USING OXYGEN DEPLETION AND CHLOROPHYLL-A AS PROXIES FOR ESTIMATES OF CYANOBACTERIA BLOOMS TO CREATE PREDICTIVE LAKE ERIE HAZARDOUS ALGAE BLOOM MODELS

Brian Alexander Jaffee

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

August 2015

Committee:

Nancy Boudreau, Advisor

John Chen

Sheila Roberts

ABSTRACT

Nancy Boudreau, Advisor

This paper examines hazardous algae blooms in Lake Erie, focusing on previously created predictive statistical models, and creating different predictive models based on two proxy measurements for hazardous algae bloom occurrences – dissolved oxygen and chlorophyll-a. While prior models have used different proxies for hazardous algae blooms, including remote sensing and boat tows, the study presented here examines whether different proxies, a larger dataset, and different independent variables create valid hazardous algae bloom predictive models and/or improve upon prior forecasting methods. More specifically, since there is no single definition for hazardous algae blooms, and no one agreed upon metric to measure them, this study examines whether the chosen proxies are suitable proxies for hazardous algae blooms in Lake Erie, using linear regression and ANOVA analyses to create a number of different models. The results from these models indicate that both dissolved oxygen and chlorophyll-a are suitable proxies for hazardous algae bloom occurrences. Further, the modeling results confirmed the Lake indicators that are the greatest contributors to hazardous algae blooms, and confirmed prior research that the Lake had changed in terms of hazardous algae bloom growth and occurrence after the mid-1990s. Following these results, the paper examines the public policy response to recent blooms. Combining the results from this and prior studies, the public policy response was scrutinized, and the paper concludes that more will likely need to be done in the future to mitigate bloom occurrences and severity.

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis chair, Dr. Nancy Boudreau, who helped guide me through the thesis process and vastly improved my understanding of applied statistics. I would like to thank Drs. Sheila Roberts and John Chen for serving on my thesis committee, and especially Dr. Chen for joining my committee very late in the process. Additionally, I would like to acknowledge Dr. Douglas D. Kane from Defiance College who provided me with guidance on my project, as well as data and research regarding Lake Erie dissolved oxygen. A large thank you goes to my parents, Dr. Laura Leventhal and Alan Jaffee, and brother, Sam Jaffee, for their insights, technical knowledge, advice, positivity, and willingness to edit many drafts throughout the process. Finally, I would like to thank my friends for their patience, motivation, and constant support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

In 2014, nearly 400,000 people in and around Toledo, Ohio were warned not to drink their tap water or shower after high levels of a dangerous toxin were discovered in the water supply.[1] This toxin was caused by a large hazardous algae bloom ("HAB") caused by the blue-green algae *Microcystis*, a type of cyanobacteria. HABs have been occurring in lakes around the world for centuries, but human activities have caused these nuisance and sometimes toxic events to increase. Eutrophication, or the enrichment of a water system with chemical nutrients, typically nitrogen, phosphorus or both, of freshwater lakes resulting from increased human nutrient loading has become a global problem.[2] Eutrophic lakes that have severe toxic cyanobacteria blooms are present throughout the world including in China (Lake Taihu), Canada (Lake Winnipeg), the Netherlands (Lake Nieuwe Meer), and the United States (Lake Erie).[3] Though controlled in Lake Erie after the 1980s, HABs have become a persistent nuisance and costly drain to the economy of Northwest Ohio and Southeast Michigan. While not the largest of the Great Lakes, Lake Erie and its health are extremely important. Lake Superior, the largest Great Lake contains 50% of all water in the Great Lake system, but only 2% of the fish. On the other hand, Lake Erie contains only 2% of the water, but the Lake has 50% of the fish in the Great Lakes.[4] Additionally, Lake Erie creates $10.7 billion in economic activity, supports more

---

[1] Atkin, Emily. "7 Things You Need to Know About the Toxin That's Poisoned Ohio's Drinking Water." Climate Progress, August 3, 2014. Available at http://thinkprogress.org/climate/2014/08/03/3467068/toledo-ohio-water-crisis/.

[2] Michalak, Anna M., et. al. "Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions." PNAS, March 4, 2013. Available at http://graham.umich.edu/scavia/wp-content/uploads/2013/04/PNAS.pdf.

[3] *Id*.

[4] LaBarge, Greg, & Hoorman, Jim. "Lake Erie and Phosphorus: What has happened since 1995?" Ohio State University, 2012. Available at http://www.oardc.ohio-state.edu/ocamm/images/MTW2012_Hoorman.pdf.

than 11,000 jobs for Ohio residents, and generates more than $750 million in tax dollars for the state of Ohio.[5]

Harmful algal blooms are episodes where large quantities of harmful forms of algae appear in parts of the lake. In freshwater ecosystems they are caused by seven species of algae of a group called cyanobacteria, also referred to as blue-green algae.[6] Although small numbers of these algae are present at all times, cyanobacteria normally require warmer temperatures (maximum growth rate occur in the 25-30 degree C range) and high levels of nutrients to stimulate growth.[7] Thus, blooms of cyanobacteria are most likely to occur in the summer and early fall.[8]

HABs can be a nuisance to human enjoyment of freshwater systems, can cause health problems in humans and wildlife, and can greatly affect the ecosystem where they are present. Cyanobacteria affect the growth of other, harmless algae, which grow rapidly, but are heavy and sink when the water is calm.[9] However, when water has excessive nutrients, the cyanobacteria grow at an increased rate, and when blooms form, they can block light from other algae, growing as thick as the nutrient load will allow.[10] In terms of human health, HABs can be toxic to ingest and can cause skin irritation. Public water suppliers that get water that is affected by lakes with HABs can manage their water intake to limit the effects of HABs. This includes filtering cells before water treatment starts, and using activated charcoal and other treatments that reduce the

---

[5] *Id*.
[6] Reutter, Jeffery, et. al. "Lake Erie Nutrient Loading and Harmful Algal Blooms: Research Findings and Management Implications." Ohio Sea Gant College Program, The Ohio State University, June 14, 2011. Available at http://worldcat.org/arcviewer/1/OHI/2011/11/08/H1320770105057/viewer/file2.pdf.
[7] *Id*.
[8] *Id*.
[9] Stumpf, Richard P., "Satellite Monitoring of Toxic Cyanobacteria for Public Health." Earthzine, March 26, 2014. Available at, http://earthzine.org/2014/03/26/satellite-monitoring-of-toxic-cyanobacteria-for-public-health/.
[10] *Id*.

toxins.[11] These procedures are highly effective, but also expensive. In addition to HABs potential toxic threat, cyanobacteria blooms can form surface scums that reduce the aesthetics of recreational waters and produce chemical compounds that cause odor issues. This issue is further exacerbated by the tendency of HABs to be blown and concentrated by wind towards harbors, shorelines, docks, and other near shore areas that are likely to be frequented by the public.[12] Eliminating cyanobacteria blooms can be difficult. Reducing HABs means reducing nutrient loading, most importantly phosphorus, to the affected water system.[13]

Being able to identify and quantify the drivers behind HABs in Lake Erie has important implications for water research management, and has been a continued area of research.[14] Working together, the National Oceanic and Atmospheric Administration ("NOAA") and local academics have researched and created models to predict the likelihood and severity of future HABs. This research is not only used to predict future HAB events, but has also been used by policy makers to help create phosphorus loading target reduction goals.[15] Separately, NOAA and the University of Michigan Water Center have created independent HAB experimental models.[16] The NOAA model comes from Stumpf, et. al., and uses spring total phosphorus load to predict bloom magnitude.[17] The University of Michigan Water Center model is based on Obenour, et.

---

[11] *Id*.
[12] Wynee, Timothy T., Stumpf, Richard P, Tomlinson, Michelle C., et. al. "Evolutoin of a cyanobacterial bloom forecast system in western Lake Erie: Development and initial evaluation." Journal of Great Lakes Research, 2013. Available at, http://www.glerl.noaa.gov/pubs/fulltext/2013/20130007.pdf.
[13] Stumpf, 2014.
[14] Obenour, Daniel R., Gronewold, Andrew D., Stow, Craig A., Scavia, Donald. "Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts." Accepted Article, doi: 10.1002/2014WR015616, 2014. Available at http://graham.umich.edu/scavia/wp-content/uploads/2014/09/Obenour-et-al-2014.pdf.
[15] NOAA. "NOAA, partners predict significant harmful algal blooms in western Lake Erie this summer." 2014. Available at http://www.noaanews.noaa.gov/stories2014/20140710_erie_hab.html.
[16] *Id*.
[17] Stumpf, Richard P., Wynne, Timothy T., Baker, David B., Fahnenstiel, Gary L. "Interannual Variability of Cyanobacterial Blooms in Lake Erie." 2012. Available at http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0042444#s3.

al., and uses a Bayesian modeling framework using a gamma error distribution, along with an untransformed response, to create a model with relatively high predictive skill compared to models based on more common statistical formulations.[18] Both models, as well as the article by Michalek et. al., found that phosphorus loads must be reduced to prevent future HAB outbreaks, but also that other variables, such as water temperature, likely play a role in the creation of HABs. Climate change models are predicting that Lake Erie temperatures are likely to increase, and increased water temperatures in Lake Erie may mean that phosphorus-loading goals created recently will not be low enough to prevent future HABs.[19]

Both of these prior models, and indeed much of the research done on Lake Erie cyanobacteria blooms, suffer from a low amount of consistent data. Additionally, a single definition of HAB, both occurrence and size, has not been agreed upon. Further, the predictive/outcome variables and other lake indicators that have formed these models have been plagued by inconsistent collection and different gathering techniques. For example, the Stumpf model uses remote sensing from satellite imagery to quantify cyanobacterial blooms from 2002 to 2011, while the Obenour model uses quantitative cyanobacteria bloom estimates for the years 2002-2013. Further uncertainty in these models comes from the techniques for making these quantitative estimates, either using satellite imagery which can only examine the surface of the Lake, and phytoplankton tows, which do sample the entire water column, but have relatively limited spatial and temporal resolution.[20]

In order to further study HAB modeling in Lake Erie, and in an attempt to improve the models presented by Stumpf and Obenour, I was interested in using two proxy measurements,

---

[18] Obenour, 2014.
[19] *Id*.
[20] *Id*.

dissolved oxygen and chlorophyll-a, to model HABs in order to increase the number of dependent variable observations. Dissolved oxygen, and the phenomenon of oxygen depletion is one related effect of HABs, such that when the large algal blooms die, they sink to the bottom, decompose, and create hypoxic zones with limited dissolved oxygen. Chlorophyll-a can be used to measure the overall presence of algae, including cyanobacteria. By using data that dates before 2002, we may be able to observe changes in variable interactions from the late 1980s to the present. From my literary review, no Lake Erie HAB modeling has been done using plankton population (chlorophyll-a) or dissolved oxygen levels as proxies for HABs. I first ran ANOVA analyses over all of my dependent and independent variables with the categorical HAB variable as the factor (either no bloom or bloom). I then ran a number of multiple linear regression analyses using oxygen depletion and chlorophyll-a as proxies for the dependent variable representing cyanobacteria blooms, using independent variables including phosphorus, dissolved reactive phosphorus and nitrogen loads from two tributaries to Western Lake Erie (Maumee and Sandusky Rivers), and water temperature. After creating my models, I compared their strength against each other, and against the Stumpf and Obenour models.

This paper will begin with a background discussion on cyanobacteria blooms and their occurrences over time in Lake Erie. Next the paper discusses public policy responses to these HABs in Lake Erie, including governmental responses and best management practices adopted by the agricultural economy. Following this background discussion, the paper will examine prior research into predicting and modeling HABs. This section will focus on possible predictor variables, will also discuss in detail both the Stumpf and Obenour models, and will examine the limitations of these models due to the data issues mentioned above. Next will be a discussion of my models and results, including comparing my models to the Stumpf and Obenour papers.

Finally, this paper will examine the results of the predictive models and how they have been incorporated or ignored in the formation of current public policy towards reducing HABs in Lake Erie.

CHAPTER II: BACKGROUND

2.1 Dangers of HABs

HABs in general, as well as in Lake Erie, can have disastrous effects on the health of the water system, humans and animals, and can negatively affect near shore economies. These blooms cloud the water and reduce oxygen levels, which threaten fish and other aquatic life.[21] Cyanobacteria are not a valuable food for zooplankton and other animals, and many consider these algae to be weeds.[22] One way they hurt the health of the water systems where they are present is by outcompeting other algae that are food for zooplankton and other animals. Cyanobacteria are not eaten by organisms that support fish production, and instead compete with those organisms for energy.[23] Additionally, when the cyanobacteria in HABs die, they sink to the bottom, are decomposed by bacteria, and the decomposition process uses up oxygen supplies and contributes to "dead zones" where other organisms cannot live.[24]

In regards to human health, the threats posed by cyanobacteria, specifically the release of the toxin *microcystis*, are vast and the toxicity of this bacteria should not be underestimated. Microcystin is a liver toxin, and the consequence of ingesting a toxic dose for people or animals is liver and kidney failure. In 1996, for example, more than 50 patients at a dialysis center in Brazil died as a result of microcystin getting into the water supply.[25] [26] Animals are also greatly affected by the microcystin bacteria. Over the last several decades, hundreds of dogs have died

---

[21] National Center for Water Quality Research. "Dissolved Phosphorus From Crop Runoff: Why it is a Big Problem!" Water Quality News and Notes. Heidelberg University. August 1, 2011, available at http://www.heidelberg.edu/sites/default/files/jfuller/images/1%20Dissolved%20P-a%20Big%20Problem%2C%2008-01-2011.pdf.
[22] Reutter, 2011.
[23] *Id*.
[24] *Id*.
[25] Stumpf, 2014.
[26] Atkin, 2014.

after exposure to cyanobacteria toxins.[27] Additionally, microcystis has been known to kill livestock.[28] Concentrations of microcystin in western Lake Erie during summer blooms have been found to exceed the 1-μg/L safety limit for drinking water established by the World Health Organization, and municipalities that use the lake as a source of drinking water may spend over $100,000 per month in additional treatment costs to ensure that the water supply is safe.[29] Along with the dangers of ingesting the toxin from cyanobacteria blooms, the algae also causes skin rashes and burns when touched.[30] Additionally, contact with the water can have even more drastic effects on people with prior liver problems. During the 2014 Toledo water crisis, people with liver problems were told not to take showers or even wash their hands with affected tap water.[31]

Due to the dangers of toxins produced by HABs, their occurrence leads to negative economic effects such as beach closures. Algal blooms can threaten other income from tourism, including boating, which is important to the economies of many cities and towns in Northern Ohio.[32] Fish die offs from too little oxygen or from the toxins themselves can have harmful effects on the tourism and commercial fishing industries.

2.2 History of HABs in Lake Erie

Cyanobacteria blooms have always been present in Lake Erie, though before human activity, these blooms were much smaller in size and less frequent in occurrence. By the 1960s and 70s increases in phosphorus and nitrogen caused by wastewater treatment plants and

---

[27] Stumpf, 2014.
[28] Atkin, 2014.
[29] Carrick, Hunter J., Moon Jessica B., Gaylord, Barrett F. "Phytoplankton Dynamics and Hypoxia in Lake Erie: A Hypothesis Concerning Benthic-pelagic Coupling in the Central Basin." Journal of Great Lakes Research, 2005. Available at http://www.sciencedirect.com/science/article/pii/S0380133005703087#.
[30] Atkin, 2014.
[31] Atkin, 2014.
[32] National Center for Water Quality Research, 2011.

industry (point sources) and agricultural runoff (non-point sources) had caused frequent and large HABs to form in Lake Erie. In 1969, the estimated total phosphorus loading to Lake Erie was approximately 29,000 metric tons.[33] Research indicated that phosphorus loadings would have to be reduced to 11,000 metric tons per year in order to reduce HABs.[34] Efforts to control these HABs focused on controlling phosphorus loading into the Lake. The original management actions that were implemented were based on the relationship between total phosphorus loading and chlorophyll-a counts.[35] Work to control point sources of phosphorus to Lake Erie began in 1972 with the signing of the Great Lakes Water Quality Agreement ("GLWQA") by the United States and Canada. Roughly two-thirds of the total phosphorus loading was coming from sewage treatment plants, thus improving sewage treatment became the primary focus of reduction efforts.[36] The major controls that were implemented were the monthly average effluent limit of 1 mg/T total phosphorus on all major sewage treatment plants (plants discharging in excess of 3,800 m^3/day)[37] and the ban of phosphorus in detergents.[38] Over $8 billion (adjusted to 1990) has been spent on building and updating sewage treatment plants, and millions more has been spent on research, outreach, and implementation of new technologies to reduce external phosphorus loading to Lake Erie.[39] By 1989, 95% of the largest municipal sewage treatment plants in the Lake Erie Basin complied with the GLQWA requirements.[40] Throughout the 1980s and early

---

[33] Reutter, 2011.
[34] *Id*.
[35] Kane, Douglas D., et. al. "Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass." Journal of Great Lakes Research, September, 2014. Available at http://www.sciencedirect.com/science/article/pii/S0380133014000768.
[36] Reutter, 2011.
[37] Dolan, David M., & McGunagle, Kevin P. "Lake Erie Total Phosphorus Loading Analysis and Update: 1996-2002." J. Great Lakes Res. 31 (Suppl. 2): 11-22. Internat. Assoc. Great Lakes Res., 2005. Available at http://www.cee.mtu.edu/~nurban/classes/ce5508/2007/Readings/dolan05.pdf.
[38] *Id*.
[39] Conroy, Joseph D., et. al. "Temporal Trends in Lake Erie Plankton Biomass: Roles of External Phosphorus Loading and Dreissenid Mussels." Journal of Great Lakes Research, 2005. Available at http://www.sciencedirect.com/science/article/pii/S0380133005703075.
[40] Dolan, 2005.

1990s, the conservation policy shifted towards limiting non-point sources of phosphorus. Goals were established for reducing sediment runoff, especially from agricultural fields, and the phosphorus attached to these sediments.[41] Thus, most conservation efforts focused on erosion control and reduction.

During most years since the 1980s the measures discussed above have worked to bring external phosphorus loading to Lake Erie below the target load established by the GLQWA.[42] The controls did not show immediate results, but by the mid-1980s, declines in Lake Erie phosphorus loadings were having an obvious effect. As annual phosphorus loads decreased, total phytoplankton biomass and decreased low oxygen events suggested that Lake Erie water quality was indeed improving.[43] The restoration of Lake Erie by the early 1990s is well documented.[44] After achieving phosphorus reduction goals from point sources, and reducing non-point loads, yearly concentrations of total phosphorus and dissolved oxygen depletion rates declined significantly.[45] Animal species that had been reduced in size returned, such as burrowing mayflies, and reductions in algal biomass, especially nuisance cyanobacteria, was observed.[46]

Despite the early success from these management actions in the 1980s, hypolimnetic oxygen depletion rates, hypoxia extent, and algal biomass have increased systematically since the mid-1990s.[47] Large blooms of cyanobacteria have returned to Lake Erie and dissolved

---

[41] Daloglu, Irem, Cho, Hyung Hwa, & Scavia, Donald. "Evaluating Causes of Trends in Long-Term Dissolved Reactive Phosphorus Loads to Lake Erie." Environmental Science & Technology, American Chemical Society. September 10, 2012. Available at http://graham.umich.edu/scavia/wp-content/uploads/2012/10/Daloglu-et-al-2012.pdf.
[42] Conroy, 2005.
[43] *Id*.
[44] Dolan, 2005.
[45] *Id*.
[46] *Id*.
[47] Michalak, 2013.

oxygen depletion continues to be problematic.[48] There has been a slight increase in overall phosphorus loadings, much of which has been attributed to non-point sources.[49] This is a change from the 1970s when point sources made up the majority of phosphorus loading to the Lake, and research has shown that point sources of phosphorus declined in importance as controls were implemented until 1991, when nonpoint sources exceeded point sources in their contribution of phosphorus to Lake Erie.[50] However, even though non-point sources have increased in importance in terms of phosphorus loading, total phosphorus levels in the Lake have remained stable.[51] A separate hypothesis for increased HABs in the 1990s centers on the introduction of invasive mussel species (referred to as Dreissenids).[52] However, Dreissenid populations, as well as total phosphorus levels in Lake Erie have stabilized, and it is hypothesized that neither of these factors is significantly contributing to recent increases in HABs.[53]

In 2011, Lake Erie experienced the largest HAB in its recorded history, with a peak intensity over three times greater than any previously observed HAB.[54] Land use, agricultural practices, and meteorological conditions are all hypothesized to have contributed to stimulating and exacerbating the bloom.[55] An important component of the total phosphorus load is the bioavailable dissolved reactive phosphorus. This component of phosphorus has been increasing in Lake Erie even as total phosphorus levels remain relatively stable. (A discussion on the importance of dissolved reactive phosphorus is below.) Regardless, it is important to note that

---

[48] Kane, 2014.
[49] Dolan, D.M., & Richards, R.P. "Analysis of Late 90s Phosphorus Loading Pulse to Lake Erie." Ecovision World Monograph Series, 2008. Available at http://www.uwgb.edu/doland/GLGrants/analysis_of_late_90s_phosphorus_loading_pulse_to_lake_erie.pdf.
[50] *Id*.
[51] Michalak, 2013.
[52] *Id.*
[53] *Id*.
[54] *Id*.
[55] *Id*.

Lake Erie is a different lake today than it was when phosphorus controls first began. Climate and meteorological changes, the introduction of invasive species, and the increased proportion of dissolved reactive phosphorus could mean that current phosphorus loading goals are no longer sufficient.[56]

Historically, point sources of phosphorus were dominated by highly bioavailable dissolved reactive phosphorus, and non-point sources were dominated by particulate phosphorus that had low bioavailability.[57] However, as mentioned above, the proportion of dissolved reactive phosphorus in total phosphorus loads from non-point sources has been increasing since the 1990s. One reason for this is the focus on erosion control, and the implementation of best agricultural management practices that may today be exacerbating the problem by increasing the amount of bioavailable dissolved reactive phosphorus that is getting into the Lake. Three management practices, Fall fertilizer application, fertilizer being placed on the surface rather than injected into the soil, and conservation tillage, can create conditions for enhanced dissolved reactive phosphorus runoff.[58] These agricultural practices have increased in the Lake Erie region over the last ten years, and consistent with these trends is the observed 218% increase in dissolved reactive phosphorus loadings between 1995 and 2011.[59] This increase occurred even though runoff only increased by 42% over the same time period.[60] Planting practices by farmers may also be exacerbating the issue of increased dissolved reactive phosphorus. Farmers are

[56] Reutter, 2011.
[57] *Id*.
[58] Michalak, 2013.
[59] *Id*.
[60] *Id*.

growing more corn today than ever, and corn is a fertilizer-intensive crop, meaning more fertilizer is used on fields growing corn.[61]

The GLWQA was updated in 2012 and included calls for minimizing the extent of low oxygen dead zones in Lake Erie associated with excessive phosphorus loading and maintaining cyanobacteria biomass at levels that do not produce concentrations of toxins that are a threat to humans or the ecosystem.[62] As a key strategy, the agreement calls for review and update of prior phosphorus loading targets. However, research has found that even though these new targets may help prevent some cyanobacteria blooms, they will not be sufficient for reducing low oxygen events.[63] These loading targets were based on the relationship between HAB size and total phosphorus loads from the Maumee River that were established by the Stumpf Model (to be discussed below).[64]

[61] *Id*.
[62] Obenour, 2014.
[63] *Id*.
[64] *Id*.

CHAPTER III: PREDICTING HABS

3.1 Important Variables

This section will discuss the different predictor and outcome variables that are hypothesized to contribute to HAB formation in Lake Erie.

3.1.1 Tributaries (Maumee and Sandusky Rivers)

The Maumee River is the largest tributary to Lake Erie, and also has the largest annual discharges and phosphorus loads.[65] Excluding contributions from the upper Great Lakes, the Maumee River watershed is the single largest external source of phosphorus to Lake Erie, contributing about 35% of the total phosphorus load in 1994 for example.[66] The River's discharge is hypothesized to contribute greatly to the supply of needed nutrients to fuel the cyanobacteria blooms.[67] Over 80% of the land within the Maumee River watershed is used for agriculture, and it discharges into the shallowest portion of Lake Erie.[68] The proximity of HABs to the inflow of the Maumee River suggests that nutrients loaded from the river may influence the development of algal blooms in western Lake Erie.[69] From 1975 through the 1990s, annual phosphorus loads from the Maumee River declined, but from about 1995 onward, annual phosphorus loads increased as dissolved reactive phosphorus loads increased.[70] Dissolved reactive phosphorus loads in 2007 and 2008 were higher than any year since measurements began in 1975.[71]

---

[65] Dolan, 2008.
[66] Bridgeman, Thomas, et. al. "From River to Lake: Phosphorus partitioning and algal community compositional changes in Western Lake Erie." Journal of Great Lakes Research, March, 2012. Available at http://www.sciencedirect.com/science/article/pii/S0380133011002164.
[67] Stumpf, 2012.
[68] *Id*.
[69] Bridgeman, 2012.
[70] *Id*.
[71] *Id*.

The Sandusky River watershed is also located in northwest Ohio and drains into Lake Erie's Western basin. Roughly 77% of the land within the Sandusky River watershed is used for agriculture.[72] The Sandusky River's discharge and phosphorus loads are smaller than the Maumee's, though it is the second largest river in terms of discharge and phosphorus loads into the Western basin. Though the Sandusky River is smaller in terms of total nutrient loading, the proportion of nutrients, follow the pattern of the Maumee River.

3.1.2 Phosphorus

As mentioned above, phosphorus is considered a pollutant in surface waters because when its concentrations grow too large, it causes excessive growth of algae.[73] The analytical method most commonly used to describe phosphorus coming from both point and non-point sources is a standardized procedure that yields what is known as "total phosphorus".[74] Total phosphorus is comprised of two major portions, dissolved reactive phosphorus and particulate phosphorus. Since most particulate phosphorus settles out of the water column when it enters lakes, particulate phosphorus can also become "positionally" unavailable to algae, and indeed most phosphorus that enters lakes and oceans eventually becomes buried in sediment.[75] Phosphorus loads to Lake Erie are not distributed evenly across the basin, and this is why many researchers focus on the Maumee and Sandusky Rivers' nutrient loads. The Western basin receives approximately 60% of the average total phosphorus loads, while the Central and Eastern basins received about 30% and 10% respectively.[76] Thus, researchers have reported that

---

[72] Daloglu, 2012.
[73] National Center for Water Quality Research, 2011.
[74] Reutter, 2011.
[75] *Id*.
[76] Scavia, Donald, et. al. "Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia." Journal of Great Lakes Research, June, 2014. Available at www.sciencedirect.com/science/article/pii/S038013304000252.

phosphorus loads to the Western basin are a very important determinant of the Western basin and Central basin eutrophication response.[77]

3.1.3 Dissolved Reactive Phosphorus

Dissolved reactive phosphorus can be measured as the phosphorus that remains in the water after that water has been filtered to remove particulate matter.[78] Dissolved reactive phosphorus is of particular concern in terms of HABs because it is highly bioavailable to algae, meaning that it supports rapid algal growth and reproduction, and dissolved reactive phosphorus remains in the water while particulate phosphorus settles to lake bottoms where it is no longer available to algae.[79] Roughly 95% of dissolved reactive phosphorus is bioavailable to algae, while only 30% of particulate phosphorus is bioavailable, and even though particulate phosphorus dominates total phosphorus loading to Lake Erie from the Maumee and other Northwest Ohio rivers, dissolved reactive phosphorus contributes more bioavailable phosphorus.[80] Additionally, dissolved reactive phosphorus has become a recent focus of HAB research because the loads of dissolved reactive phosphorus entering Lake Erie have been increasing dramatically in recent years.[81] Examining data on dissolved reactive phosphorus loads over the course of the available data shows that from the 1970s through the early 1990s, dissolved reactive phosphorus declined, but then increased since the mid-1990s.[82] This trend of increased dissolved reactive phosphorus has been shown to be particularly significant in agricultural tributaries, specifically the Maumee and Sandusky watersheds.[83] In addition to its

---

[77] *Id*.
[78] National Center for Water Quality Research, 2011.
[79] *Id*.
[80] *Id*.
[81] *Id*.
[82] Daloglu, 2012.
[83] *Id*.

relationship to HABs, the rate of oxygen depletion in Lake Erie has been strongly correlated with dissolved reactive phosphorus since the mid-1990s.[84] Research has also found that while dissolved reactive phosphorus trends are affected by river flow; similar trends are reported for dissolved reactive phosphorus concentrations in rivers, indicating that changes in the loads of dissolved reactive phosphorus are not solely a function of changes in hydrology.[85]

At the same time that dissolved reactive phosphorus loads were increasing, total phosphorus loading has not increased significantly in the last 15 years.[86] This discrepancy has caused researchers to ask what has caused the increasing dissolved reactive phosphorus problem. As mentioned above, agriculture and supposed best agricultural management practices may be contributing to the problem. Only 7% of the total phosphorus load from the Maumee River to Lake Erie is attributed to point sources, and therefore, the increases in dissolved reactive phosphorus are hypothesized to be coming from non-point sources, particularly farm fields.[87] The following changes in agricultural practices and weather have been identified as contributing to the upward trends in dissolved reactive phosphorus loads. (1) Increased broadcasting of fertilizer onto the soil surface especially in winter; (2) build-up of phosphorus concentrations due to broadcast fertilizer application, crop residue breakdown on the soil surface, and the decline of mold board plowing inverts the soil; (3) fertilizer applications even when excessive phosphorus is already available in the soil; (4) soil compaction, or packing caused by equipment traffic that increases runoff; (5) excessive phosphorus concentrations on some fields receiving animal manures; (6) increased tile drainage intensity coupled with the development of channels through the soil that convey surface water directly to drainage tile, which empty into streams thereby

---

[84] *Id*.
[85] *Id*.
[86] Scavia, 2014.
[87] National Center for Water Quality Research, 2011.

bypassing stream side filter strips; and (7) more frequent storm events with large amounts of rain over a short period, giving rise to more surface run off.[88]

3.1.4 Nitrogen

Nitrogen, like phosphorus, is another nutrient that can be washed into watersheds from fertilizer runoff. However, unlike phosphorus, the role of nitrogen in the occurrence and size of HABs is less well understood. Reducing nitrogen might also be an effective remedy to HAB's but more research is needed.[89] Some research suggests that nitrogen fixing cyanobacteria could offset the effects of any nitrogen load reductions.[90] Further, as phosphorus fertilizer runs out of favor, there is a greater need for farmers to apply nitrogen fertilizer to meet crop plant requirements. Surprisingly, when controlling HABs in near shore ocean water systems, nitrogen is most frequently the limiting nutrient, and in watersheds draining directly to the ocean, nitrogen load reductions are recommended.[91] Some research has suggested that in freshwater systems, nitrogen may be potentially capable of limiting growth of freshwater cyanobacteria blooms. Evidence has shown that for Lake Erie, at times, nitrogen can limit cyanobacteria growth and thus, the role that nitrogen loading has on Lake Erie's HABs needs to be explored further.[92]

3.1.5 Temperature

Temperature is an important driver of cyanobacteria growth and the creation of HABs. Cyanobacteria require warmer temperatures to grow. Water warmer than 15 degrees Celsius is required for cyanobacteria growth, and maximum growth rates occur in the 25-30 degree Celsius

---

[88] *Id*.
[89] Reutter, 2011.
[90] *Id*.
[91] *Id*.
[92] Kane, 2014.

range.[93] These temperature requirements mean that cyanobacteria grows rapidly in the early summer, and blooms usually end in the early fall as water temperature declines and cyanobacteria colonies sink and settle on the lake floor.[94]

3.1.6 Chlorophyll-a

Total phytoplankton biomass, which includes cyanobacteria, is often measured by chlorophyll-a concentrations. Most studies of Lake Erie phytoplankton use chlorophyll-a as a surrogate for total algal abundance because chlorophyll-a concentration is easier to determine than phytoplankton biomass.[95] Like cyanobacteria and other algae, the relationship between within-lake total phosphorus concentration and chlorophyll-a concentration is well established for a variety of lakes.[96] The predicted reduction in chlorophyll-a concentration, total phytoplankton biomass, and cyanobacteria biomass with reduction in phosphorus load is also well established.[97] Further studies in Lake Erie have shown that chlorophyll-a estimates of phytoplankton biomass does correlate with phosphorus loading, but that both the form of phosphorus (total vs. dissolved reactive) and the amount of the load are important.[98] The link between dissolved reactive phosphorus and chlorophyll-a concentrations is important because in recent years (2007 onward) Lake Erie has seen some of the largest dissolved reactive phosphorus loads since data collection began in the Maumee and Sandusky Rivers, as well as some of the largest HABs in the Lake's history.[99]

---

[93] Reutter, 2011.
[94] Bridgeman, 2012.
[95] Conroy, 2005.
[96] *Id*.
[97] *Id*.
[98] Kane, 2014.
[99] *Id*.

Recent studies have conversely shown that in recent years (1996 onward) chlorophyll-a concentrations did not predict total phytoplankton biomass well.[100] However, the use of chlorophyll-a as a surrogate for total phytoplankton biomass has been argued against for some time. Studies have found that chlorophyll-a concentration was a poor predictor of total phytoplankton biomass for many different lakes.[101] In these cases, the poor fit is commonly attributed to the overestimation of chlorophyll content, underestimation of total phytoplankton biomass, or both.[102] Further, converting from cell volume to biomass is problematic due to the inclusion of vacuoles, which do not contribute much to the functional biomass of cells, especially in diatoms.[103] Despite the possible downsides of using chlorophyll-a as a measure of HABs, it is still used by researchers because the data are easy and cheap to collect, and there is a large amount of historic data available.

3.1.7 Dissolved Oxygen

The issue of low oxygen regions at the bottom of Lake Erie is not a new phenomenon, and has been occurring in the Lake naturally for thousands of years.[104] While some hypoxia (near-absence of dissolved oxygen) is likely natural, human activities in the second half of the 20th century exacerbated the rate and extent of dissolved oxygen depletion.[105] However, too many nutrients, especially phosphorus, can make the issue much worse.[106] In the summer, the water in Lake Erie separates into two layers. The top layer is warmer than the bottom one, it

---

[100] Conroy, 2005.
[101] *Id*.
[102] *Id*.
[103] *Id*.
[104] Zhou, Yuntao, et. al. "Spatial and Temporal Trends in Lake Erie Hypoxia, 1987-2007." Environmental Science & Technology, December 13, 2012. Available at http://pubs.acs.org/doi/pdf/10.1021/es303401b.
[105] Scavia, 2014.
[106] USEPA. "Oxygen Depletion in Lake Erie." Great Lake Monitoring. June 26, 2012. Available at http://www.epa.gov/glnpo/monitoring/d_o/index.html.

receives sunlight, and this warm water mixes with oxygen from the air. The cooler water on the bottom layer is darker, and is cut off from the air so it cannot re-supply its oxygen.[107] If there is excess phosphorus in the water, the nutrient acts as a fertilizer, and more algae grow in the warm, sunlit top layer. When the algae stop growing and die, bacteria and fungi then decompose the plant organic matter. The bacteria and fungi also need oxygen to live, and as they decompose the dead algae, they consume what oxygen is available at the bottom of the water column.[108] Because the bottom layer is cut off from the air, over the summer months, less and less oxygen remains in the water. If there is a small volume of water and a lot of algae decomposition, like what occurs in Lake Erie, the oxygen will be used up faster, leading to oxygen depleted areas.[109] Research has shown that the concentration of total phosphorus has likely been influencing the rate of oxygen depletion in recent years.[110] Based on a simple dissolved oxygen model, Rucinski et al. showed that change in dissolved oxygen depletion rates reflected changes in total phosphorus loads as opposed to climate changes between 1987 and 2005.[111] Similarly, Burns et al. demonstrated that dissolved oxygen depletion rate is more closely related to the previous year's annual total phosphorus load.[112] The findings for the relationship between hypoxic zone

[107] *Id*.
[108] *Id*.
[109] *Id*.
[110] *Id*.
[111] Scavia, 2014.
[112] Burns, 2005.

area and phosphorus loading is being examined in order to create new phosphorus loading targets that would minimize the hypoxic zone coverage to the greatest extent possible.[113]

3.2 Prior Models

3.2.1 Stumpf Model

One of two well-known predictive models for Lake Erie HABs, the model created by Dr. Richard Stumpf (referred to as the "Stumpf Model" in this paper) examined the relationship between the Cyanobacteria Index ("CI"), a proxy for HABs used by Stumpf, and a number of the predicative variables discussed above.

One of the major issues in HAB modeling for Lake Erie is the lack of a precise definition of an HAB (see below for discussion), and differing metrics to measure their occurrence and size. The Stumpf model uses a proxy for HABs, the CI, which is based on remote sensing. Satellite imagery is able to provide data on the areal extent of cyanobacteria blooms, and one of the most powerful instruments is the Medium-spectral resolution imaging spectrometer ("MERIS"), which permitted quantification of blooms even in water with suspended sediments, including Lake Erie.[114] MERIS data were available since 2002, allowing the Stumpf model to compare the bloom intensity with Maumee River loads for ten years (2002-2011). Using several bands in the red and the "red edge" portion of the near-infrared, MERIS data allowed for spectral shape algorithms that targeted severe blooms.[115] Spectral shape methods use a computational equivalent to the second derivative, allowing the creation of a number of indexes, such as the florescent line height, maximum chlorophyll index ("MCI") and the CI. The MCI have been

---

[113] Rucinski, Daniel, DePinto, Joseph V., Scavia, Donald, Beletsky, Dmitry. "Modeling Lake Erie's hypoxia response to nutrient loads and physical variability." Journal of Great Lakes Research, December 16, 2013. Available at http://graham.umich.edu/scavia/wp-content/uploads/2014/03/Rucinski-et-al-20141.pdf.
[114] Stumpf, 2012.
[115] *Id*.

shown to be effective in coastal ocean algae blooms with data that have not been atmospherically corrected, while the CI has been effective at identifying cyanobacteria blooms in Lake Erie.[116] The CI is an estimate of surface concentration, which includes bloom biomass during calm winds, but can underestimate bloom biomass under high winds.[117] Stumpf used the CI as opposed to measurements of the area of HABs, finding the CI to be a more robust statistic since it estimates the total biomass of the bloom.[118]

Stumpf's methods for creating predictive models were fairly straightforward. CI was compared with Maumee River water flow, total phosphorus, and dissolved reactive phosphorus using standard least squares regression, including determinations of p-values and residual standard error ("RSE") in order to determine the significance of the created models.

Stumpf found that CI is correlated with Maumee River water flow and total phosphorus for the months of March through May, but only for March and May for dissolved reactive phosphorus.[119] Examining a linear model of the cumulative load for sequential months gave more evidence for the role of flow and loads during the spring months. Stumpf found that the relationships were strongest between CI and total phosphorus load or water flow when looking at the months from March to June, which explained 89% and 97% of the variance respectively.[120] Thus, Stumpf concluded that spring months (March-June) are needed to fully explain bloom severity and variability.[121] This makes sense since cyanobacteria favor warm temperatures. The warmest water occurs in Lake Erie from July to September, meaning that the "lag" months

---

[116] *Id*.
[117] *Id*.
[118] *Id*.
[119] *Id*.
[120] *Id*.
[121] *Id*.

between nutrient loads and blooms are the warmest in the lake, which favors the development of *Microcystis* blooms.[122] However, Stumpf also found that interannual differences in summer temperatures do not explain variations of bloom intensity within his data set.[123] Additionally, Stumpf found that nitrogen loads from the Maumee River do not show a significant influence on bloom intensity, showing that the relationship between nitrogen and CI is much poorer than the relationship between phosphorus and CI.[124]

Focusing only on bloom years, Stumpf created a number of models attempting to predict CI occurrence and intensity. Using his data set, for the six years with major blooms, spring Maumee River flow produced a stronger relationship to bloom intensity than total phosphorus or dissolved reactive phosphorus loads from the Maumee.[125] Examining the relationship for spring total phosphorus and CI, Stumpf found a strong correlation between total phosphorus (r^2=.89), but a large uncertainty (RSE of 1.8 CI).[126] Contrasting this, Stumpf found that CI against the average spring Maumee flow had an RSE of .96 CI, with a much stronger correlation (r^2=.97). Stumpf further fit an exponential model of CI against Maumee water flow (log CI vs. log flow), which had the same r^2 value, but an improved RSE of .58 CI.

From this examination, Stumpf focused on two different models to predict the occurrence and intensity of HABs (using the CI measurement as a proxy). Stumpf found that spring Maumee River flow and nutrient loads explain the severity of cyanobacteria blooms in Lake Erie. The lag of up to two months between the spring nutrient loads and peak HAB biomass allows for sufficient time for recycling of total phosphorus and dispersion in the basin to support HAB

---

[122] *Id*.
[123] *Id*.
[124] *Id*.
[125] *Id*.
[126] *Id*.

growth under optimal temperature and light conditions.[127] Stumpf's two models are a CI from Maumee River flow ("Q") in an exponential model and a linear model using total spring phosphorus loads ("TP_{June}"). (see Figure 1)

Model 1	Model 2
$CI = 1.14 \times 10^{-9} \times Q^{3.8}.$	$CI = 0.39 + 0.0173 \times (TP_{June}).$

Figure 1: The Stumpf models (Note: Q=spring Maumee River flow; TP=total spring phosphorus load)

Stumpf found that the exponential model using spring Maumee River flow captures the reasonable nonlinear response at low nutrient loads.[128] Further, answering the question as to why Maumee River flow was the more effective predictor variable, Stumpf hypothesized that dissolved reactive phosphorus loads may only promote spring algal blooms near the River's mouth, and TP, which is not immediately usable, may settle, and then be recycled to useable forms by bacteria and then be dispersed in later spring river discharge.[129] It has also been argued that the Maumee River may itself be a source of cyanobacteria, which would further explain why river flow was the most significant predictor variable, though cyanobacteria seeding in the Maumee River is debated.[130] Stumpf identified several issues with his predictive models. Most importantly was the fact that there was a relatively short period of study (10 years), which raise other questions of environmental variability, such as Lake Erie freezing over or lack of ice in certain years.[131]

[127] *Id.*
[128] *Id.*
[129] *Id.*
[130] *Id.*
[131] *Id.*

3.2.2 Obenour Model

The second popular model for predicting HABs in Lake Erie is the model created by Dr. Daniel R. Obenour (referred to as the "Obenour Model" in this paper). This model examines the same dependent and independent variables as the Stumpf model, but uses a Bayesian approach.

Obenour found that while the relationships between spring Maumee River flow and total spring phosphorus loads capture the general positive correlation, they do not address the uncertainty in these relationships explicitly. Obenour assumes that this uncertainty is expected to be substantial due to multiple other factors that affect the inter-annual variability in observed bloom size, including summer wind patterns, temperature, nitrogen co-limitation, and cyanobacteria bloom measurement error.[132] Due to the small sample size of the data for quantitative cyanobacteria estimates for western Lake Erie, Obenour argued that attempting to explicitly represent all factors affecting bloom size within an empirical model would likely result in over-parameterization and poor predicative performance.[133] Due to the small sample size, Obenour noted that in cases where historical data are limited, a parsimonious approach with an explicit representation of uncertainty is more warranted.[134] Therefore, Obenour continued to use phosphorus load as the primary bloom predictor, but re-developed the load-bloom relationship within a statistical framework where uncertainty is represented quantitatively.[135] Obenour created three formulations of the phosphorus-bloom relationship using normal, log-normal, and gamma error distributions, and evaluated these based on their predictive skill.[136]

---

[132] Obenour, 2014.
[133] *Id*.
[134] *Id*.
[135] *Id*.
[136] *Id*.

Because Obenour used complex model formulation in his study, he used the more flexible Bayesian approach.[137] Obenour noted that the Bayesian approach has been demonstrated as useful in previous studies in models used to predict algal blooms and hypoxia in other lakes.[138] Obenour also used a hierarchical approach to simultaneously calibrate the model to two different HAB estimates, noting that hierarchical modeling is an effective tool for assessing multiple sources of uncertainty allowing him to account for variability due to measurement error, prediction error, and parameter uncertainty.[139] One estimate was based on the same CI used in the Stumpf model, and the other estimate was based on in situ phytoplankton tows. The observations developed from western basin phytoplankton tows are reported in units of summer bio-volume production. Obenour explains that since there was considerable uncertainty in how to integrate the two estimates over space and time, he treated them as measurements of relative bloom intensity, and the tow data was scaled directly to the CI estimates so that the means of the two datasets matched.[140] While the scaled tow estimates did not provide Obenour with any new information regarding bloom size, they did provide an independent assessment of the year-to-year variability in relative bloom size.[141]

The three regression models created by Obenour all have the same basic form:

$$z_{I,j} = \hat{z}_I + \gamma_I + \epsilon_{I,j},$$

where (for a given year *I* and observation set *j*) $z_{i,j}$ is a bloom observation, $\hat{z}_i$ is a deterministic bloom prediction based on nutrient loads, γ_i is a year-specific stochastic error term, and $\epsilon_{i,j}$ is an

[137] *Id*.
[138] *Id*.
[139] *Id*.
[140] *Id*.
[141] *Id*.

observation-specific stochastic error term.[142] Because there are multiple bloom observations for each annual bloom and the true bloom size is unknown, the distinction between year-specific errors and observation-specific errors is important. In terms of hierarchical modeling, γ_i is a yearly "random effect" that represents deterministic model error in predicting true bloom size. Obenour noted that this approach is important because it addresses interclass correlation that exists in the Stumpf Model because Stumpf's observations for a given year are not independent.[143] The three models are that Obenour created are differentiated by their deterministic form and by the probability distributions used to represent the stochastic error terms, and it is these differences that cause different predictions and predictive uncertainties.[144]

Obenour's first model, the normal model, used the common assumption of normally distributed measurement errors and random effects. The model took the following form:

$$z_{i,j} \sim N\left(\hat{z}_i+\gamma_i, \sigma_\epsilon^2\right)$$

$$\hat{z}_i = \begin{cases} \beta_b+\beta_0+\beta_w W_i+\beta_t T_i & \text{for } \beta_0+\beta_w W_i+\beta_t T_i > 0 \\ \beta_b & \text{for } \beta_0+\beta_w W_i+\beta_t T_i < 0 \end{cases}$$

$$\gamma_i \sim N\left(0, \sigma_\gamma^2\right) \quad ,$$

where the deterministic prediction, $\hat{z}_i$, is a function of the beta parameters and the weighted total phosphorus load W_i.[145] During years with small amounts of nutrient loads, such that $\beta_0 + \beta_w W_i + \beta_t T_i < 0$, the bloom size is determined to be at background level β_b, in order to prevent negative bloom predictions. The parameter β_w refers to the rate of change in bloom size per unit of phosphorus load, and β_0 is the y-intercept that varies with time, T.

---

[142] *Id.*
[143] *Id.*
[144] *Id.*
[145] *Id.*

Obenour's second model, the gamma model, is similar to the normal model in form and structure, but it assumes measurement errors and random effects are distributed according to gamma distributions. The gamma model has the following form:

$$z_{i,j} \sim Gamma\left[(\hat{z}_i+\gamma_i)^2/\sigma_c^2,\quad (\hat{z}_i+\gamma_i)/\sigma_c^2\right]$$
$$\hat{z}_i = \begin{cases} \beta_b+\beta_0+\beta_w W_i+\beta_t T_i & \text{for } \beta_0+\beta_w W_i+\beta_t T_i > 0 \\ \beta_b & \text{for } \beta_0+\beta_w W_i+\beta_t T_i < 0 \end{cases}$$
$$\gamma_i \sim Gamma\left(\hat{z}_i{}^2/\sigma_\gamma^2,\ \hat{z}_i/\sigma_\gamma^2\right)-\hat{z}_i$$
,

where $z_{i,j}$ is modeled as a gamma distribution with shape (g_α) and rate (g_β) parameters such that the mean and variance are g_α/g_β and $g_\alpha/g_\beta{}^2$ respectively.[146] The random effects are modeled as a gamma distribution that is centered at zero by subtracting $\hat{z}_i$, which makes the gamma model formulation more comparable to the normal model formulation, with the deterministic component clearly distinguished from the random effect.[147]

Obenour's third model, the log-normal model, assumes measurement errors and random effects are normally distributed, while predicting a log-transformed response. The log-normal model takes the following form:

$$\ln(z_{i,j}) \sim N\left(\hat{z}_{i;L}+\gamma_{i;L},\sigma_{c;L}^2\right)$$
$$\hat{z}_{i;L} = \beta_0+\beta_w W_i+\beta_t T_i$$
$$\gamma_{i;L} \sim N\left(0,\sigma_{\gamma;L}^2\right)$$
,

[146] *Id.*
[147] *Id.*

where $\hat{z}_{i,L}$ is the deterministic bloom prediction on the log scale that can be back-transformed to the original scale using $\hat{z}_i = e^{\hat{z}_{i,L}}$.[148] Since there the log transformation disallows negative predictions, there is no background bloom predictor as in the other models.

Obenour also probabilistically assessed the optimal loading period for predicting bloom size by having his models use a weighted total phosphorus load W_i. The weighted phosphorus load was determined by the following:

$$W_i = \frac{1}{\sum \psi_m} \sum_{m=1}^{6} w_{i,m} \psi_m$$

$$\psi_m = \begin{cases} 0 & \text{for } m \leq (\beta_\psi - 1) \\ m+1-\beta_\psi & \text{for } (\beta_\psi - 1) < m < \beta_\psi \\ 1 & \text{for } m \geq \beta_\psi \end{cases},$$

where $w_{i,m}$ is the total phosphorus load corresponding to month (m) and year (i), φ_m is the weighting value for month, which is determined by β_φ, which is the weighting parameter.[149] Unlike the Stumpf Model, Obenour uses the months January-June in his models, though he weighted this parameter such that later months (April, May, June) are more heavily weighted.

Further distinguishing his models from the Stumpf model, Obenour used Bayesian calibration. Obenour's model parameters were estimated using a Markov Chain Monte Carlo ("MCMC") implementation of Bayes Theorem. The MCMC sampling was performed in three parallel "chains" of up to 200,000 samples each, and the first half of each chain was removed as a "burn-in period", while the remaining chain portions were thinned to 1000 samples each (to

[148] *Id.*
[149] *Id.*

reduce autocorrelation), and then checked to ensure that they had converged on equivalent posterior parameter distributions.[150]

Obenour concluded that the gamma model was the preferred model of the three. Of the three models the normal and gamma models had similar R^2 values (.93 and .91 respectively), while the log-normal performed less well in terms of R^2 (.68).[151] To differentiate between the normal and gamma models a cross validation was run, and Obenour found that the difference in skill between the full model and cross validated model results was smallest for the gamma model, suggesting that it was more robust.[152] Using gamma models to explain water quality sampling data has been used in the past, however, this may be the first study to apply the gamma distribution in order to characterize predictive uncertainty in a water quality forecasting model.[153] Based on the gamma model, Obenour found that there is a linkage between total phosphorus load and HAB size. The study also created forecasting models using dissolved reactive phosphorus and nitrogen, though these were not as strong at predicting bloom size as the total phosphorus variable. This suggests, and is consistent with the Stumpf model, that total phosphorus is the most effective predictor of HABs.[154] Perhaps more interesting than these findings were Obenour's findings that based on his modeling, the threshold loading rate of phosphorus necessary for a bloom to exceed background levels has dropped between 2002 and 2013.[155] This finding is consistent with other research (particularly Michalak et al.), suggesting that large HABs may be increasingly common in the future due to changing meteorological

[150] *Id*.
[151] *Id*.
[152] *Id*.
[153] *Id*.
[154] *Id*.
[155] *Id*.

conditions that promote cyanobacteria growth. Finally, Obenour's results were largely consistent with the Stumpf model.

3.3 Limitations

A number of issues make statistical analysis of HABs difficult. First, the data sets available to quantify Microcystis booms and environmental factors that may predict HABs have not been collected long term. The Stumpf model uses ten years of data, while the Obenour model uses twelve. Additionally, even defining an HAB or quantifying Microcystis blooms can be difficult. There are currently five common methods of quantifying HABs in Lake Erie: microcystin concentration, chlorophyll-a concentration, cell counts, quantitative real-time PCR ("qPCR"), and remote sensing.[156] Each method has advantages and disadvantages. In terms of microcystin concentration, toxin concentration may vary with the percentage of toxic vs. non-toxic Microcystis strains, making it not well correlated with cell counts, which means that it is not always a reliable surrogate for Microcystis abundance.[157] Chlorophyll-a concentration is often used as a proxy for total phytoplankton mass, but as discussed above, it cannot differentiate between the different types of taxonomic groups.[158] Cell counts are a direct measurement of abundance, but the methods to obtain cell counts are time-consuming, which limits the number of measurements that can be obtained.[159] The method referred to as qPCR has advantages in being able to distinguish between toxic and non-toxic species, but this technique requires substantial investment in analytical equipment and training.[160] Finally, remote sensing,

---

[156] Ho, Jeff C., & Michalak, Anna M. "Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie." Journal of Great Lakes Research, 2015. Available at, http://www.sciencedirect.com/science/article/pii/S0380133015000027.
[157] *Id*.
[158] *Id*.
[159] *Id*.
[160] *Id*.

particularly via satellite-based sensors, is useful in being able to delineate broad areas of surface blooms over time, however, the results from satellite images are highly dependent on the concentration of gas vacuoles within the cells of the cyanobacteria, and may also vary with the presence of other algal species that have the same pigments.[161] In Lake Erie, all of these methods have been used.

Beyond the issues of how to quantify microcystis, is perhaps the more important question of how to define what exactly an HAB is. This includes identifying a bloom and whether or not it is harmful; how to define the occurrence of an HAB; how to define the size of an HAB; and how to determine the lifetime (start, peak, decline) of an HAB. Generally, an HAB is defined by its potential harm to humans and/or the ecosystem.[162] Some work has explored what criteria algal species need to meet to be categorized as harmful, the abundance threshold that define an HAB, and the diversity of pathways that can lead to the occurrence of an HAB of a particular species.[163] Further, some groups have made a distinction between harmful booms as those HABs that can adverse health effects, and nuisance blooms as those linked to a more general class of harm.[164]

The many different measurements, metrics, and definitions suggest that what constitutes an HAB is not straightforward. As a result, conclusions between different studies using different metrics may not be immediately comparable.[165] We can see, therefore, that the simple question of what is an HAB, continues to be one that is subjective, even for a system like Lake Erie where

[161] *Id*.
[162] *Id*.
[163] *Id*.
[164] *Id*.
[165] *Id*.

the primary species of interest is known.[166] Some groups have selected subsets of harms and metrics, as well as specified thresholds, to define HABs from an operational standpoint. For example the International Joint Commission ("IJC"- formed as part of the GLWQA) and the Ohio Environmental Protection Agency ("Ohio EPA") have both defined HABs based on thresholds for microcystin, chlorophyll-a, and cell concentrations.[167]

Though thresholds have been created for certain groups, diversity among definitions still exists, and highlight that the creation of a single definition remain elusive.[168] The diversity of definitions and metrics can be problematic, especially in terms of creating predictive models. For instance, if inferences based on measurements of microcystis bio-volume differ from those based on cyanobacteria biomass, then mechanistic models validated against one of the other metric could yield substantively different results.[169] Answers to questions about HAB occurrence, size, and timing are found to strongly depend on the types of measurements used to support the analysis, and this dependence occurs due to variations in sampling frequency, temporal coverage, and thresholds for harm.[170]

---

[166] *Id*.
[167] *Id*.
[168] *Id*.
[169] *Id*.
[170] *Id*.

CHAPTER IV: METHODS & RESULTS

4.1 Rationale For Study

In order to further study HAB modeling in Lake Erie, I was interested in using two proxy measurements, oxygen depletion and chlorophyll-a counts, for HABs in order to increase the number of dependent variable observations. Both the Stumpf and Obenour models suffer from relatively small data sets (2002-2011, 2002-2013 respectively), which may fail to adequately model what has occurred in Lake Erie over the last 25 years. Generally, I am interested in seeing if different proxies for HABs, a larger data set, and different independent variables (neither Stumpf nor Obenour included temperature or Sandusky River variables) create valid HAB predictive models and/or improve upon the prior forecasting methods. More specifically, since there is no single definition of an HAB, and no one agreed upon metric to measure them, my study examines whether my proxies, oxygen depletion and chlorophyll-a, are suitable proxies for HABs in Lake Erie. Oxygen depletion is one related effect of HABs, such that when the large algal blooms die, they sink to the bottom, decompose, and create hypoxic zones with limited dissolved oxygen. Chlorophyll-a can be used to measure the overall presence of algae, including cyanobacteria. By using data that dates beyond 2002, we may be able to observe changes in variable interactions from the late 1980s to the present. From my literary review, no Lake Erie HAB modeling has been done using plankton population (chlorophyll-a) or oxygen depletion levels as proxies for HABs. I ran a number of multiple linear regression analyses using oxygen depletion and chlorophyll-a as proxies for the dependent variable representing cyanobacteria blooms, using independent variables including phosphorus, dissolved reactive phosphorus and nitrogen loads from two tributaries to Western Lake Erie (Maumee and Sandusky Rivers), and water temperature. I also ran ANOVA analyses using categorical data from the Ho & Michalak

paper, "Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie," as a dependent variable.

4.2 Data Collection

For my models I used seven independent variables and two dependent variables. My independent variables were total phosphorus, dissolved reactive phosphorus, and nitrogen loads from the Maumee and Sandusky Rivers, and water temperature. My dependent variables were chlorophyll-a and oxygen depletion, and were studied individually through multiple linear regression models.

The data for the two tributaries to Lake Erie were taken from the National Center for Water Quality Research at Heidelberg University.[171] Beginning in 1974, the Water Quality Laboratory began collecting and analyzing data for a set of Lake Erie Tributaries.[172] Data including total phosphorus, total dissolved reactive phosphorus, and total nitrogen concentrations have been collected in the Maumee and Sandusky Rivers since the late 1970s. A refrigerated auto sampler was used to collect samples in these two tributaries, with samples taken three times per day.[173] Total phosphorus and dissolved reactive phosphorus are analyzed using EPA Method 365.1 and nitrogen is analyzed using EPA Method 300.1.[174] Using the template of prior research, including both Stumpf and Obenour, I converted these concentrations into total loads from March-June during the years 1987-2011 using the following method recommended by the

[171] Heidelberg University. National Center for Water Quality Research. 2015. Available at http://www.heidelberg.edu/academiclife/distinctive/ncwqr.
[172] Heidelberg University. National Center for Water Quality Research. 2015. Available at http://www.heidelberg.edu/academiclife/distinctive/ncwqr/data.
[173] *Id*.
[174] *Id*.

Heidelberg program and also used in prior research. I used the following equation to both convert the data from concentration to load and to convert from standard to metric:

$$time * \frac{volume}{time} * \frac{amount}{volume} * coversion\ factor = load$$

This conversion from concentration to load and the conversion from standard to metric units was done for each of the three independent river variables, for both tributaries that I examined. There were missing values for the Sandusky River variables for one year, and these missing data points were estimated using the iterative regression approach, which is discussed below.

For the independent variable water temperature, I used average annual summer (May-August) temperatures of Lake Erie collected and published by NOAA.[175] The water temperatures are collected from a water treatment plant at a depth of 30 feet.

The dependent variable and possible HAB proxy chlorophyll-a was collected from data published by the US EPA.[176] The US EPA began collecting annual indicators in the Great Lakes in 1983 as a way to assess the overall health of each Lake. The sampling method is to collect water and biota samples at specific water depths from a limited number of locations in each lake two times every year. Chlorophyll-a concentrations are measured as ug/L. There were three missing years for the EPA chlorophyll-a data, 1989, 1994, and 1995, and the method of means approach was used to estimate these missing variables, which will be discussed below.

[175] NOAA. "National Weather Service Forecast Office." 2015. Available at http://www.erh.noaa.gov/buf/laketemps/laketemps.php.
[176] USEPA. "Limnology Program." 2012. Available at http://www.epa.gov/grtlakes/monitoring/limnology/index.html.

The dependent variable and possible HAB proxy dissolved oxygen was collected from data published by Scavia et al.[177] The data published by Scavia et al. was compiled from the Great Lakes National Program Office and Environment Canada, and covered the years 1987 to 2011. Annual mean dissolved oxygen concentrations, measured as mg/L, was collected at fixed water stations in Lake Erie from August to September of each year.

Finally, for the ANOVA analysis and attempted logistic regression models I used a categorical dependent variable that represented HAB occurrence. These data came from the Ho & Michalak paper, "Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie," and the authors compiled evidence as to whether there was no bloom, an ambiguous bloom, or a widespread bloom for the years 1995 to 2011.[178] The evidence for these categorical data was gathered from a literature review of prior Lake Erie HAB studies, with the authors cautiously noting that not all of the studies were done at the same time of year, and that the studies were looking at a variety of issues with regards to HABs.

Briefly, there were a number of issues with my original full dataset, most notably the missing values for dependent variable chlorophyll-a and independent variables Sandusky River total phosphorus, Sandusky River total dissolved reactive phosphorus, and Sandusky River total nitrogen. Two commonly used methods for estimating missing values, called imputation, are the method of means and iterative regression.[179] The method of means approach is straightforward – the mean of the variable is substituted for the missing value.[180] The iterative regression approach is more complicated. In the iterative regression method, one variable with missing values is

---

[177] Scavia, 2014.
[178] Ho, 2015.
[179] Rencher, Alvin C., & Christensen, William F. "Methods of Multivariate Analysis." John Wiley & Sons, Inc, Publication, Hoboken, New Jersey. 2012.
[180] *Id*.

estimated by regressing it on the other independent variables to obtain a prediction equation.[181] The missing observations are then estimated from this prediction equation. Then a second variable with missing values is selected and regressed on the other independent variables to obtain a prediction equation. This prediction equation is then used to predict the missing values for the second variable. This process is repeated for all variables with missing values. Once all the missing values are estimated, the entire process is repeated and the new prediction equations are used to re-estimate the missing values. This process is continued until the estimated values for the missing variables stabilize.[182] The iterative regression approach has been found to yield better results than the method of means if the variables are somewhat related to each other. However, if the other variables are not highly correlated with the variable containing missing values, the regression technique is essentially the same as the methods of means approach.[183] The downside of having to estimate values for missing variables is the loss in degrees of freedom. This is problematic for my dataset due to the small number of observations.

As mentioned above, I used both the method of means approach and the regression approach to estimate different missing values in my dataset. Initially, I examined the correlations among the variables of the full dataset with missing values. Based on the correlations, I chose to use the method of means approach to estimate missing values of the variable chlorophyll-a, since this variable had low correlations with other variables, and because I did not want to use the independent variables from the model to estimate the missing values of the dependent variable. I chose to use the regression approach to estimate missing values for the Sandusky River variables. Using the method of means approach, the mean for the chlorophyll-a variable was

[181] *Id*.
[182] *Id*.
[183] *Id*.

calculated to be 6.0590909, and this value was then substituted in for years 1989, 1994, and 1995. I next ran regression models to create predictive estimates for the 3 variables with missing values. Each regression model had as the dependent variable, one of the variables with missing observations, and included the rest of the independent variables as predictor variables. Using the predicted values from the regression models resulted in the following estimates: 106.1636 for Sandusky River total phosphorus; 16.5619 for Sandusky River total dissolved reactive phosphorus; and 3235.28 for Sandusky River total nitrogen.

4.3 Model Choices

I chose to use these data for the years 1987 through 2011 in order to expand the sample size beyond the 10 or 12 years used by Stumpf and Obenour. As mentioned in the "Limitations" section above, there is no one definition or metric for HABs, and similarly, there is no one metric nor is there one consistent data collection method used for any of my variables except for the river loading data. Data from a number of different sources were considered, but eventually discarded mostly because of inconsistent collection years or small sample sizes. For example, dry weight biomass of crustacean zooplankton in Western Lake Erie was considered as a proxy for HAB, but the data collected were inconsistent (covering the years 1970, 1974-75, 1984-87, 1995-2000). Similarly, I considered using spatial average phytoplankton wet weight biomass (all taxa combined) as a proxy for HAB, but again, the data collected were inconsistent (1970, 1978, 1984-87, 1991, 95-97). Several different data sets concerning dissolved oxygen were considered, however many of these data sets focused on repeated measures over a small time period, and were not appropriate for my study. Finally, a number of solid data sets with good historic data were considered for phytoplankton population estimates, dissolved oxygen, and cyanobacteria

concentrations, but many of these did not extend beyond the late 1990s or early 2000s, and thus I would not have been able to model recent changes in Lake Erie.

In order to explore whether chlorophyll-a concentrations or dissolved oxygen were suitable proxies for HABs, I created a number of models to test their relationships to the independent variables described above. Additionally, I manipulated my dataset in two different ways, both to improve the predictive power of my models, as well as to better understand the system. The first set of models that I created used the entire timespan of my dataset (1987-2011) to create two predictive models, one for each dependent variable. The second set of models that I created used a 16-year dataset (1996-2011), which not only allowed me to create another set of predictive models for each dependent variable, but also allowed me to examine the hypothesis that Lake Erie had changed during the mid-1990s in terms of HAB creation and occurrences. Additionally, but running models over the 16-year dataset, I avoided using any estimates for missing values for the dependent variable chlorophyll-a.

I began by running ANOVA analyses for each of my variables, exploring the mean values of each variable as they relate to the categorical dataset. Next, I ran a simple linear regression analysis, examining both of my dependent variables against only Maumee total phosphorus using both the full and 16-year datasets. These models allowed me to directly compare my dependent variables to the Strumpf and Obenour models, which similarly used only Maumee River total phosphorus as an independent variable. Next, I used all of the independent variables discussed above to create regression models for each of my dependent variables using both datasets. This regression analysis followed the following procedure: 1. Initial examination of variable scatter plots; 2. Create the first-order full model; 3. Use the Selection function in SAS to determine the variable combination that produces the model with the largest adjusted R^2 value;

4. Run diagnostics and analyze linear regression assumptions; 5. Make variable transformations as necessary to satisfy model assumptions; 6. Drop insignificant independent variables from model. Following this, I used the information from the one-way ANOVA analyses for each of my variables, using the information obtained to check the validity of the models created through the linear regression analysis. All of my analyses were done using the SAS 9.3 and Minitab software. The next sections will describe each of the models I created or attempted to create, as well as further analysis, in some detail.

4.4 ANOVA Analysis

In order to gain a better understanding of how my chosen dependent and independent variables relate to the categorical HAB variable from the Ho and Michalak paper, I began my analysis by examining one-way ANOVAs for each of my variables, using my variables as the dependent variables and the categorical data as the factor. In the ANOVA analysis, I ignored years where Ho and Michalak found bloom presence to be ambiguous and focused only on years where there was no bloom or a bloom presence. The ANOVA analysis can tell us the means of each variable associated with each level of the categorical dataset. Thus, we can find the mean value for each variable that is associated with the categorical no bloom or bloom presence data. These means can show us trends associated with each variable, specifically telling us if higher or lower values of a particular variable are associated with the absence or presence of blooms.

Variable	Mean No Bloom	Mean Bloom Presence	F-value	p-value
Chlorophyll-a[23]	1.416	1.5678	1.09	0.327
Dissolved Oxygen	4.28	3.3	0.91	0.368
MauTP	1031	1398	1.49	0.257
MauTSRP	145.9	270.6	4.91	0.057
MauTN	23015	18559	1.26	0.294
SanTP	252.6	316.2	0.30	0.599
SanTSRP	22.69	53.6	2.25	0.172
SanTN	4249	3674	0.57	0.472
Temp	62.656	64.203	2.39	0.161

Table 1: One-way ANOVA Analyses for individual variables

Table 1 contains the results for each individual variable's one-way ANOVA. Unfortunately none of the ANOVA tests indicated a statistically significant difference between the means at the .05 level. However, we can still learn something by looking at the trends of the means. For variables chlorophyll-a, Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Sandusky total phosphorus, Sandusky total dissolved reactive phosphorus, and temperature, higher values correspond with a higher chance of an HAB presence. On the other hand, for variables dissolved oxygen, Maumee total nitrogen, and Sandusky total nitrogen, lower values correspond with a higher chance of an HAB presence. These results are consistent with what we would expect based on prior research. See tables 15-23 for more detail.

4.5 Simple Linear Regression Analysis

So that I could directly compare my two dependent variables, chlorophyll-a and dissolved oxygen, with the other HAB proxies used in the prior models, I decided to run my dependent

variables in a linear regression analysis using only Maumee River total phosphorous as the independent variable. The results for each dependent variable in each of my two datasets are below.

4.5.1 Full dataset – dependent variable, Chlorophyll-a

The simple linear regression model with dependent variable chlorophyll-a and independent variable Maumee total phosphorus using the full dataset had an F-value of 1.85 and a p-value of .187. The R^2 value was .0746 and the adjusted R^2 value was .0343. I used the normal probability plot to check the assumption of normally distributed residuals, and found that the residuals were not normal as the residuals on the normal probability plot did not form a straight line (see Figure 2). Further, looking at the plot of residuals versus fitted values, I found that the equal variance assumption was not satisfied as the residuals were not evenly spaced above and below the zero line (see Figure 2).

In an attempt to achieve the model assumptions I tried a variable transformation on the dependent variable. I chose to transform the variable to normality to see if doing so would achieve the model assumptions. Using the Box-Cox procedure to find an appropriate transformation suggested taking the dependent variable to the power of 0.23. Using the normally transformed dependent variable and running a new regression analysis against independent variable Maumee total phosphorus had an F-value of 1.29 and a p-value of .268. The model had an R^2 value of .053 and an adjusted R^2 value of .0119. Checking the normal probability plot and the residual plot showed that the model assumptions of normality and equal variance were satisfied using the transformed dependent variable (see Figure 3). Briefly, the high p-value indicates there is not a significant linear relationship between Maumee total phosphorus and

chlorophyll-a, and therefore this model is not appropriate to predict chlorophyll-a. This is confirmed by the low R^2 and adjusted R^2 values.

Model	F-value	p-value	R^2	adj. R^2
$Chol^{.23}$=1.323+0.000129*MauTP	1.29	0.268	0.053	0.0119

Table 2: Chlorophyll-$a^{.23}$ vs. MauTP; full dataset

4.5.2 Full dataset – dependent variable, Dissolved Oxygen

The simple linear regression model with dependent variable dissolved oxygen and independent variable Maumee total phosphorus using the full dataset had an F-value of 1.21 and a p-value of .283. The R^2 value was .0499 and the adjusted R^2 value was .0086. I used the normal probability plot to check the assumption of normally distributed residuals, and found that the residuals were slightly off normal (see Figure 4). Further, looking at the plot of the residuals versus the fitted values, I found that the equal variance assumption was satisfied (see Figure 4). I attempted a number of transformations, but they were not able to better satisfy the assumptions, so we use the untransformed dependent variable. Again, the high p-value indicates that there is not a significant linear relationship between dissolved oxygen and Maumee total phosphorus, and therefore this model is not appropriate to predict our dependent variable, dissolved oxygen. This is confirmed by the low R^2 and adjusted R^2 values.

Model	F-value	p-value	R^2	adj. R^2
HypDO=2.786+.000661*MauTP	1.21	0.283	0.0499	0.0086

Table 3: HypDO vs. MauTP; full dataset

4.5.3 16-Year dataset – dependent variable, Chlorophyll-a

The next two simple linear regression analyses use the 16-year dataset. This simple linear regression model with dependent variable chlorophyll-a and independent variable Maumee total phosphorus had an F-value of 0.66 and a p-value of .43. The R^2 value was .0451 and the adjusted R^2 value was 0. I used the normal probability plot to check the assumption of normally distributed residuals, and found that the residuals were not normal (see Figure 5). Further, looking at the plot of the residuals versus the fitted values, I found that the equal variance assumption was satisfied (see Figure 5).

In an attempt to achieve the model assumptions I tried a variable transformation on the dependent variable. I again chose to transform the dependent variable to normality to see if doing so would achieve the model assumptions. I used the same transformation as in the full dataset, and transformed chlorophyll-a by taking it the power of .23. Using the normally transformed dependent variable and running a new regression analysis against independent variable Maumee total phosphorus had an F-value of 0.75 and a p-value of .401. The model had an R^2 value of .0509 and an adjusted R^2 value of 0. Checking the normal probability plot and the residual plot showed that the model assumptions were satisfied using the transformed dependent variable (see Figure 6). We see that this model has a high p-value, indicating that there is not a significant linear relationship between chlorophyll-a and Maumee River total phosphorus using the 16-year dataset. The extremely low R^2 and adjusted R^2 values confirm that this model is not appropriate to predict chlorophyll-a.

Model	F-value	p-value	R^2	adj. R^2
$Chol^{.23}$=1.331+.000138*MauTP	0.75	0.401	0.0509	0

Table 4: $Chol^{.23}$ vs. MauTP; 16-year dataset

4.5.4 16 Year dataset – dependent variable, Dissolved Oxygen

The simple linear regression model with dependent variable dissolved oxygen and independent variable Maumee total phosphorus using the 17-year dataset had an F-value of 014 and a p-value of .710. The R^2 value was .0102 and the adjusted R^2 value was 0. I used the normal probability plot to check the assumption of normally distributed residuals, and found that the residuals were slightly off normal (see Figure 7). Further, looking at the plot of the residuals versus the fitted values, I found that the equal variance assumption was possibly violated (see Figure 7). I attempted a number of transformations, but they were not able to better satisfy the assumptions, so we use the untransformed dependent variable. Again, the high p-value indicates that there is not a significant linear relationship between dissolved oxygen and Maumee River total phosphorus, and therefore this model is not appropriate to predict our dependent variable, dissolved oxygen. This is confirmed by the extremely low R^2 and adjusted R^2 values.

Model	F-value	p-value	R^2	adj. R^2
HypDO=3.244+.000276*MauTP	0.14	0.710	0.0102	0

Table 5: HypDO vs. MauTP, 16-year dataset

4.6 Multiple Linear Regression Analysis

4.6.1 Full dataset – dependent variable, Chlorophyll-a

Prior to running the multiple linear regression analysis I took a number of preparatory steps to ensure my data were appropriate for the regression model. My first step was to reexamine the correlations among the independent variables, as highly correlated independent variables may lead to multicollinearity issues later. For this data set, there were a number of highly correlated variables, specifically, the total phosphorus measurements within each river, and the variable shared between the two rivers. This would likely lead to multicollinearity issues,

but I examined this after I created my initial and reduced models. I next performed an initial examination, plotting my dependent variable against my independent variables (see Figure 8). Based on these plots, it did not appear that there was a strong linear relationship between chlorophyll-a concentrations and any of my predictor variables. However, the plots did not indicate that a second order or higher model was appropriate either. I fit a model with all seven independent variables which resulted in $F = 3.07$ and $p = .028$. This indicated that at least one of the independent variables was significant. I did a quick check of the parameter estimates' variance inflation factor ("VIF") as a check for multicollinearity. At this stage in the analysis, a number of the variables had large enough VIFs (>10) to indicate that multicollinearity was present.

My next step utilized the Selection function in SAS, which is a tool for model selection. I used the selection function in SAS to find the subset of independent variables that creates the best model for predicting the dependent variable based on the models' adjusted R^2 value. The R^2 value measures the proportion of the variance in the dependent variable that is explained by the relationship between the dependent and independent variables, and the adjusted R^2 value adjusts R^2 such that it penalizes the addition of extraneous predictors to the model. The selection function in SAS fits all one variable models, all two variables models, and continues in this fashion until all combinations of variables have been tested, with the output showing which combination resulted in the highest adjusted R^2 value. Using this function, I found that a model with the variables Sandusky total dissolved reactive phosphorus, Sandusky total nitrogen, and temperature would result in the model with the largest possible adjusted R^2 value. I next created a reduced model using these variables, which had an F value of 6.47, a p-value of .004 after adjusting for the estimated missing values, an R^2 value of .5332, and an adjusted R^2 value of

.4665. Examining the parameter estimates, Sandusky total phosphorus and Sandusky total nitrogen were found to be significant variables at the .05 alpha level, but temperature was not significant.

After creating the reduced model I checked the model assumptions, namely the assumptions that the residuals are normally distributed, that there is equal error variance, that there is no multicollinearity, and that there are no outliers. Using the normal probability plot to check for normality, I found that the residuals were not normally distributed (see Figure 11). Examining the plot of the residuals versus the predicted values showed that there may be an issue with the equal variance assumption (see Figure 11). Based on VIF values there were no issues of multicollinearity in the reduced model.

In an attempt to achieve the model assumptions a number of transformations on the dependent variable chlorophyll-a were attempted. Previously I had found that the dependent variable was not normally distributed, and to normalize this variable I took it to the power of .23. Taking this transformation normalized the dependent variable, and when the model was rerun, the residual normality assumption was satisfied. Also, the residual plot showed much better scattering (see Figure 12).

After making the transformation and analyzing their results, the final model I chose for this dataset was the model with the dependent variable chlorophyll-a transformed to normality by taking the dependent variable to the power of .23, with independent variables Sandusky total dissolved reactive phosphorus and Sandusky total nitrogen. Even though the Selection function found the model with the highest adjusted R^2include the variable temperature, temperature was found to be insignificant and was dropped from the final model. The final model had an F-value of 5.42, a model p-value of .016 after adjusting for the estimating the missing values, an R^2 value

of .3678, and an adjusted R^2 value of .3103. Both independent variables were found to be significant.

Model	F-value	p-value	R^2	adj. R^2
DV: $Chol^{.23}$ final reduced	5.42	0.016	0.3678	0.3103

Table 6: $Chol^{.23}$; full dataset; final reduced model

The final reduced model had the following form:

$$Chol^{.23} = 1.55829 + .00672 * SanTSRP - .00009713 * SanTN$$

Based on the normal probability plot the normality assumption was satisfied by this model. Similarly, based on the residual plot, the equal variance assumption was also satisfied. Examining VIF values indicated no issues of multicollinearity.

Next I briefly interpret the chosen model, which used the dependent variable chlorophyll-a transformed to normality, with independent variables Sandusky River total dissolved reactive phosphorus and Sandusky River total nitrogen. To begin with, prior research indicates that higher values of the dependent variable chlorophyll-a reflect higher chances of a HAB. This is confirmed previously in the ANOVA analysis. We can look at the chosen model to determine what causes the chlorophyll-a to increase. To begin with, we interpret the independent variable coefficients. These coefficients represent the mean change in the dependent variable for one unit change in the independent variable while holding other independent variables in the model constant. The positive coefficient for Sandusky total dissolved reactive phosphorus indicates that as this variable increases, the dependent variable also increases. Similarly, the negative coefficients for Sandusky total nitrogen indicate that as this variable increases, the dependent variable decreases. From prior research we would expect that increases in dissolved reactive

phosphorus and decreases in nitrogen would lead to HAB growth since phosphorus acts as a growth stimulant for the cyanobacteria algae, while nitrogen decelerates growth.

4.6.2 Full dataset – dependent variable, Dissolved Oxygen

I followed a similar procedure for my second model, which examined the predictor variables against the dependent variable, dissolved oxygen, using the full dataset. To begin, when I ran my initial examination and plotted the dependent variable against the predictor variables, none of the plots showed a strong linear trend, though the plots did not suggest using a different model (see Figure 9). I first ran the full model with all independent variables included which resulted in $F = 3.20$ and $p = .024$. This indicated that at least one independent variable was significant. Multicollinearity was checked, showing three variables with VIF values greater than 10, which indicated issues of multicollinearity.

I used the Selection function in SAS to find the best model for predicting the dependent variable based on adjusted R^2. Using this function I found that a model with the variables Maumee total phosphorus, Sandusky total phosphorus, and Sandusky total dissolved reactive phosphorus would result in the largest adjusted R^2 value. Using these variables to create a reduced model resulted in an F value 8.39, model p-value .001, an R^2 value of .5451, and adjusted R^2 value of .4801. Looking at the parameter estimates, variables Sandusky total dissolved reactive phosphorus, and Sandusky total phosphorus, were found to be significant, but Maumee total phosphorus was not significant.

After creating the reduced model I checked the model assumptions that the residuals are normally distributed, there is equal error variance, there is no multicollinearity, and that there are no outliers. Examining the normal probability plot indicated that the residuals were close to being normally distributed, and the residual plot indicated that the equal variance assumption

was also satisfied (see Figure 15). Based on the VIF values for the variables, multicollinearity was not present.

Independent variable Maumee River total phosphorus was found to be insignificant in the reduced model, so the final reduced model was created using untransformed dependent variable dissolved oxygen with independent variables Sandusky River total phosphorus and Sandusky River total dissolved reactive phosphorus. This model had an F-value of 9.50, a model p-value of .001 after adjusting for the estimated missing values, an R^2 value of .4749, and an adjusted R^2 value of .4272. Both independent variables were found to be significant.

Model	F-value	p-value	R^2	adj. R^2
DV HypDO final reduced	9.50	0.001	0.4749	0.4272

Table 7: HypDO; full dataset; final reduced model

The final reduced model had the following form:

$$HypDO = 2.241 + .01266 * SanTP - .0525 * SanTSRP$$

Based on the normal probability plot and the residual plot, both the normality and equal variance assumptions appeared satisfied (see Figure 16). Examining the VIF values showed no issues of multicollinearity.

Again, we can interpret the chosen reduced model for the dependent variable dissolved oxygen. Prior research indicated that lower levels of dissolved oxygen would be present during seasons with HAB occurrences. This is confirmed by the previous ANOVA analysis. Looking at the independent variable coefficients we see that the coefficient for Sandusky total dissolved reactive phosphorus is negative, while the coefficient for Sandusky total phosphorus is positive. This would indicate that as Sandusky total dissolved reactive phosphorus increases, the

dependent variable decreases. On the other hand, the model indicates that as Sandusky total phosphorus increases, dissolved oxygen also increases. Some of these results make sense, while others are more troubling. What is concerning are the coefficients for the river variables. From the literary review, we would expect dissolved oxygen values to decrease as phosphorus (both total phosphorus and dissolved reactive phosphorus) loads increase. However, the model indicates that increases in the dissolved reactive phosphorus load from the Sandusky decreases dissolved oxygen levels, but as total phosphorus from the Sandusky increases dissolved oxygen also increases. This discrepancy could be due to a number of factors. First, it may be that total phosphorus from the Sandusky River includes more phosphorus that is not soluble or usable by algae, or that more of this total phosphorus load sinks to the bottom also making it not usable by the algae. Additionally, because there is such a high correlation between the river variables, this discrepancy could be due to multicollinearity issues, even though none of the model's VIF values were greater than ten. This is also evidenced by the fact that the independent variable Sandusky total dissolved reactive phosphorus has a different sign than the correlation coefficient between this variable and the dependent variable.

4.6.3 16 Year dataset – dependent variable, Chlorophyll-a

Prior research had indicated that Lake Erie has changed dramatically over the last 30 years. One way that the Lake has changed is due to the introduction of invasive Dreissenids mussel species. Prior research has shown that by the mid-1990s the effects of these mussels in Lake Erie had changed the Lake, and I was interested in seeing if using a dataset for the years 1996-2011 could improve my modeling, since the process of HAB formations likely changed as well after this year.[184] Further, by looking at a dataset starting in 1996, models using this dataset

[184] Michalak, 2013.

should more closely resemble the Stumpf and Obenour models if my choices for dependent variables are appropriate proxies for HABs. The downside of using this dataset is the obvious reduction in observations, from 25 in my full dataset, to 16 in this dataset.

First, I examined the plots between dependent and independent variables. Similar to my models using the full dataset, these plots did not show any strong linear relationships, but did not indicate that other, higher order models would be appropriate (see Figure 17).

I began again by running a linear regression on dependent variable chlorophyll-a against all of my other independent variables which resulted in $F = 7.23$ and $p = .006$ indicating at least one of the independent variables was significant. Then, using the Selection function of SAS to find the combination of independent variables that would produce the highest adjusted R^2 indicated that using Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Maumee total nitrogen, Sandusky total phosphorus, Sandusky total nitrogen, and temperature would result in the largest adjusted R^2 value. Using these variables to create my reduced model resulted in a model with an F-value of 8.34 and a p-value of .0043, after adjusting for the estimated missing values, and an R^2 value of .8622, and an adjusted R^2 value of .7704. Examining the parameter estimates indicated that all variables used in the reduced model, except for Maumee River total nitrogen and temperature, were statistically significant.

Following the creation of this reduced model, I next checked the model assumptions. Based on the normal probability plot, the residuals did not appear to be normally distributed (see Figure 20). Also, examining the residual plot showed that the equal variance assumption was not satisfied (see Figure 20). Looking at the VIF values, multicollinearity did not appear to be an issue.

Trying to solve the model assumption issues I tried transformations on the dependent variable chlorophyll-a. I first used the same transformation that I used on the full dataset with dependent variable chlorophyll-a, taking the variable to the power of .23. This transformation normalized the residuals and the residual plot looked much better (see Figure 21).

The next reduced model that I ran used the dependent variable chlorophyll-a transformed to normal with independent variables Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Maumee total nitrogen, Sandusky total phosphorus, Sandusky total nitrogen, and temperature. This model had an F-value of 6.83 and a p-value of .0081 after adjusting for the estimated missing values. The R^2 value was .8368 and the adjusted R^2 value was .7279. All of the independent variables were significant at the .05 level except for Maumee total nitrogen and temperature.

Since independent variables Maumee total nitrogen and temperature were insignificant in the above reduced model, the final reduced model dropped these variables and has transformed dependent variable chlorophyll-a with independent variables Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Sandusky total phosphorus, and Sandusky total nitrogen. This model had an F-value of 5.42, a p-value of .0138, after adjusting for the estimated missing values, an R^2 value of .6844, and an adjusted R^2 value of .5696. All of the independent variables were significant at the .05 level.

Model	F-value	p-value	R^2	adj. R^2
DV $Chol^{.23}$ final reduced	5.42	0.0138	0.6844	0.5696

Table 8: $Chol^{.23}$; 16-year dataset; final reduced model

The final reduced model with transformed dependent variable and dropped insignificant independent variable had the following form:

$$Chol^{.23} = 1.791 - .000448 * MauTP + .002506 * MauTSRP + .001356 * SanTP - .000189 * SanTN$$

The normal probability and residual plot showed no issues with the normal and equal variance assumptions (see Figure 22). The VIF values indicated no issues with multicollinearity.

Interpreting the chosen reduced model gives us some insight into the system we are modeling. Again, prior research and the ANOVA analysis above indicate that large values of chlorophyll-a reflect higher chances of an HAB. The independent variable coefficients for Maumee River total phosphorus and Sandusky River total nitrogen, are negative, while Maumee River total dissolved reactive phosphorus and Sandusky River total phosphorus have positive coefficients. The negative coefficients indicate that as the variables Maumee River total phosphorus and Sandusky River total nitrogen increase, the dependent variable chlorophyll-a decreases. The positive coefficients indicate that as the variables Maumee River total dissolved reactive phosphorus and Sandusky total phosphorus increase, chlorophyll-a will also increase. This interpretation provides some interesting and also confusing results. The results for Sandusky total nitrogen align with prior research since nitrogen is not a nutrient that promotes HAB growth and may in fact inhibit HAB growth. Like the full dataset with the dissolved oxygen dependent variable, we have conflicting results for the river variables. The reduced model indicates that increases Maumee River total phosphorus decrease chlorophyll-a values, while increases in Sandusky River total phosphorus loads increase chlorophyll-a values. Again, these results could be due to a number of factors. First, for the Maumee River, since both total phosphorus and total dissolved reactive phosphorus are included, the model may be indicating

that total dissolved reactive phosphorus is a greater contributor to algae growth than total phosphorus from the Maumee. This could be due to dissolved reactive phosphorus being available immediately to the algae, while total phosphorus includes non-soluble forms of phosphorus that may sink or otherwise become unavailable to fuel algae growth. The other explanation for this discrepancy is that there are multicollinearity issues even though no variables in the model had VIF values greater than 10. The river variables are highly correlated, and further evidence for this is that the variable Maumee total phosphorus has a negative sign in the model, but have a positive correlation coefficient against the dependent variable.

4.6.4 16 Year dataset – dependent variable, Dissolved Oxygen

I repeated the modeling process using the 16-year dataset with dependent variable dissolved oxygen. Examining the plots between independent and the dependent variables again showed no overtly strong linear relationships (see Figure 18). The correlation between independent variables was the same as before, so issues of multicollinearity were examined.

After starting with the full model, I used the Selection Feature of SAS to find the combination of independent variables that would produce the highest adjusted R^2 which indicated that using Maumee total phosphorus, Sandusky total dissolved reactive phosphorus, and Sandusky total phosphorus would result in the largest adjusted R^2 value. Using these variables to create the reduced model resulted in a model with an F-value of 11.02, a p-value of .007 adjusted for estimating the missing values, an R^2 value of .7504, and an adjusted R^2 value of .6880. Examining the parameter estimates indicated that all independent variables were statistically significant.

Model	F-value	p-value	R^2	adj. R^2
DV HypDO final reduced	11.02	0.007	0.7504	0.6880

Table 9: HypDO; 16-year dataset; final reduced model

The reduced model had the following form:

$$HypDO = 3.58 - .001801 * MauTP + .01766 * SanTP - .0632 * SanTSRP$$

Since all of the independent variables were significant, this became the final reduced model. Following the creation of this model, I next checked the model assumptions. Based on the normal probability plot, the residuals were very questionably normally distributed. Also, the residual plot showed some issues with scattering (see Figure 24). Six transformations were attempted on the dependent variable, but none improved upon the two assumptions. Therefore, the untransformed model was used. The VIF values indicated no issues with multicollinearity.

Next we interpret the reduced model for the dependent variable dissolved oxygen. Prior research and the previous ANOVA analysis indicate that lower levels of dissolved oxygen would be present during seasons with HAB occurrences, and therefore, we can look at the reduced model to determine which independent variables cause dissolved oxygen to decrease. Examining the independent variable coefficients we see that the coefficients for Maumee River total phosphorus and Sandusky River total dissolved reactive phosphorus are negative, while the coefficient for Sandusky total phosphorus is positive. These coefficient signs indicate that as Maumee River total phosphorus and Sandusky River total dissolved reactive phosphorus increase, the dependent variable dissolved oxygen would decrease. Similarly, as Sandusky River total phosphorus increases, dissolved oxygen would also increase. From prior research we would expect dissolved oxygen levels to decrease as nutrient loads are increased. However, we again

see a discrepancy, in this model between Sandusky total phosphorus and Sandusky total dissolved reactive phosphorus. This could be due to the dissolved reactive phosphorus being more readily available for algae, while total phosphorus includes phosphorus that is not soluble or usable by the cyanobacteria. There is also a discrepancy between the sign for the coefficient of Maumee total phosphorus in the model and the sign of the correlation coefficient between Maumee total phosphorus and the dependent variable, which as mentioned supra, may be an indication of a multicollinearity issue, even though no variables in the model had a VIF greater than 10.

4.7 Model Validation

After running all of the regression models using the proxy dependent variables chlorophyll-a and dissolved oxygen, I attempted to run logistic regression models using the categorical HAB variable. However, because there were only 16 observations, the resulting models did not add anything to the study and the results are omitted. Even though the logistic models weren't able to tell us much, we can still use the categorical dataset to check the predictive power of the four created models. To do this, I examined the one-way ANOVAs for each of my variables, using my variables as the dependent variable and the ordinal data as the factor. The ANOVA analysis tells us the means of each variable associated with each level of the categorical data set. Thus, we can have a mean value for each variable that is associated with the categorical no bloom and bloom presence. After finding these means, I used the means of the independent variables found from the ANOVA analysis, putting them in the models created from the multiple linear regression analysis, and if the created models are suitable, they should predict a value of the dependent variable close to the mean value predicted from the ANOVAs.

After we used the ANOVA analyses to find the predicted means for each variable at the different levels of the ordinal dataset, we can use these means in the four multiple linear regression models to check their accuracy. We insert the predicted means of the independent variables for the bloom presence level into the four equations, and if the equations are valid, then we should see a result that is close to the predicted mean for the dependent variable. Since none of the variables showed a statistically significant difference between the two levels, the models may not result in exactly the predicted mean, but the value should still be close.

For the first equation, using the full dataset with normally transformed dependent variable chlorophyll-a, independent variables Sandusky total dissolved reactive phosphorus and Sandusky total nitrogen were included. Sandusky total dissolved reactive phosphorus had a predicted mean of 56.6 and Sandusky total nitrogen had a predicted mean of 3674 for the bloom presence level. The transformed dependent variable chlorophyll-a had a predicted mean of 1.5678 for the bloom presence level. Inserting the independent variable means into the final reduced model resulted in a prediction of 1.58071 for the transformed dependent variable. This is very close to the predicted mean for chlorophyll-a, which indicates that this model is appropriate.

The next model used the full dataset with untransformed dependent variable dissolved oxygen. Independent variables Sandusky total phosphorus and Sandusky total dissolved reactive phosphorus were included as independent variables in the model. Sandusky total phosphorus had a predicted mean of 316.2 and Sandusky total dissolved reactive phosphorus had a predicted mean of 56.6 at the bloom presence level. The dependent variable dissolved oxygen had a predicted mean of 3.3 at the bloom presence level. Inserting the independent variable means into the final reduced model resulted in a prediction of 3.272592 for the untransformed dependent

variable. Again, this value is very close to the predicted mean for dissolved oxygen, indicating that this model is appropriate.

The third equation used the 16-year dataset with normally transformed dependent variable chlorophyll-a. Independent variables Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Sandusky total phosphorus, and Sandusky total nitrogen were included in the model. Maumee total phosphorus had a predicted mean of 1398, Maumee total dissolved reactive phosphorus had a predicted mean of 270.6, Sandusky total phosphorus had a predicted mean of 316.2, and Sandusky total nitrogen had a predicted mean of 3674 at the bloom presence level. The transformed dependent variable chlorophyll-a had a predicted mean value of 1.5678 at the bloom presence level. Inserting the independent variable means into the final reduced model resulted in a prediction of 1.5772008 for the transformed dependent variable. This value is very close to the predicted mean for the transformed dependent variable, which suggests that this model is appropriate.

The final equation used the 16-year dataset with untransformed dependent variable dissolved oxygen. Independent variables Maumee total phosphorus, Sandusky total phosphorus, and Sandusky total dissolved phosphorus were included in the model. Maumee total phosphorus had a predicted mean of 1398, Sandusky total phosphorus had a predicted mean of 316.2, and Sandusky total dissolved reactive phosphorus had a predicted mean of 56.6 at the bloom presence level. Dependent variable dissolved oxygen had a predicted mean of 3.3 at the bloom presence level. Inserting the independent variable means into the final reduced model resulted in a prediction of 3.069174. This value is close to the predicted mean for dissolved oxygen, indicating that the model is appropriate.

4.8 Results

To begin this section, we will compare the four multiple linear regression final reduced models that were created. The F-values and associated p-values for each model are used in testing the null hypothesis that all of the model coefficients are zero. In other words, at alpha level .05, if the model p-value is less than .05, then the null hypothesis is rejected, and we can say that we have sufficient evidence to claim that at least one of the coefficients in the model differs from zero (i.e. some of the independent variables are statistically significant in predicting the dependent variable). On the other hand, if the model p-value is greater than .05, then we can say that we have insufficient evidence that at least one of the coefficients in the model is different than zero (i.e. none of the independent variables are statistically significant in predicting the dependent variable). The R^2 measure is the proportion of variance in the dependent variable that is explained by the relationship between the dependent variable and the independent variables. The R^2 value is an overall measurement of the strength of association and does not reflect the extent to which any particular independent variable is associated with the dependent variable. The adjusted R^2 value is an adjustment to the R^2 that penalizes the addition of extraneous predictors to the model calculated using the formula $(1 - R^2)(N - 1)/(N - k - 1)$, where k is the number of predictors. Finally, independent variables were determined to be significant or not to the model based on their p-values. The p-values for the independent variables are the 2-tailed p-values used in testing the null hypothesis that the coefficient parameter is zero. Similar to the model p-value this means that at an alpha level of .05, if the variable p-value is less than .05, then we reject the null hypothesis, and say that we have sufficient evidence that the variable's coefficient is not zero (i.e. the variable is statistically significant to the model). On the other hand, if the variable p-value is greater than .05, then we

do not reject the null hypothesis, and say that we have insufficient evidence that the variable's coefficient is significantly different than zero (i.e. the variable is not statistically significant).

Dataset	**Full dataset**	**16-year dataset**	**Full dataset**	**16-year dataset**
Dependent Var	Chlorophyll-a	Chlorophyll-a	Dissolved Oxygen	Dissolved Oxygen
F-value	5.42	5.42	9.50	11.02
Model p-value	0.016	0.0138	0.001	0.007
R^2	0.3678	0.6844	0.4749	0.7504
Adj. R^2	0.3103	0.5696	0.4272	0.6880

Table 10: Final reduced models comparison

In order to determine the "best" model, defined as the model that has the highest predictive performance, we consider the R^2 and adjusted R^2 values. The reduced models for the 16-year dataset for both dependent variables produced the strongest predictive models. The 16-year dataset reduced model for dependent variable chlorophyll-a had an R^2 value of .6844, while the 16-year dataset reduced model for dependent variable dissolved oxygen had an R^2 value of .7504. Since the R^2 is between 0 and 1, with values closer to 1 having higher predictive power, we can see that these are relatively strong prediction models. In terms of the adjusted R^2 value, the same models have the highest adjusted R^2 values among the created models. The 16-year dataset reduced models for chlorophyll-a and dissolved oxygen had adjusted R^2 values of .5696 and .6880 respectively.

Based on these categorizations, the best created models are the reduced models of the 16-year dataset for both variables. The reduced model of the 16-year dataset for variable dissolved oxygen used independent variables Maumee total phosphorus, Sandusky total phosphorus, and Sandusky total dissolved reactive phosphorus. The reduced model of the 16-year dataset for

variable chlorophyll-a used independent variables Maumee total phosphorus, Maumee total dissolved reactive phosphorus, Sandusky total phosphorus, and Sandusky total nitrogen.

Comparing my best models for each dependent variable with the Stumpf and Obenour models, we see a number of interesting results. First, my created models had R^2 lower than the Stumpf and Obenour models, meaning that my models had less predictive power. Stumpf's exponential model had an R^2 of .97, while Obenour's gamma model had an R^2 value of .91, each better than my two best models (.7504 for dissolved oxygen; .6844 for chlorophyll-a). Other differences exist between our models. First, while the Stumpf and Obenour models exclusively used Maumee total phosphorus loads as the only independent variable, my models included two Maumee River variables (phosphorus and dissolved reactive phosphorus) as well as all three Sandusky River variables (phosphorus, dissolved reactive phosphorus, and nitrogen).

The results of my models suggest that both chlorophyll-a and dissolved oxygen could be used as proxies for HAB occurrence. The dissolved oxygen proxy created stronger predictive models than the chlorophyll-a variable, likely due to the overestimation of cyanobacteria by the chlorophyll-a count. The ANOVA analyses indicated that larger values of chlorophyll-a correspond to an increased likelihood of HAB occurrences, while small values of dissolved oxygen correspond to increased HAB occurrences. Using the 16-year dataset created stronger predictive models with the relevant predictor variables. While my models indicated that the main independent variable in the Stumpf and Obenour models, Maumee River phosphorus load, was important, my models also indicated the importance of Maumee River dissolved reactive phosphorus. Further, my models suggest that phosphorus, dissolved reactive phosphorus, and nitrogen loads from the Sandusky River also play an important role in the formation of HABs in Lake Erie. Therefore, in future studies and during the creation of new predictive models, the

Sandusky River variables should be included, instead of being discarded simply because the Maumee River has higher flow and loading rates. Further, my models confirm the hypothesis that Lake Erie has changed over the last 30 years, specifically in terms of the formation of HABs. Something changed in the Lake in the mid-1990s, as shown by my improved models when only looking at the years 1995-2011. This aligns with prior research that indicated that invasive Dreissenid mussel species had changed the hydrology of the Lake enough to change the way HABs are formed.

CHAPTER V: DISCUSSION

This section takes an overarching view of the prior research concerning HABs and their modeling, my methods and results, and the public policy response to the bloom threat. To begin, why are predictive models for HABs important? First, modeling HABs gives researchers and public policy makers a better understanding of how Lake Erie is functioning, and which types of policies lead to a more, or sometimes less healthy Lake. Second, by creating predictive HAB models, researchers and NOAA have been able to create HAB forecasting. Using the Stumpf model as guidance (and the CI proxy for HABs), the NOAA disseminates weekly bulletins during the HAB season that predicts, forecasts, and tracks HABs. By making these forecasts and making them available to the public, NOAA allows for natural resource and public health managers to be better prepared to make the necessary arrangements in order to mitigate the detrimental impacts that HABs may cause.[185] This includes municipal water managers who can use the forecasts and seasonal predictions to prepare for taste and odor issues that are associated with cyanobacteria blooms, as well as be prepared for excessively large and toxic blooms that can disrupt access to tap water. Additionally, municipalities and Ohio agencies that are responsible for managing Lake Erie beaches can use the NOAA forecasts to post warning signs near public beaches that may become affected by a bloom. Further, the forecasting system is important for the scientific research community who can use the forecasts (and associated research), as I did, to help target their own research.[186] Finally, HAB modeling, and specifically the Stumpf model, has been a useful tool for public policy makers to use in determining the

[185] Wynee, 2013.
[186] *Id*.

necessary future nutrient load restrictions that should be implemented to support the goal of reducing or eliminating HABs.

Current public policy is based on HAB modeling, and following the Toledo drinking water crisis caused by a large HAB, both the Ohio government and the federal government have taken action to address the issue. This section gives a brief overview of these actions, and will be followed by a discussion on their likely usefulness based on my research on modeling.

By 2015, both the Ohio House of Representatives and the Ohio Senate had passed bills aimed at combating HABs in Lake Erie. The Ohio Senate passed Senate Bill 1 in February, and the Ohio House of Representatives passed House Bill 61 in March. Amazingly, especially for an environmental law in a state legislature dominated by Republicans, both of these bills passed their respective chambers unanimously. House Bill 61 prohibits farmers in northwestern Ohio from spreading manure and fertilizer on their fields if the ground is frozen, saturated with water, or if the weather forecast calls for a greater than 50% chance of precipitation exceeding 1 inch in a 12-hour period.[187] Like the House Bill, Senate Bill 1 prohibits the spread of manure or fertilizer on frozen and saturated soil in the western Lake Erie basin, but would also require water treatment plants to begin monthly monitoring of dissolved phosphorus and bans the dumping of dredged material in Lake Erie beginning in the summer of 2020. The Senate Bill also designates a harmful algae management and response coordinator in the Ohio EPA and prohibits the use of pipes and plumbing materials that are not lead free in water systems used to provide drinking water.[188] Both the House and Senate Bills passed their respective chambers, and the combined

[187] The Ohio Legislature. "House Bill 61." 2015. Available at https://www.legislature.ohio.gov/legislation/legislation-status?id=GA131-HB-61.
[188] The Ohio Legislature. "Senate Bill 1." 2015. Available at https://www.legislature.ohio.gov/legislation/legislation-summary?id=GA131-SB-1.

bills were referred to a conference committee.[189] On March 25, 2015 the combined Ohio House and Senate legislation, now referred to as Senate Bill 1 was reintroduced and was passed by Ohio House 96 to 0, and was passed by the Ohio Senate 33 to 0.[190] Ohio Governor John Kasich signed the bill at a state park near Lake Erie on April 2, 2015, with the legislation becoming effective 90 days after the signature.[191]

The legislation from the Ohio House and Senate have had broad support from interested groups, mainly environmentalists and farmers. Farmers and the groups that represent their interests have concerns about the cost for farmers to conform to the bills, specifically, the cost of storing manure in winter, when in the past, the manure would have been applied directly to the fields. Speaking on Senate Bill 1, the Ohio Farm Bureau supported the manure and fertilizer provisions, but was concerned about the rules taking immediate effect, and would prefer to see them phased in, as complying with the law may require engineering plans, geological surveys, or buying new equipment among other steps.[192] On the other hand, environmental groups call the bills only a first step, arguing that they do not go far enough. The Ohio Environmental Council has argued that the fertilizer-spreading prohibition contains loopholes that would largely allow farmers to continue spreading nutrients and avoid financial penalties.[193] Environmental advocates were also concerned with provisions in Senate Bill 1, particularly the fact that the fertilizer and manure protections end after five years, and that the bill was amended to prohibit

---

[189] *Id*.
[190] *Id*.
[191] Blade Staff. "Gov. Kasich signs bill aimed at protecting Lake Erie, Ohio water quality." Toledo Blade. April 2, 2015. Available at http://www.toledoblade.com/State/2015/04/02/Gov-Kasich-signs-bill-aimed-at-protecting-Lake-Erie-Ohio-water-quality.html.
[192] Borchardt, Jackie. "Bill targeting Lake Erie algal blooms passes Ohio Senate." Cleveland.com. February 18, 2015. Available at http://www.cleveland.com/open/index.ssf/2015/02/bill_targeting_lake_erie_algal_blooms_passes_ohio_senate.html.
[193] Siegel, Jim. "Bill aiming to curb toxic algae in Lake Erie passes Ohio House." The Columbus Dispatch. March 11, 2015. Available at http://www.dispatch.com/content/stories/local/2015/03/10/algae-bill.html.

spreading only manure, and not fertilizer, when the weather forecast calls for heavy precipitation.[194] Several lawmakers have acknowledged these shortcomings, reiterating that these bills, though moving in the right direction, are only a first step, and additional bills to take the next steps at reducing HABs will be coming in the future.[195]

At the federal level, legislation has also been introduced to combat HABs in Lake Erie. The Harmful Algal Bloom and Hypoxia Research and Control Act was first passed in 1998, but was amended in the summer of 2014. The new amendments add two sections to the original act. The first requires NOAA to maintain and enhance a national harmful algal bloom and hypoxia program and task force that shall coordinate interagency review of the objectives of the program; expedite the interagency review process; support implementation of the Action Strategy; and promote the development of new technologies for predicting, monitoring, and mitigating HAB and hypoxia conditions.[196] The second new section requires the task force to develop a comprehensive research plan and action strategy to address marine and freshwater HAB and hypoxia.[197] These amendments were passed by Congress and signed by President Obama into law in June of 2014.

Representative Robert Latta, of the 5th Ohio district has twice submitted the Great Lakes Algal and Fresh Water Algal Bloom Information Act. First introduced in the House as H.R. 5456 (with companion Senate Bill S. 2790) in 2014, the bill did not pass out of the House or Senate. Trying again in 2015, Representative Latta reintroduced the bill, now H.R. 349, in January of 2015. The bill is currently assigned to a congressional committee, which considers it before

---

[194] Borchardt, 2015.
[195] Siegel, 2015.
[196] Government Printing Office. "The Harmful Algal Bloom and Hypoxia Research and Control Act" June 9, 2014. Available at http://www.gpo.gov/fdsys/pkg/BILLS-113s1254eah/pdf/BILLS-113s1254eah.pdf.
[197] *Id*.

sending it on to the House to vote on as a whole. This act, if passed, would require the NOAA to create an electronic database of research and information on the causes of, and corrective actions being taken with regard to algal blooms in the Great Lakes. The database would include relevant chemical, physical, and biological data that have been collected by relevant universities, organizations, or state/federal agencies. The NOAA administrator would be required to update Congress annually on the database, and the database would be made available to the public. However, the act would not grant any authority to the NOAA Administrator, specifically no authority to require the submission of data.[198]

Representative Latta also introduced a third federal public policy response to Lake Erie HABs. The Safe Water Drinking Act, introduced in January, 2015 and passed through the House of Representatives in February, 2015, would direct the EPA to assess the risks posed by toxic algal blooms and come up with way to fight the blooms. However, as the bill's name implies, this act focuses the EPA's resources on protecting drinking water by studying the toxicity of HABs and monitoring and treating water intake used for drinking water.[199] Currently, this legislation is being considered in the Senate under a companion bill introduced to the Senate by Senator Rob Portman of Ohio.

In a separate action, the Lucas County Board of Commissioners initiated a study by University of Toledo environmental and water law professors to examine other avenues that Ohio and affected counties and cities could take in response to HABs. The report, authored by Jack Tuholske and Ken Kilbert found first and foremost that phosphorus loading into Lake Erie

---

[198] Latta, Robert. "H.R. 349." August 6, 2014. Available at http://latta.house.gov/uploadedfiles/great_lakes_and_freshwater_algal_blooms_information_act.pdf.
[199] Latta, Robert. "H.R. 212 – To amend the Safe Drinking Water Act to provide for the assessment and management of the risk of algal toxins in drinking water and for other purposes." Congress.gov. February 25, 2015. Available at https://www.congress.gov/bill/114th-congress/house-bill/212/text.

has increased steadily since the 1990s due to increased nutrient pollution from a variety of nonpoint sources, including but not limited to agricultural activities.[200] The authors argue that the lack of regulation over nonpoint sources of nutrient loads is the root of the HAB problem in Lake Erie.[201] Focusing on the Clean Water Act, the paper finds that while the Act has been successful at reducing pollution from point sources, it has been far less successful at regulating nonpoint sources, finding that federal law provides no direct regulatory authority over nonpoint source nutrient loading.[202] With no federal authority to regulate nonpoint sources, state governments need to use their regulatory power to do so, but the authors found that states, including Ohio, have traditionally shied away from nonpoint source pollution control, not wanting to harm agricultural economic interests.[203] The report noted that other stakeholders on the Great Lakes, including bordering states and provinces must work together to address the sources of nutrient pollution. The report makes a number of recommendations but overall the report concludes that "there is no simple, single legal solution to the nutrient pollution problem in Lake Erie".[204] According to the authors, nutrient pollution in Lake Erie can only be abated through legally binding requirements that address all sources of nutrient pollution, both point and nonpoint sources, throughout the entire western Lake Erie basin, including in parts of Michigan, Ohio, and Indiana.[205] Finally, the report notes that broad stakeholder participation from all sectors –

[200] Tuholske, Jack, & Kilbert, Ken. "Moving Forward: Solutions to Lake Erie's Harmful Algal Blooms." Lucas County, Ohio Board of County Commissioners. April 15, 2015 available at http://ftpcontent4.worldnow.com/wtol/pdf/KilbertAlgaeResearch.pdf.

[201] *Id*.
[202] *Id*.
[203] *Id*.
[204] *Id*.
[205] *Id*.

government, agriculture, municipal treatment plants, the affected public, and the scientific community – is critical to the success of any nutrient abatement strategy.[206]

There are a number of interesting takeaways from the public policy response. First, it is encouraging that the legislative bills, both federally and at the state level, have received bipartisan support, and even more encouraging that a Representative like Robert Latta, who has disparaged the EPA in the past, is sponsoring bills aimed at environmental protection. However, there are stark differences between the federal and state approaches. The federal action focuses mainly on creating taskforces and programs, centered on interagency cooperation, rather than focusing on actual regulation of nutrients that are causing HABs. On the other hand, the bills passed in the Ohio House and Senate do address nutrient loading, but have created loopholes and exemptions.

Another interesting aspect of the public policy response, specifically the Ohio bills, is the move away from whole-lake targets. While the GLWQA nutrient loading targets remain in place, and are targets for the entire lake, the newer Ohio bills focus exclusively on the Western basin. Moving away from whole-lake targets may be the appropriate action due to the differences and spatial scales of loading on HABs.[207] Because the major contributors to the nutrient loading issue are in the western basin, management efforts are most cost effective if they are focused on those watersheds that deliver the most nutrients. The state is able to identify the most important contributing watersheds, and this information should allow for more effective targeting of the

[206] *Id*.
[207] Scavia, 2014.

issue. Further, more effective best management practices can be encouraged in the area where the nutrient issue is the largest.[208]

One aspect missing from the public policy response is the issue of climate change. Because the results of management actions aimed at addressing the nutrient load issue from nonpoint sources can take years or decades to take effect, potential impacts of a changing climate should be taken into consideration for the action to be effective.[209] Most prior research suggest that climate change will not only exacerbate existing problems, but also make reducing loads more difficult.[210] Prior research, especially from Michalak et al., has shown that climate change is predicted to lead to temperatures and stratification patterns favorable for cyanobacteria growth.[211] Because the interaction between recent increases in large spring precipitation that flush more nutrients into Lake Erie and increased temperatures that favor cyanobacteria growth suggests that unless management actions are taken the HAB events that have re-emerged in Lake Erie will continue into the future.[212] The Lucas County Board of Commissioners study also noted the importance of climate change, finding that as Lake Erie continues to experience the effects of phosphorus loading from agricultural and urban runoff, climate change will dramatically compound these impacts through higher temperatures, increases in high precipitation events, and shifting winds that favor the production of HABs.[213]

Another criticism of the public policy response is that the current farm policy is based on volunteer, incentive-based adoptions of best management practices. Farmer adoption of best

---

[208] *Id*.
[209] *Id*.
[210] *Id*.
[211] Kane, 2014.
[212] *Id*.
[213] Tuholske, 2015.

practices will be paramount in reducing nutrient loading, and environmental groups have suggested that farmers be required to take measures against accidental nutrient runoff.[214] Farmers argue that they are interested in minimizing nutrient loads, but are fearful of mandatory regulations on an industry that has been dominated by voluntary self-regulation. Experiences in other large regions with nutrient load problems (i.e. Gulf of Mexico/Mississippi River) have shown that significantly reducing nonpoint source loads is difficult, as the sources are spatially distributed and the methods used to reduce nutrient loading are primarily voluntary and incentive based, making them difficult to target and track.[215] This is compounded by the fact that the response time between action and result can be many years or longer, and the results can only be measured cumulatively in space and through time.[216] Therefore, it has been suggested that an adaptive management approach, that sets interim targets, evaluating the results in loads and lake response on time scales (i.e. 5 year averages), and then adjusts management actions or loading targets if necessary.[217] An approach like this would be appropriate for Lake Erie, and would allow for more effective testing and post-audits of the ability of models to project the ecosystem's response and thus improve subsequent assessments and projections.[218]

Overall, the public policy response to the recent rash of HABs in Lake Erie has been encouraging. There are a number of important issues, however, that need to be considered as Ohio, other Great Lakes states, and the Federal government move forward in battling HABs in Lake Erie that have been highlighted by my research.

---

[214] Scavia, 2014.
[215] *Id*.
[216] *Id*.
[217] *Id*.
[218] *Id*.

First, consistent and quality data collection should continue, and research should also continue to examine the roles of dissolved reactive phosphorus, and the possible future changes to the Lake system that may occur due to projected climate change. Representative Latta's efforts to create a NOAA database with all relevant data related to HABs is not as flashy as the other public policy responses, but I believe that it is as important as limiting manure application. Currently, data that relates to HABs is spread out, located in different academic papers, or held onto personally by researchers. By putting this data into a database that is accessible by the public and by other academics, research into HABs in Lake Erie will become easier and more democratized. A second important issue that must be addressed is a definitive definition of what an HAB is. Researchers should define exactly what constitutes a HAB, and consistent methods for measuring occurrence, size, and severity of HABs should be developed. Finally, researchers should determine a set of proxy measurements to use when modeling, researching, or predicting HABs, which was the focus of this paper. Each proxy has advantages and disadvantages that should be considered. For example, the remote sensing CI estimate used in prior models seems to be a strong proxy for HABs, but it is an expensive method that requires near constant satellite imagery of Lake Erie over the summer. Similarly, the boat tow method used by Bridgewater is able to measure the more of the water column than the CI proxy, but it too requires man hours and money, and the method is not able to examine the entire spatial area of a HAB. While the proxies used for this paper did not create as strong predictive models as those using the CI proxy, it may be more economical to use stationary buoys to collect information regarding chlorophyll-a counts and dissolved oxygen data.

CHAPTER VI: CONCLUSION

Hazardous algal blooms, or HABs, have been an increasing issue in the western basin of Lake Erie. HABs are formed by the blue-green algae *Microcystis*, which is a type of cyanobacteria that has been present in waters around the world for centuries. However, human activities, specifically nutrient loading of freshwater lakes, cause the cyanobacteria to grow at a rapid rate, and outperform other algae that may be present. When large enough, the algae can form into expansive blooms that can release toxins, rob other species of algae of food, and cause oxygen levels to drop. HABs in Lake Erie began to cause problems in the late 1970s, however through nutrient load reduction programs, by the 1980s HABs in the Lake appeared to be under control. Recently, HABs have returned, becoming a persistent nuisance and a costly drain to the economy of Northwest Ohio and Southeast Michigan. HABs can be a nuisance to human enjoyment of lakes and rivers, can cause health problems in humans and wildlife, and greatly affect the ecosystem where they are present.

Much research has been put into understanding the drivers of HABs, and most of this work has focused on the nutrients phosphorus and nitrogen. Other variables which affect the occurrence and size of HABs include water temperature, water clarity, wind speed, and precipitation events. A number of models have been created in order to understand and predict HAB occurrences. Additionally, these models have been used by policy makers to create nutrient reduction goals. Two primary models have been created to predict HABs in Lake Erie, the Stumpf model and the Obenour model. The Stumpf model uses spring total phosphorus loads from the Maumee River to predict bloom magnitude, while the Obenour model uses a Bayesian framework using a gamma error distribution and also uses the predictor variable spring total phosphorus load from the Maumee River. A major issue with the two models is a lack of

consistent, long term data. A single definition of HAB has not been determined, and the Lake indicators that have formed these models has been plagued by inconsistent collection and different collection techniques. Further, different proxies for the dependent HAB variable have been used to create these models. The Stumpf model uses remote sensing from satellite imagery to quantify blooms from 2002 to 2011, while the Obenour model uses other estimates, including from boat tows, to quantify blooms from 2002-2013.

The purpose of this study was to generally examine if other proxies for HABs, a larger dataset, and different independent variables, could create valid HAB predictive models and/or improve upon the prior forecasting models. More specifically, since there is no single definition of an HAB, and no one agreed upon metric to measure them, my study examines whether my proxies, dissolved oxygen and chlorophyll-a, are suitable proxies for HABs in Lake Erie. To assess if these variables could be used as proxies a number of different models were created and examined. The creation and examination of these models resulted in a number of interesting conclusions. Comparing the strongest models from this study to the models of Stumpf and Obenour indicated that models using the proxy variables chlorophyll-a and dissolved oxygen were not as strong as the prior models. The models in this study and those presented by Stumpf and Obenour do differ, especially in terms of independent variables. First, the prior models focused exclusively on Maumee River total phosphorus loads, while my models examined three Maumee River nutrient variables (phosphorus, dissolved reactive phosphorus, and nitrogen) as well as these nutrient loads from the Sandusky River. Additionally, the models for this paper examined the use of temperature and water clarity as possible predictor variables. The results of this study indicate that both chlorophyll-a and dissolved oxygen could be used as proxies for HAB occurrence. Additionally, the results of this paper suggest that nutrient loads other than

exclusively Maumee River total phosphorus are important to predicting HAB occurrence, and in future studies, researchers should not discount the importance of nutrient loading from the Sandusky River. A final result from this study is a confirmation that the chemistry of Lake Erie, at least in terms of the occurrence of HABs, has changed since the mid-1990s possibly related to invasive Dreissenid mussel species.

After a number of severe HAB events in the western basin of Lake Erie between 2011 and 2014, the public policy response both at the state and national level has been quick and bipartisan. The Ohio House and Senate both unanimously passed legislation in 2015 aimed at reducing nutrient loads in the western basin of Lake Erie, primarily by restricting the time, manner, and storage of manure, as well as restricting the application of other fertilizers. At the federal level, legislation has been passed focusing on interagency cooperation in the fight against HABs. Unfortunately, legislation regarding the collection of data and requiring NOAA to have a public database of relevant data, has not moved out of the House of Representatives. Though not as directly important as legislation restricting nutrient loads, the creation of a single database would be extremely helpful, allowing researchers broader access to information to study the creation and occurrence of HABs in Lake Erie. In regards to the Ohio legislation that attempts to limit nutrient loads, both industry and environmental stakeholders are relatively satisfied. The agricultural industry is pleased that mandatory fertilizer limits have not been placed on farmers, and prefers the adoption of best management practices. On the other hand, environmental groups are happy that action is being taken, though they have voiced concerns that the legislation does not go far enough and does not have the regulatory bite necessary to solve the problem. Though agricultural interests oppose the action, it may become necessary to resort to legally regulated farming practices instead of voluntary best practices if the issue of nonpoint nutrient pollution

continues to cause harm to Lake Erie through HAB occurrences. Other issues, especially possible effects from climate change, are exacerbating the problem, and the fact that different states, provinces, and countries share the Great Lakes makes any encompassing solution even more difficult. Further action by policy makers must focus on the role of nonpoint source nutrient pollution, but also must balance the interests of stakeholders, while at the same time protecting the economic and aesthetic benefits that Lake Erie provides.

WORKS CITED

Atkin, Emily. "7 Things You Need to Know About the Toxin That's Poisoned Ohio's Drinking Water." Climate Progress, August 3, 2014. Available at http://thinkprogress.org/climate/2014/08/03/3467068/toledo-ohio-water-crisis/.

Blade Staff. "Gov. Kasich signs bill aimed at protecting Lake Erie, Ohio water quality." Toledo Blade. April 2, 2015. Available at http://www.toledoblade.com/State/2015/04/02/Gov-Kasich-signs-bill-aimed-at-protecting-Lake-Erie-Ohio-water-quality.html.

Borchardt, Jackie. "Bill targeting Lake Erie algal blooms passes Ohio Senate." Cleveland.com. February 18, 2015. Available at http://www.cleveland.com/open/index.ssf/2015/02/bill_targeting_lake_erie_algal_blooms_passes_ohio_senate.html.

Bridgeman, Thomas, et. al. "From River to Lake: Phosphorus partitioning and algal community compositional changes in Western Lake Erie." Journal of Great Lakes Research, March, 2012. Available at http://www.sciencedirect.com/science/article/pii/S0380133011002164.

Burns, Noel M., Rockwell, David C., Bertram, Paul E., Dolan, David M., Ciborowski, Jan J.H. "Trends in Temperature, Secchi Depth, and Dissolved Oxygen Depletion Rates in the Central Basin of Lake Erie, 1983-2002." Journal of Great Lakes Research, 2005. Available at http://www.slideshare.net/EarthSoft/lake-erie-trends-in-temperature-secchi-depth-and-dissolved-oxygen-rates-in-the-central-basin-of-lake-erie-19832002?type=document.

Carrick, Hunter J., Moon Jessica B., Gaylord, Barrett F. "Phytoplankton Dynamics and Hypoxia in Lake Erie: A Hypothesis Concerning Benthic-pelagic Coupling in the Central Basin." Journal of Great Lakes Research, 2005. Available at http://www.sciencedirect.com/science/article/pii/S0380133005703087#.

Conroy, Joseph D., et. al. "Temporal Trends in Lake Erie Plankton Biomass: Roles of External Phosphorus Loading and Dreissenid Mussels." Journal of Great Lakes Research, 2005. Available at http://www.sciencedirect.com/science/article/pii/S0380133005703075.

Daloglu, Irem, Cho, Hyung Hwa, & Scavia, Donald. "Evaluating Causes of Trends in Long-Term Dissolved Reactive Phosphorus Loads to Lake Erie." Environmental Science & Technology, American Chemical Society. September 10, 2012. Available at http://graham.umich.edu/scavia/wp-content/uploads/2012/10/Daloglu-et-al-2012.pdf.

Dolan, David M., & McGunagle, Kevin P. "Lake Erie Total Phosphorus Loading Analysis and Update: 1996-2002." J. Great Lakes Res. 31 (Suppl. 2): 11-22. Internat. Assoc. Great Lakes Res., 2005. Available at http://www.cee.mtu.edu/~nurban/classes/ce5508/2007/Readings/dolan05.pdf.

Dolan, D.M., & Richards, R.P. "Analysis of Late 90s Phosphorus Loading Pulse to Lake Erie." Ecovision World Monograph Series, 2008. Available at http://www.uwgb.edu/doland/GLGrants/analysis_of_late_90s_phosphorus_loading_pulse_to_lake_erie.pdf.

Government Printing Office. "The Harmful Algal Bloom and Hypoxia Research and Control Act" June 9, 2014. Available at http://www.gpo.gov/fdsys/pkg/BILLS-113s1254eah/pdf/BILLS-113s1254eah.pdf.

Heidelberg University. National Center for Water Quality Research. 2015. Available at http://www.heidelberg.edu/academiclife/distinctive/ncwqr.

Heidelberg University. National Center for Water Quality Research. 2015. Available at http://www.heidelberg.edu/academiclife/distinctive/ncwqr/data.

Ho, Jeff C., & Michalak, Anna M. "Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie." Journal of Great Lakes Research, 2015. Available at, http://www.sciencedirect.com/science/article/pii/S0380133015000027.

Institute for Digital Research and Education. "SAS Annotated Output: Ordered Logistic Regression." UCLA. 2015. Available at http://www.ats.ucla.edu/stat/sas/output/sas_ologit_output.htm.

Kane, Douglas D., et. al. "Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass." Journal of Great Lakes Research, September, 2014. Available at http://www.sciencedirect.com/science/article/pii/S0380133014000768.

LaBarge, Greg, & Hoorman, Jim. "Lake Erie and Phosphorus: What has happened since 1995?" Ohio State University, 2012. Available at http://www.oardc.ohio-state.edu/ocamm/images/MTW2012_Hoorman.pdf.

Latta, Robert. "H.R. 349." August 6, 2014. Available at http://latta.house.gov/uploadedfiles/great_lakes_and_freshwater_algal_blooms_information_act.pdf.

Latta, Robert. "H.R. 212 – To amend the Safe Drinking Water Act to provide for the assessment and management of the risk of algal toxins in drinking water and for other purposes."

Congress.gov. February 25, 2015. Available at https://www.congress.gov/bill/114th-congress/house-bill/212/text.

Ludsin, Stuart A., Kershner, Mark W., Blocksom, Karen, A., Knight, Roger L., & Stein, Roy A. "Life after death in Lake Erie: Nutrient controls drive fish species richness, rehabilitation." Ecological Applications, v. 11, iss. 3, 2011 p. 731-746. Available at https://kb.osu.edu/dspace/bitstream/handle/1811/36708/...?sequence=1.

Michalak, Anna M., et. al. "Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions." PNAS, March 4, 2013. Available at http://graham.umich.edu/scavia/wp-content/uploads/2013/04/PNAS.pdf.

National Center for Water Quality Research. "Dissolved Phosphorus From Crop Runoff: Why it is a Big Problem!" Water Quality News and Notes. Heidelberg University. August 1, 2011, available at http://www.heidelberg.edu/sites/default/files/jfuller/images/1%20Dissolved%20P-a%20Big%20Problem%2C%2008-01-2011.pdf.

NOAA. "National Weather Service Forecast Office." 2015. Available at http://www.erh.noaa.gov/buf/laketemps/laketemps.php.

NOAA. "NOAA, partners predict significant harmful algal blooms in western Lake Erie this summer." 2014. Available at http://www.noaanews.noaa.gov/stories2014/20140710_erie_hab.html.

Obenour, Daniel R., Gronewold, Andrew D., Stow, Craig A., Scavia, Donald. "Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts." Accepted

Article, doi: 10.1002/2014WR015616, 2014. Available at http://graham.umich.edu/scavia/wp-content/uploads/2014/09/Obenour-et-al-2014.pdf.

Rencher, Alvin C., & Christensen, William F. "Methods of Multivariate Analysis." John Wiley & Sons, Inc, Publication, Hoboken, New Jersey. 2012.

Reutter, Jeffery, et. al. "Lake Erie Nutrient Loading and Harmful Algal Blooms: Research Findings and Management Implications." Ohio Sea Gant College Program, The Ohio State University, June 14, 2011. Available at http://worldcat.org/arcviewer/1/OHI/2011/11/08/H1320770105057/viewer/file2.pdf.

Scavia, Donald, et. al. "Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia." Journal of Great Lakes Research, June, 2014. Available at www.sciencedirect.com/science/article/pii/S038013304000252.

Rucinski, Daniel, DePinto, Joseph V., Scavia, Donald, Beletsky, Dmitry. "Modeling Lake Erie's hypoxia response to nutrient loads and physical variability." Journal of Great Lakes Research, December 16, 2013. Available at http://graham.umich.edu/scavia/wp-content/uploads/2014/03/Rucinski-et-al-20141.pdf.

Siegel, Jim. "Bill aiming to curb toxic algae in Lake Erie passes Ohio House." The Columbus Dispatch. March 11, 2015. Available at http://www.dispatch.com/content/stories/local/2015/03/10/algae-bill.html.

Stumpf, Richard P., "Satellite Monitoring of Toxic Cyanobacteria for Public Health." Earthzine, March 26, 2014. Available at, http://earthzine.org/2014/03/26/satellite-monitoring-of-toxic-cyanobacteria-for-public-health/.

Stumpf, Richard P., Wynne, Timothy T., Baker, David B., Fahnenstiel, Gary L. "Interannual Variability of Cyanobacterial Blooms in Lake Erie." 2012. Available at http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0042444#s3.

The Ohio Legislature. "House Bill 61." 2015. Available at https://www.legislature.ohio.gov/legislation/legislation-status?id=GA131-HB-61.

The Ohio Legislature. "Senate Bill 1." 2015. Available at https://www.legislature.ohio.gov/legislation/legislation-summary?id=GA131-SB-1.

Thomas, Mike, Biesinger, Zy, Deller, John, Hosack, Mike, Kocovsky, Patrick, MacDougall, Tom, Markham, Jim, Perez-Fuentetaja, Alicia, Weimer, Eric, & Witzel, Larry. "Report of the Lake Erie Forage Task Group." Lake Erie Committee. March, 2014. Available at http://www.glfc.org/lakecom/lec/FTG_docs/annual_reports/FTG_report_2014.pdf.

Tuholske, Jack, & Kilbert, Ken. "Moving Forward: Solutions to Lake Erie's Harmful Algal Blooms." Lucas County, Ohio Board of County Commissioners. April 15, 2015 available at http://ftpcontent4.worldnow.com/wtol/pdf/KilbertAlgaeResearch.pdf.

USEPA. "Limnology Program." 2012. Available at http://www.epa.gov/grtlakes/monitoring/limnology/index.html.

USEPA. "Oxygen Depletion in Lake Erie." Great Lake Monitoring. June 26, 2012. Available at http://www.epa.gov/glnpo/monitoring/d_o/index.html.

Wynee, Timothy T., Stumpf, Richard P, Tomlinson, Michelle C., et. al. "Evolutoin of a cyanobacterial bloom forecast system in western Lake Erie: Development and initial

evaluation." Journal of Great Lakes Research, 2013. Available at, http://www.glerl.noaa.gov/pubs/fulltext/2013/20130007.pdf.

Zhou, Yuntao, et. al. "Spatial and Temporal Trends in Lake Erie Hypoxia, 1987-2007." Environmental Science & Technology, December 13, 2012. Available at http://pubs.acs.org/doi/pdf/10.1021/es303401b.

Variable Name	**Description**
MauTP	Maumee River Total Phosphorus
MauTSRP	Maumee River Total Dissolved Reactive Phosphorus
MauTN	Maumee River Total Nitrogen
SanTP	Sandusky River Total Phosphorus
SanTSRP	Sandusky River Total Dissolved Reactive Phosphorus
SanTN	Sandusky River Total Nitrogen
Temp	Temperature
Chol	Chlorphyll-a
HypDO	Dissolved Oxygen

Table 11: Variable Name & Description

Obs	id	MauTP	MauTSRP	MauTN	SanTP	SanTSRP	SanTN	Temp	Chol	HypDO
1	1	242.32	53.948	7142.93	134.852	18.942	2075.29	66.618	4.5	1.75
2	2	351.81	32.999	3127.31	25.266	1.211	691.09	64.618	3.2	2.20
3	3	1295.11	102.221	20577.05	257.595	13.006	4613.32	63.488	.	2.95
4	4	965.47	131.887	19605.12	204.181	18.331	3619.87	63.260	0.9	3.50
5	5	692.52	86.647	10051.56	55.176	4.275	960.22	66.244	5.3	1.10
6	6	576.26	60.304	12578.94	87.918	6.144	2346.05	58.919	4.8	2.90
7	7	1026.23	123.616	11250.76	227.376	21.000	2434.57	64.724	3.7	3.89
8	8	706.06	20.388	11269.63	133.119	1.885	2646.94	62.911	.	5.10
9	9	866.85	68.095	16922.03	228.480	15.412	3080.43	64.813	.	5.50
10	10	1302.60	143.455	28194.14	349.991	20.728	4314.41	62.228	6.8	6.55
11	11	1302.96	187.258	24437.40	301.739	30.778	5196.79	62.033	2.0	4.40
12	12	1306.56	170.471	14430.88	201.427	21.111	3091.91	66.008	2.3	2.75
13	13	803.35	105.376	16863.26	73.863	10.393	2397.61	65.854	1.9	2.90
14	14	940.66	145.302	21225.88	115.917	25.203	3678.71	64.545	1.1	2.70
15	15	488.56	106.856	16414.10	.	.	.	63.707	6.2	1.90
16	16	869.55	150.488	12612.76	258.904	22.079	3656.19	65.732	1.8	5.10
17	17	1408.90	322.275	27580.23	210.331	37.531	3875.06	62.033	8.9	2.15
18	18	892.48	197.060	15868.91	338.363	52.168	4578.74	63.504	4.0	5.40
19	19	308.58	84.362	6984.08	128.862	37.178	2121.76	65.211	13.6	2.85
20	20	577.12	124.058	14251.50	162.827	33.128	3279.33	66.350	3.9	3.35
21	21	1028.90	253.258	9888.79	223.408	41.336	1804.27	64.114	18.9	3.85
22	22	1315.14	264.841	13424.73	334.367	61.454	2966.93	63.732	10.6	3.95
23	23	1315.54	199.268	13680.39	137.066	19.373	1717.36	63.008	8.7	2.80
24	24	1267.64	311.538	20765.89	306.363	63.873	4835.83	66.390	6.7	2.77
25	25	2281.48	429.049	24164.31	685.403	119.721	4649.28	64.748	13.5	3.28

Table 12: Full dataset with missing observations

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations									
	MauTP	MauTSRP	MauTN	SanTP	SanTSRP	SanTN	Temp	Chol	HypDO
MauTP	1.00000 25	0.82197 <.0001 25	0.69444 0.0001 25	0.80544 <.0001 24	0.66404 0.0004 24	0.60187 0.0019 24	-0.15764 0.4517 25	0.27857 0.2093 22	0.22334 0.2832 25
MauTSRP	0.82197 <.0001 25	1.00000 25	0.53956 0.0054 25	0.75204 <.0001 24	0.86944 <.0001 24	0.50397 0.0120 24	-0.00211 0.9920 25	0.50702 0.0160 22	0.01803 0.9318 25
MauTN	0.69444 0.0001 25	0.53956 0.0054 25	1.00000 25	0.56286 0.0042 24	0.36303 0.0812 24	0.81872 <.0001 24	-0.32413 0.1140 25	-0.06970 0.7579 22	0.27759 0.1791 25
SanTP	0.80544 <.0001 24	0.75204 <.0001 24	0.56286 0.0042 24	1.00000 24	0.84899 <.0001 24	0.69863 0.0001 24	-0.05060 0.8144 24	0.35181 0.1178 21	0.41596 0.0432 24
SanTSRP	0.66404 0.0004 24	0.86944 <.0001 24	0.36303 0.0812 24	0.84899 <.0001 24	1.00000 24	0.49663 0.0136 24	0.12692 0.5545 24	0.53912 0.0117 21	0.06963 0.7465 24
SanTN	0.60187 0.0019 24	0.50397 0.0120 24	0.81872 <.0001 24	0.69863 0.0001 24	0.49663 0.0136 24	1.00000 24	-0.17642 0.4096 24	-0.13784 0.5513 21	0.43954 0.0316 24
Temp	-0.15764 0.4517 25	-0.00211 0.9920 25	-0.32413 0.1140 25	-0.05060 0.8144 24	0.12692 0.5545 24	-0.17642 0.4096 24	1.00000 25	-0.07349 0.7452 22	-0.25824 0.2126 25
Chol	0.27857 0.2093 22	0.50702 0.0160 22	-0.06970 0.7579 22	0.35181 0.1178 21	0.53912 0.0117 21	-0.13784 0.5513 21	-0.07349 0.7452 22	1.00000 22	-0.00633 0.9777 22
HypDO	0.22334 0.2832 25	0.01803 0.9318 25	0.27759 0.1791 25	0.41596 0.0432 24	0.06963 0.7465 24	0.43954 0.0316 24	-0.25824 0.2126 25	-0.00633 0.9777 22	1.00000 25

Table 13: Correlation of original data

Obs	id	MauTP	MauTSRP	MauTN	SanTP	SanTSRP	SanTN	Temp	Chol	HypDO
1	1	242.32	53.948	7142.93	134.852	18.942	2075.29	66.618	4.5000	1.75
2	2	351.81	32.999	3127.31	25.266	1.211	691.09	64.618	3.2000	2.20
3	3	1295.11	102.221	20577.05	257.595	13.006	4613.32	63.488	6.0591	2.95
4	4	965.47	131.887	19605.12	204.181	18.331	3619.87	63.260	0.9000	3.50
5	5	692.52	86.647	10051.56	55.176	4.275	960.22	66.244	5.3000	1.10
6	6	576.26	60.304	12578.94	87.918	6.144	2346.05	58.919	4.8000	2.90
7	7	1026.23	123.616	11250.76	227.376	21.000	2434.57	64.724	3.7000	3.89
8	8	706.06	20.388	11269.63	133.119	1.885	2646.94	62.911	6.0591	5.10
9	9	866.85	68.095	16922.03	228.480	15.412	3080.43	64.813	6.0591	5.50
10	10	1302.60	143.455	28194.14	349.991	20.728	4314.41	62.228	6.8000	6.55
11	11	1302.96	187.258	24437.40	301.739	30.778	5196.79	62.033	2.0000	4.40
12	12	1306.56	170.471	14430.88	201.427	21.111	3091.91	66.008	2.3000	2.75
13	13	803.35	105.376	16863.26	73.863	10.393	2397.61	65.854	1.9000	2.90
14	14	940.66	145.302	21225.88	115.917	25.203	3678.71	64.545	1.1000	2.70
15	15	488.56	106.856	16414.10	106.164	16.562	3235.28	63.707	6.2000	1.90
16	16	869.55	150.488	12612.76	258.904	22.079	3656.19	65.732	1.8000	5.10
17	17	1408.90	322.275	27580.23	210.331	37.531	3875.06	62.033	8.9000	2.15
18	18	892.48	197.060	15868.91	338.363	52.168	4578.74	63.504	4.0000	5.40
19	19	308.58	84.362	6984.08	128.862	37.178	2121.76	65.211	13.6000	2.85
20	20	577.12	124.058	14251.50	162.827	33.128	3279.33	66.350	3.9000	3.35
21	21	1028.90	253.258	9888.79	223.408	41.336	1804.27	64.114	18.9000	3.85
22	22	1315.14	264.841	13424.73	334.367	61.454	2966.93	63.732	10.6000	3.95
23	23	1315.54	199.268	13680.39	137.066	19.373	1717.36	63.008	8.7000	2.80
24	24	1267.64	311.538	20765.89	306.363	63.873	4835.83	66.390	6.7000	2.77
25	25	2281.48	429.049	24164.31	685.403	119.721	4649.28	64.748	13.5000	3.28

Table 14: Full dataset, missing values estimated

One-way ANOVA: Chol^.23 versus OrdBloom

```
Analysis of Variance

Source    DF   Adj SS   Adj MS  F-Value  P-Value
OrdBloom   1  0.04829  0.04829     1.09    0.327
Error      8  0.35399  0.04425
Total      9  0.40228


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.210354  12.00%      1.00%       0.00%


Means

OrdBloom  N    Mean   StDev       95% CI
0         3   1.416   0.211  ( 1.136,  1.696)
1         7  1.5678  0.2100  (1.3844, 1.7511)

Pooled StDev = 0.210354
```

Table 15: ANOVA Analysis – transformed chlorophyll-a versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: HypDO versus OrdBloom

```
Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
OrdBloom   1   2.031   2.031     0.91    0.368
Error      8  17.820   2.228
Total      9  19.851


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.49250  10.23%      0.00%       0.00%


Means

OrdBloom  N   Mean  StDev      95% CI
0         3   4.28   2.33  ( 2.30,  6.27)
1         7  3.300  1.079  (1.999, 4.601)

Pooled StDev = 1.49250
```

Table 16: One-way ANOVA – dissolved oxygen versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: MauTP versus OrdBloom

```
Analysis of Variance

Source    DF   Adj SS  Adj MS  F-Value  P-Value
OrdBloom   1   282659  282659     1.49    0.257
Error      8  1517192  189649
Total      9  1799851


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
435.487  15.70%      5.17%       0.00%


Means

OrdBloom  N  Mean  StDev     95% CI
0         3  1031    470  ( 452, 1611)
1         7  1398    423  (1019, 1778)

Pooled StDev = 435.487
```

Table 17: One-way ANOVA – Maumee total phosphorus versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: MauTSRP versus OrdBloom

```
Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
OrdBloom   1   32701   32701     4.91    0.057
Error      8   53248    6656
Total      9   85949


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
81.5847  38.05%     30.30%      12.32%


Means

OrdBloom  N   Mean  StDev      95% CI
0         3  145.9   40.3  ( 37.2, 254.5)
1         7  270.6   91.3  (199.5, 341.8)

Pooled StDev = 81.5847
```

Table 18: One-way ANOVA – Maumee total dissolved reactive phosphorus versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: MauTN versus OrdBloom

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
OrdBloom	1	41695200	41695200	1.26	0.294
Error	8	264530326	33066291		
Total	9	306225526			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5750.33	13.62%	2.82%	0.00%

Means

OrdBloom	N	Mean	StDev	95% CI
0	3	23015	6017	(15359, 30671)
1	7	18559	5659	(13547, 23571)

Pooled StDev = 5750.33

Table 19: One-way ANOVA – Maumee total nitrogen versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: SanTP versus OrdBloom

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
OrdBloom	1	8483	8483	0.30	0.599
Error	8	227042	28380		
Total	9	235525			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
168.464	3.60%	0.00%	0.00%

Means

OrdBloom	N	Mean	StDev	95% CI
0	3	252.6	129.1	(28.3, 476.9)
1	7	316.2	179.7	(169.4, 463.0)

Pooled StDev = 168.464

Table 20: One-way ANOVA Sandusky total phosphorus versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: SanTSRP versus OrdBloom

```
Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
OrdBloom   1    2007  2007.0     2.25    0.172
Error      8    7133   891.7
Total      9    9140


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
29.8608  21.96%     12.20%       0.00%


Means

OrdBloom  N   Mean  StDev       95% CI
0         3  22.69   7.31  (-17.07, 62.45)
1         7   53.6   34.2  (  27.6,  79.6)

Pooled StDev = 29.8608
```

Table 21: One-way ANOVA Sandusky total dissolved reactive phosphorus versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: SanTN versus OrdBloom

```
Analysis of Variance

Source    DF    Adj SS   Adj MS  F-Value  P-Value
OrdBloom   1    694889   694889     0.57    0.472
Error      8   9757422  1219678
Total      9  10452311


Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
1104.39  6.65%      0.00%       0.00%


Means

OrdBloom  N  Mean  StDev     95% CI
0         3  4249    982  (2778, 5719)
1         7  3674   1142  (2711, 4636)

Pooled StDev = 1104.39
```

Table 22: One-way ANOVA Sandusky total nitrogen versus ordinal bloom dataset removing ambiguous bloom years

One-way ANOVA: Temp versus OrdBloom

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
OrdBloom	1	5.028	5.028	2.39	0.161
Error	8	16.861	2.108		
Total	9	21.889			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.45178	22.97%	13.34%	0.00%

Means

OrdBloom	N	Mean	StDev	95% CI
0	3	62.656	0.915	(60.723, 64.589)
1	7	64.203	1.591	(62.938, 65.469)

Pooled StDev = 1.45178

Table 23: One-way ANOVA temperature versus ordinal bloom dataset removing ambiguous bloom years

Models:

Regression Analysis: Chol versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	34.18	34.18	1.85	0.187
MauTP	1	34.18	34.18	1.85	0.187
Error	23	424.07	18.44		
Total	24	458.25			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.29395	7.46%	3.43%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.48	2.08	1.67	0.108	
MauTP	0.00267	0.00196	1.36	0.187	1.00

Regression Equation

Chol = 3.48 + 0.00267 MauTP

Table 24: Regression Analysis: chlorophyll-a versus Maumee total phosphorus (full dataset)

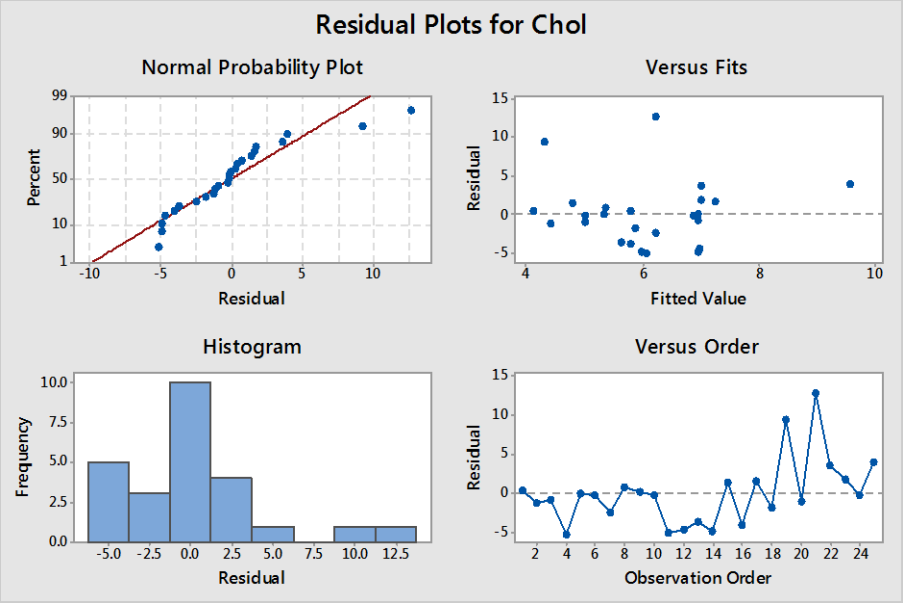

Figure 2: Residual Plots for chlorophyll-a versus Maumee total phosphorus (full dataset)

Regression Analysis: Chol^.23 versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.07934	0.07934	1.29	0.268
MauTP	1	0.07934	0.07934	1.29	0.268
Error	23	1.41674	0.06160		
Total	24	1.49608			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.248188	5.30%	1.19%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.323	0.120	11.00	0.000	
MauTP	0.000129	0.000114	1.13	0.268	1.00

Regression Equation

Chol^.23 = 1.323 + 0.000129 MauTP

Table 25: Regression Analysis: transformed chlorophyll-a versus Maumee total phosphorus (full dataset)

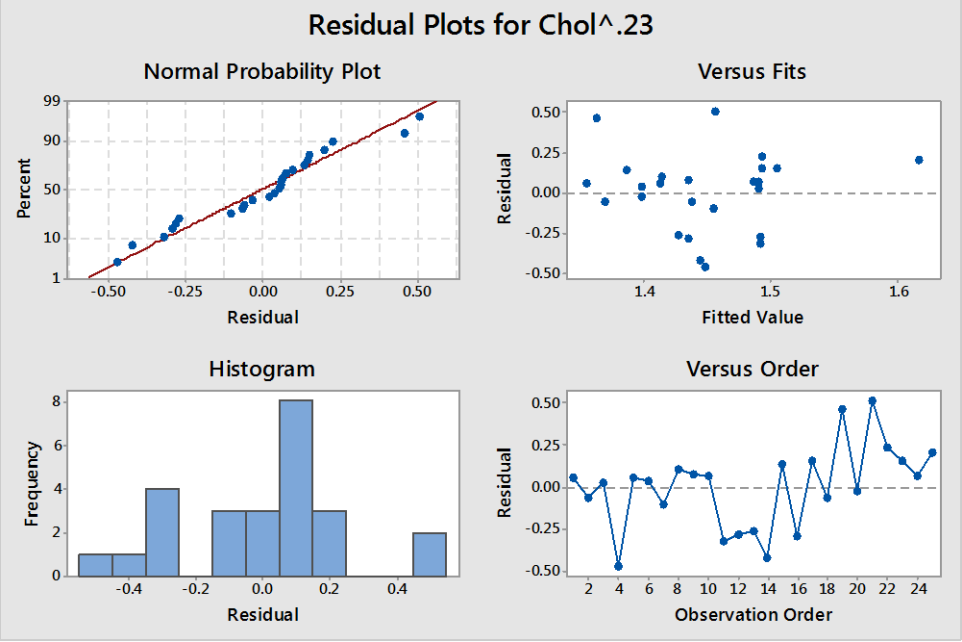

Figure 3: Residual Plots for transformed chlorophyll-a versus Maumee total phosphorus (full dataset)

Regression Analysis: HypDO versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2.085	2.085	1.21	0.283
MauTP	1	2.085	2.085	1.21	0.283
Error	23	39.724	1.727		
Total	24	41.810			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.31421	4.99%	0.86%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.786	0.637	4.37	0.000	
MauTP	0.000661	0.000601	1.10	0.283	1.00

Regression Equation

HypDO = 2.786 + 0.000661 MauTP

Table 26: Regression Analysis: dissolved oxygen versus Maumee total phosphorus (full dataset)

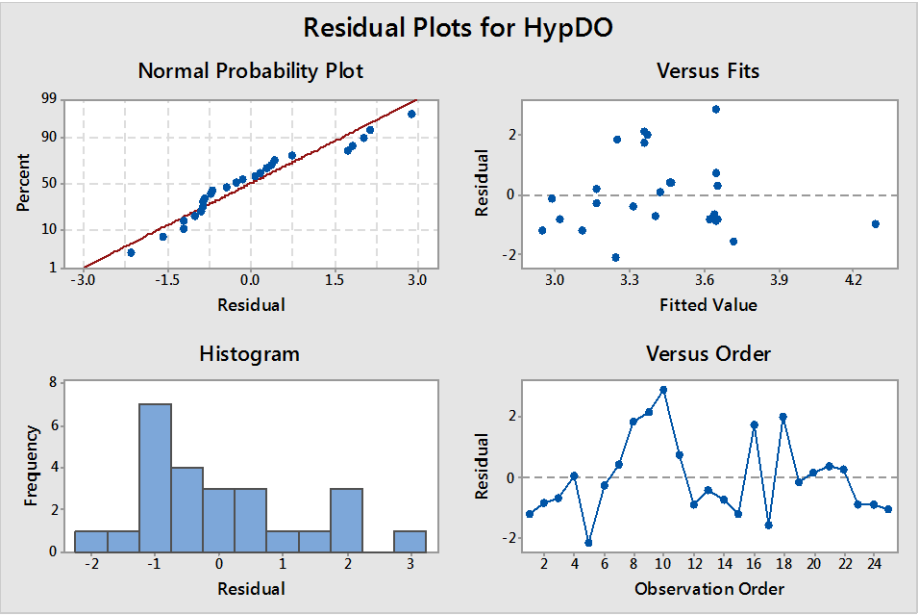

Figure 4: Residual Plots for dissolved oxygen versus Maumee total phosphorus (full dataset)

Regression Analysis: Chol versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	18.07	18.07	0.66	0.430
MauTP	1	18.07	18.07	0.66	0.430
Error	14	383.06	27.36		
Total	15	401.13			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.23081	4.51%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.35	3.43	1.27	0.225	
MauTP	0.00237	0.00291	0.81	0.430	1.00

Regression Equation

Chol = 4.35 + 0.00237 MauTP

Table 27: Regression Analysis: chlorophyll-a versus Maumee total phosphorus (16-year dataset)

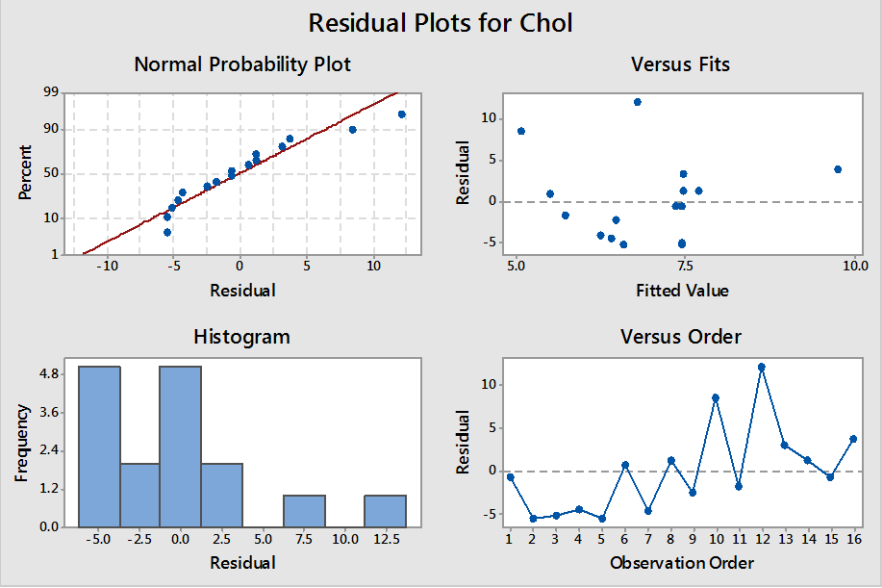

Figure 5: Residual Plots for chlorophyll-a versus Maumee total phosphorus (16-year dataset)

Regression Analysis: Chol^.23 versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.06169	0.06169	0.75	0.401
MauTP	1	0.06169	0.06169	0.75	0.401
Error	14	1.14970	0.08212		
Total	15	1.21139			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.286568	5.09%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.331	0.188	7.08	0.000	
MauTP	0.000138	0.000160	0.87	0.401	1.00

Regression Equation

Chol^.23 = 1.331 + 0.000138 MauTP

Table 28: Regression Analysis: transformed chlorophyll-a versus Maumee total phosphorus (16-year dataset)

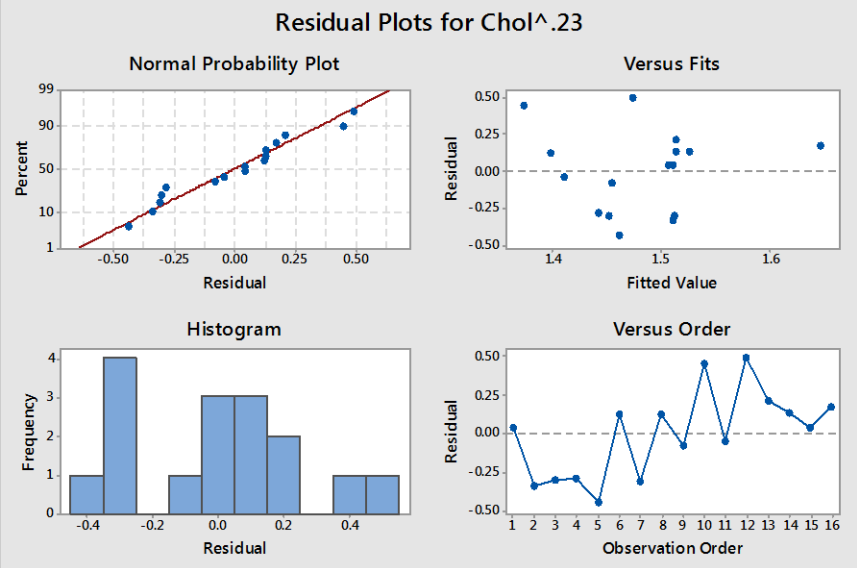

Figure 6: Residual Plots for transformed chlorophyll-a versus Maumee total phosphorus (16-year dataset)

Regression Analysis: HypDO versus MauTP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.2454	0.2454	0.14	0.710
MauTP	1	0.2454	0.2454	0.14	0.710
Error	14	23.7928	1.6995		
Total	15	24.0382			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.30364	1.02%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.244	0.855	3.79	0.002	
MauTP	0.000276	0.000726	0.38	0.710	1.00

Regression Equation

HypDO = 3.244 + 0.000276 MauTP

Table 29: Regression Analysis: dissolved oxygen versus Maumee total phosphorus (16-year dataset)

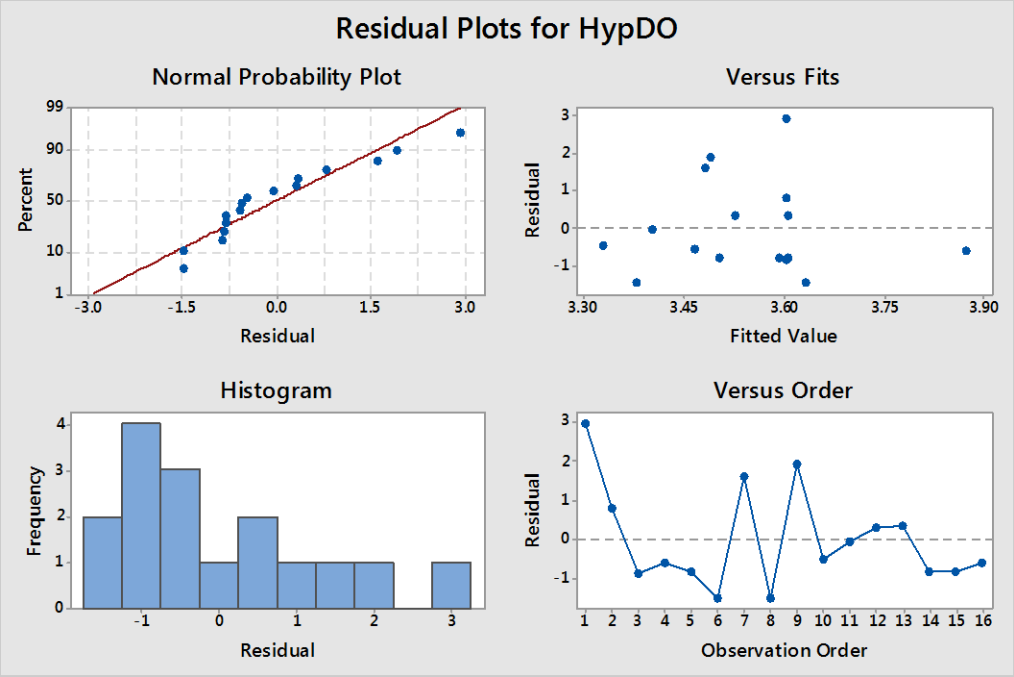

Figure 7: Residual Plots for dissolved oxygen versus Maumee total phosphorus (16-year dataset)

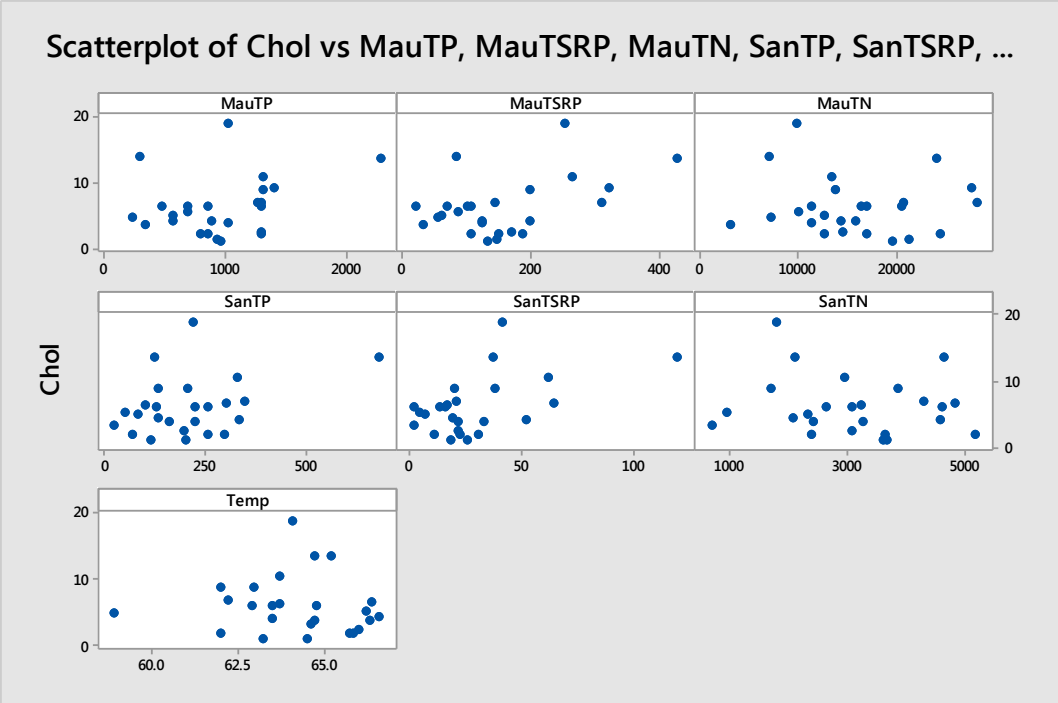

Figure 8: Scatterplot of chlorophyll-a versus all independent variables (full dataset)

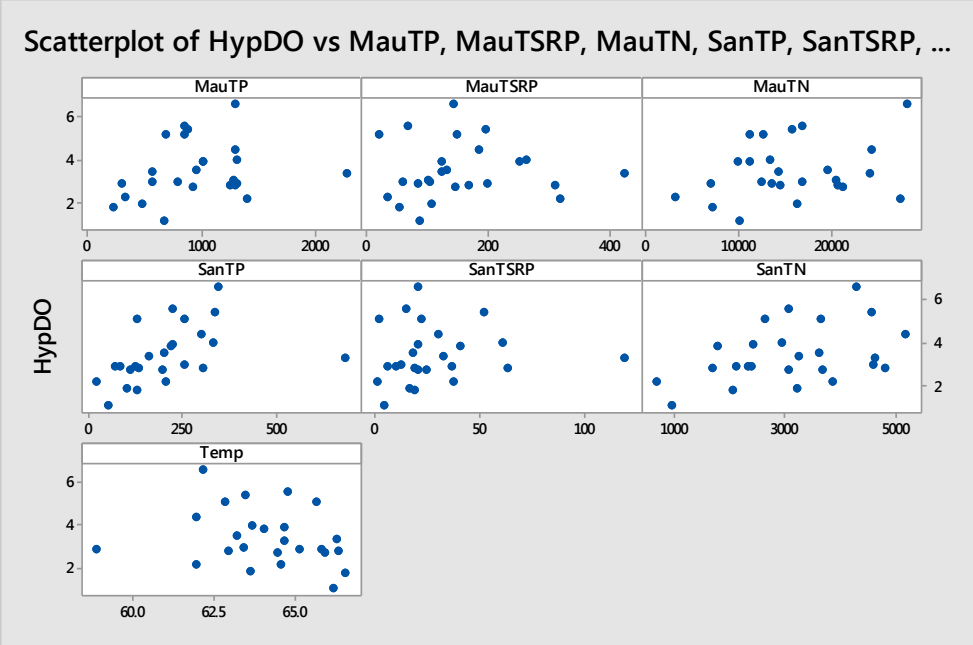

Figure 9: Scatterplot of dissolved oxygen versus all independent variables (full dataset)

Regression Analysis: Chol versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	255.844	36.549	3.07	0.028
MauTP	1	7.354	7.354	0.62	0.443
MauTSRP	1	7.856	7.856	0.66	0.428
MauTN	1	1.128	1.128	0.09	0.762
SanTP	1	8.500	8.500	0.71	0.410
SanTSRP	1	3.382	3.382	0.28	0.601
SanTN	1	49.904	49.904	4.19	0.056
Temp	1	23.855	23.855	2.00	0.175
Error	17	202.409	11.906		
Total	24	458.253			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.45057	55.83%	37.64%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	49.1	28.8	1.70	0.107	
MauTP	-0.00388	0.00493	-0.79	0.443	9.77
MauTSRP	0.0204	0.0251	0.81	0.428	12.36
MauTN	0.000083	0.000270	0.31	0.762	6.03
SanTP	0.0158	0.0188	0.84	0.410	12.92
SanTSRP	0.057	0.107	0.53	0.601	14.77
SanTN	-0.00280	0.00137	-2.05	0.056	5.52
Temp	-0.624	0.441	-1.42	0.175	1.25

Regression Equation

Chol = 49.1 - 0.00388 MauTP + 0.0204 MauTSRP + 0.000083 MauTN + 0.0158 SanTP
+ 0.057 SanTSRP
- 0.00280 SanTN - 0.624 Temp

Table 30: Regression Analysis: chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (full dataset)

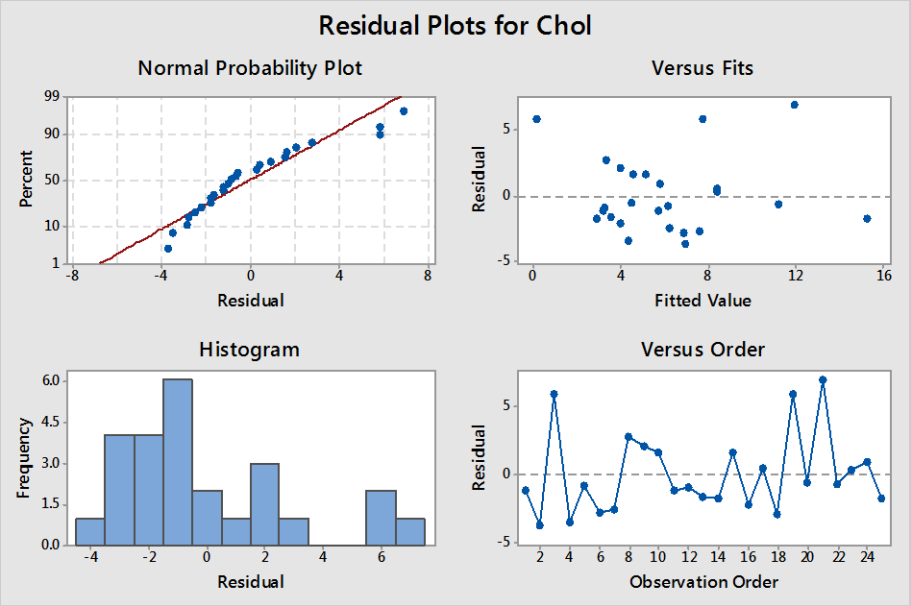

Figure 10: Residual Plots for chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (full dataset)

Regression Analysis: Chol versus SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	244.33	81.44	7.99	0.001
SanTSRP	1	231.94	231.94	22.77	0.000
SanTN	1	115.15	115.15	11.30	0.003
Temp	1	34.86	34.86	3.42	0.078
Error	21	213.93	10.19		
Total	24	458.25			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.19169	53.32%	46.65%	26.76%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	54.2	25.0	2.17	0.042	
SanTSRP	0.1454	0.0305	4.77	0.000	1.41
SanTN	-0.002170	0.000645	-3.36	0.003	1.43
Temp	-0.709	0.383	-1.85	0.078	1.10

Regression Equation

Chol = 54.2 + 0.1454 SanTSRP - 0.002170 SanTN - 0.709 Temp

Table 31: Regression Analysis: chlorophyll-a versus SanTSRP, SanTN, Temp (full dataset)

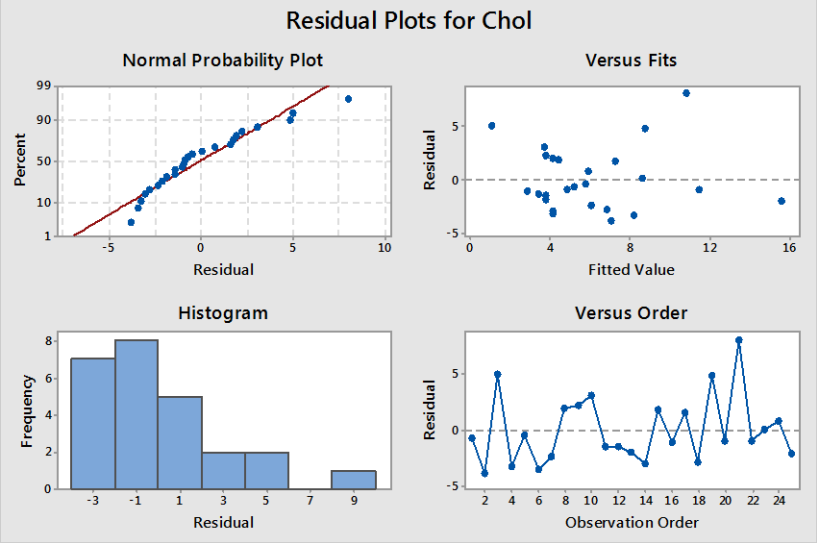

Figure 11: Residual Plots for chlorophyll-a versus SanTSRP, SanTN, Temp (full dataset)

Regression Analysis: Chol^.23 versus SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.6727	0.22422	5.72	0.005
SanTSRP	1	0.6220	0.62197	15.86	0.001
SanTN	1	0.3334	0.33336	8.50	0.008
Temp	1	0.1224	0.12244	3.12	0.092
Error	21	0.8234	0.03921		
Total	24	1.4961			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.198016	44.96%	37.10%	20.74%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.29	1.55	2.77	0.012	
SanTSRP	0.00753	0.00189	3.98	0.001	1.41
SanTN	-0.000117	0.000040	-2.92	0.008	1.43
Temp	-0.0420	0.0238	-1.77	0.092	1.10

Regression Equation

Chol^.23 = 4.29 + 0.00753 SanTSRP - 0.000117 SanTN - 0.0420 Temp

Table 32: Regression Analysis: transformed chlorophyll-a versus SanTSRP, SanTN, Temp (full dataset)

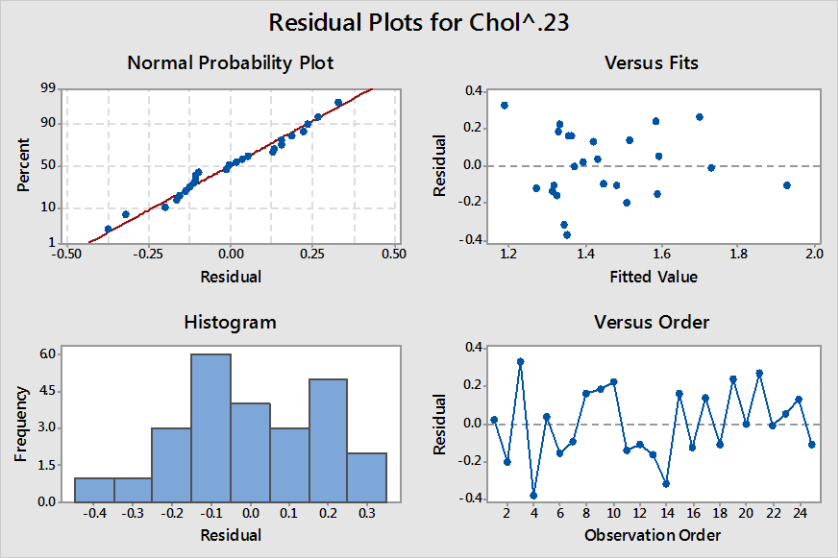

Figure 12: Residual Plots for transformed chlorophyll-a versus SanTSRP, SanTN, Temp (full dataset)

Regression Analysis: Chol^.23 versus SanTSRP, SanTN

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	0.5502	0.27511	6.40	0.006
SanTSRP	1	0.5231	0.52310	12.17	0.002
SanTN	1	0.2493	0.24927	5.80	0.025
Error	22	0.9459	0.04299		
Total	24	1.4961			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.207349	36.78%	31.03%	19.11%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.559	0.117	13.33	0.000	
SanTSRP	0.00668	0.00191	3.49	0.002	1.32
SanTN	-0.000097	0.000040	-2.41	0.025	1.32

Regression Equation

Chol^.23 = 1.559 + 0.00668 SanTSRP - 0.000097 SanTN

Table 33: Regression Analysis: transformed chlorophyll-a verus SanTSRP, SanTN (full dataset)

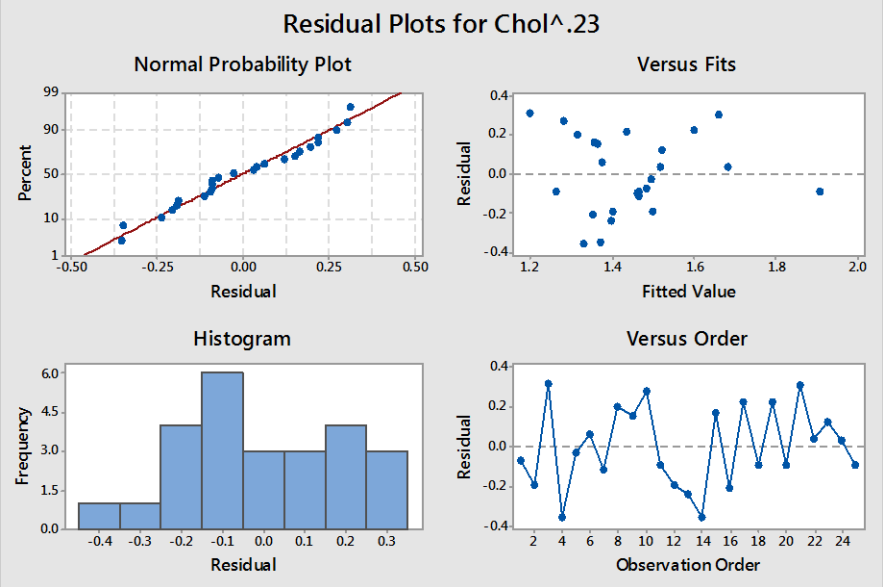

Figure 13: Residual Plots for transformed chlorophyll-a versus SanTSRP, SanTN (full dataset)

Regression Analysis: HypDO versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	23.7569	3.39384	3.20	0.024
MauTP	1	1.5310	1.53096	1.44	0.246
MauTSRP	1	0.1562	0.15622	0.15	0.706
MauTN	1	0.0360	0.03601	0.03	0.856
SanTP	1	9.4835	9.48350	8.93	0.008
SanTSRP	1	3.7616	3.76157	3.54	0.077
SanTN	1	0.0614	0.06138	0.06	0.813
Temp	1	0.6377	0.63769	0.60	0.449
Error	17	18.0531	1.06195		
Total	24	41.8100			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.03051	56.82%	39.04%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	9.33	8.62	1.08	0.294	
MauTP	-0.00177	0.00147	-1.20	0.246	9.77
MauTSRP	0.00287	0.00749	0.38	0.706	12.36
MauTN	-0.000015	0.000081	-0.18	0.856	6.03
SanTP	0.01673	0.00560	2.99	0.008	12.92
SanTSRP	-0.0599	0.0318	-1.88	0.077	14.77
SanTN	0.000098	0.000409	0.24	0.813	5.52
Temp	-0.102	0.132	-0.77	0.449	1.25

Regression Equation

HypDO = 9.33 - 0.00177 MauTP + 0.00287 MauTSRP - 0.000015 MauTN + 0.01673 SanTP
- 0.0599 SanTSRP + 0.000098 SanTN - 0.102 Temp

Table 34: Regression Analysis: dissolved oxygen versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (full dataset)

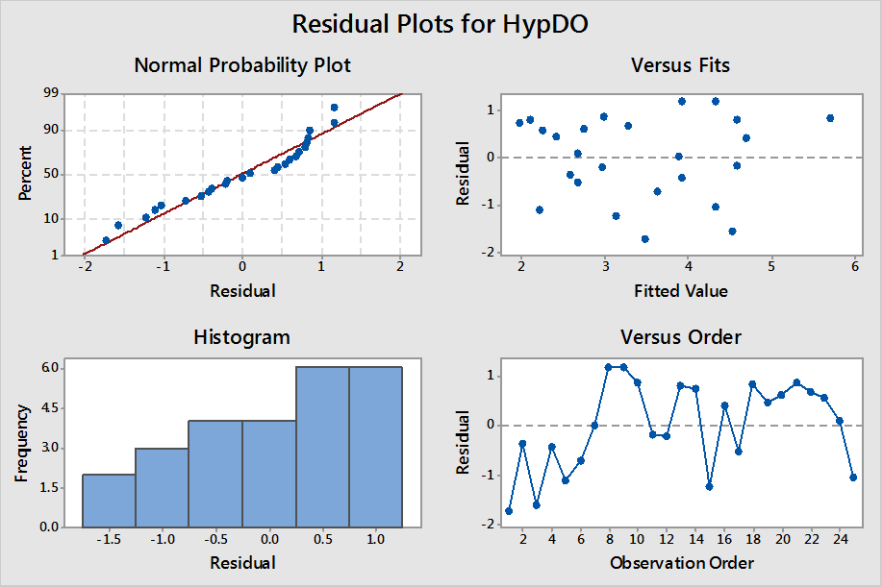

Figure 14: Residual Plots for dissolved oxygen versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (full dataset)

Regression Analysis: HypDO versus MauTP, SanTP, SanTSRP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	22.792	7.5973	8.39	0.001
MauTP	1	2.935	2.9353	3.24	0.086
SanTP	1	20.454	20.4542	22.59	0.000
SanTSRP	1	12.679	12.6793	14.00	0.001
Error	21	19.018	0.9056		
Total	24	41.810			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.951644	54.51%	48.01%	15.01%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.766	0.469	5.90	0.000	
MauTP	-0.001343	0.000746	-1.80	0.086	2.94
SanTP	0.01657	0.00349	4.75	0.000	5.87
SanTSRP	-0.0544	0.0145	-3.74	0.001	3.61

Regression Equation

HypDO = 2.766 - 0.001343 MauTP + 0.01657 SanTP - 0.0544 SanTSRP

Table 35: Regression Analysis: dissolved oxygen versus MauTP, SanTP, SanTSRP (full dataset)

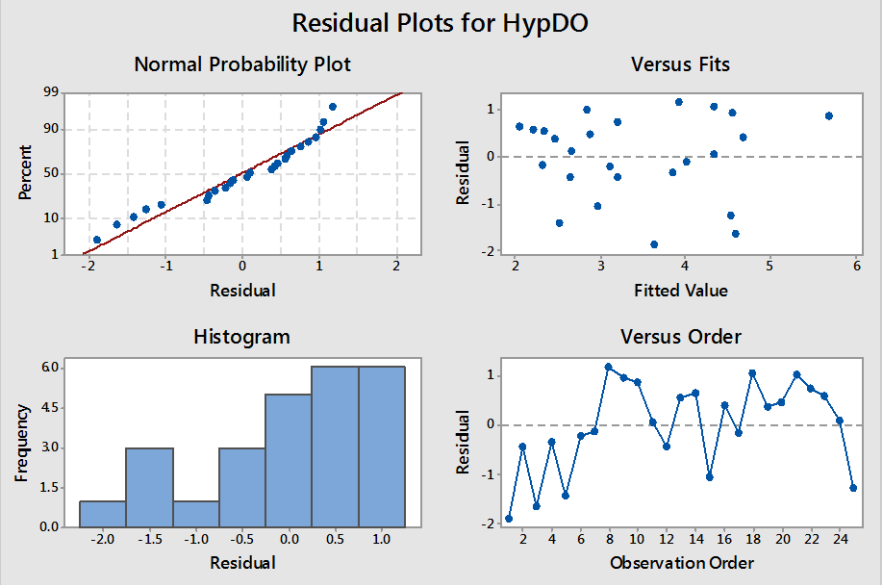

Figure 15: Residual Plots for dissolved oxygen versus MauTP, SanTP, SanTSRP (full dataset)

Regression Analysis: HypDO versus SanTP, SanTSRP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	19.86	9.9283	9.95	0.001
SanTP	1	19.51	19.5116	19.55	0.000
SanTSRP	1	11.85	11.8545	11.88	0.002
Error	22	21.95	0.9979		
Total	24	41.81			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.998940	47.49%	42.72%	9.64%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.241	0.385	5.82	0.000	
SanTP	0.01266	0.00286	4.42	0.000	3.59
SanTSRP	-0.0525	0.0152	-3.45	0.002	3.59

Regression Equation

HypDO = 2.241 + 0.01266 SanTP - 0.0525 SanTSRP

Table 36: Regression Analysis: dissolved oxygen versus SanTP, SanTSRP (full dataset)

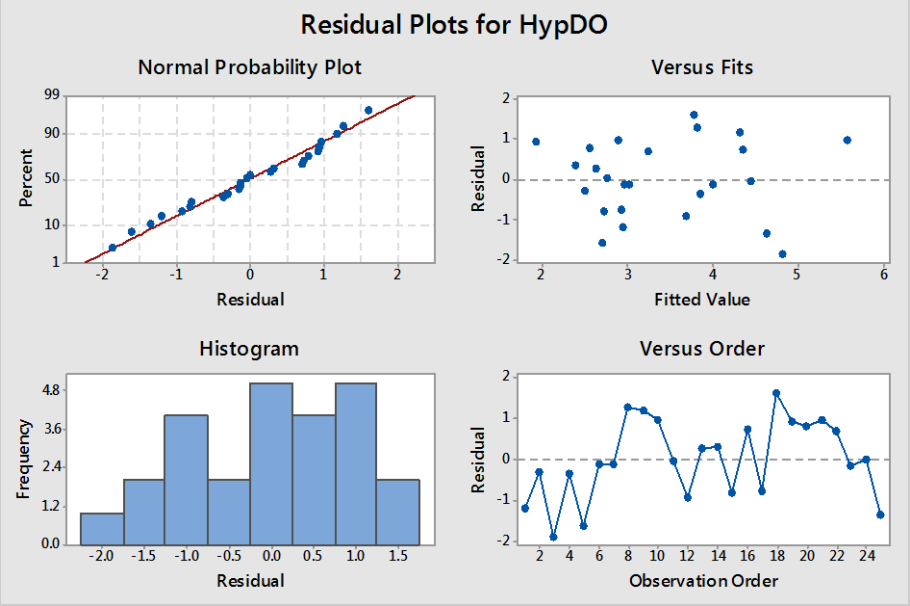

Figure 16: Residual Plots for dissolved oxygen versus SanTP, SanTSRP (full dataset)

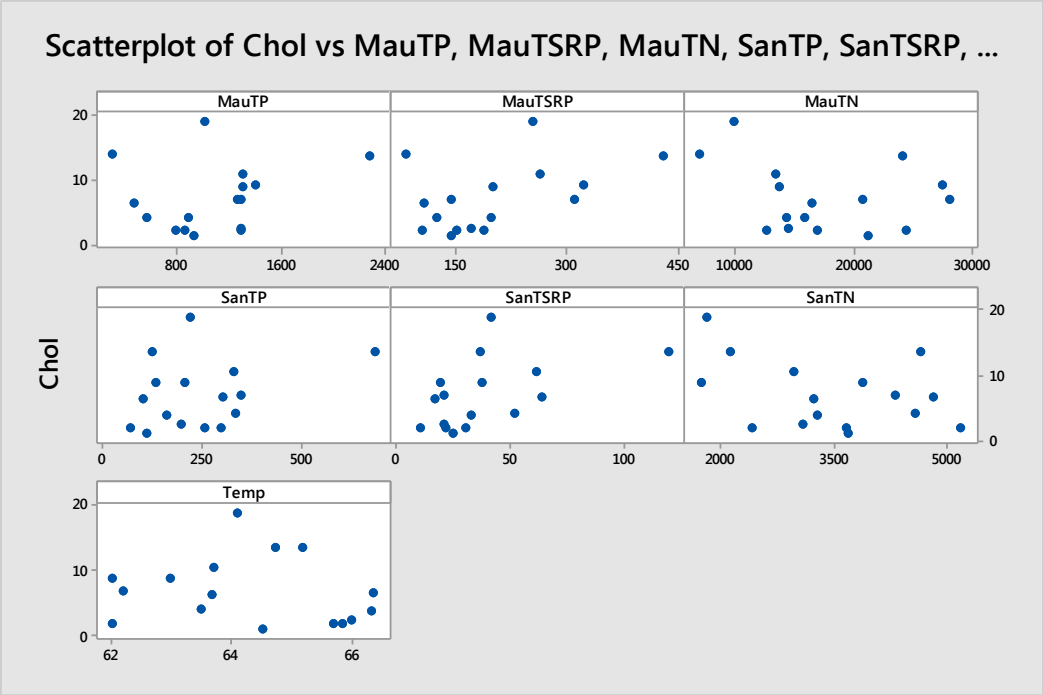

Figure 17: Scatterplot of chlorophyll-a versus all independent variables (16-year dataset)

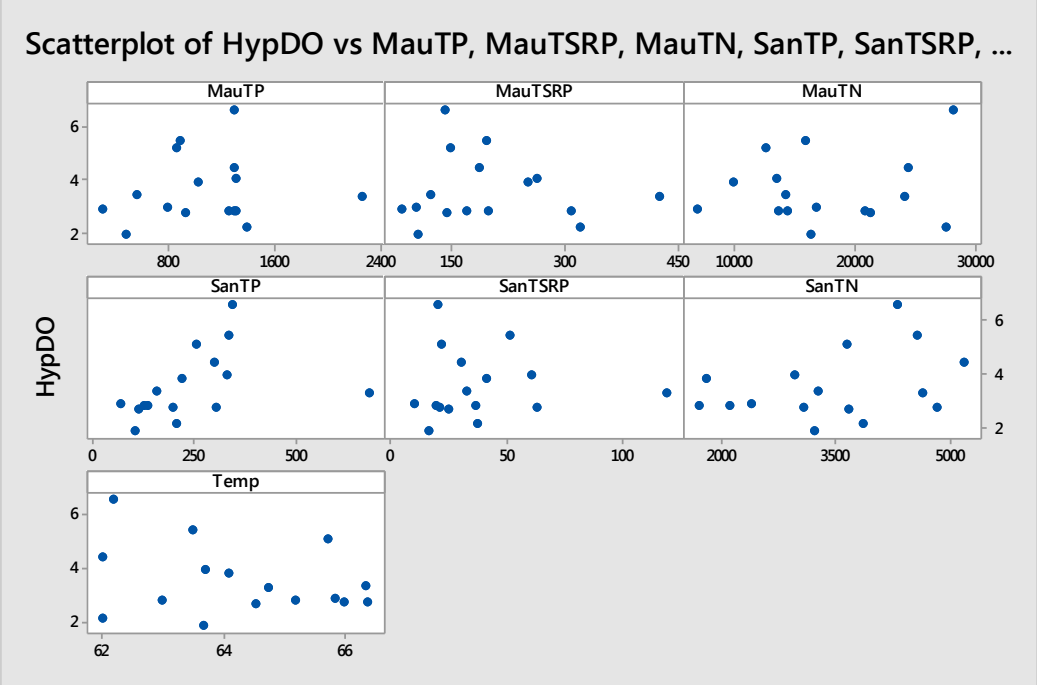

Figure 18: Scatterplot of dissolved oxygen versus all independent variables (16-year dataset)

Regression Analysis: Chol versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	346.368	49.481	7.23	0.006
MauTP	1	42.147	42.147	6.16	0.038
MauTSRP	1	26.359	26.359	3.85	0.085
MauTN	1	8.484	8.484	1.24	0.298
SanTP	1	24.334	24.334	3.55	0.096
SanTSRP	1	0.502	0.502	0.07	0.793
SanTN	1	101.371	101.371	14.81	0.005
Temp	1	15.066	15.066	2.20	0.176
Error	8	54.766	6.846		
Total	15	401.134			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.61645	86.35%	74.40%	29.94%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	71.6	38.2	1.88	0.097	
MauTP	-0.01203	0.00485	-2.48	0.038	11.07
MauTSRP	0.0447	0.0228	1.96	0.085	10.11
MauTN	0.000283	0.000254	1.11	0.298	5.48
SanTP	0.0321	0.0170	1.89	0.096	13.91
SanTSRP	0.0240	0.0885	0.27	0.793	12.31
SanTN	-0.00533	0.00138	-3.85	0.005	4.96
Temp	-0.869	0.585	-1.48	0.176	1.73

Regression Equation

Chol = 71.6 - 0.01203 MauTP + 0.0447 MauTSRP + 0.000283 MauTN + 0.0321 SanTP + 0.0240 SanTSRP - 0.00533 SanTN - 0.869 Temp

Table 37: Regression Analysis: chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (16-year dataset)

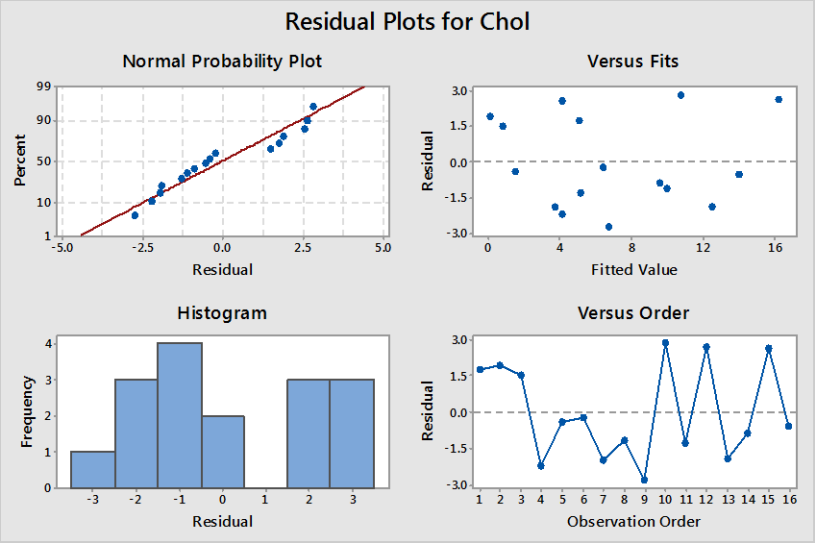

Figure 19: Residual Plots for chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (16-year dataset)

Regression Analysis: Chol versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	345.866	57.644	9.39	0.002
MauTP	1	68.320	68.320	11.13	0.009
MauTSRP	1	83.477	83.477	13.59	0.005
MauTN	1	8.763	8.763	1.43	0.263
SanTP	1	74.863	74.863	12.19	0.007
SanTN	1	109.439	109.439	17.82	0.002
Temp	1	15.452	15.452	2.52	0.147
Error	9	55.269	6.141		
Total	15	401.134			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.47810	86.22%	77.04%	50.87%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	67.6	33.3	2.03	0.073	
MauTP	-0.01276	0.00383	-3.34	0.009	7.68
MauTSRP	0.0495	0.0134	3.69	0.005	3.92
MauTN	0.000287	0.000240	1.19	0.263	5.46
SanTP	0.0356	0.0102	3.49	0.007	5.58
SanTN	-0.00541	0.00128	-4.22	0.002	4.73
Temp	-0.804	0.507	-1.59	0.147	1.45

Regression Equation

Chol = 67.6 - 0.01276 MauTP + 0.0495 MauTSRP + 0.000287 MauTN + 0.0356 SanTP
- 0.00541 SanTN
- 0.804 Temp

Table 38: Regression Analysis: chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp (16-year dataset)

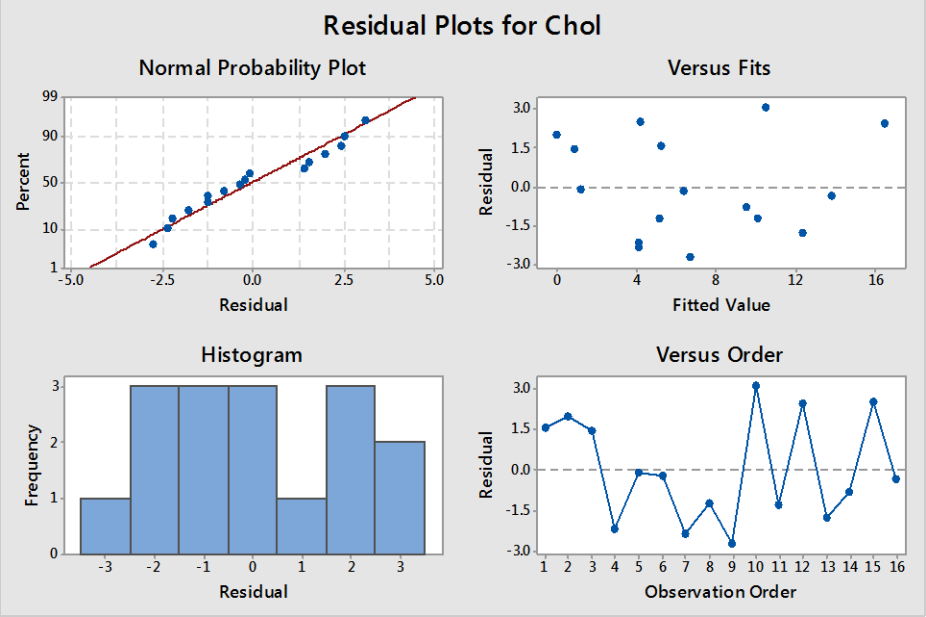

Figure 20: Residual Plots for chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp (16-year dataset)

Regression Analysis: Chol^.23 versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	1.01364	0.16894	7.69	0.004
MauTP	1	0.24713	0.24713	11.25	0.008
MauTSRP	1	0.27647	0.27647	12.58	0.006
MauTN	1	0.03877	0.03877	1.76	0.217
SanTP	1	0.23915	0.23915	10.88	0.009
SanTN	1	0.31464	0.31464	14.32	0.004
Temp	1	0.06076	0.06076	2.77	0.131
Error	9	0.19775	0.02197		
Total	15	1.21139			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.148231	83.68%	72.79%	35.65%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.17	1.99	2.60	0.029	
MauTP	-0.000767	0.000229	-3.35	0.008	7.68
MauTSRP	0.002851	0.000804	3.55	0.006	3.92
MauTN	0.000019	0.000014	1.33	0.217	5.46
SanTP	0.002013	0.000610	3.30	0.009	5.58
SanTN	-0.000290	0.000077	-3.78	0.004	4.73
Temp	-0.0504	0.0303	-1.66	0.131	1.45

Regression Equation

Chol^.23 = 5.17 - 0.000767 MauTP + 0.002851 MauTSRP + 0.000019 MauTN + 0.002013 SanTP - 0.000290 SanTN - 0.0504 Temp

Table 39: Regression Analysis: transformed chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp (16-year dataset)

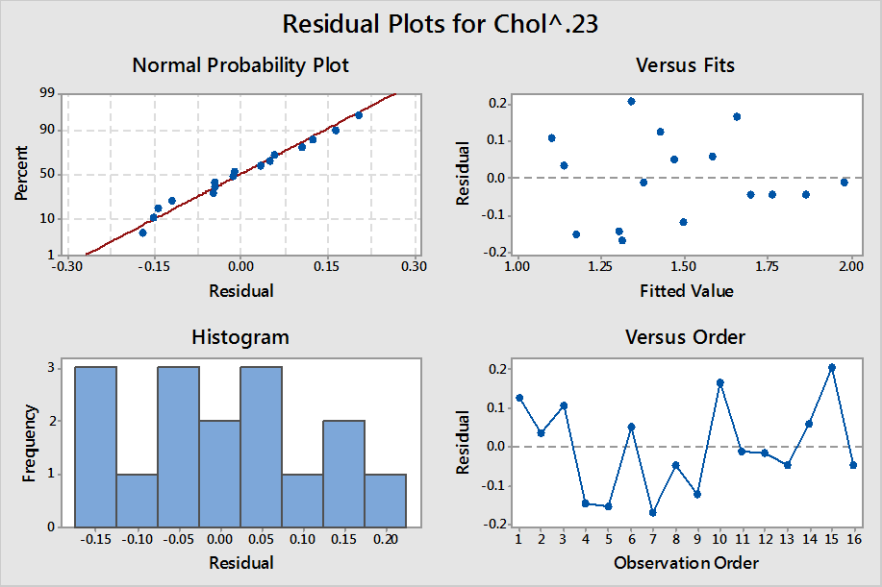

Figure 21: Residual Plots for transformed chlorophyll-a versus MauTP, MauTSRP, MauTN, SanTP, SanTN, Temp (16-year dataset)

Regression Analysis: Chol^.23 versus MauTP, MauTSRP, SanTP, SanTN

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	0.8291	0.20726	5.96	0.008
MauTP	1	0.1518	0.15179	4.37	0.061
MauTSRP	1	0.2195	0.21954	6.32	0.029
SanTP	1	0.1640	0.16397	4.72	0.053
SanTN	1	0.3998	0.39978	11.50	0.006
Error	11	0.3823	0.03476		
Total	15	1.2114			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.186435	68.44%	56.96%	30.68%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.791	0.190	9.42	0.000	
MauTP	-0.000448	0.000215	-2.09	0.061	4.27
MauTSRP	0.002506	0.000997	2.51	0.029	3.81
SanTP	0.001356	0.000625	2.17	0.053	3.70
SanTN	-0.000189	0.000056	-3.39	0.006	1.59

Regression Equation

Chol^.23 = 1.791 - 0.000448 MauTP + 0.002506 MauTSRP + 0.001356 SanTP - 0.000189 SanTN

Table 40: Regression Analysis: transformed chlorophyll-a versus MauTP, MauTSRP, SanTP, SanTN (16-year dataset)

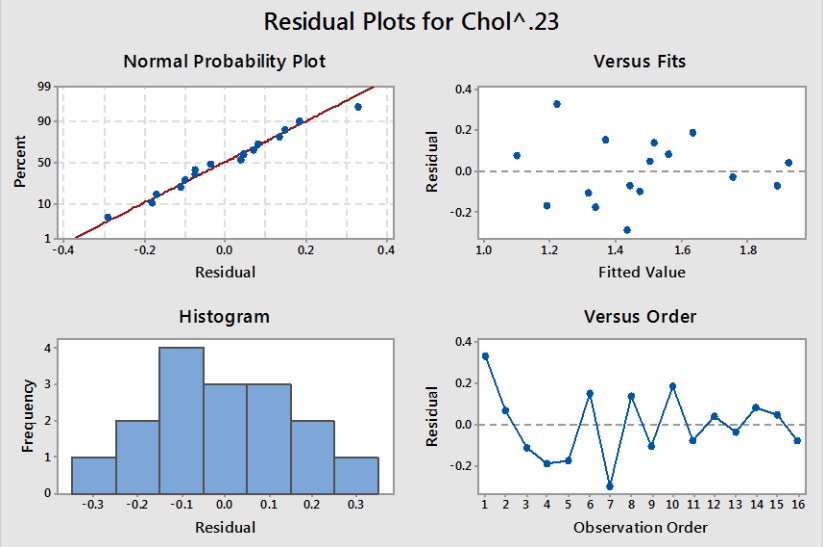

Figure 22: Residual Plots for transformed chlorophyll-a versus MauTP, MauTSRP, SanTP, SanTN (16-year dataset)

Regression Analysis: HypDO versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	18.6074	2.65820	3.92	0.037
MauTP	1	1.1610	1.16102	1.71	0.227
MauTSRP	1	0.0585	0.05852	0.09	0.777
MauTN	1	0.0841	0.08414	0.12	0.734
SanTP	1	8.4099	8.40991	12.39	0.008
SanTSRP	1	4.3100	4.31002	6.35	0.036
SanTN	1	0.0120	0.01204	0.02	0.897
Temp	1	0.1492	0.14915	0.22	0.652
Error	8	5.4308	0.67885		
Total	15	24.0382			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.823922	77.41%	57.64%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	9.6	12.0	0.80	0.447	
MauTP	-0.00200	0.00153	-1.31	0.227	11.07
MauTSRP	0.00211	0.00718	0.29	0.777	10.11
MauTN	-0.000028	0.000080	-0.35	0.734	5.48
SanTP	0.01884	0.00535	3.52	0.008	13.91
SanTSRP	-0.0702	0.0279	-2.52	0.036	12.31
SanTN	-0.000058	0.000436	-0.13	0.897	4.96
Temp	-0.086	0.184	-0.47	0.652	1.73

Regression Equation

HypDO = 9.6 - 0.00200 MauTP + 0.00211 MauTSRP - 0.000028 MauTN + 0.01884 SanTP
- 0.0702 SanTSRP - 0.000058 SanTN - 0.086 Temp

Table 41: Regression Analysis: dissolved oxygen versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (16-year dataset)

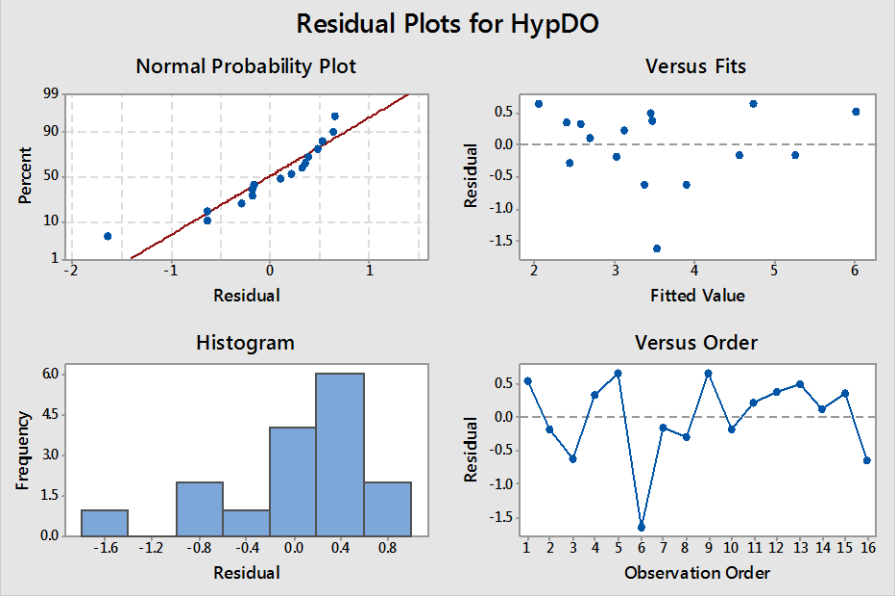

Figure 23: Residual Plots for dissolved oxygen versus MauTP, MauTSRP, MauTN, SanTP, SanTSRP, SanTN, Temp (16-year dataset)

Regression Analysis: HypDO versus MauTP, SanTP, SanTSRP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	18.038	6.0127	12.03	0.001
MauTP	1	3.965	3.9653	7.93	0.016
SanTP	1	17.671	17.6711	35.34	0.000
SanTSRP	1	11.240	11.2400	22.48	0.000
Error	12	6.000	0.5000		
Total	15	24.038			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.707104	75.04%	68.80%	34.98%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.580	0.468	7.65	0.000	
MauTP	-0.001801	0.000639	-2.82	0.016	2.64
SanTP	0.01766	0.00297	5.94	0.000	5.82
SanTSRP	-0.0632	0.0133	-4.74	0.000	3.82

Regression Equation

HypDO = 3.580 - 0.001801 MauTP + 0.01766 SanTP - 0.0632 SanTSRP

Table 42: Regression Analysis: dissolved oxygen versus MauTP, SanTP, SanTSRP (16-year dataset)

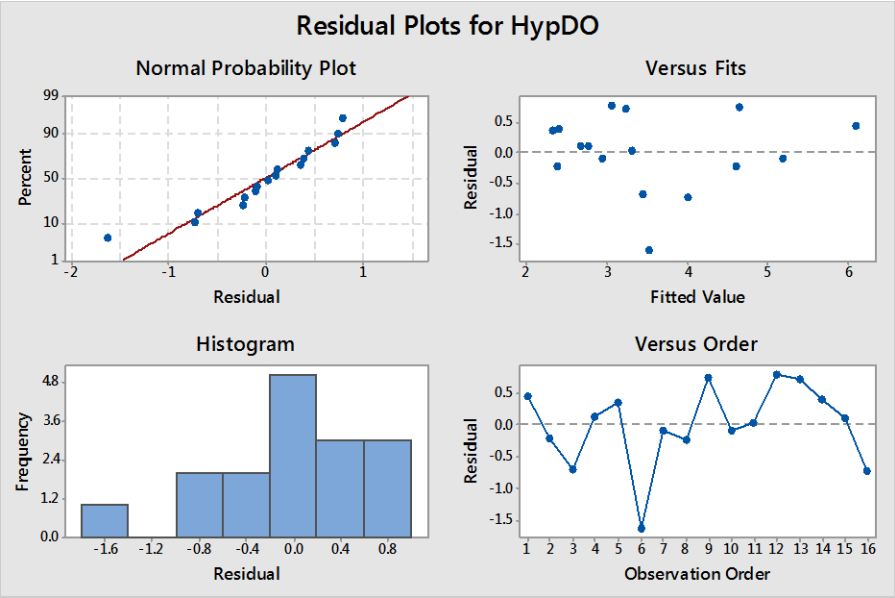

Figure 24: Residual Plots for dissolved oxygen versus MauTP, SanTP, SanTSRP (16-year dataset)