. 1991, Gilbert 1990, 1994, Hietala et al. 1995, Reinikainen et al. 1994). A meta-analysis of the many feeding assays examining the possible combinations of cyanobacterial food source and grazer species indicated the *Microcystis* genus was the most detrimental to growth rates (Tillmanns et al. 2008). Interestingly, toxicity did not have a significant negative impact on population growth rates, implying reduced growth rates are a response to the poor nutritional

value of cyanobacteria (Tillmanns et al. 2008, Wilson et al. 2006). The reduced fecundity and population growth rate of planktonic grazers has major food web implications. Reductions in a trophic level's productivity can have cascading effects reducing the overall productivity of the ecosystem along with ecologically and economically important predators (Carpenter et al. 1985, Hoagland et al. 2002). The reduced productivity of a system due to depressed herbivore growth rates may only be temporary since some zooplankton have been able to locally adapt to cyanobacterial food source (Sarnelle and Wilson 2005).

Microcystis aeruginosa

Microcystis is a well-studied clonal freshwater cyanobacterial genus. The cocci cells, roughly 2-6 μm in diameter, form amorphous colonies protected by a polysaccharide mucilage (Yang et al. 2008). The complete genome of the strain NIES-843, isolated from Lake Kasumigaura, was recently sequenced. The 5Mbp circular genome has a GC content of 42% and contains around 6,000 putative genes (Kaneko et al. 2007). Insertion sequences and miniaturized inverted repeat transposable elements composed roughly 12% of the entire genome (Kaneko et al. 2007), which is indicative of the plasticity observed within the genome (ie. *mcy* operon, (Tooming-Klunderud et al. 2008).

Colony morphology within the genus is also highly plastic, which has led to the classification of multiple species; however, low nucleotide diversity in the 16S rDNA and 16S-23S rDNA internal transcribed spacer (ITS) in addition to greater than 70% DNA-DNA hybridization among the morphologically-defined species prompted their unification under *M. aeruginosa* (Kondo et al. 2000, Otsuka et al. 2001). *M. aeruginosa* is globally distributed and often described as cosmopolitan species. The lack of a global biogeographical structure displayed by the rDNA ITS reinforces the cosmopolitan description suggesting intercontinental

dispersal of *Microcystis* is not a rare occurrence given the distance (Van Gremberghe et al. 2011). Within Japan, intraspecific lineages have been identified based on the typing of seven housekeeping loci (Tanabe and Watanabe 2011). The Japanese populations suggest local adaptation forming ecotypes and possible endemic lineages (Tanabe and Watanabe 2011). Geographical isolation is normally not considered a possible evolutionary mechanism for free-living bacteria, but other endemic clades have been described for thermophilic cyanobacteria (Hongmei et al. 2005, Ionescu et al. 2010, Papke et al. 2003) However, *M. aeruginosa* populations outside of Japan need to be examined before accepting this unlikely evolutionary mechanism. The disparity of driving forces at the global and local scales demonstrates the need for further biogeographical research. Additionally, the studies used different genomic molecular approaches which have yet to be compared possibly adding another level of complexity to the questions examined.

Microcystin

The cyanotoxin, microcystin, named after the original source *Microcystis* spp., has been shown to be produced by a variety of cyanobacterial genera, including *Anabeana* spp. and *Oscillatoria* spp. The hydrophilic cyclic polypeptide is a non-ribosomally synthesized secondary metabolite with no clear cellular regulation or function. The non-ribosomal enzyme complex that assembles microcystin is encoded by the 55 kb *mcy* operon (Tillett et al. 2000). The operon contains 10 genes which are divergently transcribed (Tillett et al. 2000). The *mcy* gene cluster consists of polypeptide synthases, polyketide synthases, tailoring genes and an ABC like transport gene (Pearson et al. 2004). This complex synthesis leads to roughly 80 different microcystin variants. Since microcystin-LR (lucine and arginine) is the most common, toxicity is

often measured and discussed in microcystin-LR equivalents. For example, widely used commercial ELISA kits are optimized for the -LR variant.

The effects of the metabolite are very clear; it is a powerful protein phosphatase inhibitor. Microcystin covalently binds protein phosphatases 1 and 2A (Labine and Minuk 2009), leading to hyperphosphorylation in the cell signaling pathway. In vertebrates, microcystin is actively transported into hepatocytes ultimately leading to hepatic hemorrhage and death (Tillett et al. 2000). At chronic sub-lethal doses, microcystin is thought to act as a tumor promoter that can increase the risk of developing hepatocellular carcinoma (Yu 1995, reviewed by Labine and Minuk 2009). Exposure to microcystin can occur through dermal, respiratory, or ingestion routes. The World Health Organization has set a safe potable and recreational thresholds of microcystin at 1μg/L and <20μg/L, respectively.

Cyanobacteria in Alabama Waterbodies

Favorable habitats for cyanobacteria in Alabama include numerous man-made impoundments given the lack of glacial relic lakes throughout the state. Small impoundments used for the state's aquaculture industry are typically stagnant and nutrient rich, an ideal setting for thick cyanobacterial blooms. Maintaining intensive aquaculture systems is not trivial given the need to balance maximum harvest yield and water quality. Nutrient inputs via fish feed not only allows for higher production yields but also creates a hypereutrophic environment supporting massive algal blooms. These thick blooms can quickly lead to anoxic conditions through intense respiration during night as well as through microbial respiration associated with the degradation of decaying algal material. Anoxic conditions can lead to catastrophic harvest losses.

Cyanobacterial dominated blooms in aquaculture systems add another level of risk through the production of cyanotoxins. These toxins may have a two-fold negative impact on aquaculture systems. Cyanotoxins have the potential to negatively impact fish health or growth at various life stages (reviewed by Malbrouck and Kestemont 2006). Many unexplained catfish deaths could be explained by liver failure due to microcystin, as observed in a Mississippi aquaculture pond (Zimba et al. 2001). Microcystin also has the potential to enter human food supply through bioaccumulation in fish liver and tissue (Freitas de Magalhães et al. 2001). When tissue with higher toxin concentrations is exposed to the cooking process, the effects of bioaccumulation can be magnified. Temperature mediated degradation of proteins in the fish tissue leads to the release of phosphatase-bound microcystins (Zhang et al. 2010). However, the cooking process reduced the effects of bioaccumulation in raw tissue with low (> 157 ng/g microcystin) toxin concentrations (Berry et al. 2011). This would imply studies measuring microcystin concentrations in uncooked tissue may not accurately estimate human consumption risk (Ernst et al. 2005, Mohamed et al. 2003, Wilson et al. 2008).

In an attempt to mitigate the negative effects of dense algal blooms, copper sulfate is a commonly used algaecide. A crude understanding of the negative effect that copper treatments have on the biomass of algal communities exists, but little is known about copper exposure influences on the microbial populations and their structure. The few available studies suggest that species exhibit varying responses (Le Jeune et al. 2006) and an increase in copper resistance (Gustavson and Wangberg 1995) when exposed to elevated copper levels. These responses are indicative of a directional selection by a destabilizing positive feedback cycle. While extensively studied, the role of microcystin is still poorly understood (Babica et al. 2006). One known transcriptional regulator of microcystin production is the ferric uptake regulator (Fur) protein,

which also regulates part of the intracellular oxidative stress response (Martin-Luna et al. 2006). Co-regulation by Fur suggests that microcystin is also part of the oxidative stress response. If this is true, cellular oxidative stress caused by copper based algaecides (Knauert and Knauer 2008) could favor directional selection towards cyanobacterial genotypes with higher toxin quotas. This in would imply current management practices are creating less desirable systems.

Cyanobacteria will continue to significantly impact ecosystems globally. They maintain crucial positive biogeochemical feedback cycles, but also have the potential to negatively affect ecosystems with their high densities and toxins. As anthropogenic eutrophication worsens, cyanobacteria have the potential to expand and dominate naive waters with negative ecological, social and economic consequences.

Objectives

The southeastern United States is experiencing a population boom requiring substantial alterations to land usage and infrastructure. These changes increase the likelihood of 1) cyanobacterial blooms through anthropogenic eutrophication, and 2) human exposure to these blooms. Given this increased risk, the prevalence and intensity of algal blooms, cyanoHABs, and cyanotoxins in the southeast should be better understood. Data collected throughout the southeast, over a period of three growing seasons, will provide a snap shot of current conditions. This data will then be used to obtain the first objective (Chapter 2): predicting environmental conditions linked to degraded waters as a result of cyanoHABs and their toxins. Additionally, the ecological processes, including intraspecific competition and diversity, leading to toxigenic cyanoHABs are often examined using molecular markers. The accurate description of a population requires an understanding of the molecular methods' sensitivity. In this case, the

second objective (Chapter 3) will compare the commonly used phycocyanin intergenic spacer to a new multilocus approach using housekeeping genes for a single geographic population.

References

Aráoz, R., Molgó, J., and Marsac, N.T.d. 2010. Neurotoxic cyanobacterial toxins. Toxicon **56**(5): 813-828.

Babica, P., Blaha, L., and Marsalek, B. 2006. Exploring the natural role of microcystins - A review of effects on photoautotrophic organisms. Journal of Phycology **42**(1): 9-20.

Backer, L., and McGillicuddy, D. 2006. Harmful algal blooms. Oceanography **19**(2): 94.

Backer, L.C., Kirkpatrick, B., Fleming, L.E., Cheng, Y.S., Pierce, R., Bean, J.A., Clark, R., Johnson, D., Wanner, A., Tamer, R., Zhou, Y., and Baden, D.G. 2005. Occupational exposure to aerosolized brevetoxins during Florida red tide events: Effects on a healthy worker population. Environmental Health Perspectives **113**(5): 644-649.

Baden, D.G. 1983. Marine foodborne dinoflagellate toxins. International Review of Cytology **82**: 99-150.

Berry, J.P., Lee, E., Walton, K., Wilson, A.E., and Bernal-Brooks, F. 2011. Bioaccumulation of microcystins by fish associated with a persistent cyanobacterial bloom in Lago de Patzcuaro (Michoacan, Mexico). Environmental Toxicology and Chemistry **30**(7): 1621-1628.

Carmichael, W., Azevedo, S., An, J., Molica, R., Jochimsen, E., Lau, S., Rinehart, K., Shaw, G., and Eaglesham, G. 2001. Human fatalities from cyanobacteria: Chemical and biological evidence for cyanotoxins. Environmental Health Perspectives **109**(7): 663-668.

Carmichael, W., and Gorham, P. 1978. Anatoxins from clones of *Anabaena flos-aquae* isolated from lakes of western Canada. Mitteilungen Internationale Vereiningung fuer Theoretische und Angewandte Limnologie **21**: 285-295.

Carpenter, S., Kitchell, J., and Hodgson, J. 1985. Cascading trophic interactions and lake productivity. Bioscience **35**(10): 634-639.

Cavalier-Smith, T. 2010. Deep phylogeny, ancestral groups and the four ages of life. Philosophical Transactions of the Royal Society: Biological Sciences **365**(1537): 111-132.

Coles, J.F., and Jones, R.C. 2000. Effect of temperature on photosynthesis-light response and growth of four phytoplankton species isolated from a tidal freshwater river. Journal of Phycology **36**(1): 7-16.

Demott, W.R., Zhang, Q.X., and Carmichael, W.W. 1991. Effects of toxic cyanobacteria and purified toxins on the survival and feeding of a *Copepod* and 3 species of *Daphnia*. Limnology & Oceanography **36**(7): 1346-1357.

Dorador, C., Vila, I., Imhoff, J.F., and Witzel, K.-P. 2008. Cyanobacterial diversity in Salar de Huasco, a high altitude saline wetland in northern Chile: an example of geographical dispersion? FEMS Microbiology Ecology **64**(3): 419-432.

Downing, J., Watson, S., and McCauley, E. 2001. Predicting Cyanobacteria dominance in lakes. Canadian Journal of Fisheries and Aquatic Sciences **58**(10): 1905-1908.

Edmondson, W. 1970. Phosphorus, nitrogen, and algae in Lake Washington after diversion of sewage. Science **169**(3946): 690-692.

Edwards, C., Beattie, K.A., Scrimgeour, C.M., and Codd, G.A. 1992. Identification of anatoxin-a in benthic cyanobacteria (blue-green algae) and in associated dog poisoning at Loch Insh, Scotland. Toxicon **30**(10): 1165-1175.

Ernst, B., Dietz, L., Hoeger, S.J., and Dietrich, D.R. 2005. Recovery of MC-LR in fish liver tissue. Environmental Toxicology **20**(4): 449-458.

Freitas de Magalhães, V., Moraes Soares, R., and Azevedo, S.M.F.O. 2001. Microcystin contamination in fish from the Jacarepaguá Lagoon (Rio de Janeiro, Brazil): ecological implication and human health risk. Toxicon **39**(7): 1077-1085.

Frias-Lopez, J., Bonheyo, G.T., Jin, Q., and Fouke, B.W. 2003. Cyanobacteria associated with coral black band disease in Caribbean and Indo-Pacific reefs. Applied and Environmental Microbiology **69**(4): 2409.

Garcia, A.C., Bargu, S., Dash, P., Rabalais, N.N., Sutor, M., Morrison, W., and Walker, N.D. 2010. Evaluating the potential risk of microcystins to blue crab (*Callinectes sapidus*) fisheries and human health in a eutrophic estuary. Harmful Algae **9**(2): 134-143.

Gilbert, J.J. 1990. Differential effects of *Anabaena affini* on cladocerans and rotifers-Mechanisms and implications. Ecology **71**(5): 1727-1740.

Gilbert, J.J. 1994. Susceptibility of planktonic rotifers to a toxic strain of *Anabaena flosaquae*. Limnology & Oceanography **39**(6): 1286-1297.

Gugger, M., Lenoir, S., Berger, C., Ledreux, A., Druart, J.C., Humbert, J.F., Guette, C., and Bernard, C. 2005. First report in a river in France of the benthic cyanobacterium *Phormidium favosum* producing anatoxin-a associated with dog neurotoxicosis. Toxicon **45**(7): 919-928.

Gustavson, K., and Wangberg, S.A. 1995. Tolerance induction and succession in microalgae communities exposed to copper and atrazine. Aquatic Toxicology **32**(4): 283-302.

Hietala, J., Reinikainen, M., and Walls, M. 1995. Variation in life history responses of *Daphnia* to toxic *Microcystis aeruginosa*. Journal of Plankton Research **17**(12): 2307-2318.

Hoagland, P., Anderson, D., Kaoru, Y., and White, A. 2002. The economic effects of harmful algal blooms in the United States: estimates, assessment issues, and information needs. Estuaries **25**(4): 819-837.

Hongmei, J., Aitchison, J.C., Lacap, D.C., Peerapornpisal, Y., Sompong, U., and Pointing, S.B. 2005. Community phylogenetic analysis of moderately thermophilic cyanobacterial mats from China, the Philippines and Thailand. Extremophiles **9**(4): 325-332.

Hughes, J.M., and Merson, M.H. 1976. Current concepts- fish and shellfish poisoning. New England Journal of Medicine **295**(20): 1117-1120.

Ionescu, D., Hindiyeh, M., Malkawi, H., and Oren, A. 2010. Biogeography of thermophilic cyanobacteria: insights from the Zerka Ma'in hot springs (Jordan). FEMS Microbiology Ecology **72**(1): 103-113.

Jensen, J., Jeppesen, E., Olrik, K., and Kristensen, P. 1994. Impact of nutrients and physical factors on the shift from cyanobacterial to chlorophyte dominance in shallow Danish lakes. Canadian Journal of Fisheries and Aquatic Sciences **51**(8): 1692-1699.

Jungblut, A.D., Lovejoy, C., and Vincent, W.F. 2009. Global distribution of cyanobacterial ecotypes in the cold biosphere. ISME Journal **4**(2): 191-202.

Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., Kawashima, K., Kishida, Y., Ono, A., Shimizu, Y., Takahashi, C., Minami, C., Fujishiro, T., Kohara, M., Katoh, M., Nakazaki, N., Nakayama, S., Yamada, M., Tabatai, S., and Watanabe, M. 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Research **14**(6): 247-256.

Knauert, S., and Knauer, K. 2008. The role of reactive oxygen species in copper toxicity to two freshwater green algae. Journal of Phycology **44**(2): 311-319.

Knoll, A.H. 2003. The geological consequences of evolution. Geobiology **1**(1): 3-14.
Kondo, R., Yoshida, T., Yuki, Y., and Hiroishi, S. 2000. DNA-DNA reassociation among a bloom-forming cyanobacterial genus, *Microcystis*. International Journal of Systematic and Evolutionary Microbiology **50**: 767.

Krienitz, L., Ballot, A., Kotut, K., Wiegand, C., Putz, S., Metcalf, J.S., Codd, G.A., and Pflugmacher, S. 2003. Contribution of hot spring cyanobacteria to the mysterious deaths of Lesser Flamingos at Lake Bogoria, Kenya. FEMS Microbiology Ecology **43**(2): 141-148.

Labine, M., and Minuk, G. 2009. Cyanobacterial toxins and liver disease. Canadian Journal of Physiology and Pharmacology **87**(10): 773-788.

Lange, W.R. 1987. Ciguatera toxicity. American Family Physician **35**(4): 177-182.

Le Jeune, A.H., Charpin, M., Deluchat, V., and Briand, J.F. 2006. Effect of copper sulphate treatment on natural phytoplanktonic communities. Aquatic Toxicology **80**(3): 267-280.

Lesser, M.P. 2004. Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. Science **305**(5686): 997-1000.

Malbrouck, C., and Kestemont, P. 2006. Effects of microcystins on fish. Environmental Toxicology and Chemistry **25**(1): 72-86.

Martin-Luna, B., Sevilla, E., Hernandez, J.A., Bes, M.T., Fillat, M.F., and Peleato, M.L. 2006. Fur from *Microcystis aeruginosa* binds in vitro promoter regions of the microcystin biosynthesis gene cluster. Phytochemistry **67**(9): 876-881.

McCauley, E., Downing, J.A., and Watson, S. 1989. Sigmoid relationships between nutrients and chlorophyll among lakes. Canadian Journal of Fisheries & Aquatic Sciences **46**: 1171.

McFarren, E., Tanabe, H., Silva, F., Wilson, W., Campbell, J., and Lewis, K. 1965. The occurrence of a ciguatera-like poison in oysters, clams and *Gymnodinium* cultures. Toxicon **3**: 111-123.

Mohamed, Z.A., Carmichael, W.W., and Hussein, A.A. 2003. Estimation of microcystins in the freshwater fish *Oreochromis niloticus* in an Egyptian fish farm containing a *Microcystis* bloom. Environmental Toxicology **18**(2): 137-141.

Negri, A.P., and Jones, G.J. 1995. Bioaccumulation of paralytic shellfish poisining (PSP) toxins for the cyanobacterium *Anabaena circinalis* by the fresh-water mussel *Alathyria condola*. Toxicon **33**(5): 667-678.

Otsuka, S., Suda, S., Shibata, S., Oyaizu, H., Matsumoto, S., and Watanabe, M. 2001. A proposal for the unification of five species of the cyanobacterial genus *Microcystis Kutzing* ex Lemmermann 1907 under the rules of the Bacteriological Code. International Journal of Systematic and Evolutionary Microbiology **51**(3): 873.

Paerl, H. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnology & Oceanography **33**(4): 823-847.

Paerl, H.W., and Huisman, J. 2008. Climate - Blooms like it hot. Science **320**(5872): 57-58.

Paerl, H.W., and Huisman, J. 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. Environmental Microbiology Reports **1**(1): 27-37.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. 2003. Geographical isolation in hot spring cyanobacteria. Environmental Microbiology **5**(8): 650-659.

Pearson, L.A., Hisbergues, M., Borner, T., Dittmann, E., and Neilan, B.A. 2004. Inactivation of an ABC transporter gene, *mcyH*, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Applied and Environmental Microbiology **70**(11): 6370.

Reinikainen, M., Ketola, M., and Walls, M. 1994. Effects of the concentrations of toxic *Microcystis aeruginosa* and an alternative food on the survival of *Daphnia pulex*. Limnology & Oceanography **39**(2): 424-432.

Robarts, R.D., and Zohary, T. 1987. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. New Zealand Journal of Marine and Freshwater Research **21**(3): 391-399.

Sakamoto, Y., Lockey, R.F., and Krzanowski, J.J. 1987. Shellfish and fish poisoning related to the toxic dinoflagellates. Southern Medical Journal **80**(7): 866-874.

Sarnelle, O., and Wilson, A.E. 2005. Local adaptation of *Daphnia pulicaria* to toxic cyanobacteria. Limnology & Oceanography **50**(5): 1565-1570.

Schindler, D. 1978. Factors regulating phytoplankton production and standing crop in worlds freshwaters. Limnology & Oceanography **23**(3): 478-486.

Schindler, D. 2009. Lakes as sentinels and integrators for the effects of climate change on watersheds, airsheds, and landscapes. Limnology & Oceanography **54**(6): 2349-2358.

Smith, V.H., and Schindler, D.W. 2009. Eutrophication science: where do we go from here? Trends in Ecology & Evolution **24**(4): 201-207.

Spivak, C.E., Witkop, B., and Albuquerque, E.X. 1980. Anatoxin-a: A novel, potent agonist at the nicotinic receptor. Molecular Pharmacology **18**(3): 384-394.

Steidinger, K. 1993. Algal Toxins in Seafoods and Drinking Water. Academic Press, London.

Tanabe, Y., and Watanabe, M.M. 2011. Local Expansion of a Panmictic Lineage of Water Bloom-Forming Cyanobacterium *Microcystis aeruginosa*. PLoS ONE **6**(2): e17085.

Thomas, P., Stephens, M., Wilkie, G., Amar, M., Lunt, G.G., Whiting, P., Gallagher, T., Pereira, E., Alkondon, M., Albuquerque, E.X., and Wonnacott, S. 1993. (+)-Anatoxin-a is a potent agonist at neutonal nicotinic acetylcholine-receptors. Journal of Neurochemistry **60**(6): 2308-2311.

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chemistry & Biology **7**(10): 753-764.

Tillmanns, A.R., Wilson, A.E., Pick, F.R., and Sarnelle, O. 2008. Meta-analysis of cyanobacterial effects on zooplankton population growth rate: species-specific responses. Fundamentals and Applied Limnology **171**(4): 285-295.

Tooming-Klunderud, A., Mikalsen, B., Kristensen, T., and Jakobsen, K.S. 2008. The mosaic structure of the *mcyABC* operon in *Microcystis*. Microbiology **154**(7): 1886-1899.

Van Gremberghe, I., Leliaert, F., Mergeay, J., Vanormelingen, P., Van Der Gucht, K., Debeer, A.-E., Lacerot, G., De Meester, L., and Vyverman, W. 2011. Lack of Phylogeographic Structure in the Freshwater Cyanobacterium *Microcystis aeruginosa* Suggests Global Dispersal. PLoS ONE **6**(5): e19561.

Watson, S., McCauley, E., and Downing, J. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnology & Oceanography **42**(3): 487-495.

Wilson, A.E., Gossiaux, D.C., Hoeoek, T.O., Berry, J.P., Landrum, P.F., Dyble, J., and Guildford, S.J. 2008. Evaluation of the human health threat associated with the hepatotoxin microcystin in the muscle and liver tissues of yellow perch (*Perca flavescens*). Canadian Journal of Fisheries and Aquatic Sciences **65**(7): 1487-1497.

Wilson, A.E., Sarnelle, O., and Tillmanns, A.R. 2006. Effects of cyanobacterial toxicity and morphology on the population growth of freshwater zooplankton: Meta-analyses of laboratory experiments. Limnology & Oceanography **51**(4): 1915-1924.

Yu, S.Z. 1995. Primary prevention of hepatocellular carcinoma. Journal of Gastroenterology and Hepatology **10**(6): 674-682.

Zehr, J.P. 2011. Nitrogen fixation by marine cyanobacteria. Trends in Microbiology **19**(4): 162-173.

Zhang, D., Xie, P., and Chen, J. 2010. Effects of temperature on the stability of microcystins in muscle of fish and its consequences for food safety. Bulletin of Environmental Contamination and Toxicology **84**(2): 202-207.

Zhang, H., Zhang, J., and Zhu, Y. 2009. Identification of microcystins in waters used for daily life by people who live on Tai Lake during a serious cyanobacteria dominated bloom with risk analysis to human health. Environmental Toxicology **24**(1): 82-86.

Zimba, P., Khoo, L., Gaunt, P., Brittain, S., and Carmichael, W. 2001. Confirmation of catfish, *Ictalurus punctatus* (Rafinesque), mortality from *Microcystis* toxins. Journal of Fish Diseases **24**(1): 41-47.

Chapter Two: Predicting Phytoplankton, Cyanobacterial, and Microcystin Concentrations throughout the Southeastern U.S.

Introduction

Safe and abundant freshwater resources are vital for properly functioning aquatic ecosystems and necessary for human survival. Climate change, eutrophication, and land use changes continue to degrade our limited freshwater systems. Phytoplankton, the dominant primary producers in pelagic systems, can strongly mediate water quality. For example, toxic, bloom-forming cyanobacteria (blue-green algae) can produce toxic secondary metabolites, such as the hepatotoxin, microcystin, which can harm zooplankton, fishes, pets, livestock, and humans (Carmichael et al. 2001, Zimba et al. 2001). Moreover, changes in water quality can profoundly affect phytoplankton species composition and abundance. For example, excessive phosphorus loading can lead to increased algal standing stock (Schindle.D.W. 1974, Smith and Shapiro 1981) which may have beneficial bottom-up effects on ecologically and economically important consumers in the food web (Carpenter et al. 1985). However, excessive nutrient loading may push ecosystems towards undesirable situations. Dense algal blooms,for example, reduce water usability for domestic, recreational, and aquaculture purposes due to associated cyanobacterial odor and flavor compounds (Graham et al. 2011, Zimba and Grimm 2003). In addition, algal blooms supported by nutrient enrichment may create deadly anoxic environments increasing overnight oxygen demands via respiration or through the decomposition of decaying algal material by bacteria (Barica 1975). Finally, increased phosphorus loading may shift phytoplankton communities toward cyanobacterial dominance (Watson et al. 1997). Global warming may further exaggerate the negative consequences of eutrophication on aquatic communities (Smith and Schindler 2009).

Understanding the prevalence and persistence of algal blooms is critical for the effective management of ecologically and economically important freshwater ecosystems. Consequently, a large number of models have been produced to understand the variables responsible for blooms of freshwater phytoplankton (Heisler et al. 2008). A majority of these studies in North America have focused on waterbodies in the midwestern U.S. (Jones et al. 1998), Florida (Bachmann et al. 2003), the Canadian shield region (Ogbebo et al. 2009), as well as the Great Lakes region (Millie et al. 2006). Although findings from these studies vary, in general, phytoplankton abundance (typically measured as chlorophyll *a* concentration or algal biomass) has been shown to be related to ambient nutrient concentrations, such as total phosphorus (Brown et al. 2000, Jones et al. 1998) or total nitrogen (Paerl 1988). Cyanobacterial biomass measured in absolute or relative concentrations has also been shown to increase along a nutrient gradient (Downing et al. 2001, Watson et al. 1997). However, the debate continues whether nitrogen and/or phosphorus concentrations limit phytoplankton growth (Paerl 2009, Schindler et al. 2008). Finally, microcystin concentrations have been shown to correlate with Secchi depth and chlorophyll *a* (Bigham et al. 2009, Giani et al. 2005). And, at least one study (Graham et al. 2004) aimed to describe environmental thresholds that are associated with microcystin concentrations above the World Health Organization (WHO) threshold for microcystin in potable water (1 μg/L).

Despite the explosive human population growth and concomitant infrastructural development throughout the southeastern U.S., the region has largely been ignored regarding the development of predictive models for freshwater algal blooms, with the exception of Florida (Bigham et al. 2009, Canfield Jr 1983). Water quality models developed for Florida freshwater systems have shown a strong relationship between total phosphorus and chlorophyll as well as

between phytoplankton and cyanobacterial biomass (Brown et al. 2000, Canfield et al. 1989). In this study, we develop linear and non-linear predictive water quality models for the southeastern U.S. which are accessible to water resource managers as well as the general public. The three general water quality models aimed to predict abundances of freshwater phytoplankton (measured as chlorophyll *a* concentration), cyanobacteria (measured as the cyanobacteria-specific pigment, phycocyanin concentration), and toxic cyanobacteria (measured as microcystin concentration).

Methods

Data used to generate the models was collected by us or our collaborators at state agencies throughout the Southeast from June to September in 2008 (Alabama (AL)), 2009 (AL and Georgia (GA)), and 2010 (AL, GA, Florida (FL), Kentucky (KY), and Tennessee (TN) (Figure 1, Table 1). The 217 waterbodies surveyed varied in size, morphology, bathymetry, and usage and included small fishing impoundments, large river reservoirs, and natural lakes. Sampling locations varied across waterbodies, but we tended to sample the deepest location in the largest basin, near the confluence of tributaries, and close to recreational areas. Sampling frequencies varied across waterbodies (range = 1-51, mean = 3, median = 1) from 2008 to 2010. Some waterbodies were sampled only once while other sites were sampled multiple times across sampling years (Table 1).

Water Sampling and Analysis

Using a handheld meter, temperature and dissolved oxygen concentration profiles were used to calculate the maximum depth of the epilimnion prior to sampling. In general, depth-integrated water samples were collected with rigid or flexible tube samplers from the surface to a depth of ≈1.5 m. Some of our agency collaborators collected discrete samples at a depth

between 0.25 – 1.0 m. No surface-collected samples were considered in this study, given that depth-integrated epilimnetic water samples well describe the near surface mixed layer and that surface algal scums can over-estimate phytoplankton and toxin concentrations for a waterbody (Tillmanns et al. 2008, Sarnelle et al. 2010). Mixed-layer water samples were stored on ice and processed within 24 hours of collection. Water transparency was measured with a Secchi desk.

Our primary response variables, including chlorophyll (µg/L), phycocyanin (µg/L), and microcystin (µg/L) concentrations, were measured by Auburn University personnel from seston collected on filters (Type A/E, Pall Corp.). Chlorophyll *a* concentrations were determined using fluorometry following a 24 hr extraction in 90% aqueous ethanol (Wilson et al. 2008); model 7200-000, Turner Design Instruments, Sunnyvale, California). Phycocyanin samples were briefly (30 sec) ground in 50mM phosphate buffer (cat. #LC18560-2, Fisher Scientific) and allowed to extract for 3 hrs prior to purifying the extract with an inline filter (0.2 µm) and analyzing the filtrate using fluorescence (Sarada et al. 1999, as modified by Randolph 2007; model 7200-000, Turner Design Instruments, Sunnyvale, California). Microcystin concentrations were determined colorimetrically using enzyme-linked immunosorbent assay (ELISA; cat #20-0068, Beacon Analytical Systems, Inc., Maine; plate reader model ELx808, BioTek , Inc., Vermont) following two 1 hr extractions in acidic 75% aqueous methanol (Wilson et al. 2008).

Water samples were also analyzed for total phosphorus (TP; µg/L), total nitrogen (TN; µg/L), and total suspended solids (TSS (mg/L). Our state collaborators used standard methods for these analyses. For water samples collected by Auburn University personnel, TP and TN were measured using a single digestion colorimetric approach (Gross and Boyd 1998; model Lambda 25 UV/Vis Spectometer, Perkin Elmer, Waltham, Mass.). TSS was determined by

calculating the weight of the material collected on pre-combusted and tared glass fiber filters (GF/C) which were dried at 55°C for a minimum of 24 hours.

Modeling

For our biotic variables of interest (chlorophyll, phycocyanin, and microcystin), we developed simple Secchi depth-based models for the general public as well as more complex models incorporating more water quality parameters (i.e., Secchi, TP, TN, N:P (by atoms), and TSS) for natural resource managers. Chlorophyll *a* concentrations were included in the complex models used to estimate phycocyanin and microcystin concentrations; however the latter parameters were not used in the complex chlorophyll *a* models. Likewise, phycocyanin was included as a possible explanatory variable in the complex microcystin models.

The data set included 217 waterbodies sampled a total of 656 times (range = 1-51). Despite the range in sampling frequencies across the waterbodies surveyed in our study, the modeling data set included only one case from each waterbody therefore equally representing each waterbody. For waterbodies that were sampled more than once, the median chlorophyll value was used to choose the case to represent the waterbody in the model building data set. When two median values were possible (i.e., even number of samplings), the most complete case (ie. case with data for the most variables) was selected. If the two median cases had equal number of measured parameters, the case for the waterbody was chosen by a coin flip. The reduced modeling data set included 217 cases including each waterbody once. Given the range and variability of water quality parameters across our study sites, all data was normalized by log transformation prior to analyses. Using these data, two modeling approaches were explored, specifically 1) multiple linear regression and 2) classification and regression tree (CART). All

models were created using only complete cases with the R statistical software package (R Development Core Team 2009; version 2.11.1).

Multiple Linear Regression

The three response variables were modeled separately using a combination forward and backward stepwise linear regression using the stepAIC function of the MASS package (Venables and Ripley 2002). The most parsimonious model was chosen based on AIC values.

CART modeling

CART modeling is non-parametric and assumes little about the normality and relationships between the response and explanatory variables (Breiman et al. 1984). A tree is constructed to reduce the residual sum of squares of the response variable by sequential binary partitioning of data based upon the suite of possible explanatory variables (Breiman et al. 1984). The resulting nodes created by the partitions increase homogeneity of the response variable within the sub-groups. Leafs, the terminal nodes, are associated with a mean of the response variable for that particular subset of data determined by the branching path. The trees were grown and pruned using the rpart package in R (Therneau and Atkinson 2010) according to Faraway (2006). To ease use and interpretation, CART models are presented using back transformed linear values.

Model Bias

Model bias was examined by randomly partitioning the data into training and testing data sets with a 70/30 split, respectively. Models were built as described above with the training data set. The model was then challenged with the testing data set to obtain the predicted values. Model bias was estimated by regressing the observed testing data set against the predicted values (Pineiro et al. 2008). A model was considered biased if a slope of 1 and an intercept of 0 did not

fall within the 95% confidence interval of the respective variable. This process was repeated 100 times for each model. If significant bias (> 25%) was not detected, the entire data set was used to create the final model.

Results

Limnological data

Limnological conditions varied significantly across our study sites (Table 2). For example, total phosphorus and total nitrogen concentrations spanned two orders of magnitude from 2.6 µg/L to 871.0 µg/L and 43.7 µg/L to 4365.1 µg/L, respectively. This variability was also reflected in the three response variables. Chlorophyll *a* spanned three orders of magnitude (0.65 µg/L to 316.2 µg/L). Phycocyanin and microcystin-LR equivalents had the widest range covering 4 orders of magnitude (Table 2). The relationships between the three response variables and abiotic variables weakened with increasing specificity (Figure 2).

Model Accuracy

Testing all twelve models (i.e., Secchi or complex models for chlorophyll *a*, phycocyanin, or microcystin) for bias using 100 random 70/30 training/testing samplings showed to have no more than 20% bias (Table 3). Tables 4 and 5 provide one example of the models' ability to accurately predict the response variable of interest. The complex linear regression and CART models' significant predictors were equally complex and able to explain a similar amount of variation in the testing data set. The Secchi and complex chlorophyll models predicted chlorophyll *a* concentrations with high accuracy (Tables 4 and 5). The error surrounding the predicted responses increased in the phycocyanin and microcystin models (Tables 4 and 5). However, the increased error did not lead to significantly different observed and predicted values ($p > 0.05$, Tables 4 and 5). Moreover, the testing data sets' model performance was not spatially

autocorrelated (Moran's I, p-value > 0.05). These findings indicate that our modeling parameters are robust, thus allowing the collapse of training and testing data sets into a single complete data set for the final models presented below.

Final Models

The Secchi models explained less variation with increasing model specificity (Table 6, Figure 3). For example, fit for the Secchi linear and CART models estimating chlorophyll concentration were much higher (R^2 = 0.74 – 0.77) than for phycocyanin or microcystin concentrations (R^2 = 0.34 – 0.43; Table 6, Figure 3). CART models always explained more variation in the response variable than the linear regression models (Table 6). The complex models always showed better fit and predictive ability than the Secchi models (R^2, Table 6, Table 7). The rigorous models show the same patterns observed for the Secchi models regarding their decreasing explanatory power with model specificity. The complexity of the models or the number of significant predictor variables was similar between the complex linear regression and CART models. Unlike the Secchi-based models, the linear regression models are better able to fit the data for chlorophyll and phycocyanin when compared to the CART models. Regardless of the number of explanatory parameters, a greater portion of the variation within the microcystin models was explained by the non-parametric CART modeling approach, suggesting non-linearities in the microcystin data set.

Discussion

Using a broad collection of diverse freshwater systems throughout the southeastern U.S., we developed a suite of models aimed at predicting concentrations of chlorophyll *a*, phycocyanin and microcystin. In addition, we tested the models' accuracy when predicting new data, an important but often overlooked necessity when evaluating applicability to water resource

managers. The CART modeling approach has similar predictive powers as the linear regression models but does not require calculations and can be easily integrated into current management practices and interpreted by scientists and non-scientists alike. Consequently, we emphasize the CART models while only mentioning results generated from the traditional linear models as a way to connect past studies.

Chlorophyll

Our data set used to train, test and build the final models had an equivalent range described for average chlorophyll *a* concentrations in Florida (2-265 µg/L; Bachmann et al. 2003). Chlorophyll *a* (µg/L) within the data set was strongly correlated with Secchi depth (r = -0.86), total phosphorus (r = 0.84), and total suspended solids (r = 0.80). The correlation of chlorophyll *a* concentration with total nitrogen (r = 0.64) or N:P ratio (r = -0.34) were significant but showed weaker correlation with chlorophyll *a* than the other variables. Despite the strong correlation between chlorophyll and total phosphorus, the complex linear chlorophyll model did not indicate total phosphorus as a significant predictor. The CART model describes a less complex relationship between chlorophyll *a* and commonly associated limnological parameters including: total phosphorus, total nitrogen and Secchi depth. Given the CART topography, in some instances chlorophyll *a* can be estimated by total phosphorus alone (TP < 26 µg/L). Total phosphorus has long been established as the driver of chlorophyll *a* within and across lakes (Jones et al. 1998, Smith and Shapiro 1981, reviewed by Smith 2003). However, studies have shown the relationship between the two variables reaches an asymptote (McCauley et al. 1989, Vollenwein et al. 1974), indicating at high levels of total phosphorus other resource(s) become limiting, such as another nutrient or sunlight. Regarding the complex CART chlorophyll model, total nitrogen becomes the driving nutrient when total phosphorus is above 26 µg/L, which is

well below the described global total phosphorus asymptote of ~100 µg/L TP (Brown et al. 2000). A similar chlorophyll *a*– total phosphorus curvilinear response has been described for large flowing systems (Van Nieuwenhuyse 2007) and lentic lakes and ponds (McCauley et al. 1989). Consequently, the decline of total phosphorus influence on chlorophyll *a* at concentrations exceeding 26 µg/L (TP) could be due to confounding effects of including a mixture of waterbodies that include rivers, reservoirs, ponds, and lakes in the model data set. The differing relationships between chlorophyll and abiotic factors between waterbody types should be explored as more data becomes available for the SE region.

Few related studies attempted to test the quality and bias of their models (McCauley et al. 1989, Van Nieuwenhuyse 2007). Such testing is necessary to determine the predictive utility of the new models. The presented example testing models (Tables 4 and 5) allowed closer examination of the predictive performance using new data. The Secchi and complex chlorophyll models have similar predictive power ($r \sim 0.90$, EF ≥ 0.75). The leaf distributions of the complex and Secchi CART models are extremely similar (e.g., the same node is used to split the two largest means Secchi depth < 0.351m, Figure 3a, Figure 4a). This similarity in leaf distribution and the congruent predictive abilities between the Secchi and complete models suggest the Secchi models have a much greater return in accuracy given the resources required, thus providing managers and the general public with an accurate and easily accessible tool for managing chlorophyll concentrations. The related strong performance of the Secchi models ($r = 0.88$) when compared to previously tested models (CHL - TP; $r = 0.86$, Brown et al. 2000) reinforces the managerial appeal of the Secchi model.

Phycocyanin

Phycocyanin was shown to be significantly correlated with several environmental variables, including chlorophyll *a* ($r = 0.72$) and Secchi depth ($r = -0.59$). Similar results have been documented in Florida lakes where a strong correlation was observed between cyanobacterial biomass and phytoplankton biomass ($r = 0.90$; Canfield et al. 1989). In general, phycocyanin is less well studied than chlorophyll which might explain the slow adoption of phycocyanin as a metric for cyanobacterial abundance. However, phycocyanin is quick to measure relative to enumerating phytoplankton samples. As more labs measure phycocyanin and use this pigment to understand dynamics in freshwater systems, a better understanding of phycocyanin analyses and concentrations will be more readily available.

The Secchi models do not predict phycocyanin as well as the complex models ($0.52 \leq r \leq 0.69$; $0.14 \leq$ EF ≤ 0.54, Table 4, Table 5). The complex phycocyanin linear regression and CART models explain a similar amount of variation in the data set (linear R^2=0.59; CART R^2 = 0.56) while using chlorophyll *a* as the sole predictor. However, the linear model proved to be a better predictor when challenged with new data (linear $r = 0.69$; CART $r = 0.53$). The overall increased performance of the linear regression models regardless of the number of predictive parameters would imply a positive parametric relationship between chlorophyll *a* and phycocyanin concentrations (linear $r = 0.57, 0.69$; CART $r = 0.52, 0.53$; Secchi and complex model respectively). This same trend has been documented previously for algal and cyanobacterial biomass (Canfield et al. 1989, Watson et al. 1997).

Since the CART Secchi and complex models have similar accuracy, it may be advantageous for water resource managers to use the simpler Secchi model. Additionally, the largest mean leaf in both CART models fall close to the cyanobacterial biovolume medium alert

level threshold of 0.2 mm^3/L (House et al. 2004) which was calculated from the previously described relationship (cyanobacterial biovolume (mm^3/L) = 0.0034 * Phycocyanin (µg/L) + 0.126, r = 0.88, derived from Fig 7a, Randolph 2007). The range of phycocyanin in the data set limits the models' ability to distinguish between medium and high risk (cyanobacterial biovolume ≥ 0.4 mm^3/L), but the models may still be useful for "risk vs. no risk" management. The phycocyanin to cyanobacterial biovolume regression is a promising relationship for managers to exploit and was recently used to describe a managerial protocol using *in-vivo* phycocyanin concentrations for a Polish drinking water source (Izydorczyk et al. 2009). Effective use of the presented models with the biovolume risk alert system relies on a manager's ability to identify dominate cyanobacterial species as toxigenic. However, once the dominate species are identified as belonging to a "toxic" or "non-toxic" genera, management becomes an issue of controlling eutrophication to reduce algal abundance.

Microcystin

The range of microcystin (0.0002 µg/L to 5.25 µg/L) which was lower than observed in a past survey of Florida lakes (<0.1 ug/L to 12 ug/L; Bigham et al. 2009) may be a result of sampling and analyses differences. While the range of microcystin was lower in our southeastern data set, the significant but weak correlations between microcystin-total phosphorus (r = 0.57) and microcystin – chlorophyll *a* (r = 0.66) remained (Bigham et al. 2009). The observed variability in the microcystin data negatively impacted the models fit and accuracy.

The complete microcystin models, like the phycocyanin models, are better predictors of new data when compared to the simpler Secchi models. The complex microcystin CART model has an increased fit of the training data (CART R^2 = 0.63; linear R^2 = 0.42). However, this increased fit does not translate to better performance when given testing data (CART r = 0.59;

linear $r = 0.66$). As in the other complex models, there is not a meaningful difference in the linear regression and CART models predictive ability. Both final complex microcystin models require five of the possible seven predictor variables (Table 7). The Secchi CART model partitions microcystin concentrations into seven leafs ranging from 0.001 µg/L to 0.1 µg/L, while the complex model creates six terminal nodes ranging from 0.002 µg/L to 0.85 µg/L. Given the data set range, the highest mean terminal nodes of the two models include values above and below the W.H.O threshold of 1.0 µg/L microcystin, reducing the practical application of the model.

The root mean square error (RMSE) surrounding the regression line of observed and predicted values does not have a large practical implication for the chlorophyll and phycocyanin models. However, the acceptable error surrounding the microcystin models is much smaller given the health implications of one unit (1 µg/L) increase or decrease. The RMSE surrounding the microcystin models ranges from 6-8 ug/L microcystin. These large errors compared to the data set mean (0.01 µg/L) and W.H.O. safe potable water threshold (1 µg/L) makes these models very impractical in application.

The poor performance of the linear microcystin models should be expected given previous findings of non-linear relationship between environmental factors and microcystin (Giani et al. 2005, Graham et al. 2004). Additionally, microcystin concentrations have been significantly associated with chlorophyll and cyanobacterial biomass (Kotak et al. 2000, Wu et al. 2006) explaining the reoccurrence of chlorophyll *a* and phycocyanin as significant predictors. In the midwestern U.S. waterbodies with maximum microcystin concentrations were found to contain between 1,500 and 4,000 µg/L TN (Graham et al. 2004), which is similar to the node ($TN \geq 2113$ µg/L) leading to the leaf with the highest median (0.856 µg/L MC) of the complex

CART model. The model presented here follows previous findings; however this is the first set of microcystin models to be tested using an independent data set.

The poor predictive nature of the microcystin models indicates further parameters need to be explored to better describe the indirect link between environmental conditions and microcystin expression. Methods for measuring microcystin-LR equivalents have been described and robust commercial kits are available leading to a plethora of microcystin data. Despite easily available data, few region specific models (China: Wu et al. 2006; Canada: Giani et al. 2005, Kotak et al. 2000; USA: Bigham et al. 2009, Graham et al. 2004) have been developed for microcystin. This gap in scientific knowledge can easily be filled, and would be eagerly applied by water resource managers.

Final Models and Application

The models presented for the southeastern United States were built using data collected from natural waterbodies, small impoundments, rivers, and large reservoirs. The inclusion of a wide range of sources allows the greatest applicability for the end user. The lack of bias within the training models allowed the collapse of the data into a final building data set, while being able to comment on the predictive nature of the models. The models are less accurate with decreasing scope from chlorophyll to microcystin due to the decreased strength of the relationship between the environmental and response variables. The diversity of waterbody type is possibly a confounding factor leading to the phycocyanin and microcystin models' poor accuracy since most regional studies to date limit the data set to a single waterbody type. This is an area to explore as we are able to describe definitive waterbody types, and grow the data set to include a variety of sites types dispersed throughout each state in the region. All three response

variables can be described as a function of Secchi depth without sacrificing meaningful accuracy when compared to the complex models.

The CART models have the potential to become valuable tools for managers. First, the CART models' binary function is similar to commonly used decision trees allowing easier integration into current managerial plans. Secondly, the CART models, in some cases, can also be useful with an incomplete set of known predictor variables. Lastly, they give defined thresholds that lead to different outcomes. These reasons along with the flexibility of input variables by the creation of the full and Secchi models should make the CART models a well-received tool by managers.

References

Bachmann, R., Hoyer, M., and Canfield, D. 2003. Predicting the frequencies of high chlorophyll levels in Florida lakes from average chlorophyll or nutrient data. Lake and Reservoir Management **19**(3): 229-241.

Barica, J. 1975. Summerkill risk in prairie ponds and possibilities of its prediction. Journal of the Fisheries Research Board of Canada **32**(8): 1283-1288.

Bigham, D., Hoyer, M., and Canfield Jr, D. 2009. Survey of toxic algal (microcystin) distribution in Florida lakes. Lake and Reservoir Management **25**(3): 264-275.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and regression trees. Wadsworth and Brooks, Monterey, CA.

Brown, C.D., Hoyer, M.V., Bachmann, R.W., and Canfield Jr, D.E. 2000. Nutrient-chlorophyll relationships: an evaluation of empirical nutrient-chlorophyll models using Florida and north-temperate lake data. Canadian Journal of Fisheries and Aquatics Sciences **57**(8): 1574-1583.

Canfield, D., Phlips, E., and Duarte, C. 1989. Factors influencing the abundance of blue-green algae in Flordia lakes. Canadian Journal of Fisheries and Aquatic Sciences **46**(7): 1232-1237.

Canfield Jr, D. 1983. Prediction of chlorophyll a concentration in Flordia lakes: The importance of phosphorus and nitrogen. Journal of the American Water Resources Association **19**(2): 255-262.

Carmichael, W., Azevedo, S., An, J., Molica, R., Jochimsen, E., Lau, S., Rinehart, K., Shaw, G., and Eaglesham, G. 2001. Human fatalities from cyanobacteria: Chemical and biological evidence for cyanotoxins. Environmental Health Perspectives **109**(7): 663-668.

Carpenter, S., Kitchell, J., and Hodgson, J. 1985. Cascading tropic interaction and lake productivity. Bioscience **35**(10): 634-639.

Downing, J., Watson, S., and McCauley, E. 2001. Predicting Cyanobacteria dominance in lakes. Canadian Journal of Fisheries and Aquatic Sciences **58**(10): 1905-1908.

Faraway, J.J. 2006. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC, Boca Raton, FL.

Giani, A., Bird, D.F., Prairie, Y.T., and Lawrence, J.F. 2005. Empirical study of cyanobacterial toxicity along a trophic gradient of lakes. Canadian Journal of Fisheries and Aquatics Sciences **62**(9): 2100-2109.

Graham, J., Loftin, K., and Meyer, M. 2011. Cyanotoxin Mixtures and Taste-and-Odor Compounds in Cyanobacterial Blooms from the Midwestern United States. Environmental Science & Technology **44**(19): 7361-7368.

Graham, J.L., Jones, J.R., Jones, S.B., Downing, J.A., and Clevenger, T.E. 2004. Environmental factors influencing microcystin distribution and concentration in the Midwestern United States. Water Research **38**(20): 4395-4404.

Gross, A., and Boyd, C. 1998. A digestion procedure for the simultaneous determination of total nitrogen and total phosphorus in pond water. Journal of the World Aquaculture Society **29**(3): 300-303.

Heisler, J., Glibert, P., Burkholder, J., Anderson, D., Cochlan, W., Dennison, W., Dortch, Q., Gobler, C., Heil, C., and Humphries, E. 2008. Eutrophication and harmful algal blooms: A scientific consensus. Harmful Algae **8**(1): 3-13.

House, J., Ho, L., Newcombe, G., and Burch, M. 2004. Management strategies for toxic blue-green algae: Literature survey. A report for the CRC for Water Quality and Treatment Project 2.0.2.4.1.2: 61.

Izydorczyk, K., Carpentier, C., Mrowczynski, J., Wagenvoort, A., Jurczak, T., and Tarczynska, 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.

Jones, J.R., Knowlton, M.F., and Kaiser, M.S. 1998. Effects of aggregation on chlorophyll-phosphorus relations in Missouri reservoirs. Lake and Reservoir Management **14**(1): 1-9.

Kotak, B.G., Lam, A., Prepas, E., and Hrudey, S. 2000. Role of chemical and physical variables in regulating microcystin-LR concentration in phytoplankton of eutrophic lakes. Canadian Journal of Fisheries and Aquatic Sciences **57**(8): 1584-1593.

McCauley, E., Downing, J.A., and Watson, S. 1989. Sigmoid relationships between nutrients and chlorophyll among lakes. Canadian Journal of Fisheries and Aquatic Sciences **46**: 1171-1175.

Millie, D.F., Weckman, G.R., Pigg, R.J., Tester, P.A., Dyble, J., Litaker, R.W., Carrick, H.J., and Fahnenstiel, G.L. 2006. Modeling phytoplankton abundance in Saginaw Bay, Lake Huron: Using artificial neural networks to discern functional influence of environmental variables and relevance to a great lakes observing system. Journal of Phycology **42**(2): 336-349.

Ogbebo, F.E., Evans, M.S., Brua, R.B., and Keating, J.J. 2009. Limnological features and models of chlorophyll-a in 30 lakes located in the lower Mackenzie River basin, Northwest Territories (Canada). Journal of Limnology **68**(2): 336-351.

Paerl, H. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnology and Oceanography **33**(4): 823-847.

Paerl, H.W. 2009. Controlling eutrophication along the freshwater-marine continuum: Dual nutrient (N and P) reductions are essential. Estuaries and Coasts **32**(4): 593-601.

Pineiro, G., Perelman, S., Guerschman, J.P., and Paruelo, J.M. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling **216**(3-4): 316-322.

R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna, Austria.

Randolph, K. 2007. Remote sensing of cyanobacteria in case II waters using optically active pigments, chlorophyll a and phycocyanin, Department of Geography, Indian University.

Sarada, R., Pillai, M., and Ravishankar, G. 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.

Sarnelle, O., Morrison, J., Kaul, R., Horst, G., Wandell, H., and Bednarz, R. 2010. Citizen monitoring: Testing hypotheses about the interactive influences of eutrophication and mussel invasion on a cyanobacterial toxin in lakes. Water Research **44**(1): 141-150.

Schindle.D.W. 1974. Eutrophication and recovery in experimental lakes - implications for lake management. Science **184**(4139): 897-899.

Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., Beaty, K.G., Lyng, M., and Kasian, S.E.M. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. Proceedings of the National Academy of Sciences of the United States of America **105**(32): 11254-11258.

Smith, V.H. 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research **10**(2) 126-139.

Smith, V.H., and Schindler, D.W. 2009. Eutrophication science: where do we go from here? Trends in Ecology & Evolution **24**(4): 201-207.

Smith, V.H., and Shapiro, J. 1981. Chlorophyll-phosphorus relations in individual lakes: Their importance to lake restoration strategies. Environmental Science & Technology **15**(4): 444-451.

Therneau, T.M., and Atkinson, B. R port by Brian Ripley. 2010. rpart: Recursive Partitioning. R package version 3.1-46.http://CRAN.R-project.org/package=rpart

Tillmanns, A.R., Wilson, A.E., Pick, F.R., and Sarnelle, O. 2008. Meta-analysis of cyanobacterial effects on zooplankton population growth rate: species-specific responses. Fundamental and Applied Limnology **171**(4): 285-295.

Van Nieuwenhuyse, E.E. 2007. Response of summer chlorophyll concentration to reduce total phosphorus concentration in the Rhine River (Netherlands) and the Sacramento - San Joaquin Delta (California, USE). Canadian Journal of Fisheries and Aquatic Sciences **64**(11): 1529-1547.

Venables, W., and Ripley, B. 2002. Modern applied statistics with S.

Vollenwein, R.A., Munawar, M., and Stadelma, P. 1974. Comparative review of phytoplankton and primary production in the Laurentian Great Lakes. Journal of the Fisheries Research Board of Canada **31**(5): 739-762.

Watson, S., McCauley, E., and Downing, J. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnology & Oceanography **42**(3): 487-495.

Wilson, A.E., Gossiaux, D.C., Hoeoek, T.O., Berry, J.P., Landrum, P.F., Dyble, J., and Guildford, S.J. 2008. Evaluation of the human health threat associated with the hepatotoxin microcystin in the muscle and liver tissues of yellow perch (Perca flavescens). Canadian Journal of Fisheries and Aquatic Sciences **65**(7): 1487-1497.

Wu, S.K., Xie, P., Liang, G.D., Wang, S.B., and Liang, X.M. 2006. Relationships between microcystins and environmental parameters in 30 subtropical shallow lakes along the Yangtze River, China. Freshwater Biology **51**(12): 2309-2319.

Zimba, P., and Grimm, C. 2003. A synoptic survey of musty/muddy odor metabolites and microcystin toxin occurrence and concentration in southeastern USA channel catfish (*Ictalurus punctatus Ralfinesque*) production ponds. Aquaculture **218**(1-4): 81-87.

Zimba, P., Khoo, L., Gaunt, P., Brittain, S., and Carmichael, W. 2001. Confirmation of catfish, *Ictalurus punctatus* (*Rafinesque*), mortality from Microcystis toxins. Journal of Fish Diseases **24**(1): 41-47.

Table 1. Geographic distribution of waterbodies and sites sampled by year

	State					Resampled		
Year	AL	FL	GA	KY	TN	Waterbodies	Total Waterbodies	Total Sites
2008	7	-	-	-	-	4	7	55
2009	60	-	9	-	-	36	69	292
2010	11	104	6	9	11	24	141	308
					total	***64***	***217***	***655***

Table 2. Summary of data set used to build models

Environmental Parameter	Annotation	n	mean	median	min	max
chlorophyll (µg/L)	CHL	203	15.10	15.85	0.65	316.23
phycocyanin (µg/L)	PHY	128	4.95	4.07	0.03	707.95
microcystin (µg/L)	MC	176	0.01	0.01	0.0002	5.25
Secchi depth (m)	SEC	197	1.06	1.00	0.15	10.47
total phosphorus (µg/L)	TP	198	35.28	33.89	2.57	870.96
total nitrogen (µg/L)	TN	198	710.53	707.95	43.65	4365.16
total nitrogen: total phosphorus by moles	N:P	198	43.91	39.81	2.00	467.74
total suspended solids (mg/L)	TSS	193	5.25	6.03	0.50	74.13

Table 3. Percent bias in models. Biased models' observed verses predicted regression line 95% confidence interval does not include a slope of 1 and an intercept of 0.

	Secchi		Full	
Response	CART	Linear	CART	Linear
Chlorophyll	5	9	15	15
Phycocyanin	5	2	14	7
Microcystin	20	7	13	1

Table 4. Example Secchi models including significant predictor and model performance based on predicted values of the testing data set

Response Variable	Predictor Variable	n	R^2	leaves	Testing Sample Size	Pearson's Correlation	EF	MSE	RMSE	Paired T-test significant	Moran's I significant
Linear											
Chlorophyll	log(SEC)	133	0.72	N/A	55	0.88	0.76	0.06	0.25	no	no
Phycocyanin	log(SEC)	83	0.38	N/A	32	0.57	0.32	0.50	0.71	no	no
Microcystin	log(SEC)	114	0.40	N/A	48	0.49	0.16	0.87	0.93	no	no
CART											
Chlorophyll	log(SEC)	133	0.77	5	55	0.88	0.76	0.06	0.25	no	no
Phycocyanin	log(SEC)	83	0.43	4	32	0.52	0.26	0.54	0.74	no	no
Microcystin	log(SEC)	114	0.46	4	48	0.49	0.13	0.90	0.95	no	no

EF; modeling efficiency
MSE; mean square error
RMSE; root mean square error

Table 5. Example full models including significant predictor and model performance based on predicted values of the testing data set

Response Variable	Significant Predictor Variable	n	R^2	leaves	Testing Sample Size	Pearson's Correlation	EF	MSE	RMSE	Paired T-test significant	Moran's I significant
Linear											
Chlorophyll	- log(SEC) + log(TN) - log(N:P)	116	0.82	N/A	48	0.90	0.81	0.06	0.243	no	no
Phycocyanin	log(CHL)	71	0.59	N/A	28	0.69	0.42	0.34	0.581	no	no
Microcystin	log(TN) + log(TSS)	66	0.42	N/A	24	0.66	0.38	0.66	0.814	no	no
CART											
Chlorophyll	log(SEC),log(TP), log(TN), log(TSS)	116	0.86	7	48	0.90	0.80	0.06	0.249	no	no
Phycocyanin	log(CHL)	71	0.56	3	28	0.53	0.14	0.50	0.707	no	no
Microcystin	log(TN), log(TSS)	66	0.63	6	24	0.59	0.32	0.72	0.851	no	no

EF; modeling efficiency
MSE; mean square error
RMSE; root mean square error

Table 6. Final Secchi models. Simplistic models were built from the entire data set, Secchi depth is the only predictor in models.

Model	n	R^2	leafs
Linear			
log(CHL) = 1.23 -1.40log(SEC)	188	0.74	N/A
log(PHY) = 0.55 -1.64log(SEC)	115	0.34	N/A
log(MC) = -2.02 – 1.84log(SEC)	164	0.34	N/A
CART			
log(CHL) ~ log(SEC)	188	0.77	5
log(PHY) ~ log(SEC)	115	0.38	3
log(MC) ~ log(SEC)	164	0.43	7

Table 7. Final full models. Models were built using the entire data set using all environmental variables as possible predictors.

Model	n	R^2	leafs
Linear			
log(CHL) = -0.41log(SEC) + 0.76log(TN) – 0.49log(N:P) + 0.24log(TSS)	164	0.82	N/A
log(PHY) = -0.82 + 1.65log(CHL)	99	0.56	N/A
log(MC) = -8.22 – 1.54log(SEC) + 1.81log(TP) + 1.27(N:P) + 1.75log(TSS)	90	0.43	N/A
CART			
log(CHL) ~ log(TP), log(TN), log(SEC)	164	0.78	5
log(PHY) ~ log(CHL)	99	0.49	2
log(MC) ~ log(CHL), log(PHY), log(TN), log(N:P)	90	0.57	6

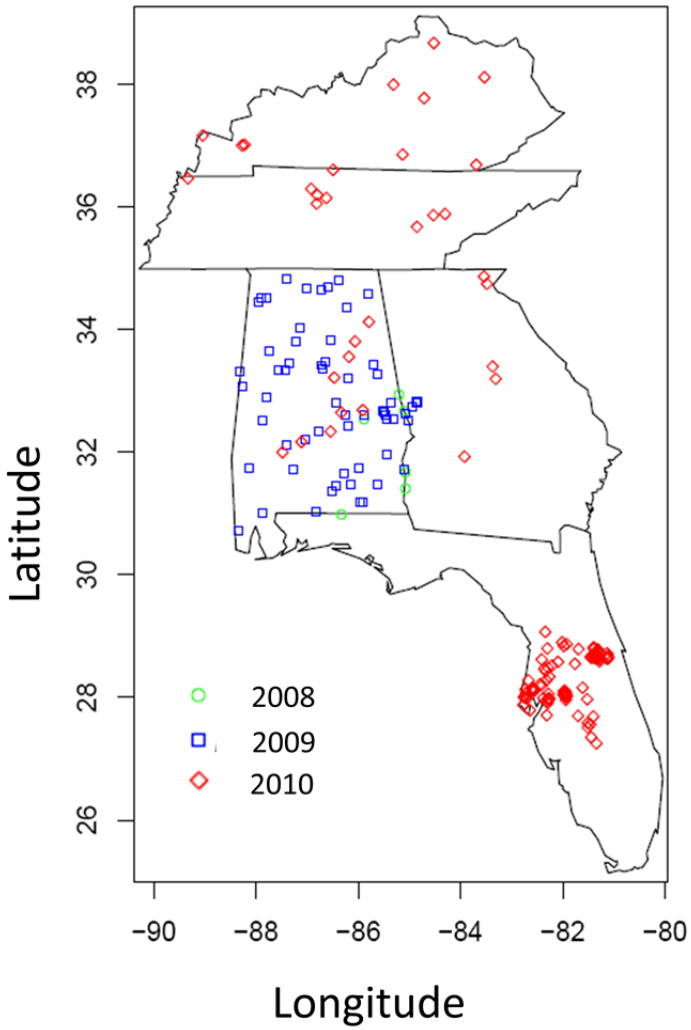


Figure 1. Distribution of waterbodies sampled in Alabama, Florida, Georgia, Kentucky and Tennessee broken down by year of case used in model building data set.

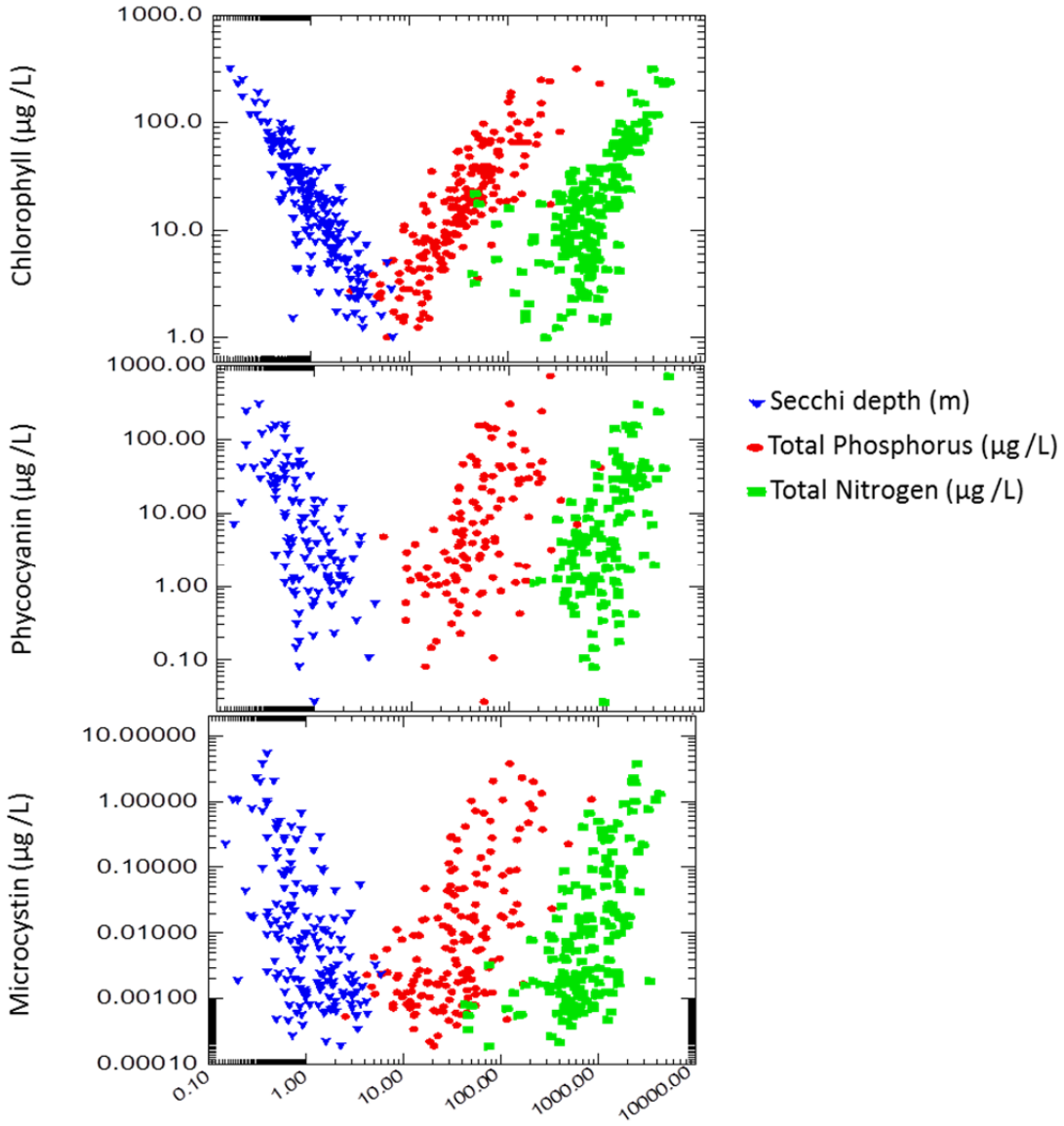


Figure 2. Trends between response and predictor variables.

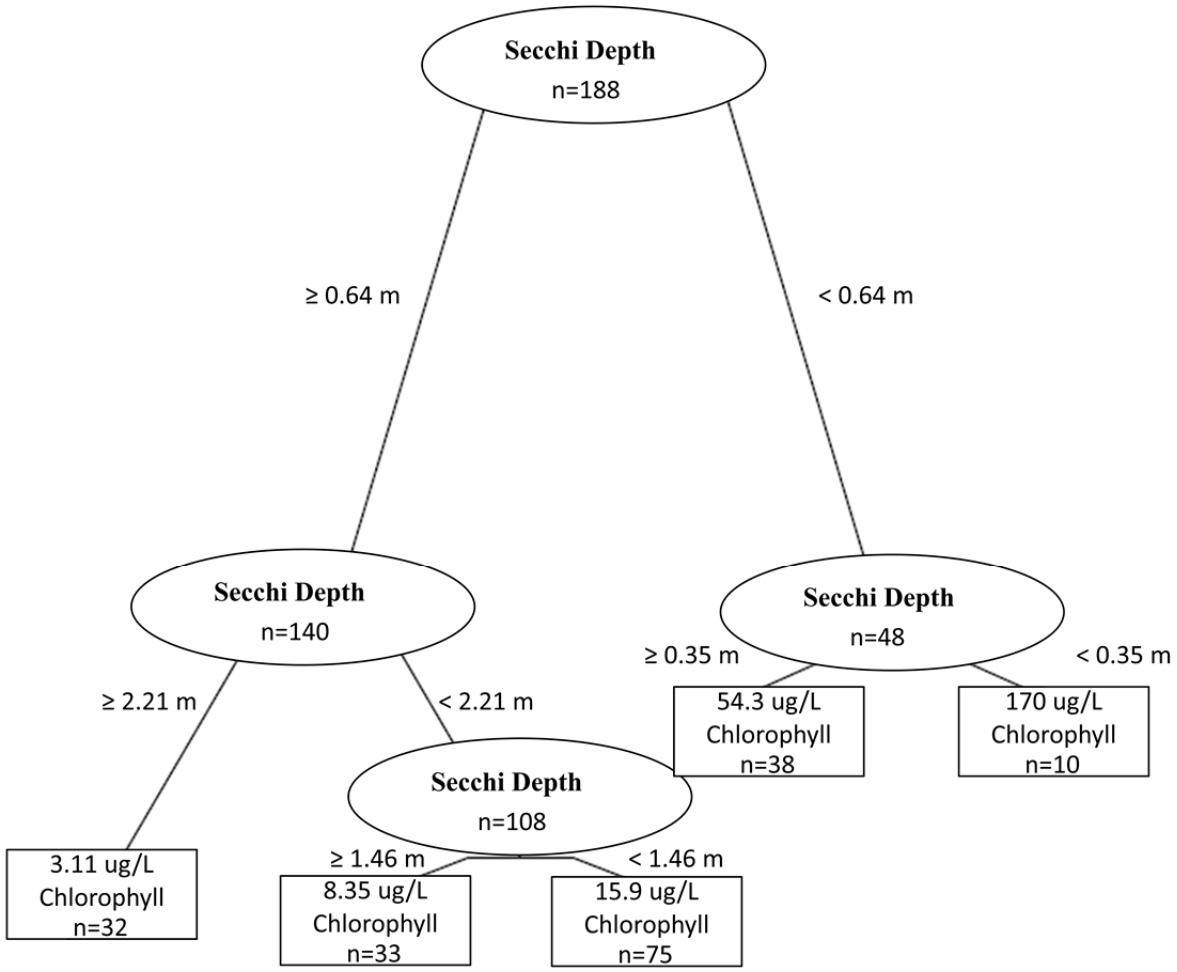


Figure 3a. Secchi chlorophyll CART model built from the complete data set.

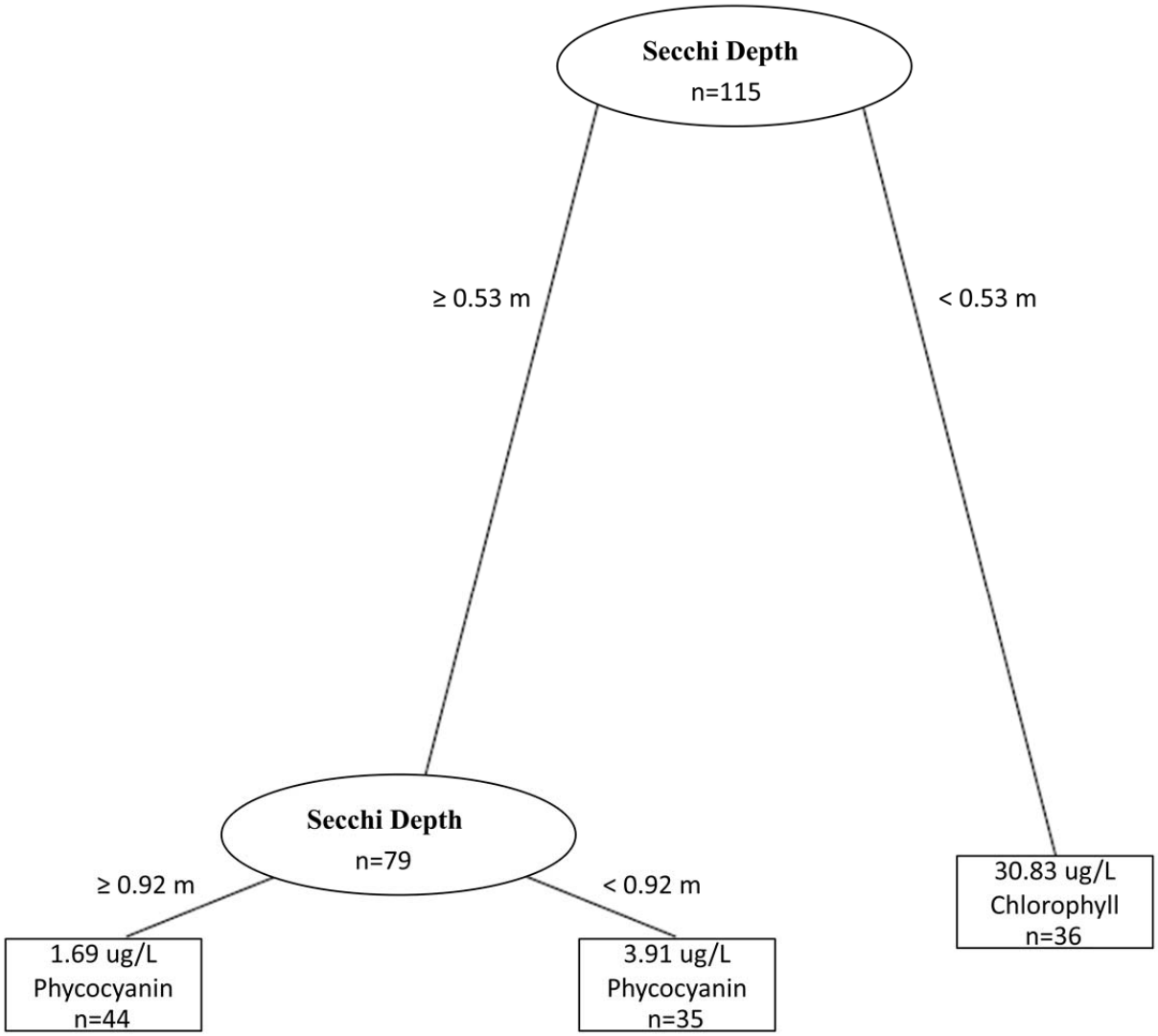


Figure 3b. Secchi phycocyanin CART model built from the complete data set.

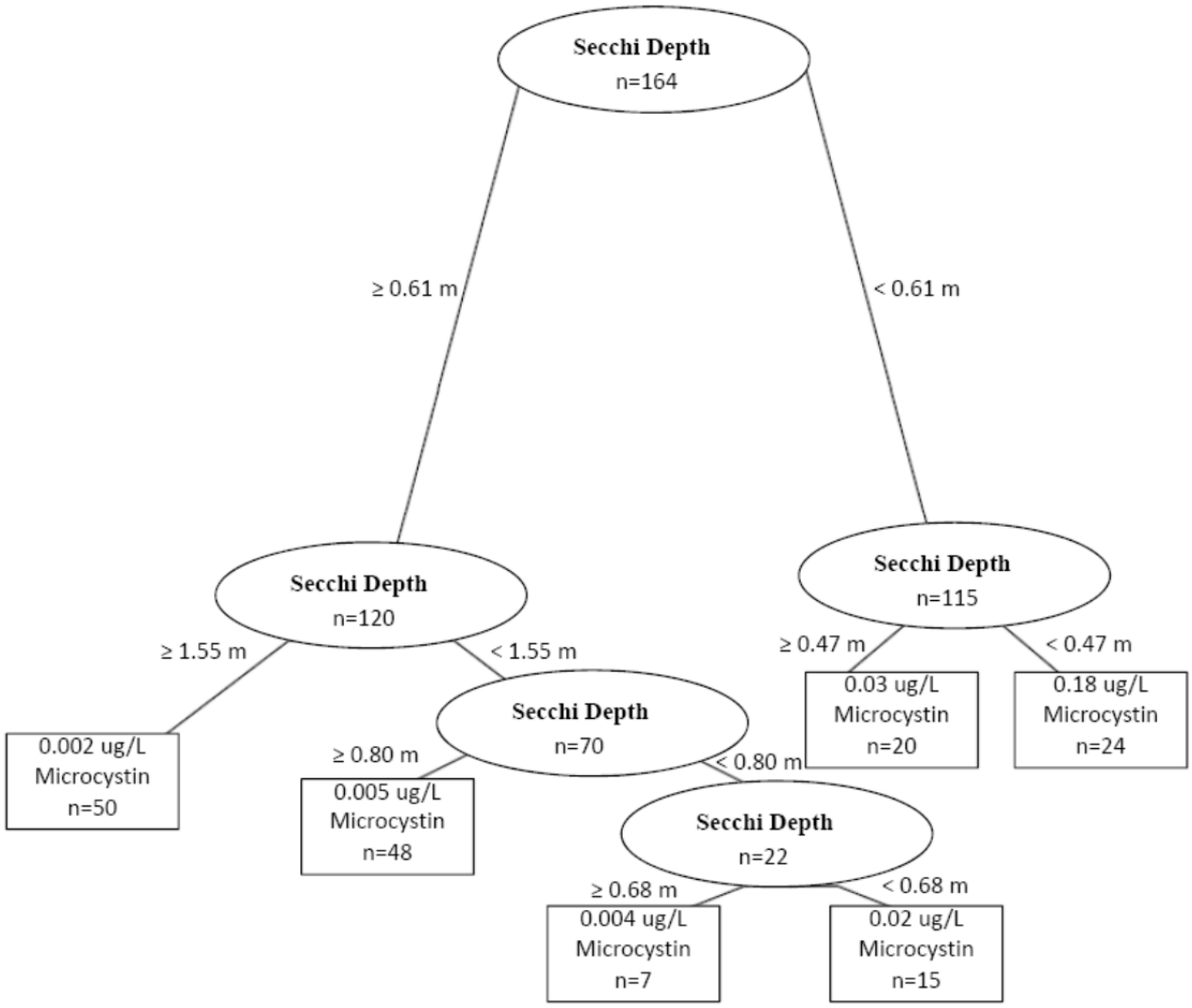


Figure 3c. Secchi microcystin CART model built from the complete data set.

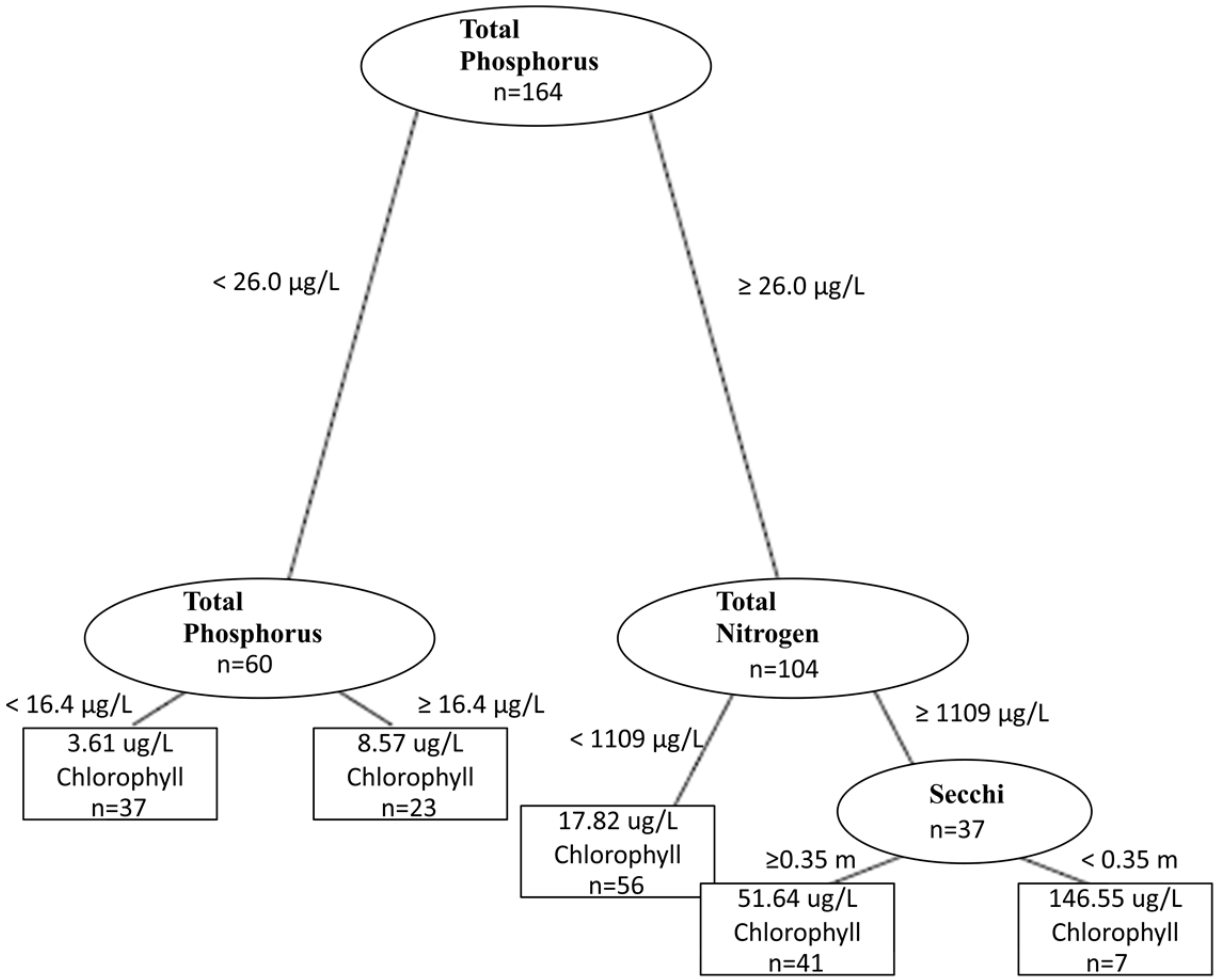


Figure 4a. Full chlorophyll CART model built from the complete data set.

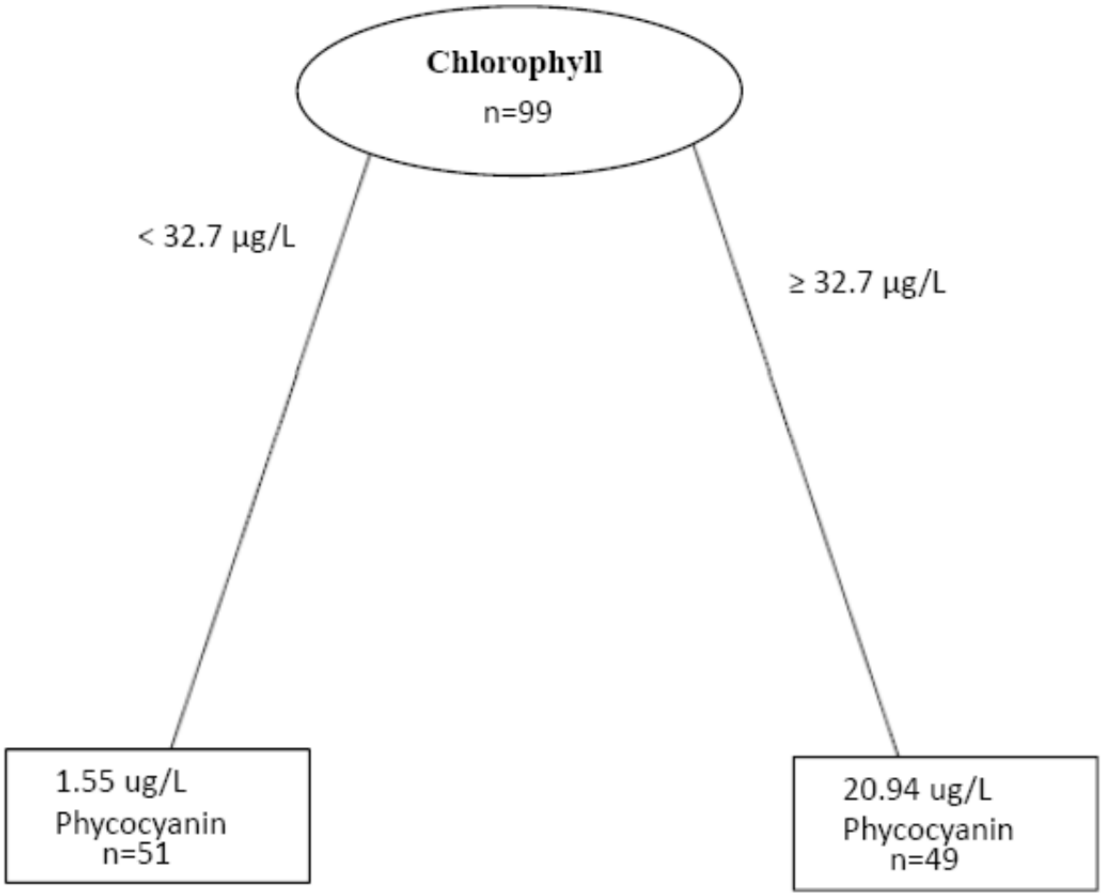


Figure 4b. Full phycocyanin CART model built from the complete data set.

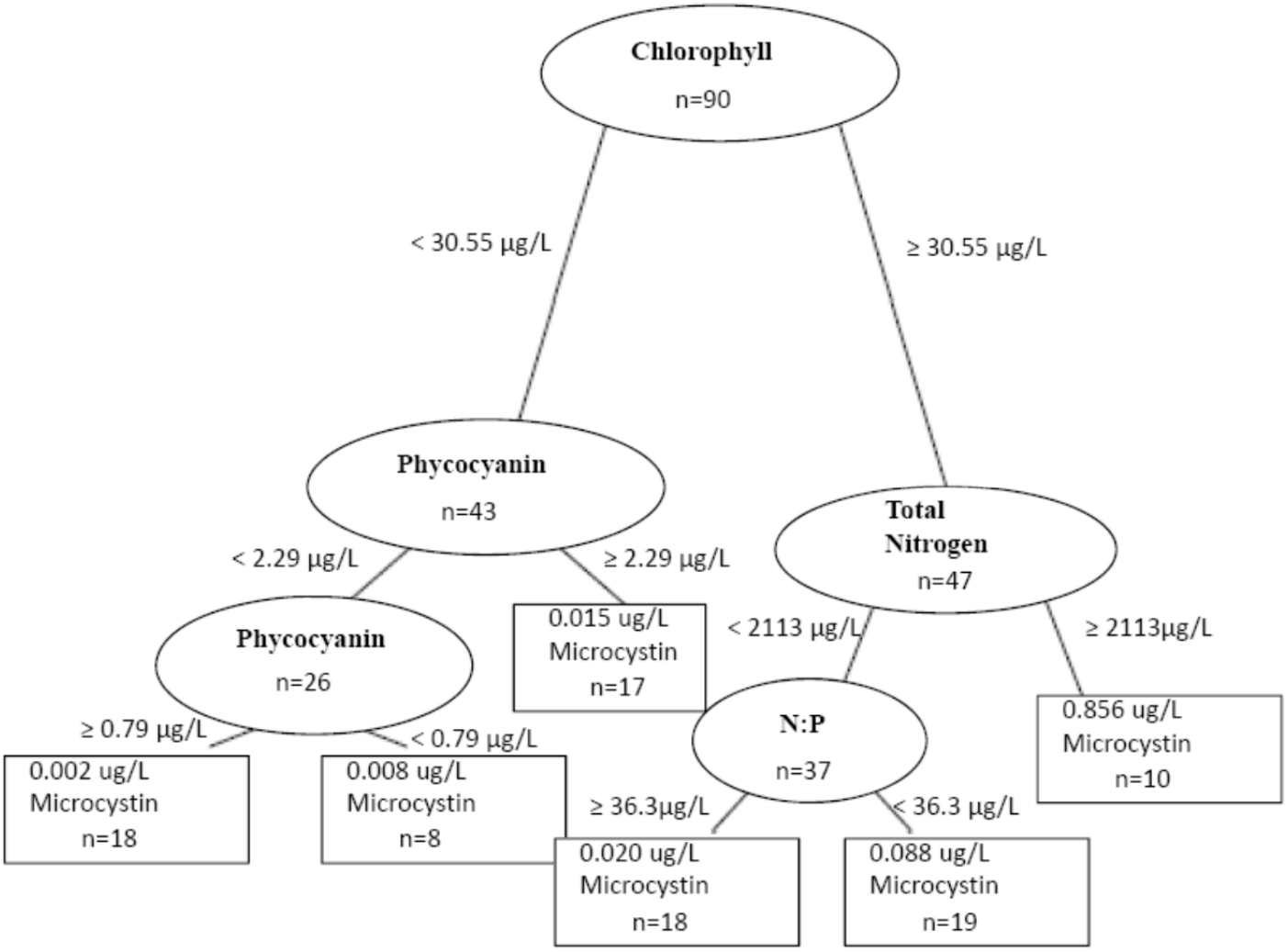


Figure 4c. Full microcystin CART model built from the complete data set.

Chapter Three: A multilocus alternative to the phycocyanin operon reveals hidden genetic diversity within a *Microcystis* population

Introduction

Microbial diversity is necessary to maintain ecosystem functions and stability (Bell et al. 2005). Within aquatic systems, increased microbial diversity has been shown to contribute to ecologically important cycles, such as decomposition (Gessner et al. 2010). Cyanobacteria, ubiquitous autotrophic prokaryotes, can have negative impacts on ecosystem services through the formation of harmful algal blooms (HABs). *Microcystis aeruginosa,* a cyanobacterium that often dominates HABs in nutrient-enriched freshwater systems, produces a hepatotoxin, microcystin, which has major public-health impacts (Carmichael et al. 2001). Molecular studies have identified high genetic diversity within *Microcystis* populations (Dadheech et al. 2010, Yoshida et al. 2008), and at least one recent study identified large variation in a critical ecological trait that appears to have a genetic basis (Bozarth et al. 2010). Given these findings, a better understanding of within-population genetic variance is needed for this important phytoplankter.

Studies to date have described genetic diversity within and between *Microcystis* populations via application of one of several polymorphic markers (Allender et al. 2009, Haande et al. 2007, Humbert et al. 2005). Since 1995, ~50 studies have used the intergenic spacer region of the phycocyanin gene (PC-igs) to quantify genetic diversity of *Microcystis* populations (Baker et al. 2001, Neilan et al. 1995, Tillett et al. 2001), making it the most popular method in active use. An alternative method originally developed for pathogenic bacteria, multilocus sequence typing (MLST), has recently been developed for *Microcystis* that may provide greater sensitivity for detecting genetic variation within and among populations of this cyanobacterium (Tanabe et al. 2007).

Multilocus typing, which examines sequence variation at five to seven core housekeeping genes, is capable of detecting clonal subdivisions as well as horizontal gene transfer (Maiden et al. 1998). The sensitivity of MLST for various pathogenic bacteria has been compared with other methods for examining clonality based on virulence (e.g., serotype, pulse field gel electrophoresis, multi-virulence gene loci sequence typing). MLST has had variable success describing unique clones when compared these other techniques (Peacock et al. 2002, Zhang et al. 2004). As MLST is applied in non-epidemiological contexts, the organisms' population level variation as described by MLST needs to be understood. We addressed this by comparing genetic diversity within a *Microcystis* population as measured by MLST versus PC-igs. We expected that MLST would reveal greater diversity given the increased sampling of the genome, relative to the single locus PC-igs approach. Our results show that MLST can detect genetic differences that the PC-igs method completely misses, a finding that can be critical when correlating ecological traits to evolutionary relationships.

Methods

Here, six clones of *Microcystis aeruginosa* were isolated over a four-year period (2006-2009) from Gull Lake (Hickory Corners, MI, USA). Colonies were collected from the mixed layer. Individual colonies were then sequentially pipetted through a series of DI washes until finally placed the growth media. Isolateswere cultured in half-strength WC-S medium (Stemberger 1981). In assessing genetic variation among the six isolates, we compared PC-igs with an abbreviated MLST scheme.

DNA was extracted from pelleted cultures using Lyse-N-Go PCR reagent (Pierce Chemical Co, Rockford, IL). The PC-igs in addition to five housekeeping genes for MLST, *ftsZ*, *glnA*, *pgi*, *gltX*, and *gyrB*, were amplified using primer and PCR conditions previously described

(Neilan et al. 1995, Tanabe et al. 2007). This is the first study of which we are aware of that uses MLST to genetically differentiate *M. aeruginosa* strains collected in the United States. Amplified PCR fragments were purified using QiagenQIAquick® PCR purification kit (Cat # 28104) and sequenced bidirectionally using an ABI 3730 Genetic Analyzer. Contiguous sequences were assembled using Sequencher 4.10.1(Gene Codes Corporation). PC-igs sequences were aligned with MEGA4 (Tamura et al. 2007). Protein coding sequences were examined for non-synonymous mutations (Tamura et al. 2007). A pair-wise comparison of the average number of nucleotide differences per a site (π) was conducted, in addition to estimating the haplotype diversity (*h*) of the PC-IGS region and MLST loci using DnaSP version 4.90.1 (Rozas et al. 2003). All sequences were submitted to GenBank (accession number JN226766-JN226772; HQ847833-HQ847857).

Results and Discussion

The PC-igs method revealed no nucleotide diversity among the six *M. aeruginosa* isolates. In contrast, the abbreviated MLST approach clearly differentiated all six isolatesas unique strains (Table 1, N) despite the fact that observed nucleotide diversity (π) was at least an order of magnitude lower than previously described for Asian *Microcystis* populations (Tanabe et al. 2007). Within our study population, the increased sensitivity of MLST with respect to PC-igs is not simply due to the increase sampling of the genome but is also a function of increased variability in housekeeping loci vs the PC-igs locus (as measured by nucleotide diversity per site (π)). For our focal population, the added investment of sampling the two additional loci (*recA* and *tpi*) described for the *Microcystis* MLST could not have differentiated more unique clones but may increase confidence of evolutionary relationships. This disparity in detected genetic diversity (Table 1, π, K_a/K_s) among housekeeping loci should be kept in mind along with the

study's purpose (describing evolutionary relationships or quantifying diversity) when choosing loci for an abbreviated MLST scheme.

MLST analysis is not ideal for all situations given the higher costs associated with increased loci sampling as well as larger amounts of required DNA template. However, the added cost associated with the sequencing of multiple loci can be overcome with next generation sequencing technology (Brockhurst et al. 2011). In contrast, the increased DNA requirement may not be as easily overcome. Since multiple loci along the genome need to be related back to a single individual or clone, MLST approaches require the successful culturing of *Microcystis* colonies. Successful isolation into culture has the potential to introduce bias with respect to the goal of describing diversity in the source population, but our isolation techniques have been refined to the point where most isolated colonies (~80%, personal observation) result in successful cultures. Although the MLST requirement of culturing individuals may not be practical for large-scale ($n \geq 1000$) population studies, the housekeeping loci are superior to the PC-igs marker for capturing fine-scale genetic diversity within *Microcystis* populations per unit effort as reflected by the observed nucleotide diversity (π) in our study.

We thank Jeff White for providing the *Microcystis* strains used in this study. Scott R. Santos aided in sequence assembly and calculating nucleotide diversity. The MLST analysis was greatly improved by assistance from Covadonga Arias.

References

Allender, C.J., LeCleir, G.R., Rinta-Kanto, J.M., Small, R.L., Satchwell, M.F., Boyer, G.L., and Wilhelm, S.W. 2009. Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells. Applied and Environmental Microbiology **75**(11): 3598-3604.

Baker, J., Neilan, B.A., Entsch, B., and Mckay, D.B. 2001. Identification of cyanobacteria and their toxigenicity in environmental samples by rapid molecular analysis. Environmental Toxicology **16**(6): 471-482.

Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L., and Lilley, A.K. 2005. The contribution of species richness and composition to bacterial services. Nature **436**(7054): 1157-1160.

Bozarth, C.S., Schwartz, A.D., Shepardson, J.W., Colwell, F.S., and Dreher, T.W. 2010. Population turnover in a *Microcystis* bloom results in predominantly nontoxigenic variants late in the season. Applied and Environmental Microbiology **76**(15): 5207-5213.

Brockhurst, M.A., Colegrave, N., and Rozen, D.E. 2011. Next-generation sequencing as a tool to study microbial evolution. Molecular Ecology **20**(5): 972-980.

Carmichael, W.W., Azevedo, S., An, J.S., Molica, R.J.R., Jochimsen, E.M., Lau, S., Rinehart, K.L., Shaw, G.R., and Eaglesham, G.K. 2001. Human fatalities from cyanobacteria: Chemical and biological evidence for cyanotoxins. Environmental Health Perspectives **109**(7): 663-668.

Dadheech, P.K., Ballot, A., Casper, P., Kotut, K., Novelo, E., Lemma, B., Proschold, T., and Krienitz, L. 2010. Phylogenetic relationship and divergence among planktonic strains of *Arthrospira* (Oscillatoriales, Cyanobacteria) of African, Asian and American origin deduced by 16S-23S ITS and phycocyanin operon sequences. Phycologia **49**(4): 361-372.

Gessner, M.O., Swan, C.M., Dang, C.K., McKie, B.G., Bardgett, R.D., Wall, D.H., and Hattenschwiler, S. 2010. Diversity meets decomposition. Trends in Ecology & Evolution **25**(6): 372-380.

Haande, S., Ballot, A., Rohrlack, T., Fastner, J., Wiedner, C., and Edvardsen, B. 2007. Diversity of *Microcystis aeruginosa* isolates (Chroococcales, Cyanobacteria) from East-African water bodies. Archives of Microbiology **188**(1): 15-25.

Humbert, J.F., Duris-Latour, D., LeBerre, B., Giraudet, H., and Salencon, M.J. 2005. Genetic diversity in *Microcystis* populations of a French storage reservoir assessed by sequencing of the 16S-23S rRNA intergenic spacer. Microbial Ecology **49**(2): 308-314.

Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J.J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Sciences of the United States of America **95**(6): 3140-3145.

Neilan, B.A., Jacobs, D., and Goodman, A.E. 1995. Genetic diversity and phylogent of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. Applied and Environmental Microbiology **61**(11): 3875-3883.

Peacock, S.J., de Silva, G.D.I., Justice, A., Cowland, A., Moore, C.E., Winearls, C.G., and Day, N.P.J. 2002. Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing *Staphylococcus aureus* isolates in a microepidemiological setting. Journal of Clinical Microbiology **40**(10): 3764-3770.

Rozas, J., Sanchez-DelBarrio, J., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**(18): 2496-2497.

Stemberger, R.S. 1981. A general approach to the culture of planktonic rotifers. Canadian Journal of Fisheries and Aquatic Sciences **38**(6): 721-724.

Tanabe, Y., Kasai, F., and Watanabe, M.M. 2007. Multilocus sequence typing (MLST) reveals high genetic diversity and clonal population structure of the toxic cyanobacterium *Microcystis aeruginosa*. Microbiology **153**: 3695-3703.

Tillett, D., Parker, D.L., and Neilan, B.A. 2001. Detection of toxigenicity by a probe for the microcystin synthetase A gene (*mcyA*) of the cyanobacterial genus *Microcystis*, comparison of toxicities with 16S rRNA and phycocyanin operon (phycocyanin intergenic spacer) phylogenies. Applied and Environmental Microbiology **67**(6): 2810-2818.

Yoshida, M., Yoshida, T., Satomi, M., Takashima, Y., Hosoda, N., and Hiroishi, S. 2008. Intra-specific phenotypic and genotypic variation in toxic cyanobacterial *Microcystis* strains. Journal of Applied Microbiology **105**(2): 407-415.

Zhang, W., Jayarao, B.M., and Knabel, S.J. 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. Applied and Environmental Microbiology **70**(2): 913-920.

Table 1. Genetic diversity by method (PC-igs and MLST) and individual genes for MLST.

Locus	bp length	N	S	$\pi * 100$	K_a/K_s
PC-igs	625	1	0	0.00	N/A
Concatenated MLST loci	2131	6	17	0.32	
ftsZ	409	4	5	0.60	0.00
glnA	452	4	5	0.46	0.07
gltX	430	3	3	0.23	0.57
pgi	424	2	3	0.24	0.15
gyrB	416	2	1	0.08	0.00

N, number of alleles

S, number of segregating sites

π, nucleotide diversity per nucleotide site

K_a/K_s, ratio of non-synonymous to synonymous mutations in protein coding regions

Appendix: Use of Secchi models by citizen scientists

What is a Secchi disk?

A Secchi disk is a tool that is used to describe water clarity based on how deep a person can see in water. The disk is normally 20 cm wide and alternates black and white. The depth of the disk in the water is determined by attaching the disk to a calibrated rope or measuring tape. Measurements can be taken from boats or docks. Once the Secchi depth of your water is measured, you can predict the amount of phytoplankton (by chlorophyll *a*) and the cyanobacterial biovolume.

How to use a Secchi disk

Measurements are least reliable on cloudy days, and near dawn or dusk. Remove sunglasses before taking measurements.

1. Lower the disk into the water on the **sunny** side of the boat or dock. Wait for clouds overhead to pass before taking measurements.
2. Lower the disk until it is out of sight. Record the depth when the disk disappears.
3. Raise the disk until reappears. Record the depth*.
4. The Secchi depth is the average of the two measurements.
5. Be sure to note unit of measurement on the measuring tape. Most metric measuring tapes will be in centimeters (cm), however Secchi depth is normally described using meters. Before predicting the amount of phytoplankton or cyanobacteria the Secchi depth must be converted to meters using the boxed equation below.

* All measurements presented here are metric, use the boxed equations below to convert inches to metric measurements.

$$\frac{Secchi\ depth\ in\ cm}{100} = Secchi\ depth\ in\ meters$$

$$\frac{Secchi\ depth\ in\ inches}{39} = Secchi\ depth\ in\ meters$$

Appendix *continued*

Making a Secchi Disk

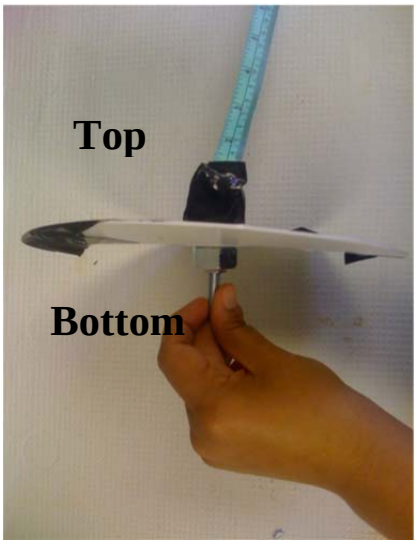
Materials

- 2 gallon plastic pail lid
- 2”-3” eye bolt, washer and large nut
- Plastic metric measuring tape (found in sewing section of store)
- Small cable ties
- Black masking tape
- Drill or hammer and nail

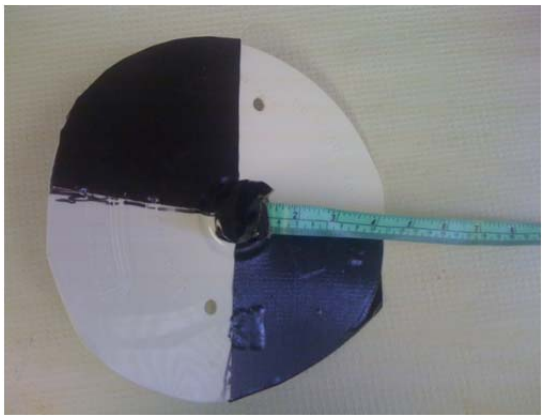
Directions

1. Cut rim off lid.
2. Drill hole in center of rim-less lid.
3. Cover 2 opposing quarters of the lid top with black masking tape.
4. Thread eye bolt through hole in center of disk. The hook should be on the top side of the lid. Add washer and tighten large nut below the disk.
5. Attach zero end of measuring tape to eye bolt using cable ties. Be sure to have the zero line flush with the lid when securing tape.
6. Wrap connection between eye bolt and measuring tape for added stability.

Finished Secchi disks



Finished disk from side



Finished disk from above

Using a Secchi disk to predict chlorophyll *a*

1. Measure Secchi depth in meters (directions above).
2. Find circle **A** in Figure 1.

Step 3
a. If your Secchi depth is greater than (<) 0.6 meters follow the right branch from circle **A** to circle **B**. Continue to **Step 4**.
-OR-
b. If your Secchi depth is less than or equal to (≥) 0.6 meters follow the left branch to circle **C.** Continue to **Step 5**.

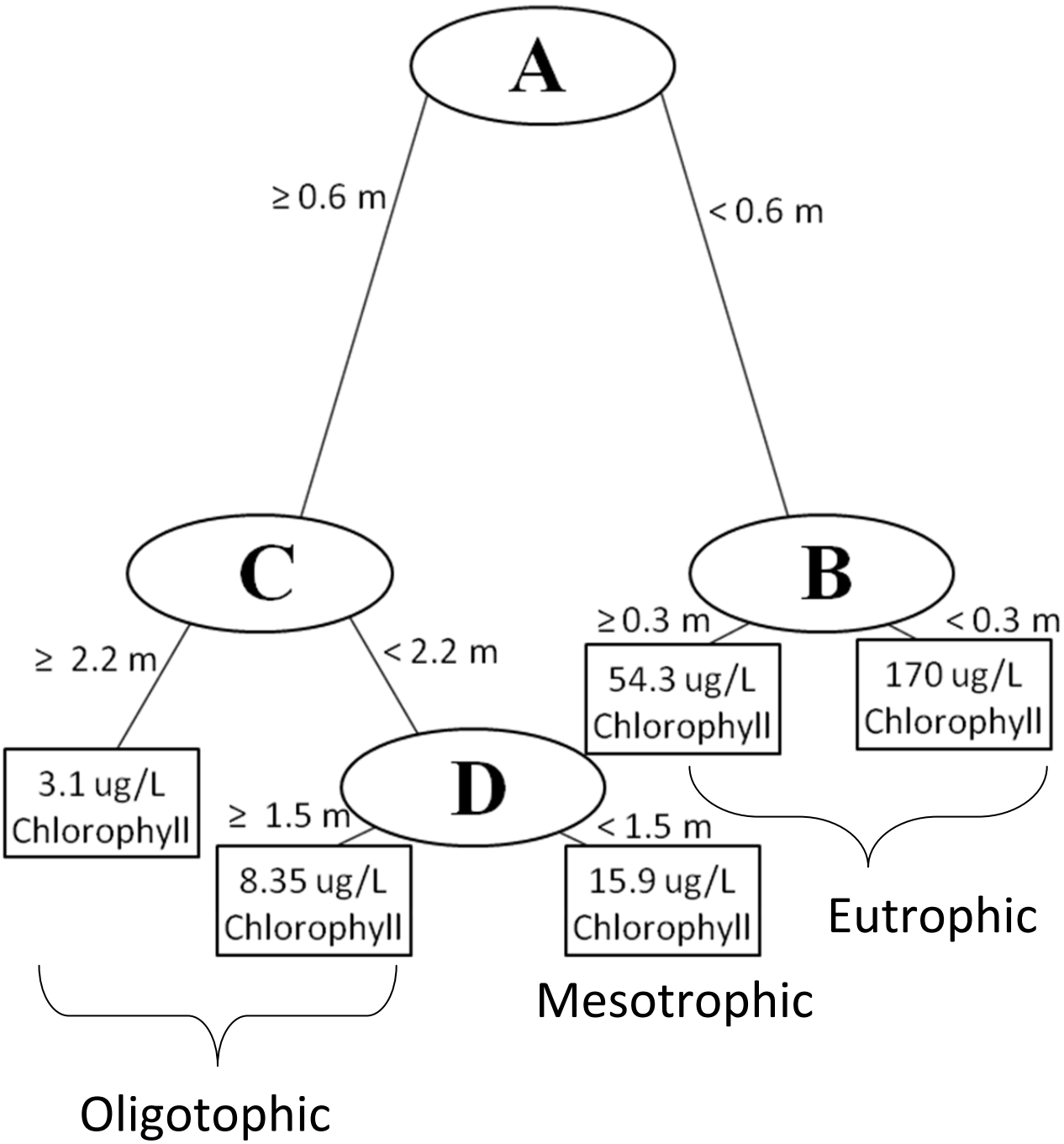
Step 4
a. At circle **B,** if your Secchi depth is greater than (<) 0.3 meters follow the right branch to the box with your estimated value.
-OR-
b. At circle **B,** if your Secchi depth is less than or equal to (≥) 0.3 meters follow the left branch to the box with your estimated value.

Step 5
a. At circle **C,** if your Secchi depth is greater than (<) 2.2 meters follow the right branch to circle **D**. Continue to **Step 6**.
-OR-
b. At circle **C,** if your Secchi depth is less than or equal to (≥) 2.2 meters follow the left branch to the box with your estimated value.

Step 6
a. At circle **D,** if your Secchi depth is greater than (<) 1.5 meters follow the right branch to the box with your estimated value.
-OR-
b. At circle **D,** if your Secchi depth is less than or equal to (≥) 1.5 meters follow the left branch to the box with your estimated value.

Appendix *continued*

Figure 1. Chlorphyll *a*



Appendix *continued*

Using a Secchi disk to predict risk of cyanobacteria

1. Measure Secchi depth in meters (directions above).
2. Find circle **A** in Figure 2.
3a. If your Secchi depth is greater than (<) 0.5 meters follow the right branch from circle **A** to the boxed estimated value.

-OR-

3b. If your Secchi depth is less than or equal to (≥) 0.5 meters follow the left branch to circle **B.** Continue to **Step 4**.

4a. At circle **B,** if your Secchi depth is less than or equal to (≥) 0.9 meters follow the left branch to the boxed estimated value.

-OR-

4b. If your Secchi depth is greater than (<) 0.9 meters follow the right branch to the boxed estimated value.

Figure 2. Cyanobacterial biovolume