Environmental Drivers of Harmful Algal Blooms in Michigan Inland Lakes

THESIS FOR THE DEGREE OF
MASTER OF SCIENCE IN CHEMISTRY

HAMZAH DANYAL ANSARI

OAKLAND UNIVERSITY

2018

Environmental Drivers of Harmful Algal Blooms in Michigan Inland Lakes

by

HAMZAH DANYAL ANSARI

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN CHEMISTRY

2018

Oakland University
Rochester, Michigan

Thesis Advisory Committee:

Roman Dembinski, Ph.D., Chair
David C. Szlag, Ph.D.
Linda Schweitzer, Ph.D.
Thomas Raffel, Ph.D.

I dedicate this thesis to Carl Sagan who greatly inspired my ambition to think critically and meticulous, ultimately leading me to hone my skills in the pursuit of science. The memory of his voice will forever echo within my conscious

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. David Szlag, for providing me the opportunity to grow in my laboratory skills and critical thinking. He has believed in my diverse ability to apply my knowledge in the pursuit of science. I also would like to thank Annie Ettinger for pushing me to be an effective scientist, instilling me with great laboratory management tips. Also I would like to thank Brian Spies and Andrew Herppich for aiding me in all the chemical analysis on our collected samples. I also want to give special thanks to Marc Lucido, Brayden Metcalf and Mikaela Cantu for the lending their support in conducting sample preparation and analysis. I also would like to thank Dr. Raffel's and his lab group, especially Ryan Mcwhinnie and Jason Sckrabulus for doing the bulk of the field sampling and mussel counts. I also would like to thank the Oakland University Chemistry Department. I want to give special thanks to Dr. David Newlin and Dr. Linda Schweitzer who has personally helped me push along my graduate program and gave me hope to strive on.

I also would like to acknowledge the Michigan Department of Environmental Quality for funding this research (Grant # 36636). My research would not have happen without their financial support.

Hamzah Danyal Ansari

ABSTRACT

Environmental Drivers of Harmful Algal Blooms in Michigan Inland Lakes

by

Hamzah Danyal Ansari

Adviser: David C. Szlag, Ph.D.

The increase in harmful cyanobacteria blooms threatens freshwater ecosystems and presents a risk to human health. The rapid growth of cyanobacteria can cause rapid declines in water quality if left unchecked. A survey of 29 inland lakes was conducted to investigate microcystin. Liquid chromatography-mass spectrometry (LC-MS/MS) was used to identify and quantify the microcystin variants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Harmful Algal Blooms

Water is indisputably vital for life on Earth. One of the major threats to water quality is the ongoing onslaught of Harmful Algal Blooms (HABs), a disastrous phenomenon which have impacted multiple areas around the world. With HABs, algae and cyanobacteria can grow out of control, often "painting" the water green. HABs are a result of over-productivity of phytoplankton biomass which mostly resides near-shore in the epipelagic zone often forming a thick layer [**?**]. HABs are dangerous due to the ecological damage from its rapid growth and the toxins they create. Unfortunately, the intensity, extent, and spatial coverage of HABs has been increasing globally due to more ecological disturbances [**?**]. In most cases, HABs are found in coastal regions, streams, and freshwater lakes [**?**]. HABs is a broad term which includes a large variation of different generas, depending on the location and which effected water body. The composition of HABs are diverse ranging from different species of cyanobacteria, diatoms, algae, and dinoflagellates found worldwide [**?**]. In the context of this paper, we focus on HABs that affects freshwater systems, more specifically in the state of Michigan.

HABs has become a national concern to public health and their occurrences. One of the dangerous qualities of HABs are the toxin compounds that are released. Cyanobacterial toxins (cyanotoxins) are diverse as over 600 peptides have been discovered [**?**]. In response to their occurrences, the Harmful Algal Bloom and Hypoxia Research and Control Amendments Act (HABHRCA) of 2014, originally

enacted in 1998, requires national programs to research and monitor HABs and to mitigate the harmful effects [**?**]. The Contaminant Candidate List (CCL) required by the Safe Drinking Water Act (SDWA) has listed cyanobacteria and cyanotoxins to be investigated by the United States Environmental Protection Agency (USEPA) to have a set Maximum Contaminant Level (MCL) which is subject to strict regulations [**?**].

One of the most prevalent toxin, Microcystin (MC), is a small cyclic peptide having a large range of different structure. MC is hepatotoxic which inhibits protein phosphatases 1, 2A and 3 [**?**]. MC is produced mostly within the *Microcystis* genera but also by other generas such as *Anabaenopsis, Nostoc,* and *Planktrothrix* [**?**, **?**]. For MC, the World Health Organization (WHO) have a guidance level of 1 μg/L for drinking water and 20 μg/L for recreational water [**?**, **?**]. The USEPA recently announced a new drinking water Health Advisory (HA) for MC with a guidance level of 4 μg/L for recreational surface water [**?**]. Methods of quantifying MC will often measure Microcystin-LR (MC-LR). Levels above those guidelines requires state officials to issue a public health advisory not to swim in the effected area. The common routine methods employed by state agencies are ADDA-ELISA kits by Abraxis[1]. Enzyme-linked Immunosorbent Assay (ELISA) is an antibody method that detects and measures the 3-amino-9-methoxy-10-phenyl-2,6,8-trimethyl-deca-4,6-dienoic acid (ADDA) moiety of the MC. The ADDA moiety interacts with the Y-shape active site of protien phosphatase [**?**] The test kit has good cross-reactivity with other congeners, and it use MC-LR for its calibration standards, thus the reported value is in terms of MC-LR equivalence.

[1]Abraxis, Inc 124 Railroad Drive, Warminster, PA 18974

Cylindrospermonsin is an alkaloid produced by *Cylindrospermopsis, Anabaena, Oscillatoria* and other species of cyanobacteria [**?**]. In addition to the working draft by the USEPA, Cylindrospermonsin also have a HA with a guidance level of 8 μg/L not to be exceeded for more than one day [**?**]. Cylindrospermopsin is a hepatotoxin and potentially carcinogenic [**?**]. The mechanism of toxicity is not understood, however some studies have shown it to be genotoxic, causing breakage of DNA strands [**?**, **?**]. Another dangerous group are called saxitoxins and are sodium channel blockers, a potent neurotoxin which paralyze and lead to death by respiratory failure [**?**]. Anatoxins are also very dangerous exhibiting potent toxicity. It is known as Very Fast Death Factor due to its ability to irreversibly bind to nicotinic acetylcholine receptors which consequently leading to respiratory failure in a very quick manner [**?**, **?**]. Currently, anatoxin is listed in the CCL to be considered for regulation [**?**].

Microcystins are uniquely synthesized in cyanobacteria by a mix of two systems, Polyketide synthase (PKS) and Nonribosomal peptide synthetase (NRPS) [**?**]. This is different from how other proteins are normally synthesized ribosomally . The genetic mechanism of MC synthesis involves multiple protein modules spanning a 48 kilobase pair gene cluster which are responsible for incorporating different amino acids, ultimately creating the cyclic peptide [**?**, **?**]. The major amino residues of MC are of D-erythroβ-methylaspatric acid (D-MeAsp), ADDA, *N*-methyldehydro-alanine (Mdha) and other possible variable amino acids [**?**, **?**].

There are over 100 known variants of congeners of MC with 6 congeners recognized by the USEPA as chemical contaminants which are Microcystin-LA (MC-LA), Microcystin-LF (MC-LR), MC-LR, Microcystin-LY (MC-RR), Microcystin-RR (MC-RR), and Microcystin-YR (MC-LR) [**?**]. The most frequently occurring and potent congener variant of MC is MC-LR [**?**]. Figure **??** shows the

structure of MC-LR which contains L-luecine (L) and L-Arginine (R) in its variable positions. Other amino acids such as alanine (A), tryptophan (W), tyrosine (Y), and phenylalanine (F) can be substituted for other variable congeners. With the known variants, MC is roughly around 1000 Da [**?**]

The harm caused by HABs does not necessarily come entirely from the toxins. In cases of HABs where there has been a large accumulation biomass, they can quickly die off due to limiting resources and cause eutrophication [**?**]. The dead biomass is quickly consumed by other aquatic microorganisms which increases respiration rates, depletes dissolved oxygen creating an unsuitable habitat for other aquatic organisms. [**?**]. The layer of scum formed from HABs also block light for submerged aquatic macrophytes, killing them by preventing photosynthesis [**?**]. HABs can also be a major nuisance and a cost to the community, as this create odor and limit recreation [**?**, **?**].

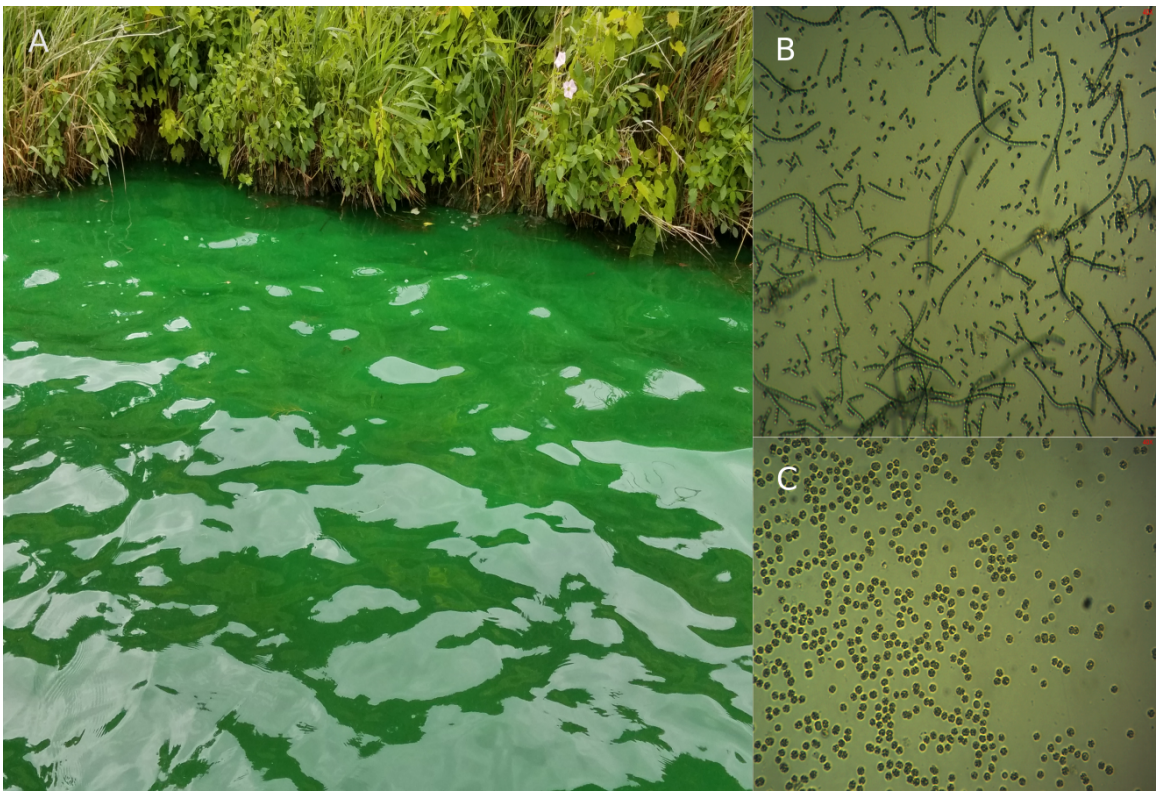

Figure 1.1: Examples of Harmful Algal Blooms. (A): A bloom in Manitou Lake. Picture taken by Fred Farcus. (B): Anabaena (C): Microcystis

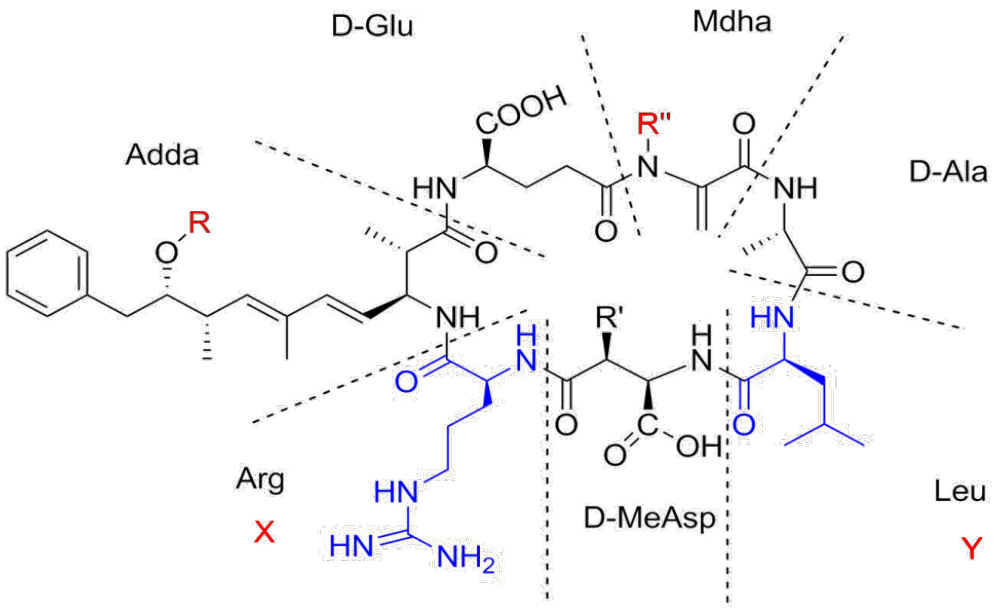

Figure 1.2: Structure of MC-LR

The most common structure of microcystin. The seven major moieties are separated by dashed lines. Position X and Y can be interchangeable with other variable amino acids. R and R" can also be found demethylated in other congeners.

Swimming or contact in any waterbody with HABs can pose a health risk as all HABs are irritants and some can have toxin producing species. Many of the species that are producing hepatotoxins and neurotoxins in high amounts can kill livestock, pets and wildlife [**?**]. The possible route of exposure for humans can be from dermal contact, accidental ingestion, breathing in lake spray aerosols and failure of purification in drinking water plants [**?**, **?**]. As stated, exposure through dermal contact can be an irritant. HABs can create lipopolysaccharide, an endotoxin, which create rashes after skin contact as it triggers a inflammatory response [**?**].

Ingestion is the most dangerous route of exposure for any cyanotoxins. Usually, complete immersion and prolonged exposure to affected waters increases the chance of ingestion of water with significant levels of cyanotoxins. Although the case is rare for humans, this is more dangerous to animals and pets [**?**]. Cyanotoxins can also be consumed in fish and wildlife that has drank enough amounts where it can bioaccumulate in their fatty tissues, eventually leading to ingestion by humans [**?**]. Drinking water from an affected lake could lead to acute toxicity or other symptoms [**?**]. Unfortunately, HABs can be found in storage reservoirs and source waters in regions that do not have sophisticated drinking water facilities in providing clean water.

In addressing the removal of toxins, a water treatment facility needs to understand where most of the toxins resides.Cyanotoxins are mostly intracellular, except for the case of cylindrospermonspin [**?**]. With a health risk management plan, an analysis of water input (raw water) needs to identify the genera and the toxins levels intracellular and extracellular concentrations [**?**]. The information of whether the toxin is intact within the cell or not can suggest different treatments. Intact cells should be removed as much as possible at the intake without cell lysis

[?]. Different treatment can be used with ozone, chlorination, activated carbon and advance oxidation process [?, ?]. However ozone and chlorination carries risk of cell lysis and disinfection by-products. The destruction of MC by ultraviolet radiation coupled with H_2O_2 is proven to be an effective treatment method [?]. The use of activated carbon or membrane filteration is effective but costly to implement as a routine [?]. Removal of toxins can be treated by oxidants such as $KMnO_4$ which oxidizes the dissolved toxins and prevent the cells to lyse [?]. An effective health risk managment plan would implement a multipoint check system in the water treatment process and also be economically viable [?, ?].

1.2 Environmental Drivers of Harmful Algal Blooms

Cyanobacteria in it's natural environment is complex and often unexpected. The mechanism for what drives the proliferation is not fully understood as some studies have attempted to explain what factors are responsible of their explosive growth [?, ?]. As autotrophs, HABs can rapidly grow under warm and nutrient-rich conditions [?]. HABs are most likely to occur in the summer or warmer months where primary productivity is most likely to peak due to increase daylight, warmer temperature and low water flow in streams and rivers [?, ?].

Excess nutrients is long believed to significantly impact the growth of HABs [?, ?]. However, the relationship is not as clear as some expect. Urbanization and agriculture has increased the frequency of conditions in lakes and coastal environments, often regarded as cultural eutrophication[?]. Organic and inorganic forms of nutrients play a role in biomass production. Nitrogen usually from non-point sources such as septic tanks, animal waste and agricultural runoff which contributes to HABs. Phosphorus is the main culprit in freshwater in causing HABs [?]. For the bloom in Lake Erie, some studies suggests the nutrient runnoff into the

Maumee river, which has and its watershed mostly comprised of agriculture [?, ?]. Other studies in other areas worldwide, similar to Lake Erie, have shown nutrient enrich conditions usually from agricultural runoff or disturbance of ecological conditions that impacts nutrient cycle to cause blooms [?, ?, ?, ?].

Ecological factors could be a major factor in some cases. Previous studies in inland Michigan lakes are finding *Dreissena polymorpha* (zebra mussels) to have an impact on finding blooms [?]. In a study done by Michigan State University [?] found zebra mussels can promote phytoplankton growth due to their effect on bloom ecology. Some studies suggests that of zebra mussels should be incorporated in building a predictive model as they have a significant impact on HABs with their presence [?, ?, ?]. In our survey, we will investigate whether zebra mussels have an influence on MC concentration or cyanobacteria population.

1.3 Application of Technology

Statistical predictive models are increasingly being used to forecast HABs. Models coupled with weather data have been increasingly successful with wind direction, speed, temperature and precipitation being the best predictor of HABs. Along coastal environment, National Oceanic and Atomospheric Administration (NOAA) uses real-time data from satellite to predict HABs on the Gulf of Mexico, Lake Erie, and other coastal environments based on satellite data and weather models [?]. Remote sensing instruments such as buoys equipped with sensors have signifigantly improved, capable of collecting multiple water characteristics in real-time. In predicting HABs, fluorescence of chlorophyll-a levels can be measured continuously which can indicate growth of cyanobacteria. In addition, phycocyanin, an accessory pigment found in cyanobacteria is also an effective parameter to uniquely identify and measure the relative biomass [?, ?]. This data can provide

more information in understanding what seems to drive their growth. Most importantly, the HABs forecast warnings can prevent exposure and improve communication with the public. They have provide comprehensive information continuously available for the public on their website [2].

Inland lakes, satellite imagery is not ideal cloudy conditions. Forecasting for inland lakes does not work with satellite imagery as the extent of the lake's area will often limit the predictability. An effective predictive models should be based on features that are found to contributes to HABs with measured abiotic and biotic factors. Understanding the main drivers for cyanobacteria can help to predict HAB occurrences.

One of the challenges in assessing water quality of a lake is the frequency of sampling. HABs is sporadic and often can be missed if sampling regime is too sparse. Monitoring for HABs requires frequent sampling regime inorder to not to miss a HABs events. Concentrations of cyanotoxins may fluctuate through time and sampling at a large interval may not capture the reality of the lake's condition.

Solid Phase Adsorbtion Toxin Tracking (SPATT) is a unique method of monitoring waterbodies in a more time-integrative approach. A porous bag or a sachet is filled with a resin inside a permeable bag. SPATT is used in for monitoring other analyte of interest such as diarrhetic shellfish poisoning [**?**]. The SPATT bag is submerged in a waterbody of interest for a period of time. During this period, free-floating compounds will adsorb onto the polymer beads. SPATT can are then retrieved and analyzed for chemical analytes of interest. This technique can be useful if sampling frequency is financially limited.

[2] `https://tidesandcurrents.noaa.gov/hab_info.html`

1.4 Goals and Aims

My goal is to explore what drives HAB development and build a predictive model based from our collected data. With the collected observations, I investigated the best possible predictive model from our dataset. Eventually with the built model based on each lake's unique geological characteristics and be ranked by the likelihood of HABs. Some studies build predictive models based using cyanobacterial cell count or mass, concentrations of chlorophyll-a and MC concentrations as a response variable as its most likely associated with HABs [**?**, **?**, **?**, **?**, **?**]. The total MC concentration measured by Liquid Chromatography/Tandem Mass Spectrometry (LC-MS/MS) was used as the main predictor variable of interest. The *16s rRNA* gene copies measured by Quantitative Polymerase Chain Reaction (QPCR) was also observed as a response variables as well as this measures relatively amount cyanobacteria. In our survey on 29 inland lakes in Michigan, we seek to understand what drives HABs and build a predictive model on MC. In addition, we also analyzed for cylindrospermopsin and anatoxin in our surveyed lakes. Before the survey, I hypothesized if the lake's watershed is more urbanized areas, I would expect higher MC concentrations. Developed land can increase nutrient runoff which can increase algal blooms. Previous studies have shown developed areas as having a major influence on the occurrence of HABs because of more possible sources like applied fertilizer or leaky septic tanks [**?**, **?**]. Lakes with higher developed areas may also have a possible influence nutrient mobility, which in turn drive MC production. I expect this relationship to also be similar with total cyanobacteria measured by QPCR, chloraphyl, *mcyE*.

With the extent of the study, I would expect to find some key features that can reveal a definite pattern which can describe what drives HABs. Prior to the study, I hypothesized lakes with dominant urbanization would see an increase in

MC concentrations. In addition, I also would expect the presence of Zebra mussels to have a positive influence as well. My research objectives:

1. Harmful algal blooms are influenced by developed/urban land, which can be used to predict MC concentrations
2. Lakes with the presence of zebra mussel will have a higher concentration of MC than lakes with none found.
3. Nutrient concentrations can be explained by land use characteristics
4. Identify important features that influence HABs.

CHAPTER 2

SURVEY DESIGN AND METHODS

2.1 Survey Study of Michigan Inland Lakes

As a grantee of Michigan Department of Environmental Quality (MDEQ), our main objective is to develop a predictive model. We address the challenge by assessing land use information and collected cyanotoxin data and investigate statistical relationship that potentially drives HABs. For the summer of 2017, a total of 29 inland lakes were sampled. Prior to sampling, permission of riparian owner was obtained for lakes which did not have public access. Sample surveying began in the month of June 2017 until October 2017. Each month, every lake one was sampled once. Sampling locations were chosen from known list of lakes reported with HABs given by Aaron Parker from MDEQ and existing collaborative partners with different lake associations. In addition, we also chose lakes that are reasonably close to I-75 expressway for the ease of transportation. See figure **??** for a map of our lakes sites and for more detail, see table **??**.

Cyanotoxins were analyzed primarily by LC-MS/MS. For MC, 12 different congeners along with nodularin were identified and quantified by LC-MS/MS. Total MC is the sum of the 12 congeners and nodularin analyzed by LC-MS/MS measured in the sample. In parallel, ELISA was used to quantify total MC and nodularins. we compared the results of total MC from both LC-MS/MS and ELISA and assessed the agreement between the two. Cylindrospermopsin and Anatoxin-a were also analyzed with LC-MS/MS. In addition, we also used QPCR which detects and quantifies *16s rRNA*, *mcyE/nadF*, *sxtA*, and *cyrA* gene. The *16s rRNA* gene is a

common gene in QPCR to uniquely identify bacterial phylogeny and taxonomy [**?**]. The *mcyE/nadF*, *sxtA* and *cyrA* are gene clusters responsible of synthesizing MC, nodularin, saxitoxin-a and cylindrospermonsin. A commercial kit from Phytoxigene Inc. [1] is used for QPCR which contains the neccessary primers and probes.

At the beginning of our survey, a constructed sampler float was installed at each lake. The sampler was constructed by the Dr. Raffel's team, which consisted of 3 plexi-glass plates and placed at each sampling location for the purpose of collecting zebra mussels (see figure **??**). The stack of three square plexiglass sheets were about 15cm,20cm, and 25cm in diameter which gives them a total surface area of 0.23 m^2 per sampler. In addition, also installed a slotted PVC pipe which is used for SPATT, a beta test of a new method of monitoring toxins. A majority of samplers were installed on the riparian owner's dock, or as a float. HOBO™pendant temperature and light logger were installed on floats at each lake site. In October, we collected the samplers and scraped all mussels into a glass mason jar for analysis of biomass.

We analyzed different water parameters that could lead us to understand key drivers of HABs. Upon arrival at each lake pH, conductance and dissolved oxygen. A portable nephelometer and flurometer was used to measure turbidity, chloraphyl-a and phycocyanin. Dissolved nitrate+nitrite, orthophosphate, ammonia, total phosphorus and nitrogen were analyzed colorimetrically on the collected water samples. Land use for each lake's watershed was calculated using Geographic Information System (GIS) software. Data was compiled for statistical analysis.

[1] Phytoxigene Inc. Diagnostic Technology, 7 Narabang Way, Belrose 2085. Australia.

Table 2.1: Geographic information of sampling points at each surveyed lakes

Name of Lake	Shorten Code	County	Longitude	Latitude	HUC 14 Reachcode
Bear Lake	BEA	Kalkaska	-84.9438079727	44.7286139551	04060103001048
Belleville Lake	BEL	Wayne	-83.4663770506	42.2145253455	04090005001822
Bogie Lake	BOG	Oakland	-83.5054334514	42.6188513679	04090005001348
Brighton Lake	BRI	Livingston	-83.7958137995	42.5169054061	04090005001500
Coldwater Lake	COL	Isabella	-84.9565922285	43.6613607551	04080202000902
Deer Lake	DEE	Charlevoix	-84.9770123186	45.166441811	04060105001116
Ford Lake	FOR	Washtenaw	-83.5849122567	42.2159133043	04090005001823
Houghton Lake	HOU	Roscommon	-84.7262816343	44.3385407778	04060102002461
Hudson Lake	HUD	Lenawee	-84.2545514803	41.835000535	04100002001317
Intermediate lake	INT	Antrim	-85.2293359783	45.0265435299	04060105003435
Lake Cadillac	CAD	Wexford	-85.4266252378	44.2410192547	04060102001951
Lake Margrethe	MAR	Crawford	-84.7830175986	44.6464747348	04060103001058
Lake Nepessing	NEP	Lapeer	-83.3728265865	43.0161554865	04080204001601
Lime Lake	LIM	Hillsdale	-84.3791188315	41.7861576065	04100006000872
Little Glen Lake	LGL	Leelanac	-85.963633169	44.8687577197	04060104000456
Little Round Lake	LRO	Lenawee	-84.3527742524	41.9093334799	04100006000858
Manitou Lake	MAN	Shiawassee	-84.2038069227	42.925537136	04050005000939
Ore Lake	ORE	Livingston	-83.7959940227	42.4805569493	04090005001574
Paradise Lake	PAR	Emmett	-84.7512093045	45.6872890124	04060105001063
Platte Lake	PLA	Benzie	-86.092789204	44.6900468421	04060104000558
Pontiac Lake	PON	Oakland	-83.451096479	42.6664394508	04090005001288
Posey lake	POS	Lenawee	-84.3007962072	41.8970465491	04100006000857
Round Lake	ROU	Lenawee	-84.1318219224	42.0712488438	04100002001130
Sanford Lake	SAN	Midland	-84.3860517762	43.7104273774	04080201001468
Silver Lake	SIL	Grand Traverse	-85.687150728	44.6980286859	04060105003542
Stony Creek Lake	STO	Oakland	-83.0870627175	42.7260717429	04090003001029
Sugden Lake	SUG	Oakland	-83.4972563639	42.6173106359	04090005001347
West Twin Lake	WTL	Montmorency	-84.3501403918	44.8762035424	04070007001271
Wixom Lake	WIX	Gladwin	-84.3537506311	43.8276751177	04080201001442

HUC=Hydrological Unit Code

GPS Coordinates are in decimal degrees, North American Datum of 1983 (NAD83)

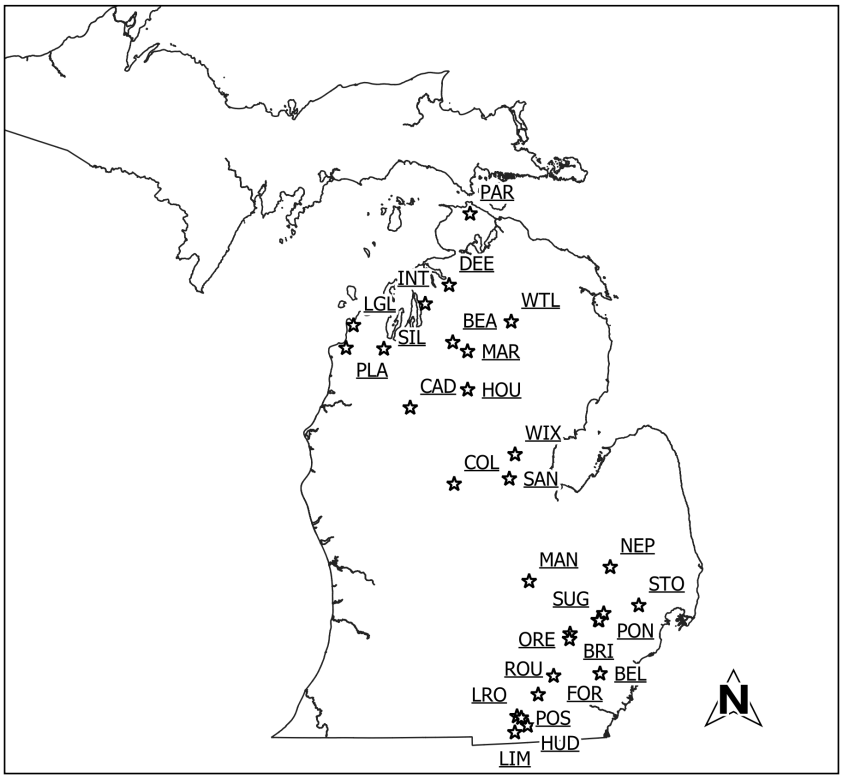

Figure 2.1: Map of Sampled Lake Sites in Michigan

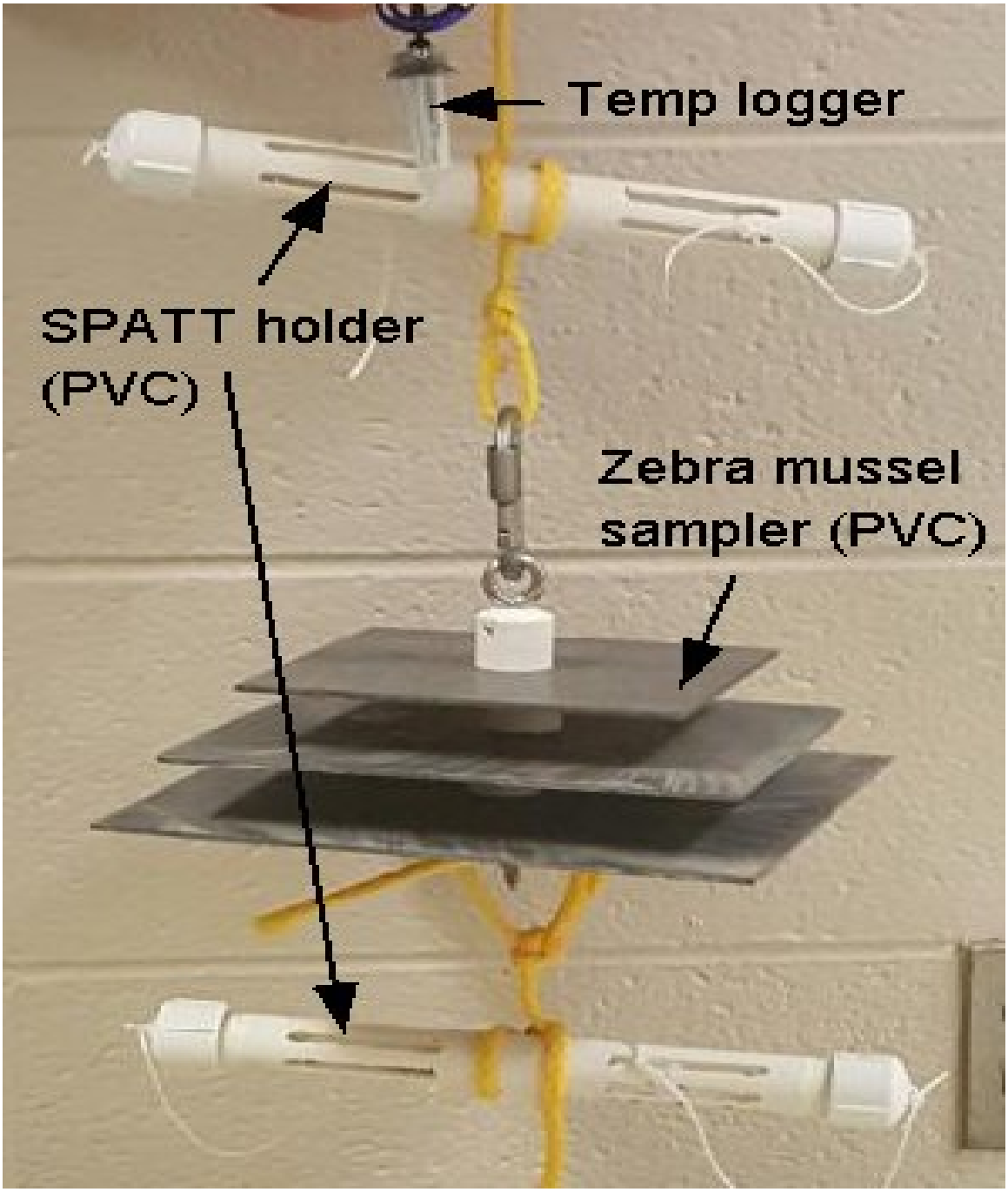

Figure 2.2: Picture of the constructed sampler installed at each lake. HOBO Pendant is attached with a secure carabiner. SPATT PVC holders and Zebra mussel sampler is shown

2.1.1 Water Sampling

Water samples were collected by wading in toward the center of the lake until water height reached waist height. All water grab samples were taken roughly one foot below water surface. Each water collecting vessel was rinsed 3 times with the lake water before obtaining final sample. A total of 4 team of field surveyors sampled each with designated lakes. At each lake, a hand-held multi-meter was used to measure pH, conductivity (μS/cm), dissolved oxygen (mg/L) and temperature (°C). Phycocyanin and chlorophyll fluorescence were also measured using an portable fluorometer by Amiscience with the optical excitation of 470 nm and 590 nm and the emission is read at 685 nm, which is measured in Relative Fluorescence Units (RFU). Each fluorometer for each surveyor was calibrated against rhodamine WT[2] dye as a secondary calibration standard, which ensured the calibration is relative to other fluorometer and to prevent drift. A portable meter from Hach was used to measure turbidity in Nephelometric Turbidity Unit (NTU). Formazin standards were used to calibrate the turbidity meter.

Water sampling kits were prepared by storing pre-labeled water vessels in zip-lock bags for each lake to prevent cross-contamination between different lake water samples during sampling transport and storage. Each sampling kit contained 60mL Polyetheylene Terephthalate (PETG) vials for MC analysis, 100mL sterile IDEXX bottles for QPCR analysis, 50mL polypropylene centrifuge vials and 250 mL High-Density Polyetheylene (HPDE) Nalgene bottles for nutrient analysis. Each kit also provided alkaline Lugol's iodine solution for preserving cyanobacteria samples for identification and 3M H_2SO_4 for acid preservation of nutrient samples. The 3M H_2SO_4 is in a separate zip-lock bag with roughly 20g of $NaCO_3$ wrapped in paper towel to neutralize the sulfuric acid incase of a spill or a leak. Lugol's iodine

[2]Sodium chloride 4-[3,6-bis(diethylamino)-9-xantheniumyl]isophthalate (2:1:1)

solution was prepared by dissolving 100g of KI, 100g of I_2 and 100g of sodium acetate dissolved in 1 liter of water. See figure **??** for an example of a sampling kit.

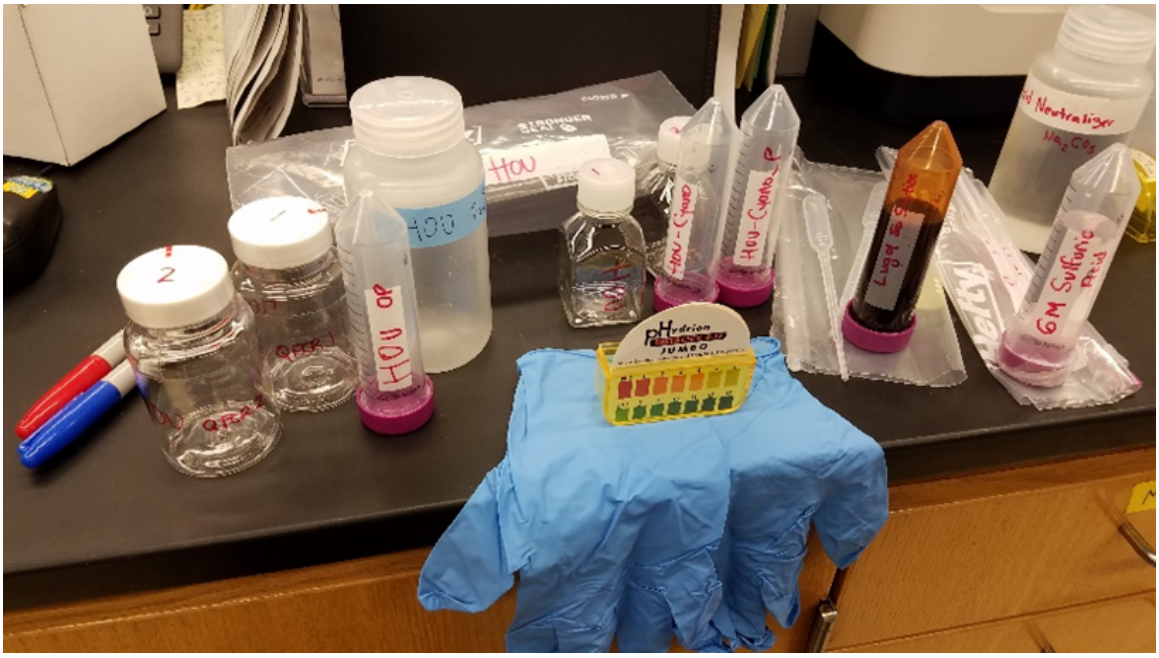

Figure 2.3: An example of sampling kit and the contents

2.2 Analytical Methods

2.2.1 Liquid Chromatography Mass Spectrometry

2.2.1.0.1 Grab Samples Water samples were collected in 60mL PETG vials upon arrival at each lake. Within 3 days from sampling, the water samples were freeze-thawed for 3 cycles for cell lysis. Water samples are thawed slowly in a heated water bath at 37°C, then frozen at -20°C. Once finally thawed, an AcroPrep 96-well plate with glass fiber was used to filter the water samples. Once filtered, a 3.5 mL aliquot of each sample was transferred to glass vials suitable for the Thermo Scientific EQuan MAX (online sample concentrator). The samples were transported to Wayne State University and analyzed for 12 MC congeners, nodularin, anatoxin-a and cylindrospermopsin. The Westrick group at the WSU Lumigen Instrument Center has developed a high-throughput LC-MS/MS analysis for MC in surface and drinking water analyses. The analysis done by the Westrick group's LC-MS/MS platform includes a Thermo Scientific EQuan MAX (online sample concentrator) and ThermoFisher's UltiMate 3000 Ultra-High Performance Liquid Chromatography (UHPLC) system and a Triple-Stage Quadrupole (TSQ) Quantiva. Their method is similar to EPA method 544 with the addition of 5 more congener analytes [**?**]. Figure **??** shows a standard chromatogram of all 12 MC, nodularin, and the ethylated internal standard ($[C_2D_5]$ MC-LR) eluting between 2.2 and 5.2 minutes allowing for the total analyses time to be less than 12 minutes. The Minimum Detection Limit (MDL) is 0.030 μg/L for all cyantoxin analysis.

2.2.1.0.2 Solid Phase Adsorption Toxin Tracking The SPATT bag is constructed with Nitex[3] and filled with Diaion™ HP-20[4], a non-polar resin

[3]Purchased from Dynamic Aqua-Supply Ltd. `http://www.dynamicaqua.com/nitex.html`
[4]Purchased from Sigma Aldrich: CAS Number 9052-95-3

(styrene-divinylbenzene copolymer), To construct the SPATT bags, a 1 meter x 5 centimeter strip of Nitex mesh were cut with a sharp blade. The Nitex strip was sewn by folding half length-wise (or *hot dog* style). With tape holding the fold, the end of the strip was sewn 0.5cm from the edge. Stitching design was tight to ensure no leakage of polymer beads. Approximately 9-10cm of sewn Nitex strips were cut and zip-tied about 0.5cm at one end. With one end open, 3.00-3.01 grams of HP-20 resin was carefully filled using a funnel. The open end is zip-tied once the Nitex bag is full. Prior to deployment,the filled SPATT bags were activated by soaking in 100% methanol for 24 hours under 4°C. Next the SPATTs were rinsed with Milli-Q water and then soaked for 24 hours in Milli-Q water under 4°C before deploying the SPATT bag in our target sample lakes.

At each lake site, two SPATT bags were loaded into the slotted PVC pipe on the constructed float. The SPATT bags are left for about a month at each lake. When SPATT are retrieved, they are carefully removed and rinsed with Milli-Q water upon arrival and stored in a 15mL centrifuge vial with a plastic spacer on the bottom. SPATT are stored at 4°C during transport back to the lab. The SPATT are centrifuged at 8000rpm. The spacer allows liquid to pool on the bottom when centrifuged. When centrifuged, the SPATT bags are cut open and the resin is poured into a 50mL centrifuge tube. Milli-Q water is used to rinse the SPATT bags to effectively transfer all the resin. About 30mL of Milli-Q water is used. The solution is allowed to rest so the resin settles to the bottom. Using a pipet, the water is carefully decanted until the total volume is 5mL. A solution of 80% methanol with 10μM ammonium formate is added to the tube until the total volume is 45mL. The solution is gently mixed and then allowed to settle for 30 minutes. A 3.5mL aliqout of the supernatant is transferred to glass vials and analyzed by LC-MS/MS by the Westerick group. Similiar to our analysis of the

grab sample, the SPATT were analyzed for all 12 congeners of MC and nodularin. The final reported value is calculated to give the amount of MC per gram of resin per day. It is calculated by this equation:

$$(\frac{\text{ng of MC}}{\text{g of resin per day}}) = (\frac{\mu\text{g of MC}}{L}) \times (\frac{1000\text{ng}}{\mu\text{g}}) \times 0.045L \times \frac{1}{\text{3g of resin}} \times \frac{1}{\text{days deployed}}$$

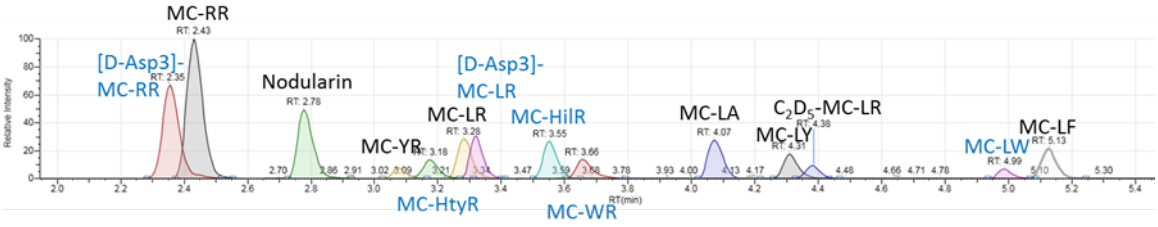

Figure 2.4: Liquid chromotography-mass spectrometry chromatogram of the MC congeners. Chromatogram provided by Westrick Group

2.2.2 Enzyme-Linked Immunosorbant Assay

A commercial Microcystin/Nodularins ADDA ELISA kit was used from Abraxis to analyze total microcystin[5]. The analysis uses a polyclonal antibody which specifically binds to the ADDA moiety found in MC. However it does not distinguish between different congeners and the given results are in terms of MC-LR equivalance. The analysis followed the reccommended guidelines provided by the EPA [**?**]. In preperation of loading the plate, 100μL of standards, controls, blanks and samples are aliquoted into a seperate sterile 96-well plate. To minimize assay drift caused by slow plate loading, a multi-channel pipettor was used to load the the standards, controls, blanks and samples to the final 96-well plate. The assay procedures were carried out and read by Synergy H1 microplate reader from Biotek.

2.2.3 Quantitative Polymerase Chain Reaction

Phytoxigene™CyanoDTec cyanobacteria and toxin test kit was preformed with Applied Biosystem StepOnePlus™QPCR. The kit provides two separate assay mixes. Total cyanobacteria assay will quantify the 16srRNA gene copies found in the water sample. Both the total cyanobacteria and toxin gene assay were analyzed in parallel for each month of grab samples. The primer/probe sequence is unknown. The PCR reaction mix contained 5 μL of template/sample extracts and 20 μL of rehydrated mastermix. Each sample were run in singlicate due to limited resources. Positive standards for target genes were run on each PCR analysis. Phytoxigene™CyanoNAS nucleic standards were used to generate standard curves for quantification of gene copies. The CyanoNAS was removed from -20°C and allowed to thaw prior to analysis. Standards were run in duplicates.

[5]https://www.abraxiskits.com/products/algal-toxins/

Samples for QPCR were filtered either on site with portable Santino pump, or brought back to the lab for filtration within 8 hours from sampling. At each lake, 100mL or more of water sample was collected in a sterile IDEXX vessel then filtered through a 0.4μm pore size polycarbonate membrane and stored at -20°C until QPCR. Once filtered, they are immediately transferred into BioGX vials. BioGX vials are stored at -80°C until analysis. BioGx vials contains 500 uL of lysis buffer, lysis beads and filtrate. For cell lysis, vials were vigorously shaken by bead beater on the highest setting for 2 minutes. After bead beaten, sample vials were centrifuged for 1 min. After centrifuge, 50 μL of the supernatant was transferred to a microcentrifuge tube and centrifuged for 5 min, then roughly 25 μL of the final supernatant to another set of microcentrifuge tubes for PCR template. Sample extracts are stored at 4°C and analyzed within 4 hours.

From following the recommended guide by Phytoxigene, the PCR heat cycles were programmed with initial denaturing step at 95°C for 2 min, then a repeating of 95°C for 15 seconds and 60°C for 30 seconds reaching a total of 40 cycles. The appropriate gene target filters were manually set to match the emission spectra of each probe. For each PCR run, a standard curve was generated from within the StepOnePlus software. CT threshold and baseline were manually assigned for each run by visually assessing each target run. The calculated gene copies are done automatically by the StepOnePlus software, expressed in gene copies/μL of lysate. The final reportable value is calculated by this equation:

$$\frac{Genecopies}{mL} = (\frac{Genecopies}{\mu L\ of\ lysate}) \times (\frac{500 \mu L\ of\ lysate}{\text{mL of Sample Volume}})$$

2.2.4 Nutrients

Two 125-mL HPDE Nalgene bottles were used to collect acid-preserved water samples with 2 mL of 3M H_2SO_4, resulting to pH <2. A 50-mL centrifuge tube is

used to collect water samples without acid preservation for orthophosphate. One of the two Nalgene bottles is allocated for ammonia-N and nitrate+nitrite-N by our lab at Oakland University, and the other is for total phosphorus and total nitrogen run by Ben Southwell and his team at Lake Superior State University. Samples were kept at 4 °C during transport and stored at -20°C. Upon receiving samples from field samplers samples are thawed if frozen and cool at 4°C. All lake water samples were homogenized by inverting 8 times and aliquoted into 15-mL centrifuge vials and centrifuged at 3000rpm for 45 seconds. The supernatant was collected into a clean 3-mL vial to be prepared for the AQ1 auto sampler. All samples were analyzed within the appropriate time frame from time of sampling collection.

Nutrient concentrations are quantified by colorimetric analysis with AQ1 from SEAL Analytical[6]. Ammonia-N (NH_3) is quantified by a reaction with dichloroisocyanurate and dissolve ammonia to create chloramines which forms a blue-green color with salicylate which is measured at 660 nm [**?**]. The range of application is between 0.02-1.0 mg N/L with a mininum detection limit of 0.006 mg N/L for quantifying ammonia. Nitrate+nitrite-N ($NO_3^- + NO_2^-$) is analyzed with an open tube copperized cadmium coil which the pH buffered sample water will have nitrate reduced to nitrite. The reduced water sample is then reacts with sulfanilamide with the presence of *N*-(1-napthyl)-ethylenediamine dihydrochloride to form a reddish color measured at 520 nm [**?**]. The range of application for analyzing nitrate+nitrite is between 0.25-15 mg N/L with a detection limit of 0.04 mg N/L. Orthophosphate-P (PO_4^{-3}) is analyzed with acidic molybdate solution with antimony potassium tartrate to form a complex with dissolved orthophosphate. The complex is reduced with ascorbic acid to create a blue color measured at 880 nm [**?**]. The range for orthosphosphate is between 0.003-0.3 mg P/L with 0.008 mg P/L as

[6]SEAL Analytical Inc. 6501 West Donges Bay Road Mequon, Wisconsin 53092

the detection limit. Total Kjeldahl nitrogen-N (organic nitrogen) is analyzed by sample digestion with copper(II) catalyst at 380°C. Nitrogen containing compounds such as amino acids and peptides are converted to ammonia which is then reacted with hypochorite to create chloramine, which is then reacted with salicylate at a pH of 12.6 with the presence of nitroferricyanide to form a green-blue color measured at 670 nm [**?**]. The range for total Kjeldahl nitrogen is between 0.2 to 4.0 mg N/L with 0.07 mg N/L as the detection limit. Total phosphorus (polyphosphates and some organic phosphorus) is analyzed by acid-persulfate digestion which water sample with ammonium persulfate and sulfuric acid is autoclaved at 121°C for 30 minutes which organic phosphorus is converted to orthophosphate. After digestion, orthophosphate is reacted with acidic molybdate which is reduced by ascorbic acid to create a blue color measured at 880 nm [**?**]. The range for total phosphorus is between 0.01-1.0 mg P/L with 0.02 mg P/L as the detection limit.

2.2.5 Geographic Information System Analysis

Watershed delineation and calculation of land use were done using Quantum Geographic Information System (QGIS) [**?**]. Elevation data was downloaded in bulk by an FTP client as mosaic raster files for the state of Michigan downloaded from USGS [7]. Elevation data prepared by using $r.fill.dir$ function from Geographic Resources Analysis Support System (GRASS) which fills sinks or depressions [**?**]. A flow accumulation raster map is generated from this command. The value of each cell designates the amount of flow based on drainage characteristics the elevation data. Visually viewing the histogram of the distribution of flow accumulation values, selecting the highest values displays will display the most probable areas the flow of water will be. The pour point is where the lake's outlet, where the water is

[7] `https://earthexplorer.usgs.gov/`

most likely to leave. This provided visual aid in selecting the pour point of each lake. A new shapefile was created and selected each lake's pour with the visual aid of stream flow lines. Using the *r.distance* function from GRASS, it snapped each pour point to the proper place to help the delineation step. Each lake's watershed was delineated using *r.drain* to create a elevation model map derived from the flow accumulation raster file. The drainage raster file is then used as an input for function *r.water.outlet* along with the coordinates of fixed pour point location, which gives the shape of each lake's watershed extent.

Land use data was downloaded from the 2006 National Land Cover Database [**?**]. The land use data were classified at Anderson level-II, which has 20 different classification of land distinguishing different biomes and regions. To simplify the land use data, the raster is reclassified into 8 Anderson level-I categories using *r.recode* tool from GRASS. The 8 reclassified Anderson level-I classes with band ID are water (11, 12), developed (21,22,23,24), barren, (31,32,33), shrubs (41,42,43), forest(52), agriculture (71), herbaceous (81,82) and wetlands (90,95). The land use raster file was transformed into a vectorized shapefile. The shapefile was merged by union (or dissolved) by each lake's watershed, which resulted area of each land use class within each lake's watershed. This data was exported as a .csv file and prepared for statistical analysis.

Precipitation data were retrieved from the Global Historical Climatology Network (GHCN) database from NOAA [**?**]. Daily precipitation data was downloaded from NOAA's FTP server [8]. The geolocation of each rain gauge station were imported into QGIS and mapped. The distribution of the rain gauges were not uniformally distributed. Thiessen/Voronoi polygons for each station were generated and overlayed on each watershed. The area of each thiessen/voronoi polygon's

[8] `ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/`

intersection with the corresponding catchment is divided by the area of the lake's watershed to give a weighted value. The mean areal precipitation for each lake's watershed is calculated by taking each station's measurements and multiplying by the weighted value, then averaged together. Ambient air temperature for each watershed is simply averaged together with their intersection of the lake's watershed. Precipitation data with each sampled lake is joined by lakes watershed. Averaged 3, 5, 7, and 30 days lagged precipitation and ambient air temperature were calculated for our analysis.

2.2.6 Statistical Analysis

Each analytical measurement was compiled and organized by each sampling event. We have data sampled from Lake Superior, Lake St. Clair and Lake Erie, however with my discussions with Dr. Szlag and Dr. Raffel, we decided to exclude them in our analysis. Their unique geology and lake morphology does not fit our focus on inland lakes. Data manipulation and analysis was done in Program R, a statistical computing language [**?**]. The "dplyr" package was primarily used for data cleaning, compiling and preparation to have our dataset ready for statistical analysis [**?**]. We also used other packages with program R to for additional tools to rearrange our data matrix [**?**], display our graphs[**?**, **?**, **?**, **?**] and create statistical summary tables [**?**, **?**, **?**, **?**, **?**] for exploratory analysis.

The requirements for building our model using linear regression assumes the distribution of explanatory and response variables to follow a normal distribution [**?**]. The compiled dataset contained in total of 115 observations from the 29 inland lakes. From our collected dataset, we assessed each variable's distribution and $log10$-transformed to fit a normal distribution. In order to solve the problem of data values that are zero, we added the corresponding minimum detection limit first,

then applied a log transformation. See table **??** for details of which variable was transformed and the shorten variable name.

For selecting the best predictor variables, a best subset linear regression analysis was used to find good predictors that can potentially explain our response variables using the "leaps" package from R [**?**]. Measurements from each lake is a factor that may contribute as a random effect. This can be an issue where measurements from each lake is pseudo-replicated [**?**]. Best subset and correlation matrix analysis is done on an averaged dataset based on each lake which works around this issue. With the best variables from the regression subset, backward step-wise regression will be preformed to further refine the best fit model. Variables will be backwardly selected by F-test using simple linear regression [**?**]. Finnaly, a linear mixed effect analysis is used to verify our best models as it allows to account the variance of each sample site without taking the average of each lake site. The predictor variables are set as fixed effects and lake site as random effects with varying intercepts. A visual inspection of the residual plots is done to check if the models deviate from homoscedasticity. The linear mixed effect models were built on the full dataset as this accounts for the variance of each of our lake site[**?**]. The library package "lme4" is used for our linear mixed modeling [**?**]. Each non-nested models are rank by the lowest Bayesian Information Criterion (BIC) being our best model.

Table 2.2: Table Summary

Measured Variable (Units)	**Shortened Code Name**	**Transformation**
Total microcysin of all 12 congeners (μg/L)	SUM	$log10$(SUM+0.03)
Cyano *16s rRNA* gene copies (cp/L)	X16SRNA	$log10$(X16SRNA+45)
mcyE gene copies (cp/L)	MCYE	$log10$(MCYE+45)
Ortho-P (mg-P/L)	OP	$log10$(OP+0.003)
Nitrate/Nitrite (mg-N/L)	NO3	$log10$(NO3+0.04)
Ammonia (mg-N/L)	NH3	$log10$(NH3+0.006)
Total nitrogen (mg-N/L)	TN	$log10$(TN+0.116)
Total Kjeldahl nitrogen (mg-N/L)	TKN	$log10$(TKN+0.07)
Total phosphorus (mg-P/L)	TP	$log10$(TP+0.002)
Total nitrogen to total phosphorus ratio	TNTP	None

Continued on next page

Table 2.2 – continued from previous page

Measured Variable (Units)	Shortened Code Name	Transformation
Measured pH of Lake	pH	None
Dissolved oxygen (mg/L)	DO	$log10$(do+0.01)
Conductance (uS/cm)	conduc	$log10$(conduc+0.01)
Turbidity (NTU)	turb	$log10$(turb+0.01)
Chloraphyll-a (RFU)	chloro	$log10$(chloro+0.01)
Phycocyanin (RFU)	phyco	$log10$(phyco+0.01)
Maximum depth of lake (meters)	Max_-Depth	None
Lake area (sq Km)	LkArea	$log10$(LkArea+1)
Watershed Area (sq Km)	WtWhArea	$log10$(WtWhArea+1)
Lake area to watershed area ratio	LkWshRatio	$log10$(LkWshRatio+1)
Water Land-Use (%)	Water	None
Developed Land-Use (%)	Developed	None

Continued on next page

Table 2.2 – continued from previous page

Measured Variable (Units)	Shortened Code Name	Transformation
Barren Land-Use (%)	Barren	None
Forest Land-Use (%)	Fores	None
Shrubs Land-Use (%)	Shrubs	None
Herbaceous Land-Use (%)	Herbaceous	None
Agriculture Land-Use (%)	Agriculture	None
Wetlands Land-Use (%)	Wetlands	None
Average precipitation 3 days prior (mm)	precip3	$log10$(precip3+1)
Average precipitation 5 days prior (mm)	precip5	$log10$(precip5+1)
Average precipitation 7 days prior (mm)	precip7	$log10$(precip7+1)

Continued on next page

Table 2.2 – continued from previous page

Measured Variable (Units)	Shortened Code Name	Transformation
Average precipitation 30 days prior (mm)	precip30	$log10$(precip30+1)
Water temperature at time of sampling (Celcius)	wtemp	None
Average temperature 3 days prior from GHCN (Celcius)	temp3	None
Average temperature 5 days prior from GHCN (Celcius)	temp5	None
Average temperature 7 days prior from GHCN (Celcius)	temp7	None
Average temperature 30 days prior from GHCN (Celcius)	temp30	None

Continued on next page

Table 2.2 – continued from previous page

Measured Variable (Units)	Shortened Code Name	Transformation
Average Temperature from Hobo pendant 30 days prior (Celcius)	hobotemp	None
Average light intensity from Hobo pendant days prior (lux)	hobolight	$log10$(hobolight+1)
Zebra mussel Mass (grams)	MusselMass	$log10$(MusselMass+1)
Zebra mussel (counts)	MusselNum	$log10$(MusselNum+1)