STRATEGIES FOR DEFINING CHEMICAL CONNECTIVITY BETWEEN STREAMS AND WETLANDS

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

JOHN BRITO

Dr. Enos C. Inniss, Thesis Supervisor

MAY 2013

The undersigned, appointed by the dean of the Graduate School,

have examined the thesis entitled

STRATEGIES FOR DEFINING CHEMICAL CONNECTIVITY

BETWEEN STREAMS AND WETLANDS

presented by John Brito,

a candidate for the degree of master of science

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Enos C. Inniss

Dr. Kathleen M. Trauth

Dr. Timothy C. Matisziw

DEDICATION

First off I would like to thank God for blessing me with countless opportunities. I would like to dedicate this thesis to my mother Paula Brito and my late father Antonio Brito Sr. Without their continued love and support I would not be the person I am today. I would also like to dedicate this thesis to my family and friends who have always believed in me and pushed me to strive for the best. Finally to all the teachers and professors throughout my education who have imparted me with a great lessons.

ACKNOWLEDGMENTS

First off I would like to express my sincere gratitude to Dr. Enos C. Inniss. He has truly been a great advisor and mentor to me these last two years. He has taught me valuable lessons on how to succeed as a research assistant and future professional. I would also like to thank Dr. Timothy Matisiziw for his continued support and guidance in my research project. It was a pleasure working with Dr. Matisiziw and he has always been patient and supportive of my endeavors. Finally I would like to thank Dr. Kathleen Trauth. Dr. Trauth has offered me great guidance and advice from the beginning of my masters programs.

I thank my committee members for taking the time to reading my thesis and giving me guidance on how to improve the work. In addition to my above committee members I would like to thank Dr. Raymond Semlitsch, Steve McIntosh and John Hudson from the MODNR, Dr. Bin Hua from Lincoln University, Guocheng Zhu, and my fellow researchers and lab mates.

This research was made possible by a grant from the U.S. Environmental Protection Agency (EPA), Award # CD-97723401. Entitled, *Modeling Stream-Wetland Connectivity and Vulnerability to Non-Point Source Pollution*. Thank you EPA for the support and opportunity to do exiting research.

DISCLAIMER

Although the information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency (EPA) under assistance agreement CD97723401 to University of Missouri-Columbia, it has not been subjected to the agency's publications review process and, therefore, may not necessarily reflect the views of the agency; hence no official endorsement should be inferred.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

STRATEGIES FOR DEFINING CHEMICAL CONNECTIVITY BETWEEN STREAMS AND WETLANDS

John Brito

Dr. Enos C. Inniss, Thesis Supervisor

ABSTRACT

Methods for characterizing wetland-to-stream chemical connectivity were developed to help regulators define jurisdiction and wetland managers prioritize their use of resources. Chemical connectivity here is based on historic stream water quality found using publically available databases and analyzed in a Geographic Information System (GIS). Although stream data was readily available, wetland data was not abundant. Interpolation methods were therefore developed using observed water quality data points in streams and projecting them to adjacent wetlands. The different interpolation methods were compared using their root mean square error (RMSE) and local polynomial interpolation (LPI) for chloride data and Universal Kriging for specific conductance data was found to typically yield the best results.

To also address the lack of actual wetland water quality data a relative site suitability analysis approach using GIS was developed to determine monitoring locations for generating additional data. The key environmental variables identified that relate to

connectivity and potential health of the wetlands include hydrology and land use. In additional to conservative water quality parameters, the idea of measuring fluorescence-based excitation and emission matrix (EEM) is introduced. This is a rapid assessment source tracking tool for identifying the natural organic matter (NOM) in different waters. EEM trends through two drinking water treatment plants demonstrate the consistency of the EEM signature and shows that only the more aggressive chemical activities cause notable changes.

CHAPTER 1 : INTRODUCTION TO WETLANDS AND PERTINENT FEDERAL WETLANDS REGULATIONS

1.1 Introduction

Understanding the flux in an open system is crucial in order to elucidate the variables impacting the system. This is especially true in natural water systems in which connectivity to other water bodies can serve as a conduit for pollution or contamination. The definition of connectivity is broad and can refer to a landscape level in which different catchments are connected (Brierley et al. 2006) or to a river system level in which different streams connect (Hooke 2003) . While the definition of connectivity can vary, here connectivity is defined based on its chemical connections, hydrologic connections, and biological connections. These three categories can be used to characterize the connectivity of wetlands to streams as well as wetlands to wetlands. The determination of this connectivity is important in order to classify the wetlands as protected wetlands under federal regulations.

1.1.1 Objective

This research effort places particular emphasis on defining ways to characterize chemical connectivity in waterways using publicly available data. Geographically referenced data on wetlands and different land types are collected and analyzed in a Geographic Information System (GIS). The collected data attempts to depict if wetlands and streams share any similar chemical signatures by tracking specific water quality parameters (Singh et al. 2004) . This involves using multiple spatial tools like

interpolation and raster calculators to develop a methodology. In addition to using GIS techniques, the research develops a preliminary method that correlates fluorescence signatures in waters to different water quality parameters (Baghoth et al. 2011). The fluorescence signatures are captured using a fluorescence spectrometer and then plotted using the Excitation Emission Matrix (EEM) and subject to a Parallel Factor Analysis model (PARAFAC) in MatLab to determine the significant natural organic matter (NOM) in the samples. This thesis shows that the addition of EEM data to already existing data can provide the wetland manager with the additional information needed to better understand the connectivity of the waters of interest. To further assist with decision making and effective use of available resources, a suitability analysis of optimal sampling locations is presented, which provides direction for locating monitoring stations.

1.1.2 Scope

The first objective was to develop a methodology for displaying publicly available data. Completion of the first task involved locating different sources from which to download wetland and stream chemical data. Data from the State of Missouri was used as a case study to show where in records (primarily managed by the State) chemical data could be found and transformed into a format compatible with a GIS. The second task involved determination of the best way to display the data and fill in data gaps using interpolation techniques.

The *second objective* was to create a suitability analysis to allow wetland managers to better evaluate optimal monitoring locations. This objective was broken into three tasks: Task 1: Determine which GIS layers are needed; Task 2: develop a methodology to give weights and values to map layers; Task 3: run analysis and display wetland areas which are suggested for monitoring.

The *third objective* was to make use of EEM with PARAFAC analysis to correlate NOM to different water quality parameters. The three tasks associated with this objective are: 1) Run EEM on different water samples in order to begin to establish a baseline of information, 2) collect samples and analyze for select water quality variables, and 3) correlate EEM and water quality variables to determine the utility of adding this parameter to the suggested date collection.

1.2 Wetland Definitions

Defining and delineating wetlands has been a divisive area in science due to their dynamic nature and variability across different landscapes. Currently, there is no one universal definition of wetlands (Turner et al. 2000). Wetland definitions vary based on regional, climatic, and political boundaries. An accepted concept in wetland definitions is the need for the land to be inundated for a specific period of time. According to federal regulations, it must be inundated 5% or more of the growing season in 5 out of 10 years (Vepraskas et al. 2004). This continued saturation causes the biological structure of the soils to change into a hydric soil. These soils are usually anaerobic and provide a habitat for bacteria and organisms necessary for the function of a wetland

(Vepraskas et al. 2004). Hydric soils can be used as an indicator to the required duration and frequency of surface saturation (Vepraskas et al. 2004). In addition to hydric soils, wetland plant indicator species that require a specific habitat to grow and flourish also indicate the existence of a wetland. Wetland indicator plants have been used recently to assist scientists in identifying wetland areas. The United States Department of Agriculture (USDA) has listed plants that are indicative of wetlands health and function (USDA 2012). Consequently, hydrology, hydric soils, and plant species can be used concurrently to define wetland areas.

While scientific entities struggle to reach a consensus on a wetland definition, government agencies are equally divided. The following agencies work closely with wetlands and wetland regulations, but they each have their own definitions:

Table 1-1: Wetland Definitions (Tiner 1994)

U.S. Fish and Wildlife Service (1979)	"Wetlands are lands transitional between terrestrial and aquatic systems where the water table is usually at or near the surface or the land is covered by shallow water. For the purposes of this classification wetlands must have one or more of the following three attributes: (1) at least periodically, the land supports predominantly hydrophytes; (2) the substrate is predominantly undrained hydric soil; and (3) the substrate is nonsoil and is saturated with water or covered by shallow water at some time during the growing season of each year."
U.S. Environmental Protection Agency / U.S. Army Corps of Engineers (33 CFR 328.3) taken from EPA regs. (40 CFR 230.3)	"Wetlands are those areas that are inundated or saturated by surface or groundwater at a frequency and duration sufficient to support, and that under normal circumstances do support, a prevalence of vegetation typically adapted for life in saturated soil conditions. Wetlands generally include swamps, marshes, bogs, and similar areas."
U.S. Soil Conservation Service (National Food Security Act Manual 1988)	"Wetlands are defined as areas that have a predominance of hydric soils and that are inundated or saturated by surface or ground water at a frequency and duration sufficient to support, and under normal circumstances do support, a prevalence of hydrophytic vegetation typically adapted for life in saturated soil conditions, except lands in Alaska identified as having high potential for agricultural development and a predominance of permafrost soils."

The variability of the government's wetland definitions requires that land owners and developers must understand how their actions are interpreted by regulatory agencies. Regardless of how these agencies define wetlands, they agree that wetlands offer important ecosystem services and are vital to many animal and plant species.

1.3 The Value of a Wetland

Historically in the United States, wetlands have been perceived as waste lands that needed to be drained for farming or other land uses (Woodward and Wui 2001). Wetlands are complex systems, which make it difficult to succinctly define their relationship to the ecosystem (Brody et al. 2008). Wetlands provide a multitude of functions for the overall health of the environment and humans. However, these services are difficult to quantify which seemingly manifests as a disconnection between their services and their value to society (Woodward and Wui 2001). They are key habitats for migratory birds, of which many are endangered species. They serve as habitat for amphibians and act as nurseries for fish and other wildlife during earlier stages of growth. Wetlands are classified as some of the most biodiverse landscapes in the world and are vital to the water cycle in a watershed; they are involved in water purification and water storage. Wetlands along streams and tributaries function as sinks for sediments and nutrients that can otherwise cause a stream to become impaired (Almendinger 1998) . The plant communities in wetlands remove excess nutrients like phosphorus, which if left untreated in the stream can harm aquatic resources downstream. In addition to purifying water, wetlands may also act like a sponge for excess flood water, which contributes to flood abatement. Some scientists attribute increased flood damage in urban areas to the removal of the wetlands that historically functioned as buffers to slow down flood waters reducing destructive erosion. The sponge effect further allows water to infiltrate into the subsurface and ultimately migrate to the groundwater table (Bullock and Acreman 2003). Additionally, wetlands

can also be used for recreation like boating, fishing, and swimming. Many schools use wetlands as educational tools in the classrooms (Sim 2011).

The wetlands off the coast of the Gulf of Mexico provide much of the United States fertile fishing grounds (Chesney et al. 2000). The value of ecosystem services provided by wetlands is hard to succinctly quantify and the impacts of removing these services would be disastrous. One study put the value of the world's wetland ecosystem services at $4.9 trillion/yr. (Costanza et al. 1997). Other studies have estimated a range of wetland per acre values from $0.06 to $22,050 (Woodward and Wui 2001). These figures point to the necessity to continue to protect as well as enhance wetlands, because they are a key component to many different ecosystem functions which are vital to the biosphere.

Wetlands are also being used as a sustainable alternative in treating wastewater effluent. For example, the Columbia, Missouri wastewater treatment plant has four functioning wetland treatment cells that are cited in their permit as the method for disinfection and nutrient reduction (Kadlec et al. 2010). Wetlands can also be used to treat nonpoint source pollution from storm water, which accounts for 50% of the pollution entering most waterways. The nutrient uptake by plants as well as the hydraulic retention time allow for the pollutants to be removed (Novotny and Chesters 1981).

1.4 Background: Regulatory Powers to Protect Wetlands

1.4.1 Clean Water Act (CWA)

The main federal agencies that regulate wetlands are the Environmental Protection Agency (EPA) and the U.S. Army Corps of Engineers (ACE). The main tool used by these agencies to regulate the integrity of US waters is the Clean Water Act (CWA). EPA and ACE have categorized their wetland goals into three main functions: protect, restore, and mitigate wetland resources (Copeland 2010). The requirement to mitigate wetland impacts came with the 1972 amendments to the Federal Water Pollution Control Act, which would later be known as the Clean Water Act in its 1977 amendments (Hough and Robertson 2008). The CWA section that holds the key to protecting wetlands is the regulatory program spelled out in Section 404, which prohibits the dumping of dredged and fill material into the waters of the US without a permit. The program is administered under a partnership between the EPA and the ACE. The ACE is responsible for issuing permits on a day-to-day basis, while the EPA develops the criteria used by the ACE to make permitting decisions. The division of power comes from the fact that the ACE had historical precedence in permitting beginning with the River and Harbors Act of 1899.

1.4.2 Wetland Permitting

Issues arose early because the ACE was issuing the majority of permits that were requested, and the EPA was not rejecting any of the ACE permits. The 1977 amendments gave the ACE authority to issue General permits as a way to manage the overwhelming number of complex permit applications. The 1977 amendments allowed

for additional types of permits. These are Individual permits or Letters of Permission and General permits which are either Nationwide permits or Regional permits. General permits are designed for activities that are determined to be similar in nature and cause minimal environmental harm (Copeland 2010).

Permitting is an important part of the CWA. It ensures that projects which may dredge or fill a wetland are reviewed. The review process allows for a comment period and makes environmentally controversial projects come to light. However, permits are only needed if a wetland is connected to a "water of the US" and can be regulated. The permitting process is in place to protect surrounding waters and ensure that sustainable practices are used during construction.

1.4.3 Section 404: Minimize, Mitigate or Restore

The 1978 regulations issued by the Council on Environmental Quality (CEQ) under the authority of National Environmental Policy Act (NEPA) introduced a way to replace the functions of filled wetlands where permit denials were unlikely. These changes have allowed Section 404 to also become a way to mitigate wetlands. Wetland mitigation in today's context was not the original intended consequence of Section 404. The EPA's definition of mitigation has changed over the years. Currently under mitigation, permit holders are required to: (1) avoid impacting wetlands or streams when practicable, (2) minimize unavoidable impacts, and (3) provide compensation for unavoidable impacts in the form of ecological restoration, enhancement, or preservation of similar, alternate resources. Wetland mitigation sometimes requires the

use of a wetlands bank system. The banks can be run by a range of agencies from regional ACE and EPA offices to local or county-level water management districts. These governing entities are used when wetland destruction cannot be avoided and constructed wetlands must be created. Constructed wetlands (CW) have been in use for a short period of time in the US, which makes data acquisition and the analysis of their effectiveness questionable (Campbell et al. 2002; Haberl et al. 1995). Certain natural aspects of wetlands like hydric soils are difficult to engineer into CWs and take years to develop (Pontius 2008).

Another important area in wetlands regulation is whether to define a wetland as a "water of the US." The determination of whether or not a wetland is a water of the US can be strenuous. In recent years, certain court cases have addressed some of the issues associated with defining wetlands as waters of the US. If a wetland is not adjacent to a water of the US then it cannot be regulated under the CWA. The court cases which address the issue of wetland regulation will be further discussed in Chapter 2.

1.5 Statement of Problem

The association between policy, resource management, and engineering sets the stage for these current research problems. The project goal is to develop a methodology to determine stream and wetland connectivity by examining chemical, biologically, and hydrological connectivity parameters. A logical solution creating this methodology is to look at the problem within a spatial context. This involves the use of engineering tools such as a geographic information system (GIS), which can be used to analyze data within

that spatial context. GIS allows users to input data onto discrete points on a map and then use analyst tools to produce specific information. In this project, the analyst tools allow a closer look at the connectivity paradigm between streams and wetlands. Thus, GIS was used alongside the available data to develop a methodology which can be used to determine stream and wetland connectivity and allow wetland managers to protect wetlands under current policy. In order to create a more robust solution to this problem, the need for current wetland water quality data has always been needed. This prompted the researchers to take a closer look at how to develop a methodology to allow wetland managers to select the most optimal wetland monitoring sites.

Problems within resource management can be alleviated with the use of mapping tools like GIS. Currently the state of Missouri does not have an active wetland management plan which leads to a lack of available data. The lack of data may lead to ill-informed decisions being made pertaining to wetland understanding. This study's research team identified this gap early in the research formulation process. The proposed remediation to the problem includes developing a GIS methodology to help wetland managers select the best wetland monitoring locations. The tool uses multiple wetland layers to create a raster overlay and identify key wetlands that may hold a better potential for robust wetland water quality monitoring. The goal is to include this tool in the wetland management plan, which will allow wetland managers to make and defend informed decisions with this support tool. Alongside tools like GIS, the project develops a methodology to look at the use of fluorescent spectroscopy to determine chemical connectivity in water.

Fluorescent spectroscopy analyzes the fluorescence of a sample using UV light in order to characterize the natural organic matter (NOM) in water. The goal achieved by this research was using fluorescent spectroscopy to analyze the excitation and emission (EEM) spectrum of water samples thereby determining the different NOM. The EEM was analyzed and correlated to over 24 water quality parameters such as total organic carbon (TOC), disinfectant by-products (DBPs), ammonia, total nitrogen (TN), pH, UV254, and total bromine. Correlations were discovered showing how EEM can serve as a precursor to some of the above parameters and can further serve to determine chemical connectivity in waterways.

1.6 Thesis Organization

After the introduction and a brief overview of the project, Chapter 2 presents a detailed literature review discussing the intersection between resource management and the implementation of scientific solutions like GIS. It also examines court cases impacting wetland regulation, past research on source tracking and the use of EEM as an applied method for identifying linkages between waters.

Chapter 3, entitled “Displaying Water Quality Data Using GIS and Different Interpolation Methods,” begins with the description of the Lower Grand Watershed, which is the site location suggested by The Missouri Department of Natural Resources (MODNR) and EPA Region 7 for a case study application of the tools developed in this research. Next, this chapter discusses how to locate chemical data from publicly available sources such as government databases, academic sources, and non-profit /

volunteer monitoring programs. Then the chapter presents methods on how ArcGIS is applied to the research problem in order to display and analyze the data on landscape maps.

Chapter 4, entitled “Suitability Analysis for Wetland Monitoring,” continues to look at issues in wetland management and the use of GIS as a practical solution. The chapter was inspired by the lack of current wetland water quality data and a need for dynamic surface water quality site selection tools. The beginning of the chapter discusses current surface water monitoring initiatives and trends, followed by a discussion of the site suitability analysis and the use of GIS and raster data manipulations. Finally, a description of the publicly available maps used and the corresponding results are presented.

Chapter 5, entitled “Potential Application of EEM to Wetland Connectivity,” examines EEM with parallel factor analysis (PARAFAC) and its viable use in fingerprinting water sources. The chapter begins with a description of the experiment, which used drinking water facilities in Missouri as closed systems to observe the methods of using PARAFAC. The experiment was followed by correlation of different water quality parameters with samples’ EEM signals. Finally, the researchers determined if application on a larger scale relating to stream-to-wetland connectivity issues. Collectively the three projects were used to begin developing a stream and wetland connectivity paradigm.

Chapter 6 presents the conclusion as well as future recommendations for further research.

Chapter 2 LITERATURE REVIEW

2.1 GIS Water Quality Wetland Section

Wetlands have recently been in the spotlight of US regulatory agencies. They are now considered an asset to ecosystems, but they were once considered as land that could be reclaimed for agricultural usage (Davis and Froend 1999). This change in perspective has been a catalyst for new research in wetland science. Researchers are trying to understand how beneficial wetland ecosystems are and how urbanization impacts them (Owen 1999). Wetlands can help improve overall watershed health and biodiversity so protecting them is crucial. However wetlands are not directly regulated by the EPA unless a biological, chemical, or hydrologic connection can be determined between the wetland and a regulated water of the US. The question of what defines proper connection between a wetland and navigable waters of the US is based on the "Rapanos vs. United States" court cases discussed in the following section alongside additional court cases.

2.2 Court Cases

Wetlands can only be regulated when they are adjacent to "waters of the US." Some examples of waters of the US are: waters used for commerce, waters used for commercial fishing, interstate and intrastate waters such as lakes or rivers, and wetlands adjacent to the specified waters. The uncertainty is in how to assess if a wetland is adjacent to waters of the US. In the past it was based solely on clear

hydrologic connections but the following court cases have tried to clarify the legal understanding of wetland adjacency.

2.2.1 1985: United States vs. Riverside Bayview Homes

The case began when developers for Riverside Bayview Homes (RBH) decided to place their fill material near Lake St. Clair, MI. The ACE stated that the developers needed a permit to dump the fill material. RBH believed that the permit was not necessary because the lands were not part of the "waters of the US" (Guttery et al. 2004). The ACE filed a lawsuit against RBH and won in the District Court but the decision was later reversed by the Court of Appeals. Finally, the Supreme Court held that wetlands adjacent to traditional navigable waters were properly considered to be "waters of the US." (EPA).The expansive definitions of "waters of the US" allowed the ACE to protect the wetlands in this court case and set future precedence.

This court case was very significant in the realm of wetland protection. It was the first case that involved the ACE's authority to regulate and issue permits for wetlands. It proved that the ACE could add its interpretations to the permitting of wetlands. Second, it settled the case on whether or not the waters of the US could be further defined by more than just navigable waters. It, therefore, set a precedent that property owners and developers could not ignore the CWA in order to increase profits.

2.2.2 2001: SWANCC vs. Army Corps of Engineers

"In *Solid Waste Agency of Northern Cook County (SWANCC) v. U.S. Army Corps of Engineers*, the Court addressed the question of CWA jurisdiction over isolated, non-

navigable, intrastate ponds" (EPA). SWANCC is a consortium of 23 Chicago-based cities and towns that wanted to develop a landfill on land that was previously used as a gravel mine. The ACE initially stated they did not have any jurisdiction over the land. Shortly thereafter, the ACE was contacted by the Illinois Nature Preserve Commission and informed that migratory birds used the waters and insisted that the ACE should not issue a permit to SWANCC (Christie and Hausmann 2003). The ACE reversed itself and stated that the waters were indeed waters of the US because the land had not been used as gravel mines for many years, and the area had developed natural characteristics that were used by migratory birds. In 1986, the ACE, in an attempt to further clarify its jurisdiction, insisted that it had jurisdiction to permit for intrastate waters. The agency stated that any waters used by migratory birds could be regulated. This determination would later be known as the "Migratory Bird Rule." (Christie and Hausmann 2003) . SWANCC decided to go to the Court of Appeals and argue against the ACE. The District Court and the US Court of Appeals ruled in favor of the ACE. In 2001, the case ended by a Supreme Court ruling in which it voted 5-4 in favor of SWANCC.

Some environmentalists believe that the court misinterpreted Congressional intent in the matter (Copeland 2010). However, the Supreme Court did not overrule the court case of *RBH vs. US*. The Supreme Court did imply that wetlands had to be connected to navigable waters with a "significant nexus" in order to be regulated (Downing et al. 2003). The case made the migratory bird rule invalid and caused regulators to question what was defined as an "adjacent" wetland (Alexander et al. 2007; Christie and Hausmann 2003). Thus, in 2001 a joint legal memorandum issued by

the EPA and the ACE provided their legal interpretation of the ruling. This guidance document was issued to help federal agencies better understand the SWANCC ruling. EPA also released an advance notice of a proposed rule [EPA–HQ–OW–2011–0409; FRL–9300–6] (ANPR) on how to define waters under jurisdiction, receiving a flood of negative comments. The public overwhelmingly opposed any rule changes because they believed that the current definition helped policy makers protect wetlands. If one defines them too specifically, then one leaves out the possibility of expanding the definition of waters of the US. In late 2003, the EPA and Corps released a statement saying that they would not attempt a rule change because it would cause too much debate and controversy.

2.2.3 2006: Rapanos vs. United States

The Supreme Court ignited doubt in the power of the CWA to protect wetlands with the ruling SWANCC vs. ACE. The next significant court case on wetlands regulation was *Rapanos vs. US* with an associated case of *Carabell vs. US Army Corps of Engineers* attached to the decision. Rapanos was a private land owner who faced civil charges because of his actions in filling and draining wetlands. Rapanos filled and drained 22 acres of wetlands after the Michigan Department of Natural Resources had warned him not to (Macdonald 2007). He believed he did not need a wetlands permit and continued with the destruction of his wetlands. The project area was adjacent to channels which were either ditches or intermittent streams that drained to navigable waters. Rapanos tried to argue that there was no hydrologic connectivity between his wetland and the

channel. However, it was proved that the wetland did empty into some tributaries that later flowed to regulated waters.

Lower courts had ruled in favor of the ACE. However in June of 2006, the Supreme Court ruled in a 5-4 decision that the lower courts "had applied an incorrect standard in determining whether the wetlands were in fact covered by the CWA" (Copeland 2010). Justice Scalia opened his analysis with the ACE's gaffes in the regulatory program. He pointed out the burdensome cost, unpredictable regulatory regime, and its limitless scope (Macdonald 2007). Scalia rejected the ACE's interpretation of "waters of the US" stating that Congress utilized the word waters instead of water to indicate that water had to be a relatively permanent body of water—not a tributary or intermittent stream. He continued to assert that the ACE is only permitted to regulate wetlands that have permanent surface water. Scalia concluded by saying that he believed the ACE had a limitless view of power regarding permits and this would lead to more court cases. Justice Kennedy concurred with the ruling, but he had a different reasoning for doing so. Kennedy felt that the pending limitations set for application to the ACE were not going to alleviate the problem of regulating wetlands. Their power was not limitless, and they did in fact have a standard in selecting wetlands to be regulated (Macdonald 2007). However some direction was needed to determine regulated wetlands and Justice Kennedy recommended using the criterion that a wetland must have a "significant nexus" to waters of the US in order to determine connections either hydrological, chemical, or biological (Blumm and Bosse 2007). The nexus could be a hydrologic connection, by having clear connections above

or below the surface of the waters. Biological connectedness could be established by demonstrating that animal life located in two waterways share the same family genes. Chemical connectedness could be demonstrated by two waterways sharing the same chemical parameters (Leibowitz et al. 2008). Kennedy said that the ACE had good reason to protect the wetlands adjacent to waterways because they enhanced the surrounding waters with their ecosystem services. The majority opinion of the Supreme Court did not provide any clear guidance and thus Kennedy's statements have become the standard.

The *Rapanos* case, unlike previous cases, forced regulators to begin to try to answer the question of what makes wetlands a "water of the US". A major advancement came from Justice Kennedy's statements. Most people agree that Justice Kennedy's opinion was more significant than Justice Scalia's. He mentioned the "significant nexus" that used three criteria to determine connection, but they were not covered in depth. This left the ACE free to determine how it would proceed in defining the nexus. The scientific community plays a pivotal role in helping the EPA and the ACE properly define wetland connections, as we continued to clarify the definition of a significant nexus indicated in the Rapanos opinion. This case sets the backdrop for the current research project for determining measures for stream and wetland connectivity, with emphasis on chemical measures.

2.3 Connection Types

Biologic connections

Wetland ecosystems are hosts to a diverse group of plants, animals, and microorganisms. The loss of biodiversity in wetlands can be a clear indication of reduced health and potential decline of a wetland ecosystem. Researchers have developed the Index of Biological Integrity (IBI) as a tool to evaluate the health of a wetland ecosystem (Miller et al. 2006) based on anthropogenic influences and biological activity. King and Richardson (2003) used the presence of plants and animals as an indication of the health of a wetland system (King and Richardson 2003). Amphibians are thought to be one of the best species to use as indicators because they require wetland habitat and terrestrial landscapes to survive. Alam (2012) used salamanders in a recent study to determine biological connectivity of wetland ecosystems. This connection was determined by creating corridors between different wetlands which would be ideal habitats for salamanders (Alam 2012).

Hydrologic connections

Hydrologic connections can be above or below ground. On the surface, if clear drainage ditches, perennial, intermittent or ephemeral streams exist between two waterways then a connection can be inferred. However groundwater connections between a losing stream and a local aquifer may not be as easily deduced. It is understood that water bodies in a delineated watershed have a high probability of interacting on an ecosystem level. This interaction can occur through the movement of

storm water runoff or seasonal flooding with absence of barriers (Toner and Keddy 1997).

It is understood that water moves downhill following the contours of the terrain. Thus the use of digital elevation models (DEMs) can be a logical place to begin analyzing hydrological connections above or below ground. This gives modelers a way to look at connection by determining stream flow path and potential runoff ditches. Also because water is the universal solvent many different organic and non-organic chemicals follow its movement. Chemical data can be used as tracers to determine hydrologic connectivity (Cox et al. 2007).

Chemical connections

Developing a methodology to determine chemical connectivity is made easier with prior knowledge of the landscape and possible chemical influences. If a known pollutant has been leaching from a discrete location and entering waterways, it can be a possible tracer (Heberer 2002). However, in situations without a clear pollutant, the researcher must use common water quality parameters as tracers. This requires collecting data for a range of water quality parameters and then trying to locate trends in the data (Zhao et al. 2011). Certain water quality parameters can potentially reveal a problem in the ecosystem when the concentration of the parameter spikes. For example eutrophication of ponds or lakes is due to an increased amount of phosphorus or nitrogen (Gupta et al. 2012). However, while there are vast amounts of data collected indicating the presence of nitrogen and phosphorus, it may be difficult to trace their

travel paths in surface waters. Nitrogen has many natural and some anthropogenic non-point source emitters into the environment like fertilizer application and confined feeding operations (CAFOs) waste. Nitrogen flow in watersheds is uncertain due to the many processes in the nitrogen cycle such as fixation, mineralization, denitrification, and uptake by plants. Thus other water quality parameters which are more sensitive to change must be used as potential chemical tracers. These may include acidity, alkalinity, chloride, specific conductance, and hardness. The following discussion will explore each individually.

2.3.1 Potential Water quality parameters

Acidity

Acidity (low pH level) is a problem in surface waters and can lead to degradation of the natural habitat. EPA has measured acidity in surface waters in order to gauge the impact of pollutants like acid rain or mine leachate contamination, which can cause acidification in surface waters. Due to low buffering capacity of surface waters and high pollution the northeast is a problem area for acidification. While acidity can be traced in waterways and can be localized to point source pollution, it may be difficult when dealing with atmospheric deposition of acid. The fact that acid rain can cross boundaries and change acidity levels in multiple regions could make it difficult to use acidity as a tracer in certain areas (Driscoll et al. 2001; Kirby and Cravotta Iii 2005; Kirby and Cravotta Iii 2005).

Alkalinity

Alkalinity is a common parameter measured by government agencies during water quality monitoring plans. It represents the buffering capacity of a water body. Alkalinity is important in order to allow the water body to resist drastic change in pH when an acid is introduced into the water either through acid rain or inputs from wastewater. Water becomes more alkaline based on the natural geology of the area and the weathering of rocks, soils, and different salts. Alkalinity is typically reported as mg/l of calcium carbonate ($CaCO_3$) which is usually introduced through limestone deposits. Unnatural impacts can also change alkalinity like wastewater or industrial waste with detergents or soaps. A rapid change in alkalinity in waterways may be a sign of potential problems in non-point pollution and can be used to trace connectivity (Kirby and Cravotta Iii 2005; Kirby and Cravotta Iii 2005).

Chloride

Chloride (Cl^-) is commonly used as a conservative tracer because it is considered hydrologically and chemically inert. It is highly mobile in aqueous solutions and is not readily adsorbed onto surfaces (Peters and Ratcliffe 1998). However recent research has shown that chloride does interact in the soils' complex biogeochemical cycle which means it can be a sink or source for Cl^- (Svensson et al. 2007). Cl^- is still a common tracer used based on available data and known methodologies. Cl^- is linked to natural salt deposits which can be released through weathering of specific rocks like halite. However, chloride increases in the environment can usually be attributed to human

impacts such as impervious surfaces attributed to urbanization (Kelly et al. 2007). The application of road salt increases Cl^- concentrations near roads and surrounding surface waters (Svensson et al. 2007). Cl^- concentrations can be elevated in waters that have been stagnant for some time due to evaporation and longer residence times (Peters and Ratcliffe 1998).

Specific conductance

Specific conductance (SC) is the measurement of the ability of water to conduct an electrical current. It can be used indirectly as a measure of the presence of total dissolved solids (TDS), like salts, in water. For example ocean water has a high SC while distilled water is very low. High SC can be an indicator of pollution and is a problem in drinking water facilities due to the difficulty of removing TDS. SC can be caused naturally through the weathering of rocks like calcite and calcareous shales. However, if an area is devoid of these minerals, then SC will usually be linked to point and non-point pollution (Dow and Zampella 2000).

Hardness

Water hardness is typically determined by the concentration of multivalent cations such as Ca^{2+} and Mg^{2+}. Hardness is naturally occurring and is affected by the geology of the region. Flowing water may pick up dissolved minerals from the soil and surrounding rocks. Hard water is common in the United State and does not pose a negative health effect. However, it is commonly measured in drinking water sources and

can be detrimental to piping systems and industrial factories. Due to it commonly being measured, it was examined as a potential water quality tracer (Hudak 2001).

"Data rich information poor"

The government's water quality monitoring programs are usually constrained by available resources. The amount of monitoring sites and the frequency of monitoring is not sufficient enough to make the most statistically valid decisions according to EPA (EPA 2008). This means that existing water quality parameter databases may have some data gaps spatially or temporally. Some researchers believe there is a disconnection between information users and information producers of water quality data. Literature often refers to this disconnection as the data-rich-but-information-poor syndrome (Timmerman et al. 2010). The reason for this disconnection is the lack of communication between water quality data users and collectors. In the realm of wetland management this is extremely evident as there is little to no wetland water quality data available. When data is unavailable different tools and methods have to be used to deduce missing links.

2.4 Introduction to Geographic Information System (GIS) tools

Stream-wetland connectivity is a spatial problem and GIS is a tool commonly used to solve spatial problems. Engineers use GIS for the management, visualization and analysis of monitored data (Kumar et al. 2007) . GIS has seen many uses as a tool to deal with problems that have a spatial context. It is also a great tool to perform a gap analysis of current data. The versatility of GIS has allowed it to be used in wetland

science as a tool to help researchers attempt to fill in the wetland data gaps on a global and local scale (Rebelo et al. 2009). GIS spatial analyst tools provide researchers an array of different methods such as interpolation to help fill in these data gaps.

2.4.1 Interpolation methods

Interpolation is a mathematical function that estimates values at locations where no measured data is currently available. Interpolation uses known discrete data points to create new data points at these data barren locations. Interpolation has been especially useful for creating elevation maps because cartographers did not have an elevation for every discrete point. If one was trying to create an elevation map of a state, a data point at every city would not be sufficient enough to create an accurate map. Success and accuracy of the interpolation depends on the type of interpolation method used and the density of discrete points already obtained (Luo et al. 2008). Each interpolation method comes with its own internalized assumptions that control how the function will perform. Some common interpolation paradigms are global vs. local, deterministic vs. stochastic, exact vs. approximate, point vs. areal, and gradual vs. abrupt (Akkala et al. 2010).

Kriging

Kriging is a geostatisical interpolation method available on most GIS software. The data used in kriging is assumed to be stochastic in nature, which means it should be random variables that are ordered in a spatial scale such as elevation measurements. Kriging is a statistical model that includes probability in its predications which allows you

to assess the error of the predications. This allows researchers to justify the predictions and methods used. Kriging relies on autocorrelation to create the predicated values based on measured values. Autocorrelation in kriging is a function of distance and it assumes things closer together tend to be more similar than things far apart. The spatial correlations can be quantified by assessing the semi-variance (ESRI 2011). The semi-variance can be a function of both distance and direction so it can account for direction-dependent variability in anisotropic spatial problems.

The kriging types available in ArcGIS are ordinary, simple, universal, indicator, probability, and disjunctive. However the three types most commonly utilized are ordinary, simple, and universal kriging. Simple kriging is mathematically the simplest and depends on the mean of predicated values being known beforehand even though they are usually unknown. Ordinary kriging is the most popular type of kriging and assumes a constant but unknown mean. Universal kriging is similar to ordinary kriging but assumes the data are deterministic and tries to fit a polynomial trend (Luo et al. 2008). Kriging has been used frequently in ecological studies, and most of these studies have concluded that the most accurate results are obtained by using kriging when compared to other interpolation types like inverse distance weighting (IDW), global polynomial interpolation (GPI), and local polynomial interpolation (LPI).

Inverse Distance Weighted (IDW)

IDW is an interpolation method that assumes that things close to one another are more likely to be more similar than things far away. IDW creates predicated values

by using the measured values surrounding the prediction location. IDW assumes that the measured value closer to the prediction location carries a heavier influence than those farther away. IDW gives the greatest weight to point's closet to the prediction location, and the weights diminish as a function of distance.

The value of each weight is proportional to the inverse distance between the known data point and the prediction point. The inverse of the distance is raised to the power *p*, and *p* is a value selected by the user. If *p* = 0 then there will be no decrease of influence with distance, and the prediction point will be the mean of the measured values in the search neighborhood. The optimal *p* value can be determined by previewing the output and examining the cross-validation statistics of each output. Users should look at the root mean square predication error (RMSPE) value which quantifies the error of the prediction surface (ESRI 2011).

The search neighborhood of the measured points can be manipulated by the users. Data points farther away from prediction points hold little influence in their calculation. Thus, users can increase efficiency of the model by limiting the number of measured values used in the calculation to a minimum or maximum. Users can also select the shape of the neighborhood, and directional influences like wind can be included.

IDW is an exact interpolator, which means the maximum and minimum values in the interpolated surface can only occur at measured sample points. IDW is very sensitive to clustering of values and outliers can also disrupt the model. IDW assumes that the

model is driven by local variations. It is difficult to justify the output in IDW because it does not provide prediction standard errors in the cross validation (ESRI 2011).

Global and Local Polynomial Interpolation (GPI and LPI)

Global polynomial interpolation (GPI) is an inexact interpolator that attempts to fit a smooth surface which is defined by a polynomial function. GPI is commonly used in large coarse scale spatial problems. The changes in the measured data points should follow gradual trends on a large scale. GPI can be used for pollution problems which show slow varying measurements; however, outliers tend to disrupt GPI outputs. GPI uses polynomials to fit the data. For example, a first order polynomial fit would demonstrate a linear trend on a large scale (ESRI 2011).

Local polynomial interpolation (LPI) is similar to GPI but it fits many polynomials within specified search neighborhoods which are defined by the user. This allows LPI to pick up local trends that GPI analysis would simply smooth over in fitting the data. LPI relies on the assumptions that the samples were taken on an equally spaced grid, and the data values within the search neighbors are normally distributed. LPI gives users standard errors to evaluate the uncertainty associated with values predicated at each location (ESRI 2011).

Applied Interpolation methods

Akkala et al. (2010) analyzed a list of 11 different interpolation methods including kriging, IDW, GPI, and LPI. They did an extensive analysis on the different

methods and compared pros and cons of each. They also mentioned which software packages offered the diffident interpolation methods. The Akkala paper presents a list of research papers and respective topics which employed the different interpolation methods. Akkala et al. (2010) conducted a case study with radon data collected by the University of Toledo throughout Ohio. The university collected samples in 1,262 out of the 1,492 ZIP code locations in Ohio (Akkala et al. 2010). For the remaining 230 ZIP codes, they employed six interpolations techniques which included: Kriging, IDW, GPI, LPI, radial basis functions (RBF), and artificial neural networks (ANN) to determine which one is most effective for this environmental data. ANN received the lowest average error of 4.64%, but for our purpose kriging received the lowest of the four previous methods with 7.18% error. IDW received 7.48%, LPI received 8.76% and GPI received 9.50%.

Lou et al. (2008) also compared different interpolation methods in an experiment to develop surface maps for continuous wind speeds using irregularly distributed data from the UK. Wind speed data was collected from 1998 until 2002 at approximately 560 locations across the UK. The researchers analyzed the data with IDW, LPI, thin plate spline (TPS), trend surface analysis (TSA), ordinary kriging, universal kriging, and ordinary cokriging. Cokriging is considered a multivariate version of kriging that uses additional covariates, which would be ideally sampled at the same location as the estimated variables to assist in prediction. Cross-validation tools from the output in ArcGIS were used to compare each method. ArcGIS computes the mean error (ME) and

root mean square error (RMSE) in the output in order to measure the accuracy of the interpolation. ME is used to detect bias, and it should be zero if predications are centered on the measurement values. The RMSE is used to determine how closely predicated values match measured values; the smaller the RMSE--the better the interpolation (Luo et al. 2008). In the end the cokriging methods had the lowest RMSE value of 1.47, followed by ordinary kriging with 1.61, LPI with 1.69, universal kriging with 1.71, and IDW with 1.74.

Additional Methods

Kriging proves to be an effective interpolation method for certain environmental data but it does have its limitations. Application of kriging to hydrological interpolation may prove to be difficult because chemicals in a stream tend to follow drainage paths determined by topography. This understanding gave rise to the idea of using sub watersheds as potential areal interpolation surfaces in this thesis. A smaller watershed within a larger watershed is called a sub watershed. These sub watersheds are created based on the topographic contours in a specific area. A sub watersheds size depends on the drainage point used to delineate the watershed. ArcGIS comes equipped with tools to delineate sub watersheds based on DEMs and a selected drainage point. The idea of using sub watersheds as potential drainage points comes from research conversations with Dr. Timothy Matisziw and his prior work with interpolation. When Dr. Matisziw conducted research on crop pricing based on county boundaries, some counties did not have data on crop prices, so Dr. Matisziw used surrounding county data to interpolate

these prices (Meeting 2012) . In this thesis, sub watersheds are considered the boundaries, and the data from the discrete water quality points are assigned to each respective sub watershed. Other researchers have used area-point downscaling, which is a reciprocal to the method discussed above (Yoo and Kyriakidis 2008).

EPA BASINS and HSPF

The program known as Better Assessment Science Integration point and Nonpoint Sources (BASINS) is an open source free GIS system developed by EPA and its affiliates. BASINS is a tool developed for use by regional, state, and local agencies for water quality and watershed analysis. It integrates GIS data with water quality point source data from current government sources. The current version, BASINS 4.0, allows users to download water quality data directly from USGS and EPA databases. EPA currently suggests the use of BASINS to help solve Total Maximum Daily Load (TMDL) issues on the state and local level. It can be used to delineate watersheds and use nonpoint source models like Hydrological Simulation Program-Fortran (HSPF) to perform model analysis.

BASINS, when coupled with HSPF, gives users the ability to input time series water quality and precipitation data to model changes in the watershed (Filoso et al. 2004). Metrological data was downloaded based on available rain gauges in the watershed and inputted into the program through plugins like WDM utility. While BASINS is open source it requires a certain expertise to properly model and understand the output data. Users have also complained about the lack of available water quality

data for the BASINS program. This means it has not been as widely used as EPA had previously hoped (Whittemore and Beebe 2000).

2.5 Suitability analysis

A suitability analysis is a tool that can assist land managers in their future decisions (ESRI 2006). It allows the managers to choose the most optimal location for specific land development based on criteria of his/hers choosing. The goal of a suitability analysis is to create a land map which demonstrates the best location based on the overlaying of specific criteria onto a single map. The criteria are usually in the format of raster map data but can also be in the vector form. For example, a retail store must be accessible to customers; hence, the land manager will want to include road locations and location of known population. Suitability analysis has been around since the times of hand drawn maps, but they continue to show great potential with the ongoing development of computerized GIS software (Collins et al. 2001). When developing a suitability analysis the common steps in (Figure 2-1) should be followed.

2.5.1 Suitability Decision Making Methods

There are different types of decision making methods used within a suitability analysis. Some suitability projects use the multi-criteria evaluation (MCE) method. MCE utilizes multiple factors or criteria into the final overlay process. This is used for a problem that has a multitude of factors influencing the final solution. A user could also create a suitability analysis with only one factor if it is the defining criteria. Within the suitability analysis framework, users must select their overlay process; Boolean overlay,

or Fuzzy logic (Collins et al. 2001). Boolean overlay creates a yes or no definition of whether or not an area is suitable. It uses the GIS operators such as intersection and union to determine if the criteria are met on the location. The location either receives a score of 1 which means it is acceptable or 0 which means it has not reached the standard. Fuzzy logic alternatively allows for the user to create a range of acceptable locations. They are scored from 1 to 0 with the possibility of scores in-between and thus being partially acceptable as an outcome. This allows the user flexibility in selecting a range of land areas that can be possible solutions.

Selecting Weighted Values

Within the fuzzy logic paradigm comes the notion of ordered weighted averages (OWA). Unlike Boolean, fuzzy logic allows for a range of possible values in the final outcome. OWA allows for the creation of two sets of weights for each map/criteria. The first set of weights controls the relative contribution of each specific raster cell type in the map, while the second weight controls the aggregation of the total map in the final analysis. Selection of the weighted values directly impacts the outcome of each suitability analysis. Selecting the weighted values can be difficult, but one solution is to survey experts on the topic. However, the quality of the weighted values is directly related to the quality of knowledge each respondent has on that specific topic and the use of GIS technology (Palmeri and Trepel 2002). Another option in selecting the weights is the pairwise method. Pairwise compares each criterion in a pair to determine which one holds more importance. After all the criteria are compared the value from

the pairwise matrix is used to develop the weight. Pairwise is common in the analytic hierarchy process (AHP) method; (White and Fennessy 2005) used a pairwise comparison in their analysis to determine weights for each map criteria.

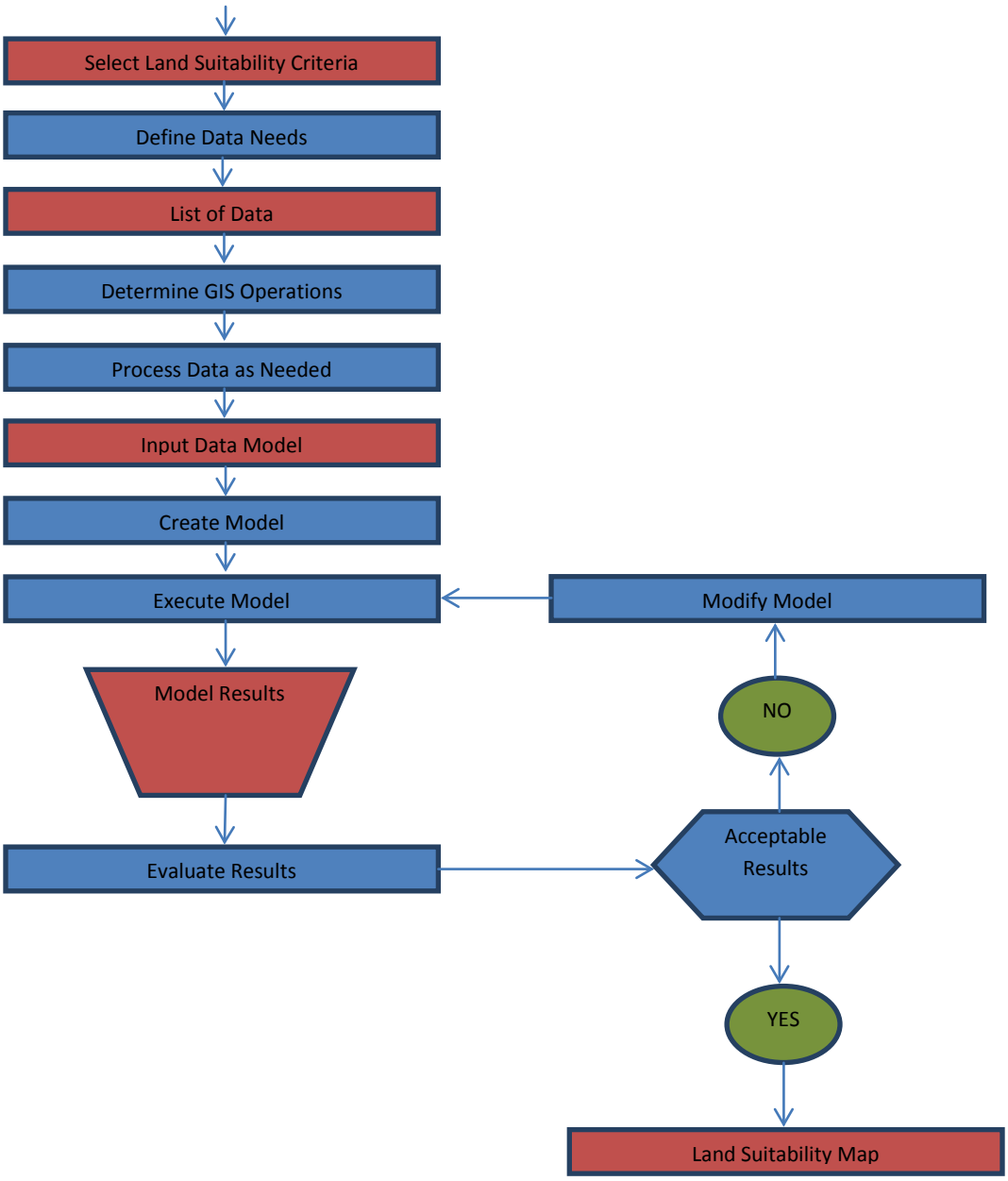

Figure 2-1- Suitability analysis flow chart

2.5.2 Suitability Analysis: Wetland Management

Suitability analysis and its use with wetland management have been explored in recent studies (Palmeri and Trepel 2002; Van Lonkhuyzen et al. 2004; White and Fennessy 2005) . Wetlands are sensitive ecosystems, and when they are kept healthy and functioning, the surrounding biosphere benefits. Thus, federal and state governments protect wetlands under laws like Section 404 of the Clean Water Act (CWA) which requires contractors who destroy wetlands to reconstruct the wetlands in new locations, most commonly in wetland mitigation banks. Site location of wetlands requires an understanding of the links between wetlands and their surrounding landscapes in order to identify the hydrogeological settings necessary to support healthy wetland development (White and Fennessy 2005). Thus the use of a suitability analysis can help wetland managers identify these key landscape features which contribute to a successful constructed wetland location.

White and Fennessy et al. (2005) located new wetland mitigation sites using a site suitability analysis. They decided to locate the wetland sites by analyzing the watershed as a whole site instead of specific areas within the watershed. This scale was used so as to enhance the entire watershed's ability to meet US CWA goals of supporting aquatic life. The researchers used MCE within the GIS framework to identify the areas with the highest likelihood to support a healthy wetland system. They used a pairwise comparison in the study to develop the weights for each criterion in the MCE. A map of site characteristics was created for each of the following criteria: stream order,

overland flow length, saturation index (without permeability), saturation index (with permeability), land use type, and use attainment. The project focused on only using readily available data in their analysis. DEM and land use layers were taken from available sources and used in the spatial analyst process to create all the criteria maps listed above. They created three versions of the suitability analysis based on how the factors were calculated and different weighting methods. This was done because they understood that different weights or priorities assigned to each model have a significant effect on the final results (White and Fennessy 2005). Overall this project successfully utilized a suitability analysis with readily available GIS data.

Van Lonkhuyzen et al. (2004) created a suitability analysis to locate wetland mitigation sites on a federally owned location near the University of Chicago. This study unlike the previous project by White and Fennessy et al. (2005) used a site specific scale in a more urban location to indicate sites that can be restored as wetlands to improve ecological function in the federal site. The variables used to create the site specific maps in this analysis were: hydrology, soils, historic condition, vegetation cover, adjacent vegetation, and land use. This project placed a strong emphasis on the hydrology of the location for the restored wetland. This study clearly states: “Hydrology has perhaps the most important influence on project success because the magnitude and duration of inundation or saturation affects the success of hydrophytic vegetation establishment as well as hydric soil development”(Van Lonkhuyzen et al. 2004). Hydric soils are one of the variables researchers used in the analysis because the occurrence of hydric soils can

potentially indicate the presence of past wetlands as well as a hydrologic regime that can support healthy establishment of a wetland ecosystem in the future (Van Lonkhuyzen et al. 2004). The quality of the water the wetland site received was also important to researchers, and they stated that a site which receives water from surrounding developed or disturbed areas would have a higher probability of invasive species such as cattails (Typha spp.); thus, adjacent land use and vegetation was also important. Van Lonkhuyzen et al. (2004) used weights applied to each variable to determine its importance; the higher the weight the stronger the influence. Weights as are follows: hydrology, soil, and historic condition received a 3; adjacent vegetation received a 2, and vegetation cover and land use received a 1. For land use, soil, historic condition and vegetation cover, the vectors where converted into 3m x 3m rasters for proper analysis. They also created 15m buffers of water bodies and 6 m buffers of local depressions so as to embrace wetland locations that surrounded these features. (Van Lonkhuyzen et al. 2004) desired for this model to be viewed as a screening tool where further evaluation should be done before a final decision is made by wetland managers.

Both studies successfully used a suitability analysis to help planners and wetland managers make objective decisions regarding wetland management. These tools are extremely important in resource management because they give planners data to illustrate to policy makers the logic behind their choices. These same analysis methods can be used to locate wetlands, which are ideal for water monitoring regimes. Using the MCE method alongside selected weighted maps can give managers a range of wetlands

to select from when they are developing monitoring regimes to increase understanding of wetlands quality in state wetland management plans. White et al. (2005) used hydrology in their analysis by including stream order and overland flow length. Instead of overland flow length, however, this thesis used flow accumulation in a suitability analysis to determine the potential for a site to receive water. Both studies used land use as a variable which can impact the potential for a wetland to be a good candidate for restoration. Land use can impact the quality of water in the wetland and in our monitoring analysis, it was selected as criterion of concern. Van Lonkhuyzen et al. (2004) used hydric soils and historic conditions. Stein et al. (2010) determined the potential for a site to promote healthy wetland ecosystems. In Van Lonkhuyzen's 2004 study, hydrology analysis flood plains were included as a variable that received a high score. Instead of using flood plains, the flooding frequency of a site was used to determine potential for overland flow in a more discrete fashion. However both analyses by Van Lonkhuyzen (2004) and White and Fennessy (2005) did not consider proximity of wetlands to other wetlands which can demonstrate the larger wetland ecosystem framework (Alam 2012). To include this connectivity between wetlands in this thesis, a hotspot analysis (Zhang et al. 2008) was performed based on the size of the wetlands. This analysis indicated clusters of large wetlands which can act as a sink for water quality data. Also a proximity buffer was used for each wetland to determine the amount of neighboring wetlands and thus give a quantitative figure to a wetlands potential connectivity in a wetland ecosystem.

Designing water quality monitoring networks has always been difficult and researchers have tried to make the task easier with the use of tools such as GIS. Water resource managers have to select sites, develop sampling frequencies, select variables to be sampled, and select the period or duration of sampling. The initial and most crucial part of the network design is selecting the sampling locations (Harmancioglu and Alpaslan 1994). GIS coupled with a site suitability analysis is a great way to justify the selection of their monitoring locations and can be used as a tool in water resource management.

2.6 Source Tracking

The closure of beaches due to pollution and subsequent drinking water advisories are a common problem in water resource management. When a body of water is considered out of compliance by EPA, local authorities must determine the source of this contamination. The contaminants usually follow natural streams and drainage paths and pick up indicators from the surrounding environment that scientists can use to determine the contaminants' origin. Some processes (e.g., tracking source of contamination with organic dyes) requires officials to constantly monitor downstream areas for the appearance of the dye. However new methods are being developed for more rapid assessment of these connections, e.g., using the water's Natural Organic Matter (NOM) in the water as a water "fingerprint." Regardless of the specific use of source tracking there are different methods to determine the connectivity. Each method

has its pros and cons and deciding which one is more effective depends on the environment in which you are experimenting.

Tracers are used to elucidate a connection which is difficult to determine due to complexity of the system or inaccessibility to sampling points (Flury and Wai 2003). The use of tracers dates back about 2000 years ago when Philip, the tetrarch of Trachonitis, threw chaff into a crater lake and reported the chaff down gradient in a spring at the headwaters of the Jordan River (Davies et al. 1980). Tracers can be either in-situ or introduced by experiments. The selection of which tracer to employ depends on the environment.

2.6.1 Tracking parameters

Temperature

Tracking water temperature is common in ground water studies due to the ground water temperature difference when compared to surface waters. Water has a high specific heat capacity which means temperature change in water does not occur rapidly. Sampling a water source for temperature may require continuous sampling of an area to notice any changes (Cox et al. 2007). Monitoring a water source for an extended period can determine if a spike in temperature has occurred due to interference by another water source. There is a gradual change in temperature due to change in seasons, and that is why an extended monitoring regime is required for temperature gauging. Anderson et al. (2005) preformed a great review on the use of heat as a tracer in past and present groundwater studies (Anderson 2005). Temperature

is a natural tracer that does not require introducing any artificial tracers into the water. Temperature can be an initial indicator of interference but it is also wise to couple temperature readings with other tracer methods to get a larger range of data (Cox et al. 2007). Cox et al. (2007) coupled temperature with chloride and conductivity to create a better connectivity methodology which should be used to determine the source of pollution.

Organic Dyes

Dyes are effective tracers in investigating surface and subsurface hydrologic connections and have been used for over a century (Flury and Wai 2003). They are employed to analyze flow pathways, velocities, travel times, hydrodynamic dispersion, recharge, and discharge (Flury and Wai 2003). Organic dyes have a relatively low toxicity, are easily detected and cheap (Davies et al. 1980). The dye selected must be insensitive to changes in water chemistry such as pH, alkalinity, or ionic strengths. Dyes interact differently based on the environment they are used in so a pre assessment is necessary. Smart and Laidlaw (1977) evaluated characteristics of eight fluorescent dyes used in surface water tracing and their sorption onto organic and inorganic sediments (Smart and Laidlaw 1977).

Microbial Source Tracking (MST)

Microbial source tracking (MST) is a process commonly used to identify the sources of fecal contamination in waterways. MST is grouped into two major method types, library dependent and library-independent. The former depends on using

bacteria cultures from a water source and matching them with corresponding bacteria in a catalog. The latter uses sample-level detection methods to determine specific DNA genetic markers taken from the sample. Most MST studies are done by government agencies which are trying to regulate unhealthy concentrations of fecal coliform in public waters which are not meeting their designated uses. MST is able to link fecal bacteria to a specific source, human or animal, and then the proper solution can be devised by the agencies. Validation and improvement of MST methods are needed and are an ongoing process (Stoeckel and Harwood 2007).

Fluorescence spectroscopy

Natural Organic Matter (NOM) in aquatic waters is usually caused by the degradation of terrestrial plant matter. This NOM is transported through stream systems and estuaries out into marine environments (Stedmon et al. 2003). Tracing NOM in aquatic environments has been tested and promises to be a potential rapid assessment tool in the future. Some studies have looked at reviewing the structural characterization of NOM by laborious and time consuming methods (Leenheer 2004). However, for faster results, fluorescence spectroscopy can be beneficial in characterization of NOM because it is inexpensive and requires little to no pre-treatment. Fluorescence spectroscopy is the process of exciting a molecule with a beam of light (usually UV light); then, allowing the molecule to reach an excitation stage and finally re-emitting some of the light when the molecule leaves this excited stage. Measurements are taken at the range of wavelengths when the molecule becomes

excited and re-emits the energy. This range of wavelengths is known as the excitation and emissions matrix (EEM). The use of fluorescence in water science has grown extensively in the last 50 years and continues to grow (Hudson et al. 2007).

Fluorescence has been used to detect trace chemicals that may be leaching from a surrounding landfill or pollutants. It can be used to detect the impact of anthropogenic sources on fresh and marine waters. New studies have been looking at using EEM in water treatment plants (Bro and Andersen 2003). Because drinking water treatment plants must understand their source waters to maximize their treatment, new viable sources in tracking NOM must be employed in these facilities. Tracking NOM at each stage in the water treatment plant is becoming a more common practice because of new Stage 2 disinfection by-product (DBPs) regulations(EPA 2012). These DBP regulated treatment plants remove NOM through coagulation, sedimentation, and filtration basins, all of which should change the EEM signature of the NOM. This proves to be a good place to begin understanding how EEM can be used as a tracking tool through natural water bodies. Also if the use of EEM increases in water treatment plants, it may be possible for this technology to employed more readily in water resource management.

2.6.2 Parallel Factor Analysis (PARAFAC)

Recently EEM with Parallel Factor Analysis (PARAFAC) has been used to characterize NOM. In the past, most EEM analyses were performed by visual identification of peaks and ratios of fluorescence in different regions of the spectrum.

However multivariate data analysis has allowed researchers to study NOM using much easier techniques. Some common methods are Principal Component Analysis (PCA), Tucker3, and PARAFAC. Tucker3 and PARAFAC are decomposition methods of PCA. This means there are constrained models that can allow for more guided analysis. PARAFAC has recently been the selected method of choice and is a tri-linear decomposition method that creates a 3-way array to allow for analysis in MatLab or other tools. The three arrays are excitation, emission, and sample number (Stedmon and Bro 2008).

Stedmon and Bro's (2008) article entitled "Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial" is a great introduction to EEM with PARAFAC analysis. EEM analyzes samples by combining the fluorescence (emission) spectra measured from a series of different excitation wavelengths. When running EEM, excitation has a testing range of 250 to 400 nm and emission spectra from 350 to 500 nm (Stedmon and Bro 2008). An EEM analysis with PARAFAC allows the user to characterize and quantify changes in NOM. PCA analysis is not capable of quantifying any type of fluorescence data and is only a qualitative model; this is why PARAFAC has been used so widely. The PARAFAC model integrates three-way data; excitation, emission, and sample number as can be seen in the following equation:

$$X_{ijk} = \sum a_{if} b_{jf} c_{kf} + e_{ijk} \qquad i = 1, \ldots., I: j = 1 \ldots . J: k = 1 \ldots K \qquad \text{Eq. 1}$$

Where a=relative concentration of analyte f in sample i, b=excitation of analyte f, c=emission of analyte f , e= residual element

It is assumed that the Equation 1 behaves according to the Beer-Lambert's law and there is no interaction between the different fluorophore. The fluorescence characteristic (peak) is assumed to shift, but it should only increase its intensity based on the concentration of each fluorophore. Also, any changes in the analytes' local environment (temperature, metal concentrations, etc.) are assumed to be minimal (Stedmon and Bro 2008).

Sample Size/ Pre-treatment

One of the initial questions with PARAFAC is how many EEM samples are required; the answer being--there is no exact number. However, 20-100 samples are preferred and a sampling numbering upward toward 100 makes the model easier to converge (Stedmon and Bro 2008). According to Stedmon and Bro (2008), the best samples follow a gradient (e.g., mixing), or follow a process (e.g., drinking water facility). This allows for easier interpretation of results because you can follow the change in NOM. Initially the instrument must be calibrated with either pure water or quinine sulfate. This increases the accuracy of the quantitative (signal intensity) and qualitative (spectral shape) outputs. Instruments may have small imperfections in their ability to transmit light and a proper calibration should remove them. The other requirement in pre-treatment is the removal of the effects of Raman and Rayleigh scatter (Rinnan et al. 2005; Thygesen et al. 2004), this is done by subtracting the pure water spectrum from the sample spectrum (Stedmon and Bro 2008).

Selecting Components

The PARAFAC model requires fitting the data using one, two, three or more components. This is done by starting at a low number of components and moving up until the right number of components is reached. Components indicate the different fluorophores present in the samples and selecting the wrong number of components would rescind the analysis. In addition to selecting the correct number of components, outliers must be identified and removed if they distort the data. According to Stedmon and Bro (2008), an outlier does not have to be incorrect data but can be data that is correct but abnormal to the average. The next step involves model validation and data interpretation. The model must be validated in order to determine if the correct number of components was used. Stedman points to using split half, which splits the samples into two different halves and then models them independently. The components are derived for each half and if they are the same then the model is considered robust. Finally Stedmon and Bro (2008) go into data interpretation. The PARAFAC model states that if the identity of the component is unknown, then the concentration also cannot be known. However the fluorescence intensity at the maximum can be derived for each component.

2.6.3 PARAFAC: Rapid water monitoring tool

Practical use of PARAFAC requires an understanding of the constraints and the capabilities of the model. EEM's first applications were in chemometrics and food sciences. EEM has yet to spread to more main stream analytical chemistry due to the

complexity of modeling the data (Bro and Vidal 2011). However, it has been widely used in fingerprinting water samples from fresh water, marine water, drinking water, and industrial facilities. Hua et al. (2007) looked at fingerprinting four specific water sources; landfill leachate, wastewater treatment plant outflow, lakes, and rivers.

The motivation for this thesis focused on development of a rapid water monitoring and source identification tool. Missouri residents had previously experienced flooding which had contaminated drinking wells and surface water with landfill leachate and wastewater. It was difficult to identify which wells and water sources where polluted. However, EEM can be used to properly identify water which has been contaminated. Each water source has a distinct fluorescence signature when analyzed using EEM. Humic acid-like fluorescence occurs at 420-450 nm from excitation at 230-260 nm and 320-350 nm. Protein or amino acids fluorescence occurs with maximal emissions between 300-305 nm and 340-350 nm and is found commonly in uncontaminated river water, ground water and sea water (Hua et al. 2007). Sewage wastewater and landfill leachates also have distinct fluorescence substances which are used to detect their presence in surface waters such as humic and lignin substances, steroids, phenols, non-volatile acids, oils, and trace quantities of surface-active agents (Hua et al. 2007). Studies have looked at the impacts of wastewater on surface waters by examining fluorescence of upstream and downstream samples from the point source. It was discovered that a fulvic-like centre and a tryptophan fluorescence center were much more prevalent on downstream portions (Hua et al. 2007).

Hua et al. (2007) attempted to help Missouri Department of Natural Resources (MODNR) explore the use of EEM with a PARAFAC analysis. MODNR had previously explored fluorescence to trace landfill leachates (in the 250-450 nm fluorescence range) in surface water due to the floods of 1993 which decimated Missouri. Their technique was a two-step process that involved characterization of known water sources and then classification of unknown waters by comparison with known sources (Hua et al. 2007). Self-identification of EEM by comparison of known and unknown substances is difficult and has limitations which can cause subjectivity in the results. Hua et al. (2007) employed the PARAFAC model in order to fingerprint different water sources and develop a more objective method for MODNR.

PARAFAC: Post processing of data

Hua at al. (2007) analyzed four different waters using EEM and PARAFAC to extract their fluorescence signature. Dilution of the high concentration substances like wastewater and landfill leachate is necessary to keep the range of EEM absorbance scans less than .05 nm. One method to determine the correct component in the EEM is to convert the data signature data into Excel spreadsheets and then use N-Way toolbox in MATLAB to fit the PARAFAC model. The N-Way toolbox was developed by Raham Bro (Bro 2004), and has been made available online. The toolbox includes an online tutorial which guides the user through the steps in a PARAFAC model with examples and descriptions.

The determination of the ideal component number can be difficult when dealing with field-collected samples because it is impossible to know characteristics of the fluorophore present in the samples. However there are commonly accepted methods to determine the component number which are to: 1) test the effect of increasing the number of components on the number of iterations used to fit the model; 2) use a split half analysis for two sample sets to describe the similarities or variations, and (3) calculate the core consistency diagnostic (CORCONDIA) as a function of the component number (Bro 2004). Hua et al. (2007) used the CORCONDIA method in order to fit the model, and according to this method, the PARAFAC model is valid if the core consistency is close to 100%. If the value is close to 0%, then the model considered invalid and if it is around 50%, then the model is considered unstable. Their CORCONDIA values were 47, 75, and 92% when the number of components used was 5, 4, and 3, respectively. However, when the number of components was 3, two of the three emissions loadings were bimodal, which indicates that the number of components chosen was too small. When the number 4 was used, only one of the four emissions loadings were bimodal; thus, 4 components were selected as optimal (Hua et al. 2007). Each component has specific characteristics that were investigated and noted by researchers.

According to Hua et al. (2007), some of the recognized fluorescing components were previously identified in natural and anthropogenic waters. Components are identified based on the range where they exhibit their peak. These ranges are either exported in Excel graphs or displayed on graphs. Component 1 is the dominant fluorophore in river waters; however, in wastewater, Component 2 is the dominate

fluorophore. For landfill leachate, Component 3 dominates, and in lake waters, Components 1, 3, and 4 are present at almost the same percentage which reflects a complex source water mixture.

Final validations can be done with PARAFAC to confirm that the model fit correctly. In Hua et al (2007), one lake's water sample, one treated wastewater sample, and two river samples were selected and then classified as having unknown sources. The remaining 11 samples were used to fit the PARAFAC model using the previous method. Then the unknown samples were cross referenced with the outputs from the PARAFAC model, and it was confirmed that the unknown sources fit in with the known sources in a manner that was similar to the previous PARAFAC fit using all of the samples. Overall, this study showed how EEM with PARAFAC can be used to analyze multiple water source types and fingerprint specific fluorophores that may lead to proper tracking of contamination from different sources.

2.6.4 Water quality parameter correlations

Studies have shown that the Total Organic Carbon (TOC) concentrations can be precursors for Trihalomethanes (THM) formation. TOC is the measurement of carbon bound in an organic compound and is an indicator of water quality in drinking water facilities. Bieroza et al. (2009) is able to use fluorescence to characterize NOM in a drinking water facility and then use that data to predict TOC removal efficiencies. Based on the collected data, the researchers received a correlation coefficient value of R^2 = 0.90, which is significantly linear. Previous approaches to inferring TOC removal at a

drinking water facility used standard jar tests, ultra violet absorbance, and zeta potential measurements. However these processes had limitations in precision and ease of use (Bieroza et al. 2009). Studies have shown that fluorescence is more stable than UV in detecting Dissolved Organic Carbon (DOC) removal from reservoir water (Baghoth et al. 2011). Bieroza et al. (2009) points to the possibility of using EEM and correlating to other water quality parameters commonly collected.

The use of PARAFAC has been demonstrated in the above literature, and its application to drinking water facilities has been explored (Baghoth et al. 2011). It is evident based on these studies that EEM is often more accurate and more timely than other methods in describing NOM in waters. Baghoth et al. (2010) showed significant correlation ($P<0.01$) between PARAFAC components, UV_{254} and DOC concentrations. While this study successfully demonstrated these correlations, work must still be done on correlating the PARAFAC components with other water quality parameters that are commonly tested in drinking water facilities (Baghoth et al. 2011)**.** Past studies have correlated dissolved organic matter (DOM) signatures using EEM with THM precursors in 55 lake samples collected in Missouri (Hua et al. 2010). In this thesis EEM did prove to have significant correlations with some commonly tested water quality parameters and thus its use in elucidating connectivity in surface waters can be further explored.

CHAPTER 3 : DISPLAYING WATER QUALITY DATA USING GIS AND DIFFERENT INTERPOLATION METHODS

3.1 Introduction

Wetlands have been in the spotlight for many organizations that deal with environmental regulations. Recent cases such as “Rapanos vs. United States” have intensified the debate on what constitutes a wetland which can be regulated under the Clean Water Act (CWA). This case stated that if a biological, chemical, or hydrological connection could be determined between an area designated as “waters of the US” and a wetland, then that wetland may be regulated under the CWA (Macdonald 2007). However, the case did not state the method to take in determining these connections, and, thus, researchers must develop new methodologies in determining the stream to wetland connectivity.

One way to elucidate a connection between waterways is to track the change in water quality. This requires water quality data at two discrete locations to determine any similarities in the parameters. Previous EPA water quality monitoring initiatives have not included collecting wetland data as a high priority (Dahl and Watmough 2007). EPA initiatives have pushed for water quality monitoring at locations of immediate concern such as impaired streams or drinking water reservoirs. Wetlands, therefore, are seldom sampled for water quality parameters. However, surrounding areas may have water quality data that can be applied to the wetlands if needed.

Resource managers must be innovative in their use of currently available data in the areas surrounding wetlands. This project examines different interpolation methods which may be used to create additional data to be applied to these wetlands. Spatial interpolation methods create data in barren locations from observed data in adjacent areas. Spatial interpolation is most effective when the density of discrete points is significant (Luo et al. 2008). However this is not always the case because monitoring and sampling is an expensive practice.

Online databases were data mined to acquire pertinent water quality information. These databases are run by federal agencies such as the Environmental Protection Agency (EPA) and the United States Geological Survey (USGS) as well as smaller state and local entities. Overall the goal of the project was to determine new methodologies to display stream and wetland connectivity by using publically available data.

3.2 Methods

3.2.1 Site Location

The Lower Grand watershed (8-digit HUC 10280103) was selected as the pilot area for development of the stream-wetland connectivity methodology. The Missouri Department of Natural Resources (MODNR) and EPA suggested the Lower Grand in order to complement state initiatives. MODNR has designated the Lower Grand as one of three pilot watersheds being studied in order to improve management and decision making tools at a watershed level (MODNR 2012). The watershed is approximately 1.5

million acres located in Northern Missouri and a portion in Southern Iowa (Figure 3-1). The Lower Grand watershed drains into the Grand River and the Missouri river following drainage paths formed by Medicine Creek, Locust Creek, and West Yellow Creek. The Lower Grand watershed contains 303d listed impaired water bodies which include the three creeks and a portion of the Grand River, Salt Creek, and Marceline New Lake. 303d listed waters are water bodies designated by the EPA as not meeting their appropriate designated uses.

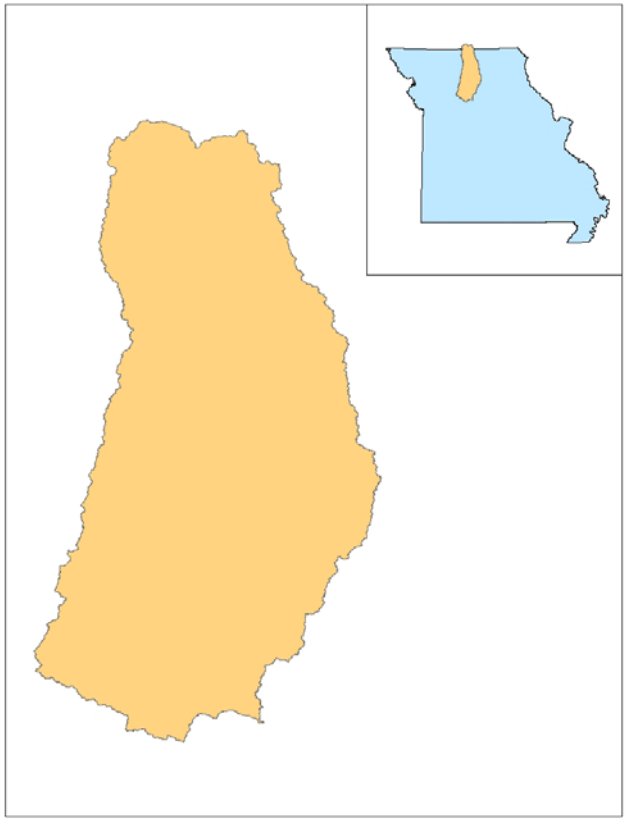

Figure 3-1: Lower Grand Watershed Location

3.2.2 Water Quality Parameters

The goal of this project is to develop a methodology to determine stream-wetland connectivity using publically available water quality data. This requires locating chemical data at wetlands and streams and then analyzing this data for spatial and temporal trends. Specific water quality parameters are selected from the databases as potential tracers. The parameters include the presence of chlorides (Cl^-), acidity, hardness, salinity, specific conductance, alkalinity, phosphorus, nitrogen, and dissolved oxygen. Water quality parameters are normally classified by their degree of degradation in the environment. Conservative parameters are usually physically or chemical inert in the environment. These parameters include but are not limited to salts or metals. Non-conservative parameters are transformed in the environment through physical and biological processes, and these usually include organic compounds. Thus the more conservative parameters (i.e., chlorides (Cl^-), acidity, hardness, salinity, specific conductance, and alkalinity) are seen as superior tracers in the environment (Cox et al. 2007; Driscoll et al. 2001).

All available data was downloaded online from the MODNR water quality database and organized into Excel spreadsheets for each water quality parameter. Ultimately chloride and specific conductance were selected as the trace chemicals to observe because they are considered conservative tracers (Cox et al. 2007). Cl^- had 34 sampling sites while specific conductance had 63 sampling locations in the Lower Grand compared to only 10 for hardness and 7 for alkalinity. Chlorides are considered

chemically and hydrologically inert and are not easily adsorbed to surfaces (Peters and Ratcliffe 1998). Cl^- can occur naturally from salt deposits but are more commonly linked to road salt applications (Kelly et al. 2007). Cl^- and specific conductance have been coupled in tracer studies in order to develop better methods in determining connectivity (Cox et al. 2007). Cl^- and specific conductance can sometimes show correlations due to the way specific conductance is measured using Total Dissolved Solids (TDS) in the environment like salts. The parameters do share a positive correlation of ***r term .49***.

Water quality data preparation

The chloride and specific conductance data were exported from the MODNR website and converted into Excel files for GIS analysis. The Excel spreadsheets were set up with sampling points, Missouri ID number, stream name, GPS location, sample type, analyte measured, unit of measurement used, and date measured. Each spreadsheet was organized by GPS location and date. A logical method for organizing the data had to be followed in order to reduce misinterpretation when converted into GIS. Some of Codd's Relational rules (1980) were followed in order to minimize any complications like one value per cell, unique rows, and no significance in row sequence (Belohlavek and Vychodil 2006). Each location received a yearly average for that parameter as well as monthly averages within each year. This was done to help show any seasonal changes which may have occurred for the parameter. Correct GPS data for each sampling point was and remains essential for input into GIS.

3.2.3 Available Databases

This project used publicly available data in all of the analyses. EPA, USGS, and other state and local entities have collected a multitude of water quality data over the years. This data is usually used by these government agencies for regulatory actions and to track historical changes in the environment. Most of the data collected is available in online databases for the benefit of researchers in academia or industry. While these databases hold a tremendous amount of information, some of them have received criticism based on misleading user interfaces. For example, in 2004, the US Chambers of Commerce complained that various EPA databases were beset with errors and uncertainties in some of the essential physical-chemical constants used (Renner 2005). They stated that in over 16 databases, 24 chemicals where listed with different values for the same physical-chemical constant. In one study the vapor pressure and octanol-water partition coefficient (K_{ow}) varied by orders of magnitudes. This means that users must understand each online database used and have an understanding of the possible benefits and pit falls. Thus specific databases were evaluated based on their ability to fit our water quality data needs. A brief description of each of the databases used and a discussion of the subsequent evaluation follows.

(a) STORET

STORET is a US Environmental Protection Agency (EPA) water quality database which was originally designed in the 1960s. It is the EPA's principal repository for federal, state, and local government monitoring data pertaining to marine, freshwater,

and biological parameters. The database is used by academics, the private sector, and watershed volunteer groups. The search frame is based on an interactive web based search engine that begins with specifying a location based on many factors (ZIP code, hydrologic unit codes (HUC), etc.). A database query can return extensive information about a specific location based on the watershed in which it is located. Some of the information includes impaired waters in the watershed, citizen groups that are testing watershed waters, and counties, which are located in the watershed.

Database use and evaluation

Data added to STORET is from volunteer and government organizations. Groups ranging from non-profit watershed organizations to state governments can upload data to STORET, but they must indicate the lab and procedure of collection for authentication. The collection organization name is included in the STORET data. This means the data may be peer reviewed but inaccurate data can still be uploaded. All citizens groups with contact information are on the "Surf Your Watershed" page. STORET can be a good initial database used for water quality data as it gives an accurate overview of the watershed, including water bodies in the watershed, counties, and impaired waters. STORET also provides links to important information pertaining to the watershed. However, since the database is national, it does not have a specific focus on Missouri. In contrast a state-focused database such as the MODNR water quality assessment system often contains a larger amount of data for specific regions.

(b) MODNR Water Quality Assessment System

The DNR water quality assessment system is an online database of water quality data. This data is generated by the DNR Environmental Services Program as well as other organizations which utilize DNR data. The database can be queried by the water body name, water body ID, and County name. The easiest method is to query the water body name and county. It is necessary to include the county in the search since the same names are often used on several water bodies in MO. After the water body is located, a certain media type (e.g., water, sediment, tissue) and or a specific location which includes site code can be searched. After the exact data is located it can be exported to outputs such as Excel spreadsheets.

Database use and evaluation

The amount of data available in the MODNR database as well as the simplicity of the search engine makes this database the most effective database for water bodies in Missouri. It should be used after the names of the selected water bodies are known because it makes searching watersheds much easier. A wide range of parameters are listed and are varied based on the water body and whether it was on the 303(d) list. This database allows one to compile a list of water body data for a specific county which can be used to develop a watershed overview.

(c) United States Geological Services (USGS) National Water Information System (NWIS)

The USGS water quality data was the third database considered for use in this thesis. This website provides data collected by the USGS and partners to meet federal

requirements of waters. The site offers real-time daily, monthly, and annual data. The data selected were taken from the field/lab samples area. There are more than 300,000 locations sampled nationwide. The NWIS site prompts you to select the state from which you want to retrieve samples. Then, the best way to narrow down the search is to choose the data to be displayed by HUC. The data can be exported as "unknown file type," which can be opened in Excel. However, the easiest method is to select the program to open the file in the browser and copy and paste into Excel since it is already in tabs and columns.

Database use and evaluation

The Lower Grand watershed has approximately 75 sample sites recorded on the NWIS website. The parameters were mostly Total Nitrogen (TN), Total Phosphorus (TP), and dissolved oxygen (DO) percent of saturation in mg/l. Even though there are 75 sites from the Lower Grand only 14 have been recently monitored. The USGS site is not a great site for obtaining parameters other than nitrogen or phosphorus. Therefore, this database for the USGS site was the least used for data acquisition and implementation in this thesis.

Database water quality collection methods

To better understand the quality of the data presented by the databases used in this thesis each organization received an email from the research group inquiring about who collected the data. EPA STORET and USGS stated that the data had variability in terms of who collected it. Some of it was collected by government officials, while other

data were collected by private companies or volunteer groups. MO DNR replied that they had most of the data collected by federally funded government organizations or by public works personnel of large municipalities like St. Louis or Kansas City.

Data processing and analysis using Geographic Information Systems Layers

Water quality trends are both a temporal and spatial problem which can be analyzed using GIS. GIS has increasingly been used in studies to effectively depict spatial water quality trends (Gupta et al. 2012; Rabah et al. 2011). This thesis utilized ESRI's ArcGIS 10 to perform these tasks. The chemical connectivity methodology is developed to fit primarily into the framework offered by ArcGIS 10. Each spatial analysis begins with specific base layers relevant to the study. These geospatial layers include landscape features, topography, and state and local boundaries. All georeferenced datasets are projected or converted into North American Datum (NAD)_1983_ Universal Transverse Mercator (UTM)_Zone_15N. The following data layers were obtained from the MSDIS website as well as the U.S. Fish & Wildlife Service :1:24,000 MSDIS Rivers and Streams (Vector-line shapefile) ,1:24,000 MSDIS Watershed boundaries (Polygon shapefile), 1:24,000 Wetlands: U.S. Fish & Wildlife Service, National Wetlands Inventory (Polygon shapefile), 1:24,000 MSDIS Streams in Lower Grand Watershed (Vector-line shapefile), MSDIS Missouri Counties (Polygon shapefile), 1:24,000 MSDIS Digital Elevation Model (DEM) of Missouri (60m x 60m continuous field shapefile, and 1:24,000 Transportation corridors like roads and railways (Vector-line shape file).

GIS offers an array of tools to properly asses the quality and quantity of water quality data. First, the tabular water quality data obtained from MSDIS was used to create a point representation for 24 sampling sites. This was done using the latitude and longitude values in the tabular data and converting them to points in GIS. The monitoring points were all located on major streams which are considered important based on their designated uses. The 303(d) impaired water bodies were consistently the most sampled sites. However, outside of the streams there was no water quality data on wetlands or surrounding lakes. Proper wetland data is necessary in order to create a stream-wetland connectivity methodology. GIS spatial analyst methods such as interpolation can be used to model the water quality data at these data barren wetlands. Below Table 3-1 summarizes the pros and cons of each database.

Table 3-1: Databases – Pros and Cons Summary

Database Name	Pros	Cons
USGS National Water Information System (NWIS) (http://water.usgs.gov/owq/)	Extensive online database with over 300,000 sites sampled nationwide	Some data was to outdated to use. Database deals with all 50 states and thus is not catered to MO
EPA STORET (http://www.epa.gov/storet/)	Gives clear overview of the watershed with Surf Your Watershed. A multitude of data has been uploaded from government, academia, and private entities.	Not all data collected in states were uploaded to STORET promptly. Credibility of samplers vary
MODNR Water quality assessment program (http://www.dnr.mo.gov/mocwis_public/wqa/waterbodySearch.do)	Data collected specifically in MO. Simple search interface and extensive data available.	Need to know which water bodies to search before coming to website. No overview of watersheds available.

3.2.4 GIS Interpolation methods

Deterministic methods

Global Polynomial Interpolation (GPI)

GPI is an inexact interpolator that attempts to fit a smooth surface which is defined by a polynomial function. GPI is commonly used in large, coarse scale, spatial problems. These problems commonly include air pollution predictions. GPI uses polynomials to fit the data. So for example a first order polynomial fit would demonstrate a linear trend on a large scale. Outliers in the dataset greatly decrease the accuracy of GPI. (ESRI 2011)

Local Polynomial Interpolation (LPI)

LPI is an inexact interpolator that also fits data into polynomial functions similar to GPI. However LPI fits this data into localized search neighborhoods as defined by the user (Luo et al. 2008). LPI fits each localized area with a specific polynomial function corresponding to the general trend in that area. This gives the model increased accuracy when dealing with localized trends that GPI would simply overlook as it fits trends over large areas (ESRI 2011).

Inverse Distance Weighted (IDW)

IDW is an interpolation method that assumes that things close to one another are more likely to be similar than things far away (Luo et al. 2008). IDW interpolates values at a location based on observed values surrounding the location. IDW gives the

greatest weight (w) to observations closer to the location, and the weights diminish as distance (d_i) increases as shown in Equation 3.1. The value of each weight is proportional to the inverse distance between the known data point and the location to be interpolated. IDW is an exact interpolator meaning the maximum and minimum values interpolated are the same as the observed values (ESRI 2011).

$$Z'(s_0) = [\sum_{i=1}^{N} w(d_i)Z(s_i)] \div [\sum_{i=1}^{N} w(d_i)] \quad \text{Eq. 3.1}$$

where $Z'(s_0)$ = predicted value, $Z(s_i)$ = observed value, N is the number of measured sample points, w(d) is the weighting function and d_i is the distance from s_0 to s_i.

Geostatistical methods

Kriging

Kriging is a geostatistical interpolation method that can be used to fit stochastic data. This means the variables should be ordered randomly on a spatial scale. Kriging relies on autocorrelation to interpolate values based on observed samples. Autocorrelation in kriging is a function of distance and assumes things closer together tend to be more similar than things far apart (ESRI 2011). In this thesis, ordinary, simple, and universal kriging are considered. Ordinary kriging is the most popular type of kriging and assumes a constant but unknown mean. Simple kriging is mathematically the simplest and depends on the mean being known beforehand even though it is usually

unknown. Universal kriging is similar to ordinary kriging but assumes the data is deterministic and tries to fit a polynomial trend (Luo et al. 2008).

Validation of interpolation Methods

The four interpolation methods are cross-validated based on ArcGIS outputs. Each interpolation method outcome is associated with a mean error (ME) and a root mean square error (RMSE) which is used in cross validation (ESRI 2011). The ME is used to detect bias and the value should be close to zero if it is unbiased. Biased interpolations are models where the interpolated values are over or below the observed values. The RMSE indicates how accurately each method interpolates the values relative to the observed values. The best interpolate method should ideally correspond with the smallest RMSE (Luo et al. 2008). The following equations show the mean error (ME) and root mean square error (RMSE) which are used to validate the different interpolation methods.

$$\mathrm{ME} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}[z'(s_i) - z(s_i)] \qquad \text{Eq. 3.2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}[z'(s_i) - z(s_i)]^2} \qquad \text{Eq. 3.3}$$

Where: ME=mean error, RMSE=root mean square error, N=number of interpolated values Z' (s_i) = predicated value, z (s_i) = measured value.

3Model Optimization

ArcGIS gives users the ability to increase the accuracy of the models by changing specific attributes in the analysis. When an interpolation model is initiated, ArcGIS automatically populates the model assuming different values and trends. However, the easiest way to change the model is to simply click the optimize button located in the model options. This tool recalculates values in the model to directly represent the data.

Semivariogram and Covariance

In the kriging model, the users can select the autocorrelation method. In kriging, autocorrelation among observations is a function of distance, assuming that things nearby tend to be more similar than things that are father apart. The semivariogram and covariance are mathematical representations of autocorrelation. In the semivariogram, if two locations are close together in terms of distance then the difference in the Z (s_i) - Z (s_j) will be small. As they get farther apart, they become less similar and their difference Z (s_i) - Z (s_j) becomes larger. Because the variance of the difference increases with distance, semivariograms can be thought of as a dissimilarity function. The semivariogram is defined in Equation 3.4, where *var* is the variance (ESRI 2011):

$$\gamma(s_i, S_j) = \frac{1}{2}\mathrm{var}\left(Z(s_i) - Z(s_j)\right) \quad \text{Eq. 3.4}$$

Covariance is a scaled version of correlation. When the two locations s_i and s_j are close together they will be similar and their covariance will be large. As s_i and s_j move apart and become less similar, their covariance becomes zero. Given the covariance function decreases with distance, it can be thought of as a similarity function. Both

covariance and semivariograms are used to fit the data and create a more accurate prediction. The covariance function is defined in Equation 3.5, where *cov* is the covariance (ESRI 2011).

$$C(s_i, s_j) = \text{cov}\left(Z(s_i), Z(s_j)\right) \quad \text{Eq. 3.5}$$

Anisotropy

Anisotropy is another interpolation parameter that can be changed to increase the accuracy of the model to increase the accuracy of the model. This parameter takes into consideration potential directional trends in the data. This is especially helpful in interpolation data that follows a gradient such as wind or water flow. In this thesis the data does follow stream flow so choosing to use anisotropy actually reduced the RMSE in some of the models. The model gives you an option for the major and minor axis and by altering these values you incorporate directional bias in the model. Some models' RMSE can be reduced with this method while others will not change. The data demonstrating these affects is presented in the appendix.

Lag Distance

Lag distance is the neighborhood search distance used to constrain the set of observed values used to interpolate a value at a location. It is important that the correct lag distance is selected to optimize the model. If a small distance is selected, surrounding influences may be left out and if a large distance is selected then all the trends will be lumped together. One method to determine the lag distance is to use the

Average Nearest Neighbor tool in spatial statistics tools. The optimize button also creates a lag distance for each map (ESRI 2011).

There are additional options to optimize the models. The kriging model offers the most optimization options. Our data set was quite small and changing some of the options did not alter the RMSE or ME. Interpolation can be a very complicated tool and can have large amounts of different outputs based on optimization features.

Subwatershed Interpolation

Chemicals in a stream will tend to follow drainage paths determined by topography. ArcGIS allows users to upload digital elevation models (DEM) to display and analyze these topographic changes. Based on the understanding that chloride (Cl^-) tends to be chemically inert and moves with the flow, we decided to take an interpolation approach using subwatersheds as boundaries. Each subwatershed was assigned a value based on the corresponding Cl^- or specific conductance drainage point used to create the watershed. These subwatersheds are based on the topographic flow of water in a specific area. A subwatershed's size depends on the drainage point used to delineate the watershed.

The method proposed in this thesis involves delineating the watersheds using ArcGIS Hydrology tools (Sun et al. 2008) . Chloride and specific conductance values are considered for each data point because of the known correlations between chloride

levels and conductance. The year 2004 is selected for analysis because it had the most data values at 24. The following steps are used to create the subwatersheds:

a) fill the sinks in the DEM,

b) create flow direction,

c) create flow accumulation,

d) create watershed pour points, and

e) finally delineate the watersheds.

After the watersheds are delineated then each watershed can be assigned their respective chloride and specific conductance values. Figure 3-2 displays the first 4 steps from A-D.

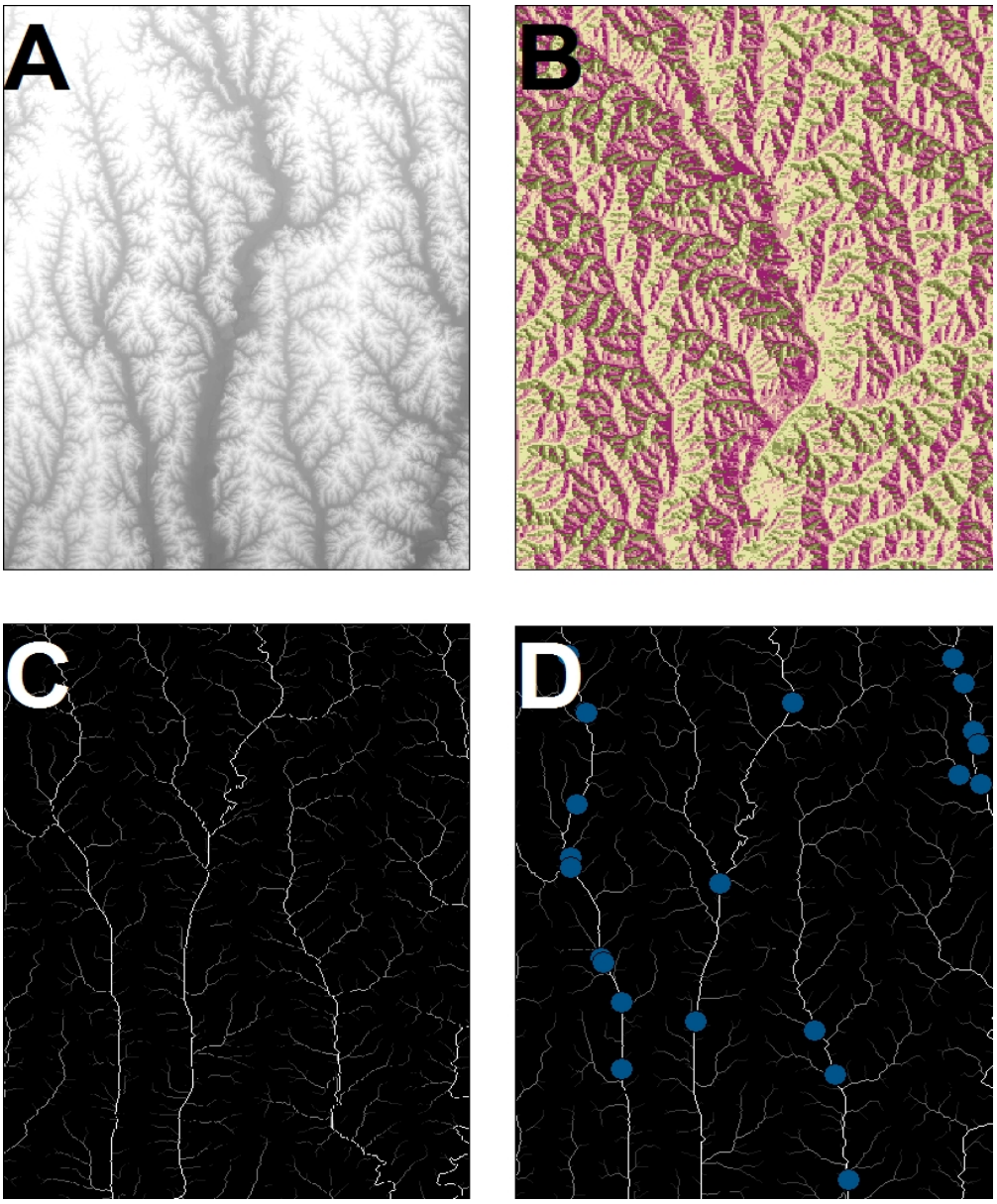

Figure 3-2: Layers used to create subwatersheds: a) fill the sinks in the DEM, b) create flow direction layer, c) create flow accumulation layer, and d) create flow accumulation layer with watershed pour points

Filling Sinks in the DEM

First, the DEM layer must be filled using ArcGIS in order to remove any imperfections in the model. The DEM layer must be depressionless for all the hydrological analysis to work. If there is an internal drainage area on the DEM, the

process will not create proper data. This tool is found in spatial analyst tools > Hydrology > fill. The DEM files are the data input and the output file was a DEM raster with filled in depressions.

Create flow Direction

The next step involves using another ArcGIS hydrology tool to determine flow direction. Flow direction is needed to determine the destination of the surface flow. The flow direction tool analyzes the neighborhood of cells in the DEM to determine each cells steepest adjacent cell. The tool is found in Spatial Analysis Tools > Hydrology > Flow Direction. The input fill was the filled DEM from step a, and the output file is a raster file demonstrating flow direction.

Creating Flow Accumulation

This next tool is a function to calculate the flow in each cell by accumulating the cells that flow into each down sloped cell. The tool is located in Spatial Analysis Tools > Hydrology > Flow Accumulation of ArcGIS. The input is the layer produced from the flow direction layer, and the output displays the accumulated flow direction of raster cells. The initial output did not represent the flow accumulation clearly. The symbology needs to be changed in order to clearly understand the flow accumulation. Therefore, the data values are reclassified from 5 classes to 10 classes 100, 200, 400, 800, 1600, 3200, 6400, 12800, 256000, and 256000 to the maximum amount of accumulated cells to represent the amount of cells flowing into each location respectively.

Creating Watershed Pour Points

The watersheds are delineated based on the location of each monitoring point. The monitoring points had to be created into a new shapefile and snapped to the flow direction layer. Pour points were then selected on the flow accumulation raster as the outlet points were used to create each subwatershed. Each pour point was placed adjacent to the known water quality sample locations. The created pour points were then snapped to the flow accumulation raster to ensure the points were located on the highest flow accumulation cell within the neighborhood. This step also served to convert the pour points into a raster format. A snap distance of 100 m limited the search for the highest accumulated flow in the vicinity of the pour points.

Delineating watersheds

To finally create the subwatesheds, the flow direction layer and the snapped pour points were inserted into the watershed tool in ArcGIS. After analysis, a total of 24 watersheds were delineated, corresponding to the 24 chloride/conductivity values for the year 2004. The output file was in raster format which was subsequently converted to polygon features for further analysis.

Assigning water quality values to wetlands

The watersheds were delineated based on the drainage to each sampled chloride location. Next, the watersheds were assigned their respective chloride/conductivity values by computing a new attribute field. Initially, zonal statistics

methods were attempted in order to assign each chloride value. However, the output was similar to a DEM and did not seem to interact well with the watershed polygons. Next, the chloride pour points were joined based on 2004 chloride values with the actual corresponding chloride point. The water quality attribute data was then joined with the watershed for each pour point. This allowed water quality monthly values, stream names, and MO_ID numbers to also be associated with each subwatershed. The final step required using the geometric intersect tool in ArcGIS to intersect the watershed polygons with the NWI wetland inventory polygons to associate each wetland polygon with the watershed in which it was located..

3.3 RESULTS

The results from the interpolation models were evaluated using spatial analyst tools in ArcGIS. Each of the six initial methods underwent extensive runs and alterations to determine the best model based on low RMSE scores. Specifically, chloride data and specific conductance data for the year 2004 were modeled because they had the most data available from any year.

Ordinary Kriging

The chloride Ordinary Kriging ideal output was the optimized model given by ArcGIS fitted with a semivariogram not using anisotropy with a lag distance of 3900 m. The specific conductance ideal model is the covariance optimized with anisotropy true and a lag distance of 3900 m. The chloride and specific conductance interpolations are listed in Figure 3-3 & 3-4 denoted as C. Ordinary kriging had the lowest ME for the

specific conductance interpolation, which implies that this interpolation is the least biased. However, it received the second largest RMSE value for the chloride interpolations and the third lowest RMSE values for conductivity interpolations. Ordinary kriging assumes an unknown constant mean, which accepts that the data follows a set trend. It is difficult to assume an average value in environmental data when it is so variable.

Simple Kriging

The chloride simple kriging ideal model was optimized with covariance, anisotropy true, and a lag distance of 3900 m. Also the specific conductance ideal model was optimized with covariance, anisotropy true, and unaltered lag distance. Simple kriging received the second lowest RMSE score for chloride interpolations. Simple kriging interpolations in Figure 3-3 & 3-4 are denoted by F. These figures show the most unpronounced trends of all the interpolations. This is because simple kriging selects a known constant mean which forces the interpolation to take a “simple” approach.

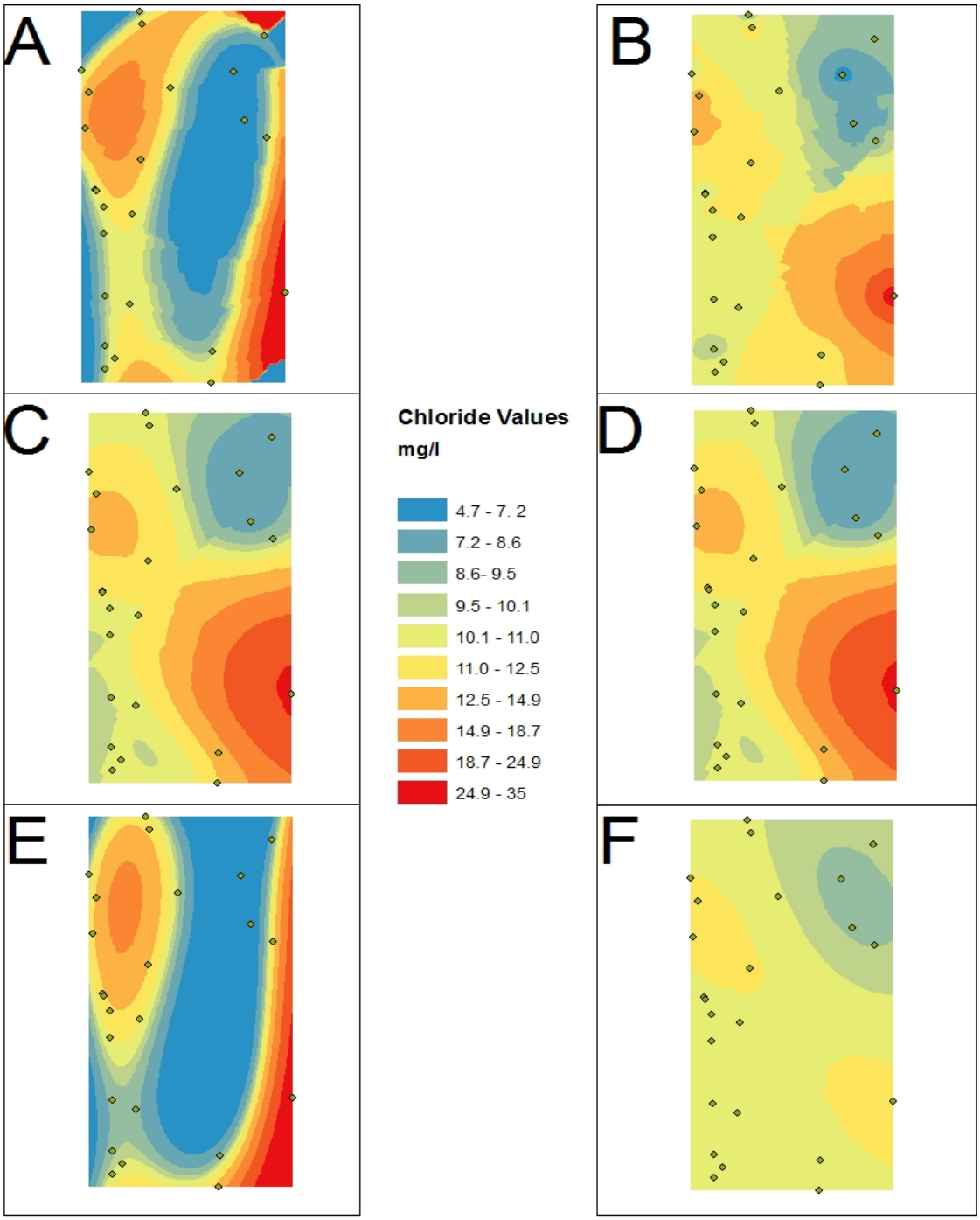

Figure 3-3: Chloride Interpolation model outputs with data point's locations shown. A = LPI, B = IDW C = ordinary kriging, D = universal kriging, E = GPI, F = simple kriging

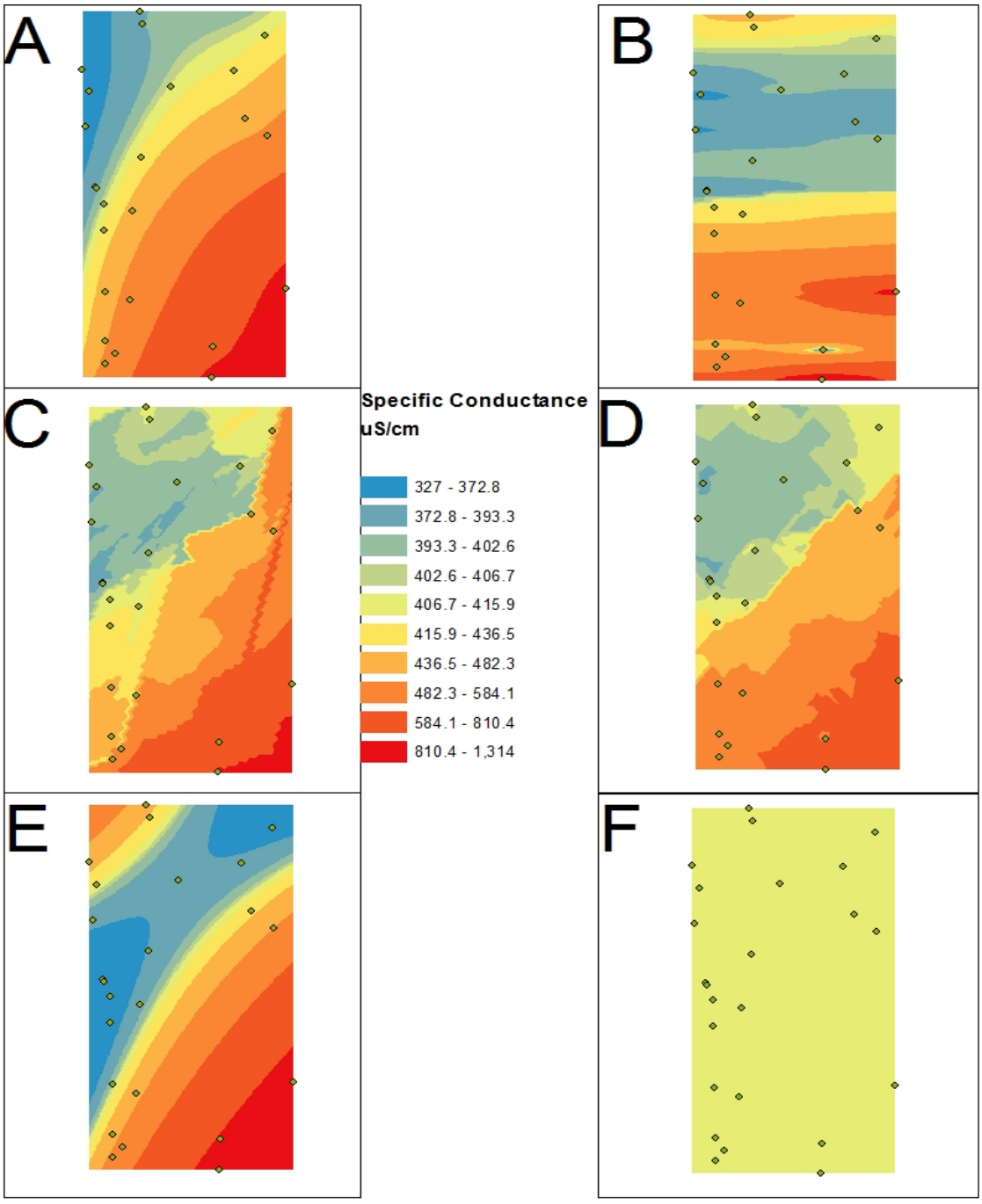

Figure 3-4: Specific conductance interpolation model outputs with data point's locations shown. A = LPI, B = IDW, C = ordinary kriging, D = universal kriging, E = GPI, F = simple kriging

Universal Kriging

The chloride universal kriging ideal model is the optimized model with semivariogram, anisotropy false and lag distance unaltered. The specific conductance ideal model is optimized with covariance, anisotropy false, and lag distance unaltered. Universal had the second lowest RMSE value for the specific conductance interpolations. Both of the interpolations in Figure 3-3 & 3-4 (D) seem to clearly demonstrate high- and low-value hotspots. Universal kriging was the optimal interpolation model for the specific conductance. It received the lowest RMSE value for specific conductance followed by Ordinary Kriging.

Deterministic models

The following deterministic models did not have semivariogram or covariance but instead the models could be fit to different polynomial powers. Also the anisotropy was not simply set to true or false. Instead a factor had to be selected based on the ratio to major and minor axes. A measured point near the axis with the higher value was given larger weights in determining the interpolation.

Inverse Distance Weighed (IDW)

IDWs ideal chloride model was optimized to the power of 1.0 with an anisotropy factor of .909. The specific conductance model was also optimized to the power 1.0 but with an anisotropy factor of .139. IDW received the highest RMSE value for the specific conductance interpolations making it the worst model in that category. IDW is extremely sensitive to outliers and the specific conductance values had two potential

outliers that would have fooled the IDW interpolation. This is made clear in Figure 3-4 (B) as the other interpolators incorporated diagonal trends while IDW only displayed horizontal trends.

Global Polynomial Interpolation

GPI only had the option to change the polynomial order number. For chloride, the ideal number was found to be 3.0 and for specific conductance the ideal number was found to be 2.0. In Figure 3-3 & 3-4 (E), GPI exhibited trends unlike the other interpolations. This is due to the way GPI smoothed over any local trends and fit the data into coarse scale trends. For local phenomenon like water quality changes, GPI would not be able to display the discrete data changes needed for these studies.

Local Polynomial Interpolation

LPI ideal chloride model is the optimized model fitted with a third order polynomial and anisotropy factor of 1.0. The ideal specific conductance model is optimized with a first order polynomial fit and anisotropy factor of 1. Originally the polynomial fit of 3.0 was the best model for specific conductance. However, after further inspection, it was discovered that the optimized model actually removed a data point from the interpolation. The data point had the largest value, and the model decided to remove it to make the interpolation smoother. This data point was not an outlier because it was high in the previous years as well as the following years. The LPI model specific conductance model also removed the largest value for the optimized

third order polynomial. The smallest RMSE value before the values were removed was used for both data fits.

Subwatershed Interpolations

Areal interpolation is common when dealing with spatial units of analysis such as Census blocks and ZIP codes. Data must be scaled up and down in some instances in order to convey a message. This project scaled up point data into subwatersheds to display a non geostatistical interpolation method. Figure 3-5 displays the specific conductance interpolations and Figure 3-6 displays the chloride interpolations. There were no quantitative comparisons done neither with these interpolations nor with the other methods. This method is only compared qualitatively. The two legend scales used for chloride and specific conductance do not have any correlation so the color differences should not be used as a comparison for the methods. However, when overlaying the six interpolation methods to their corresponding subwatershed interpolation, you can see some similarities in the interpolation methods. However, it is understood that creating such a coarse scale interpolation will not always be the most effective method to take. Researchers examine how changing data from fine scale into coarse scale impacts its validity (Kar and Hodgson 2012). This problem, known as the modified areal unit problem (MAUP) introduces a statistical bias based on the size of the aggregation unit. The size of the subwatershed can be minimized in the future to create finer scales of interpolation. The average size of the subwatersheds is 13,492 acres, the minimum size is 85 acres, and the maximum size is 51,174 acres.

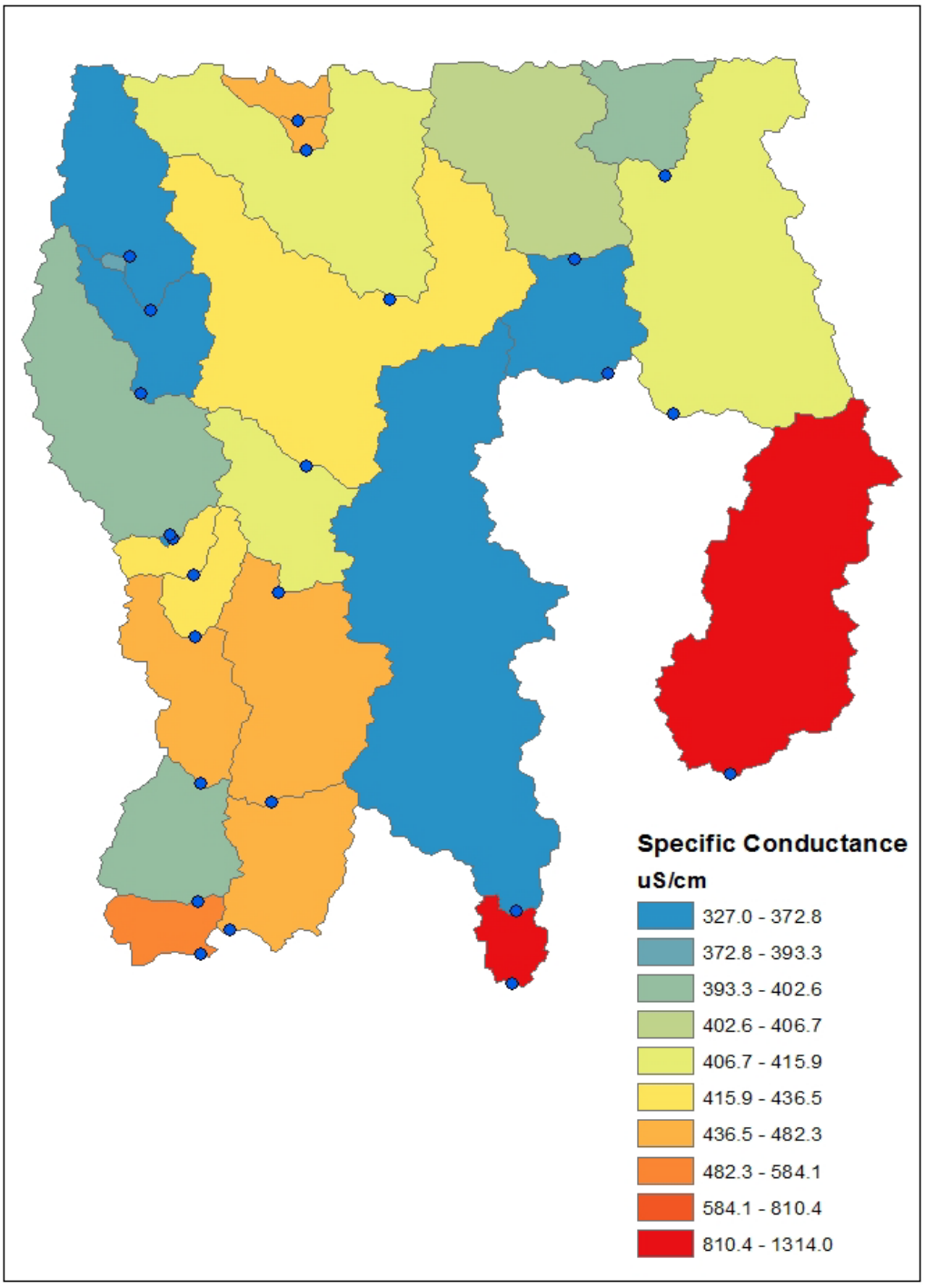

Figure 3-5: Subwatershed interpolations for specific conductance

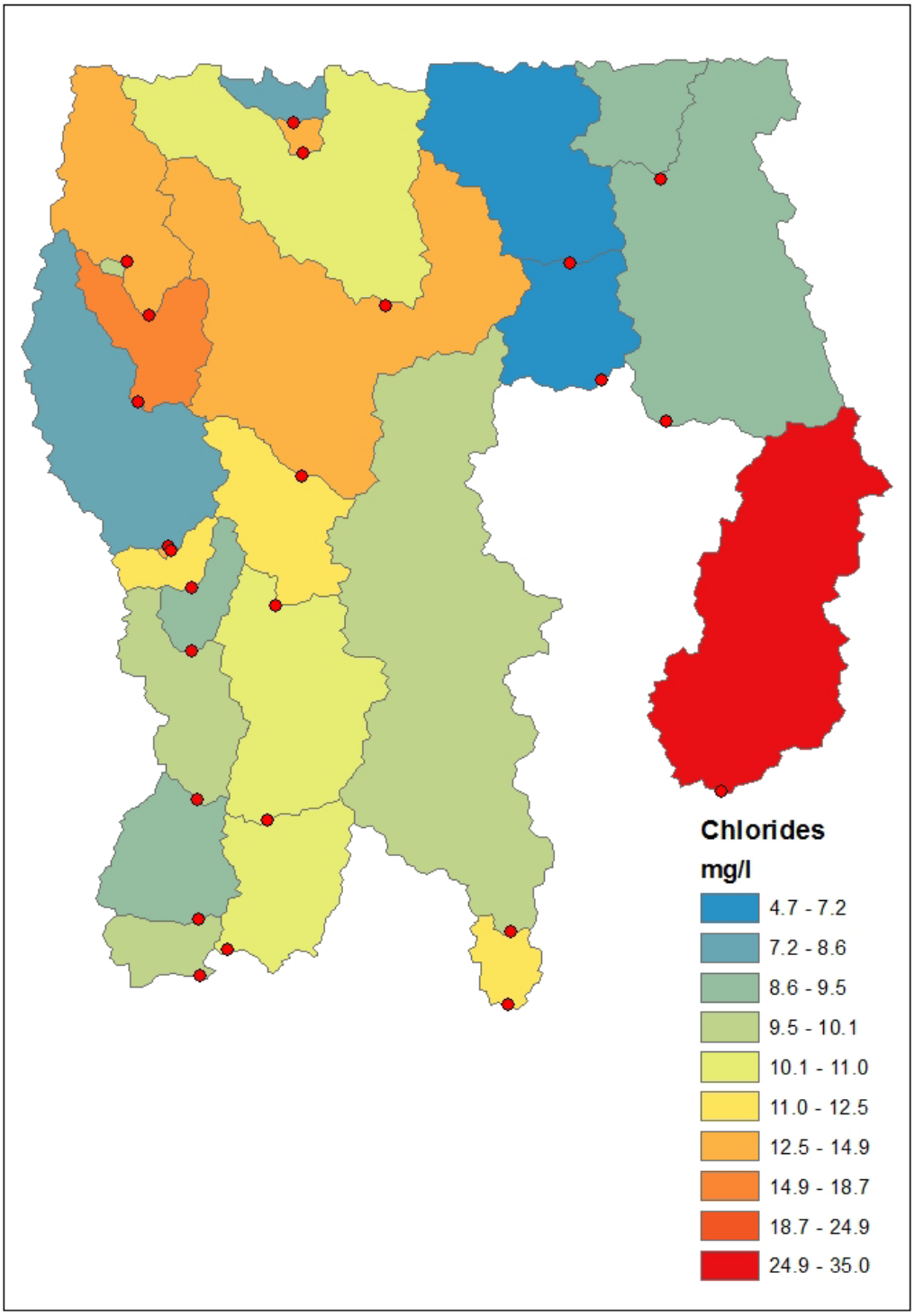

Figure 3-6: Chloride subwatershed interpolations

3.3.1 Accuracy of Results

Based on previous understanding of the interpolation models, it was expected that Kriging would receive the best model fits. For specific conductance this was indeed true with Universal Kriging receiving the lowest RMSE followed by Ordinary Kriging. However, looking at the chloride interpolations, LPI received a considerably lower RMSE value when compared to the other methods. Lou et al. (2008) did state that even though LPI can sometimes give lower RMSE values, it should be not favored over the more accurate Kriging methods. This is because in data sparse regions, LPI may create misleading results because of the lack of measured values in the search neighborhoods. Figure 3-7 displays the measured values vs. the predicted values for six chloride methods. The graph displays LPI at station # 3 following the rapid trend shift in the measured values. GPI, on the other hand, actually predicted a negative value at one of the stations, illustrating the inaccuracy in fitting data with GPI. Figure 3-8 displays the specific conductance measured vs. predicted values. It is clear see that all of the interpolation methods have difficulty predicting values when abnormal measured values occur. Knowing this information, users can increase the dependability of the models by understanding their limits. Tables 3-2 and 3-3 display the ME and RMSE values for each interpolation method.

Table 3-2: Specific Conductance RMSE and ME optimized model outputs

Specific Conductance	ME	RMSE
Ordinary Kriging	-0.02544345	204.2452983
Simple Kriging	-4.754973131	216.1761583
Universal Kriging	-15.63517289	202.6154323
Inverse Distance Weighting	-23.48720683	217.5846693
Global Polynomial Interpolation	-8.186412695	217.0125589
Local Polynomial Interpolation	-3.997707589	215.8956021

Table 3-3: Chloride RMSE and ME optimized model outputs

Chloride Values	ME	RMSE
Ordinary Kriging	-0.564369	5.725482469
Simple Kriging	-0.126297	5.392064287
Universal Kriging	-0.077669	5.725482469
Inverse Distance Weighting	-0.625106	5.680884523
Global Polynomial Interpolation	-0.990216	5.649918887
Local Polynomial Interpolation	-0.331895	3.288882366

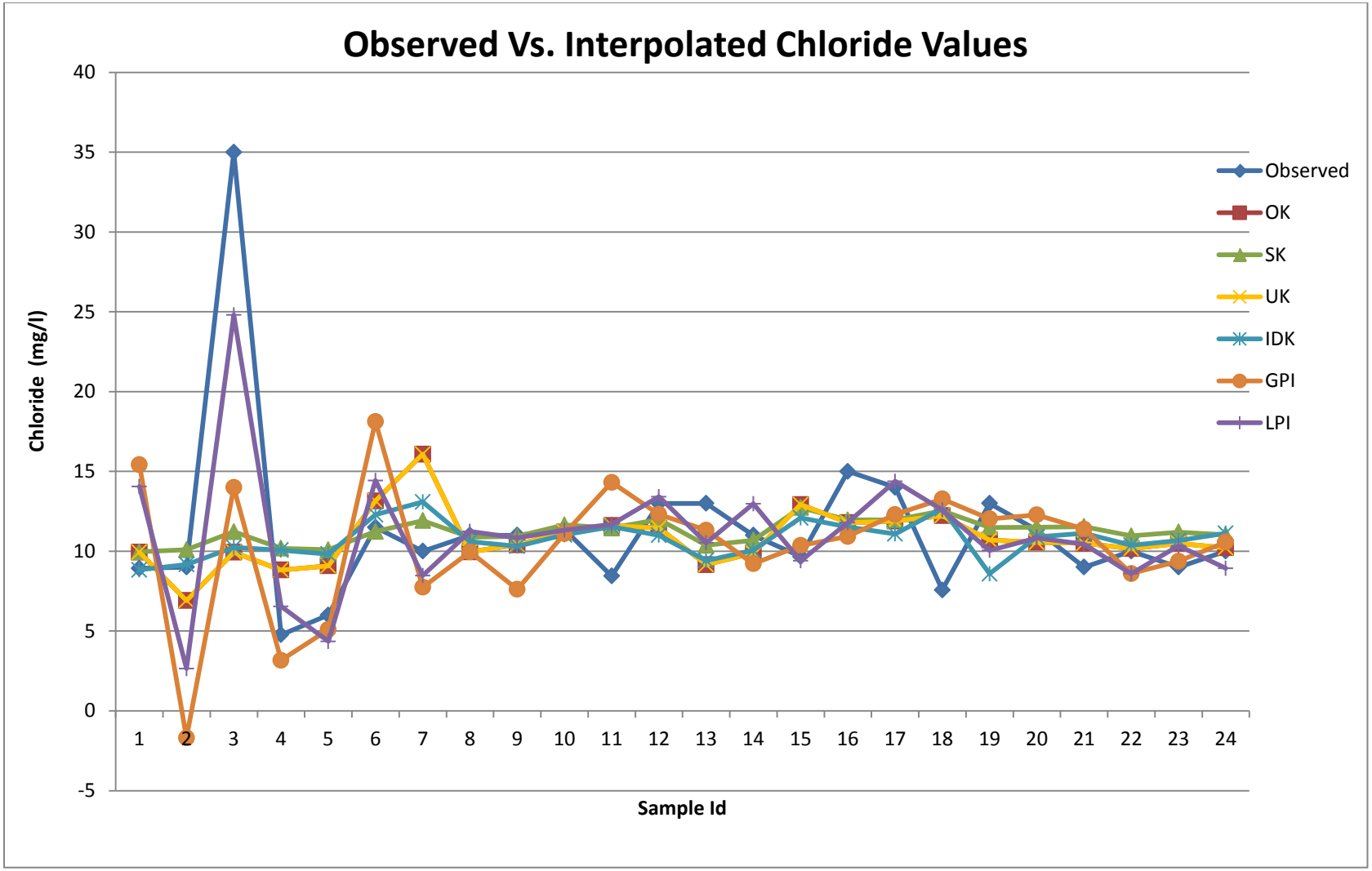

Figure 3-1: Chloride Observed vs. Interpolated values

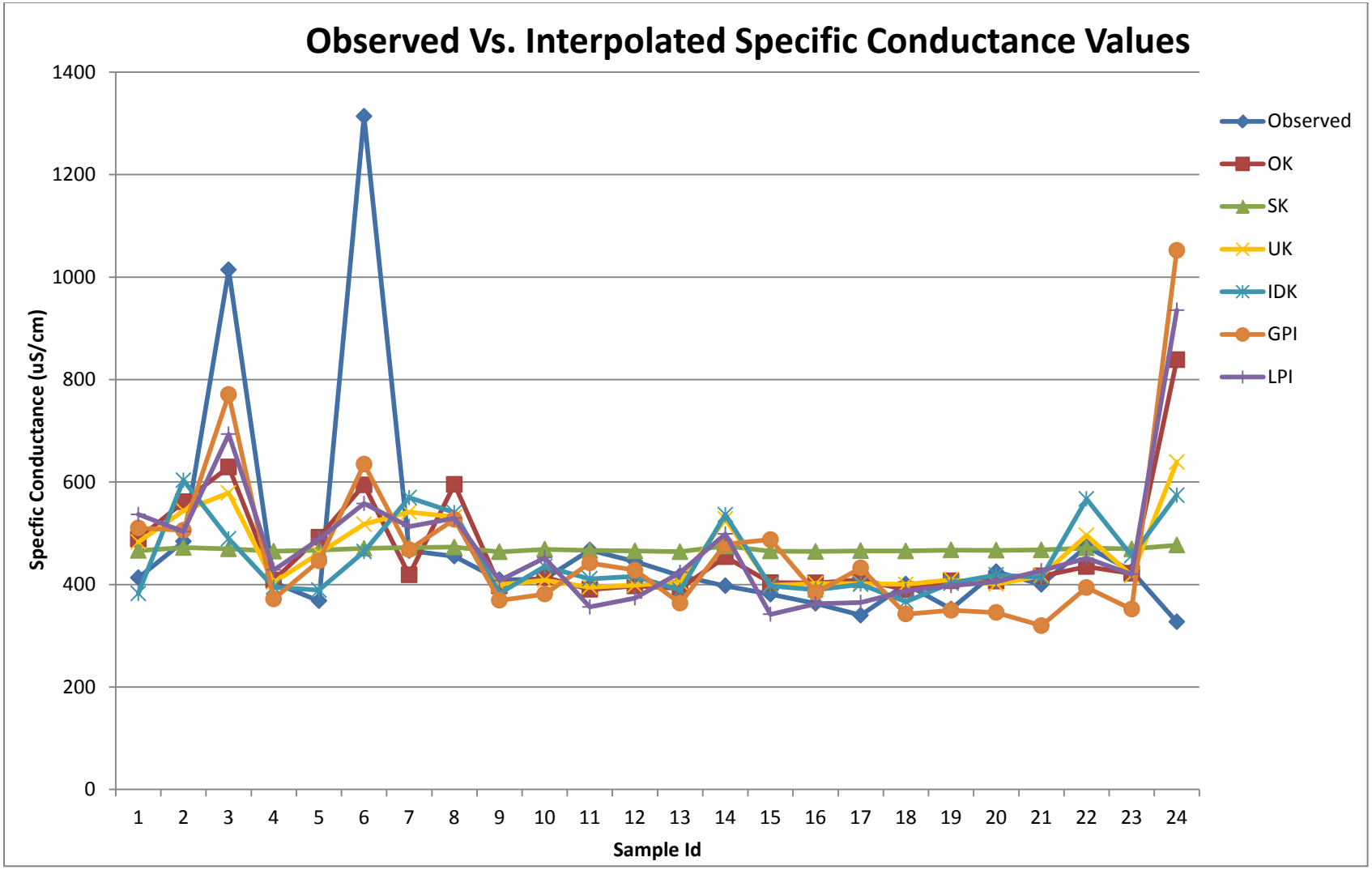

Figure 3-2: Specific Conductance Observed vs. Interpolated values

3.4 Discussion

Spatial variability is at the core of most ecological sampling studies. To add to the variability most sampling areas tend to be small and sometimes randomly placed. Robertson (1987) presented an overview of how geostatistics methods can help scientists interpolate these values to larger areas with a certain degree of error. (Robertson 1987) There are a variety of geostatisical methods used to interpolate the data and each dataset responds differently to the known methods. Geostatistics have been used to predict things with extreme spatial variability from snow distribution in mountains to annual precipitation (Erxleben et al. 2002; Vicente-Serrano et al. 2003). The interpolation model is only as good as the data and method used.

The ArcGIS Geostatisical analyst tool gives users validation and prediction errors like RMSE and ME which can be used to validate the selection of the most optimal model. The model cross validation outputs vary from method to method with the kriging output giving additional standardized errors which demonstrates over- and underestimation of the prediction values. The user must understand how to select the best interpolation model through their expertise with the software and knowledge of the environmental data. Users must decide if they want to remove outliers from the methods and if the dataset is sufficient to provide accurate interpolations. Some of the methods can only undergo a qualitative analysis like the subwatershed interpolations. The efficiency of the subwatershed interpolation can be improved by reducing the scale of the boundaries and incorporating more data points. Regardless, the constraints of each interpolation must be fully understood before a method is selected.

3.5 Conclusions

This thesis analyzed seven methods of interpolation for both specific conductance and chloride values using ArcGIS and publically available water quality data. The data collected from these databases has a variety of limitations as was duly noted. Both deterministic and geostatisical models proved to be effective in their own right. This depended on data values and the location of the data points. Some of the optimized models attempted to remove data points from the analysis. When dealing with such a small sample set (N=24), removal of data points was not beneficial to the model.

Overall, the goal of these comparisons was to demonstrate the different interpolation methods for water quality data and how to optimize and validate them. No one specific method was selected as the solution because further analysis must be done to validate the models. Additional sampling on both a spatial and temporal scale would likely increase the accuracy of the models. The need for engineers to develop methods to accurately interpolate these water quality parameters is immediate in the face of changing regulations. In answer to the question: “Can this method can be used in a regulatory case to determine stream and wetland connectivity using chemical data?”—the conclusion at this time is that while this method alone may not provide all the answers for stream-wetland connectivity--if it is coupled with other models that look at biological interactions in wetlands, its use in a regulatory setting could be justified. For future projects, researchers may want to look at more water quality parameters other than just chloride and specific conductance. Studies would also

benefit by collecting additional data to couple with online databases in order to paint a more vivid picture.

CHAPTER 4 SUITABILITY ANALYSIS FOR WETLAND MONITORING

4.1 Introduction

Because wetland functions are diverse, their environmental, economic, and social values are important. The ecosystem services provided by wetlands include surface water storage, water quality improvement, flood abatement, erosion control, and recreation. The total ecosystem services of wetlands worldwide is estimated at over 4.9 trillion dollars a year (Costanza et al. 1997). However, the value of wetlands has not always been recognized in the United States. Between the 1950s and the 1970s, there was an annual net loss of 458,000 acres of wetlands, while between 1998 and 2004 there has been an annual net gain of 32,000 acres. The recent amelioration of wetland landscapes can be attributed to states trying to restore and mitigate wetland habitat with management plans (Stave 2001).

A research poll developed by (La Peyre et al. 2001) attempts to assess all 50 states' current wetland management practices and over 75% of the states had established state wetland goals of no net loss of wetlands (La Peyre et al. 2001). Some states like Tennessee established initial goals of determining the quantity and quality of their wetlands. The quantity of wetlands has been documented by the Fish and Wildlife Service's (FWS) National Wetlands Inventory (NWI) but the quality of wetlands remains unknown. If a wetland's health degrades and loses its ecosystem services, this can have the same impact as physically removing the wetland (Davis and Froend 1999). To

understand the health of a wetland, a proper monitoring program is needed to collect biological and chemical data for continued analysis (Cvetkovic and Chow-Fraser 2011).

The Environmental Protection Agency (EPA) specifies that current state water monitoring data is insufficient. Available data are not able to support the full range of water programs, to make statistically valid statements about the condition of waters, or to track significant changes over time. The EPA's response to this lack of data is the Clean Water Act Section 106 Monitoring Initiative to fund the enhancement of state monitoring programs through the development and implementation of monitoring strategies and through the formation of state-EPA partnerships to survey the nation's waters (EPA 2008). In 2005, the US federal government spent $18.5 million in state grants for this program. Currently, all the states are participating in the program in order to increase their monitoring capabilities. The goal is increased monitoring to accomplish our nation's directives under the Clean Water Act.

Initial surface water monitoring objectives were mainly focused on rivers and streams regulated under the Clean Water Act. These regulated waters are known as "Waters of the US." EPA has developed a specific criteria to determine waters of the US, such as if the waters are used for navigation. Most wetlands are not considered "Waters of the US" unless a clear connection is found between the wetland and the regulated waters. This connection can be a hydrological, chemical, or biological connection that must be demonstrated by the party attempting to demonstrate a connection (Leibowitz et al. 2008). The EPA has designed the National Aquatic Resource Surveys to collect data

on lakes, rivers, streams, coastal waters, and most recently wetlands. The recent need for wetland monitoring comes from recognizing the demand for accurate wetland data in order to properly manage wetlands.

The EPA released a 2006 report titled "Elements of a State Water Monitoring and Assessment Program for Wetlands." The document was designed to help states develop wetland monitoring by following EPA suggestions in areas such as: monitoring objectives, monitoring design, and core-supplemental indicators and methods. The document made it clear that states need help developing wetland monitoring programs due to the lack of current management tools and strategies. A wetland manager may need help in determining monitoring objectives which should be in accordance with the current initiatives being promoted on the state and federal level. This requires a manager to use the proper scientific tools and personal judgment to select the objectives. Some common monitoring objectives are (Park et al. 2006):

- To understand long and/or short-term trends of water quality variations;
- To monitor violations of water quality standards;
- To identify external causes and sources of water quality changes;
- To support the utilization of water resources;
- To examine an acute variation in water quality through a targeted investigation during a given temporal period;
- To estimate pollution loads for Total Maximum Daily Load (TMDL) analysis and reporting;

- To develop water quality modeling to support TMDL and water quality management functions;
- To establish information systems for water resource management.

Selection of water quality monitoring locations should coincide with the objectives of the water quality program. The data from the sites should help develop the water quality database and resolve any pertinent regulatory issues (Harmancioglu and Alpaslan 1994). When monitoring locations are randomly selected without the use of a scientific method, collection of inaccurate or incomplete data can result. The EPA suggests three ways to select wetland monitoring sites(EPA 2008).

Method 1: Examining every wetland within an area of specific interest. The method is classified under special area management plans, which are employed to identify significant wetlands in need of specific regulatory attention.

Method 2: Use of a probability approach to understand wetlands within a watershed basin. This means creating a list of all the wetlands; then selecting a spatially distributed and random sample of units from that list and monitoring these sites. This approach should allow an agency to get a cumulative unbiased understanding of the condition of the wetlands.

Method 3: Rely on best professional judgment to target wetland monitoring sites. This approach is used to compare wetlands along a gradient of human disturbance in order to establish a reference wetland condition.

Table 4-1: Pros and Cons of EPA wetland monitoring site selection approaches

	Pros	Cons
Method 1	Locations are specific to current needs, helps policy makers reach decisions, saves money by only sampling when there is a need	Doesn't give a holistic picture of an ecosystem, not suited for long-term planning, important wetlands may be left out if they do not currently have a regulatory need
Method 2	Examines data on a watershed level, represents all different types of wetlands; simple approach	Wetlands selected at random and are not optimal wetland locations; the selection of wetlands should be specific to optimize data--not "unbiased"
Method 3	Uses wetland manager expertise to select sites, looks at wetlands in different levels of human disturbance	Does not use a scientific method to select wetland locations. Selected wetlands are based on personal knowledge of wetlands which can be incorrect

These methods on their own may not address all the objectives in a wetland monitoring plan. One may need additional scientific tools to select wetland locations and then have a wetland manager use his expertise to go out and evaluate the locations. The EPA's National Wetland Condition and Assessment (NWCA) survey sampled over 1,100 wetlands nationwide and attempted to incorporate random sampling with the help of state wetland experts to examine site locations. The EPA initiated the NWCA in 2007 and is scheduled to release a report in 2013 (EPA 2012). The program's goals are to produce a national report to describe the quality of the nation's wetlands while helping states implement monitoring and assessment programs to guide policy making. The program seeks to answer questions about the quality of wetlands and how they degrade over time. The program selected wetland sites to sample across

the nation. The selection was processed using a probability based network that selects wetlands in each state. The designers used stratification and unequal probability to select the sites. The wetlands were stratified by states, while seven wetland types were used for the unequal probability (Estuarine Intertidal Emergent (E2EM), Estuarine Intertidal Scrub-Shrub (E2SS), Palustrine Emergent (PEM), Palustrine Scrub-Shrub (PSS), Palustrine Forested (PFO), Pf Palustrine farmed and Palustrine Unconsolidated Bottom/Palustrine Aquatic Bed (PUBPAB)). This allows the survey to represent the different wetland types and locations in the nation without a bias. The project accessed an extremely large dataset across multiple eco-regions which may not be ideal for a state monitoring plan. The survey was selected to be random in order to remove any bias in the selection of the wetlands. However a state program would want to select specific wetlands in order to follow their monitoring objectives.

All of the above EPA initiatives point to a caveat in wetland management plans that indicate a lack of tools to help locate monitoring sites. The states may have difficulty developing a methodology for selecting monitoring locations. While EPA offers suggestions to the NWCA, the suggested plans may not be easily scaled down to the state level. This thesis presents methodology for selecting suitable wetland monitoring locations using Geographic Information Systems (GIS). A relative site-suitability analysis was used to determine monitoring locations in sub-watershed locations. Site suitability involved the overlay of multiple GIS layers to create a composite representation of suitability sites to accomplish a goal. Missouri's Goose-Lake West Yellow Creek

subwatershed was selected as a pilot location. The site has been extensively used when officials are locating sites for wetland mitigation or restoration (Palmeri and Trepel 2002; Van Lonkhuyzen et al. 2004; White and Fennessy 2005) but not necessarily for wetland sampling locations.

4.2 Methods

To construct a GIS suitability analysis, layers of geographic data relevant to suitability needed to be identified. In order to define the extent of the study area, Hydrologic Unit Codes (HUC), which pertain to watershed and subwatersheds and were developed by the United States Department of Agriculture (USDA) was used. HUCs are arranged from largest to smallest based on unique 8-digit, 10-digit and 12-digit codes, respectively. The large 8-digit HUC (Lower Grand Watershed) was set as the primary extent of the study area, which was further refined to a 12-digit HUC within the 8-digit as the final study area. The 12-digit HUC (Goose Lake-West Yellow Creek) was used in order to investigate the hydrologic and physical properties of the study area with more precision (Chang 2009). This study area in northern Missouri exemplifies what a wetland manager looks for when trying to optimize wetland monitoring locations.

Site Location

Selection of the study area was based on suggestions from the EPA and the Missouri Department of Natural Resources (MODNR) in order to complement state initiatives. Both agencies had expressed interest in the Lower Grand Watershed, 8-digit HUC 10280103, illustrated in Figure 4-1. MODNR has designated the Lower Grand as

one of three pilot watersheds being studied in order to improve management and decision-making tools at a watershed level (MODNR 2012). The watershed is approximately 1.5 million acres and is mostly located in Missouri, with a small portion in Iowa. The Lower Grand watershed drains into the Grand River and the Missouri river following drainage paths formed by Medicine Creek, Locust Creek, and West Yellow Creek. The Lower Grand contains 303(d) listed impaired water bodies which include the three creeks and a portion of the Grand River, Salt Creek, and Marceline New Lake. The subwatershed selected was Goose Lake-West Yellow Creek, whose 12-digit HUC is 102801031206. The area is approximately 14,000 acres and is located in Linn County in the Southeastern portion of the Lower Grand HUC (refer to Figure 4-1). It is the drainage basin for a portion of the 303(d) stream, West Yellow Creek, and home to roughly 2000 citizens according to the 2000 US census. This watershed was selected due to currently available data in the area which is in the process of being developed for the whole state of Missouri. The watershed represents an interesting study site with a larger percentage of wetlands and urban area when compared to the Lower Grand HUC. Figure 4-2 displays the percentage of land use comparisons between the Lower Grand Watershed and the Goose Lake- West Yellow Creek sub-watershed.

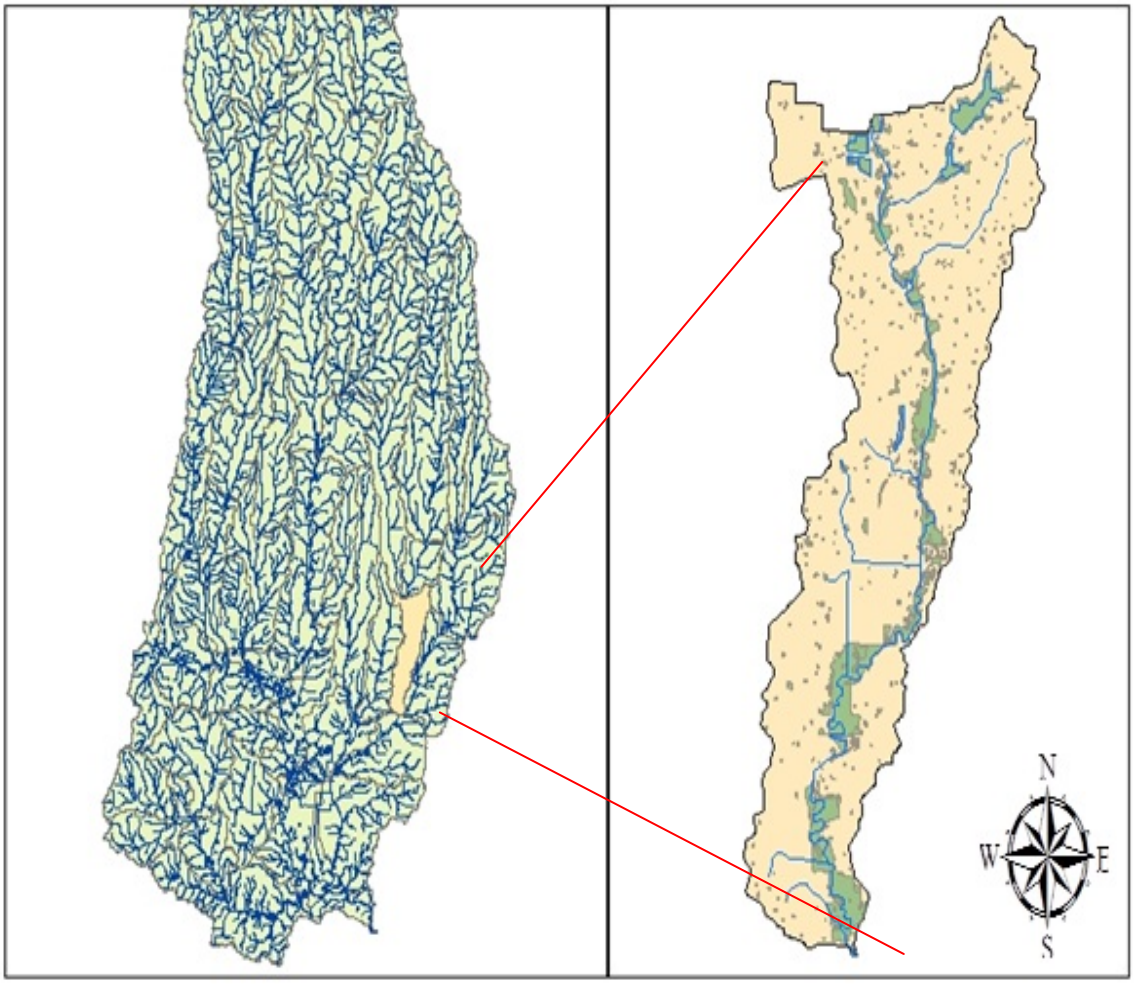

Figure 4-1: The left is the Lower Grand Watershed with streams in blue and Goose Lake sub-watershed in beige. On the right is the Goose Lake Subwatershed with green polygons representing wetlands (Source ArcGIS 10)

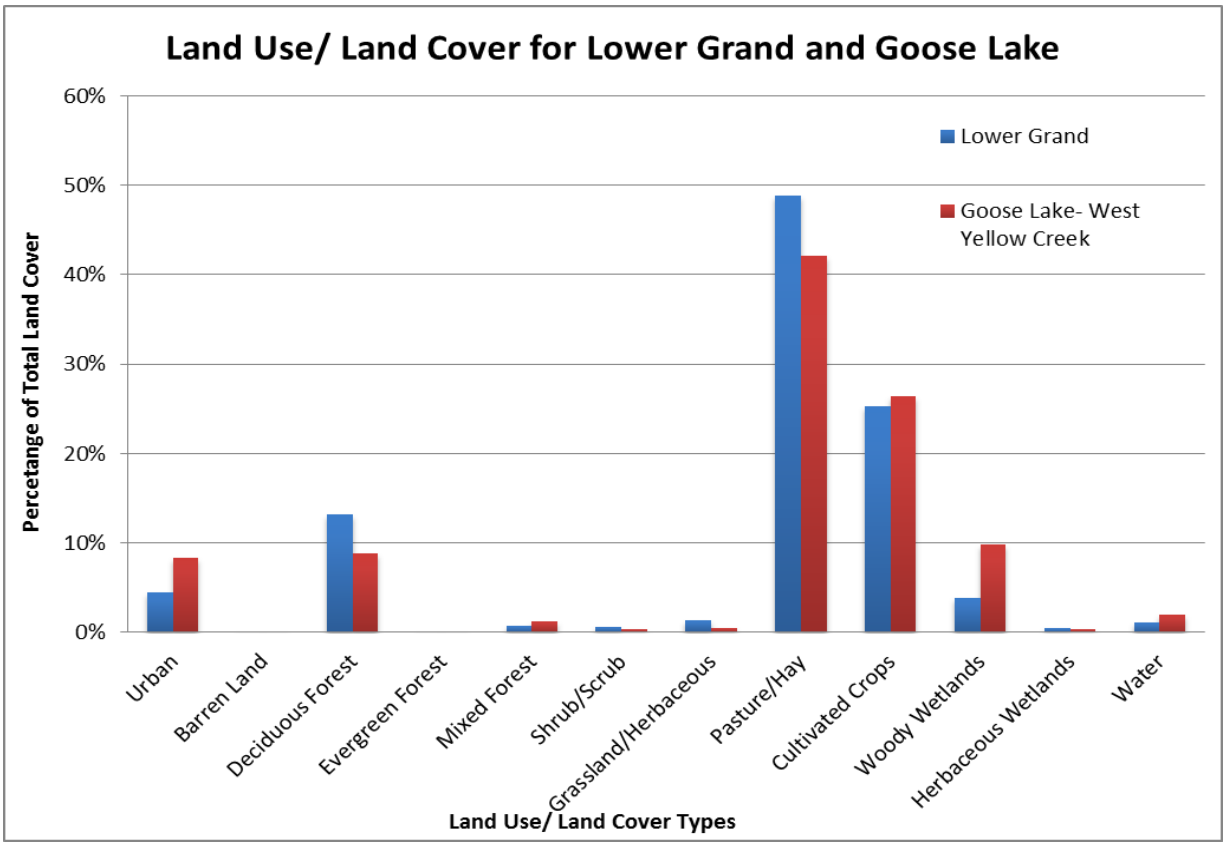

Figure 4-2: Land Use/ Land Cover comparisons Lower Grand Watershed – Goose Lake Subwatershed

4.2.1 Phase 1: Suitability Analysis GIS Layers

The goal is to develop a GIS methodology for selecting the most suitable wetland monitoring location using specific variables selected by the user. The suitability analysis will follow the multiple criteria evaluation (MCE) method which is a decision-making concept that incorporates several different variables which are expressed by weights, values or intensities of preference (Ceballos-Silva and López-Blanco 2003). It allows a user to develop a range of alternative options based on user assigned critical variables and weights. The GIS approach creates specific layers that illustrate a variable's utility for an activity by assigning a score ranging from 'most' suitable to 'least' suitable. In this

thesis, each layer of geographic information is in raster format. The raster calculator function of ArcGIS can then be used to combine each value in the overlaying raster cells in order to determine which cells have the highest scores and thus highest suitability for some activity (Wayne 2003).

This stage of the thesis is comprised of two phases of analysis. The first phase involves a suitability analysis using physical and topographical parameters in the landscape. These parameters demonstrate different environmental indicators that may be ideal for a water quality sampling location based on hydrology and land use/ land type. The second phase will take an additional look at proximity of neighboring wetlands to each monitoring location as well surrounding land use impacts. Geographic layers used in the Phase 1 calculation are described below.

- Historical conditions are used to locate sites which are prime wetland habitat (historic wetland areas).
- Wetland proximity to streams is factored in by creating multiple buffers at 100 m increments from the stream.
- Drainage path/flow accumulation illustrates the inflow of water into wetland locations.
- Hydric soils are used to demonstrate potential wetland health because it is one of the requirements for a functioning wetland.
- LU/LC (land use/land cover) is used to also show potential health of a wetland and its ability to stay connected with surrounding water bodies.

- Hotspot analysis will be used to determine spatial hotspots of large wetlands which could be a sink for water quality data in the surrounding areas.

Each layer will be further discussed in detail below with Figures (4-10) displaying the geographic layers and Table 4-2 displaying suitability values for each layer.

Base geographic indicators of suitability

In this thesis the base geographic indicators of suitability include: 10 meter DEM raster data, National Wetland Inventory (NWI) vector data, river/stream vector data, and 12 digit HUC layer vector data. These layers were all projected in NAD 1983 UTM Zone 15N, which is an accurate projection for Missouri. Each base layer except the stream layer was obtained from the MSDIS (Missouri Spatial Data Information Service) website. The stream layer was obtained from the United States Geological Services (USGS) National Hydrography vector data set at a 1:24,000 scale. The USGS National hydrography data is available for the continental US including Hawaii, and Alaska.

Historic Wetlands

Understanding the historic changes and past anthropogenic impacts on a landscape can be an indicator of the functionality of the surrounding ecosystem (Rhemtulla and Mladenoff 2007). These changes can be the underlying reason why a specific ecosystem has flourished or perished when it was introduced into the landscape. In the case of wetlands, the historic land use can be an indicator of how suitable the land is for functioning and interacting with surrounding wetlands that have been in the area for centuries. Historic wetland sites have been used as indicators for

potential locations for successful re-establishment of wetlands (Stein et al. 2010). In this study, historic wetland sites are an indicator of present and future functionality of a wetland and as a gauge of possible past wetland hydrologic interactions or connections. The historic wetlands layer (shown in Figure 4-4) was obtained from the MSDIS website in the format of vector polygons.

Flooding Frequency

Floodplains are areas of low lying land adjacent to water bodies which allow periodic inundation of the land (Tockner and Stanford 2002). Floodplains are the ecotones between terrestrial and aquatic regions much like wetlands are. Wetlands are synonymous with floodplains and require the hydrologic connections provided by flooding. Flooding frequency determines the vegetation and landscape heterogeneity (Toner and Keddy 1997). The spatial and temporal range of flooding determines the fauna and animal life required for a healthy wetland (Tockner and Stanford 2002). Initially, the 100 year floodplain according to Federal Emergency Management Agency (FEMA) was to be used; however, flooding frequency maps (provided in Figure 4-5) obtained from (SURGO/CARES) allow for better elucidation of flooding impacts on wetlands. The data come in vector format and are identified with the frequency of the flooding.

Buffer of 1st order streams

Wetlands and streams are important parts of a heterogeneous landscape especially when they are connected hydrologically. Wetlands at headwaters are known

to increase water quality of downstream segments. However, according to Whingham et al. (1988) wetlands are more hydrologically connected at lower order streams, specifically first order streams. As wetlands move into areas surrounding second to fifth order streams, there is a probability of interaction only during flooding events (Whigham et al. 1988). Stream order was determined using the Strahler stream order methodology (Strahler 1952). This system uses a hierarchy of tributaries to determine the stream order of each segment. For example, a headwater stream with no upstream tributaries is considered a first order stream, and if two first order streams intersect, it creates a second order stream. However, if a second and a first order stream intersect, it would still become a second order stream and not a third order (Strahler 1952). Thus, according to Whingham et al. (1988), a higher percentage of water flows through wetlands from lower order streams. This is important because when choosing a monitoring location a suitable wetland would be a wetland that is highly connected hydrologically. Based on the hydrologic connectivity, the wetland would also have a higher probability to increase water quality in surrounding areas downstream. The buffer tool in ArcGIS spatial analyst is used to create specific buffers at the distance of 50, 100, 150, and 200 meters (m) for each first order stream. The buffers of the first order stream segments can then be used to determine if wetlands are encompassed in any of the buffer zones displayed in Figure 4-6.

Hydric soils

Hydric soils are formed under conditions of inundation or long term saturation which develop anaerobic conditions in the upper part of the soil layer. Hydric soils are one of the parameters required for functioning wetland landscapes. Wetland hydrology requires the land to be inundated for 5% of the growing season with a frequency of at least 5 of the 10 years (Clausnitzer et al. 2003). Wetland regulators use hydric soils as a parameter to delineate wetland areas. The hydric soils layer is included in this thesis to demonstrate a wetland's potential health and hydrologic connection. Here it is assumed that a wetland not located in hydric soils would receive less precipitation or was more disconnected from surrounding water sources. The hydric soils layer downloaded from (SURGO/CARES) is in the format of vector polygons and is illustrated in Figure 4-7.

Land Use/ Land Cover (LU/LC)

LU/LC data has been utilized extensively in environmental research. The data displays the predominant land cover in a 30 m x 30 m raster cell. LU/LC determines the hydrology, connectivity, and health of a wetland and surrounding water bodies (Owen 1999). The data can also demonstrate anthropogenic impacts on an environment, and the heterogeneity or homogeneity of a landscape. In each of the wetland mitigation suitability analysis articles, LU/LC was commonly used (Palmeri and Trepel 2002; Van Lonkhuyzen et al. 2004; White and Fennessy 2005). The importance of LU/LC is relevant to water quality and is included in the suitability analysis as one of the indicators. The layer is available from MSDIS as a raster dataset as illustrated in Figure 4-8.

Flow accumulation layers

Flow accumulation is a hydrology geoprocessing methodology that allows a user to render the contributing drainage upstream of a raster cell. For example, for equal sized raster cells, there is one (1) downstream raster cell with five (5) upstream cells. If a value of 1 is given to the cells that flow into that downstream raster, the downstream cell would be given the value of 5 corresponding to the summation of the upstream raster cells. This process is done with digital elevation model (DEM) data obtained from MSDIS. This methodology requires first filling in the DEM in order to remove any imperfections in the model. The DEM layer must not have depressions within it in order to remove the chance of the hydrology following incorrect paths. If there is an internal drainage area on the DEM it will not execute properly. Next, the ArcGIS hydrology tool is used to determine flow direction. Flow direction is needed to determine the destination of the surface flow based on topography. The neighboring of cells in the DEM is used to determine each cell's steepest adjacent cell. The tool is able to calculate the flow in each cell by accumulating the cells that flow into each down slope cell and then displaying that as a final raster-based layer as is displayed in Figure 4-9. Table 4-2 demonstrates the different values for suitability, 10 is given to cells with the most contributing upland cells and 1 to the pixels with the least.

Hotspot analysis

Integrating geostastical tools in ArcGIS allows a user to develop spatial statistics on a study area. Spatial clustering or hotspot analysis is used in this thesis to evaluate

the location of clusters of large wetlands. Hot Spot Analysis (Getis-Ord Gi*) has been used in studies looking at biological or environmental hotspots. Before running the analysis tool the optimal fixed distance had to be determined using GIS Spatial Autocorrelation (Global Moran's I). This tool gives an output of the Moran's Index, Expected Index, Variance, z-score, and p-value. While this tool can be used to evaluate if "you can reject the null hypothesis," in this study it is used to determine the optimal distance based on the z-score output. While the P score was indeed significantly low indicating that clustering was occurring it was not used as a variable to determine optimal distance. The Z-score which indicates standard deviation of the values was plotted at different distances in order to determine different peaks in the data. These peaks indicate optimal distances that can be used for the hotspot analysis in order to represent the data correctly. The hotspot tool (Getis-Ord Gi*) also gives an output Z value which is used to determine the hot or cold spots. A large Z value and a small p-value indicate a spatial clustering of large values, which in our case are large wetlands. A low Z value indicates a clustering of small wetlands. Smaller wetlands are usually not as connected by surface hydrology as is the case with larger wetlands. This smaller wetlands' disconnection issue could result in them not being an accurate representation of surrounding wetlands (Tiner 2003). The preference for larger wetlands does not mean that isolated smaller wetlands are not important (Semlitsch and Bodie 1998). This is only being used to optimize monitoring locations that may contain consistent water and biota to sample. Refer to Figure 4-10 for these hotspot analysis results map layer.

Map Algebra

Each of the seven maps is first converted into 10m x 10m raster cells if they were originally vector data. The LU/LC layer was 30 m x 30 m raster cell; however, when reclassifying the different land types, the cell size was changed to 10 m x 10 m for better precision. This is significant because some wetlands are relativity small and using a finer resolution would allow managers to capture more discrete impacts. Figure 4-3 is an example of how map algebra sums pixel values while calculating for suitability. In the original analysis, map weights were all given an equal value; however, in the second analysis, weights for each layer were allowed to vary. Displayed in Equation 4.1 is the suitability measure used to compute suitability of each pixel. Also refer to Table 4-2 to view the values given for each pixel value in the analysis.

$$Si = \left((W_x * V_x) + (W_y * V_y) + \ ..) \right. \qquad \text{Eq. 4.1}$$

Where

S_i= Suitability Score at pixel

W_x=Weight of $Layer_x$

V_y=Value of pixel in $Layer_y$

Layer 1

10	10	9	5	5
10	10	0	0	5
5	7	0	0	1
5	3	1	1	1
4	4	4	4	0
4	4	2	2	0

Layer 2

3	3	1	7	7
3	1	1	7	7
2	1	1	0	0
2	2	10	10	0
5	5	8	10	0
5	5	8	8	0

Map Algebra tool output

13	13	10	12	12
13	11	1	7	12
7	8	1	0	1
7	5	11	11	1
9	9	12	14	0
9	9	10	10	0

Figure 4-3: GIS Map algebra pixel analysis when all the weights are equal to 1.0

Table 4-2: Suitability Pixel Values

	Scores	Layer Name	Scores
Historic Wetlands		*Land Use*	
Within Historic wetland	10	Wetlands	10
Outside Historic wetland	0	Open Water	9
		Forest	6
Hotspot Analysis		Grassland	5
Standard Deviation > 2.58	10	Cropland	3
Standard Deviation 1.96-2.58	8	Barren or Sparsely Vegetated	1
Standard Deviation 1.65-1.96	6	Urban	0
Standard Deviation -1.65 - 1.65	4		
Standard Deviation -1.96 - -1.65	3	*Hydric Soils*	
Standard Deviation -2.58 - -1.96	1	Hydric Soil	10
Standard Deviation < -2.58	0	Partially Hydric Soil	5
		Not a Hydric Soil	0
Flow Accumulation			
579,015,312- 911,413	10	*Flooding Frequency*	
471,790-579015	9	Frequent	10
407,455-471,790	8	Occasional	6
328,823-407,455	7	Rare	3
260,914-328,823	6	None	0
207,301-260,914	5		
121,521-207,301	4	*First Order Stream Buffers*	
67,909-121,521	3	within 50 meters	10
28,593-67,909	2	within 100 meters	8
7,148-28,593	1	within 150 meters	6
0-7,148	0	within 200 meters	4
		outside 200 meters	0

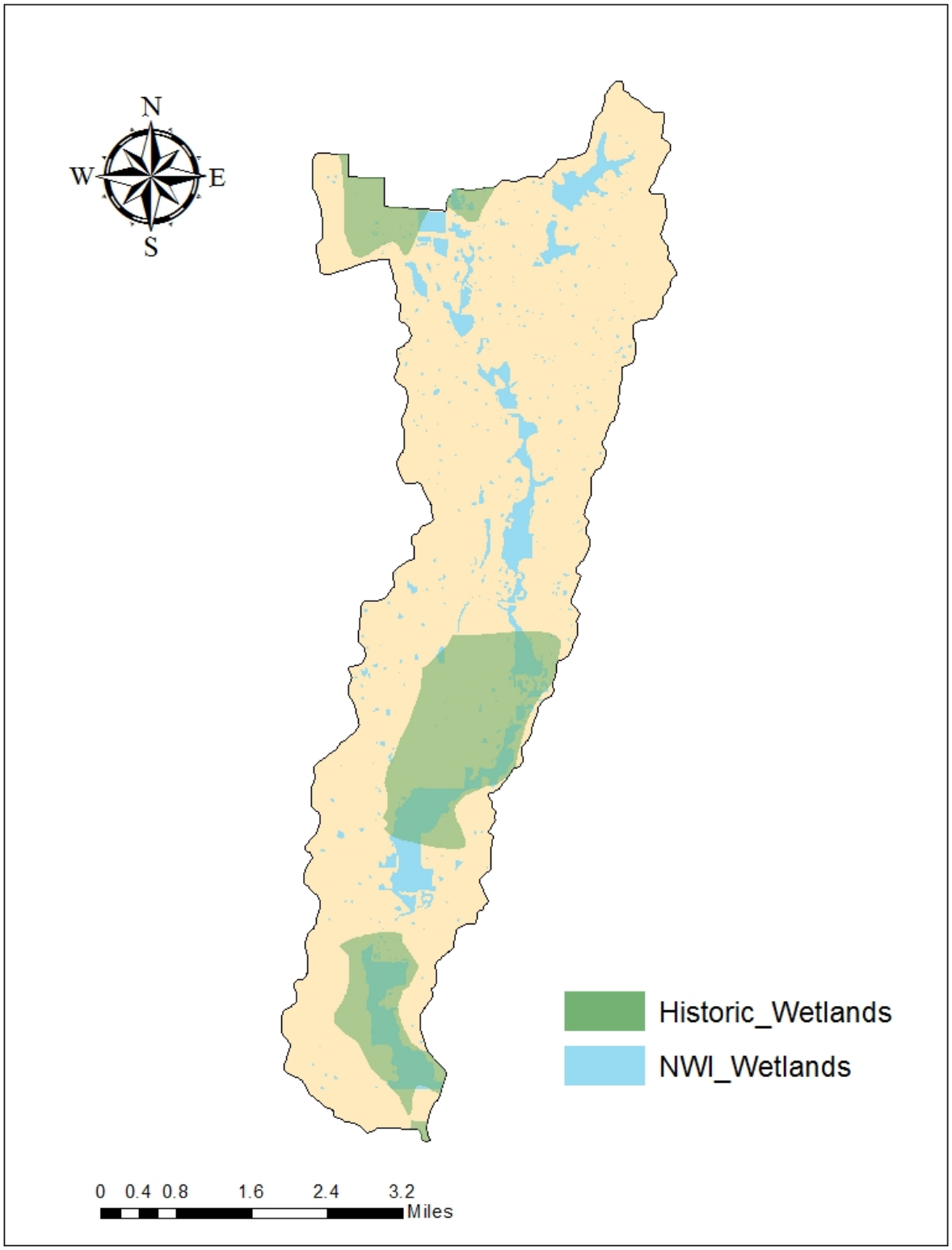

Figure 4-4: Historic Wetlands Layer

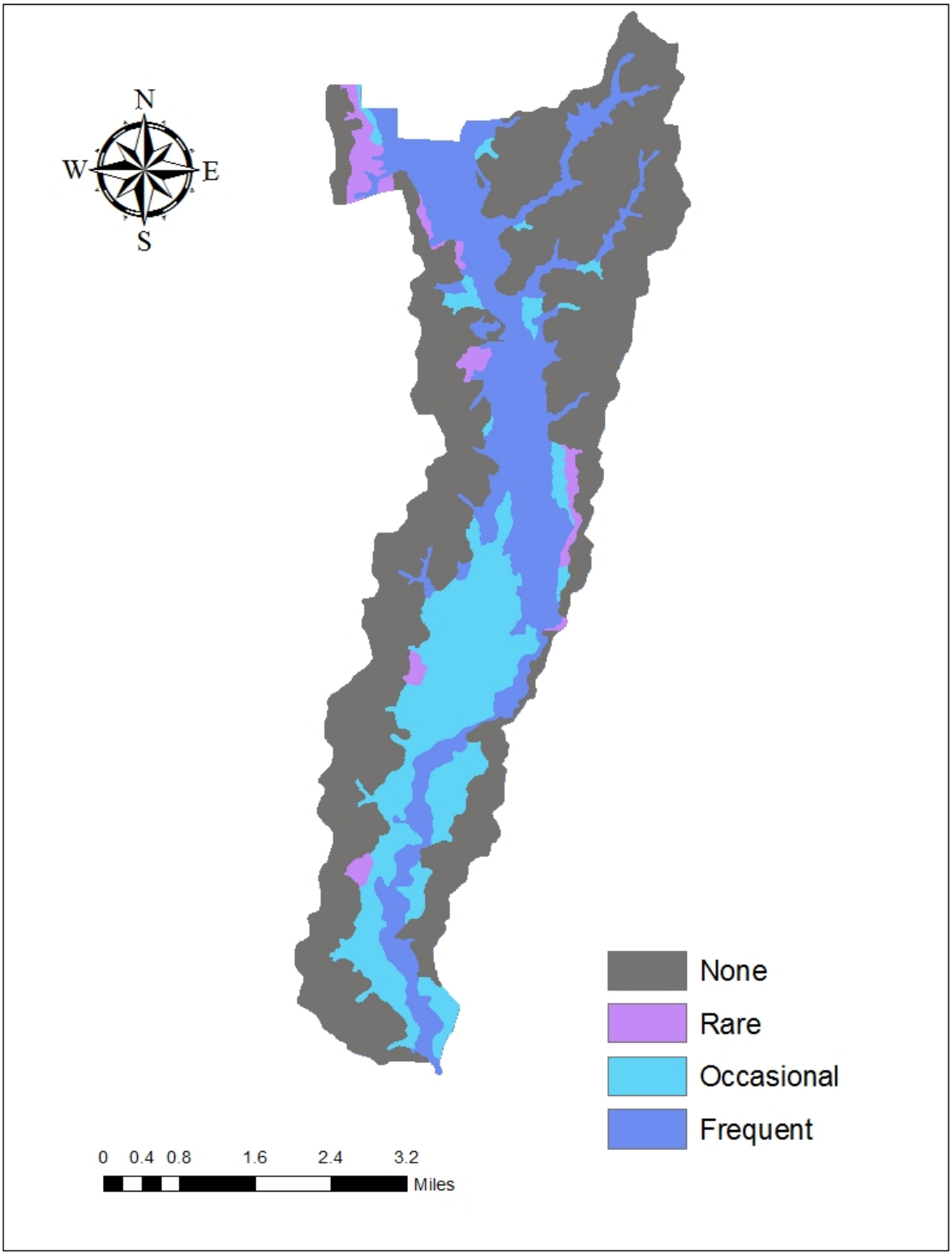

Figure 4-5: Flooding Frequency Layer

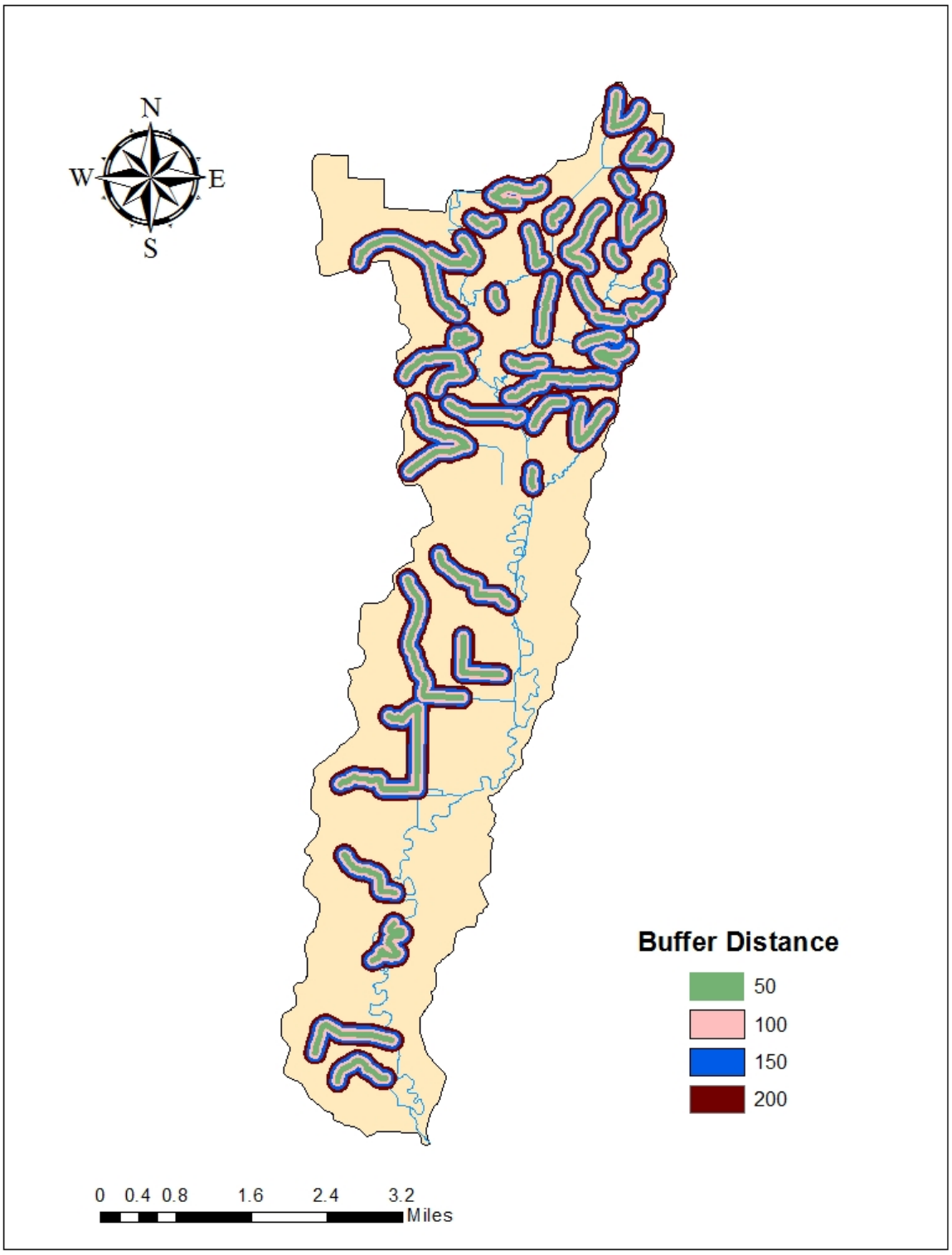

Figure 4-6: First order stream buffer layer

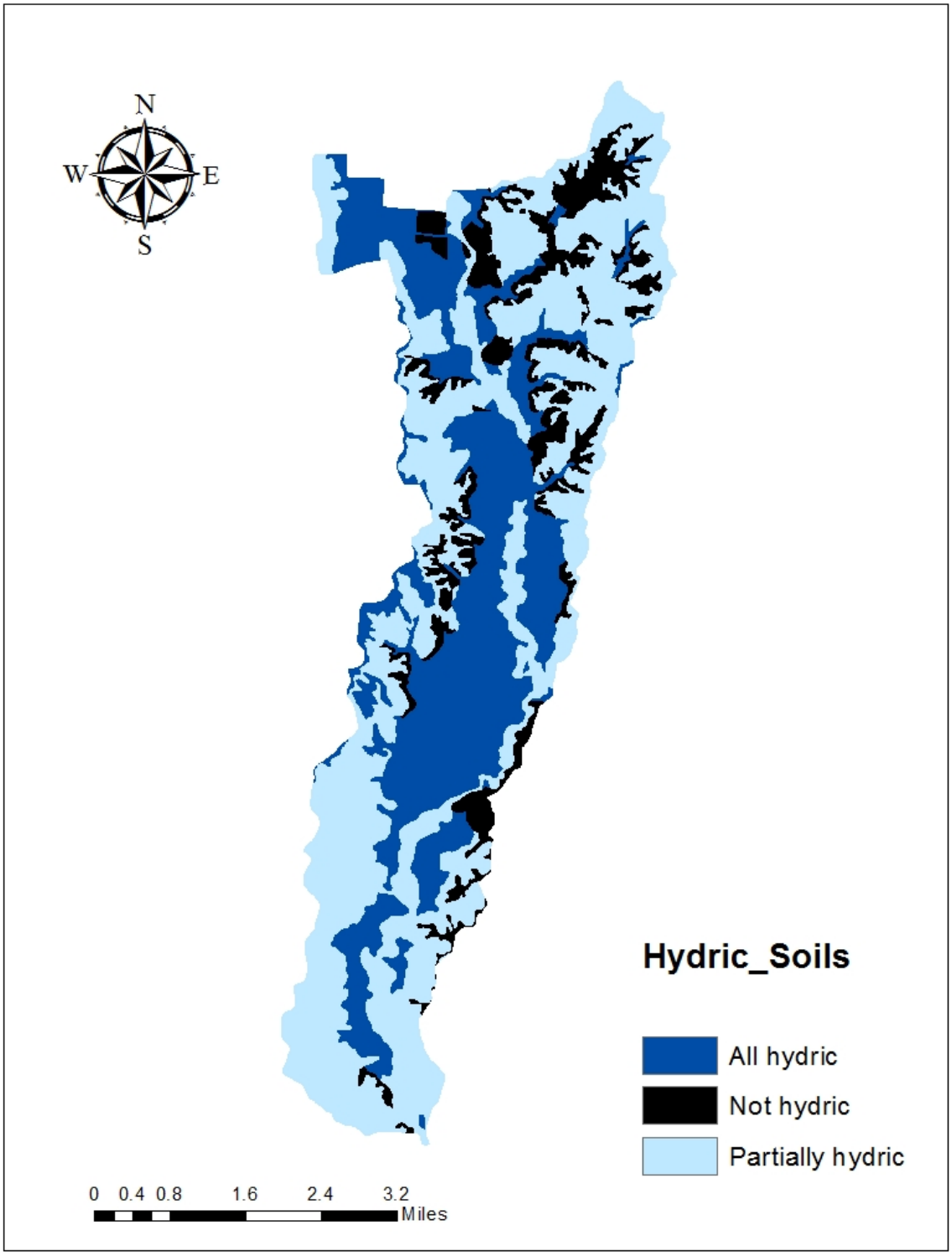

Figure 4-7: Hydric soils layer

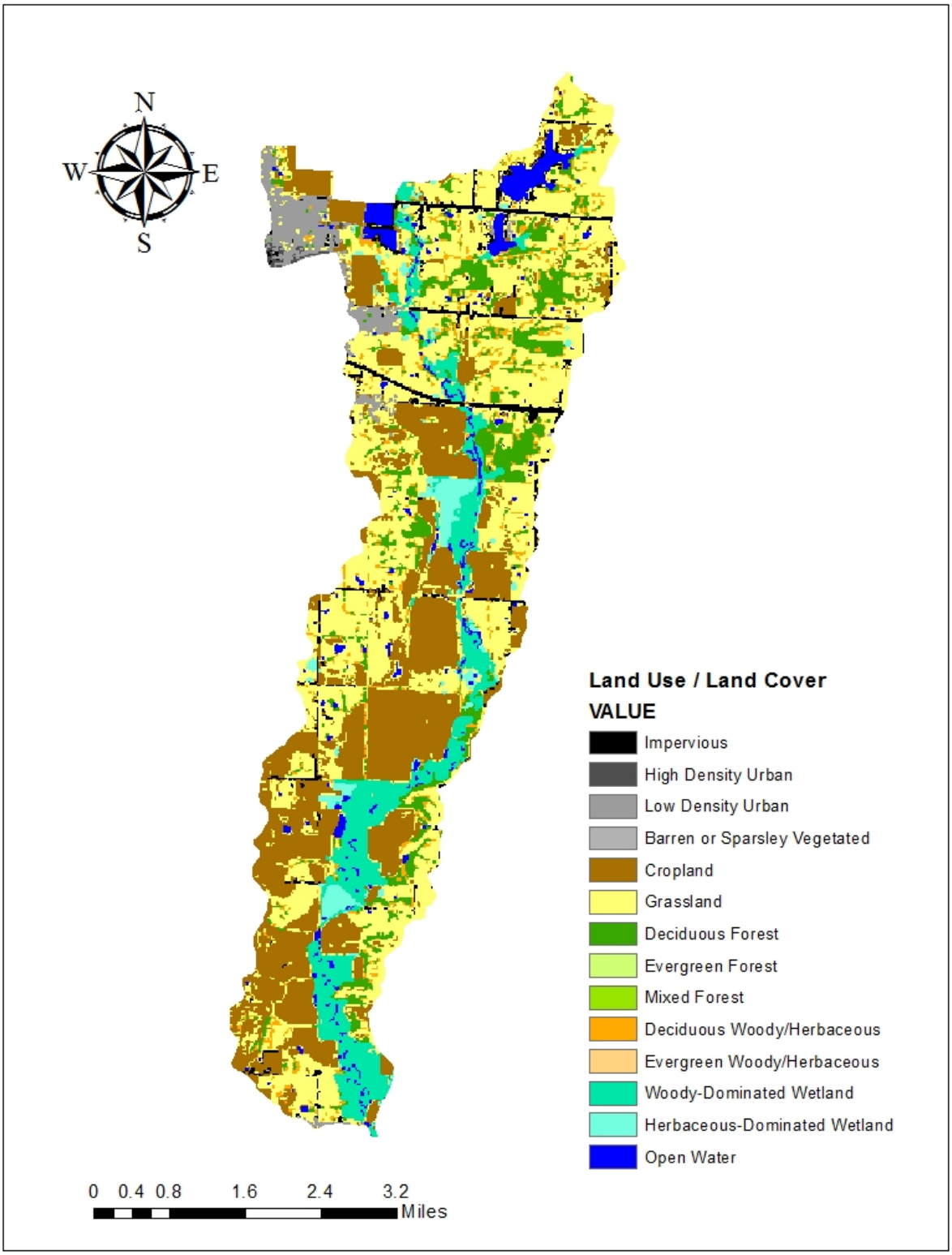

Figure 4-8: Land use / land cover Goose Lake subwatershed

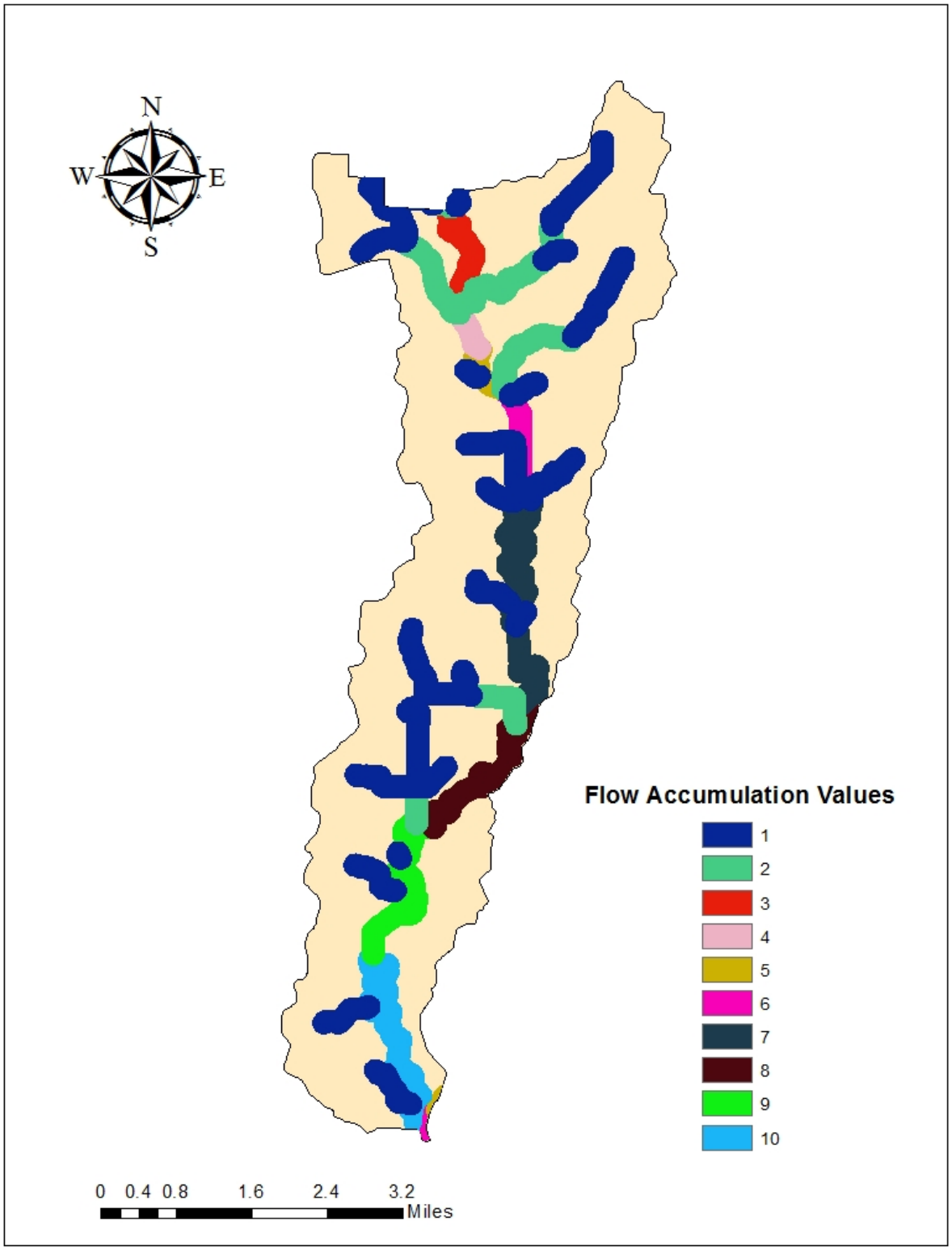

Figure 4-9: Flow accumlation layer

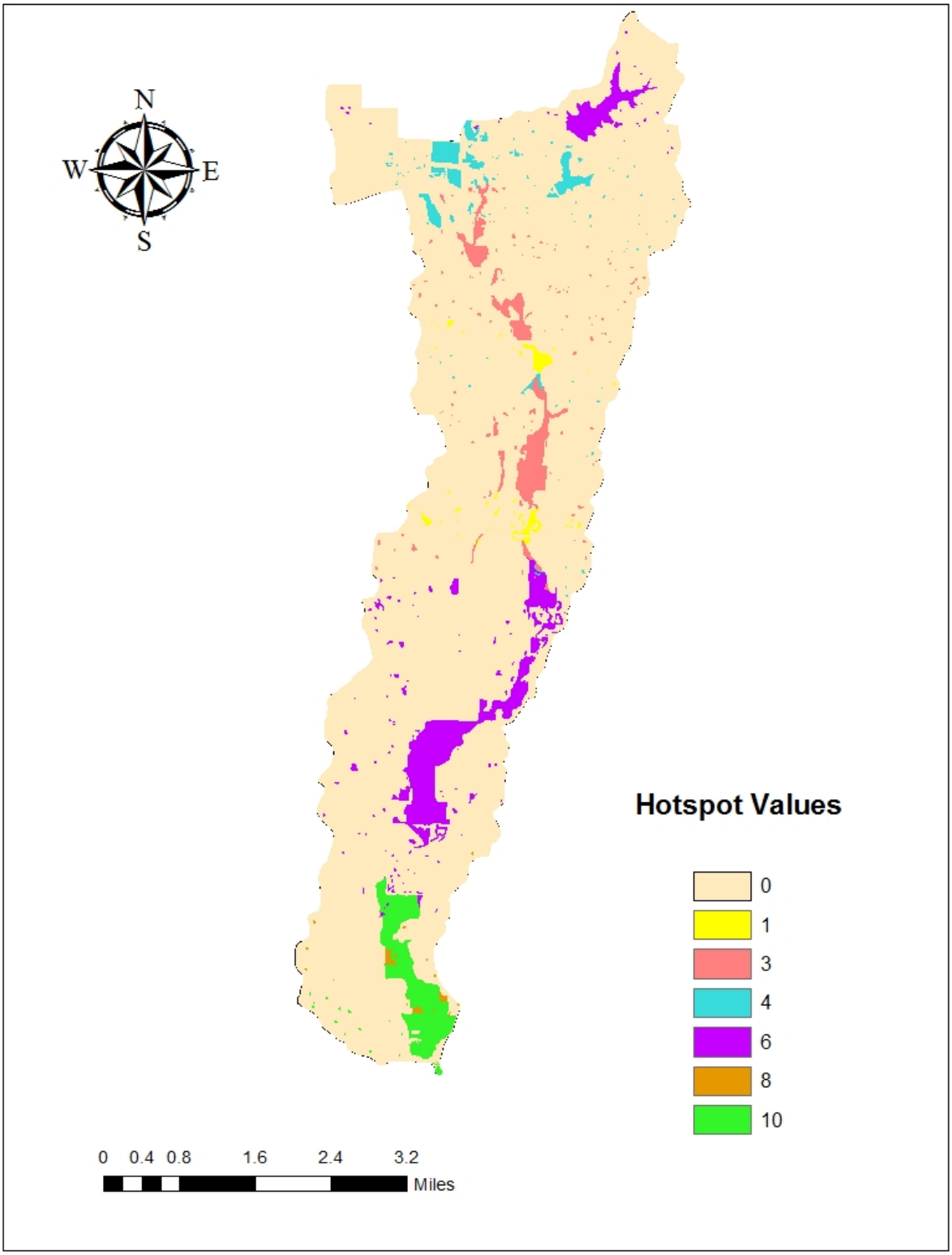

Figure 4-10: Hotspot analysis layer

4.2.2 Phase 2 Layers

Phase 2 combines the Phase 1 analysis of the seven maps with two additional maps: Wetland Neighbor analysis (Figure 4-11) and Majority Land Use Buffer analysis (Figure 4-12). These three maps are combined to create the final suitability layer. The aggregation of the first seven utility layers receives a weight of 0.60; the Wetland Neighbor analysis receives a weight of 0.25, and the Majority Land Use Buffer layer received a weight of 0.15.

Wetland Neighbor Analysis

The first part of phase 2 analyzes the proximity of each wetland to its neighboring wetlands. If a manager can only select one wetland to monitor, he would likely want to select a wetland that is well connected with surrounding wetlands. This would allow the water quality data to represent a larger dataset of potential surrounding influences. One way to look at wetland connections is to examine the distance between the wetland polygons in ArcGIS. Euclidian distance is the straight line distance measured between two points using the Pythagorean formula. Wetlands with more surrounding wetlands can be considered well connected wetlands. Not many dye tracer studies have been conducted on natural wetlands and their potential connections. However, studies using other waterways such as streams have discussed their interconnectivity in catchments through subsurface flows (Bencala 1993). Biological connections can also be considered; for example certain amphibians like Tiger Salamanders have dispersal ranges between 245 m to 2,830 m (Consentino and Philip

2011). This means that any wetland within that given range can be a potential habitat for the Tiger Salamander in its lifetime. Based on previous work done by our research group, a distance of 2,000 m was selected as the distance each wetland will be buffered to determine its potential connections with neighboring wetlands. This distance was selected because 2,000 m falls within the migratory range of the Tiger Salamander's potential habitats.

A member of the research group, Mahabub Alam developed a Python script that can be executed in ArcGIS to determine the number of neighboring wetlands with 2,000 m of wetland. This script was run using the wetlands in our study site and each wetland was assigned a number according to the amount of wetland neighbors within 2,000 m.

Majority Land Use Buffer analysis

"Healthy ecosystems are defined as those with integrity and sustainability, which correlates to limited development in the surrounding landscape and that maintain of ecosystem structure and function even when stressors are present." (Reiss and Brown 2007). Studies have demonstrated that anthropogenic land use changes affect the quality of the ecosystem (Allan et al. 1997). Land use impacts are identified in water bodies by the lack of biodiversity and high levels of chemicals such as chlorides. Knowing the impact of land use on surrounding environments, researchers have developed methods such as the Landscape Development Intensity Index (LDI). LDI attempts to quantify the human disturbance gradient on ecosystems by analyzing land use/land

cover maps using GIS. For the site suitability analysis, LDI was considered as a viable option but in order to keep the analysis simple another option was used.

"Researchers in Florida have shown that areas within 100 m of an isolated wetland or forest patch sufficiently capture the anthropogenic impacts caused by land use." (Brown and Vivas 2005). Thus, a 100 m buffer was created around each wetland in order to determine what type of land use may be impacting the wetland as a possible monitoring location. The land use/ land cover maps from Missouri were then reclassified and assigned utility scores based on a score of 1-10 similar to the LDI analysis (Brown and Vivas 2005). Then using the spatial analyst tools in ArcGIS, zonal statistics were run for each wetland buffer using the land use raster pixels. The buffer was then given a value based on the majority of encompassing land use pixel in each wetland buffer.

The land use buffer analysis is a good way for managers to assess surrounding impacts on their monitoring sites. If a wetland is adjacent to a road or crop field, the manager can determine this using GIS and appropriate map layers. This ability to determine the nature of surrounding land use with GIS gives managers better discernment in the selection of their wetland monitoring locations.

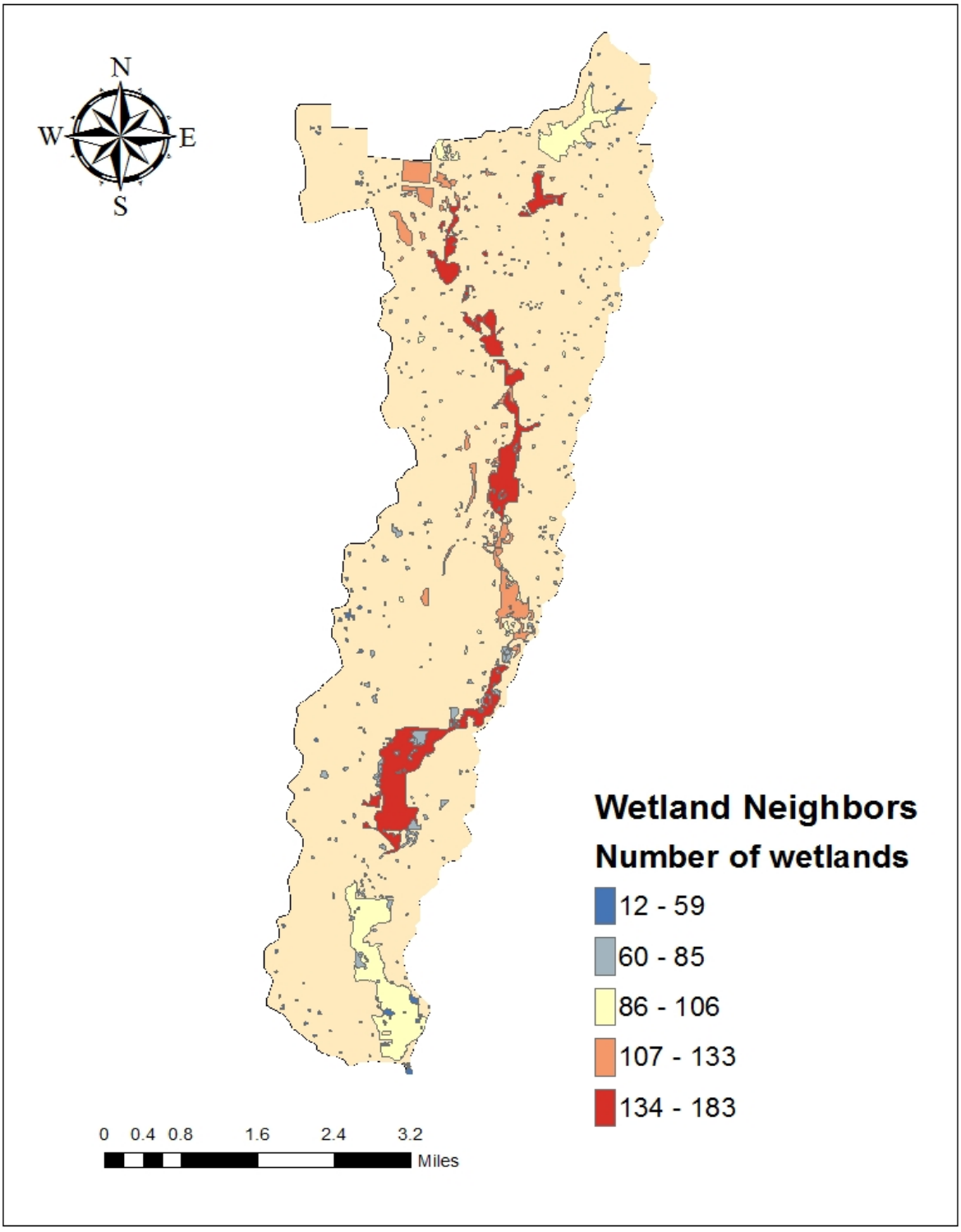

Figure 4-11: Wetland Proximity Output

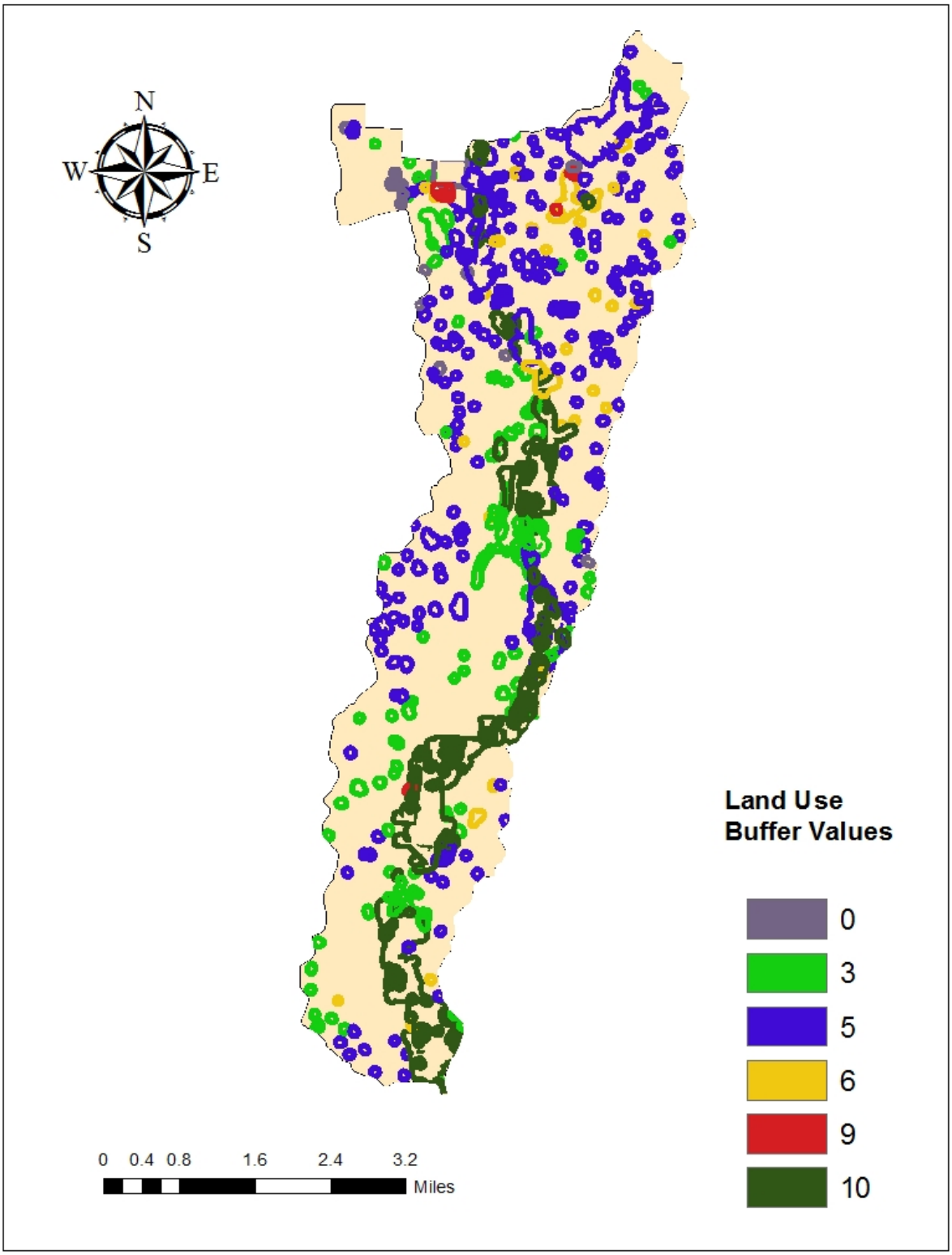

Figure 4-12: Majority Land Use Buffer Output

4.3 RESULTS

Phase 1 Base Model Results

This suitability analysis is a two phase approach. Phase 1 used the seven environmental indicator maps to create a raster representation of potential suitability for monitoring locations. The user can employ the Phase 1 analysis to locate areas which are deemed most suitable for the Phase 2 analysis. In this thesis, wetlands were not filtered out after the Phase 1 analysis because we wanted to take a whole watershed approach to the problem. However, a wetland manager with time and budgetary constraints might want to filter out wetlands which are clearly unsuitable. The Phase 2 analysis expanded on the knowledge from the environmental indicator maps by accounting for wetland proximity (Figure 4-11) and a majority land use buffer analysis (Figure 4-12). These two layers were combined with the output from Phase 1 to create the final suitability layer. The two-phase approach was also employed by White and Fennessy (2005) to prioritize the potential for wetland restoration sites on a watershed scale.

Phase 1 is the raster addition output of Figure 4-10. Each of the seven suitability utility layer pixels had a range of values from 0-10 as displayed in Table 4-2. Phase 1 analysis was initially performed with each layer receiving an equal weight of 0.1428 (Figure 4-11) in order for the seven map weights to sum up to one and reduce complexity. For the second Phase 1, analysis, the weights all varied based on perceived importance retrieved from a pairwise comparison. The pairwise method compares each utility layer to the other layers in order to determine which is perceived as most

important. The pairwise comparison was sent to members of the MODNR Wetlands Department and a former researcher involved with this project. The pairwise comparison scores and the weights are displayed in Tables 3 and 4.

The additional Phase 1 analysis was performed in order to examine the impacts of selecting different weights in the analysis. The two layers (Figure 4-13 & Figure 4-14) demonstrate differences in the areas which are considered most suitable. The pairwise comparison surveyors attributed the strongest weights to historic wetlands, hydric soils, and flooding frequency with values of 0.28, 0.24 and 0.17, respectively. This change in weights also results in areas, which were considered more suitable when compared to the unweighted maps. The weighted layer had 22% of the wetlands with a score greater than 8, while the unweighted map had only 13%. Also 39% of wetlands received a different score when comparing both Phase 1 analyses. This finding suggests that using a weighted output can greatly alter the output of the map to favor the user's preferences. Figure 4-15 compares the scores from the two Phase 1 outputs in a bar graph. The x-axis displays the Phase 1 suitability score that each wetland can receive, and the y-axis displays the number of wetlands which received that score. We can see that the weighted layer denoted by the blue bar contained a larger amount of wetlands receiving a score of 8 or 9.

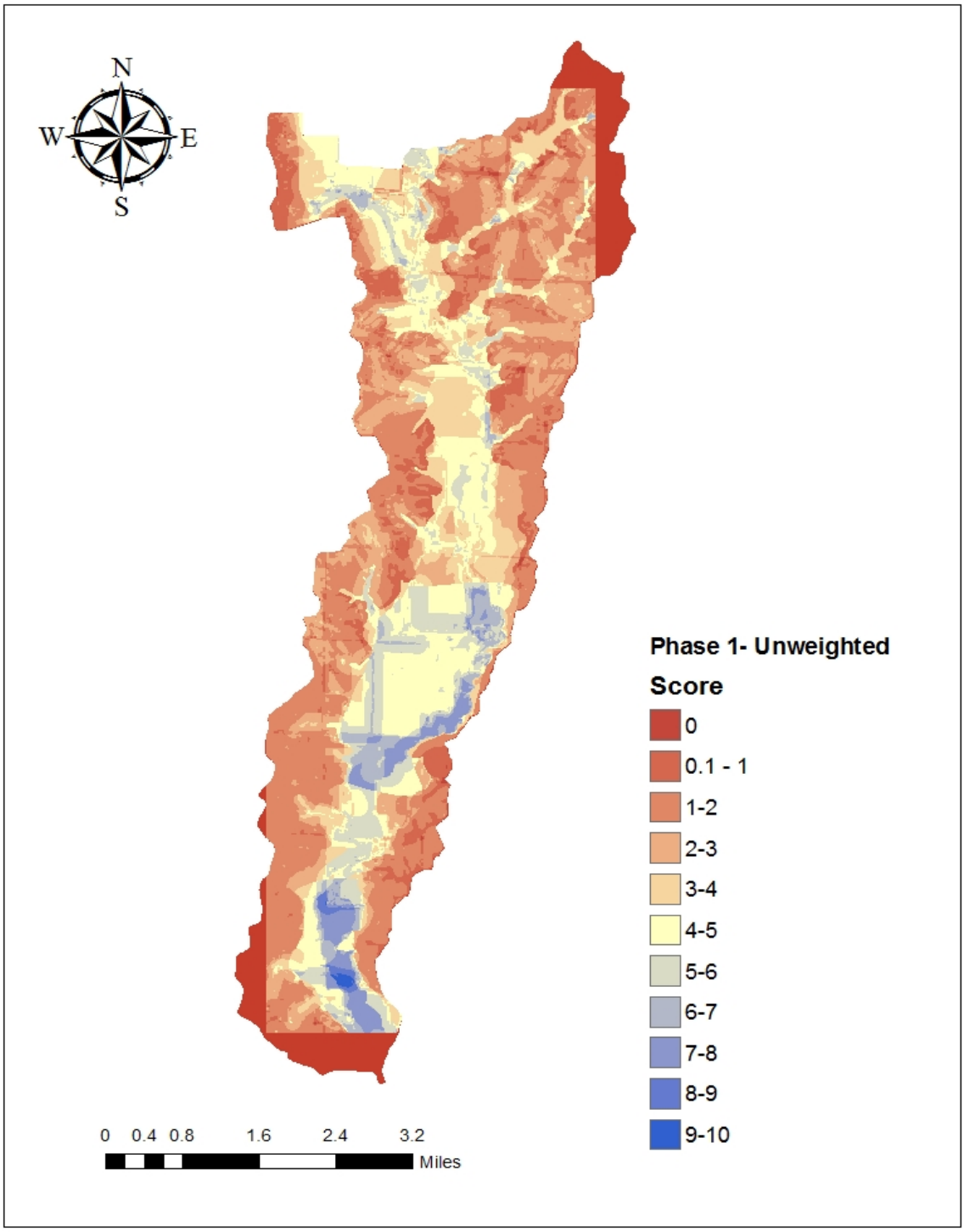

Figure 4-13: Phase 1 Site Suitability- Unweighted

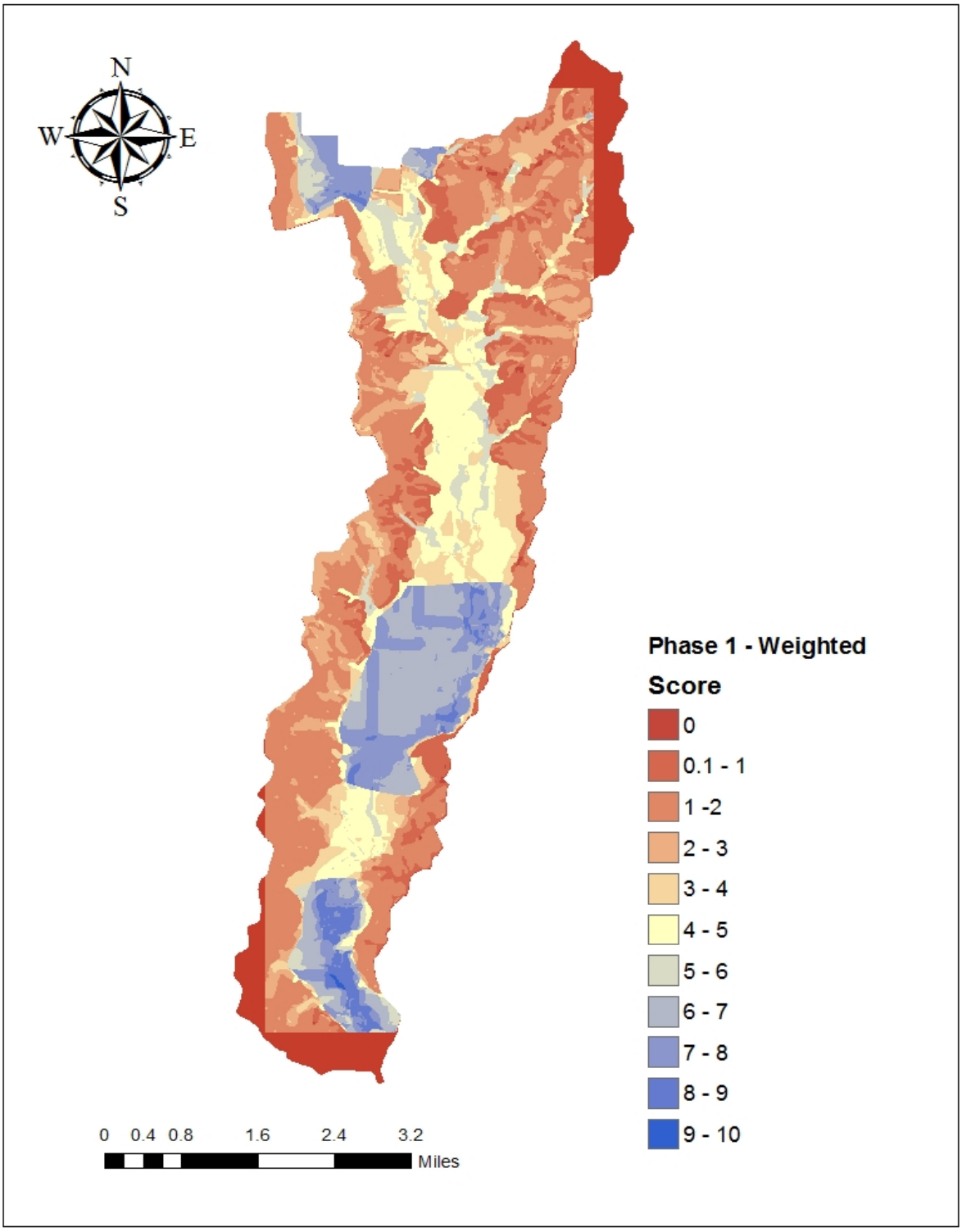

Figure 4-14: Phase 1 Site Suitability- Weighted.

Table 4-3: Pairwise Scores

1	Equally preferred
2	Equally to moderately preferred
3	Moderately preferred
4	Moderately to strongly preferred
5	Strongly preferred
6	Strongly to very strongly preferred
7	Very strongly preferred
8	Very to extremely strongly preferred
9	Extremely preferred

Table 4-1: Pairwise Survey Template with Final Weights

	Historic wetlands	Flooding Frequency	Hydric Soils	1st order stream buffer	Land use/ Land Cover	Hotspot Analysis	Flow accumulation Layer	Weights 1	Weights 2	Weights 3	**Total Average weight**
Historic wetlands								0.19	0.30	0.36	**0.28**
Flooding Frequency								0.20	0.09	0.24	**0.17**
Hydric Soils								0.25	0.32	0.15	**0.24**
1st order stream buffer								0.09	0.06	0.08	**0.07**
Land use/ Land Cover								0.09	0.03	0.05	**0.06**
Hotspot Analysis								0.10	0.13	0.09	**0.11**
Flow accumulation Layer								0.09	0.07	0.04	**0.07**

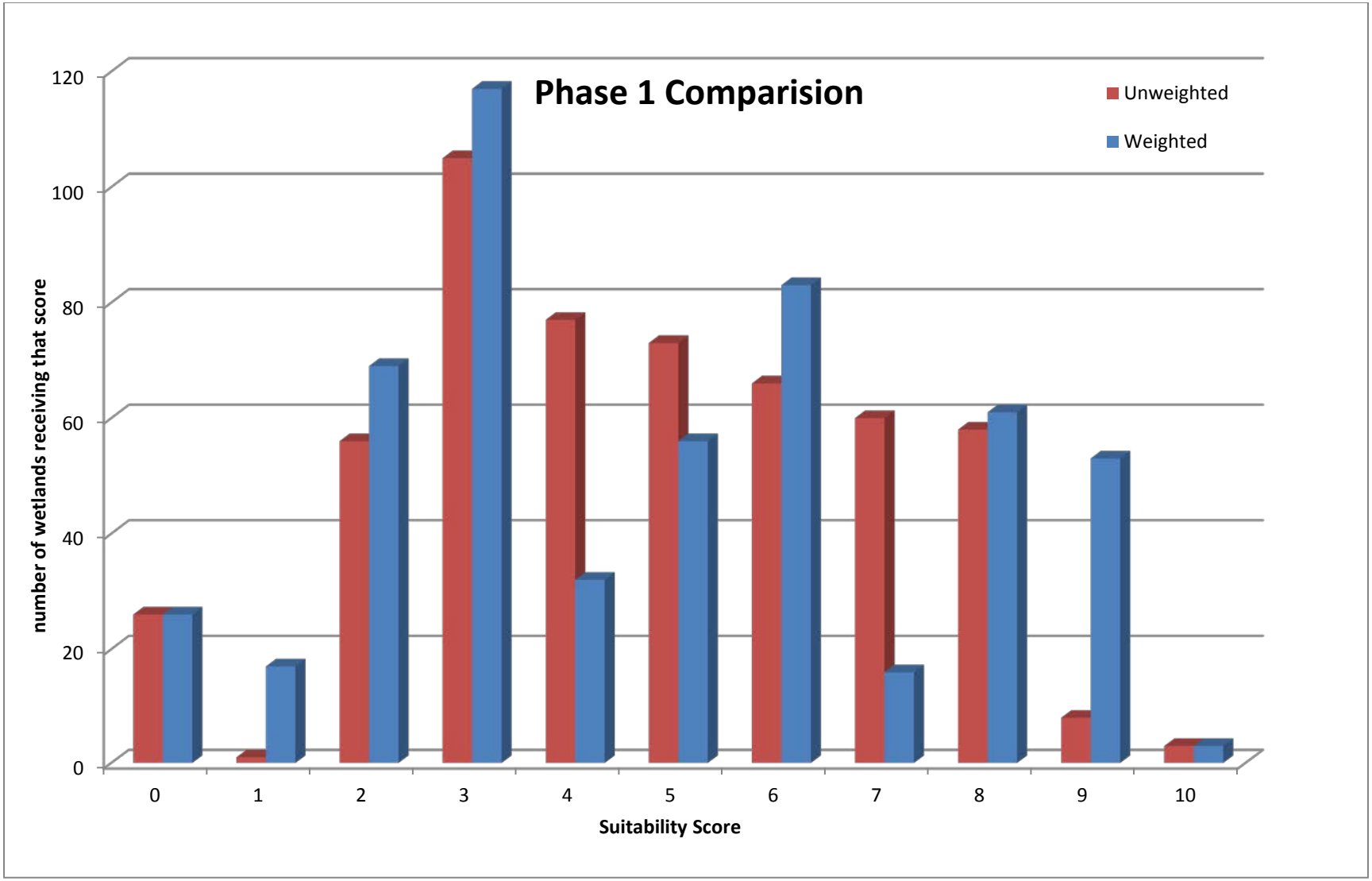

Figure 4-1: Phase 1 Comparision--Unweighted vs. Weighted

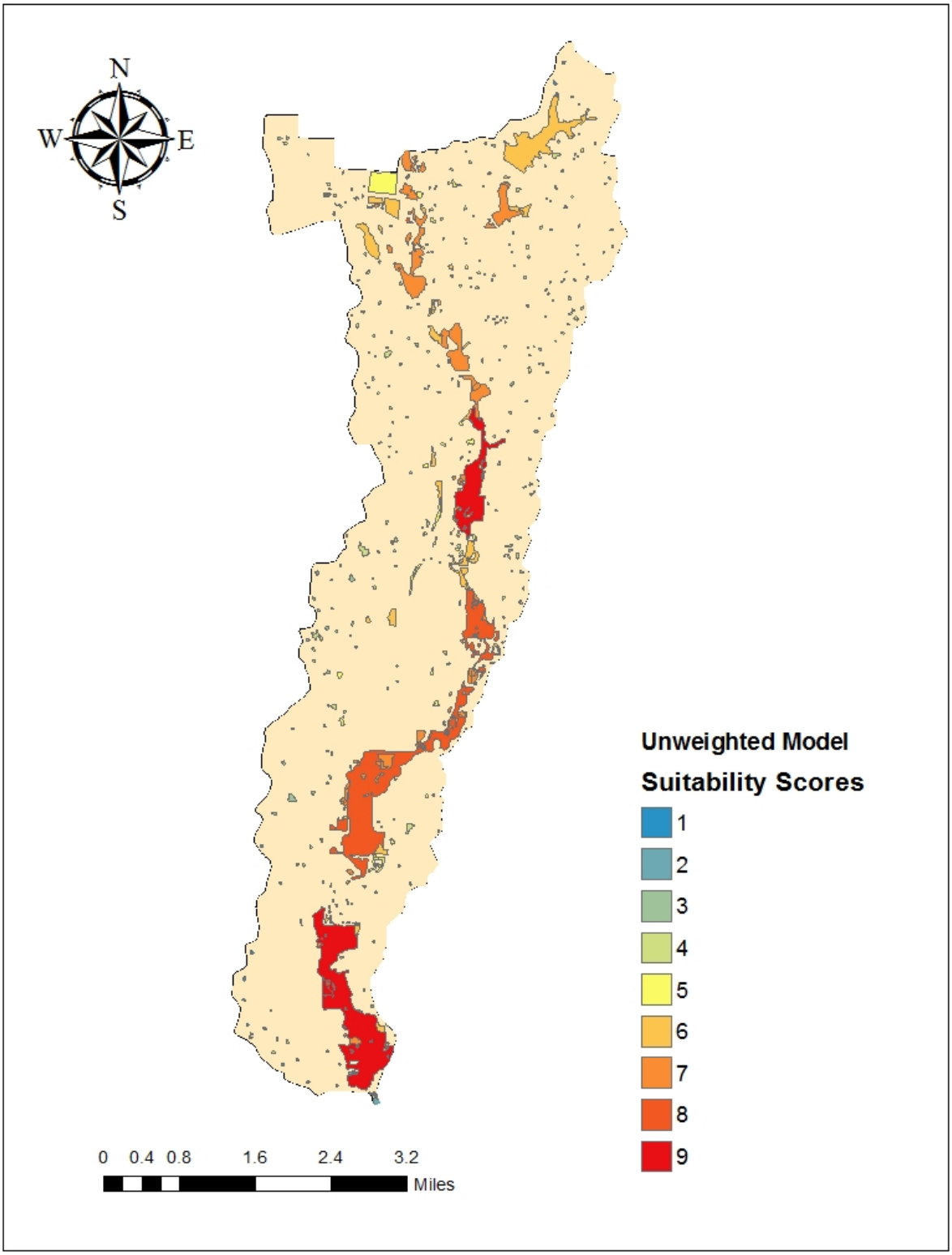

Figure 4-16: Final Site Suitability--Unweighted Output

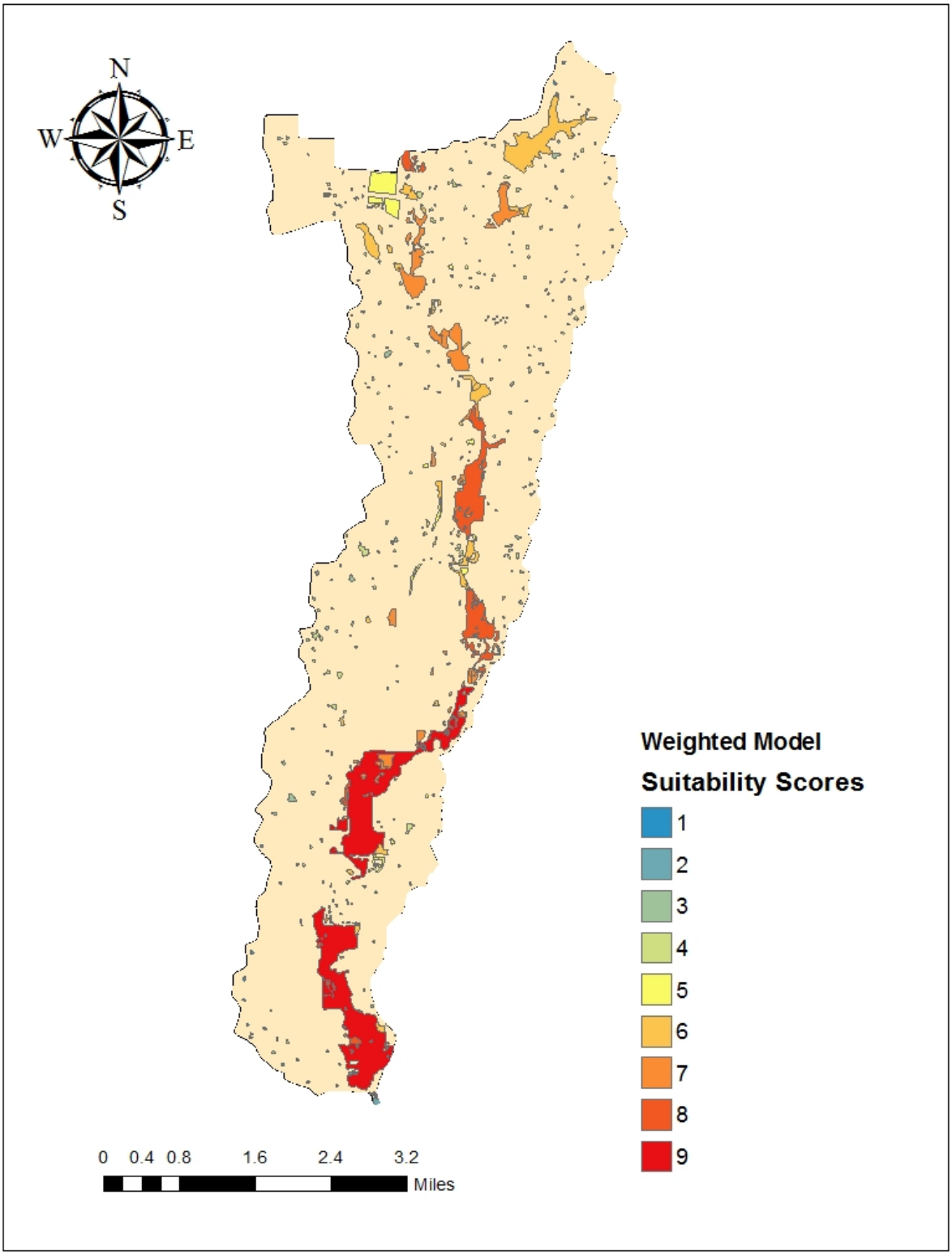

Figure 4-17: Final Site Suitability--Weighted Output

4.3.1 Final Suitability Results

The final analysis is presented in Figures 16 & 17. The first map (Figure 4-13) depicts results from the Phase 1 output, which included no weighted values, and the second map (Figure 4-14) shows the output from using the weighed values. The two maps illustrate each wetland in the subwatershed and assign it a score from 1 to 9 with 9 being the most suitable wetlands. Some wetlands overlapped multiple score pixels and were assigned the higher score pixel in their polygon. Both final analyses look similar but looking further at the statistics we see that they are not. When comparing Figures 16 & 17, approximately 16% of wetlands received different scores. When comparing which model created wetlands which were deemed more suitable, the weighted model had 9% and the unweighted model had only 4% of wetlands with scores of 8 or greater. Additional statistics are presented in Table 4-5**.** According to the analysis, the most suitable wetlands tended to be larger in size and located near the outlet of the watershed. There were some smaller wetlands in the northern portion of the watershed which received high scores by both the models as well. Figure 18 shows the center of the watershed at a more detailed level for unweighted (a) and weighted (b) final outputs. The difference in this picture stems largely from the historic wetland location which is highlighted on the weighted map (b). Historic wetlands received the highest weight from the surveyors at 0.28.

Table 4-5: Final model comparison Unweighted vs. Weighted

	Unweighted Suitability	Weighted Suitability
Percent of wetlands receiving score ≥ 8	4%	9%
Average score	4.87	4.93
Standard deviation	1.68	1.83
Maximum score	8.8	9
Majority score	4	4

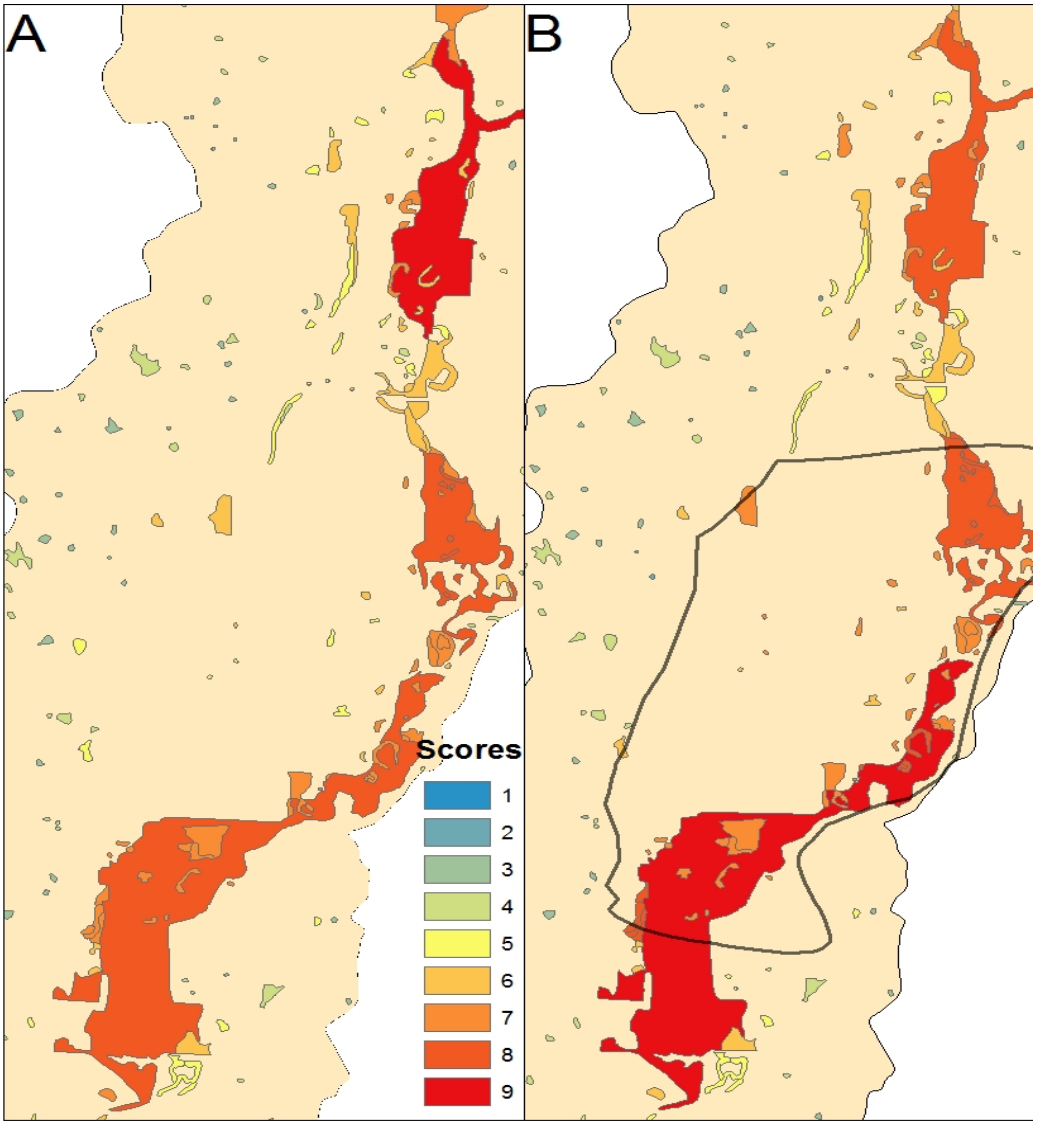

Figure 4-18: Final Phase Comparison--A = Unweighted output; B = Weighted output

4.4 Discussion

Decision-making tools must be dynamic so as to deal with ever changing circumstances. The suitability model proposed gives an output of optimal monitoring sites based on scientific wetland knowledge as well as the expertise of the wetland manager employing it. The manager decides what types of monitoring conditions are ideal (vegetation, microbial, amphibian, water quality, etc.). This allows the user to determine the layers used as well as the scores and weight values. This suitability analysis was designed for water quality sampling; thus, a large part of the map layers included hydrologic parameters. The model had many components, and changing a specific map layer value or weight impacted the output received.

Selecting weights and values

Selecting and prioritizing the importance of each GIS layer by attributing specific weights and values is difficult and quite subjective. This is commonly done by using accepted literature or expert opinion. However, every site is unique; thus, determining appropriate values should be left to local experts who are familiar with the landscape. It is understood that this portion of the suitability analysis is the most subjective and can have the greatest impact in the outputs as demonstrated in Figure 4-13 & 14 (White and Fennessy 2005). Understanding the variability in a model such as this allows the manager to understand its limitations. There is no panacea in resource management, and managers must understand the tools they are using.

Model Constraints

The current methodology only used publicly available data for Missouri to create the analysis. This means that there was a lot of time put into data mining this information and finding viable sources. The model can only be as effective as the data collected and analysis methods used. The model only gives managers a tool to try to filter out large wetland data sets. It does not take away the need for the manager to go out into the field and survey the land area for additional data collection. Of course, there are layers that were left out of the analysis, and a manger has the freedom to add or remove layers as needed.

Additional monitoring goals

While the layers selected in this analysis focused primarily on hydrology, soils, and land use there are other factors to consider when looking for optimal monitoring locations. Some wetlands, which are located in highlands and considered isolated, may also be effective monitoring locations based on needs. These wetlands may be located in areas that increase the overall understanding of a watershed area. For example, selecting a water monitoring location in a data barren site can help increase the validity of some of the interpolation methods explored in chapter 3.

Also In recent years, the index of biological integrity (IBI) has been used to monitor wetland health. This looks at the factors affecting plant life such as species richness, diversity, and evenness. Using the IBI, one can indicate anthropogenic stressors in the area that can be impacting the wetland (Miller et al. 2006). Depending

on the goal of the monitoring, having this information can be very useful for the wetland manager.

4.5 Conclusions

The motivation for the approach proposed in this thesis was due to a lack of current wetland data in Missouri. It is clear that current EPA initiatives like the NWCA are pushing for more understanding of the health of our wetland systems. While this initiative is needed, there must also be objective tools available for managers to utilize in selecting the wetland monitoring locations. This tool does just that by providing the managers with an array of environmental indicator maps which can demonstrate the potential for a wetland to be a monitoring location. The indicator maps in this thesis had a strong focus on the ability of the monitoring site to be hydrologically connected to surrounding water bodies. Most of the maps used in the Phase 1 analysis shared the goal of locating wetlands that had potential for being water quality sampling locations. When all these indicators of monitoring suitability are merged, the range of potential wetlands for ideal monitoring regimes can be better understood. Wetland managers must still select how often to monitor at these locations and what type of parameters should be involved in rendering. Nonetheless the location of monitoring suitability sites can be assisted with a tool such as this.

CHAPTER 5 : POTENTIAL APPLICATION OF EEM TO WETLAND CONNECTIVITY

5.1 Introduction

The health of aquatic systems is a constant concern for regulators and resource managers. This has caused some governments to develop constant monitoring regimes at major rivers and waterways in order to monitor changes (Singh et al. 2004; Zhao et al. 2011). Some of the sampling methods are laborious and can take a week or longer for the data to be properly processed like biological oxygen demand (BOD). A week's delay may be too late to act accordingly and detect the source of pollutants. Pollution originates from both point and a non-point source, which means regulators must understand how local water bodies are connected so that they can consider options to reduce the risk of contamination.

Current methods for sampling and analysis of water quality parameters are not always time or cost effective. Other measurements like organic dye concentrations and temperature changes are also not fast enough for rapid assessment (Cox et al. 2007; Flury and Wai 2003). However, recently the use of tracking natural organic matter (NOM) in waterways has been gaining momentum in water resource science (Baker 2001; Hudson et al. 2007; Hur and Cho 2012). NOM occurs from the natural degradation of plant and organic matter and can be used to fingerprint the origin of the water. NOM contains a mixture of complex organic matter which can be classified as humic substances (fulvic acid, humic acid, and humin), amino acids, lipids, and polysaccharides

(Thurman 1985). Understanding the concentration and type of NOM can help scientists solve problems relating to surface water connectivity and water quality in drinking water facilities (Baghoth et al. 2011; Timofeyev et al. 2004).

There are multiple methods, which attempt to detect types of NOM in water samples. For example, some methods require the addition of "mild" chemicals to allow fluorescent chemodosimeters to properly identify NOM (Du et al. 2012). However, there are more non-evasive and less time consuming methods in use such as fluorescence spectroscopy (Bro and Andersen 2003). NOM contains aromatic functional groups and conjugated double bonds which emit fluorescence when excited by light in the UV and blue regions (Stedmon et al. 2003). Within the realm of fluorescence spectroscopy, there is a method which uses the excitation and emission matrix (EEM) of the NOM to determine the different organic components in water samples.

Recently scientist have used multivariate statistical models such as the Parallel Factor Analysis (PARAFAC) to analyze the EEM of NOM. PARAFAC has proven to be a more accurate way of identifying NOM using EEM output (Bro and Andersen 2003). PARAFAC is a three-way decomposition method used for identifying the different loading and score factors of the NOM components in water samples (Bro and Andersen 2003). PARAFAC users can utilize score factors as potential concentrations of the NOM components. Recently, the use of PARAFAC in water treatment plants has seen a rapid increase. This is due to PARAFACs rapid assessment capabilities for characterizing NOM. PARAFAC has been used to characterize different water sources such as wastewater,

landfill leachate, and natural sources (Hua et al. 2007). Studies have even correlated the EEMs to different water quality parameters in drinking water facilities as well as urban rivers (Baghoth et al. 2011; Bieroza et al. 2009; Bieroza et al. 2010; Hur and Cho 2012; Sanchez et al. 2013). Using PARAFAC along a gradient which incorporates altering or mixing of water sources (a common practice at drinking water facilities) presents an ideal case study (Stedmon and Bro 2008). This portion of the project used PARAFAC in an attempt to identify which water quality parameters correlate best with EEM outputs. This will further advance the understanding of PARAFAC and its potential as a rapid assessment tool to elucidate water way connectivity. This technique can be used in a regulatory setting to help determine water body connectivity.

5.2 Methods

Two drinking water treatment facilities were used to demonstrate the utility of PARAFAC analysis to track water quality measurements. Since these processes aggressively change water quality, the effects of these changes should be evident from the EEM signatures. This confirms our understanding of similarities and differences in water quality that can be detected by EEM. The treatment process includes both physical changes (filter bed) and chemical changes (disinfection). This thesis shows how the physical changes can potentially be used to develop an understanding of how EEM is altered in the natural environment. For this project, the Boonville water treatment facility was the main location for developing the methodology to determine potential differences in EEM and water quality parameter correlations. A second treatment plant

in Vandalia was used as a source for water contrast; both plants use surface waters as their primary source waters and are located in Missouri. Boonville uses the Missouri river as its primary source while Vandalia uses a local reservoir. By using two treatment facilities which utilize different methods in treating their water we demonstrated new ways to incorporate EEM as a water quality tracking tool.

Boonville Site Location

The Boonville water treatment plant (MO3010089) was selected as the first case study sampling site. The treatment facility is located in Boonville, MO and uses the Missouri river as its source water. The population of Boonville is roughly 8,000 residents, and the treatment facility has a capacity of 4.6 MGD (million gallons per day). Using the Missouri river as its source water means that the treatment facility deals with fluctuations in turbidity and organic matter depending on upstream changes. This provides the research group with an interesting water treatment facility to test the viability of EEM analysis.

In 2007, Boonville changed their treatment process from chlorination to chloramination. Since 2005, Boonville has only had five drinking water violations, specifically with their Trihalomethane (TTHMs), a disinfection byproduct. Previous to the change, Boonville used chlorine as the primary disinfectant in their process. Chloramination involves the use of both ammonia and chlorine to disinfect water. The ammonia reacts with the free chlorine to create chloramines, specifically the monochloramine species, which are known to produce less disinfectant byproducts

(DBPs) caused by chlorination. They also renovated some of the treatment processes by adding additional steps (Figure 5-1).

Figure 5-2 displays Boonville's current treatment process. The red stars symbolize locations where samples were taken with plastic bottles for analysis of water quality parameters. Table 5-1 displays the chemicals added before the sampling points.

Table 5-1: Boonville Process descriptions

ID	Stage	Physiochemical processes
3-1	Source Water	Missouri River water- average TOC 5 mg/L, pH~8
3-2	Aerator outlet	oxidization of iron and manganese by air
3-3	Carbon settling basin outlet	aluminum coagulant + copper sulfate + powder activated carbon
3-4	primary sedimentation outlet	permanganate (NaMnO4) + aluminum coagulant
3-5	secondary sedimentation outlet	emulsion polymer
3-6	filter outlet	chlorine (OCl^-); disinfection
3-7	clear well outlet	chlorine (OCl^-) + fluoride + caustic soda; free chlorine residual
3-8	storage tank outlet	ammonia; chloramine residual

Old treatment process:
- (1) primary sedimentation,
- (2) secondary sedimentation,
- (3) carbon contact chamber,
- (4) tertiary sedimentation
- (5) clearwell

New treatment process:
- (6) aerator,
- (7) primary rapid mix, flocculation, sedimentation,
- (8) secondary rapid mix, flocculation, sedimentation,
- (9) dual media filters (4),
- (10) clearwell
- (11) process lab

Figure 5-1: Boonville site location

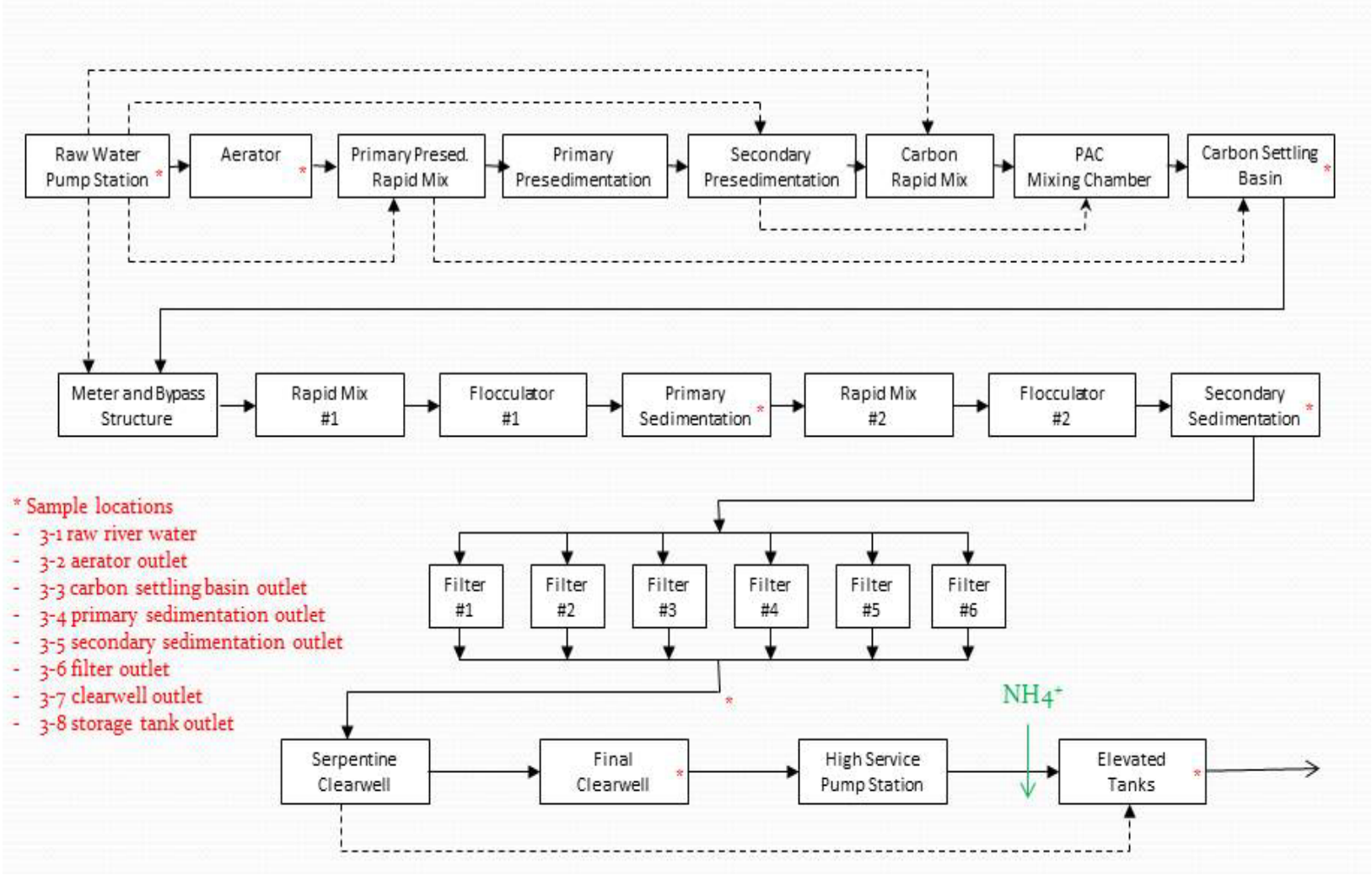

Figure 5-2: Boonville treatment process

Vandalia Site Location

The Vandalia water treatment plant (MO2010812) serves 2,529 users. The source water is retrieved from the Vandalia reservoir located approximately 6.3 miles northeast of the city. The average daily output is 220,000 GPD (gallons per day or 0.22 MGD), and they have received 16 violations since 2005 in both TTHM and Total Haloacetic Acids (HAA5), another disinfection byproduct. The treatment process includes adding copper sulfate ($CuSO_4$) during seasonal increases of algae. Table 5-2 indicates the different chemicals added at the treatment stage and Figures 5-3 and 5-4 displays the treatment process and site location.

Table 5-2: Vandalia Process descriptions

ID	Stage	Physiochemical processes
2-1	Raw pond/reservoir water	copper sulfate; algae control; TOC range is 4.5 to 9 mg/L; pH~7.7
2-2	Primary sedimentation outlet	powder activated carbon + aluminum sulfate ($Al_2(SO_4)_3$*$18H_2O$)
2-3	Secondary sedimentation outlet	lime (CaOH)+polymer+chlorine (OCl^-); raise pH to between 7.7 and 8
2-4	Filter effluent	chlorine (OCl^-) + fluoride; disinfection + residual
2-5	Clear well outlet	free chlorine contact time
2-6	Storage tank outlet	2-3 day detention time; sample collected from hydrant near tank

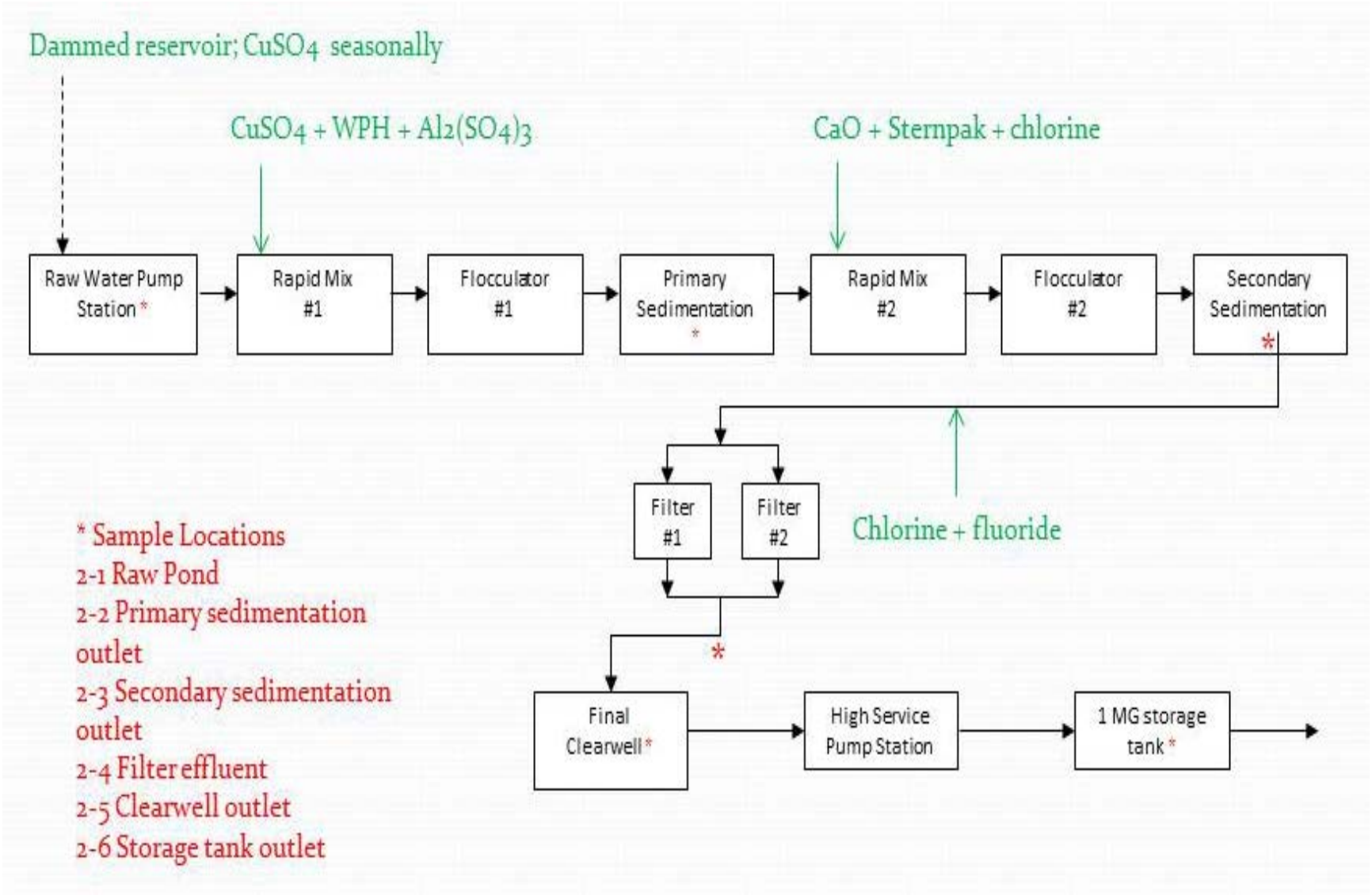

Figure 5-1: Vandalia treatment process

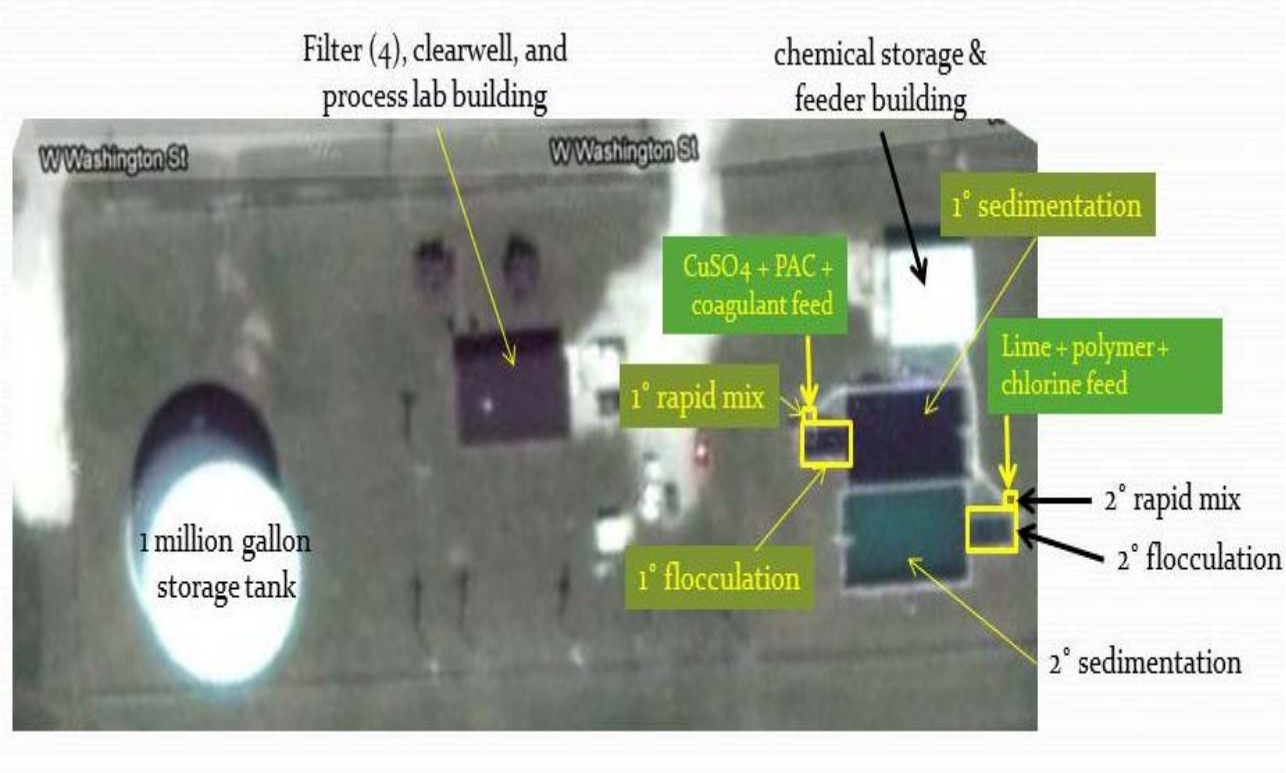

Figure 5-4: Vandalia site location

Measurement of different parameters

Each sampling location received extensive analysis of different water quality parameters. The analyses were done at Missouri S&T, Lincoln University, and University of Missouri-Columbia. Each sample was tested for the following parameters: EEM, pH, UV-254, DOC (ppm), Total Bromine (ppb) using an ICP-MS, Ammonia (mg/l), NO_2^- (mg/l), NO_3^- (mg/l), total inorganic Nitrogen (mg/l), total nitrogen (mg/l), DON+chloramine (mg/l), anion concentrations of NO_2^- , NO_3^- , Br^-, SO_4^{2-} (ppm), Chlorine concentrations (ppm), and THM (trihalomethane) concentrations of $CHCL_3$, $CHBrCL_2$, $CHBr_2Cl$, $CHBr_3$ (ppb) using a SPME-GC-ECD.

Fluorescence measurement

Each sample taken during the water treatment stages underwent fluorescence analysis using a Hitachi F-4500 Spectrograph (Hitachi Co.). The samples were originally refrigerated so they were allowed to reach a constant room temperature of around 21-+ 2 ˚C before analysis. Ambient temperature samples were placed into a 10 ml syringe and filtered through a single use 0.45 µm nylon filter into the 10 ml quartz cuvette. The cuvette was previously cleaned using DI water and Kimwipes. The sample in the cuvette is placed in the Hitachi and the xenon lamp set at a constant 700 V for each sample. The fluorescence spectra was measured at an excitation range of 200-400 nm with a 2 nm step, whereas the emission range of 200-500 nm with 3nm step was used. The instrument data was exported onto a Dell Dimension 8100 Pentium 4 with 1.48 GHz processor and 256 MB RAM. FL solutions 2.0 (Hitachi Co.) was used to display the data and to export the data as a text file which was later imported into a Microsoft Excel file.

5.2.1 PARAFAC modeling

After the data is exported from FL Solutions, it must be decomposed in order to determine the different fluorescence signatures. PARAFAC was selected as the preferred method of analysis for EEM decomposition based on its ability to analyze the data and give a solution as a three-way array. PARAFAC is a tri-linear decomposing method (Equation 5.1) that uses alternating least squares (ALS) to accurately fit the EEM data (Bro and Andersen 2003).

$$X_{ijk} = \sum a_{if} b_{jf} c_{kf} + e_{ijk} \qquad i = 1, \ldots., I: j = 1 \ldots . J: k = 1 \ldots K$$

Eq. 5.1

In Eq. 5.1, a = relative concentration of analyte f in sample i, b=emission of analyte f, c=excitation of analyte f , e= residual element

PARAFAC reduces the EEM into sets of one score vector and two loadings vectors for each analyte present (analytes can also be known as components or fluorophores). The score vector is sometimes assumed to be the concentration of the analyte which is "a" in the equation while the two loading vectors are excitation "b" and emission "c". "e" is the residual element which represents the unexplained signal (residuals containing noise and other un-modeled variation) (Stedmon et al. 2003). However, while some components can be ascribed to specific organic matter in the water samples, they are more likely a group of organic compounds having similar fluorescence properties (Baghoth et al. 2011). Additional research needs to be done with EEM in order to determine if these components are indeed an arrangement of organic matter. Nonetheless, the score vector can be used to illustrate variations in the organic composition of water samples in a dataset (Baghoth et al. 2011).

The PARAFAC model was fitted in MatLab R2007B using the *N*-way toolbox developed by Rasmus Bro. The tool box can be found on the following website http://www.models.life.ku.dk/nwaytoolbox. MatLab is a programming language that can be used to solve complex matrices, so it requires some level of understanding of

programming. Developers of the *N*-way toolbox also created a tutorial for the toolbox, which can be found at http://www.models.kvl.dk/~pih/parafac/chap0contents.htm.

Pre-Processing data

Before the data is run with the PARAFAC model it must be optimized and certain spectra must be removed. One such optimization is the removal of the Rayleigh and Raman scatters. Rayleigh and Raman scatters cannot be modeled correctly in PARAFAC because they do not comply with the premise of trilinearity and so they must be removed prior to the PARAFAC model calculation (Bro and Andersen 2003; Hua et al. 2007). Also the scatter signal from Rayleigh is usually very large and should be deleted for efficient modeling. Rayleigh scatter is elastic which means the photons frequency and wavelength do not change during excitation and emission and show up in the EEM as straight lines. Raman scattering is inelastic and causes the photons to vibrate and change energy states after they are excited. The fluorescence peak of the samples is anchored at a specific frequency which allows us to identify the NOM. However, Raman scattering can emit at more frequencies and is considered disconnected from the excitation frequency (Hua et al. 2007). One method to mitigate the impact of Raman scatter is to collect an EEM of a blank sample (Mill-Q water) and subtract it from the EEM of every sample. The MatLab code given by Dr. Bin Hua removes the Rayleigh scatter and replaces the missing values with NaN (Not a Number). Also the excitation-emission pairs with emission wavelengths 0-10 nm higher than excitation wavelengths of the first order Rayleigh scatter and 0-50 nm lower than twice the excitation

wavelengths of the second-order Rayleigh scatter, which were also replaced with NaN (Hua et al. 2010).

PARAFAC constraint options

The N-way toolbox PARAFAC model gives users a variety of different options to optimize their output. One such option is the ability to constrain the PARAFAC model. According to the N-way tutorial model, constraints are helpful to display specific spectra that are disrupted by noisy data. Model constraints are used for the following reasons as well: “Obtain parameters that do not contradict with a priori knowledge (Ex.: Require chromatographic profiles to have but one peak). Obtain unique solution where otherwise a non-unique model would be obtained (Ex.: Use selective channels in data to obtain uniqueness). Test hypotheses (Ex.: Investigate if specific NOM is present in sample). Avoid degeneracy and numerical problems (Ex.: Enabling a PARAFAC model of data otherwise inappropriate for the model). Speed up algorithms (Ex.: Use truncated bases to re-express problem by a smaller problem). Enable quantitative analysis of qualitative data (Ex.: Incorporate sex and job type in a model for predicting income)(Bro 2004)”. Some argue that constraining models that should be unique, like PARAFAC is superfluous. However, sometimes the model may not provide a satisfactory description of the data and constraints must be used.

The PARAFAC model gives users the ability to constrain the model using some preinstalled vector options. Users can make the model orthogonal, nonnegative, and input unimodality with nonnegativity. Each option is assigned a specific number that

when inputted in a vector format with MatLab constrains that specific output. For this project non-negativity constraint is employed as an indirect solution to not having enough samples. If there were more samples it might be possible to not employ non-negativity because the model would converge easier.

Validating the number of components

The PARAFAC model lets you determine the best number of components (factors) to describe your samples. When these samples are field-collected this can be a difficult task based on the complexity of NOM and the possibility of different sources. However, if the user considers and accepts that each component obtained from the PARAFAC corresponds to a defined group of fluorophores, then there are some tools that can be applied to determine the component number. Some of the tools include: (1) test the effect of increasing the number of components on the number of iterations used to fit the model. For example if n=2 gives 65 iterations and then n=3 gives 1,000 iterations, then n=2 may be the correct component number. Iteration fluctuations may also be a sign of ill-defined data such as loadings being too correlated. (2) Conduct a split-half analysis for two sample sets displaying the same common variations in data; and (3) calculate the core consistency diagnostic (CORCONDIA) as a function of the component number (Bro 2004; Hua et al. 2007). Bro states that the PARAFAC model is valid when the CORCONDIA value is closer to 100%; if too many components are used, the CORCONDIA will be close to zero; and when the CORCONDIA is close to 50% the model is unstable (Bro 2004). Once the CORCONDIA output rapidly decreases, then the

user must use the number of components prior to the decrease. For this study, CORCONDIA is used as the method to select the appropriate amount of components.

5.3 Results and Discussion

5.3.1 Boonville Water Treatment Facility

Selecting Components

Two model outputs are examined in each experiment: a constrained, non-negativity model and an unconstrained model. Below are the outputs from the models which help users select the best number of components to use. The model gives users a results page which includes the number of iterations (it), the summation of the squared errors (err), and the CORCONDIA value. For this experiment, the CORCONDIA value was used as the validation tool. Table 5-3 & 5-4 display the data from the outputs and illustrate the difference between the two models. Based on the constrained model, the CORCONDIA value drops after the two component models which implies that two components is the best fit. For the unconstrained model it drops after the one component model which means one component is the best fit for this model (Bro 2004). Fitting the data as a one component model for a natural water source did not seem like a valid selection based on previous work (Hua et al. 2007). This made the constrained model with two components the better choice.

Table 5-3: Constrained Model Outputs for Boonville

Constrained- Non-Negativity				
Model components	1	2	3	4
Iterations (it)	56	293	291	2215
Summation of squared errors (err)	197.0241	61.7061	29.6863	20.6501
CORCONDIA	100	100	-6.5641	5.0793

Table 5-4: Unconstrained Model Outputs for Boonville

Unconstrained				
Model components	1	2	3	4
Iterations (it)	55	1375	1232	798
Summation of squared errors (err)	197.0242	99.4197	29.5973	19.3004
CORCONDIA	100	-0.3633	-2.7622	8.5806

After the components were selected the PARAFAC model was prompted to display the values from the three loading vectors which are a=concentration, b=emissions, and c=excitation. In our model, there were two components so PARAFAC first displays them in order of concentration. C1 makes up 72% of the total concentration loading while C2 makes up 28% of the concentration loading. This means C1's loadings where displayed and then C2's. The loadings are displayed by each stage in the drinking water process in a bar graph in Figure 5-5. C1 dominates at every stage except for Stage 7 when C2 concentration surpasses. Stage 7 sees the introduction of chlorine (OCl-), fluoride, and sodium hydroxide (caustic soda). The caustic soda is used to regulate the pH of the water and the fluoride is commonly added to prevent tooth decay from drinking the

water. The introductions of these chemicals are the likely cause of the C2 change in concentration.

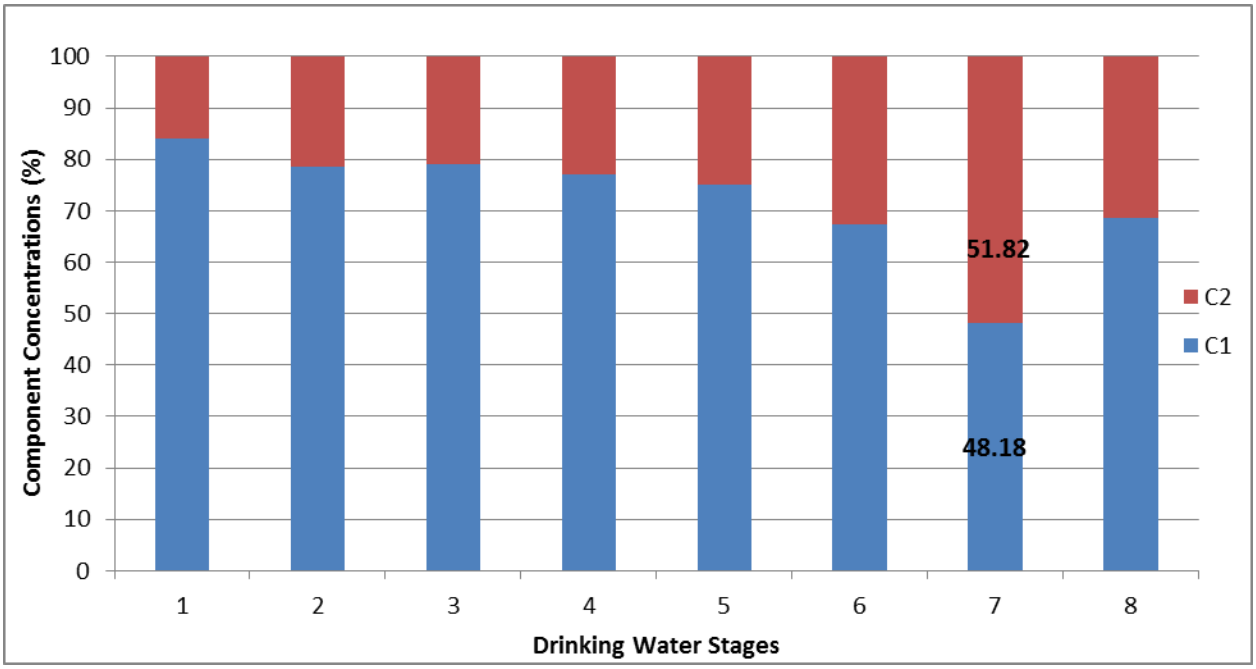

Figure 5-5: Component Concentrations as a function of Boonville drinking water treatment stage (1 source water, 2 aerator, 3 carbon basin, 4 primary settling, 5 secondary settling, 6 filter, 7 clearwell, 8 storage tank)

Characterizing NOM from components

One of the benefits of the EEM is the ability to characterize the NOM from the output data. Users can examine the peaks of the exicitation and emissions to cross reference their samples to previous studies (Coble 1996). Peak identification between articles can vary slightly based on instrument noise and natural sample influences. However, identifying the NOM components is beneficial in the tracking of changes and source of water. Literature has regulary classified the EEM spectras into the following catergories: humic substances (fulvic acid, humic acid, and humin), amino acids, lipids,

and polysaccharides (Thurman 1985). These NOMs can have terrestrial or marine origin which allows users to classify them. This study used a range of articles to help identify components based on the EEM Figures 5-6 & 5-7.

Component 1 (Figure 5-7) had an Emission range of 410-450 nm and an Excitation range of 230-250 nm. Yan et al. (2000) identified a similar component as humic acid- with emission of 420–450 nm and excitation at 230 –260. Hua et al. (2010) also identified his third factor as a terrestrial humic matter similar to our sample with 230–250 nm/emission: 400–410 nm/excitation. Finally Stedmon et al. (2003) also found similar EEM with Em 400-420 nm and Ex 250 which is also from terrestrial humic matter. From the literature, there is high confidence that our component 1 is indeed a humic like terrestrial substance which would be expected in the Missouri River (Hua et al. 2010; Stedmon et al. 2003; Yan et al. 2000).

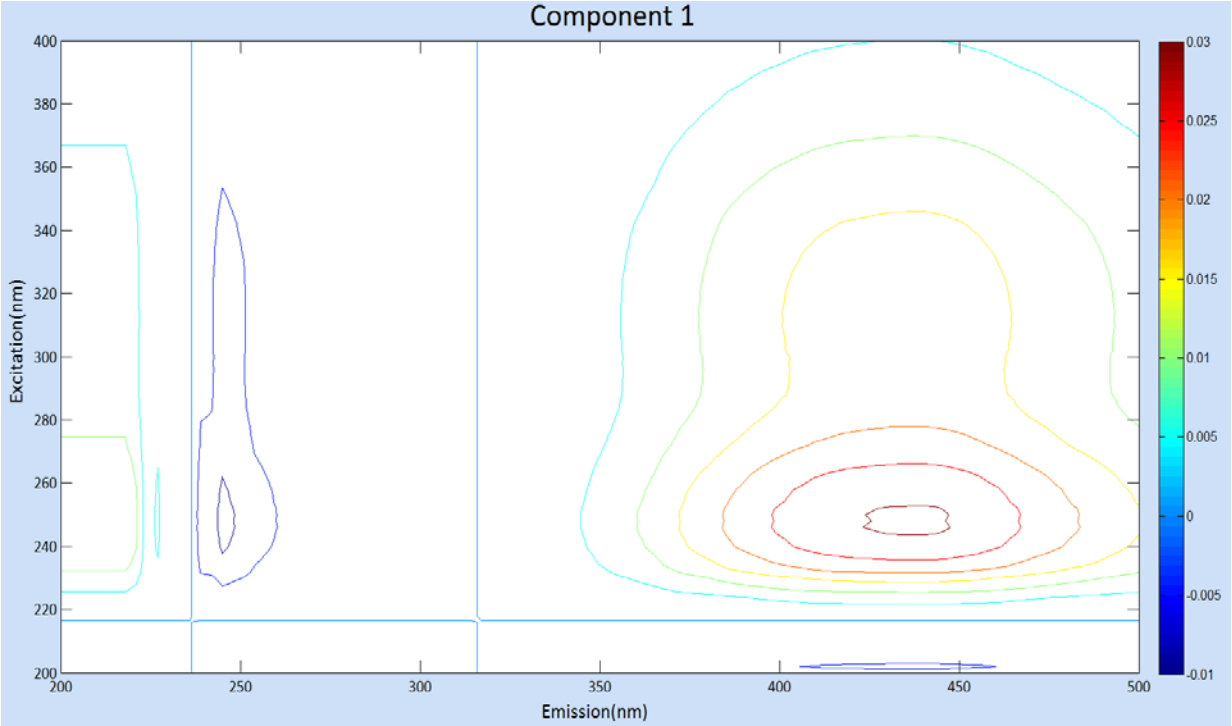

Figure 5-6: Component 1 EEM graph for the Boonville water treatment plant

Component 2 (Figure 5-7) displays a peak at the emission range of 360-390 nm/excitation of 220-240 nm and another peak at the emission range of 200-220 nm/excitation of 220-240 nm. This component was slightly harder to identify based on recognized values. It was not one of the common humic substances that have extensive literature information. However the first peak of Em 360-390 nm/ Ex 220-240 nm was actually identified as the T2 peak in Hambly et al. (2010) as a protein-like substance. Yan et al. (2000) also identified similar peak ranges of 340–350 nm with excitation at 220 and 275 nm as a protein- or amino acid-like substance. Finally, Hua et al (2010) found a similar substance but did not confidently identify it and instead suggested it was an amino acid. Therefore, the second component is likely a protein- or amino acid-like substance even though the second smaller peak remains unexplained. This may be

caused by instrument noise or the small sample set used (Hambly et al. 2010; Hua et al. 2010; Yan et al. 2000).

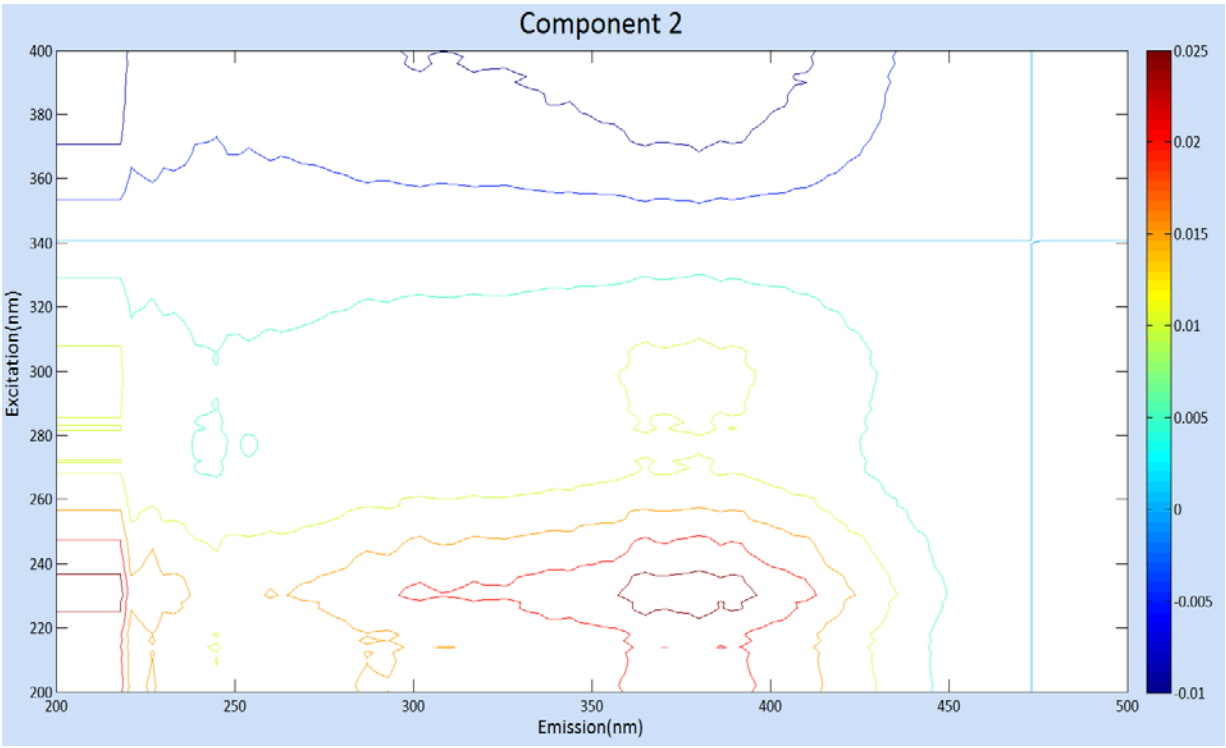

Figure 5-7: Component 2 EEM graph for the Boonville water treatment plant

Boonville EEM-Water Quality Correlation Results

Pearson's product-moment correlation (Pearson's r) is used to determine the relationships between EEM components and various water quality parameters. Pearson's r is a measure of the linear dependence between two known variables scoring them from +1 to -1. If the r value is +1 or close, then it is positively correlated, and if it is -1 or close it is negatively correlated (Hur and Cho 2012). Table 5-5 gives ranges of r values, which allow the correlations to be grouped into areas ranging from strong to weak correlations. The results of the Pearson's r correlations for both EEM components

are presented in Table 5-6. IBM's SPSS software was used to determine the correlations between the variables.

Table 5-5: Pearson's r value ranges

Pearson's r values	
+0.70 or >	Very strong positive relationship
+0.40 to +0.69	Strong positive relationship
+0.30 to +0.39	Moderate positive relationship
+0.20 to +0.29	Weak positive relationship
+0.01 to +0.19	No or negligible relationship
-0.01 to -0.19	No or negligible relationship
-0.20 to -0.29	Weak negative relationship
-0.30 to -0.39	Moderate negative relationship
-0.40 to -0.69	Strong negative relationship
-0.70 or <	Very strong negative relationship

Table 5-6: Correlation between EEM components and water quality parameters in Boonville

Component 1 Correlations	R value	Component 2 Correlations	R value
pH	0.797	pH	-.550
UV-254 Average	0.763	UV-254 Average	-0.721
DOC	-.036	DOC	-.064
SUVA	0.902	SUVA	-0.81
Ammonia-N	-.302	Ammonia-N	.018
NO_2^- -N	-0.825	NO_2^- -N	.672
NO_3^- -N	.544	NO_3^- -N	-.468
Total inorganic N	.028	Total inorganic N	-.293
Total N	-.224	Total N	-.094
DON+chloramine N	-.433	DON+chloramine N	.335
NO_2^- (IC)	-0.829	NO_2^- (IC)	.677
Br^- (IC)	0.650	Br^- (IC)	-.407
NO_3^- (IC)	.544	NO_3^- (IC)	-.468
SO_4^{2-} (IC)	.393	SO_4^{2-} (IC)	-.426
NO_2^- -N (IC)	-0.829	NO_2^- -N (IC)	.677
NO_3^- -N (IC)	.544	NO_3^- -N (IC)	-.468
NPOC "original"	-.034	NPOC "original"	-.066
TN "original"	-.226	TN "original"	-.091
NPOC "disinfectant"	-.078	NPOC "disinfectant"	.057
TN "disinfectant"	-.241	TN "disinfectant"	.224
$CHCl_3$	.347	$CHCl_3$	-.154
$CHBrCl_2$	0.934	$CHBrCl_2$	-0.807
$CHBr_2Cl$	.203	$CHBr_2Cl$	-.290
TTHMs	.469	TTHMs	-.276

Component 1 Correlations

Component 1(C1) was the dominate component in the analysis in all but Stage 7. C1 also had the most variables with a "very strong relationship" of >0.70. In order of most significant r values (positive or negative): $CHBrCl_2$, SUVA, NO_2^- (anions), NO_2^- N (anions), NO_2^- N, pH, and UV-254. Bromodichloromethane ($CHBrCl_2$) is a carcinogen and a common TTHM measured in drinking water facilities. It was originally used in flame retardants and chemical manufacturing. It received the highest correlation with a value

of .934. However the other 2 TTHMs and total TTHMs received much lower scores but still showed slight correlations. The next highest value .902 was the parameter known as SUVA or "Specific UV absorbance". SUVA is the measure of the aromatic characteristics of dissolved organic matter and is calculated by dividing UV-254 readings by the dissolved organic carbon (D0C) concentration. SUVA has been used in correlation studies dealing with EEM and water quality parameters so the high correlation was expected (Hua et al. 2010). In this thesis, SUVA was highly correlated to C1 which contained humic-like substances**.** UV-254 was also well correlated at .763, but DOC had no significant correlation. Nitrate (NO_2^-) measured as an anion using Ion chromatography (IC) received a correlation of -0.829 and Nitrate measured in mg/l using a Spectrophotometer had a negative correlation of -.825. Lastly, pH showed a strong positive correlation with a value of .797. C1 had four parameters show no significance, four showed "weak relationships", three showed "moderate relationships", and six showed "strong relationships'.

Component 2 correlations

Component 2 (C2) had less parameters with "very strong relationships" when compared to C1. In order of most significant r values (positive or negative): SUVA, $CHBrCl_2$, and UV-254. Also, all of the strong correlations where negative correlations, so as C2 increased in concentration, the other parameters decreased in concentrations. SUVA and UV-254 still showed fairly strong negative correlations with -0.810 and -0.721, respectively. Also $CHBr_2Cl$ showed strong correlations once again with .807. C2 did have

9 parameters with “strong relationships” (.4 to .69). In order to find potential correlations, one should look at these “strong relationships” instead of the “very strong relationships.” C2 also had seven parameters which showed no correlations, four with “weak correlations,” and one with “moderate correlations.”

5.3.2 Vandalia Water treatment facility results

Selecting Components

The same methods used to determine the components from Boonville were applied in the Vandalia process. This means that the constrained model and unconstrained models where analyzed in PARAFAC. Tables 7 & 8 show the outputs for the constrained and unconstrained-non negativity models. Unlike the previous models which had low CORCONDIA values for the unconstrained model, after two components were selected, this model actually showed viable solutions for up to two components for both constrained and unconstrained. The three component models also showed significantly higher CORCONDIA values when compared to the Boonville three component models. In the end, however, the constrained two-component model was selected because the CORCONDIA value was 100%.

Table 5-7: Constrained Model Outputs for Vandalia

Constrained- Non-Negativity				
Model components	1	2	3	4
Iterations (it)	57	205	277	306
Summation of squared errors (err)	446.3766	150.2804	95.0439	38.7948
CORCONDIA	100	100	83.3988	2.3588

Table 5-8: Un-Constrained Model Outputs for Vandalia

Unconstrained				
Model components	1	2	3	4
Iterations (it)	57	206	265	329
Summation of squared errors (err)	446.3766	150.2804	61.8617	33.1231
CORCONDIA	100	100	95.6345	56.6968

Once again, with the two-component model, we had two concentration outputs from MatLab. The concentrations showed some similar trends to Boonville results with C1 dominating at the beginning of the treatment process and then C2 dominating at the later stages of the process. C1 contributed 71% of the total concentration values while C2 accounted for 29%. The bar graph in Figure 5-8 shows the concentration gradient change from Stage 1 to 6. C2 dominates at the clearwell outlet and the storage tank outlet where the interaction with chlorine and the addition of fluoride are the likely cause of change in the concentrations.

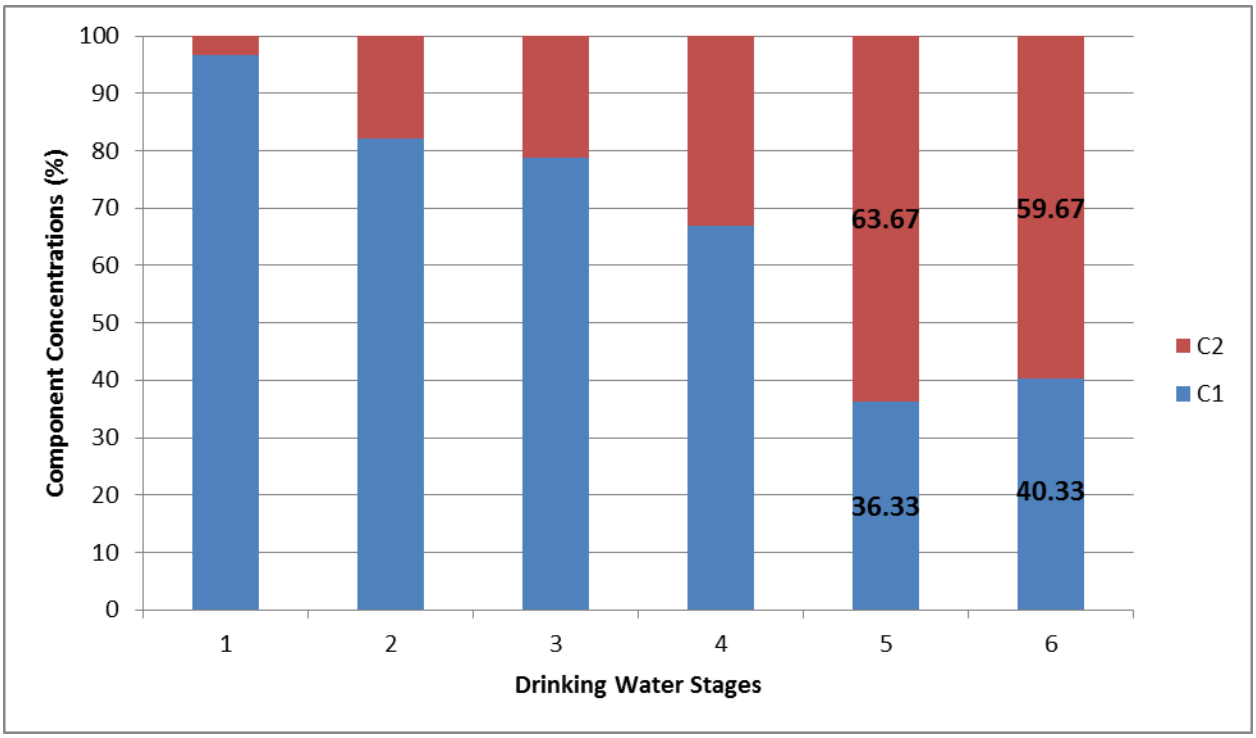

Figure 5-8: Component Concentrations as a function of Vandalia drinking water treatment stage (1 source water, 2 Primary Sedimentation, 3 Secondary Sedimentation, 4 filter effluent, 5 clearwell outlet, 6 storage tank)

Characterizing NOM from Components

Component 1 (Figure 5-9) had an Emission range of 400-460 nm and Excitation range of 230-250 nm. It was identified as a humic-like substances originating from terrestrial matter. This component is closely related to Boonville's component 1 which was also identified as a humic-like substance. Multiple sources have cited the EEM ranges as the humic component, and we are confident in the identification (Hua et al. 2010; Stedmon et al. 2003; Yan et al. 2000).

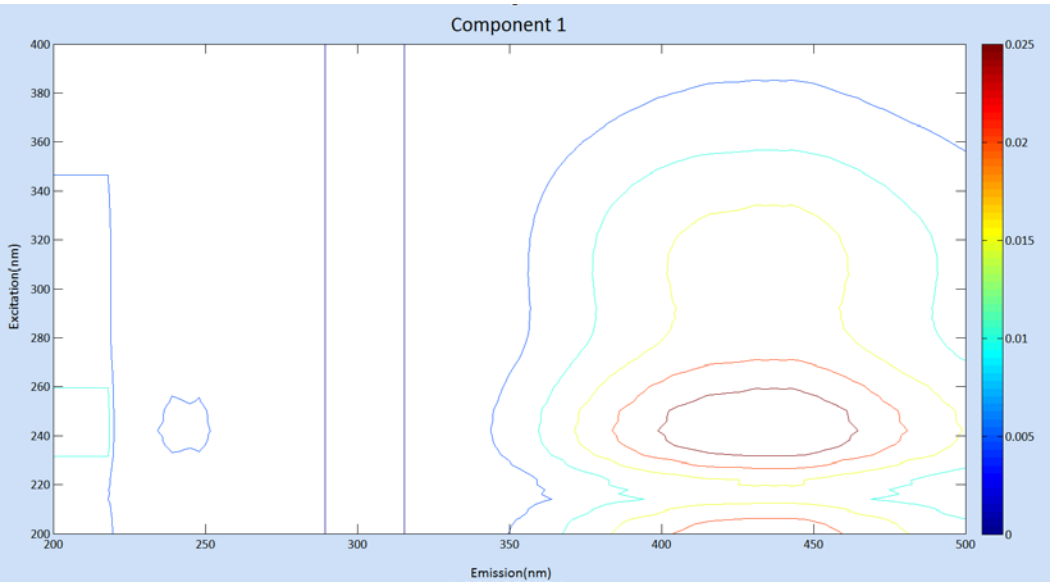

Figure 5-9: Component 1EEM for the Vandalia water treatment plant

Component 2 (Figure 5-10) displays an Emission peak at 310 nm and a duel Excitation peak at 230 nm and 280 nm. This component differs from the Boonville treatment plant's C2 (Figure 5-7). Using previous literature as a foundation to identify the component helped us identify it as an amino acid, protein-like substance. Baghoth et al. (2010) identified their C7 with similar attributes as found in our C2 with an Emission at 306 nm and Excitation at 270 nm (Baghoth et al. 2011). Stedmon and Markager (2005) also identified their C8 with an Emission of 306 nm and Excitation of 275 nm as an amino acid which is either free or protein bound (Stedmon and Markager 2005). However we do have two peaks and so while it is similar to that identified in literature there is a degree of error.

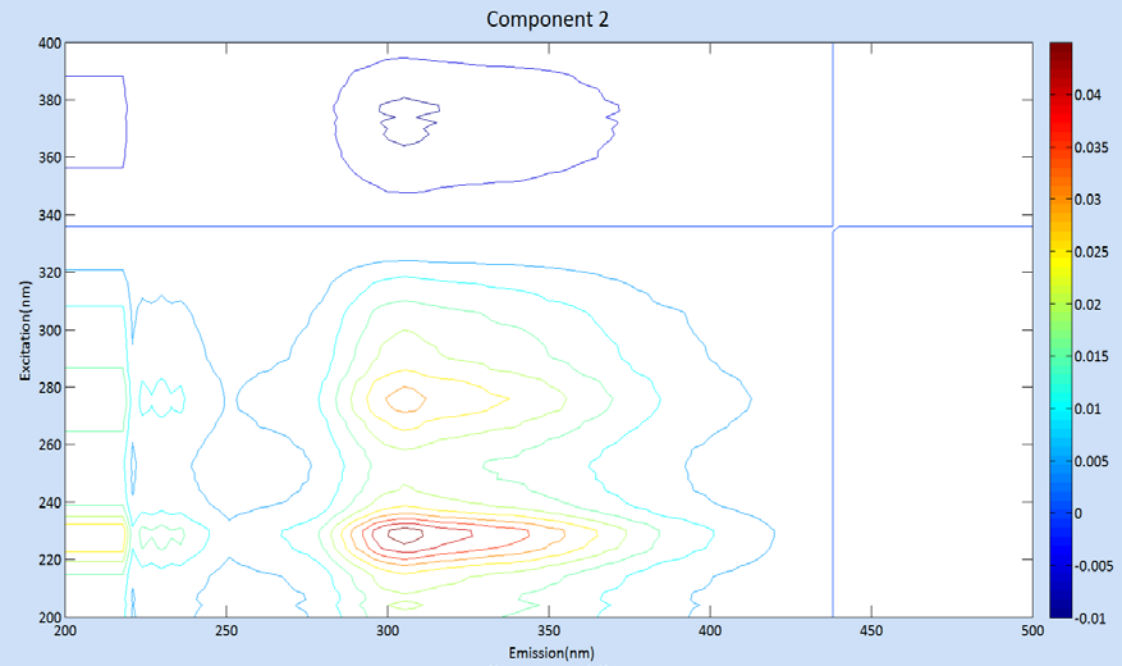

Figure 5-10: Component 2 EEM for the Vandalia water treatment plant

Vandalia EEM-Water Quality Correlation Results

Our correlation process followed the methods used for the Boonville correlations. Pearson's r was used and the value ranges in Table 5-5 was also utilized to organize the output. Unlike the Boonville correlations, C1 and C2 in Vandalia showed more similarities in their r values for each water quality parameter. Bromide (Br^-) received Pearson's r of not available (NA) because only one reading was above the detectable limit.

Component 1

C1 had nine water quality parameters show significant correlation (> 0.70). Out of the nine values, four had r values greater than 0.90. Ammonia-N received the highest value with 0.998, an almost perfect correlation. However, the last three ammonia

readings at Stage 4-6 were below the detectable limit and were not included in the correlation method. This suggests the correlation may change in the future if the values reach the detectable limits. The next significant values are from NO_2^- N and NO_2^- (IC), both with .995. These parameters also had the last three values under detectable limits. However NO_3^- N and NO_3^- (IC) received a value of -0.888, and all the stages stayed above detectable limits which means that it's likely the NO_2^- and ammonia will still receive high r values when all the stages' outputs are readable. UV-254 and SUVA had values of 0.729 and 0.727, respectively.

Component 2

C2 has ten water quality parameters with a significant correlation (>0.70) displayed in Table 5-9. While C1 mostly had positive correlations with parameters, C2 parameters received mostly negative correlations. For example, UV-254 was positively correlated (0.729) with C1 but negatively correlated (-0.780) with C2. Unlike C1, there was also a slightly more significant correlation of SO_4^{2-} with C2. Copper Sulfate ($CuSO_4$) was added in the early stages of the drinking water process which may lead to the higher correlation in C2. C2 had almost identical correlations with NO_3^- as C1 did. However, the other water quality parameters had a difference of roughly (0.01 - 0.10) between C1 and C2 correlations.

Table 5-9: Correlation between EEM components and water quality parameters in Vandalia

Component 1 Correlations	R value	Component 2 Correlations	R value
pH	-.269	pH	.277
UV-254 Average	.729	UV-254 Average	-.780
DOC	.647	DOC	-.663
SUVA	.727	SUVA	-.797
Ammonia-N	.998*	Ammonia-N	-.951
NO_2^- -N	.995	NO_2^- -N	-.963
NO_3^- -N	-.888*	NO_3^- -N	.889*
Total inorganic N	.115	Total inorganic N	-.210
Total N	.183	Total N	-.158
DON+chloramine N	.184	DON+chloramine N	-.152
NO_2^- (IC)	.995	NO_2^- (IC)	-.963
Br^- (IC)	NA	Br^- (IC)	NA
NO_3^- (IC)	-.888*	NO_3^- (IC)	.889*
SO_4^{2-} (IC)	-.670	SO_4^{2-} (IC)	.712
NO_2^- -N (IC)	.995	NO_2^- -N (IC)	-.963
NO_3^- -N (IC)	-.888*	NO_3^- -N (IC)	.889*
NPOC "original"	.647	NPOC "original"	-.663
TN "original"	.184	TN "original"	-.160
NPOC "disinfectant"	.642	NPOC "disinfectant"	-.670
TN "disinfectant"	-.034	TN "disinfectant"	.062
$CHCl_3$	.663	$CHCl_3$	-.638
$CHBrCl_2$	.662	$CHBrCl_2$	-.584
$CHBr_2Cl$	-.567	$CHBr_2Cl$	.637
TTHMs	.673	TTHMs	-.645

5.4 Discussion

Vandalia and Boonville use two different types of surface waters for their treatment process. Boonville uses the Missouri River which is a fast moving large river that carries a large quantity of suspended sediments and dissolved organic matter. Vandalia uses a reservoir which is supplied by an upstream river and has seasonal issues with algae blooms. Both Boonville and Vandalia EEM analysis identified humic-like substances and amino acid-protein like substances. C1, which was the humic-like material, is usually derived from the terrestrial ecosystem and introduced to the water body through natural processes like erosion (water, and wind) (Massicotte and Frenette 2011). Vandalia's and Boonville's C1 component demonstrated similar EEM and were confidently identified as the humic-like substance. C2, the protein-like substance usually originates through the extracellular release by photosynthetic organisms such as macrophytes and phytoplankton within the water body (Massicotte and Frenette 2011). Vandalia and Boonville's C2 EEM are vastly different in their peaks. They are both identified as protein-like substances but current PARAFAC modeling understanding is not advanced enough to give more detailed information such as origin.

The bar graphs (Figures 5-5 & 5-8) indicate the differences in the C1 and C2 concentrations along the drinking water process. Stage 1 in the process is the natural source water used in the treatment facility. Boonville had an initial C1 concentration of 84% while Vandalia had concentrations of 97%. This can be attributed to the fact that the Missouri River is dynamic while Vandalia has to use a reservoir which is static and

accumulates sediments. EEM ratios between C1 and C2 changed throughout the treatment process in both Boonville and Vandalia. This demonstrates how EEM corresponding to the water quality can change based on physical and chemical processes. For example, surface water percolating through the vadose zone into groundwater can mimic the natural sand filters in a treatment process. However, the more rapid changes, which occur during the disinfection stage, would not be expected in a natural environment unless a pollutant was being introduced into the water bodies.

One of the initial changes in the signatures occurs from the addition of coagulants in the treatment process which may be used to settle out NOM. After this stage, the ratio of C1 to C2 decreases as the reduction of the larger humic-like C1 coagulates. This has also been noted in research using drinking water facilities and EEM by Baghoth et al. (2010). The final stages in the drinking water treatment also showed dramatic changes with the addition of disinfectants and fluorides. In both facilities, the C2 concentration surpassed the C1 concentration in the later stages.

We were successfully able to use drinking water facilities as case studies to examine the change of EEM and water quality parameters along the treatment stage. This allowed us to develop our methodologies and also indicate potential parameters that can be correlated to EEM concentrations. Both treatment plants demonstrated different correlations between EEM components and the water quality parameters. In the future, these correlations may also be applied in natural waters to help determine surface water connectivity.

5.5 Conclusion

EEM analysis clearly has application potential in water resources. The PARAFAC analysis identified NOM that can be used to track source waters. Then the component concentrations also proved usable for creating correlations between known water quality parameters. EEM has only recently been seen as a feasible tool to be applied to engineering problems. However, while EEM has proven to be a great tool, the user must understand the model's limitations based on the strength of the data collected. In our analysis, we understand that our data set was limited, and we have addressed the issue accordingly. Stedmon et al (2008) suggests that at a minimum, between 20-100 samples should be collected. Collecting 100 or more samples is preferred and would allow the model to converge more effectively (Stedmon and Bro 2008). This means that our dataset of only one sampling date should be viewed with a certain degree of potential error. However, we have attempted to mitigate this error by implementing the non-negativity constraint.

The PARAFAC analysis can be applied in drinking water facilities (Baghoth et al. 2011; Bieroza et al. 2009; Bieroza et al. 2010). Previous studies focused on correlations between EEM and TOC, UV-254, DOC, SUVA, and other common organic measurement surrogates. This project analyzed 24 parameters with the intention of locating any other potential correlations. We did identify new potential correlations that can help these facilities remain in compliance. These additional correlations can be useful to water treatment operators and adds to the reason why EEM analysis should be used in these

facilities to improve operation efficiency. When dealing with TTHMs the high C1 concentrations (humic acid) correlated to high $CHBrCl_2$. Understanding this relationship can lead to new ways of removing more of the NOM from the stages before chlorination.

EEM has plenty of application potential in natural surface waters as well. PARAFAC is a potential monitoring tool because it can track very small variations in EEM datasets and resolves this data into independent groups of NOM (Hur and Cho 2012). Hur and Cho (2012) discovered high correlations between EEM, biochemical oxygen demand (BOD), chemical oxygen demand (COD), and total nitrogen. These parameters were linked to the organic content in the water, and EEM proved to be a good surrogate to the parameters. They successfully showed that EEM can be used to monitor an urban river which contained wastewater discharge (Hur and Cho 2012). Therefore, an apparent application for EEM is its use in pollution monitoring of natural water bodies, particularly when needing to confirm connectivity between water bodies in order to develop an appropriate pollution prevention strategy.

Massicotte et al. (2011) used EEM analysis to look at the spatial connectivity in a large river system and how the sources and fate of dissolved organic matter (DOM) changes. They evaluated how environmental heterogeneity and spatial connectivity may influence the processing of DOM in the St. Lawrence River (SLR). Previous studies by Giorgio et al. (2008) determined that as a river system moves linearly, the concentration of DOC is reduced by bacterial degradation unless new inputs occur from

surrounding tributaries. In Massicotte et al. (2011) the SLR's EEM signatures were found to be dominated by terrestrial humic-like substances at points where tributaries like wetlands and surrounding landscape drainage intersected. Their study also pointed out that the local environmental conditions accounted for more of the dynamics of the protein like DOM. Parameters, such as temperature, impact the growth rate of the photosynthetic organism which represented a major source of protein like DOM (Massicotte and Frenette 2011). This suggests that both the C1 and C2 components observed in our study could be directly impacted by a multitude of spatial and environmental factors, which can be used to help identify spatial connectivity.

From the initial data collected and correlated, it does seem like EEM has potential to be used as a rapid water quality assessment tool. It can be used to determine connectivity between waterways and help regulators understand which waterways to protect with the CWA. If waters show similar composition and concentration of specific NOM, one can potentially infer connections. In the stream to wetland connectivity paradigm, it is clear that the introduction of high levels of NOM found in some wetlands can change the streams NOM concentration and makeup (Massicotte and Frenette 2011). In addition, the suspected inlet EEM and outlet EEM in a wetland could potentially have different signatures based on the natural removal of suspended solids and other NOM in wetlands. NOM is a very complex compound and additional information on potential environmental and spatial impacts as evaluated by Massicotte et al. (2011) helps increase confidence in the analysis. When the NOM

signature is rapidly changing from these environmental and spatial impacts, it may be easier to infer connections. For this project, future work must be done to look at other water sources such as groundwater and mixing zones along coastal wetlands. By developing a larger understanding of the EEM signatures of different water sources, the user may become more confident in their ability to identify components and the environmental changes that impact these components.

CHAPTER 6 : CONCLUSION AND FURTHER RECOMMENDATIONS

6.1 Overall Goal of the project

The overall goals of this thesis were to develop methodologies in determining connectivity between surface waters. The first task employed GIS and available water quality data to interpolate water quality. The second task attempted to alleviate a caveat in surface water monitoring by introducing a method on selecting optimal wetland monitoring locations. Finally, the last task used EEM and PARAFAC to fingerprint water sources by their NOM signatures while also exploring any potential correlations between specific water quality parameters. All three methods attempted to create solutions to help resource managers navigate EPA regulations in determining stream and wetland connectivity.

6.2 Wetland GIS water quality work

Determination of chemical connectivity using available data from online databases was challenging. The MODNR water quality database was most helpful for collecting up-to-date and accurate data. Based on the quality and availability of data, chlorides were selected as the ideal parameter to help with the development of the connectivity model. Specific conductance was selected as another parameter to model due to its correlation to chloride values in the data. GIS Interpolation methods were used to increase the spatial data range. The interpolation methods explored included

kriging, LPI, GPI, IDW, and subwatershed-based interpolation. Cross validation using the RMSE and ME were used to identify the LPI model as the more appropriate interpolation method for chloride and universal kriging for the specific conductance model.

6.3 Site suitability analysis tool

Locating water quality data for the first task brought to light a caveat in water quality sampling which was the lack of sufficient wetland chemical data. Wetland water quality sampling is minimal which makes it difficult for resource managers to determine connectivity between streams and wetlands. The site suitability tool allows managers to locate optimal wetland monitoring locations based on publically available GIS map information. The suitability model was assessed by creating two outputs, one using weighted values and the other using unweighted values. The weighted model rendered a more precise depiction of potential monitoring locations. The vision is for this site suitability tool to be included as a resource in the upcoming Missouri wetlands plan which is being developed by MODNR.

6.4 EEM signatures

The final task demonstrated the potential for the use of NOM to track connectivity between waters. We used two drinking water plants (Boonville and Vandalia) as case studies to develop methods on how to use EEM and PARAFAC to fingerprint water sources. At the plants, specific treatment stages were sampled in order to display the difference in NOM concentrations based on water quality changes

due to physiochemical processes. There was a clear difference between the NOM components in the two sites as well as the concentration of the NOM components from their source waters. In addition to determining the NOM, component correlations were made between the NOM concentrations and common water quality parameters. The Boonville plant's NOM concentrations were found to be highly correlated to Bromodichloromethane ($CHBrCl_2$) and SUVA. In contrast, the Vandalia NOM concentrations were found to be highly correlated to ammonia (NH_4^+), nitrite (NO_2^-) and nitrate (NO_3^-).This may prove to be useful in understanding drinking water facility efficiencies in removing NOM and also points to potential organic compounds that may interact with the NOM signatures in natural waters.

6.5 Future recommendations

Each project has its own highlights and areas where it could be improved. The following discussion lists some future recommendations to build on the work presented in this manuscript.

Wetland GIS water quality work

- Field validate the accuracy of the interpolation data by sampling discrete points.
- Couple the chemical connectivity model with additional biological or hydrological models to create a more robust model.
- Examine other GIS applications such as BASINS/HSPF to help regulators model data without having to purchase expensive programs.

- Develop methods to incorporate groundwater connectivity into the model

Suitability analysis tool

- Explore additional GIS layers such as depressions in DEM which can be potential areas for water accumulation and wetland formation
- Send the model to MODNR officials and explore the possibility of it being used by wetland managers in the state.

EEM

- Analyze additional water sources such as groundwater wells and wetland types to develop a better understanding of the differences in EEM signatures based on water type.
- Increase the amount of literature collected on identifying components to make it easier to identify new components in water samples.
- Apply the water quality correlations to natural waterways to identify any other potential correlations which may help determining connectivity.
- Increase the amount of sampling done so the model converges more accurately.

Appendix

Chloride Interpolation RMSE and ME outputs for multiple model configurations

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Ordinary kriging Semivariogram	6.274009927	**5.725482469**	5.725482469	5.725482469	5.761852036	5.725482469
Ordinary kriging - Covariance	6.268059055	5.872939646	5.872939646	5.872939646	5.81750484	5.872939646
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Ordinary kriging Semivariogram	-0.103439433	**-0.56436913**	-0.56436913	-0.56436913	-0.459532995	-0.56436913
Ordinary kriging - Covariance	-0.134914283	-0.359819983	-0.359819983	-0.359819983	0.026863216	-0.359819983

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Simple kriging Semivariogram	5.488097284	5.41594081	5.41594081	5.41594081	5.417138197	5.41594081
Simple Kriging - Covariance	5.400100074	5.399944828	5.399944828	5.399944828	**5.392064287**	5.399944828
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Simple kriging Semivariogram	-0.171378347	-0.215546491	-0.215546491	-0.215546491	-0.219305463	-0.215546491
Simple Kriging - Covariance	-0.129795805	-0.127259734	-0.127259734	-0.127259734	**-0.126296726**	-0.127259734

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Universal kriging Semivariogram	5.982812611	**5.725482469**	5.725482469	5.725482469	5.761852036	5.725482469
Universal Kriging - Covariance	5.905157524	5.872939646	5.872939646	5.872939646	5.817518293	5.872939646
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy-False
Universal kriging Semivariogram	-0.364471539	**-0.07766886**	-0.07766886	-0.07766886	-0.459532995	-0.07766886
Universal Kriging - Covariance	-0.318460948	-0.359819983	-0.359819983	-0.359819983	0.026895253	-0.359819983

RMSE					
	Unoptimized-Power 3	Unoptimized-Power 2	Optimized-Power 1	anistropy-.9098368	anisotropy- 1
Inverse Distance Weighting	5.812781693	5.784745633	5.737378547	**5.680884523**	5.737378547
Mean Error					
	Unoptimized-Power 3	Unoptimized-Power 2	Optimized-Power 1	anistropy-.9098368	anisotropy- 1
Inverse Distance Weighting	-1.068061559	-0.911629242	-0.625606934	-0.625105678	-0.625606934

RMSE				
	Order of Polynomial-1	Order of Polynomial-2	**Order of Polynomial-3**	Order of Polynomial-4
Global Polynomial Interpolation	6.709915473	6.702213419	**5.649918887**	12.31681964
Mean Error				
	Order of Polynomial-1	Order of Polynomial-2	Order of Polynomial-3	Order of Polynomial-4
Global Polynomial Interpolation	-0.105432562	-0.465067855	-0.990216291	-2.443915237

RMSE						
	Unoptimized	**Optimized- Power 1**	Optimized- Power 2	Optimized- Power 3	Optimized- Power 4	
Local Polynomial Interpolation	6.010447238	**3.008485593**	3.72515347	3.288882367	10.46582121	
Mean Error						
	Unoptimized	Optimized- Power 1	Optimized- Power 2	Optimized- Power 3	Optimized- Power 4	
Local Polynomial Interpolation	-0.743014798	0.429022859	0.04523419	-0.331895238	-1.264765232	
RMSE						
	Kernel function- altered	anistropy-1	anistropy-2	anistropy-0.5	anistropy-0.34	anistropy-0.77
Local Polynomial Interpolation	6.093019395	3.008485593	6.260632799	3.336391733	3.618545377	3.060833654
Mean Error						
	Kernel function- altered	anistropy-1	anistropy-2	anistropy-0.5	anistropy-0.34	anistropy-0.77
Local Polynomial Interpolation	-0.62830765	0.429022859	-0.966384568	0.395512958	0.894950945	0.602527655

Specific Conductance Interpolations RMSE and ME outputs for multiple model configurations

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy- False
Ordinary kriging Semivariogram	249.4860132	223.1895007	223.1895007	223.1895007	229.3559292	223.1895007
Ordinary kriging - Covariance	205.3177502	202.6154323	202.6154323	202.6154323	**204.2452983**	202.6154323
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy- False
Ordinary kriging Semivariogram	-24.80245799	-17.33239817	-17.33239817	-17.33239817	-0.083105012	-17.33239817
Ordinary kriging - Covariance	12.91034098	-15.63517289	-15.63517289	-15.63517289	-0.02544345	-15.63517289

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy- False
Simple kriging Semivariogram	221.3184788	219.3349268	219.3349268	219.3349268	219.3349268	219.3349268
Simple Kriging - Covariance	217.4856015	216.4159117	216.4159117	216.4159117	**216.1761583**	216.4159117
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True	anisotropy- False
Simple kriging Semivariogram	-19.17652338	-33.91989944	-33.91989944	-33.91989944	-33.91989944	-33.91989944
Simple Kriging - Covariance	-6.314250864	-4.603370866	-4.603370866	-4.603370866	-4.754973131	-4.603370866

RMSE						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True .5	anisotropy- False
Universal kriging Semivariogram	249.4860132	223.1895007	223.1895007	223.1895007	218.2220952	223.1895007
Universal Kriging - Covariance	205.3177502	**202.6154323**	202.6154323	202.6154323	204.2452983	202.6154323
Mean Error						
	Unoptimized	Optimized	Lag distance	3900	anistropy- True .5	anisotropy- False
Universal kriging Semivariogram	-24.80245799	-17.33239817	-17.33239817	-17.33239817	-10.85915535	-17.33239817
Universal Kriging - Covariance	-12.91034098	-15.63517289	-15.63517289	-15.63517289	-7.858131322	-15.63517289

RMSE							
	Unoptimized- Power 3	Unoptimized- Power 2	Optimized- Power 1	anistropy- .5	4000 major ,0.279	2000 major, .139 ratio	anisotropy- 1
Inverse Distance Weighting	279.7254127	251.4477748	221.5790653	218.3575934	217.8295828	**217.5846693**	221.579065
Mean Error							
	Unoptimized- Power 3	Unoptimized- Power 2	Optimized- Power 1	anistropy- .5	4000 major ,0.279	2000 major, .139 ratio	anisotropy- 1
Inverse Distance Weighting	-26.45708229	-30.77137144	-34.13316879	-26.94249574	-25.69943094	-23.48720683	-34.133169

RMSE				
	Order of Polynomial-1	Order of Polynomial-2	Order of Polynomial-3	Order of Polynomial-4
Global Polynomial Interpolation	216.7100932	**217.0125589**	344.7403231	589.6642995
Mean Error				
	Order of Polynomial-1	Order of Polynomial-2	Order of Polynomial-3	Order of Polynomial-4
Global Polynomial Interpolation	-3.604381377	-8.186412695	29.90366548	-118.1949768

RMSE				
	Unoptimized	Optimized- Power 1	Optimized- Power 2	Optimized- Power 3
Local Polynomial Interpolation	215.8956022	**215.8956022**	254.4297057	~~161.48196~~
Mean Error				
	Unoptimized	Optimized- Power 1	Optimized- Power 2	Optimized- Power 3
Local Polynomial Interpolation	-3.99770759	-3.99770759	-16.24460496	29.33387273
RMSE				
	Optimized- Power 4	Kernel function- altered	anistropy-1	anistropy-2
Local Polynomial Interpolation	583.4402189	265.2872649	161.48196	269.3045006
Mean Error				
	Optimized- Power 4	Kernel function- altered	anistropy-1	anistropy-2
Local Polynomial Interpolation	-128.5041575	-23.58444634	29.33387273	-17.53390539

Bibliography

Akkala, A., Devabhaktuni, V., and Kumar, A. (2010). "Interpolation techniques and associated software for environmental data." *Environmental Progress & Sustainable Energy*, 29(2), 134-141.

Alam, M. (2012). "MODELING WETLAND CONNECTIVITY AND VULNERABILITY TO WETLAND-CORRIDOR LOSS." MS, University of Missouri- Columbia, University of Missouri-Columbia.

Alexander, R. B., Boyer, E. W., Smith, R. A., Schwarz, G. E., and Moore, R. B. (2007). "The Role of Headwater Streams in Downstream Water Quality." *Journal of the American Water Resources Association*, 43(1), 41-59.

Allan, J. D., Erickson, D. L., and Fay, J. (1997). "The influence of catchment land use on stream integrity across multiple spatial scales." *Freshwater Biology*, 37(1), 149-161.

Almendinger, J. E. (1998). "A method to prioritize and monitor wetland restoration for water-quality improvement." *Wetlands Ecology and Management*, 6(4), 241-251.

Anderson, M. P. (2005). "Heat as a ground water tracer." *Ground Water*, 43(6), 951-968.

Baghoth, S. A., Sharma, S. K., and Amy, G. L. (2011). "Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC." *Water Research*, 45(2), 797-809.

Baker, A. (2001). "Fluorescence Excitation-Emission Matrix Characterization of Some Sewage-Impacted Rivers." *ENVIRONMENTAL SCIENCE & TECHNOLOGY* 35(5), 948-953.

Belohlavek, R., and Vychodil, V. "Relational model of data over domains with similarities: An extension for similarity queries and knowledge extraction." 207-213.

Bencala, K. E. (1993). "A perspective on stream-catchment connections." *Journal of the North American Benthological Society*, 12(1), 44-47.

Bieroza, M., Baker, A., and Bridgeman, J. (2009). "Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment." *Science of The Total Environment*, 407(5), 1765-1774.

Bieroza, M. Z., Bridgeman, J., and Baker, A. (2010). "Fluorescence spectroscopy as a tool for determination of organic matter removal efficiency at water treatment works." *Drinking Water Engineering and Science*, 3(1), 63-70.

Blumm, M. C., and Bosse, S. L. (2007). "Justice Kennedy and the Environment: Property, States' Rights, and a Persistent Search for Nexus." *Washington Law Review*, 82(3), 667-736.

Brierley, G., Fryirs, K., and Jain, V. (2006). "Landscape connectivity: the geographic basis of geomorphic applications." *Area*, 38(2), 165-174.

Bro, R. (2004). "N-Way PARAFAC toolbox." <http://www.models.kvl.dk/source>. (11/15, 2012).

Bro, R., and Andersen, C. M. (2003). "Practical aspects of PARAFAC modeling of fluorescence excitation-emission data." *Journal of Chemometrics*, 17, 200-215.

Bro, R., and Vidal, M. (2011). "EEMizer: Automated modeling of fluorescence EEM data." *Chemometrics and Intelligent Laboratory Systems*, 106(1), 86-92.

Brody, S., Davis, S., Highfield, W., and Bernhardt, S. (2008). "A Spatial-Temporal Analysis of Section 404 Wetland Permitting in Texas and Florida: Thirteen Years of Impact Along the Coast." *Wetlands*, 28(1), 107-116.

Brown, M. T., and Vivas, M. B. (2005). "Landscape development intensity index." *Environmental Monitoring and Assessment*, 101(1-3), 289-309.

Bullock, A., and Acreman, M. (2003). "The Role of Wetlands in The Hydrological Cycle." *Hydrology and Earth System Sciences*, 7(3), 358-389.

Campbell, D. A., Cole, C. A., and Brooks, R. P. (2002). "A Comparison of Created and Natural Wetlands in Pennsylvania, USA." *Wetlands Ecology and Management*, 10(1), 41-49.

Ceballos-Silva, A., and López-Blanco, J. (2003). "Delineation of suitable areas for crops using a Multi-Criteria Evaluation approach and land use/cover mapping: a case study in Central Mexico." *Agricultural Systems*, 77(2), 117-136.

Chang, C. L. (2009). "The impact of watershed delineation on hydrology and water quality simulation." *Environmental Monitoring and Assessment*, 148(1-4), 159-165.

Chesney, E. J., Baltz, D. M., and Thomas, R. G. (2000). "Louisiana Estuarine and Coastal Fisheries and Habitats: Perspectives From A Fish's Eye View " *Ecological Applications*, 10(2), 350-366.

Christie, J., and Hausmann, S. (2003). "Various State Reactions to the SWANCC Decision." *Wetlands*, 23(3), 653-662.

Clausnitzer, D., Huddleston, J. H., Horn, E., Keller, M., and Leet, C. (2003). "Hydric Soils In A Southeastern Oregon Vernal Pool." *Soil Sci. Soc. Am. J.*, 67(3), 951-960.

Coble, P. G. (1996). "Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy." *Marine Chemistry*, 51(4), 325-346.

Collins, M. G., Steiner, F. R., and Rushman, M. J. (2001). "Land-use suitability analysis in the United States: Historical development and promising technological achievements." *Environmental Management*, 28(5), 611-621.

Copeland, C. (2010). "Wetlands: An Overview of Issues." C. R. Service, ed.D.C. , 1-22.

Costanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., and Van Den Belt, M. (1997). "The Value of the World's Ecosystem Services and Natural Capital." *Nature*, 387(6630), 253-260.

Cox, M. H., Su, G. W., and Constantz, J. (2007). "Heat, chloride, and specific conductance as ground water tracers near streams." *Ground Water*, 45(2), 187-195.

Cvetkovic, M., and Chow-Fraser, P. (2011). "Use of ecological indicators to assess the quality of Great Lakes coastal wetlands." *Ecological Indicators*, 11(6), 1609-1622.

Dahl, T. E., and Watmough, M. D. (2007). "Current approaches to wetland status and trends monitoring in prairie Canada and the continental United States of America." *Canadian Journal of Remote Sensing*, 33(SUPPL. 1), S17-S27.

Davies, S. N., Thompson, G. M., Bentley, H. W., and Stiles, G. (1980). "Ground-water tracers - a short review." *Ground Water*, 18(1 , Jan.-Feb. 1980), 14-23.

Davis, J. A., and Froend, R. (1999). "Loss and degradation of wetlands in southwestern Australia: Underlying causes, consequences and solutions." *Wetlands Ecology and Management*, 7(1-2), 13-23.

Dow, C. L., and Zampella, R. A. (2000). "Specific Conductance and pH as Indicators of Watershed Disturbance in Streams of the New Jersey Pinelands, USA." *Environmental Management*, 26(4), 437-445.

Downing, D., Winer, C., and Wood, L. (2003). "Navigating Through Clean Water Act Jurisdiction: A Legal Review." *Wetlands*, 23(3), 475-493.

Driscoll, C. T., Lawrence, G. B., Bulger, A. J., Butler, T. J., Cronan, C. S., Eagar, C., Lambert, K. F., Likens, G. E., Stoddard, J. L., and Weathers, K. C. (2001). "Acidic Deposition in the Northeastern United States: Sources and Inputs, Ecosystem Effects, and Management Strategies." *BioScience*, 51(3), 180-198.

Du, J., Hu, M., Fan, J., and Peng, X. (2012). "Fluorescent chemodosimeters using "mild" chemical events for the detection of small anions and cations in biological and environmental media." *Chemical Society reviews*, 41(12), 4511-4535.

EPA "Draft Guidance on Identifying Waters Protected by the Clean Water Act." U. E. P. A. (EPA), ed.http://www.epa.gov/indian/pdf/wous_guidance_4-2011.pdf, 1-39.

EPA (2008). "Fact Sheet on CWA Section 106 Monitoring Initiative." <http://water.epa.gov/grants_funding/cwsrf/upload/2008_07_18_cwfinance_fact-sheet-monitoring-initiative.pdf>. (2/13, 2013).

EPA (2012). "National Wetland Condition Assessment." <http://water.epa.gov/type/wetlands/assessment/survey/index.cfm>. (September 30th 2012).

EPA (2012). "Water: Stage 2 DBP Rule." <http://water.epa.gov/lawsregs/rulesregs/sdwa/stage2/basicinformation.cfm#one>. (11/24, 2012).

Erxleben, J., Elder, K., and Davis, R. (2002). "Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains." *Hydrological Processes*, 16(18), 3627-3649.

ESRI (2006). "ArcGIS Spatial Analyst: Advanced Raster Spatial Analysis." <http://www.esri.com/library/brochures/pdfs/spatialanalystbro.pdf>. (February 12, 2013).

ESRI (2011). "An overview of the Interpolation toolset." <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z0000006900000.htm>. (February 13, 2013).

Filoso, S., Vallino, J., Hopkinson, C., Rastetter, E., and Claessens, L. (2004). "MODELING NITROGEN TRANSPORT IN THE IPSWICH RIVER BASIN, MASSACHUSETTS, USING A HYDROLOGICAL SIMULATION PROGRAM IN FORTRAN (HSPF)1." *JAWRA Journal of the American Water Resources Association*, 40(5), 1365-1384.

Flury, M., and Wai, N. N. (2003). "Dyes as tracers for vadose zone hydrology." *Reviews of Geophysics*, 41(1), 2-1 - 2-37.

Gupta, N., Aktaruzzaman, M., and Wang, C. (2012). "GIS-based Assessment and Management of Nitrogen and Phosphorus in Rönneå River Catchment, Sweden." *Journal of the Indian Society of Remote Sensing*, 40(3), 457-466.

Guttery, R. S., Poe, S. L., and Sirmans, C. F. (2004). "An Empirical Investigation of Federal Wetlands Regulation and Flood Delineation: Implications for Residential Property Owners." *The Journal of Real Estate Research;*, 26(3), 299-315.

Haberl, R., Perfler, R., and Mayer, H. (1995). "Constructed Wetlands in Europe." 305-315.

Hambly, A. C., Henderson, R. K., Baker, A., Stuetz, R. M., and Khan, S. J. (2010). "Fluorescence monitoring for cross-connection detection in water reuse systems: Australian case studies." *Water Science & Technology*, 61(1), 155-162.

Harmancioglu, N. B., and Alpaslan, N. (1994). "Basic approaches in design of water quality monitoring networks." *Water Science and Technology*, 30(10 pt 10), 49-56.

Heberer, T. (2002). "Tracking persistent pharmaceutical residues from municipal sewage to drinking water." *Journal of Hydrology*, 266(3–4), 175-189.

Hooke, J. (2003). "Coarse sediment connectivity in river channel systems: a conceptual framework and methodology." *Geomorphology*, 56(1–2), 79-94.

Hough, P., and Robertson, M. (2008). "Mitigation Under Section 404 of the Clean Water Act: Where it Comes From, What it Means." *Wetlands Ecology and Management*, 17(1), 15-33.

Hua, B., Dolan, F., McGhee, C., Clevenger, T. E., and Deng, B. (2007). "Water-source characterization and classification with fluorescence EEM spectroscopy: PARAFAC analysis." *International Journal of Environmental Analytical Chemistry*, 87(2), 135-147.

Hua, B., Veum, K., Yang, J., Jones, J., and Deng, B. (2010). "Parallel factor analysis of fluorescence EEM spectra to identify THM precursors in lake waters." *Environmental Monitoring and Assessment*, 161(1-4), 71-81.

Hudak, P. F. (2001). "Water hardness and sodium trends in Texas aquifers." *Environmental Monitoring and Assessment*, 68(2), 177-185.

Hudson, N., Baker, A., and Reynolds, D. (2007). "Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review." *River Research and Applications*, 23(6), 631-649.

Hur, J., and Cho, J. (2012). "Prediction of BOD, COD, and total nitrogen concentrations in a typical urban river using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices." *Sensors*, 12(1), 972-986.

Kadlec, R. H., Cuvellier, C., and Stober, T. (2010). "Performance of the Columbia, Missouri, Treatment Wetland." *Ecological Engineering*, 36(5), 672-684.

Kar, B., and Hodgson, M. E. (2012). "A Process Oriented Areal Interpolation Technique: A Coastal County Example." *Cartography and Geographic Information Science*, 39(1), 3-16.

Kelly, V. R., Lovett, G. M., Weathers, K. C., Findlay, S. E. G., Strayer, D. L., Burns, D. J., and Likens, G. E. (2007). "Long-Term Sodium Chloride Retention in a Rural Watershed: Legacy Effects of Road Salt on Streamwater Concentration." *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 42(2), 410-415.

King, R. S., and Richardson, C. J. (2003). "Integrating Bioassessment and Ecological Risk Assessment: An Approach to Developing Numerical Water-Quality Criteria." *Environmental Management*, 31(6), 795-809.

Kirby, C. S., and Cravotta Iii, C. A. (2005). "Net alkalinity and net acidity 1: Theoretical considerations." *Applied Geochemistry*, 20(10), 1920-1940.

Kirby, C. S., and Cravotta Iii, C. A. (2005). "Net alkalinity and net acidity 2: Practical considerations." *Applied Geochemistry*, 20(10), 1941-1964.

Kumar, A., Maroju, S., and Bhat, A. (2007). "Application of ArcGIS geostatistical analyst for interpolating environmental data from observations." *Environmental Progress*, 26(3), 220-225.

La Peyre, M. K., Reams, M. A., and Mendelssohn, I. A. (2001). "Linking actions to outcomes in wetland management: An overview of U.S. state wetland management." *Wetlands*, 21(1), 66-74.

Leenheer, J. A. (2004). "Comprehensive assessment of precursors, diagenesis, and reactivity to water treatment of dissolved and colloidal organic matter." *Water Science & Technology: Water Supply*, 4(4), 1-9.

Leibowitz, S. G., Wigington, P. J., Rains, M. C., and Downing, D. M. (2008). "Non-navigable Streams and Adjacent Wetlands: Addressing Science Needs Following the Supreme Court's Rapanos Decision." *Frontiers in Ecology and the Environment*, 6(7), 364-371.

Luo, W., Taylor, M. C., and Parker, S. R. (2008). "A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales." *International Journal of Climatology*, 28(7), 947-959.

Macdonald, M. A. (2007). "Rapanos V. United States and Carabell V. United States Army Corps of Engineers " *Harvard Environmental Law Review*, 31, 321-332.
Massicotte, P., and Frenette, J.-J. (2011). "Spatial connectivity in a large river system: resolving the sources and fate of dissolved organic matter." *Ecological Applications*, 21(7), 2600-2617.
Meeting, R. (2012). "Conversatoin with Dr. Timothy Matisiziw."
Miller, S. J., Wardrop, D. H., Mahaney, W. M., and Brooks, R. P. (2006). "A plant-based index of biological integrity (IBI) for headwater wetlands in central Pennsylvania." *Ecological Indicators*, 6(2), 290-312.
MODNR (2012). "Missouri DNR Watershed Pilot Program ", <http://dnr.mo.gov/omwi.htm>. (July 15, 2012).
Novotny, V., and Chesters, G. (1981). *Handbook of Nonpoint Pollution: Sources and Management*, Van Nostrand Reinhold, New York.
Owen, C. R. (1999). "Hydrology and history: land use changes and ecological responses in an urban wetland." *Wetlands Ecology and Management*, 6(4), 209-219.
Palmeri, L., and Trepel, M. (2002). "A GIS-based score system for siting and sizing of created or restored wetlands: Two case studies." *Water Resources Management*, 16(4), 307-328.
Park, S. Y., Choi, J. H., Wang, S., and Park, S. S. (2006). "Design of a water quality monitoring network in a large river system using the genetic algorithm." *Ecological Modelling*, 199(3 SPEC. ISS.), 289-297.
Peters, N. E., and Ratcliffe, E. B. (1998). "Tracing hydrologic pathways using chloride at the Panola mountain research watershed, Georgia, USA." *Water, Air, and Soil Pollution*, 105(1-2), 263-275.
Pontius, F. (2008). "Wetlands Mitigation Rules Finalized." *American Water Works Association*, 100(5), 12-16.
Rabah, P. K. J., Ghabayen, S. M., and Salha, A. A. (2011). "Effect of GIS Interpolation Techniques on the Accuracy of the Spatial Representation of Groundwater Monitoring Data in Gaza Strip." *Journal of Environmental Science & Technology*, 4(6), 579-589.
Rebelo, L. M., Finlayson, C. M., and Nagabhatla, N. (2009). "Remote sensing and GIS for wetland inventory, mapping and change analysis." *Journal of Environmental Management*, 90(7), 2144-2153.
Reiss, K. C., and Brown, M. T. (2007). "Evaluation of Florida palustrine wetlands: Application of USEPA levels 1, 2, and 3 assessment methods." *EcoHealth*, 4(2), 206-218.
Renner, R. (2005). "Environmental news: EPA's physical-chemical data criticized." *Environmental Science and Technology*, 39(13), 274A-275A.
Rhemtulla, J., and Mladenoff, D. (2007). "Why history matters in landscape ecology." *Landscape Ecology*, 22(0), 1-3.
Rinnan, A., Booksh, K., and Bro, R. (2005). "First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes." *Analytica Chimica Acta*, 537(1-2), 349-358.
Robertson, G. P. (1987). "Geostatistics in Ecology: Interpolating With Known Variance." *Ecology*, 68(3), 744-748.
Sanchez, N. P., Skeriotis, A. T., and Miller, C. M. (2013). "Assessment of dissolved organic matter fluorescence PARAFAC components before and after coagulation-filtration in a full scale water treatment plant." *Water Research*, 47(4), 1679-1690.
Semlitsch, R. D., and Bodie, J. R. (1998). "Are Small, Isolated Wetlands Expendable?" *Conservation Biology*, 12(5), 1129-1133.

Sim, A. (2011). "Wetland Microbial Diversity and Function: A Literature Review " PhD Dissertation University of Missouri- Columbia

Singh, K. P., Malik, A., Mohan, D., and Sinha, S. (2004). "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study." *Water Research*, 38(18), 3980-3992.

Smart, P. L., and Laidlaw, I. M. S. (1977). "EVALUATION OF SOME FLUORESCENT DYES FOR WATER TRACING." *Water Resources Research*, 13(1), 15-33.

Stave, K. A. (2001). "DYNAMICS OF WETLAND DEVELOPMENT AND RESOURCE MANAGEMENT IN LAS VEGAS WASH, NEVADA1." *JAWRA Journal of the American Water Resources Association*, 37(5), 1369-1379.

Stedmon, C. A., and Bro, R. (2008). "Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial " *Limnology and Oceanography: Methods* 572–579.

Stedmon, C. A., and Markager, S. (2005). "Tracing the Production and Degradation of Autochthonous Fractions of Dissolved Organic Matter by Fluorescence Analysis." *Limnology and Oceanography*, 50(5), 1415-1426.

Stedmon, C. A., Markager, S., and Bro, R. (2003). "Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy." *Marine Chemistry*, 82(3-4), 239-254.

Stein, E., Dark, S., Longcore, T., Grossinger, R., Hall, N., and Beland, M. (2010). "Historical Ecology as a Tool for Assessing Landscape Change and Informing Wetland Restoration Priorities." *Wetlands*, 30(3), 589-601.

Stoeckel, D. M., and Harwood, V. J. (2007). "Performance, design, and analysis in microbial source tracking studies." *Applied and Environmental Microbiology*, 73(8), 2405-2415.

Strahler, A. N. (1952). "Hypsometric (Area-Altitude) Analysis of Erosional Topography " *Geological Society of America Bulletin*, 63(11), 1117-1142.

Sun, Q. Y., Yu, X. X., Hu, S. P., Xiao, Y., Li, J. H., and Wu, J. (2008). "Extraction and application of hydrological information based on DEM in ArcGIS environment." *Beijing Linye Daxue Xuebao/Journal of Beijing Forestry University*, 30(SUPPL. 2), 144-147.

Svensson, T., Sandén, P., Bastviken, D., and Öberg, G. (2007). "Chlorine transport in a small catchment in southeast Sweden during two years." *Biogeochemistry*, 82(2), 181-199.

Thurman, E. M. (1985). *Organic geochemisty of natural waters.* , Martinus Nijhoffy/Dr W. Junk, Dordrecht, Boston

Thygesen, L. G., Rinnan, Å., Barsberg, S., and Møller, J. K. S. (2004). "Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area." *Chemometrics and Intelligent Laboratory Systems*, 71(2), 97-106.

Timmerman, J. G., Beinat, E., Termeer, K., and Cofino, W. (2010). "Analyzing the data-rich-but-information-poor syndrome in dutch water management in historical perspective." *Environmental Management*, 45(5), 1231-1242.

Timofeyev, M. A., Wiegand, C., Kent Burnison, B., Shatilina, Z. M., Pflugmacher, S., and Steinberg, C. E. W. (2004). "Impact of natural organic matter (NOM) on freshwater amphipods." *Science of The Total Environment*, 319(1–3), 115-121.

Tiner, R. W. (1994). "Wetland Defintions and Classifications in the United States " *United States Geological Survey Water Supply*, <http://water.usgs.gov/nwsum/WSP2425/definitions.html>. (August 11 2012).

Tiner, R. W. (2003). "Geographically isolated wetlands of the United States." *Wetlands*, 23(3), 494-516.
Tockner, K., and Stanford, J. A. (2002). "Riverine flood plains: present state and future trends." *Environmental Conservation*, 29(03), 308-330.
Toner, M., and Keddy, P. (1997). "River Hydrology and Riparian Wetlands: A Predictive Model for Ecological Assembly." *Ecological Applications*, 7(1), 236-246.
Turner, R. K., van den Bergh, J. C. J. M., Söderqvist, T., Barendregt, A., van der Straaten, J., Maltby, E., and van Ierland, E. C. (2000). "Ecological-economic analysis of wetlands: scientific integration for management and policy." *Ecological Economics*, 35(1), 7-23.
USDA (2012). "Wetland Indicator Status." <http://plants.usda.gov/wetland.html>. (June 30, 2012).
Van Lonkhuyzen, R. A., Lagory, K. E., and Kuiper, J. A. (2004). "Modeling the suitability of potential wetland mitigation sites with a geographic information system." *Environmental Management*, 33(3), 368-375.
Vepraskas, M. J., He, X., Lindbo, D. L., and Skaggs, R. W. (2004). "Calibrating hydric soil field indicators to long-term wetland hydrology." *Soil Science Society of America Journal*, 68(4), 1461-1469.
Vicente-Serrano, S. M., Saz-Sánchez, M. A., and Cuadrat, J. M. (2003). "Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature." *Climate Research*, 24(2), 161-180.
Wayne, C. (2003). "Suitability analysis with raster data." *ArcUser*, 6(April-September, 2-3), 54-57.
Whigham, D. F., Chitterling, C., and Palmer, B. (1988). "Impacts of freshwater wetlands on water quality: a landscape perspective." *Environmental Management*, 12(5), 663-671.
White, D., and Fennessy, S. (2005). "Modeling the suitability of wetland restoration potential at the watershed scale." *Ecological Engineering*, 24(4), 359-377.
Whittemore, R. C., and Beebe, J. (2000). "EPA'S BASINS MODEL: GOOD SCIENCE OR SERENDIPITOUS MODELING?1." *JAWRA Journal of the American Water Resources Association*, 36(3), 493-499.
Woodward, R. T., and Wui, Y. S. (2001). "The economic value of wetland services: A meta-analysis." *Ecological Economics*, 37(2), 257-270.
Yan, Y., Li, H., and Myrick, M. L. (2000). "Fluorescence fingerprint of waters: Excitation-emission matrix spectroscopy as a tracking tool." *Applied Spectroscopy*, 54(10), 1539-1542.
Yoo, E. H., and Kyriakidis, P. C. (2008). "Area-to-Point Prediction Under Boundary Conditions." *Geographical Analysis*, 40(4), 355-379.
Zhang, C., Luo, L., Xu, W., and Ledwith, V. (2008). "Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland." *Science of The Total Environment*, 398(1–3), 212-221.
Zhao, J., Fu, G., Lei, K., and Li, Y. (2011). "Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management." *Journal of Environmental Sciences*, 23(9), 1460-1471.