# 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

TABLE OF CONTENTS—Continued

TABLE OF CONTENTS—Continued

LIST OF TABLES

LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ADDA | 3-amino-9-methoxy-10-phenyl-2,6,8-trimethyl-deca-4,6-dienoic acid |
| BIC | Bayesian Information Criterion |
| D-MeAsp | D-erythro$\beta$-methylaspatric acid |
| USEPA | United States Environmental Protection Agency |
| GHCN | Global Historical Climatology Network |
| GRASS | Geographic Resources Analysis Support System |
| HABs | Harmful Algal Blooms |
| HPDE | High-Density Polyetheylene |
| LC-MS/MS | Liquid Chromatography/Tandem Mass Spectrometry |
| MC | Microcystin |
| MC-LA | Microcystin-LA |
| MC-LR | Microcystin-LF |
| MC-LR | Microcystin-LR |
| MC-LR | Microcystin-YR |
| MC-RR | Microcystin-LY |
| MC-RR | Microcystin-RR |
| MDEQ | Michigan Department of Environmental Quality |
| Mdha | $N$-methyldehydro-alanine |
| MDL | Minimum Detection Limit |
| NOAA | National Oceanic and Atomospheric Administration |

| | |
|---|---|
| NRPS | Nonribosomal peptide synthetase |
| NTU | Nephelometric Turbidity Unit |
| ODK | Open Data Kit |
| PETG | Polyetheylene Terephthalate |
| PKS | Polyketide synthase |
| QGIS | Quantum Geographic Information System |
| QPCR | Quantitative Polymerase Chain Reaction |
| RFU | Relative Fluorescence Units |
| RSA | Rivest-Shamir-Adleman |
| SPATT | Solid Phase Adsorbtion Toxin Tracking |
| TSQ | Triple-Stage Quadrupole |
| UHPLC | Ultra-High Performance Liquid Chromatography |
| WHO | World Health Organization |

# 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 [1]. 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 [2]. In most cases, HABs are found in coastal regions, streams, and freshwater lakes [3]. 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 [4]. 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 [5]. 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 [6]. 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) for potential future regulations [7].

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 [1]. MC is produced mostly within the *Microcystis* genera but also by other generas such as *Anabaenopsis, Nostoc,* and *Planktrothrix* [3, 8]. For MC, the World Health Organization (WHO) have a guidance level of 1 $\mu$g/L for drinking water and 20 $\mu$g/L for recreational water [9, 10]. The USEPA recently announced a new drinking water Health Advisory (HA) for MC with a guidance level of 4 $\mu$g/L for recreational surface water [11]. 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 [12] 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.

Cylindrospermonsin is an alkaloid produced by *Cylindrospermopsis, Anabaena, Oscillatoria* and other species of cyanobacteria [4]. [11].

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

Another dangerous group are called saxitoxins and are sodium channel blockers, a potent neurotoxin which paralyze and lead to death by respiratory failure [1]. 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 [2, 1].

## 1.2 Synthesis and Structure of Microcystin

Microcystins are uniquely synthesized in cyanobacteria by a mix of two systems, Polyketide synthase (PKS) and Nonribosomal peptide synthetase (NRPS) [13]. 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 [14, 15]. The major amino residues of MC are of D-erythro$\beta$-methylaspatric acid (D-MeAsp), ADDA, $N$-methyldehydro-alanine (Mdha) and other possible variable amino acids [16, 15].

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) [17]. The most frequently occurring and potent congener variant of MC is MC-LR [3]. Figure 1.2 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 [4]

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 [18]. 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. [19]. The layer of scum formed from HABs also block light for submerged aquatic macrophytes, killing them by preventing photosythesis [20]. HABs can also be a major nuisance and a cost to the community, as this create odor and limit recreation [21, 22].

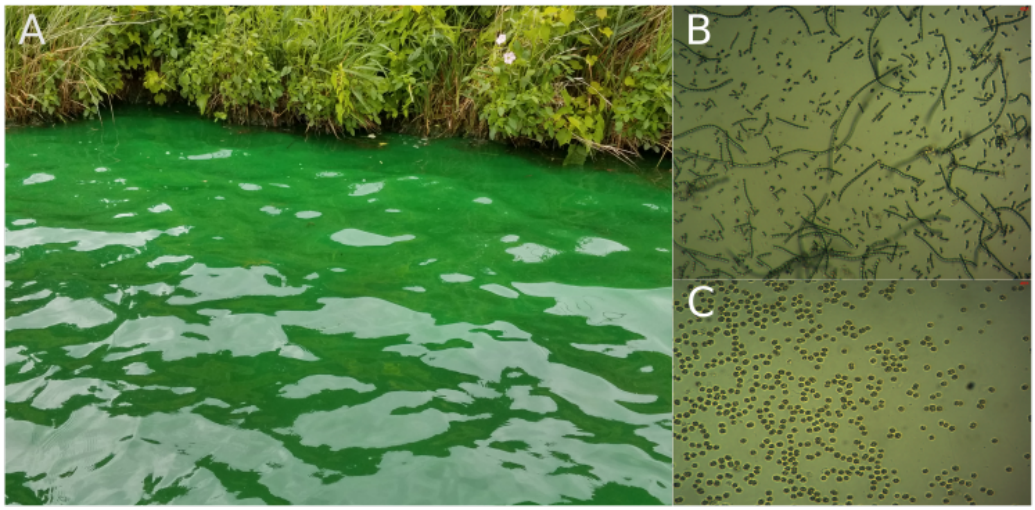


Figure 1.1: Examples of Harmful Algal Blooms. (A): A bloom in Manitou Lake. Picture taken by Fred Farcus. (B): Anabaena (C): Microcystis

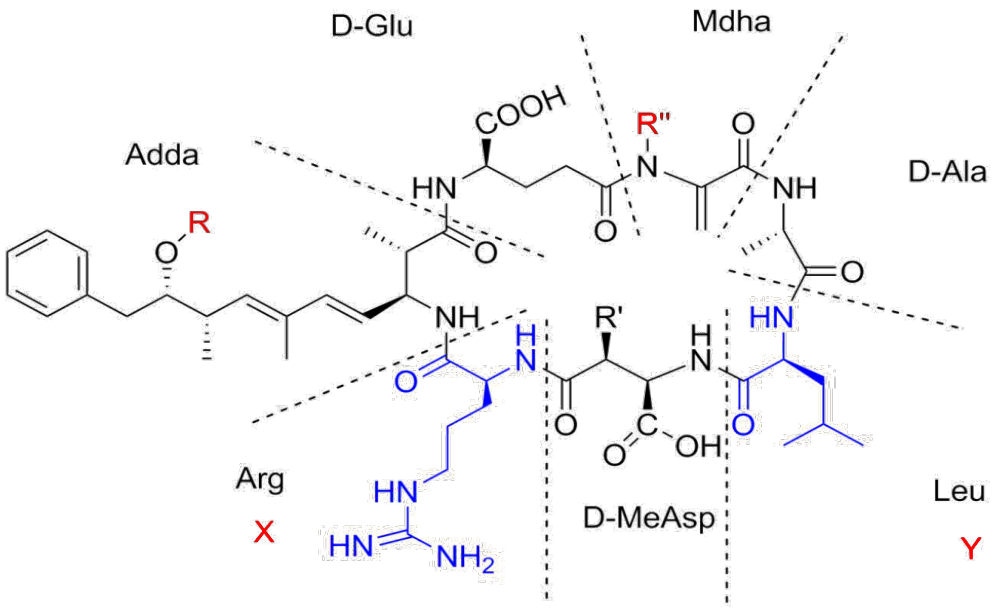


Figure 1.2: Structure of MC-LR

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.

## 1.3 Fate and Transport of Cyanotoxins

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 [19]. 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 [23, 2]. 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 [1].

Although a rare case, accidental ingestion of cyanotoxin by directly drinking water from an affected lake could lead to acute toxicity or other symptoms [24]. 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 [3]. 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 [25]. 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 [26]. Different treatment can be used with ozone, chlorination, activated carbon and advance oxidation process [27, 28]. However ozone and chlorination carries risk of cell lysis and disinfection by-products. The use of activated carbon or membrane filteration is effective but costly to implement as a routine [27]. Removal of toxins can be treated by oxidants such as $KMnO_4$ which oxidizes the dissolved toxins and prevent the cells

to lyse [28]. An effective health risk managment plan would implement a multipoint check system in the water treatment process and also be economically viable [26, 25].

## 1.4 Environmental Drivers of Harmful Algal Blooms

The mechanism for what drives the proliferation is not fully understood [4]. As autotrophs, HABs can rapidly grow under warm and nutrient-rich conditions [3]. 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 [29, 30].

Excess nutrients is long believed to significantly impact the growth of HABs [19, 19]. 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[31]. 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 [19]. 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 [32, 33]. 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 [34, 35, 19, 36].

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 [37]. In a study done by Michigan State University [38] 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 [39, 40, 38]. In our survey, we will investigate whether zebra mussels have an influence on MC concentration or cyanobacteria population.

## 1.5 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 [41]. These warnings to the public can prevent exposure and improve communication with the public. They have provide effective forecast, 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.

## 1.6 Survey Study of Michigan Inland Lakes

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

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

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 Michigan Department of Environmental Quality (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 1.3 for a map of our lakes sites and for more detail, see table A1.

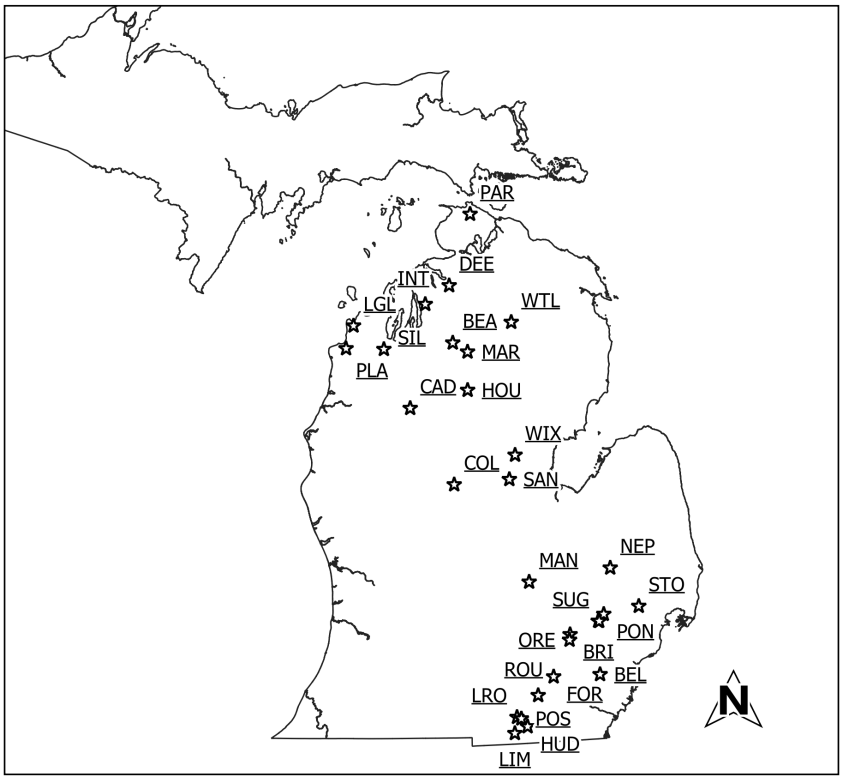


Figure 1.3: Map of Sampled Lake Sites in Michigan

## 1.7 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 chloraphyll-a and MC concentrations as a response variable as its most likely associated with HABs [1, 34, 36, 42, 43]. 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 [44, 19]. 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 defininate 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 hypotheses:

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

## Harmful Algal Blooms and Correlation to Geochemical Features

## 2.1 Introduction

In the pursuit to understand what drives HABs, an extensive survey was conducted on 29 inland lakes. Every month between July to October of 2017, a sample is taken at every lake.

## 2.2 Methods

### 2.2.1 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 ($\mu$S/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[1] 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

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

used to measure turbidity in Nephelometric Turbidity Unit (NTU). Formazin standards were used to calibrate the turbidity meter.

In July 2017, 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 2.1). 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 contains a Solid Phase Adsorbtion Toxin Tracking (SPATT), a beta test of a new method of monitoring toxins. This method will be disscussed in chapter 3. 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.

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 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 2.2 for an example of a sampling kit.

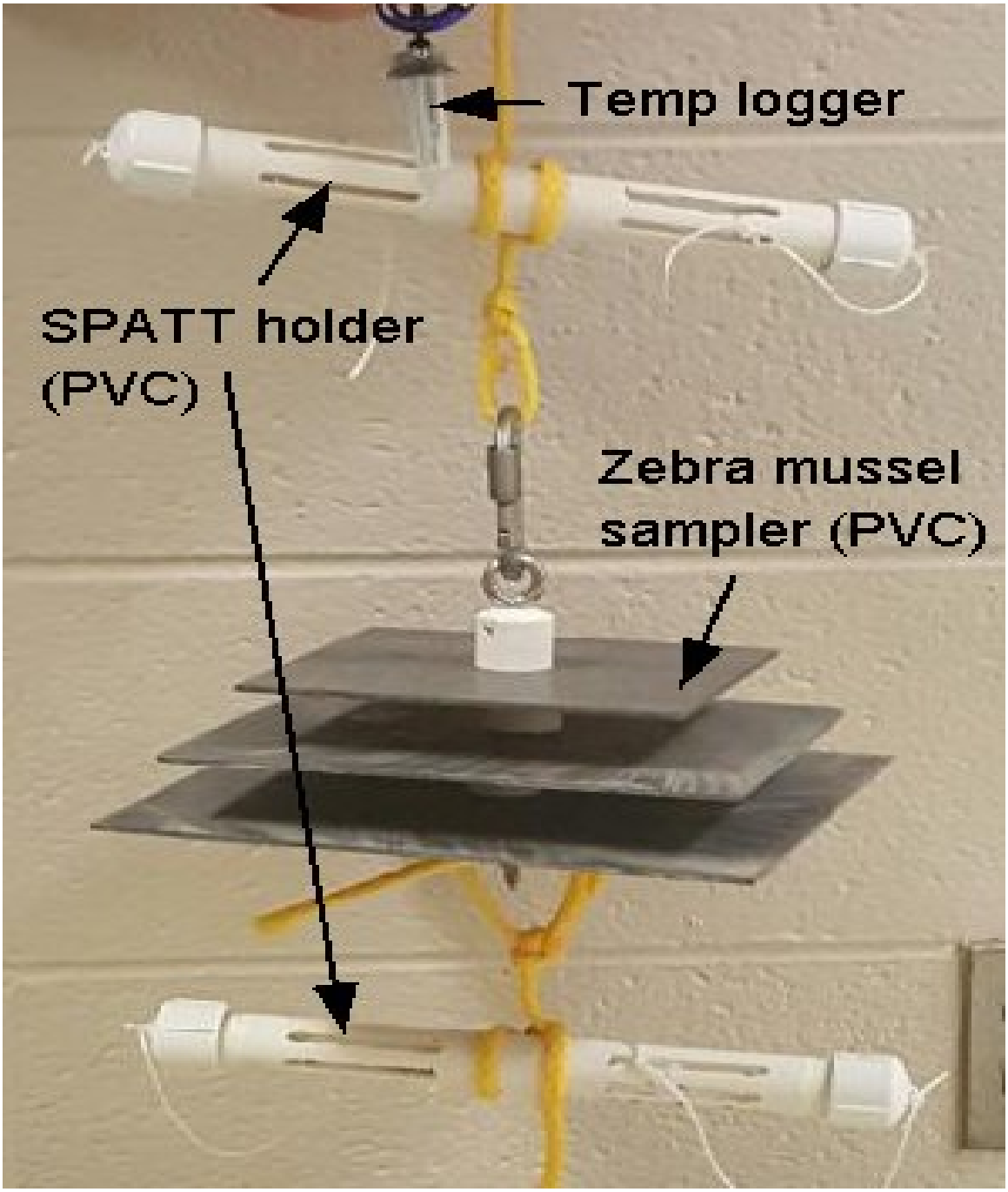


Figure 2.1: Picture of the constructed sampler installed at each lake. HOBO Pendant is attached with a secure carabiner. SPATT PVC holders and Zebra mussel sampler is shown

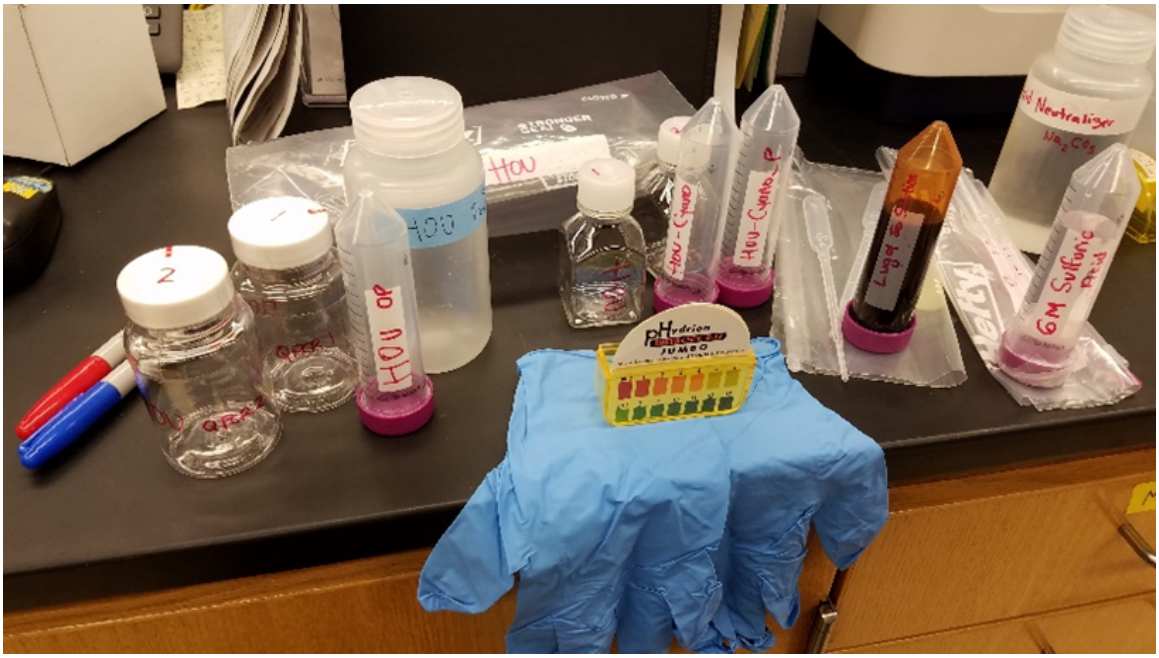


Figure 2.2: An example of sampling kit and the contents

### 2.2.2 Liquid Chromatography Mass Spectrometry

The collected water samples were stored in 60mL PETG vials. 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, a 3.500 ml aliquot of each sample above was transferred to glass vials suitable for the Thermo Scientific EQuan MAX (online sample concentrator). These sample were transported to Wayne State University and analyzed for 12 MC congeners and nodularin. The Westrick group at the WSU Lumigen Instrument Center has developed a high-throughput LC-MS/MS analysis for microcystins 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 [45]. Figure 2.3 shows a standard chromatogram of all 12 microcystins, 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) are 0.030 $\mu$g/L for microcystins.

### 2.2.3 Enzyme-Linked Immunosorbant Assay

A commercial Microcystin/Nodularins ADDA ELISA kit was used from Abraxis to analyze total microcystin[2]. The analysis uses a polyclonal antibody which specifically binds to the ADDA moiety found in MC. However it does not

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

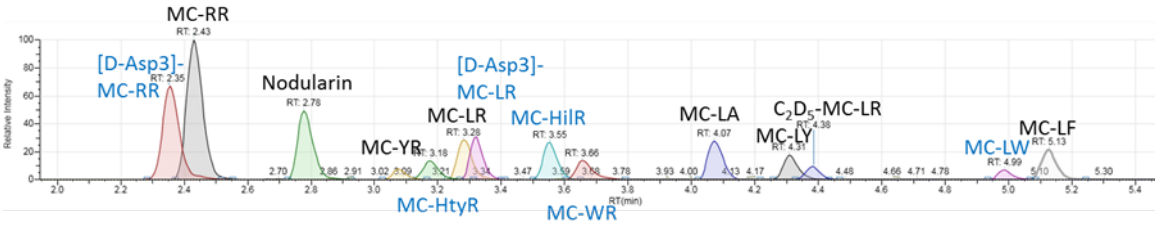


Figure 2.3: Liquid chromotography-mass spectrometry chromatogram of the MC congeners. Chromatogram provided by Westrick Group

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 [46]. In preperation of loading the plate, 100$\mu$L 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.

### 2.2.4 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[3], a discrete colorimetric analyzer. 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 [47]. 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 [48]. 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 [49]. The range for orthosphosphate is between 0.003-0.3 mg P/L with 0.008 mg P/L as 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 [50]. The range for total Kjeldahl nitrogen is between 0.2 to 4.0 mg N/L

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

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 [49]. 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 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 $\mu$L of template/sample extracts and 20 $\mu$L 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.

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 0.4$\mu$m 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 $\mu$L of the supernatant was transferred to a microcentrifuge tube and centrifuged for 5 min, then roughly 25 $\mu$L 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/$\mu$L of lysate. The final reportable value is calculated by this equation:

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

### 2.2.6 Geographic Information System Analysis

Watershed delineation and calculation of land use were done using Quantum Geographic Information System (QGIS) [51]. Elevation data was downloaded in bulk by an FTP client as mosaic raster files for the state of Michigan downloaded from USGS [4]. Elevation data prepared by using $r.fill.dir$ function from Geographic Resources Analysis Support System (GRASS) which fills sinks or depressions [52]. 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

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

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 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 [53]. 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 [54]. Daily precipitation data was downloaded from NOAA's FTP server [5]. The geolocation of each rain gauge station

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

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 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.

### 2.2.7 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 [55]. The "dplyr" package was primarily used for data cleaning, compiling and preparation to have our dataset ready for statistical analysis [56]. We also used other packages with program R to for additional tools to rearrange our data matrix [57], display our graphs[58, 59, 60, 61] and create statistical summary tables [62, 63, 64, 65, 57] 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 [66]. 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 2.1 for a summary 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 [67]. 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 [68]. 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 [69]. 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[70]. The library package "lme4" is used for our linear mixed modeling [66]. Each non-nested models are rank by the lowest Bayesian Information Criterion (BIC) being our best model.

Table 2.1: Table Summary

| **Measured Variable (Units)** | **Shortened Code Name** | **Transformation** |
|---|---|---|
| Total microcysin of all 12 congeners ($\mu$g/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 2.1 – 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 2.1 – 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 2.1 – 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 2.1 – 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) |

## 2.3 Results

### 2.3.1 General Trends

In our sampled lakes for the summer season of 2017, $log10$(MC) concentration and *16s rRNA* were slighly different between each monthly sample (see figure 2.4). We found the MC concentrations to be the highest in the month of August with an average of 1.13 $\mu$g/L compared to 0.182, 0.357 and 0.326 $\mu$g/L for the months of July, September and October. The pattern for the concentration of *16s rRNA* followed similarly, but peaking in September with an average of $6.72 \times 10^5$ cp/L compared to $2.37 \times 10^4$, $5.30 \times 10^5$ and $4.01 \times 10^5$ cp/L for July, August and

October. Water temperature when measured both at the time of sampling and the HOBO loggers we observe a steady decline with each month, however, we did not observe a noticable fluctuation with light intensity (see figure 2.5). With our sampling regime, we did not sample in the spring or early summer months and did not sample frequently to observe a clear temporal trend.

Of all the sampled lakes, we did not detect any cylindrospermonsin in our surveyed lakes. Moreover, QPCR analysis did detect genes responsible for producing cylindrospermonsin (*cyrA*) and saxitoxin (*sxtA*). We detected varying amounts of *mcyE* gene copies in each lake. From the LC-MS/MS we found one instance of anatoxin-a detected at Brighton lake in August 2017 which was sampled during an visible bloom (2.80 $\mu$g/L of Anantoxin-a).

In our analysis of MC, MC-RR, MC-LR and MC-LA were the most detected congeners throughout the summer. MC-RR was found with the highest concentrations 8.55 $\mu$g/L at Brighton Lake in August (see figure 2.6). Majority of our samples were below the EPA guidance level. From our sampled lake sites, Belleville, Brighton, Ford, Hudson, and Wixom Lake had instances of MC concentrations that exceeded the EPA guidance level of 4 $\mu$g/L both from LC-MS/MS and ELISA analysis (see figure 2.7).

### 2.3.2 Comparison of Enzyme-linked Immunosorbent Assay and Liquid Chromatography/Mass-spec

LC-MS/MS and ELISA are compared there are discprenancies between the two, analysis from ELISA tends to be report higher values than LC-MS/MS. The difference is more apparent in the month of September and October, where the difference can be quite large (see figure 2.8). The ELISA generally presents a higher result than LC-MS/MS. This is in part due to the detection limit and reportable

limits being much higher for the ELISA, 0.15 and 0.1 $\mu$g/L. However, our analysis in the month of August had good agreement between LC-MS/MS and ELISA.

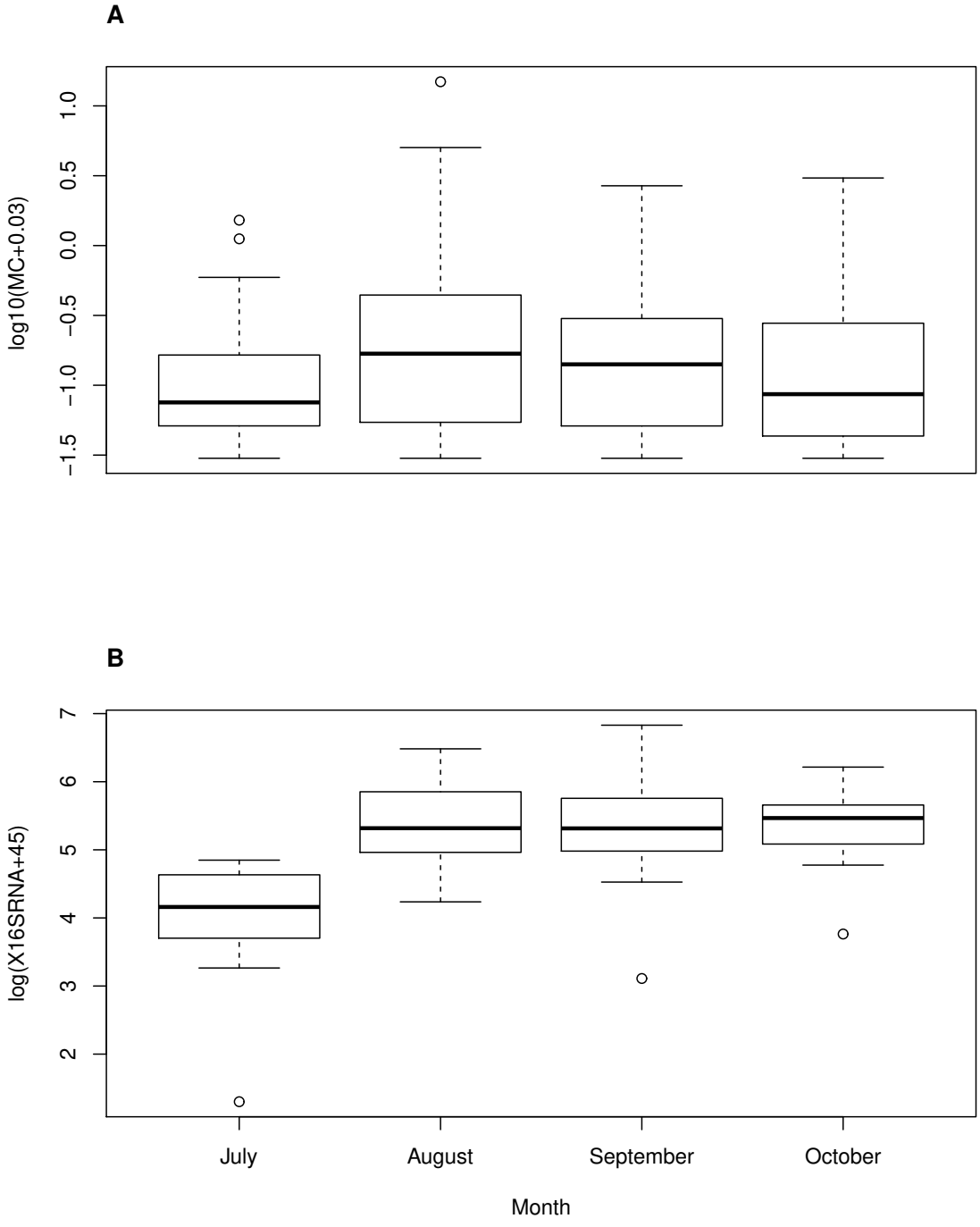


Figure 2.4: (A): Boxplot summary of the average total MC for each month. (B): Boxplot summary of average *16s rRNA* genecopies for each month.

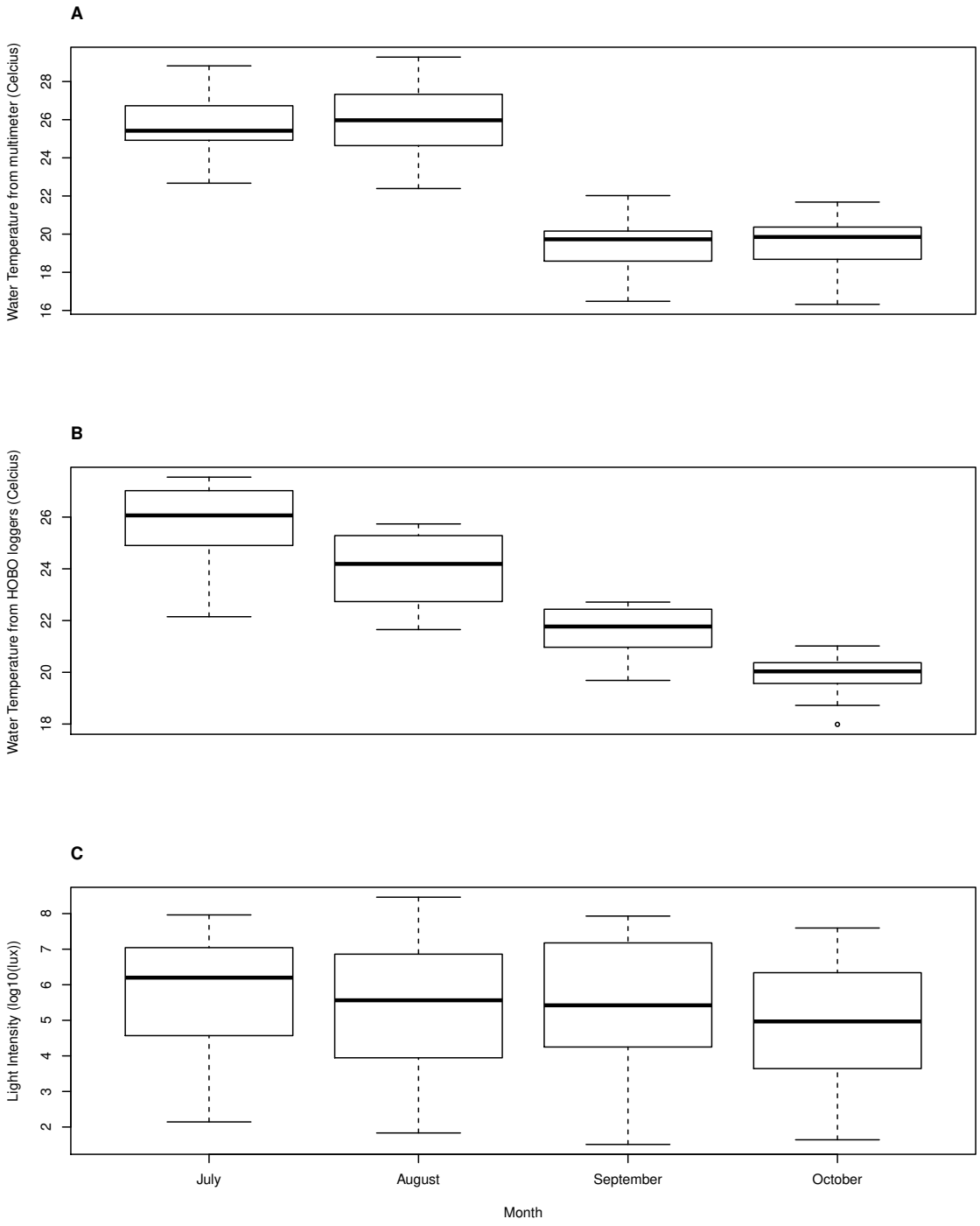


Figure 2.5: (A): Boxplot summary of the average lake temperature measured at the time of sampling with hand-held multimeter (B): Boxplot summary of average lake temprature from HOBO loggers. (C): Boxplot summary of light intensity also measured by HOBO loggers.

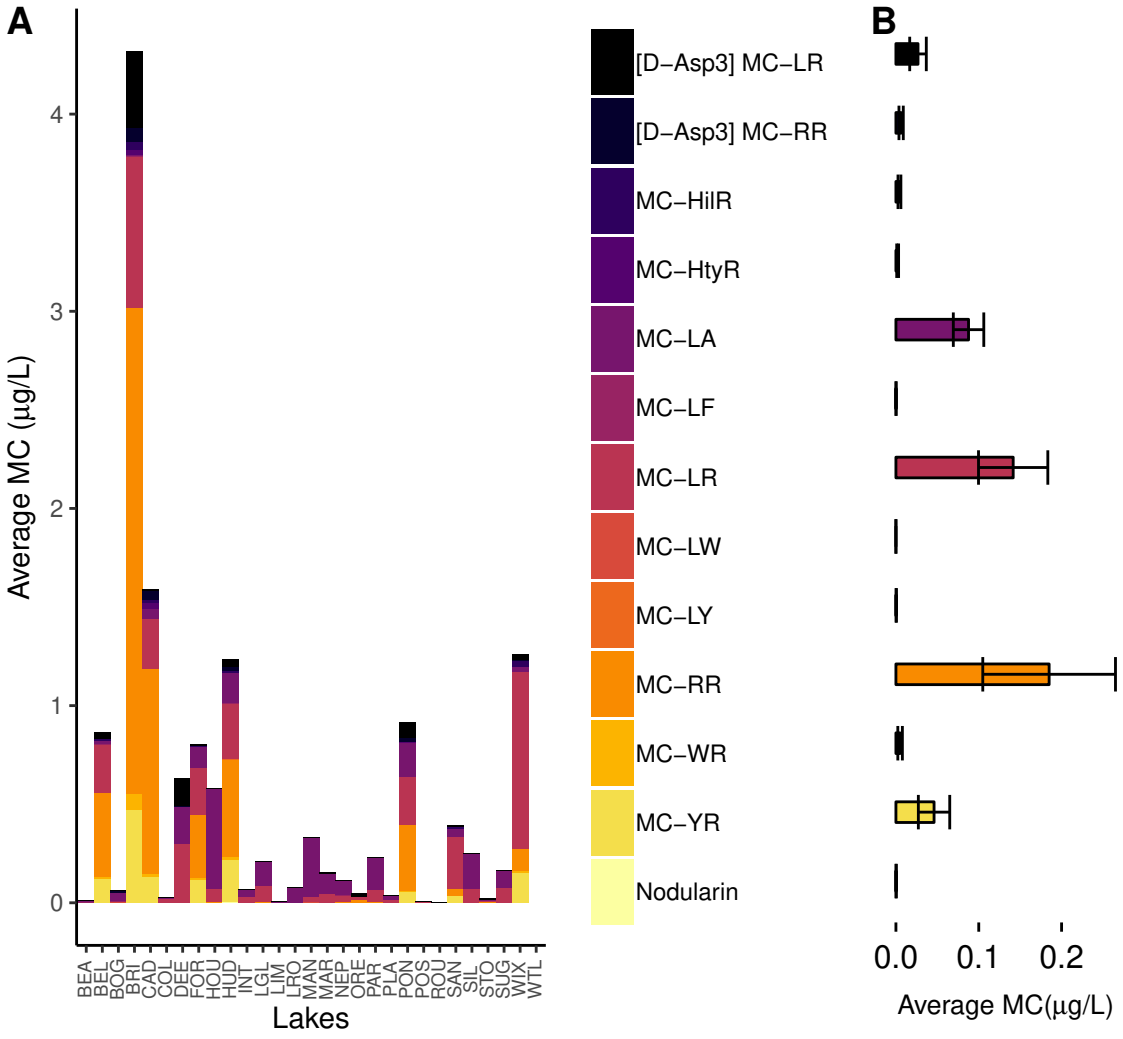


Figure 2.6: (A): Average MC found at each lake in all four months. The height of each bar represents the total average of MC. The proportion of each congener is represented by color in each bar (B): Average MC congener with the included error bars represents one standard deviation of the mean

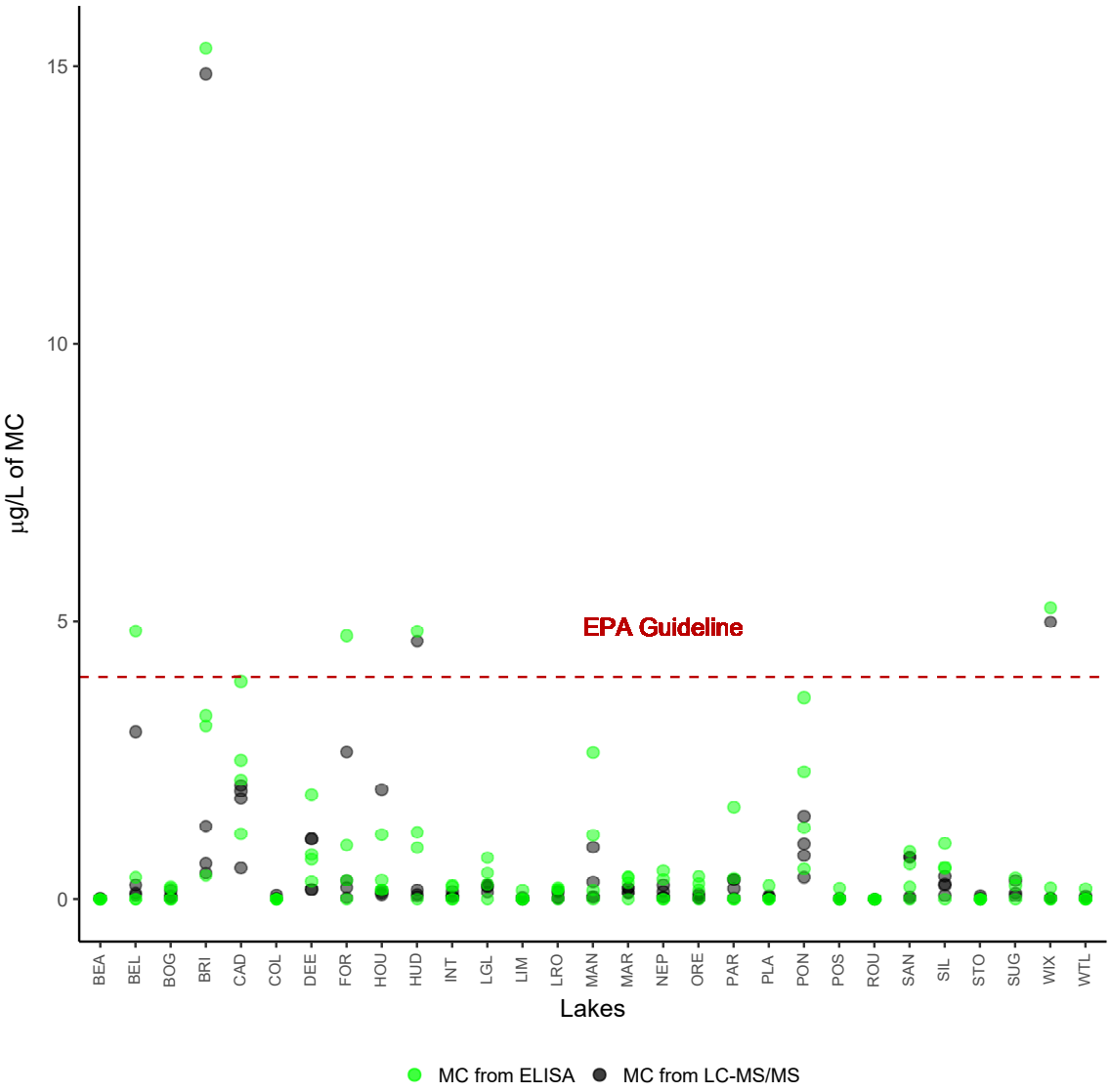


Figure 2.7: Total MC with all results from ELISA and LC-MS/MS plotted by each lake

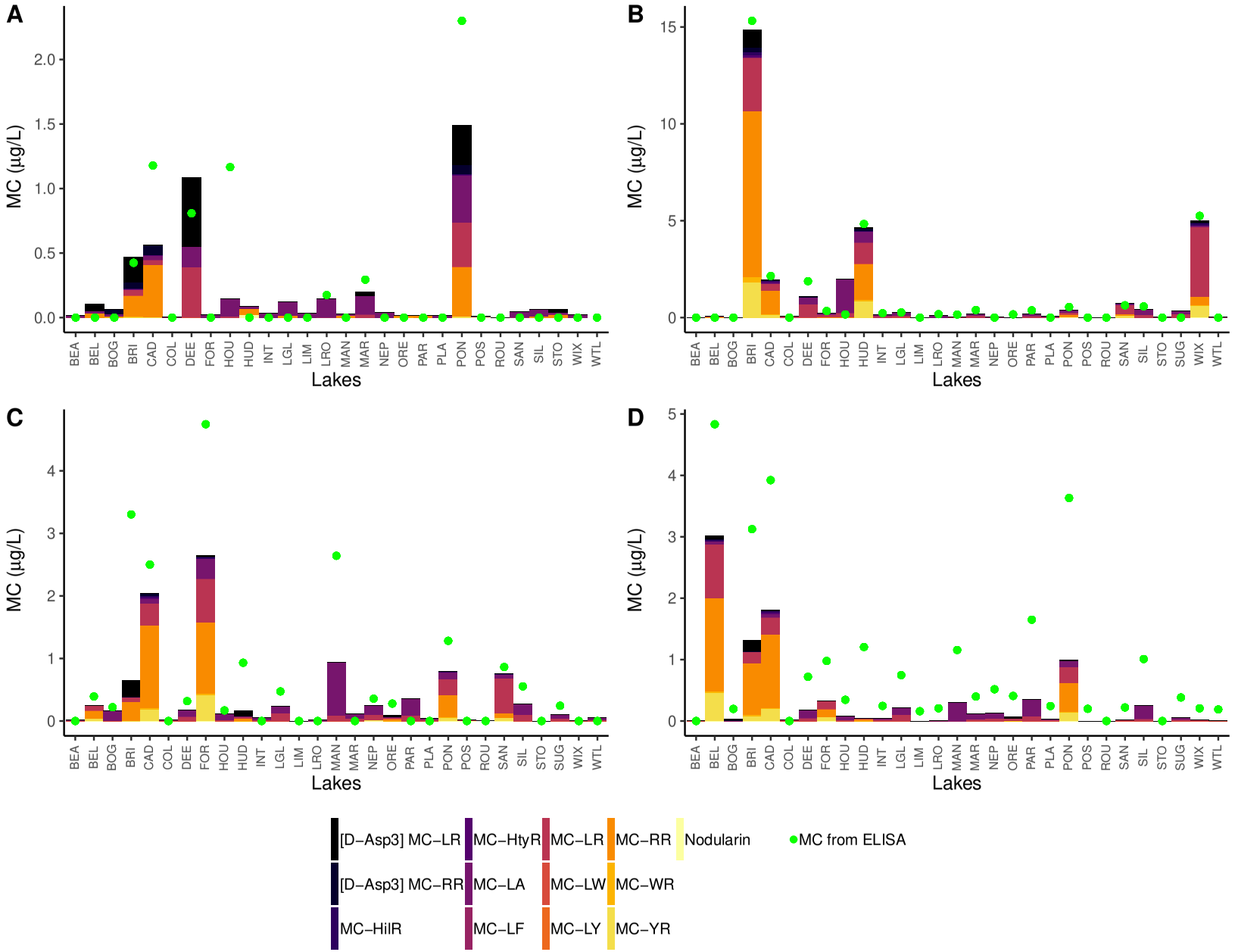


Figure 2.8: Barplots of MC and their congeners from LC-MS/MS and ELISA for each lake for the month of (A): July, (B): August, (C): September, and (D): October

### 2.3.3 *A Priori* Hypothesis Test

In my hypothesis, I expected developed land-use percentage to have a positive influence on MC concentrations. Much of the southern lakes were vastly dominated by developed land such as Belleville, Ore, Brighton, Ford and Bogie Lake. Developed land-use percentage plotted had a slight positive effect with $log10(\text{MC})$ concentration , however the relationship was not significant ($\beta = 0.59$, $F_{1,27} = 1.75$ , $p = 0.20$)(see figure 2.9) . There were no significant relationship with developed land-use percentage in predicting $log10(\textit{16s rRNA})$ gene copies ($\beta = 0.48$, $F_{1,25} = 1.75$ , $p = 0.27$). However, developed land-use had a positive significant relationship with log10(nitrate+nitrite) ($\beta = 0.75$, $F_{1,27} = 1.08$, $p = 0.018$)

Forest land-use percentage had a significant negative relationship with $log10(\textit{16s rRNA})$ gene copies ($\beta = -1.42$, $F_{1,25} = 7.08$, $p = 0.013$)(see figure 2.10). Albiet, forest land use did not have a significant effect on $log10(\text{MC})$ concentrations ($\beta = -0.34$, $F_{1,26} = 0.32$, $p = 0.57$).

With the surveyed lakes, 17 out of 29 had Zebra mussels found in the month of October. The average MC concentrations with of Zebra mussles present was 0.568 $\mu$g/L, and absent with 0.305 $\mu$g/L. Using a two sample Student's t-test, we fail to reject the null where difference of the mean is greater than 0 ($t = 1.14$, $df = 107$, $p = 0.13$). We also did not find any significant relationship with the mussel counts and mass with predicting MC concentration.

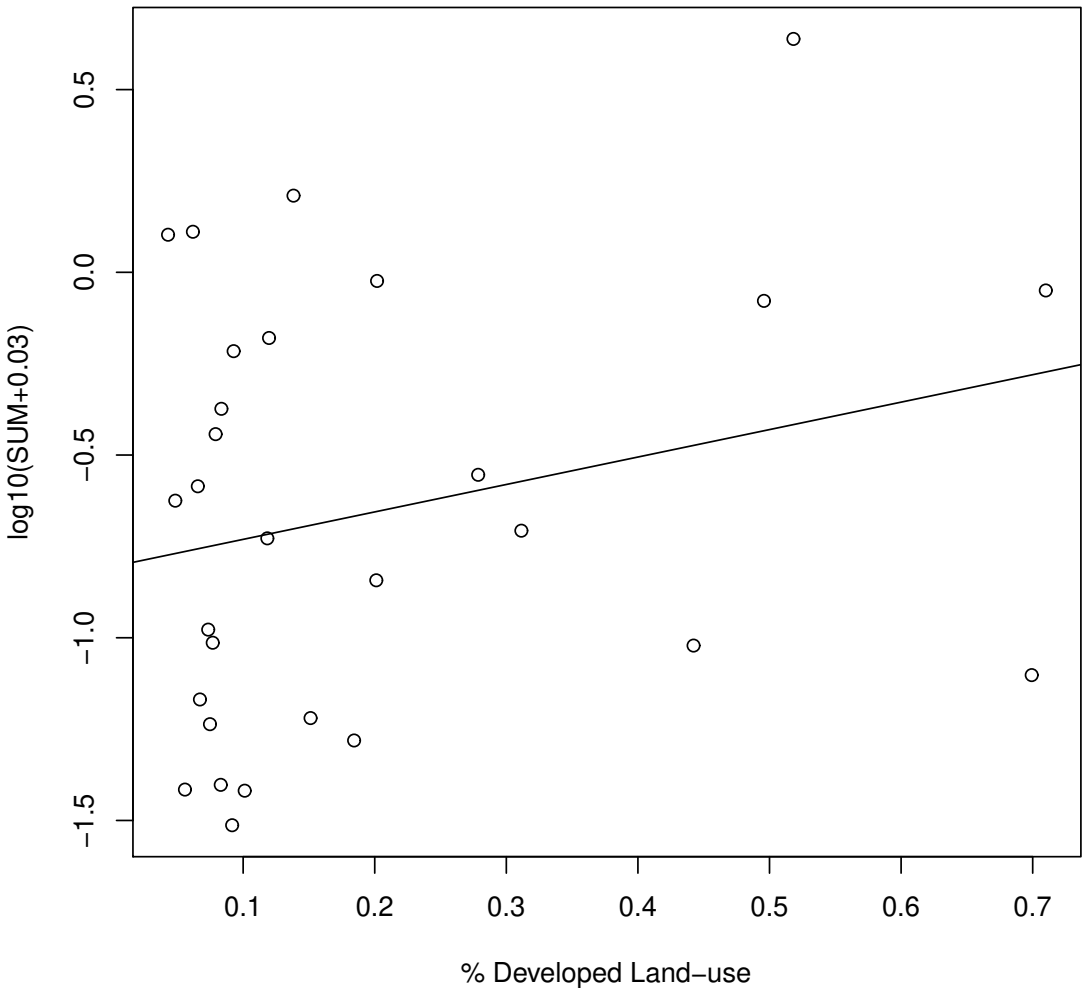


Figure 2.9: A slight positive relationship between $log10(MC)$ and developed land use. ($\beta = -0.59$, $F_{1,27} = 1.75$, $p = 0.20$)

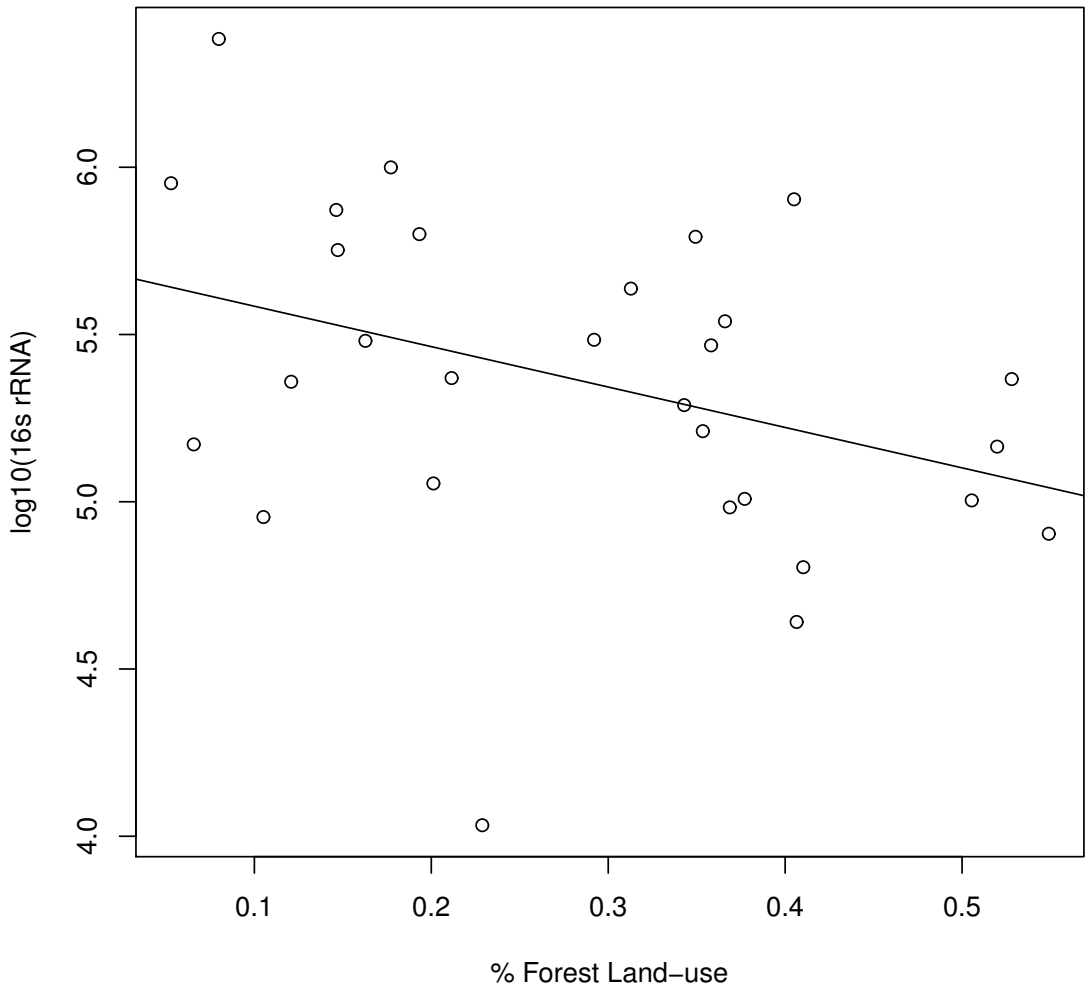


Figure 2.10: A negative relationship between $log10(MC)$ and forest land use. ($\beta = -1.42$, $F_{1,25} = 7.08$, $p = 0.013$)

### 2.3.4 Exploratory Analysis

With our model selection, a correlation matrix analysis was done to observe any possible collinear relationships between our predictor variables. The correlation matrix analysis showing only significant relationships and arranged by the anglular order of eigenvectors and identified few correlated variables (see 2.11, and see A.1 for the full correlation values). In our correlation matrix analysis, we observed a few relationships that may be meaningful. We found turbidity positively correlated with orthophosphate, total phosphorus, total Kejldahl nitrogen, water temperature, phycocyaninin and chloraphyll. Chlorophyll is positively correlated with *mcye*, phycocyanin, *16S rRNA*, total Kejldahl nitrogen and orthophosphate. Conductance was found to have negative relationship with precipitation. Developed land-use is positively correlated with nitrate+nitrite, total nitrogen and conductance. Watershed area was positive correlated with Zebra mussel mass. Agriculture had a positive relationship with nitrate+nitrite, Forest land-use had negative correlation with total Kejldahl nitrogen, total nitrogen, conductivity, ammonia, nitrate+nitrite and total nitrogen to total phosphorus ratio. Precipitation had negative correlation with total nitrogen, orthophosphate and conductivity.

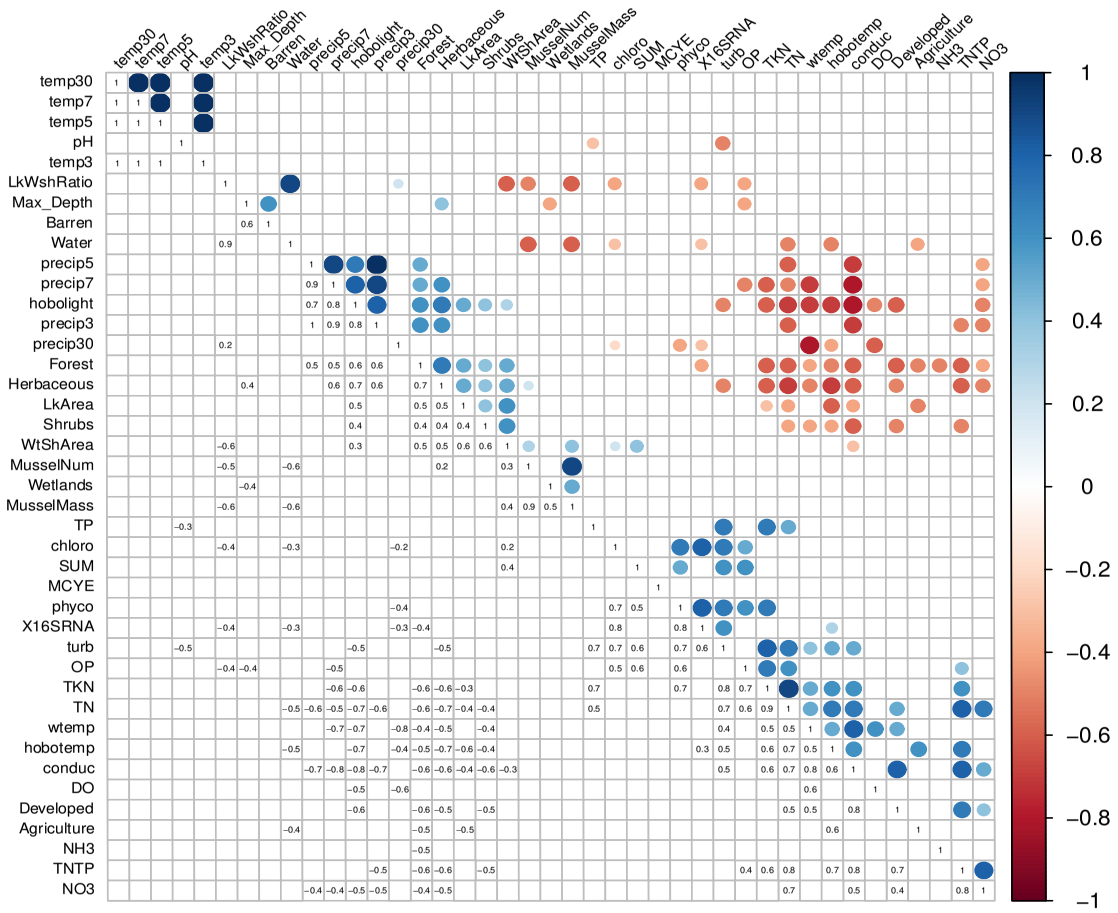


Figure 2.11: Correlation matrix with calculated Pearson coefficient in the lower triangle, and a graphical representation of coefficient value in the upper triangle. Each pair was tested for association between paired variables with Pearson's product moment correlation with relationships not shown if $\alpha > 0.05$. Data matrix was arranged by the angular order of the eigenvectors.

### 2.3.5 Feature Selection

From our best subset analysis with the largest subset size of 3 (nvmax=3), orthophosphate, *mcyE*, turbidity, max depth, lake area to watershed ratio, barren land-use and precipitation 3 days prior to sampling were frequently chosen predictor variables for MC concentration (see figure 2.12).

From our data, $log10$(Orthophosphate) was shown to be positively correlated with $log10$(turbidity) from our exploratory analysis, with a simple linear regression analysis, the relationship is slightly significant ($\beta = 0.57$, $F_{1,26} = 3.13$, $p = 0.08$). The regression did have outliers, with the two data points removed the relationship is significant ($\beta = 0.17$, $F_{1,25}$, $p = 0.01$) (see 2.13). We avoid including either $log10$(orthophosphate) or $log10$(turbidity) in our models together. With $log10(mcyE)$ as a single predictor in a linear mixed model shown to have a significant relationship with $log10$(MC), however we are missing data from July ($N = 91$) which does not allow us to use the full dataset ($\beta = 0.15$, $F_{1,72}$, $p = 0.002$).

We found $log10$(turbidity) as the best single predictor. The relationship was significant with predicting $log9$(MC) concentration in a linear mixed-effect model ($\beta = 0.55$, $F_{1,27} = 5.90$, $p = 0.02$) (see 2.14). However, with $log10$(orthophosphate) as a single predictor the relationship is almost significant ($\beta = 0.35$, $F_{1,26} = 3.13$, $p = 0.07$) (see 2.14). The best model for predicting MC is turbidity as the sole predictor. Adding additional predictors did not significantly improve the model.

For predicting the $log10(16s\ rRNA)$ gene copies, chloraphyll, agriculture land-use and average temperature 3 and 30 days prior to sampling and precipitation 5 days before sampling were investigated for building our model (see figure 2.15). Precipitation and temperature data did not have significant relationship and was

eliminated from the model. Agriculture percent land-use found to be the best predictor for *16s rRNA* ($\beta = 0.70$, $F_{1,26} = 5.10$, $p = 0.03$)(figure 2.16)

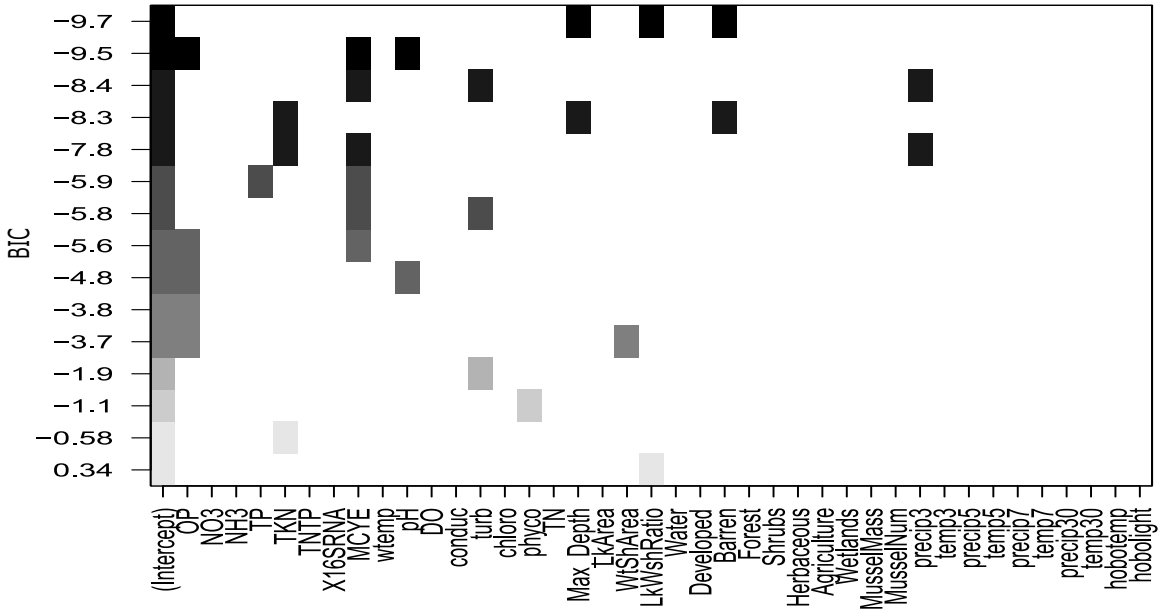


Figure 2.12: Subset regression analysis with MC Sum from LC-MS/MS as response variable. Each row is a model. Variable is included in the model it is represented as a black rectangle. The BIC is plotted on the y axis where the lowest value is higher up on the axis.

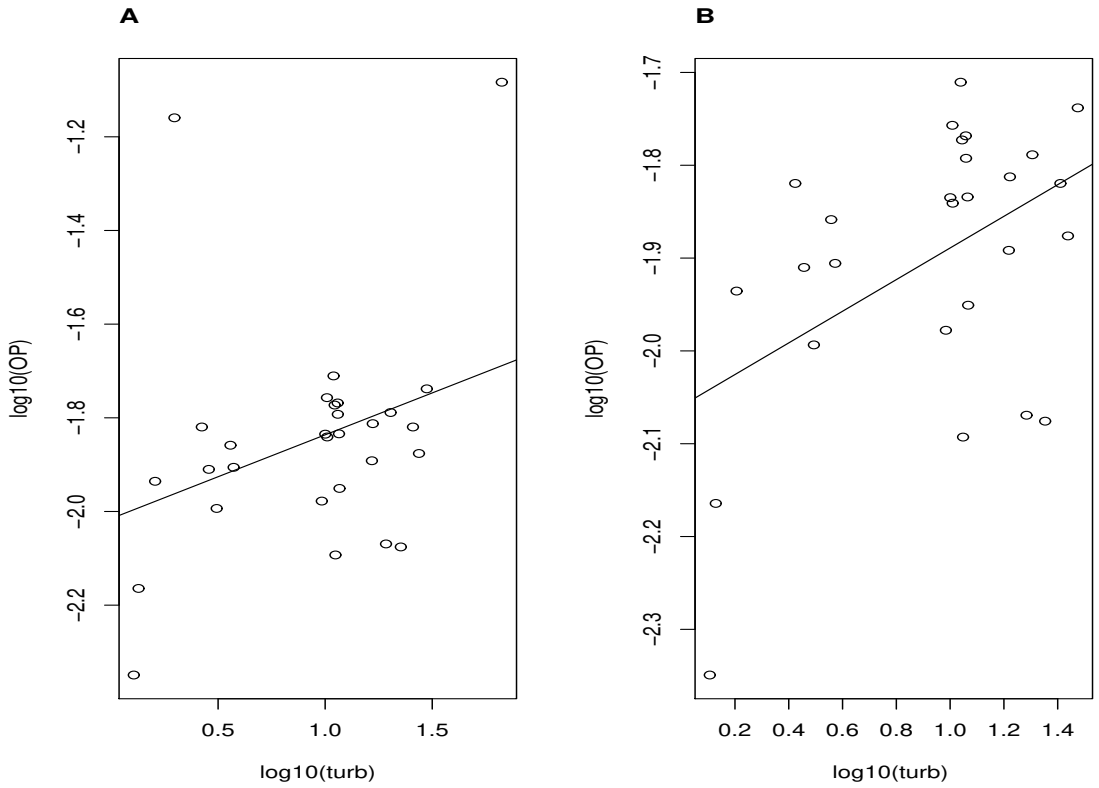


Figure 2.13: (A): Positive relationship with average $log10(\text{OP})$ with average $log10(turb)$ ($\beta = 0.57$, $F_{1,26} = 3.13$, $p = 0.08$). (B): With two outliers removed, the relationship between $log10(OP)$ and $log10(turb)$ was signifigant ($\beta = 0.17$, $F_{1,25}$, $p = 0.01$).

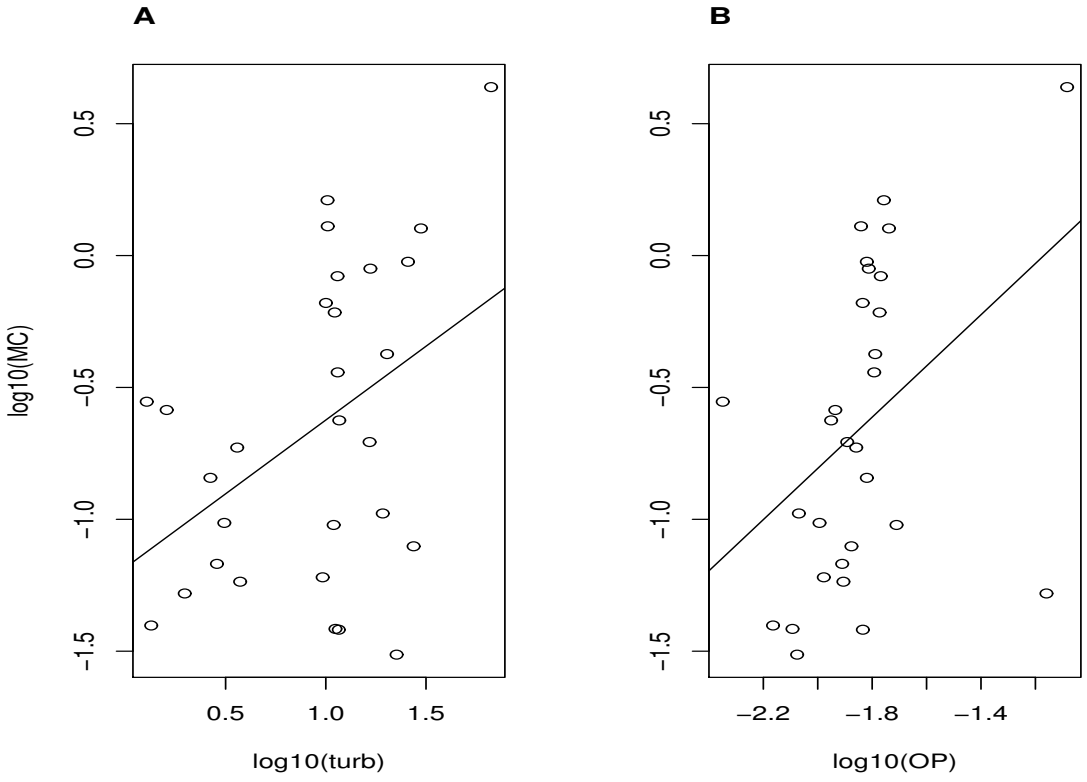


Figure 2.14: (A): Positive relationship between average $log10(\text{MC})$ with average $log10(turb)$ ($\beta = 0.55$, $F_{1,27} = 5.90$, $p = 0.02$) (B): Positive relationship between average $log10(MC)$ and average $log10(\text{OP})$ ($\beta = 0.35$, $F_{1,26} = 3.13$, $p = 0.07$). Turbidity is our best predictor variable for MC. Orthophosphate is nearly significant predictor

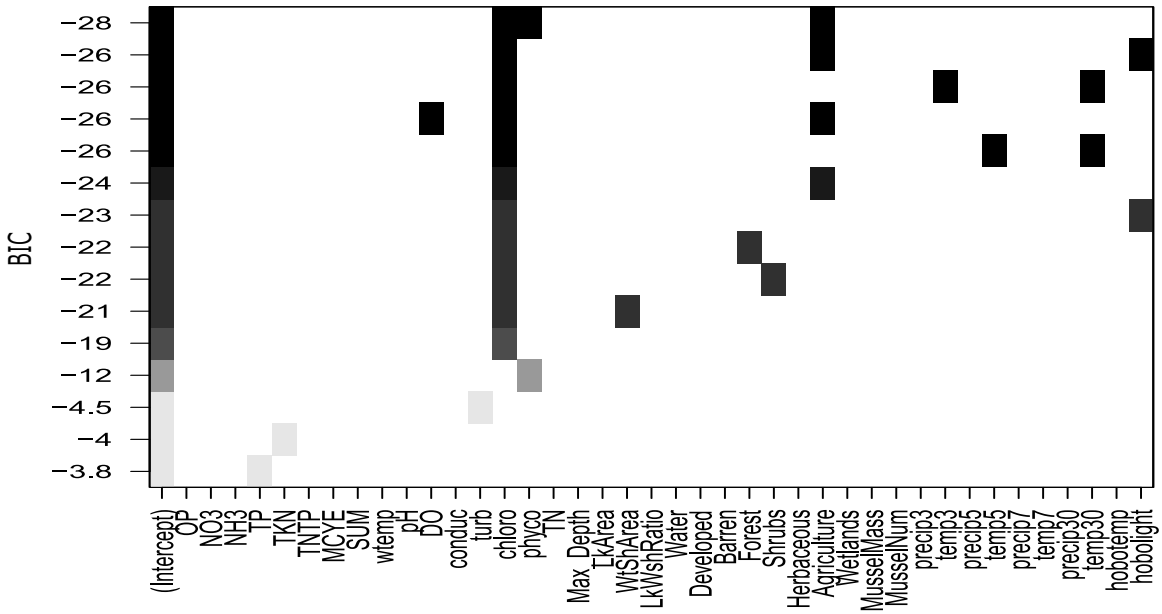


Figure 2.15: Best Subset: *16s rRNA* Gene copies as response variable

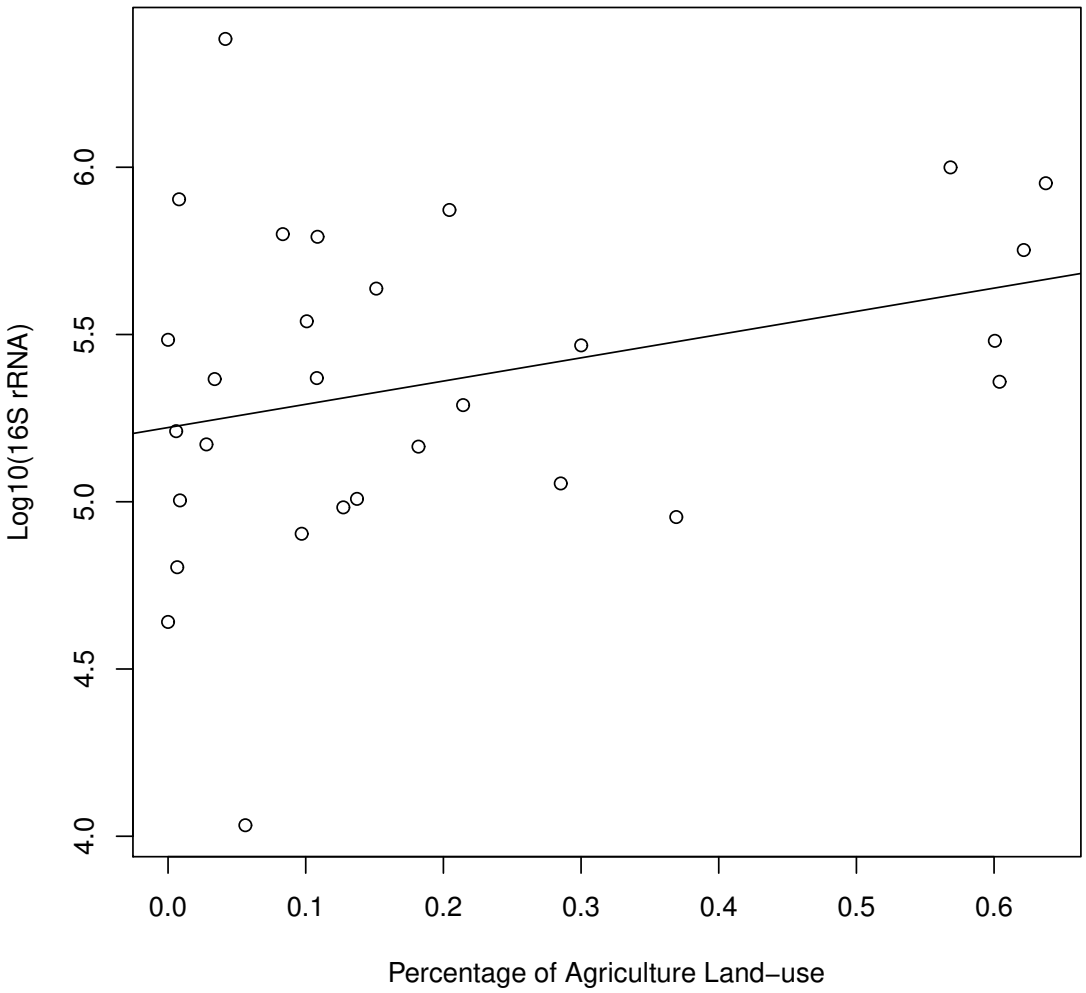


Figure 2.16: Positive relationship between percent agriculture land-use and $log10(16s\ rRNA)$ ($\beta = 0.70$, $F_{1,26} = 5.10$, $p = 0.03$)

### 2.3.6 Discussions

Comparing the results between MC concentrations and *16s rRNA*, we can see a difference in figure **??**. Lake turnover, winter stable, spring mix, summer stable, fall mix. Microcystin concentration tells a different story compared to *16s rRNA*. This curve is to be expected due to warmer temperature as our data does follow this pattern (see figure 2.5).

In our first hypothesis, I wanted test if disturbance from developed land is associated with HABs and if it has an affect on nutrient concentration. From our data, we did not find a significant relationship between developed land use and MC concentrations or *16s rRNA*. However, we did find developed land use having some effect with nitrate+nitrite.

Forest has negative correlation with NUTRIENTS

Disturbances of flora contributes to lower nutrient retention and hydrological impacts that can cause more blooms [19, 71, 72]. MC could increase as developed is more prevalent in the lake's watershed.

Bioavailability of phosphorus is pH dependent, where most is available between when in alkaline conditions greater [73]. Mostly all of our sampled lakes were mostly alkaline conditions averaging around 8.53±0.37 pH ranging from 7.44 to 9.94.

Wixom lake is very unique, it has the largest watershed amongst all our surveyed lake and has visually one of the largest extent

In controlled lab experiments, nutrients such as inorganic nitrogen or phosphorus are found to be limiting growth factors. In a when *Microcystis aereginosa* grown in the laboratory [74, 75].

Cyanobacteria are versatile as they can acquire nutrients under extreme conditions. They can utilize a process called quorom sensing which they coordinate with each other by using a signaling hormone acylated homoserine lactones [76]. They also can create a network of cells which work together as a unit, often as seen as a layer of green goo floating on top of water. There are very robust organisms, as some can change their buoyancy by modulating the intracellular gas vesicles gives them a competitive advantage over other species for seeking light [77].*Microcystis* can form a complex colony made by a mucilage structure which can be bouyant due to high concentration of dissolved oxygen [74]. Most of the cyanotoxin are intracellular, however with increase turbulence from wind and precipitation can release more toxins due to cell lysis either from cell apoptosis or from mechanical abruption which release [78].

# CHAPTER 3

## Solid Phase Adsorption Toxin Tracking

### 3.1 Introduction

Monitoring waterbodies for HABs require frequent visits to water sampling at a point in time can miss HABs events. Concentrations may fluctuate through time and sampling at a large interval may not capture the reality of the lake's condition. Solid Phase Adsorption Toxin Tracking (SPATT) is new unique method of monitoring waterbodies in a more time-integrative approach. SPATT is used as a bag or a sachet is filled with a porous resin inside a permeable bag. SPATT is used in for monitoring other analyte of interest such as diarrhetic shellfish poisoning [79]. 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.

### 3.2 Methods

The SPATT bag will be make with Nitex[1] and filled with HP-20[2], a non-polar resin (styrene-divinylbenzene copolymer), To construct the SPATT bags, a 1 meter x 5 centimeter strip of Nitex mesh were cut with 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.

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

9-10cm of sewn Nitex strips were cut and zip-tied about 0.5cm at one end. 3.00-3.01 grams of HP-20 resin was filled using a funnel. The other end is zip-tied once the Nitex bag is full. SPATT bags are 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 a slotted PVC pipe. At each lake, a float was installed as described in section 2.2.1. SPATT are left for about a month at each lake. When SPATT are retrieved, they are carefully removed and rinsed with Milli-Q water 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 10$\mu$M 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. 3.5mL of the supernatant is transferred to glass vials and analyzed by the Westerick group. See section 2.2.2

## 3.3 Results

Out of all 12 congeners, MC-LA, MC-LR, and MC-RR were the most frequently detected congener (see 3.1). Microcystin concentrations from SPATT was mostly detected in lake Paradise, Intermediate Lake, Lake Cadillac and Wixom

Lake. Comparing the average results from SPATT to the grab sample, we identified a major difference (see 3.1). When average concentration of MC from SPATT and grab sample is ordered by latitude, we found the relative mangitude of microcystin is higher found in the upper latitude of Michigan compared to regular grab sample. In our grab sample, Brighton, Pontiac, Wixom Lake and Lake Cadillac had high average MC with notable HABs.

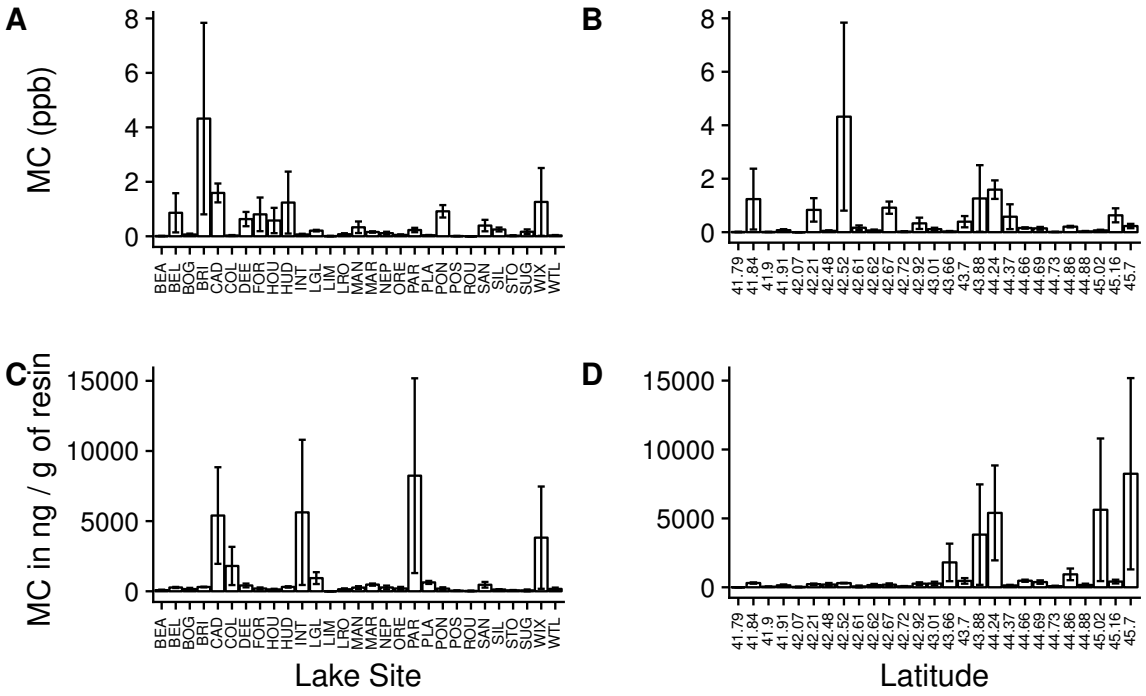


Figure 3.1: Microcystin measured from SPATTS compared to grab sample: and Grab Samples. Average concentration of MC are plotted as bar graphs. Microcystin concentrations analyzed from grab sample are shown in figure (A) arranged by each lake and (B) arranged by latitude. Microcystin concentrations analyzed from SPATT samples are shown in figure (C) arranged by each lake and figure (D) arranged by latitude

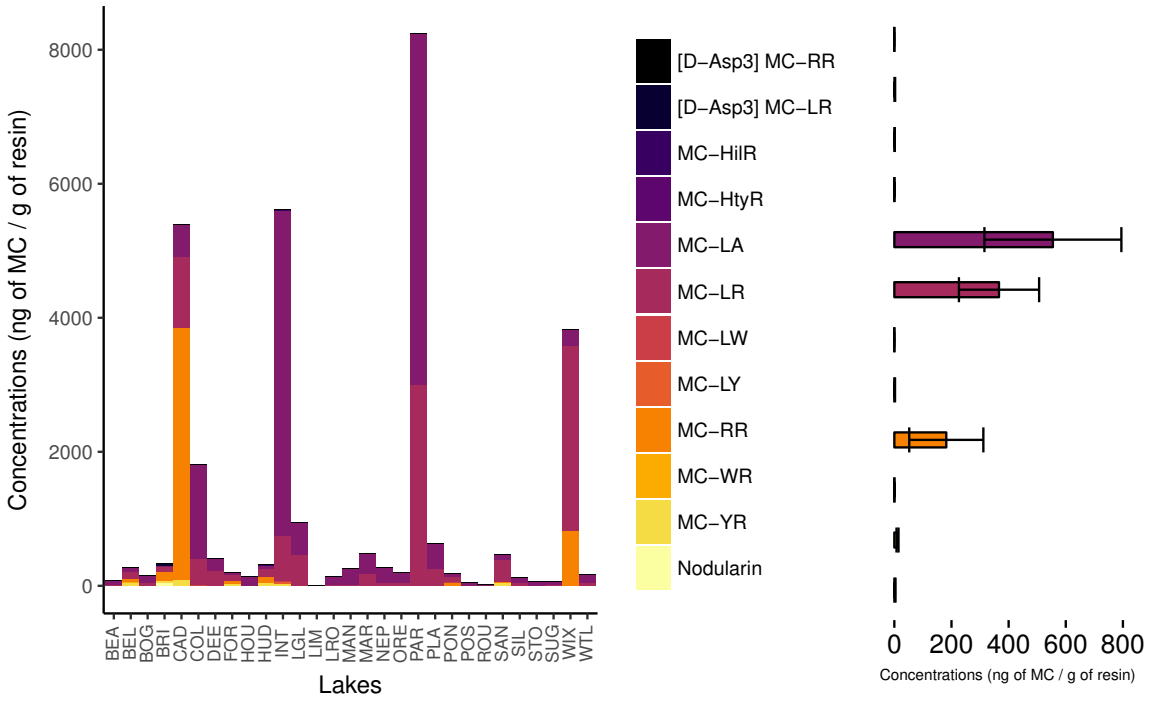


Figure 3.2: Average MC congeners at each lake. Error bars represent one standard deviation of the mean

Table 3.1: Microcystin Congener from SPATT

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| [D-Asp3]MC-RR | 91 | ND | ND | ND | ND |
| MC-RR | 91 | 164 | 1,116 | ND | 10,380 |
| Nodularin | 91 | 2 | 17 | ND | 160 |
| MC-YR | 91 | 10 | 30 | ND | 145 |
| MC-HtyR | 91 | 0 | 2 | ND | 18 |
| MC-LR | 91 | 353 | 1,211 | ND | 8,146 |
| [D-Asp3]MC-LR | 91 | 2 | 10 | ND | 92 |
| MC-HilR | 91 | 1 | 6 | ND | 53 |
| MC-WR | 91 | 0 | 1 | ND | 7 |
| MC-LA | 91 | 534 | 2,063 | ND | 13,977 |
| MC-LY | 91 | 1 | 11 | ND | 101 |
| MC-LW | 91 | 0 | 0 | ND | 0 |
| MC-LF | 91 | 0 | 2 | ND | 18 |
| Total MC | 91 | 1,066 | 3,273 | ND | 22,124 |

Values are expressed as (ng of MC / gram of resin)
ND=No Detects

## 3.4 Discussion

The results from SPATT seemed to suggest the lakes we found to have low microcystin from our grab sample missed other HABs. When lakes were ordered by latitude, it also suggest lakes found in the upper latitude of Michigan may had higher frequency of HABs.

We suggest there is another factor that explains the higher magnitude of microcystin. We started to observe biofilm to accumulate on the SPATT bags. The biofilm was more notable in lakes that had significant blooms such as Brighton, Pontiac, and Wixom Lake. Initially we did not expect this would effect the capacity of the SPATT, but this data may suggest this. In our next year of survey we will test the hypothesis if the biofilm has a negative effect of microcystin adsorption. We may need to dispatch the SPATT bags more often to prevent biofilm to encase and clog the Nitex mesh bag.

# CHAPTER 4

# CONCLUSION

## 4.1 Environmental Drivers

Using a predictive model is a vital utility for protecting the public from HABs. Addressing public health is a balance of scientific knowledge, risk assessment of the situation and maximizing available tools at our disposal. It is apparent that it is difficult to model HABs as it involves more complex dynamic. In general there are no best models in an ideal sense, but statistical modeling can help identify meaningful relationships. Point sampling does not encapsulate all the complex relationship and does not explain subtle changes. Confounding variables such as unaccounted disturbances not measured from our survey is not accounted.

Its important to understand why HABs is occurring. There are other factors to consider that our study did not observe. Urbanization also have effect with releases of heavy metals such as copper, lead, iron and zinc [80]

Hydrodynamic Modeling

## 4.2 SPATT

# APPENDIX A

# TABLES AND FIGURES

Figure A.1: Correlation matrix displaying Pearson's coefficient on the full data.

Table A1: 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)

Table A2: Microcystin congener statistical summary

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Nodularin | 114 | 0.0003 | 0.002 | ND | 0.021 |
| [D-Asp3]MC-RR | 114 | 0.006 | 0.028 | ND | 0.255 |
| MC-RR | 114 | 0.185 | 0.855 | ND | 8.552 |
| MC-YR | 114 | 0.046 | 0.202 | ND | 1.799 |
| MC-HtyR | 114 | 0.002 | 0.012 | ND | 0.107 |
| MC-LR | 114 | 0.142 | 0.447 | ND | 3.570 |
| [D-Asp3]MC-LR | 114 | 0.027 | 0.107 | ND | 0.902 |
| MC-HilR | 114 | 0.004 | 0.019 | ND | 0.150 |
| MC-WR | 114 | 0.005 | 0.029 | ND | 0.302 |
| MC-LA | 114 | 0.088 | 0.196 | ND | 1.729 |
| MC-LY | 114 | 0.0004 | 0.002 | ND | 0.024 |
| MC-LW | 114 | ND | ND | ND | ND |
| MC-LF | 114 | 0.0002 | 0.001 | ND | 0.012 |
| MC Sum from LC MS/MS | 114 | 0.505 | 1.580 | ND | 14.857 |
| MC from ELISA | 115 | 0.747 | 1.784 | ND | 15.320 |

Values are expressed as($\mu$g of MC*$L^{-1}$)
ND=No Detects

Table A3: QPCR statistical summary table

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| 16S rRNA | 112 | 405,761 | 798,946 | ND | 6,765,631 |
| mcyE | 91 | 8,517 | 49,396 | ND | 467,174 |
| cyrA | 93 | ND | ND | ND | ND |
| sxtA | 93 | ND | ND | ND | ND |

Values are expressed as Gene copies/mL
ND=No Detects

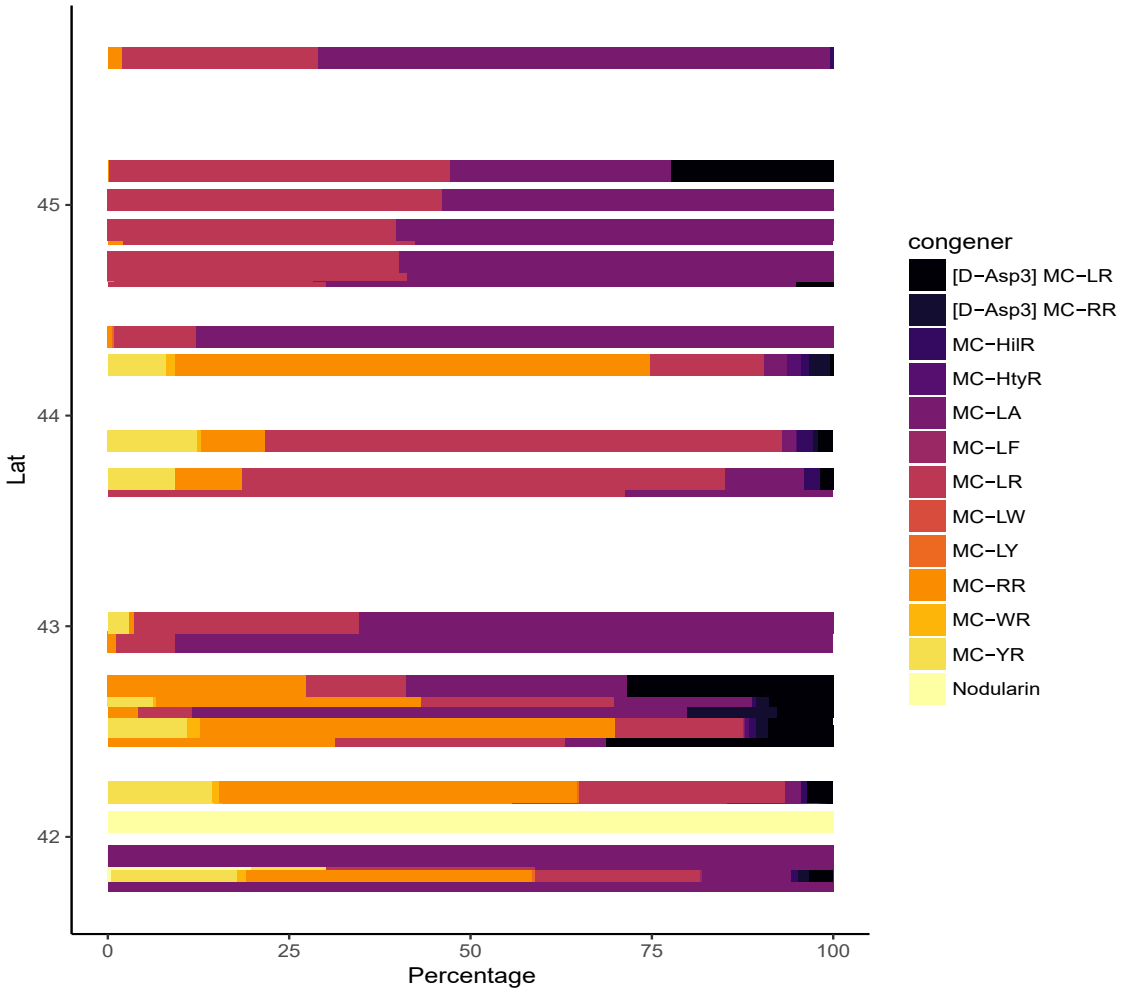


Figure A.2: Proportion of MC congeners plotted by latitude

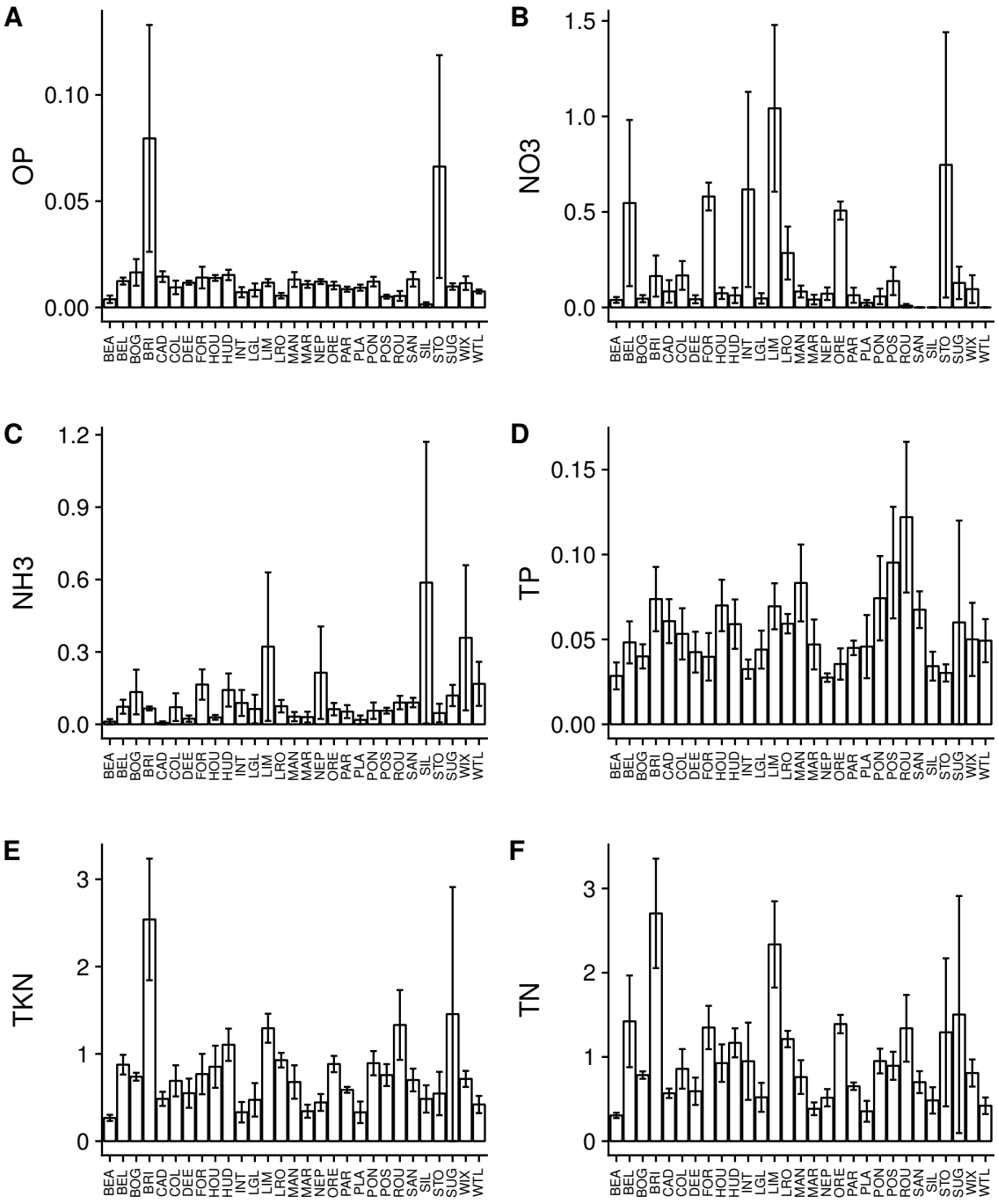


Figure A.3: Average nutrient concentrations for each lake. Orthophosphate (mg-P/L) in figure (A), nitrate+nitrite (mg-N/L) in figure (B), ammonia (mg-N/L) in figure (C), total phosphorus (mg-P/L) in figure (D) and total Kjeldahl nitrogen (mg-N/L) in figure (E). Error bars represents one standard deviation of the mean.

Table A4: Lake nutrients statistical summary

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Orthophosphate (mg P/L) | 114 | 0.015 | 0.030 | ND | 0.237 |
| Nitrate+Nitrite (mg N/L) | 115 | 0.199 | 0.443 | ND | 2.827 |
| Ammonia (mg N/L) | 115 | 0.112 | 0.281 | ND | 2.338 |
| Total Phosphorus (mg P/L) | 114 | 0.055 | 0.037 | ND | 0.239 |
| Total Kjeldahl Nitrogen (mg N/L) | 114 | 0.763 | 0.602 | ND | 4.555 |
| Total Nitrogen | 114 | 1.074 | 0.870 | 0.103 | 4.717 |

ND=No Detects

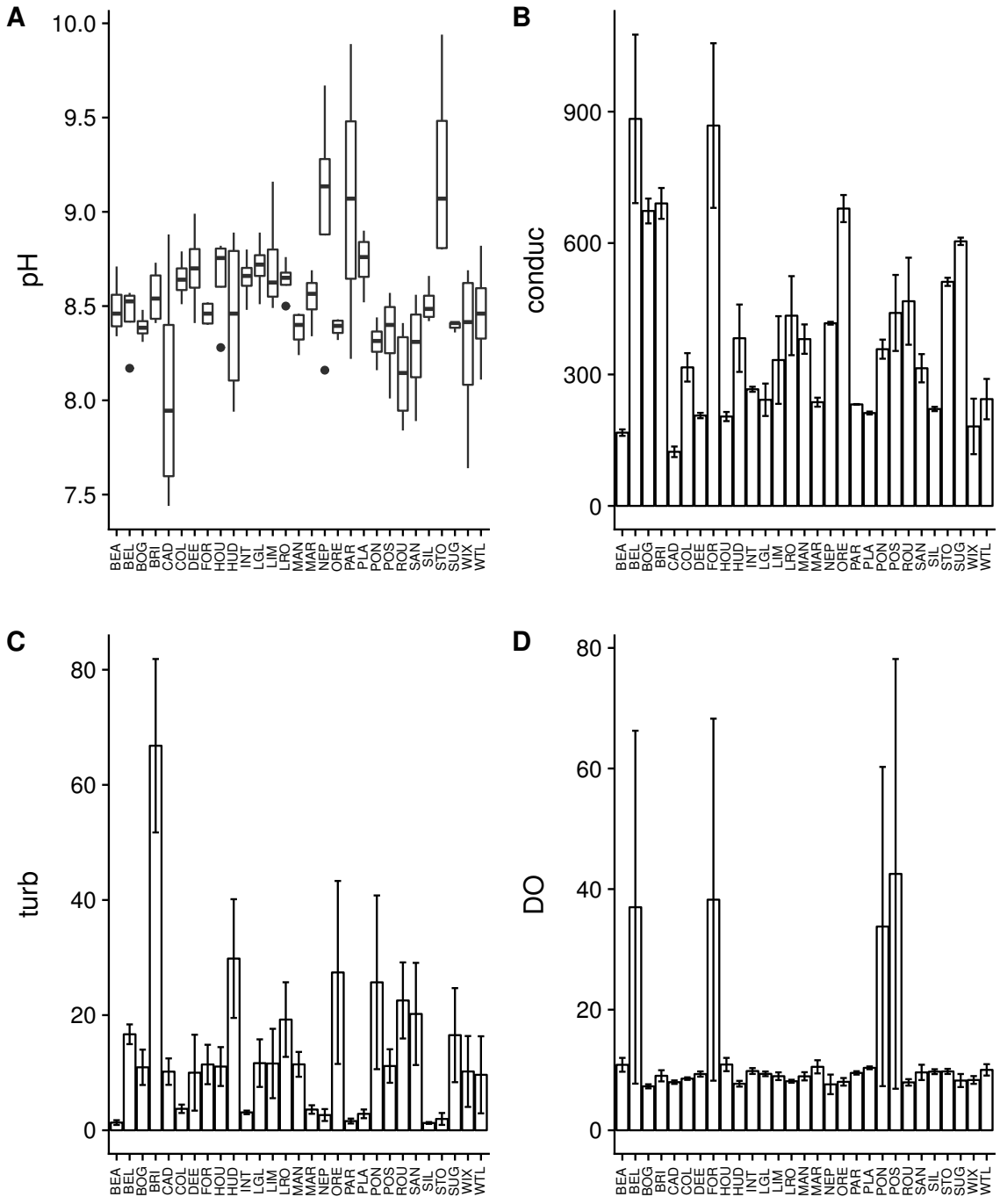


Figure A.4: Summary of measured water chemical parameters: A box and whisker plot of pH in figure (A). Bar plots of average conductance (B), average turbidity (C), and average dissolved oxygen (D) plotted by each lake.

# APPENDIX B

# GEOSPATIAL DATA COLLECTION GUIDE USING OPEN DATA KIT

## B.1 Introduction

Collaborative work is the fundamental part of a large scale research project. Data aggregation and structural collection is one of the few key factors that determines the quality of the study. Large scale studies are often done by a team of multiple groups and organizations. There are available commercial tools which can aid in data collection and aggregation automatically. ArcGIS [1] coupled with their Android/IOS app called Collector for ArcGIS [2].

From my experience I found a viable alternative which does not require an expensive license. I wish to help other research studies with a cheaper alternative and greatly facilitate their pursuit. I have written this guide to allow anyone to build from the ground up on setting a database and providing a viable way to organize an enterprise. This guide is aimed to initially setup a database for non-profit organization and academic research with a limited budget. The initial cost is $5 per month for a virtual private server (VPS) which has enough storage for a small database. The software tool is from an open-source software by Open Data Kit (ODK)[3].

## B.2 Required Components

Here you must setup a server that stores you data and be accessible through the Internet. Having your own computer machine as a server is a viable solution but it requires more extensive networking knowledge. An alternative solution is to purchase a rent a server from a third party which automatically sets up and quickly

[1] ESRI 2017. ArcGIS Desktop: Release 10.5.1 Redlands, CA: Environmental Systems Research Institute.

[2] https://www.esri.com/products/collector-for-arcgis

[3] Open Data Kit 2.0: Expanding and Refining Information Services for Developing Regions `http://www.hotmobile.org/2013/papers/full/2.pdf` Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, Gaetano Borriello In HotMobile, 2013. `http://dl.acm.org/citation.cfm?id=2444790`

gets connected to the Internet. A Virtual Private Server (VPS) are cheap and flexible servers to set up a database up on the cloud. For as little as $5 dollars a month one can have a server with everything they need to start collecting data.

DigitalOcean, Amazon (free trial for a year), Google VPS can work. I suggest anyone of them, but in this guide we will choose to setup from DigitalOcean as its the cheapest available server. I recommend Debian or Ubuntu linux distribution to setup your database. In this guide, we will choose Debian 9 (Stretch).

## B.3 Syntax

In the terminal, you will need bash commands to setup the server's configuration. In this guide, bash commands will be represented the $ symbol. Each line of code is represented with a single $ at the beginning. Enter the text following anything after the $. With some lines of code, there will be included comments which will be represented by the #.

```
$ run this command # This is a comment
```

Note that the first line starts with a#. This is not a command but a comment. In bash language, this is usually ignored. In this guide, a line starting with $ designates a command. Do not include the $ as its merely a symbol to designate a command. Running lines will not have $ as its indented to be one line of code.

## B.4 Server Setup

You must get an SSH client to access the server for setting up ODK Aggregate. For Linux or Mac, use the native terminal application to access the server. For windows users, I recommended program is PuTTy `https://www.putty.org/`. Install this program for windows. With PuTTy, open up PuTTy and enter in the IP address

and click “Open”. You will enter “root” and the password given by the online service or by the user installation setup. Usually when using VPS services, the IP address and the password for user *root* information is given by email. The command for linux or mac users, they can use Terminal and use this command:

```
$  ssh root@[EnterYourIp] # SSH with root as user
#  replace [EnterYourIp] with the given IP address
```

From a fresh installation, the first prompt when accessing the server will be to change your password. Choose a good strong password for the root username and have it written down. After setting the password, preform any upgrades and install sudo which allows a normal user to have elevated privileges and install nano, a text editor. This may be already installed but some distributions might not have them. To do this, follow these commands.

```
$  apt update          # Use the apt package manager and update repository
$  apt upgrade -y      # Preform any upgrades
$  apt dist-upgrade -y # Distribution upgrades
$  apt install sudo nano # Install the sudo and nano packages
```

Next, we will set a separate username. The *root* user is only used to administer the server. Choose any name as you see fit and something you can remember. For our example, for the case of this guide, *dummy* will be the username.

```
$  adduser dummy
```

You will be prompted to create a password. When asked for name and other user information, hit “Enter” to leave it blank. This does not need to be filled out.

Here, we will also assign this new user as a sudo user. This will enable the new user to have elevated privileges. This will allow to run command lines with root access if "sudo" is run before the line of command. Be cautious with commands with sudo.

```
$ usermod -a -G sudo dummy # Add the user dummy to sudo group
```

Its highly advisable to setup a pubic/private Rivest-Shamir-Adleman (RSA) key pair for more secure login setup. This prevents unauthorized access to the server and prevents the basic brute-force attacks. Here are the steps necessary to secure the server against malicious attacks. We will also setup a firewall called ufw (Uncomplicated Firewall) which is very simple to setup and secure your server. We will also generate a key-pair for this user. You can have a passphrase which is an additional layer of security, or leave it blank. When generated, the public key will be located at `/home/dummy/.ssh/id_rsa.pub`. Follow this command:

```
$ su dummy           # Switch out from root and change the user to dummy
$ sudo apt install ufw # Install uncomplicated firewall
$ sudo ufw allow 22  # Allow SSH ports
$ sudo ufw allow 8080 # Allow aggregate server
$ sudo ufw allow 5432 # Allow PostgreSQL server
$ ssh-keygen         # Generate the keys for SSH
```

Next, we will move the generate public key to the authorized key folder. We will also limit the permissions of the authorized keys for only the user dummy to access, install putty tools and use it to convert the private key for windows client to use:

```
$ sudo mv ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
```

```
$ chmod 600 ~/.ssh/authorized_keys
$ sudo apt install putty-tools
$ cd ~/.ssh
$ puttygen id_rsa -o id_rsa.ppk
```

Next, you have to retrieve the converted key file from the server to your own local machine. With windows, I recommend using WinSCP for secure file transfer `https://winscp.net/eng/index.php`. When prompted, enter the IP address in host name, port 22, and enter the new username and password (dummy). In WinSCP, hidden files are not shown by default. Hit Ctrl+Alt+H to show hidden files. On the right side panel, you will find the servers directory. The left side is the local computer's directory. Here we can transfer files between the machines securely. On the right side, locate .ssh folder and find the id_rsa.ppk file and download it to your document folder. Simply double-click the file to download it to your local machine.

This file can be pre-loaded int PuTTy. Double click the id_rsa.ppk file, open up PuTTy and type the IP address. Entering in the username will automatically log in without password as the key authenticates you.

Next, we will disable login by password which only allows users with the key file to access the server. This secures the server as it prevents malicious bots from brute-force attacks:

```
$ sudo nano /etc/ssh/sshd_config
```

The command will open up *nano* text editor. Scroll down the config file by pressing the down arrow key. Find the line that contains

```
>    PermitRootLogin yes
```

Change it to:

```
>    PermitRootLogin no
```

Scroll all the way to the end and find the line containing:

```
>    PasswordAuthentication yes
```

And change it to:

```
>    PasswordAuthenticion no
```

Make sure there is no # in the beginning of the changed lines. Exit nano text editor by pressing "Ctrl+X". Press "Y" to save file and hit "Enter" to write file.
Next restart ssh:

```
$  sudo service ssh restart
```

Next, you will install ODK Aggregate. ODK Aggregate is a Java web applet that handles the data and stores into a database. The web applet is hosted as a website and can be accessed through any web browser. On the host machine, we will need to setup tomcat server that hosts ODK Aggregate along with an SQL server. PostgreSQL is used as we can also extend this to work with spatial data and be accessible from QGIS. As of writing this guide, ODK Aggregate is version 1.5.0. You may need to find the current version. Go to `https://github.com/opendatakit/aggregate`. If the version is different than 1.5.0, change the link following after the "wget" command.

```
$ sudo apt install tomcat8 unzip
$ wget https://github.com/opendatakit/aggregate/releases/download/v1
   .5.0/ODK-Aggregate-v1.5.0-Linux-x64.run.zip
$ unzip ODK-Aggregate-v1.5.0-Linux-x64.run.zip
$ chmod +x ODK-Aggregate-v1.5.0-Linux-x64.run
$ ./ODK-Aggregate-v1.5.0-Linux-x64.run
```

You will see a prompt when the ODK Aggregate Setup Wizard appears. Follow these steps:

1. Read the license. Hit "Enter" to scroll through the document. Type "Y" at the end of the document to accept the license terms.
2. Type `ODK` to create a folder where the necessary files will be created.
3. Type "3" to setup a PostgreSQL platform setup
4. Type "Y" to downloaded
5. Type "N" for ssl
6. Type "1" for no ssl
7. Type "Y" for port config
8. Hit "Enter" to default 8080
9. Type the IP address of the server.
10. Hit "Y" for PostgreSQL
11. Hit "Enter" to set the default PostgreSQL port to 5432

12. Hit "Enter" for default username *odk_user*

13. Make up a password for this user login

14. It will ask the name of the database *odk_prod*

15. Then it will ask to setup a name for your schema. *odk_prod*

16. The "ODK Aggregate Instance Name" is used as a display for your users. This can typically be set for a project name or title. For our purpose, it will be set to *demo*.

17. Create a super user name, most likely whomever is managing the data. We will set ours to be called *super*.

Next, you must setup PostgreSQL:

```
$ echo deb http://apt.postgresql.org/pub/repos/apt/stretch-pgdg main |
  sudo tee -a /etc/apt/sources.list
$ wget --quiet -O https://www.postgresql.org/media/keys/ACCC4CF8.asc |
  sudo apt-key add -
$ sudo apt update
$ sudo apt install postgresql-9.4
```

Next you need to create postgres as a username and setup a password. Here you will be switching into *psql* command prompt to set things up. Once you are done, you will exit *psql* by typing \q which will lead into the regular bash home directory. Follow these command:

```
$ sudo -u postgres psql postgres # This will log you into psql command
  prompt.
```

```
  % postgres=# \password postgres
  % postgres=# \q
$                              #  Back to bash command prompt
```

Now you must set up the configuration of PostgreSQL to allow remote connections. You will edit the `/etc/postgreql/9.4/main/pg_hba.conf` file and the `/etc/postgresql/main/postgresql.conf`:

```
$  sudo nano /etc/postgresql/9.4/main/pg_hba.conf
```

Change the line that has this:

```
>   local  all   all     peer
```

To this:

```
>   local  all   all     md5
```

Also add this line at the very end of the file:

```
>  host   all    all     0.0.0.0/0 md5
```

Then exit by pressing "Ctrl+X", then hit "Y" to confirm, then hit "Enter". Lets edit the other file:

```
$  sudo nano /etc/postgresql/9.4/main/postgresql.conf
```

Find this line:

```
>   #listen_addresses = 'localhost'
```

Replace it with this, note the removal of # at the beginning of the line:

```
>   listen_addresses = '*'
```

Now you will run the SQL configuration file generated from ODK Aggregate to setup the database in PostgreSQL. This will create the user *odk_user* and the database and schema *odk_prod*. Once this is done, you will restart PostgreSQL and have the database online.

```
$  sudo -u postgres psql postgres
    % postgres=# \cd '/home/dummy/ODK/ODK\ Aggregate'
    % postgres=# \i create_db_and_user.sql
    % postgres=# \q
$  sudo systemctl restart postgresql
```

ODK Aggregate run file from the previous step generated a .war file that needs to be copied to apache tomcat webapps folder. With tomcat8 installed on debian, the webapps folder is located at `/var/lib/tomcat8/webapps`.

```
$ cd $HOME/ODK/ODK\ Aggregate/
$ sudo cp ODKAggregate.war /var/lib/tomcat8/webapps
$ sudo service apache restart
```

Now you should have the ODK Aggregate server running. You can simply access the server by going to your web browser on your local computer and access it by replacing your given ip address of the server in this form: `http://xxx.xxx.xxx.xxx:8080/ODKAggregate` Log into your super user login and click "Site Admin". Here you can add more users to administer the site or be a data collector. Site administrators have the ability to create more users and have all other power as a data collector. For a simple user who is just collecting data, you can have them set as "Data Collector" and not be able to change any data that has already been collected. Next, we must create a form for the data collectors to use.

Before you start collecting data, you must set up a predefined form which consists of variables of interest. The ODK Collect app can collect spatial data in a form of GPS points, path of coordinates or trace a spatial polygon. ODK Aggregate can accept .xml files which needs to be created. The ODK Suite provides an online tool to create a desired form `https://build.opendatakit.org/`. When you first go to the site, you will be asked to sign up for a login. You do not have to create an account and can simply click cancel.

Here you can design your survey form. On the top-right hand corner, you must name this form. This will be displayed for the user, so choose something that pertains to what kind of data is being collected. Next, you will choose what kind of data are going to be collected. The most important are usually the date/time and GPS coordinates. On the bottom, there is an option to select location. When selected, a configuration window will appear on your right-hand side. For the name, name it location. You can have it configured to be a single point, a path which creates a polyline, or a shape which creates a polygon shape. For the date and time, you can configure the full date and time or just the date. Depending on your project, you can select a variety of data ranging from audio or video data or a design questionnaire. For audio or video there is a media option. You can collect audio, video and picture or have all three options if you wish. The questionnaires can be designed by clicking the select multiple option. Here you can have options for the data collector to choose from. This can be configured to have a follow-up questions depending on what options the data collector initially selects. More information can be found by clicking help for more advance configurations.

Once you have a satisfied form, you must save the form as an .xml file. On the top-right hand side, click File and select Export to XML. Once saved, head over you

your ODK Aggregate website and click on "Form Management". Click "Add New Form" and select the .xml file to upload to the server.

Next, the data collectors need to have the ODK Collect app to start collecting data. Android users can download the ODK Collect app either through the Google Play store or directly download the .apk file from `https://github.com/opendatakit/collect/releases/tag/v1.15.1`. Unfortunately there is no app for the IOS platform.

First, you must set the settings. Go to the "general settings" and tap on "server" settings. For the option "Type" choose the *ODK Aggregate* option. The URL will be the the same as the ODK Aggregate web address. Username and password will be the users that are created from ODK Aggregate website, or the super user. Head back to the main menu and click on "Get Blank Form". This will download the forms you have uploaded to the ODK Aggregate server. The app can collect data offline and send the filled forms once connected online. For each entry of data, tap the "Fill Blank Form" option and select the appropriate form. Here you will start filling out the form. Once completed, you will save the form offline. If you are connected to the Internet, you will have to upload the filled forms to the ODK Aggregate server. To do so, head to the main menu and select the "Send Finalized Form" option and simply tap on the form that is saved. This will now upload the data to your server.

## B.5 Conclusion

The utility from using ODK tools is beneficial for large scale studies, especially handing with geospatial data. ODK Aggregate is very flexible in working with other forms of applications. With PostgreSQL database, the stored data can be accessed either throught the ODK Aggregate webserver or by connecting to the

PostgreSQL database connection. QGIS ,ArcGIS Desktop or Microsoft Access can connect to this database and pull data. For more advance configuration, see the ODK documentation: `https://docs.opendatakit.org/`. You may want to consider registering a domain name which allows you to enter a human-recognizable address instead of an IP address.

REFERENCES


[1] Moore, Richard and Ohtani, Ikuko, "Cyanobacterial Toxins," *Gazzetta chimica Italiana*, vol. 123, no. 6, pp. 329–336, 1993.

[2] G. A. Codd, S. G. Bell, K. Kaya, and C. J. Ward, "Cyanobacterial toxins, exposure routes and human health," *European Journal of Phycology*, vol. 34, pp. 405–415, Oct. 1999.

[3] R. P. Rastogi, R. P. Sinha, and A. Incharoensakdi, "The cyanotoxin-microcystins: current overview," *Reviews in Environmental Science and Bio-Technology*, vol. 13, pp. 215–249, June 2014.

[4] E. Dittmann, D. Fewer, and B. Neilan, "Cyanobacterial toxins: Biosynthetic routes and evolutionary roots," *FEMS microbiology reviews*, vol. 37, Sept. 2012.

[5] M. Welker and H. von Dohren, "Cyanobacterial peptides - nature's own combinatorial biosynthesis," *FEMS Microbiol Rev*, vol. 30, pp. 530–63, July 2006.

[6] *Harmful Algal Bloom and Hypoxia Research and Control Amendments Act of 2014*. June 2014. Public Law No: 113-124.

[7] USEPA, *Drinking Water Contaminant Candidate List 4-Final*. USEPA, Nov. 2016. 81 FR 81099.

[8] T. W. Davis, S. B. Watson, M. J. Rozmarynowycz, J. J. H. Ciborowski, R. M. McKay, and G. S. Bullerjahn, "Phylogenies of Microcystin-Producing Cyanobacteria in the Lower Laurentian Great Lakes Suggest Extensive Genetic Connectivity," *PLoS ONE*, vol. 9, p. e106093, Sept. 2014.

[9] *Guidelines for drinking-water quality. Vol. 2, Health criteria and other supporting information: addendum*, vol. 2. World Health Organization, 1998.

[10] *Guidelines for safe recreational water environments*. World Health Organization, 2003.

[11] USEPA, *Draft Human Health Recreational Ambient Water Quality Criteria and/or Swimming Advisories for Microcystins and Cylindrospermopsin Documents*. Office of Waters, Dec. 2016. EPA 822-P-16-002.

[12] S. R. Pereira, V. M. Vasconcelos, and A. Antunes, "Computational study of the covalent bonding of microcystins to cysteine residues – a reaction involved in

the inhibition of the PPP family of protein phosphatases," *The FEBS Journal*, vol. 280, pp. 674–680, Jan. 2013.

[13] D. Tillett, E. Dittmann, M. Erhard, H. von Döhren, T. Börner, and B. A. Neilan, "Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system," *Chemistry & Biology*, vol. 7, pp. 753–764, Oct. 2000.

[14] M. C. Moffitt and B. A. Neilan, "Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins," *Applied and Environmental Microbiology*, vol. 70, pp. 6353–62, Nov. 2004.

[15] T. Nishizawa, M. Asayama, K. Fujii, K. Harada, and M. Shirai, "Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in Microcystis spp," *Journal of Biochemistry*, vol. 126, pp. 520–529, Sept. 1999.

[16] G. B. Trogen, A. Annila, J. Eriksson, M. Kontteli, J. Meriluoto, I. Sethson, J. Zdunek, and U. Edlund, "Conformational studies of microcystin-LR using NMR spectroscopy and molecular dynamics calculations," *Biochemistry*, vol. 35, pp. 3197–205, Mar. 1996.

[17] J. Puddick, M. R. Prinsep, S. A. Wood, S. C. Cary, and D. P. Hamilton, "Modulation of microcystin congener abundance following nitrogen depletion of a Microcystis batch culture," *Aquatic Ecology*, vol. 50, pp. 235–246, June 2016.

[18] M. N. Charlton, "Oxygen Depletion in Lake Erie: Has There Been Any Change?," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 37, pp. 72–80, Jan. 1980.

[19] D. M. Anderson, P. M. Glibert, and J. M. Burkholder, "Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences," *Estuaries*, vol. 25, pp. 704–726, Aug. 2002.

[20] T. Bucak, D. Trolle, U. N. Tavsanoglu, A. I. Cakıroglu, A. Ozen, E. Jeppesen, and M. Beklioglu, "Modeling the effects of climatic and land use changes on phytoplankton and water quality of the largest Turkish freshwater lake: Lake Beysehir," *Science of The Total Environment*, vol. 621, pp. 802–816, Apr. 2018.

[21] J. L. Graham, K. A. Loftin, M. T. Meyer, and A. C. Ziegler, "Cyanotoxin Mixtures and Taste-and-Odor Compounds in Cyanobacterial Blooms from the Midwestern United States," *Environmental Science & Technology*, vol. 44, pp. 7361–7368, Oct. 2010.

[22] W. W. Carmichael and G. L. Boyer, "Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes," *Harmful Algae*, vol. 54, pp. 194–212, Apr. 2016.

[23] N. W. May, N. E. Olson, M. Panas, J. L. Axson, P. S. Tirella, R. M. Kirpes, R. L. Craig, M. J. Gunsch, S. China, A. Laskin, A. P. Ault, and K. A. Pratt, "Aerosol Emissions from Great Lakes Harmful Algal Blooms," *Environmental Science & Technology*, vol. 52, pp. 397–405, Jan. 2018.

[24] N. R. Monks, S. Liu, Y. Xu, H. Yu, A. S. Bendelow, and J. A. Moscow, "Potent cytotoxicity of the phosphatase inhibitor microcystin LR and microcystin analogues in OATP1b1- and OATP1b3-expressing HeLa cells," *Mol Cancer Ther*, vol. 6, pp. 587–98, Feb. 2007.

[25] A. Saoudi, L. Brient, S. Boucetta, R. Ouzrout, M. Bormans, and M. Bensouilah, "Management of toxic cyanobacteria for drinking water production of Ain Zada Dam," *Environmental Monitoring and Assessment*, vol. 189, pp. 1–11, July 2017.

[26] J. A. Westrick, D. C. Szlag, B. J. Southwell, and J. Sinclair, "A review of cyanobacteria and cyanotoxins removal/inactivation in drinking water treatment," July 2010.

[27] J. Koreiviene, O. Anne, J. Kasperoviciene, and V. Burskyte, "Cyanotoxin management and human health risk mitigation in recreational waters," *Environ Monit Assess*, vol. 186, pp. 4443–59, July 2014.

[28] J. A. Westrick and D. Szlag, "A Cyanotoxin Primer for Drinking Water Professionals," *Journal - American Water Works Association*, vol. 110, pp. E1–E16, Aug. 2018.

[29] R. L. Vannote, G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing, "The River Continuum Concept," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 37, pp. 130–137, Jan. 1980.

[30] S. C. Chapra, B. Boehlert, C. Fant, V. J. Bierman, J. Henderson, D. Mills, D. M. L. Mas, L. Rennels, L. Jantarasami, J. Martinich, K. M. Strzepek, and H. W. Paerl, "Climate Change Impacts on Harmful Algal Blooms in U.S. Freshwaters: A Screening-Level Assessment," *Environmental Science & Technology*, vol. 51, pp. 8933–8943, Aug. 2017.

[31] V. H. Smith and D. W. Schindler, "Eutrophication science: where do we go from here?," *Trends Ecol Evol*, vol. 24, pp. 201–7, Apr. 2009.

[32] A. M. Michalak, E. J. Anderson, D. Beletsky, S. Boland, N. S. Bosch, T. B. Bridgeman, J. D. Chaffin, K. Cho, R. Confesor, I. Daloğlu, J. V. DePinto,

M. A. Evans, G. L. Fahnenstiel, L. He, J. C. Ho, L. Jenkins, T. H. Johengen, K. C. Kuo, E. LaPorte, X. Liu, M. R. McWilliams, M. R. Moore, D. J. Posselt, R. P. Richards, D. Scavia, A. L. Steiner, E. Verhamme, D. M. Wright, and M. A. Zagorski, "Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 16, pp. 6448–6452, 2013.

[33] J. D. Chaffin, D. D. Kane, K. Stanislawczyk, and E. M. Parker, "Accuracy of data buoys for measurement of cyanobacteria, chlorophyll, and turbidity in a large lake (Lake Erie, North America): implications for estimation of cyanobacterial bloom parameters from water quality sonde measurements," *Environmental Science and Pollution Research*, pp. 1–15, June 2018.

[34] C.-Y. Ahn, H.-M. Oh, and Y.-S. Park, "Evaluation of Environmental Factors on Cyanobacterial Bloom in Eutrophic Reservoir Using Artificial Neural Networks1," *Journal of Phycology*, vol. 47, pp. 495–504, June 2011.

[35] C.-Y. Ahn, A.-S. Chung, and H.-M. Oh, "Rainfall, phycocyanin, and N:P ratios related to cyanobacterial blooms in a Korean large reservoir," *Hydrobiologia*, vol. 474, pp. 117–124, Apr. 2002.

[36] Y. Jiang, B. Ji, R. N. S. Wong, and M. H. Wong, "Statistical study on the effects of environmental factors on the growth and microcystins production of bloom-forming cyanobacterium—Microcystis aeruginosa," *Harmful Algae*, vol. 7, pp. 127–136, Feb. 2008.

[37] H. A. Vanderploeg, J. R. Liebig, W. W. Carmichael, M. A. Agy, T. H. Johengen, G. L. Fahnenstiel, and T. F. Nalepa, "Zebra mussel (Dreissena polymorpha) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 58, pp. 1208–1221, June 2001.

[38] D. F. Raikow, O. Sarnelle, A. E. Wilson, and S. K. Hamilton, "Dominance of the noxious cyanobacterium Microcystis aeruginosa in low-nutrient lakes is associated with exotic zebra mussels," *Limnology and Oceanography*, vol. 49, pp. 482–487, Mar. 2004.

[39] P. J. Lavrentyev, W. S. Gardner, J. F. Cavaletto, and J. R. Beaver, "Effects of the zebra mussel (Dreissena polymorpha Pallas) on protozoa and phytoplankton from Saginaw Bay, Lake Huron," *Journal of Great Lakes Research*, vol. 21, pp. 545–557, 1995.

[40] L. B. Knoll, O. Sarnelle, S. K. Hamilton, C. E. H. Kissman, A. E. Wilson, J. B. Rose, and M. R. Morgan, "Invasive zebra mussels (Dreissena polymorpha)

increase cyanobacterial toxin concentrations in low-nutrient lakes," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 65, pp. 448–455, Mar. 2008.

[41] K. E. Kavanaugh, K. Derner, K. M. Fisher, E. Davis, C. Urizar, and R. Merlini, "Assessment of the Eastern Gulf of Mexico Harmful Algal Bloom Operational Forecast System (GOMX HAB-OFS)," July 2013.

[42] M. Beaulieu, F. Pick, and I. Gregory-Eaves, "Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set," *Limnology and Oceanography*, vol. 58, pp. 1736–1746, Sept. 2013.

[43] Z. E. Taranu, I. Gregory-Eaves, R. J. Steele, M. Beaulieu, and P. Legendre, "Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework for modelling an important health risk factor," *Global Ecology and Biogeography*, vol. 26, no. 6, pp. 625–637, 2017.

[44] J. R. Beaver, E. E. Manis, K. A. Loftin, J. L. Graham, A. I. Pollard, and R. M. Mitchell, "Land use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation," *Harmful Algae*, vol. 36, pp. 57–62, 2014.

[45] J. Shoemaker, *Method 544. Determination of Microcystins and Nodularin Iin Drinking Water by Solid Phase Extraction and Liquid Chromotography/Tandem Mass Spectrometry (LC/MS/MS).* USEPA, 2015.

[46] USEPA, *Method 546: Determination of Total Microcystins and Nodularins in Drinking Water and Ambient Water by Adda Enzyme-Linked Immunosorbent Assay.* U.S. EPA, Office of Ground Water and Drinking Water, Standards and Risk Management Division, Aug. 2016.

[47] USEPA, *Method 350.1, Revision 2.0: Determination of Ammonia Nitrogen by Semi-Automated Colorimetry.* Environmental Monitoring Systems Laboratory, Office of Research and Development, USEPA Cincinnati, OH, 1993.

[48] USEPA, *Method 353.2., Revision 2.0: Determination of nitrate-nitrite nitrogen by automated colorimetry.* Environmental Monitoring Systems Laboratory, Office of Research and Development, USEPA Cincinnati, OH, 1993.

[49] USEPA, *Method 365.1, Revision 2.0: Determination of Phosphorus by Semi -Automated Colorimetry.* Environmental Monitoring Systems Laboratory, Office of Research and Development, USEPA Cincinnati, OH, 1993.

[50] USEPA, *Method 351.2, Revision 2.0: Determination of total Kjeldahl nitrogen by semi-automated colorimetry.* Environmental Monitoring Systems Laboratory, Office of Research and Development, USEPA Cincinnati, OH, 1993.

[51] QGIS Development Team, *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2009.

[52] GRASS Development Team, *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*. Open Source Geospatial Foundation, 2017.

[53] C. Homer, C. Huang, L. Yang, B. Wylie, and M. Coan, "Development of a 2001 National Land-Cover Database for the United States," *Photogrammetric Engineering & Remote Sensing*, vol. 70, pp. 829–840, July 2004.

[54] M. J. Menne, I. Durre, B. Korzeniewski, S. McNeill, K. Thomas, X. Yin, S. Anthony, R. Ray, R. S. Vose, B. E. Gleason, and T. G. Houston, "Global Historical Climatology Network - Daily (GHCN-Daily), Version 3," *J. Atmos. Oceanic Technol.*, vol. 29, pp. 897–910, 2012.

[55] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018.

[56] H. Wickham, R. Francois, L. Henry, and K. Müller, *dplyr: A Grammar of Data Manipulation*. 2017.

[57] D. Robinson, *broom: Convert Statistical Analysis Objects into Tidy Data Frames*. 2018.

[58] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[59] B. Schloerke, J. Crowley, D. Cook, F. Briatte, M. Marbach, E. Thoen, A. Elberg, and J. Larmarange, *GGally: Extension to 'ggplot2'*. 2017.

[60] S. Garnier, *viridis: Default Color Maps from 'matplotlib'*. 2018.

[61] T. Wei and V. Simko, *R package "corrplot": Visualization of a Correlation Matrix*. 2017.

[62] P. Leifeld, "texreg: Conversion of Statistical Model Output in R to \LaTeX and HTML Tables," *Journal of Statistical Software*, vol. 55, no. 8, pp. 1–24, 2013.

[63] H. Wickham, *tidyverse: Easily Install and Load the 'Tidyverse'*. 2017.

[64] H. Zhu, *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. 2018.

[65] M. Hlavac, *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI), 2018.

[66] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[67] T. L. Miller, *leaps: Regression Subset Selection*. 2017.

[68] C. Eisenhart, "The Assumptions Underlying the Analysis of Variance," *Biometrics*, vol. 3, no. 1, pp. 1–21, 1947.

[69] M. G. Kenward, "A Method for Comparing Profiles of Repeated Measurements," *Applied Statistics*, vol. 36, no. 3, p. 296, 1987.

[70] M. J. Crawley, *The R Book*. England: John Wiley and Sons Ltd,, 2007.

[71] G. A. Codd, "Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control," *Ecological Engineering*, vol. 16, pp. 51–60, Oct. 2000.

[72] J. M. Fraterrigo and J. A. Downing, "The Influence of Land Use on Lake Nutrients Varies with Watershed Transport Capacity," *Ecosystems*, vol. 11, pp. 1021–1034, Nov. 2008.

[73] R. E. Lucas and J. F. Davis, "Relationships between ph values of organic soils and availabilities of 12 plant nutrients," *Soil Science*, vol. 92, no. 3, pp. 177–182, 1961.

[74] M. Xiao, M. Li, and C. S. Reynolds, "Colony formation in the cyanobacterium Microcystis," *Biological Reviews*, vol. 93, pp. 1399–1420, Aug. 2018.

[75] L. Yema, E. Litchman, and P. de Tezanos Pinto, "The role of heterocytes in the physiology and ecology of bloom-forming harmful cyanobacteria," *Harmful Algae*, vol. 60, pp. 131–138, Dec. 2016.

[76] B. A. Van Mooy, L. R. Hmelo, L. E. Sofen, S. R. Campagna, A. L. May, S. T. Dyhrman, A. Heithoff, E. A. Webb, L. Momper, and T. J. Mincer, "Quorum sensing control of phosphorus acquisition in Trichodesmium consortia," *The ISME journal*, vol. 6, no. 2, p. 422, 2012.

[77] T. Feng, C. Wang, P. Wang, J. Qian, and X. Wang, "How physiological and physical processes contribute to the phenology of cyanobacterial blooms in large shallow lakes: A new Euler-Lagrangian coupled model," *Water Research*, vol. 140, pp. 34–43, Sept. 2018.

[78] T. Rohrlack and P. Hyenstrand, "Fate of intracellular microcystins in the cyanobacterium Microcystis aeruginosa (Chroococcales, Cyanophyceae)," *Phycologia*, vol. 46, no. 3, pp. 277–283, 2007.

[79] L. MacKenzie, V. Beuzenberg, P. Holland, P. McNabb, and A. Selwood, "Solid phase adsorption toxin tracking (SPATT): a new monitoring tool that simulates the biotoxin contamination of filter feeding bivalves," *Toxicon*, vol. 44, pp. 901–18, Dec. 2004.

[80] J. C. Clausen, *Introduction to Water Resources*. Long Grove, IL: Waveland Press, 2018.