
SCHOOL of GRADUATE STUDIES
EAST TENNESSEE STATE UNIVERSITY

East Tennessee State University
Digital Commons @ East Tennessee State University

Electronic Theses and Dissertations | Student Works

12-2018

The Ecology of Fecal Indicators

Dennis A. Gilfillan
East Tennessee State Unviersity

Follow this and additional works at: https://dc.etsu.edu/etd

 Part of the Environmental Indicators and Impact Assessment Commons, Environmental Public Health Commons, and the Water Resource Management Commons

Recommended Citation

Gilfillan, Dennis A., "The Ecology of Fecal Indicators" (2018). *Electronic Theses and Dissertations.* Paper 3521. https://dc.etsu.edu/etd/3521

This Dissertation - Open Access is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.

The Ecology of Fecal Indicators

_____

A Dissertation

presented to

the faculty of the Department of Environmental Health Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy in Environmental Health Sciences

_____

by

Dennis Andrew Gilfillan

December 2018

_____

Phillip Scheuerman, Ph.D., Chair

Timothy Andrew Joyner, Ph.D.

Kimberlee Hall, Ph.D.

Ying Li, Ph.D.

Kurt Maier, Ph.D.

Keywords: Microbial Ecology, Machine Learning, Multivariate Statistics, Fecal Pollution

ABSTRACT

The Ecology of Fecal indicators

by

Dennis Gilfillan

Animal and human wastes introduce pathogens into rivers and streams, creating human health and economic burdens. While direct monitoring for pathogens is possible, it is impractical due to the sporadic distribution of pathogens, cost to identify, and health risks to laboratory workers. To overcome these issues, fecal indicator organisms are used to estimate the presence of pathogens. Although fecal indicators generally protect public health, they fall short in their utility because of difficulties in public health risk characterization, inconsistent correlations with pathogens, weak source identification, and their potential to persist in environments with no point sources of fecal pollution. This research focuses on characterizing the ecology of fecal indicators using both modeling and metabolic indicators to better understand the processes that drive fecal pollution. Fecal indicator impairment was modeled in Sinking Creek, a 303 (d) listed stream in Northeast Tennessee, using the ecological niche model, Maxent, for two different fecal indicators. While the use of Maxent has been well demonstrated at the macroscale, this study introduces its application to ecological niches at the microscale. Stream impairment seasonality was exhibited in two different indicators over multiple years and different resolutions (quarterly versus monthly sampling programs). This stresses the need for multiple year and month sampling to capture heterogeneity in fecal indicator concentrations. Although discharge is strongly associated with dissolved solutes, fecal indicator impairment was governed by other ecological factors such as populations of heterotrophic bacteria, enzyme activity, nutrient conditions, and other metabolic indicators. This research also incorporated metabolic indicators to characterize spatiotemporal variability in microbial community function, making connections to fecal and other pollution gradients. Communities differed in their ability to use a wide variety of substrates, and metabolic inhibition in sediments captured most of the interaction of aquatic and benthic communities. Sediment substrate activity was also indicative of degrees of pollution, suggesting that sediment is a potential reservoir for *Escherichia coli* in this stream, and there is possibility for resuspension, extended residence times, and increased

duration for exposure. This research highlights the benefit of using models and microbial indicators to better understand how environment shapes the niche of fecal indicators.

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee, especially Dr. Andrew Joyner and Dr. Phillip Scheuerman, for their guidance and good humor throughout the duration of this research. I am also grateful for Brian Evanshen and Chuck Patton for their support and advice during my time here at East Tennessee State University. Since my students had to deal with an overwhelmed instructor juggling a heavy work load, a special thanks goes out to them for their patience and positive attitude through my laboratories and lectures. I am thankful for the Research Development Committee at East Tennessee State University to have supported my data collection.

There are a few groups that have been supportive who exist outside of the University, and I would like to acknowledge these people as well.

- My family who have always trusted and supported my journey to find my weird. I love you and am so thankful to have you in my life.
- My friends who provided positive outlets during the good and bad times of the dissertation process. Thanks for the meals, the advice, the laughs, and most importantly, the shreds.
- All of my previous high school students who shaped my teaching style. Although it seems like another lifetime ago, the trials and tribulations of the high school classroom made me a stronger person, especially my “Experts in Algebra I.”
- Zach Boone for allowing me to chase my dream while leaving him in his senior year of running. I am glad to have experienced something unique coaching you, and look forward to hearing about your future endeavors.

TABLE OF CONTENTS

LIST OF TABLES

Page

Table

LIST OF FIGURES

Page

Figure

CHAPTER 1

INTRODUCTION, LITERATURE REVIEW, AND GOALS OF STUDY

Introduction

Animal and human wastes introduce pathogens into rivers and streams, creating human health and economic burdens. Pathogens represent the primary cause of impairment for United States' surface waters, impacting over 170,000 miles of rivers and streams (United States Environmental Protection Agency 2017). While direct monitoring for pathogens is possible, it is impractical due to the sporadic distribution of pathogens, cost to identify, and health risks to laboratory workers (Savichtcheva and Okabe 2006; Field and Samadpour 2007). To overcome these issues, fecal indicator organisms (FIOs) are used to estimate the presence of pathogens. These FIOs should exist whenever pathogens are present, be versatile in their use, not reproduce in the water column, and have an enteric origin (Cimenti et al. 2007; Maier et al. 2009). Elevated levels of FIOs should correlate with the presence of pathogens, protecting public health and identifying locations with sources of fecal pollution.

Impairment from pollution reduces a riverine system's utility, so strategies are needed to minimize exposure to harmful pollutants and improve water quality. The Clean Water Act outlines the process for identifying impairment, listing of polluted watersheds, and developing a total maximum daily load (TMDL) for a watershed to determine acceptable pollutant concentrations (United States Environmental Protection Agency 2001a). A TMDL includes the amount of pollutants from point sources, typically identified through the national pollutant discharge elimination system (NPDES), and nonpoint sources (Borsuk et al. 2002; Shirmohammadi 2006). Nonpoint sources represent a substantial challenge in managing fecal pollution, necessitating creative strategies for source appropriation to reduce fecal loadings and determine responsible parties (Duda 1993; Meays et al. 2004; Field and Samadpour 2007).

Although fecal indicators generally protect public health (Wade et al. 2003), they fall short in their utility because of difficulties in public health risk characterization, inconsistent correlations with pathogens, weak source identification, and their potential to persist in environments with no known sources of fecal pollution (Byappanahalli, Fowler, et al. 2003;

Savichtcheva and Okabe 2006; Field and Samadpour 2007; Yates 2007). This drives development of alternate indicators to improve on these weaknesses. A variety of alternate indicator organisms have been evaluated, and differentiation between human and animal sources have been suggested using host-specific microbes, genetic markers, functional characteristics such as carbon utilization, antibiotic resistance, and chemical markers (Scott et al. 2002; Cimenti et al. 2007; McLellan and Eren 2014). All of these can improve on the single indicator paradigm to monitor for impairment.

In addition to these novel biochemical indicators as a means to improve monitoring and management, the use of geographic data and statistical modeling can improve on the single indicator paradigm (Nevers et al. 2016). Utilizing geographic data can allow for causal inferences into land use for source tracking, determining natural and anthropogenic drivers of fecal pollution(Smith et al. 2001; Eleria and Vogel 2005; Coulliete et al. 2009; Vitro et al. 2017). The use of statistical modeling has been applied to fecal pollution in a variety of ways: optimizing source tracking (Brion and Lingireddy 1999; Brion et al. 2002; Belanche-Muñoz and Blanch 2008; Ballestè et al. 2010), identifying environmental factors contributing to FIO persistence (Wilkes et al. 2011; Piorkowski et al. 2013; Hall et al. 2014), and creating predictive models to estimate conditions that exceed water quality criteria (Kim and Grant 2004; Eleria and Vogel 2005; Gonzalez et al. 2012; Gonzalez and Noble 2014). Using geographic data and statistical modeling present flexible strategies that can inform decision-makers on the sources of fecal pollution, identify fate and transport mechanisms, and are applicable on a universal scale. Although the single indicator paradigm is inherently limited, simultaneously incorporating alternative indicators, source tracking techniques and creative modeling strategies can improve its utility, aiding in understanding and reducing impairment due to fecal pollution.

Public Health and Economic Burden of Pathogens

A variety of pathogens are associated with waterborne and foodborne illnesses, contributing to morbidity and mortality globally.(Pandey et al. 2014) A serious concern for decision-makers and stakeholders, pathogens represent the primary reason for impairment within the United States, contaminating over 170,000 miles of riverine systems(United States

Environmental Protection Agency 2017). More than 100 different types of pathogenic microorganisms exist in aquatic environments, and waterborne illness from fresh water originates from animal and human waste via the fecal-oral route of transmission, dermal contact, or inhalation of bioaerosols (Soller et al. 2015). Each pathway results in a variety of clinical manifestations, such as gastrointestinal illness, respiratory problems, fever, inflammation of brain and meninges due to brain consumption, organ damage, respiratory distress, necrotizing fasciitis, Naegleriasis, and even death in highly susceptible populations (Hofstra 2011; United States Environmental Protection Agency 2012).

Waterborne pathogens enter waterbodies through inadequately treated sewage, stormwater runoff, and various agricultural practices, but also can be naturalized members of microbial communities (Ferguson et al. 2003; Jamieson et al. 2004; Lasalde et al. 2005). These sources can be divided into two general types of pollution; point and nonpoint source. Point source pollution often derives from treated wastewater effluent and storm sewers, both which operate under the NPDES permitting process. Nonpoint sources are less characterized, but wildlife defecation, stormwater runoff, faulty septic systems, manure spreading and spraying, overflow from confined animal feeding operation waste lagoons, general livestock waste runoff, and other sources all contribute to microbial contamination (Savichtcheva and Okabe 2006; Boehm et al. 2009; Ibekwe et al. 2013). Pathogens that are of greatest concern exist in both recreational and drinking water, are highly infectious at low doses, possess traits of environmental resilience, survive for extended durations within nonhost environments, and are resistant to some water treatment processes (Field and Samadpour 2007; Ashbolt et al. 2010; Soller, Schoen, et al. 2010). For example, protozoan parasites are notorious for disease outbreaks in swimming pools due to their resistance to chlorination, and certain pathogens are able to survive in drinking water systems (Ashbolt et al. 2010). These require alternate methods of elimination for drinking water, increasing cost of treatment to prevent illness. Disease outbreaks in the United States are driven by many types of pathogens, including bacteria, protozoa, and viruses (Arnone and Walling 2007).

Waterborne disease outbreaks create both public health and financial burdens on communities. Globally, an estimated 13 million people die each year from waterborne pathogens, and waterborne illness causes approximately 900,000 incidents of disease and 900 deaths due to exposure to contaminated water (Arnone and Walling 2007). From 1986 to 2000, there were 95 outbreaks associated with recreational waters in the United States, and 48 outbreaks associated with drinking water (Arnone and Walling 2007). These outbreaks resulted in 5,095 total cases of disease occurrence due to recreational exposure, and 437,082 cases of disease due to drinking water contamination (Arnone and Walling 2007). During this time, the largest waterborne disease outbreak occurred in Milwaukee, Wisconsin, with faulty filtration processes leading to cryptosporidiosis in over 403,000 residents (MacKenzie et al. 1994). These outbreaks also create economic burdens due to medical costs and loss of productivity. In the 1993 Milwaukee *Cryptosporidium* outbreak, the estimated costs of illness were $96.2 million; this included $31.7 million in medical costs and $64.6 million in productivity losses (Corso et al. 2003). The estimated individual costs for illness ranged from $116 to $7,808, depending on severity of the illness (Corso et al. 2003). The burdensome nature of disease outbreaks requires diligent monitoring of pathogens to reduce this impact.

Characterizing human health risks associated with waterborne pathogens is difficult because of differences in risk based on source, degree of exposure, and individual susceptibility. Human and nonhuman wastes possess unique assemblages of microorganisms; as a result, the risk of infection is dependent on the distribution, diversity, and number of pathogens (Soller, Schoen, et al. 2010). Animal and human sources of fecal pollution also have different pathways for exposure; pathogen introduction in wastewater effluents is continuous, with increases of poorly treated wastewater during rain events, but pathogen loading from manure and other animal sources is largely stormwater driven (Soller et al. 2015). Susceptibility to infection is a spectrum, and risks are higher for immunocompromised populations, the elderly, and children (Nwachuku and Gerba 2004; de Man et al. 2014). These factors make identifying universal conditions of elevated risk challenging, but essential to consider in monitoring for pathogens to protect human health.

Direct monitoring for pathogens would be ideal to mitigate human health risk, but the sporadic distribution of pathogens, their diversity in type and number, costly identification procedures, and risks to exposed laboratory workers makes this an impractical option (Savichtcheva and Okabe 2006; Cimenti et al. 2007; Field and Samadpour 2007; Maier et al. 2009). Indicator organisms are used instead to estimate the relative density of enteric pathogens in water bodies, alleviating some of the difficulties of direct monitoring (Cimenti et al. 2007). FIOs should exist whenever pathogens are present, be versatile in their use, not reproduce in the water column, and have an enteric origin (Savichtcheva and Okabe 2006; Cimenti et al. 2007; Yates 2007). Current water quality standards for pathogens need to be scientifically defensible, universal in implementation across all societies and geographies, and properly protect human and ecosystem health (Boehm et al. 2009; Maier et al. 2009). Table 1-1 identifies some of the characteristics of an ideal fecal indicator.

Indicators historically used have been total coliforms, fecal coliforms, and fecal streptococci (Yates 2007). The indicator selected is highly dependent on potential use of the water bodies; for example, in Tennessee, drinking water is monitored using total coliform presence while recreational waters are monitored using *E. coli* (Tennessee Department of Environmental and Conservation 2015a). Nationally, recreational water quality criteria were updated in 2012 to support the use of enterococci for fresh and marine waters and *E. coli* for fresh waters (Boehm et al. 2009). Elevated FIO concentrations were correlated to gastrointestinal illness rates of either 32 or 36 per 1000 people for *E. coli*. In 2012, a statistical threshold value was also introduced, which is approximately the 90^{th} percentile of the distribution of the water samples taken (United States Environmental Protection Agency 2012). In general, FIOs have been successful in alerting populations when potential for gastrointestinal illness is present (Wade et al. 2003), and can identify potential impairment due to human and animal wastes.

Table 1.1. Characteristics of an ideal fecal indicator. Adapted from Maier et al. (2015) and Cimenti et al. (2007)

	Characteristic
a.	The indicator should exist in the presence of fecal contamination, and be absent when there is not fecal contamination
b.	If the indicator is a microorganism, it should be the member of the gut microflora of warm-blooded animals, and should not grow in the environment
c.	If the indicator is a chemical substance it should be associated with fecal discharges
d.	The indicator should be useful for all types of waters
e.	The concentration of the indicator should be greater than or at least equal to the amount of pathogens

f.	The indicator should persist in the environment for a longer time than the most resilient pathogen
g.	The quantification of the indicator should be faster, easier to perform, and more sensitive than quantification of pathogens
h.	The quantification of the indicator should be less expensive than the quantification of pathogens

Management of Impaired Watersheds

The need for clean water sources and the control of water pollution has been a part of public policy since the Federal Water Pollution Control Act of 1948 (Adler et al. 1993; Copeland 1999). Growing public awareness and concern for managing water pollution led to the metamorphosis of the Federal Water Pollution Control Act into what is now known as the Clean Water Act of 1972 (Adler et al. 1993). The amendments included establishing a basic structure for regulating pollutant discharges, setting wastewater and surface water quality standards, and recognized the need to address the critical problems posed by nonpoint source pollution (Copeland 1999). The Clean Water Act requires the regular monitoring of surface waters to identify potential contamination, and creation and maintenance of the 303(d) list for impaired waters.

As required by section 303(d) of the Clean Water Act, water bodies identified as impaired are listed. In the 2012 updated recreational water quality criteria for fecal pollution, states are given governance concerning the number of samples to be taken within a 30 day period to determine impairment (United States Environmental Protection Agency 2012). Tennessee uses a 5-sample 30-day geometric mean to determine pathogen impairment. Once a stream has been identified as impaired, a TMDL is developed (Hall et al. 2014; United States

Environmental Protection Agency 2017). A TMDL characterizes the point and nonpoint sources of pollution in a watershed, incorporates a margin of safety to account for variability, and is used to guide remediation efforts to return waterbodies to their intended use (United States Environmental Protection Agency 2001a; United States Environmental Protection Agency 2017). A reduction in load is required in many of these watersheds, and this is determined through monitoring, modeling, or a combination of the two to fully characterize the watershed sources of impairment (United States Environmental Protection Agency 2001a). States and municipalities develop and implement TMDLs with varying degrees of success (United States Environmental Protection Agency 2012). States are given creative leeway for development of TMDLs; some operate by designing them at the stream section level, whereas states like Tennessee create watershed TMDLs. The concept of a watershed could be defined in terms of drainage ditches, small farm ponds, or even large rivers; in the case of Tennessee 8-digit hydrologic unit codes are used as the demarcation of a watershed (Cohen and Davidson 2011). These TMDLs are designed to appropriate sources of pollution in a watershed, and effectively plan best management practices (BMPs) to mitigate these pollutants.

The United States Environmental Protection Agency (EPA) distributes a variety of aquatic models for developing TMDLs and pollutant monitoring. In many of these models, surface flow is the only source of runoff into streams. Although landscape runoff has been shown to contribute the greatest microbial loads to water bodies, hyporheic exchange and subsurface transport of microbes can be significant, needing consideration in certain environments (Hunter et al. 1992; Jamieson et al. 2004). Hydrograph modeling of fecal indicators is best used in extreme conditions such as flooding or drought conditions (Ghimire and Deng 2013); fate and transport of FIOs is not consistent with the build-up/wash-off theory used in most distributed watershed models (Benham et al. 2006; Surbeck et al. 2006; Drummond et al. 2015). Wash-off from land surfaces is unlikely to be at consistent land use specific rates, and differential survival of FIOs based on source are realistic issues, but are not considered in the current modeling practices (Surbeck et al. 2006). Most water quality models treat microorganisms as free-floating colloids with neutral buoyancy, despite the consensus that bacteria associate with sediment in stream environments (Jamieson, Doug M. Joy, et al.

2005). Although process-based models have been used for development of TMDLs with varying degrees of success, they are complex, cumbersome to use, and require a great deal of data to calibrate (Borah and Bera 2003; Shirmohammadi 2006). In addition, the complexity of the models does not necessarily improve simulation accuracy (Stow et al. 2003). These are difficulties and shortcomings of process models, stressing the need for alternate approaches to infer fate and transport processes of FIOs as well as the creation of simple but accurate models in data-sparse watersheds.

Limited resources in the forms of finances, staff, and water quality modeling expertise are prevalent within all levels of the TMDL program. Financial resources are the strongest limitation because they reduce the amount of staff necessary to implement program goals, and reduce the amount of data that can be collected within impaired watersheds. Currently in the United States, each of the 10 regions is given an allotment to distribute to their member states, but this currency is not enough to fulfill the exhaustive requirement set out by the Clean Water Act. External factors create imbalances within the regions, especially in watersheds with vested interests from industry (Neilson and Stevens 2002; Maguire 2003). For example, the EPA regions in the eastern United States have considerably more funds due to local revenue from large numbers of point source dischargers desiring appropriate load allocations to reduce the financial burden of treatment (Neilson and Stevens 2002). Resource availability also can be limited by the variation in the number of listed streams within state, differential viewpoints on the need for environmental protection, and local political climates (Shirmohammadi 2006). Politicians have been critical of the TMDL program, at some points needing scientific proof that the TMDL process was a valid approach to reduce water pollution (Neilson and Stevens 2002; Elshorbagy et al. 2005; Shirmohammadi 2006). Support of this program is needed from all levels, and the lack of adequate resources nationwide makes proper implementation and development of TMDLs challenging.

Another difficulty of developing TMDLs is the volume of impaired watersheds within the United States. To alleviate the stress of developing an exhaustive amount of unique TMDLs, many states have adopted a watershed approach as a feasible alternative (Elshorbagy et al. 2005). The watershed approach is a coordinated framework designed to restore aquatic

systems and protect human health more effectively (Cohen and Davidson 2011). Encouraging collaborations between point source dischargers and citizens in a collaborative atmosphere provides an environment that recognizes dischargers and citizens as integral parts of the solution, focusing on long-term comprehensive solutions for both point and nonpoint sources of pollutants. In Tennessee, 55 watersheds are identified using 8-digit hydrologic unit codes, and these are evaluated on a five year cycle (Tennessee Department of Environmental and Conservation 2015b). The watershed approach fosters a better understanding of how physical and biochemical changes affect watersheds, allowing agencies and citizens to focus on those solutions most likely to be effective.

A successful strategy in TMDL implementation is active involvement of stakeholders within all program stages (Neilson and Stevens 2002; Maguire 2003; Elshorbagy et al. 2005). When resources to properly monitor and manage watersheds are scarce, having surrounding community involvement invokes environmental groups to help achieve the objectives of the TMDL (Neilson and Stevens 2002). Involvement of stakeholders influences resource availability, with areas of support advocating for more data collection, and going beyond the limitations of the state (Maguire 2003). Stake-holders and the public should be brought in early to TMDL development to voice their concerns, and participation should be encouraged to make decision-makers accountable for a successful TMDL. Increased public participation should be welcomed throughout the process, from development to implementation of best management practices.

Although the TMDL program is admirable in its goals, limitations abound within the program. Modeling FIO fate and transport is challenging due to flow-independent survival, differential adaptive mechanisms based on source, and other ecological factors (Byappanahalli, Fowler, et al. 2003; Benham et al. 2006; Surbeck et al. 2006; Surbeck et al. 2010; Berthe et al. 2013). Compounding these issues are the lack of resources for appropriate monitoring, modeling and remediation, weakening the ability to improve water quality in the United States (Neilson and Stevens 2002; Elshorbagy et al. 2005; Shirmohammadi 2006). Developing watershed-sized TMDLs optimizes efforts for maximum strategic impact, while stake-holder involvement stimulates cooperation, allocation of additional resources, and improve potential for successful implementation (Tennessee Department of Environmental and Conservation

2015b). However, there is a need for alternate modeling strategies to evaluate source in data-sparse watersheds to better characterize loads and mechanisms driving fecal impairment, and a call for policies that encourage cooperation among decision-makers and stake holders within a watershed.

The Fecal Indicator Paradigm

Characterizing Public Health Risks

Identifying public health risks based on FIOs is difficult because of geographic variability and the site-specific nature of epidemiologic studies. Moe et al. (1991) found that although significant differences in illness rates could be identified from highly contaminated water (>1000 *E. coli* per 100 mL), disease threshold risk varies based on local climate and cultural conditions (Moe et al. 1991). Colford et al. (2007) found that fecal indicator bacteria did not predict health effects at a marine bathing beach, cautioning that their results may be site-specific, and suggesting that results were related to the lack of human sources and negative detection of enteric viruses (Colford et al. 2007). Fujioka et al. (2015) stated that the 2012 Recreational Water Quality Criteria did not improve strategies to assess bathers' health risks in all types for recreational waters (Fujioka et al. 2015). This may be attributed to the fact that these criteria were based on 7 marine bathing beaches, 2 freshwater beaches, and no riverine systems (United States Environmental Protection Agency 2012; Fujioka et al. 2015). However, it is infeasible to collect enough epidemiologic data in watersheds to identify regionally specific disease rates, and it is challenging to categorize an exposed population in many of recreational water bodies. However, Wade et al. (2002) found that although significant heterogeneity existed in epidemiologic studies, FIOs such as Enterococci and *E. coli* were found to be consistent predictors of elevated gastrointestinal illness in multiple geographic regions.(Wade et al. 2003) This supports their use as a FIO in light of the issues of site specific results, geographic variability, and local climate and culture.

Alternate Indicators and Correlations with Pathogens

The diverse ecology of pathogens makes it difficult to use a single indicator to mimic all pathogens, increasing the challenge to find consistent relationships between pathogens and levels of FIOs. This has spurred interest in the development of alternate indicators for use in watersheds, improving on the bacterial single indicator paradigm. Alternate indicators include fecal anaerobes, viral indicators, and chemical compounds. Each of these has certain strengths as well as inherent weaknesses, and some need further testing to characterize their utility.

The fecal anaerobes *Bifidobacterium* and *Bacteriodes* have been suggested as potential indicators but do not survive as long as *E. coli* in the environment, meaning they only indicate recent microbial contamination (Kreader 1998; Savichtcheva and Okabe 2006). One benefit of *Bacteriodes* and *Bifidobacterium* is that certain species are host specific (Kreader 1995; Bonjoch et al. 2004; Simpson et al. 2004). However, only a few studies have used these indicators, and there is need for studies at larger scales that compare these alternate indicators with specific pathogens and waterborne disease (Savichtcheva and Okabe 2006). *Clostridium perfringens* has been implemented as a fecal indicator for sewage contaminated streams, ocean environments, and sea waters (Bisson and Cabelli 1980; Savichtcheva and Okabe 2006). Because *C. perfringens* is resistant to environmental stress in comparison to other indicators, it represents one of the most conservative indicators of fecal pollution (Davies et al. 1995). However, *C. perfringens* is useful for fate determination of sewage and assessment of effectiveness of disinfection in drinking water systems (Payment and Franco 1993; Payment et al. 2000). One criticism is of *C. perfringens* is their extended viability in aquatic sediments; they are able to be detected long distances from fecal discharges, indicating either remote or old fecal pollution (Sorensen et al. 1989; Desmarais et al. 2002).

Bacterial indicators do not serve well as indicators for viruses because of their different ecologies. Bacteriophages infecting *Bacteriodes fragilis* and coliphages (F-specific RNA coliphage) have been suggested as two potential viral indicators (Borrego et al. 1987; Borrego et al. 1990; Chung and Sobsey 1993). *B. fragilis* bacteriophages were detected in human samples but not in animal samples, hinting at use for source tracking (Tartera and Jofre 1987). Survival of these bacteriophages is comparable to or better than enteric viruses in surface

waters and occur in greater numbers (Tartera et al. 1988; Chung and Sobsey 1993). Difficulty in recovery of this bacteriophages from waters without human sources limits its usefulness, warranting further methodology development (Savichtcheva and Okabe 2006) Coliphages have different serotypes for animal and human feces that could be used for source tracking.(Scott et al. 2002) Male-specific (F+) coliphages represent promising viral models because of their physical resemblance to human enteric viruses, their stability in aquatic environments, and their resistance to water treatment processes (Chung and Sobsey 1993; Sinton et al. 2002; Savichtcheva and Okabe 2006). F-specific RNA coliphage is also suggested as a useful viral indicators in freshwaters, where they have the highest environmental resistance (Havelaar and Pot-Hogeboom 1988; Sinton et al. 2002). F-specific coliphages are promising indicators, but there is an urgent need to simplify the methodology to concentrate and recover these viruses in environmental samples (Savichtcheva and Okabe 2006; Field and Samadpour 2007). There is also a need for more complete and detailed genetic characterization of different coliphage groups (Savichtcheva and Okabe 2006).

In environments in which traditional and alternate indicators might exist as natural microflora, fecal organic compounds such as fecal sterols could be used as alternative indicators (Dutka et al. 1974; Isobe et al. 2002; Isobe et al. 2004). Coprostanol is one of the major fecal sterols excreted by humans and animals; it is microbially degraded under aerobic conditions with half-lives of less than 10 days, suggesting the presence of fresh fecal pollution (Isobe et al. 2002). Other structurally related fecal sterols could provide conclusive information concerning source (Leeming et al. 1996; Isobe et al. 2002; Isobe et al. 2004). An issue with coprostanol and other fecal sterols is that these chemicals are easily incorporated into sediments where they degrade relatively slowly (stable for 450 days at $15^{O}C$) (Isobe et al. 2002). Future studies also need to evaluate host specificity, detection limits, and correlation with known pathogens (Field and Samadpour 2007).

While the alternate indicators hold promise to improve on the single indicator paradigm, current research questions the utility of traditional and alternate indicators to correlate with pathogens (Savichtcheva and Okabe 2006; Wu et al. 2011). Coliform groups typically demonstrated poor correlations with pathogens, partially explained by different

survival and persistence patterns in diverse environments such as rainforests and grey water (Carrillo et al. 1985; Ottoson and Stenström 2003). High densities of FIOs have been correlated with *Salmonella* spp., and the persistence of these pathogens was similar in both marine and freshwater (Morincigo et al. 1989). However, there is debate about the efficacy of the use of fecal coliforms to identify the occurrence of *Salmonella* (Savichtcheva and Okabe 2006). FIOs do not correlate with the protozoan pathogens *Cryptosporidium* and *Giardia* (Ferguson et al. 2003). Harwood et al. (2005) evaluated the use of total coliform bacteria, fecal coliform bacteria, *C. perfringens*, and F-specific coliphages to predict presence of pathogens at wastewater reclamation facilities (Harwood et al. 2005). Although pathogens were also detected, no strong correlations were identified with the indicators. Because of the sporadic distribution of pathogens, many correlation studies have insufficient data regarding both the number of samples collected and the number of positive samples for pathogens (Wu et al. 2011). The search for an indicator that reliably correlates with pathogens is on-going, but requires exhaustive direct monitoring of pathogens to be successful; this might be a faulty approach because of resource constraints in certain areas.

In addition to the pursuit of alternate indicators of fecal pollution, rapid indicators using molecular techniques are being developed to provide decision-makers near real-time estimations of levels of traditional FIOs. Wade et al. (2006) evaluated the performance of rapid quantitative polymerase chain reaction (qPCR) for detecting levels of *Bacteroides* and Enterococcus at two Great Lakes recreational beaches (Wade et al. 2006). Enterococcus concentrations were found to correlate with gastrointestinal illness rates at both beaches; however, mixed results were found for the *Bacteroides* species at both beaches (Wade et al. 2006). Gonzalez and Noble (2014) developed predictive models to compare how environmental conditions predicted both qPCR and cultured-based methods of Enterococcus and *E. coli* (Gonzalez and Noble 2014). While qPCR models showed high accuracy when applied to management decisions, inhibition was an issue that confounded the results; the authors developed an inhibition regression model to address these concerns (Gonzalez and Noble 2014). Although qPCR methods have been approved by the United States as an alternate indicator to be used in conjunction with cultured-based methods, there is a need for further

evaluation of these methods in multiple geographic contexts (United States Environmental Protection Agency 2012).

High throughput sequencing techniques, also known as next generation sequencing (NGS), present powerful approaches to shift the paradigm of microbial water quality management from fecal indicator control to assessing true pathogenic risk in watersheds. Ibekwe, Leddy, and Murinda (2013) used pyrosequencing (a NGS technique) to observe pathogen presence in a variety of urban and agricultural environments (Ibekwe et al. 2013). The results indicate that within this watershed, potential pathogens represented the greatest percentage of total operational taxonomic units in urban runoff water(7.94 %), agricultural runoff sediment (6.52%), and recreational park sediment (6.00%)(Ibekwe et al. 2013) NGS studies have been used to discover new indicators of sewage contamination, identifying bacterial taxa associated with human sources (McLellan et al. 2010; Unno et al. 2010; Cai and Zhang 2013). NGS has been applied to multiple drinking water safety studies, including source waters (Chao et al. 2013), various stages of the drinking water treatment and distributions processes (Shi et al. 2013), and investigation of the effects of biofilms on microbial diversity(Huang et al. 2014). The NGS methodology needs to be fine-tuned and standardized to make application of these techniques ubiquitous for indicator use (Tan et al. 2015), but these represent an exciting transition in indicators of fecal pollution.

Source Identification

The ubiquitous nature of *E. coli* and Enterococcus in the guts of warm-blooded animals makes them impractical for source identification, necessitating development of source tracking techniques (Yates 2007; McLellan and Eren 2014; Blount 2015). Fecal source tracking is important for the following reasons; it helps investigate causes of high levels of FIOs (Kirschner et al. 2017), identifies potential pathogens that may exist within a watershed (Ibekwe et al. 2013), and assists in estimating human health risk from exposure to pathogens (Field and Samadpour 2007; Soller, Schoen, et al. 2010; Soller et al. 2014). The principle underlying fecal source tracking is the assumption that different sources of fecal waste have unique, detectable identifiers that connect to the host (Savichtcheva and Okabe 2006; Field and Samadpour 2007).

Varieties of culture independent and dependent techniques have been suggested throughout the literature, some requiring a library to be developed for their use. These can be further divided into culturing, phenotypic, genetic, and chemical methods (Cimenti et al. 2007). Some examples of these are shown in Table 1-2.

Although the use of single FIOs does not usually identify source, some researchers still use FIO ratios to infer source and age of pollution. The application of ratios of fecal coliforms to fecal *streptococci* have historically been used, with ratios greater than 4 indicating human fecal matter and less than 0.7 indicating animal fecal matter. However, these ratios or shifts in ratios are not recommended for universal application due to different survival rates between coliforms and *streptococci* that cause complex changes in ratios over time (Scott et al. 2002; Simpson et al. 2002). Other microorganisms that can be cultivated for source-tracking include previously mentioned alternate indicators *Bifidobacteria* spp. and *Bacteroides* spp (Carrillo et al. 1985; Kreader 1995; Kreader 1998; Savichtcheva and Okabe 2006; Field and Samadpour 2007). Bacteriophages of *B. fragilis* have been used as a human specific viral indicator because they are more persistent in water than *B. fragilis* (Tartera and Jofre 1987; Cimenti et al. 2007). Despite their high specificity in source tracking, there is uncertainty associated with overall reliability of this source tracking mechanisms (Sinton et al. 1998; Maier et al. 2009) .Other phages used for human and animal source tracking are the F-RNA phages; subgroups II and III have been isolated only in human wastes, while subgroup I was only found in non-human mammals (Calci et al. 1998; Cole et al. 2003; Cimenti et al. 2007).The main issue with these phages is the detection methods are highly complex (Cimenti et al. 2007).

Phenotypic methods hold potential for source tracking because different functional characteristics are exhibited by lineages of microbes from different hosts. These dissimilar metabolic characteristics are because unique environmental factors control microbial populations within the host (Scott et al. 2002). One of the primary drawbacks of phenotypic methods is multiple microbial species can show similar responses to biochemical gradients, potentially confounding unique source-specific fingerprints, but utilizing multiple phenotypic characteristics increases the ability to create unique physiognomies (Field and Samadpour

2007). Common phenotypic methods include antibiotic resistance analysis (ARA), serotyping, and carbon substrate utilization profiles.

ARA can be used to differentiate between bacteria based on varying responses to antibiotic treatments. Differential uptake patterns of antibiotics and other pharmaceuticals elicit unique patterns in animal and human microflora, giving a characteristic fingerprint to identify fecal sources (Cimenti et al. 2007). A simple but time consuming procedure, isolates are plated on media with increasing concentrations of antibiotics and a profile is generated that can be compared to known profiles typical of the strains in question (Whitlock et al. 2002; Wiggins et al. 2003). These have been used to evaluate indicators of pig manure application (Huysman et al. 1993), fecal *streptococci* strains (Wiggins 1996), and Enterococci (Graves et al. 2002; Booth et al. 2003); however, the need for a database is the strongest limitation to this method (Cimenti et al. 2007).

Serotyping involves identifying the presence of different somatic antigens and has been successfully used to fingerprint non-overlapping serotypes from different fecal sources (Parveen et al. 2001). The ability of different fecal bacteria to utilize multiple sources of carbon has also been used to discriminate between Enterococcus species (Hagedorn et al. 2003). Community Level Physiological Profiles (CLPPs) based on these substrates patterns can be used to identify over 2000 species of microorganisms using the Biolog software (Cimenti et al. 2007). Both of these methods suffer from the need to have a geographically specific database, and additional studies to evaluate the overall effectiveness (Field and Samadpour 2007).

Genotypic profiles of enteric bacteria can be used to discriminate between sources of fecal pollution (Cimenti et al. 2007; Field and Samadpour 2007; McLellan and Eren 2014). The genetic marker that is analyzed needs to be either host-specific or generic; non-specific indicators, however, must have characteristic DNA fingerprints to discern source. Two popular methods of these genetic methods are ribotyping and pulse field gel electrophoresis (PFGE). Ribotyping involves examining the rRNA in each bacterial isolate using probes after treating genomic DNA with restriction endonuclease (Scott et al. 2002). This technique has been used to discriminate human and animal sources of *E. coli* (Parveen et al. 1999; Carson et al. 2001), and identify genetic variations in wildlife and geographic locales (Hartel et al. 2002; Scott et al.

2003). PFGE is the process of separating DNA fragments by alternating an electric current in more than one direction to obtain a DNA fingerprint of bacterial isolates (Myoda et al. 2003). Both of these are time-consuming to develop a geographic-specific database of isolates, but are promising source tracking methods based on comparison testing (Field and Samadpour 2007).

Molecular assays, either as specific genetic markers or DNA extracted from a water sample, have been proposed as improved source tracking indicators because there is no intervening culturing step; this speeds up the source tracking process and allows access to markers that would be difficult or impossible to detect using culturing methods. Assays of specific marker genes using a polymerase chain reaction (PCR), referred to as host-specific PCR, can be used to identify sources and these genes can include anaerobic bacterial genes, toxin or virulence genes, and host mitochondrial sequences. Host-specific PCR methods for viruses have been developed for humans (Shanks et al. 2007), pigs (Jiménez-Clavero et al. 2003; Hundesa et al. 2006), and cattle (Fong and Lipp 2005); however, the enteroviruses that infect bovine species are not host-specific because they are also found in deer, sheep, horse and geese (Jiménez-Clavero et al. 2005). Although these methods have promise for source-tracking viral contamination, they require large samples that, when concentrated, can also concentrate PCR-inhibiting substances, interfering with detection (Jiang et al. 2001; Surbeck et al. 2006). This can be compensated for by using nested PCR, but this compromises quantitative detection (Jiang et al. 2001; Maluquer de Motes et al. 2004).

Many fecal anaerobic bacteria have host-specific distributions and exist at much higher densities than coliform species and enterococci (Kreader 1995; Savage 2001), but these anaerobes are not commonly used as indicators due to difficulty in cultivating bacteria from the genera *Bifidobacterium* and *Bacteroidales* (Eckburg et al. 2005; Field and Samadpour 2007). These anaerobes are expected to have minimal reproduction in secondary habitats such as water bodies, but the advent of molecular detection allows for host-specific markers such as ribosomal RNA genes and protein gene targets.(Shanks et al. 2006; Shanks et al. 2007) Currently *Bacteroidales* host-specific PCR primers can identify feces from ruminants, humans, dogs, pigs, horses, and elk (Bernhard and Field 2000; Dick, Bernhard, et al. 2005; Dick, Simonich, et al.

2005; Okabe et al. 2007), with comparable detection limits to *E. coli* (Bernhard and Field 2000; Dick and Field 2004). These assays are geographically stable, have been used in North America (Kreader 1995; Bernhard and Field 2000), northern Europe,(Seurinck et al. 2005), Japan(Okabe et al. 2007), Hawaii(Betancourt and Fujioka 2006), and New Zealand (Gilpin et al. 2003). These host-specific markers also correlate well with sewage and FIOs(Dick and Field 2004) as well as some zoonotic pathogens.(Walters et al. 2007) Limitations of this approach include lack of wildlife host-specific markers and horizontal transfer of fecal bacteria of organisms in close contact such as humans and their pets (Dick, Bernhard, et al. 2005). Host-specific methods have also been developed for *Bifidobacterium* species (Bonjoch et al. 2004), but studies have shown mixed sensitivity and their limited environmental persistence reduces their usefulness as an indicator (Bernhard and Field 2000).

Host-specific assays of toxin genes associated with *E. coli* and enterococci are suggested as potential source tracking tools because these toxin-containing strains occur worldwide, providing geographic stability (Field and Samadpour 2007). Toxins identified for *E. coli* include the human-specific STIb toxin (Oshiro and Olson 1998), the pig-specific STII toxin (Khatib et al. 2003), and the cattle-specific LTIIa toxin (Khatib et al. 2002); additionally, the human specific virulence gene (esp) has been identified for *Enterococcus faecium* (Scott et al. 2005). One drawback of this method is that the target genes are rare, requiring enrichment of FIOs (Scott et al. 2005), nested PCR, or magnetic bead capture (Tsai et al. 2003). Detection of these genes is semi-quantitative, and if enrichment is used, it becomes a culture dependent technique (Field and Samadpour 2007). Horizontal gene transfer is also a concern of this method, lowering sensitivity of detection (van den Bogaard et al. 2002). Mitochondrial gene sequences from blood and intestinal cells in theory make excellent host-specific targets because these cells are not found in multiple hosts and cannot spread among species, except transiently after meat consumption (Martellini et al. 2005; Caldwell et al. 2007). However, Martellini et al. 2005 had problems with specificity and detection limits in an initial study (Martellini et al. 2005), and a follow-up study by Caldwell et al. 2007 demonstrated that large samples sizes (at least 0.2 g of feces per 100 mL of water) are required to correctly discriminate between human, pig, and

bovine mitochondrial gene sequences (Caldwell et al. 2007). This amount of feces within samples would be difficult to find in all but the most polluted waters.

Chemicals markers can be used to assess human versus nonhuman sources in a watershed. Chemicals used include caffeine (Daneshvar et al. 2012), fecal sterols (Sinton et al. 1998; Black et al. 2007; Fahrenfeld et al. 2016), laundry brighteners (Hayashi et al. 2002), human pharmaceuticals (Daneshvar et al. 2012), fragrances (Standley et al. 2000; Peck and Hornbuckle 2004), long-chain alkyl benzenes (Sinton et al. 1998; Martins et al. 2002), and animal growth promotors (Boxall et al. 2004; Zhou et al. 2013). A combined index of caffeine and fragrances can indicate the presence of human sewage, while a ratio of certain steroids can identify agricultural and wildlife inputs (Standley et al. 2000). Sterols have been used to suggest human, dog, and bird fecal impacts in Australia (Suprihatin et al. 2003), and relative concentrations of two sterols, 24-ethylcoprostaonol and coprostanol, reliably discriminated between human and non-human sources (Blanch et al. 2006). Although chemical indicators hold promise as culture independent source tracking methods, the spread, transport, and persistence of these chemicals may not correlate with pathogens and FIOs (Field and Samadpour 2007). Ecological and physical factors affect survival of microorganisms and will impact chemical indicators differently, such as settling, insolation, UV irradiation, nutrient conditions, and grazing (Düreth et al. 1986; Sinton et al. 2002; Whitman et al. 2004; Brookes et al. 2005).

Table 2.2. Types of fecal source tracking techniques.

Group of source tracking techniques	**Examples**
Culturing methods	Fecal coliforms-Fecal streptococci ratios and shifts
	Fecal streptococci species identification
	Bifidobacteria spp.
	Bacteroides spp.
	Bacteriophages
Phenotypic methods	Antibiotic resistance

	Multiple antibiotic resistance analysis
	Serogrouping
	Community level physiologic profiling using carbon utilization
Genetic methods	Ribotyping
	Pulse-field gel electrophoresis
	PCR
	Host-specific molecular markers
Chemical methods	Fecal sterols
	Caffeine
	Fragrances
	Human pharmaceuticals
	Long-chain alkyl benzenes
	Animal growth promotors

Each of these source tracking techniques has their strengths and weaknesses, but much of the problem in source tracking is within validation and standardization. Many methods have not extended past the proof of concept, feasibility, or biological likelihood stage of development, and few comparative studies exist that identify the best performing source tracking tools (Field and Samadpour 2007). There is a limited amount of both comparative and blind sampling validation for many of these methods, whether culture dependent or not (Field and Samadpour 2007). The Southern California Coastal Water Research Project (SCCWRP) and the US EPA participated in a comparison study in 2003 to assess effectiveness of source tracking techniques across multiple labs. Identical, blind samples containing human, cattle, dog, or gull feces, sewage/human wastewater, or mixtures of each were given to several labs compare accuracy between methods (Field et al. 2003; Harwood et al. 2003; Myoda et al. 2003; Noble et al. 2003; Stoeckel and Harwood 2007). Methods were assessed based on correct classification of samples, quantification of fecal contributions, and ability to handle different matrices. Only ribotyping, PFGE, and host-specific PCR were notably accurate within this study, but none of the

methods identified all the sources in every sample (Field and Samadpour 2007; Stoeckel and Harwood 2007). Performance was variable between different investigators, stressing the need for standardization, and none of the methods were able to accurately quantify the sources (Field and Samadpour 2007; Stoeckel and Harwood 2007).

Smaller studies have compared library dependent methods of *E. coli*; results indicated that ribotyping and PFGE performed well, but only a few of the isolates could be classified using ribotyping (Stoeckel et al. 2004). Single enzyme (HindIII) ribotyping has been compared to antibiotic resistance, but neither performed well at identifying the isolates (Moore et al. 2005), and another ribotyping and antibiotic resistance study demonstrated problems with antibiotic resistance stability (Samadpour et al. 2005). The difference of these smaller studies could be due to study design, methodological considerations, poor choice of analysis, and operator error (Field and Samadpour 2007; Stoeckel and Harwood 2007). Shanks et al. (2016) recently developed standardization techniques for human associated source tracking indicators, demonstrating that research is moving towards developing data acceptable criteria for universal use (Shanks et al. 2016).

In looking for the best practices associated with source tracking, a catch-all one indicator approach is just as limiting as fecal indicators, but the development of suites of indicators or microarrays holds promise for improving source tracking. Statistical models were used to evaluate the use of multiple molecular techniques to identify source, with the goal to create suites of indicators to optimize source tracking (Ballestè et al. 2010). McLellan and Eren (2014) suggested next generation sequencing, microbiome arrays, and better understanding of gut microbiomes may improve source tracking through using the most informative taxonomic groups as indicators (McLellan and Eren 2014).

Adaptive Mechanisms

E.coli is a well-studied bacterium in its host environment, but the ecology of *E.coli* in secondary habitats such as aquatic systems is less well-characterized (Blount 2015). The conditions of the nonhost environment (i.e., aquatic and benthic systems) were once considered too harsh for *E. coli* to survive, making it an attractive indicator organism. However,

research has shown that *E. coli* has potential to grow, persist, and evolve within nonhost environments such as on the surface of *Cladaphora* (Byappanahalli, Shively, et al. 2003), temperate soils (Byappanahalli, Fowler, et al. 2003; Whitman et al. 2006), beach sand (Alm et al. 2003; Whitman and Nevers 2003; Cloutier et al. 2015), and lake bottom sediments (LaLiberte and Grimes 1982). Even pathogenic indicator bacteria, such as *E. coli* O157:H7, survive in oligotrophic environments (Vital et al. 2008). Environmental *E. coli* strains exhibit diverse functions from their enteric counterparts, suggesting evolution in secondary habitats (Winfield and Groisman 2003; Luo et al. 2011). *E. coli* populations were discovered in non-human impacted environments in Puerto Rico, suggesting that these strains of *E. coli* were part of the native microflora (Lasalde et al. 2005). This phenomenon leads itself to the hypothesis that once released from the host, selective pressure is exerted on *E.coli* populations through environmental conditions; part of the population is lost in this secondary habitat while those *E.coli* with adaptive advantages persist (Luo et al. 2011; Berthe et al. 2013).

Understanding what environmental factors are associated with *E. coli* persistence can help characterize the niche for *E. coli* in aquatic environments. *E. coli* survival was strongly dependent on concentrations of dissolved organic carbon and phosphorus in microcosm experiments (Surbeck et al. 2010). Wild isolates of *E.coli* were found to survive for longer periods of time by possessing minimal virulence, adapting to low temperatures, and coexisting with low levels of fecal bacteria (Berthe et al. 2013). There is an active need in research to understand whether FIOs such as *E. coli* are competitively excluded from long-term residence, or if FIOs outcompete certain members of the community due to their metabolic plasticity (Souza et al. 2002). Persistent *E. coli* populations in excess of water quality criteria would suggest either diverse adaptations to facilitate long-term survival, continual inputs of fecal pollution, or a reservoir of resuspended material contaminated with *E. coli*.

Microbes are essential to ecosystem function because of their role in secondary production of organic matter and cycling of nutrients to higher trophic levels; however, little is known about the relationships between these assemblages and FIOs (Cloutier et al. 2015). Changes in these communities could act as a signal of environmental alterations such as inputs of fecal pollution, benefitting microbial source tracking and water quality management (Boivin

et al. 2002). Cloutier et al. (2015) examined these relationships, and found that the structure of these assemblages was similar among the same beach zone, and fine scale differences distinguished communities from different beaches, suggesting a biogeographic effect (Cloutier et al. 2015). Relationships between community metabolism and FIOs have not readily been investigated; however, community level physiological profiles (CLPPs) using carbon substrate utilization have been used to identify sources of pollution in surface waters (Hagedorn et al. 2003). These metabolic and functional characteristics are sensitive to environmental changes, and could be used as an early indicator of harmful anthropogenic inputs.

The Watershed Approach and Source Tracking

Characterizing landscape geography is an important component of watershed approaches to evaluate sources and mechanisms of fecal pollution. Land use within a catchment alters water chemistry in specific ways, shifting available nutrients and allowing formation of niches that endorse pathogen survival (Mawdsley et al. 1995; Williams et al. 2012). Surface flow is the primary transport mechanism of microbial pollution; however, different landscapes possess different microbial assemblages that can be shed through runoff (Jamieson et al. 2004; Wilkes et al. 2011). If a majority of nonpoint source fecal pollution stems from agricultural practices and livestock access to streams, proximity of agricultural land use near streams beds could provide insight into reasoning why certain watersheds continually violate water quality standards (Mawdsley et al. 1995; Wilkes et al. 2011). The identification of urban and residential land use with increased FIOs exceedances might point to either violations of the NPDES permits, faulty septic tanks systems, or illicit urban discharges (Duda et al. 1982). This type of source identification would be difficult to experimentally determine, but geographic information can be used to guide decision making.

Heavy rainfall typically precedes waterborne disease outbreaks; over 50% of waterborne disease outbreaks in the United States from 1948 to 1994 were associated with extreme rainfall events (Curriero et al. 2001). Heavy rainfall causes increased shredding of pathogens from the landscape, and overloads wastewater treatment systems. Turbulence associated with flooding also resuspends pathogens in sediments, increasing the distribution of pathogens within a

watershed; the association between FIOs, pathogens, and suspended solids has been documented (Characklis et al. 2005; Fries et al. 2006; Piorkowski et al. 2013; Sterk et al. 2013; Drummond et al. 2015). Rainfall has been associated with increased likelihood of detecting *Giardia* and *Cryptosporidium* in river water as well as enteric viruses (Atherholt et al. 1998; Hunter 2003; Sterk et al. 2013). The heaviest rainfall in a 50-year period preceded a 1993 *Cryptosporidium* outbreak in Milwaukee (Patz et al. 2008). Factors such as soil composition, topography, and amount of impervious surface govern how microbes partition into runoff, driving magnitude and extent of runoff mediated fecal pollution (Frey et al. 2013; Wilkes et al. 2013; Martinez et al. 2014). Steep slopes cause increased erosion due to stormwater, and in the presence of recent fecal contamination can increase number of pathogens in runoff as well as allow for longer distance FIO transport (Jamieson et al. 2004; Martinez et al. 2014). Impervious surfaces cause decreased infiltration, altering flow regimes and increasing shedding of pathogens from these and nearby landscapes (Mawdsley et al. 1995; Brabec 2002).

In order to use geographic data appropriately, researchers must identify both natural and anthropogenic drivers of fecal pollution to assist in source allocation. The soil survey geodatabase (SSURGO) is an effective tool to explore soil data for identifying natural drivers of fecal pollution much like topography. This can help determine differential effects on fecal partitioning into runoff, where infiltration into soils is to occur, and what areas have potential for long-term residences in soils (Whitman et al. 2006; Cloutier et al. 2015). Aquatic sediments are a reservoir for FIOs and pathogens because of favorable nutrient conditions, protection from UV inactivation, and safety from protozoan grazing (Alm et al. 2003; Ferguson et al. 2003). Widespread soil sampling is impractical for most watersheds, but the soil database allows for inferences into which soil characteristics are worthy of further investigation in controlled microcosm experiments.

One of the current issues for public health is the effect of climate change, and how this influences environmental transmission of pathogens. Increased temperatures and shifts within the hydrologic cycle will alter fate and transport of environmentally transmitted pathogens regardless of whether they are waterborne or vector spread (Sterk et al. 2013). Increased frequency of heavy rainfall events is expected in many regions of the United States increasing

potential for waterborne disease transmission (Curriero et al. 2001; Hofstra 2011). Increasing temperature can have a drastic effect on distribution and persistence of environmental pathogens, but effects arepathogen dependent (Hofstra 2011; Sterk et al. 2013). Even in the face of uncertainty regarding the scale of climate change, it is essential to explore how this increases potential for exposure from waterborne pathogens and how policy needs adjustment to limit adverse health effects due to climate change.

Use of Modeling to Improve Fecal Indicator Management

Although the advent of alternate indicators of fecal pollution and the accessibility of geographic datasets can improve understanding source of FIOs and pathogens in a watershed, there is a need to develop analytic techniques that incorporate this information constructively for decision-makers (Benham et al. 2006). Statistical models can extract information concerning environmental factors driving *E. coli* impairment, allowing inference into sources of fecal pollution; this can be used to assess the degree of compliance, identify dominant factors associated with impairment, and as a communication tool for stakeholders. While many TMDL models are process-based models, meaning that they are supposed to simulate the fate and transport of FIOs within the environment, statistical modeling can be used to optimize TMDLs in terms of source appropriation,identifying key processes integral to chronic fecal indicator impairment (Nevers et al. 2016). These techniques can also be utilized in a QMRA framework to adequately assess public health risks in areas with mixed sources of pollution (Ashbolt et al. 2010). Although one indicator cannot mimic the plethora of pathogens in water, incorporating modeling can fine tune an indicator's utility, inform the public concerning health risks, enhance understanding of the ecology of waterborne pathogens, and optimize source tracking and development of microarrays while increasing confidence in decision making.

Informing the Public Concerning Health Risks

Statistical models have been implemented to provide information concerning public health risks of pathogens through the use of predictive modeling. The use of these models can be multifaceted; provide early warning signs for recreational beach closures, provide

estimations of FIOs in times between sampling periods, and to identify hotspots of fecal pollution (Gonzalez et al. 2012; Herrig et al. 2015). To characterize water bodies impacted by multiple sources and transport mechanisms, it is infeasible to collect the number of samples needed to describe exposure and appropriately quantify source contribution; modeling fills in the gaps to better inform the public, decision-makers, and managers (Ashbolt et al. 2010).

Quantitative microbial risk assessment (QMRA) is a framework that allows for modeling a diverse range of scenarios concerning potential illness due to exposure to pathogens (Haas et al. 1999; Ashbolt et al. 2010). This can inform management actions and decision making concerning microbial water quality by simulating different sources of pollutants (Soller et al. 2014), resuspension of sediments (Abia et al. 2016), and stormwater events (McBride et al. 2013; Soller et al. 2015). QMRA approaches have also been used to infer etiologic agents of disease, and human enteric viruses including *Norovirus* were identified as the probable agents of gastrointestinal illness based on time of onset of illness (Soller, Bartrand, et al. 2010). Another important finding from QMRA studies is the concept that certain animal sources pose significantly lower risks than human sources of pollution; however, some species like cattle pose similar risks as human impacted waters (Soller, Bartrand, et al. 2010).

One of the benefits of risk assessment modeling is that it provides clear guidance of identifying research gaps as well as defining management actions; standard compliance monitoring does not support this agenda (Ashbolt et al. 2010). Since many QMRA approaches use Monte-Carlo simulations, this has the added benefit of being a probabilistic approach, which allows for estimation of uncertainty in models (Donald et al. 2011). This improves the traditional grab sample approach by providing a distribution of risks rather than a point estimate based on 1 or 2 statistics, allowing for assessment of long-range trends or degree of compliance for water quality managers (Ashbolt et al. 2010).

Enhancing Understanding of the Ecology of Fecal Indicators and Pathogens

FIOs and pathogens are unique water contaminants because of their ability to survive and grow in the environment, confounding the management processes. However, by modeling the ecology of these organisms, environmental factors driving impairment can be identified,

aiding in determining appropriate best management practice (BMP) for a given watershed. Issues such as presence of enteric viruses in surface waters, particle attachment of FIOs, and connecting land use, water chemistry, and pathogens can all be investigated using modeling approaches.

Predicting the presence of enteric viruses in surface waters is complex, and requires understanding of the presence, load, and age of fecal material within the environment. A novel application of the atypical coliform ratio was incorporated into multivariate logistic regression models to represent fecal age in predicting detection of viruses (Black et al. 2007). Other parameters within the best performing logistic models included the presence of a human-sterol to indicate source, and some FIOs to provide information concerning loads. This research highlights the importance of fecal age with regards to identifying the presence of fecal pollution.

Particle attachment of FIOs such as *E. coli* has been demonstrated to be dependent on a variety of factors such as particle size (Soupir and Mostaghimi 2011), suspended sediment loads (Garcia-Armisen and Servais 2009), organic matter(Guber et al. 2007), water chemistry (Park et al. 2008), and stormflow conditions (Characklis et al. 2005). This cacophony of influences makes it difficult to determine constant attachment percentages for watershed TMDL models (Piorkowski et al. 2013). Classification and regression trees (CART), regularized regression using a least angle shrinkage and selection operator (LASSO), and multivariate adaptive splines (MARS) have been used to investigate factors concerning *E. coli* attachment to particles and virulence (Piorkowski et al. 2013). The benefit of these types of models is that they are not confounded by overfitting, multicollinearity, or restriction to linear relationships (Friedman 1991; Tibshirani 1996; Parkhurst et al. 2005). MARS was the highest performing model based on accuracy statistics. Hydrological and meteorological variables tended to have minor influences on particle attachment except in the LASSO models, with land use and particle properties being included in the majority of models (Piorkowski et al. 2013). In terms of virulence, factors such as residential land use, electrical conductivity, and water and air temperatures were found to be associated with the presence of *E. coli* virulence markers within the attached fraction, in addition to the particle properties within the previously mentioned

attachment models. Similar results were found within the unattached models for presence of virulence factors without the particle properties.

Multiple linear regressions using climate variables, measurements easily obtained using field equipment, or land cover datasets derived from remote sensing have been employed with varying success to develop predictive models for watershed managers (Eleria and Vogel 2005; Gonzalez et al. 2012; Herrig et al. 2015). Regression has also been used to compare new molecular techniques with their cultural counterparts to identify if qPCR was appropriate for the Neuse River estuary in North Carolina (Gonzalez et al. 2012). Most predictive models find that fecal pollution is governed by precipitation, landscape, and water quality parameters (Francy et al. 2003; Gonzalez et al. 2012; Herrig et al. 2015). In the Charles River, Massachusetts it was found that rainfall amount and intensity governed fecal coliform levels, with previous day's coliform levels also accounting for some variation (Eleria and Vogel 2005); in other studies, landscape indicators such as agriculture on steep slopes, urban land, and natural stream cover predicted watershed impairment (Smith et al. 2001). Fecal pollution has also been modeled using a Bayesian maximum entropy approach which incorporates spatial covariance as well as specific measurements at sites to produce informative predictive maps of water quality (Coulliete et al. 2009; Money et al. 2009). CART were used to identify distinct environmental and land use indicators of the sporadic distribution of *Cryptosporidium* oocysts and *Giardia* cyst densities, as well as the presence of *Salmonella enterica*, *Campylobacter* spp., *Listeria monocytogenes*, and *Escherichia coli* O157:H7 (Wilkes et al. 2011). Season, stream order, turbidity, and discharge were strong recurring predictors. Land use was found to not be a strong predictor, although densities of pathogens were higher near dairy operations. Multivariate statistical techniques such as canonical correlation analysis and canonical discriminant analysis can reveal spatiotemporal variability within parameters affecting water quality, and these can be used to infer mechanisms driving fecal pollution; this can ultimately be used to identify the sources of this contamination (Hall et al. 2014). Some other applications of modeling to understand how environmental factors affect FIOs are estimating *E. coli* loads using physical, chemical, and biological factors within a neural network (Dwivedi et al. 2013),

and hyporheic-groundwater interactions associated with transport of *E. coli* within sediments porewater (Dwivedi et al. 2016).

Optimizing Source Tracking

In looking for best practices for source tracking, a catch-all indicator approach is just as limiting as the single indicator paradigm, but using statistical models to create suites of indicators holds promise to improve source tracking. Discriminant analysis, k-nearest neighbors, decision trees, and naïve Bayes classifiers were used as predictive models to evaluate the use of multiple molecular techniques to identify source (Ballestè et al. 2010).*Bifidobacterium adolescentis* and host-specific mitochondrial DNA markers for bovine (Bomito) and swine (Pomito) sources of pollution were identified to differentiate between human, swine, poultry, and bovine sources with 75.7% accuracy; when just discriminating between human and animal sources, only *B. adolescentis* and Pomito were needed to achieve 84.6% accuracy.

Neural networks are machine learning models that are intriguing for complex modeling problems such as source tracking because these systems can estimate complex, non-linear relationships characteristic of riverine systems, including fate and transport of fecal pollution (Brion and Lingireddy 1999; Basheer and Hajmeer 2000). Brion et al. (2002) used neural networks to sort between animal runoff and human sewage; additionally, the relative age of the fecal pollution was evaluated within the watershed (Brion et al. 2002). Gram-negative and gram-positive bacteria were required to sort sewage from runoff, and turbidity was found to be relatively unessential in source tracking. Neural networks have also been used to identify sources of intrusions in water supplies to assess the vulnerability of water systems, and develop early warning systems to protect human health and the environment (Kim et al. 2008).

Summary of Literature Review

Fecal pollution is a consequence of human ecology, and FIO monitoring provides a protective measure against exposure to pathogens surviving in these wastes. Waterborne disease outbreaks are costly, both from a public health and economic standpoint, so protecting from this exposure is necessary to reduce burden. However, this problem is ubiquitous, causing water quality degradation in a substantial portion of our rivers and streams. Alternative methods are needed to utilize this monitoring data and identify strategies to reduce sources of fecal pollution, and improve a riverine system's overall utility.

Although the spirit of the TMDL is geared towards improving the quality of water in the United States, program implementation has several deficiencies. Inadequacies in current modeling approaches, resource limitation, and the sheer number of impaired watersheds make it difficult to develop successful TMDLs. While the policy behind TMDL development is progressive, the science associated with the policy is inadequate to effectively implement the Clean Water Act. Strategies to improve modeling and stimulate resources can boost the effectiveness of TMDLs, successfully fulfilling the objectives of the Clean Water Act to rehabilitate, protect, and enhance surface water quality nationally.

While fecal indicators provide a protective measure against exposure, they fall short because they do not appropriately characterize public health risks, relationships between fecal indicators and pathogens are ill-defined, and they do not provide information about source. In addition, adaptive mechanisms exist in fecal indicators to persist in secondary habitats, leading to unique monitoring challenges in chronically impaired streams. A deeper issue is that policy makers focus on controlling fecal indicators, losing sight that the main purpose of fecal indicators is to reduce exposure to pathogens. Because of this misdirected focus, much of the management of watersheds concerns control of fecal indicators rather than understanding and limiting exposure to pathogens.

These shortcomings have stimulated development of a plethora of alternate indicators and source-tracking techniques using both culture dependent and independent techniques. While culture-based are relatively inexpensive, require minimal processing, and are easily available, the need for a library is time consuming, requiring extensive sampling to create a

cosmopolitan database. This limits these methods to smaller geographic areas. The problem of culturing bias can be reduced through using culture independent techniques, allowing sampling of the full population in question. These reduce the time needed for analysis, are typically simpler, and are not restricted to easily cultivatable organisms. However, host-specific markers may not be present in all individuals of a species, validation is needed when applied to different regions, and large samples are potentially required for rare toxin genes. Only a few animal species are currently available, and wildlife species are poorly represented, identifying a definitive research gap. Additionally, correlations between FIOs, pathogens, and culture independent markers either are inconsistent or not well-characterized. Since current regulation is based on FIOs, any source tracking marker used must correlate with FIOs for them to be useful.

For reasons mentioned above, managing fecal pollution is a complicated process, needing creative strategies in monitoring and assessment; statistical modeling presents ways to extract information concerning water quality to better characterize source, transport mechanisms, and ecological drivers of chronic *E. coli* impairment. These applications are multi-faceted; identify key sources of impairment, predict levels of FIOs in between sampling periods, fill knowledge gaps concerning hydrologic dependence, and connect existing water quality to degree of fecal pollution. Modeling using multivariate statistics affords the ability to synthesize information from a variety of molecular, metabolic, chemical, and geographic datasets, enhancing and optimizing monitoring and identifying key remediation strategies.

Historically, sanitary engineers needed to understand pollution sources and environmental dynamics because traditional microbial methods were either inadequate or time-consuming. It seems counter-intuitive that this approach has been traded in the effort to develop high-tech rapid indicators without acknowledging the wider environmental context. A multitier approach refocuses attention to pathogens, incorporating multiple source tracking techniques and water quality monitoring to paint a more comprehensive picture of the pathogen distribution and survival. The best approach to source tracking is to develop blended indicators of specific types of pollution, and develop microarray methods to reduce cost and resources in analysis of data. In addition, efforts to coordination NGS research, FIOs, and source

tracking can optimize monitoring strategies. Quantitative microbial risk assessment (QMRA) can assist in developing risk scenarios and models to estimate human health risk in a universal framework. Watershed decision-makers also need to evaluate potential sources within a watershed using modern geospatial techniques, and the use of statistical modeling should not be undervalued in a watershed approach characterize and communicate fate and transport of fecal pollution.

Goals of Study

This dissertation presents three manuscripts to aid in understanding the ecology of fecal indicators in a lower order stream in the Watauga River watershed. The papers integrate multivariate modeling, multiple FIOs, and microbial functional diversity to characterize the niche of fecal indicators within their secondary habitat, extracting information concerning source and mechanisms driving their presence along Sinking Creek, a 303(d) listed stream. Since fecal indicators such as *E. coli* exist as both naturalized and invasive species within the microbial community, this dissertation also explores connections between microbial activity, fecal pollution, and other types of pollution along this stream continuum. Incorporating these techniques can improve decision making in chronically impaired watersheds, guiding management and remediation strategies most appropriately.

Chapter 2, **Maxent Estimation of Aquatic *Escherichia Coli* Stream Impairment**, explores the use of Maxent, a commonly used ecological niche modeling, to identify environmental factors associated with fecal indicator impairment. The goal of this research was to utilize long-term water quality data to extract information concerning what water quality parameters are associated with the probability of impairment; impairment in this case is defined as one sample being in violation 2012 recreational water quality criteria standard for *E. coli* (United States Environmental Protection Agency 2012). Using Maxent to model how water quality influences *E. coli* impairment aids in inferring source and mechanisms of fecal pollution. This approach allows for estimation of both linear and non-linear effects of water quality, demonstrates a probabilistic method for variable selection, and reframes the question from “How much *E. coli* in our watershed?” to “what factors separate *E. coli* impairment from compliance?”, which is

useful when evaluating watershed decisions. This manuscript was published in the open source journal *PeerJ* on September 13th, 2018.

Chapter 3, **Canonical Variable Selection for Ecological Modeling of Fecal Indicators**, extends the use of Maxent by incorporating canonical correlation analysis as a variable selection procedure in probabilistic models of F+ somatic bacteriophage detections and *E. coli* impairment. The goal of the study was to determine if fecal indicators share common ecological niches, and if not, what differences are observed between impairment and water quality. This study used a single year of monitoring data to identify chemical and microbial parameters most associated with these fecal indicators, indicating distinct ecological drivers of the probability of impairment. This manuscript was published in a special issue of *Journal of Environmental Quality* titled "Microbial Water Quality—Monitoring and Modeling" on September 20th, 2018.

Chapter 4, **Microbial Community Metabolism Associated with Pollution along a Stream Continuum,** introduces the use of microbial metabolism, indicated by the degradation of single sources of carbon, to identify spatiotemporal changes and connections to fecal and other types of pollution within Sinking Creek. The goal of this study was to observe whether changes in the activity of the microbial community could be connected to different sites and pollution gradients, and what is the seasonal variation of this activity. Unique patterns of substrate utilization existed, and these were further investigated using canonical correlation analysis and multiple linear regression. Canonical correlation analysis was used to summarize interactions between aquatic and benthic microbial communities to utilize various carbon substrates, better characterizing metabolic exchange between these environments. Dominant degraded substrates patterns were then used to inform regression models on three pollutants; fecal pollution as measured by *E. coli* concentrations, nutrient pollution in the form of nitrates (NO_3^-), and organic pollution in the form of biochemical oxygen demand (BOD_5).

Chapter 5, **Conclusions and Recommendations**, presents a critical analysis of the achievement of the objectives, and limitations of the study were mentioned. Connections between chapters were exposed, and merits of the methodology were summarized. The conclusions from the research were stated to review the information gained from this research, and future directions for research were developed as well.

CHAPTER 2

MAXENT ESTIMATION OF AQUATIC *ESHERICHIA COLI* STREAM IMPAIRMENT

DENNIS GILFILLAN, ANDREW JOYNER, PHILLIP SCHEUERMAN

Abstract

Background. The leading cause of surface water impairment in United States' rivers and streams is pathogen contamination. Although use of fecal indicators has reduced human health risk, current approaches to identify and reduce exposure can be improved. One important knowledge gap within exposure assessment is characterization of complex fate and transport processes of fecal pollution. Novel modeling processes can inform watershed decision making to improve exposure assessment. **Methods.** We used the ecological model, Maxent, and the fecal indicator bacterium *Escherichia coli* to identify environmental factors associated with surface water impairment. Samples were collected August, November, February, and May for 8 years on Sinking Creek in Northeast Tennessee and analyzed for 10 water quality parameters and *E. coli* concentrations. Univariate and multivariate models estimated probability of impairment given the water quality parameters. Model performance was assessed using area under the receiving operating characteristic (AUC) and prediction accuracy, defined as the model's ability to predict both true positives (impairment) and true negatives (compliance). Univariate models generated action values, or environmental thresholds, to indicate potential *E. coli* impairment based on a single parameter. Multivariate models predicted probability of impairment given a suite of environmental variables, and jack-knife sensitivity analysis removed unresponsive variables to elicit a set of the most responsive parameters. **Results.** Water temperature univariate models performed best as indicated by AUC, but alkalinity models were the most accurate at correctly classifying impairment. Sensitivity analysis revealed that models were most sensitive to removal of specific conductance. Other sensitive variables included water temperature, dissolved oxygen, discharge, and NO_3^-. The removal of dissolved oxygen improved model performance based on testing AUC, justifying development of two optimized multivariate models; a 5-variable model including all sensitive parameters, and a 4-variable model that excluded dissolved oxygen. **Discussion.** Results suggest that *E. coli* impairment in

Sinking Creek is influenced by seasonality and agricultural runoff, stressing the need for multi-month sampling along a stream continuum. Although discharge was not predictive of *E. coli* impairment alone, its interactive effect stresses the importance of both flow dependent and independent processes associated with *E. coli* impairment. This research also highlights the interactions between nutrient and fecal pollution, a key consideration for watersheds with multiple synergistic impairments. Although one indicator cannot mimic the plethora of existing pathogens in water, incorporating modeling can fine tune an indicator's utility, providing information concerning fate, transport, and source of fecal pollution while prioritizing resources and increasing confidence in decision making.

Introduction

Rapid urbanization of rural areas causes deterioration in water quality, rendering many water bodies unfit for their domestic and recreational use. An assortment of contaminants is introduced into aquatic systems, but pathogens represent the major cause of stream impairment in the United States (United States Environmental Protection Agency 2017). Pathogens are difficult to measure directly because of their sporadic distribution, costly identification, and potential health risks to laboratory workers (Field and Samadpour 2007). Most pathogens in aquatic systems stem from human and animal fecal wastes, including direct deposition of feces in water (Vidon et al. 2008), runoff from land with fecal deposits (Tyrrel and Quinton 2003; Jamieson et al. 2004), and sanitary sewer malfunctions (Ferguson et al. 2003; McLellan and Eren 2014). To address the difficulties in monitoring specific pathogens, fecal indicator organisms (FIOs) are commonly used to assess the presence of fecal pathogens.

An effective fecal indicator is associated with the presence of specific pathogens, with a straightforward method for enumeration that correlates with magnitude and age of fecal pollution (Savichtcheva and Okabe 2006; Maier et al. 2009). The use of FIOs such as fecal coliform bacteria and *Escherichia coli* are traditionally used for determining surface water pathogen impairment (Yates 2007; United States Environmental Protection Agency 2012). Although these indicators assist in alerting populations when exposure to pathogens is likely, the current approach is limited by using a single indicator such as *E. coli* for a designated use

(Wade et al. 2003; Savichtcheva and Okabe 2006; Field and Samadpour 2007). The cosmopolitan nature of *E. coli* in warm-blooded animals makes them impractical for source identification (Field and Samadpour 2007; Yates 2007; McLellan and Eren 2014; Blount 2015). The ability of *E. coli* to survive in soils (Lasalde et al. 2005; Ishii et al. 2006), algae (Byappanahalli, Shively, et al. 2003), and sediments (LaLiberte and Grimes 1982; Alm et al. 2003; Drummond et al. 2015) provide a reservoir for continued persistence and potential to naturalize (Winfield and Groisman 2003; Lasalde et al. 2005; Luo et al. 2011). These characteristics and deficiencies emphasize the difficulty of single standard FIO monitoring for impairment, stressing the need for additional methods to evaluate source and mechanisms of FIO impairment.

In addition to the above issues, appropriately characterizing FIO impairment for regulation and decision making is difficult due to complex fate and transport processes (Benham et al. 2006; de Brauwere et al. 2014; Drummond et al. 2015). These complex fate and transport processes include transport through runoff and stormwater (Lipp et al. 2001; Kistemann et al. 2002; McKergow and Davies-Colley 2010), remobilization from sediments and hyporheic exchange (Drummond et al. 2015; Dwivedi et al. 2016), particle attachment (Characklis et al. 2005), and UV light exposure (Sinton et al. 2002). Additionally, ecological processes control FIO fate and transport through variable survival patterns of indicators and pathogens,(Anderson et al. 2005; Stott et al. 2011) availability of nutrients and organic matter ,(Surbeck et al. 2010; Perkins et al. 2016) and predation.(McCambridge and McMeekin 1980) Appropriately characterizing the physics and ecology driving fate and transport can better inform management decisions for total maximum daily load (TMDL) development, reduction of pollution, and allocation of resources.

Modeling provides flexible approaches to infer sources and processes associated with FIOs and other pathogens, overcoming some of the issues of the single indicator paradigm. Various statistical and machine learning models have been used to approach such problems of incorporating age of fecal pollution for source tracking or detection of viruses (Brion et al. 2002; Black et al. 2007); identifying land use, environmental, and water quality parameters associated with FIOs and pathogens (Brion and Lingireddy 1999; Viau et al. 2011; Wilkes et al. 2011;

Gonzalez et al. 2012; Gonzalez and Noble 2014; Hall et al. 2014; Herrig et al. 2015; Lušić et al. 2017); determining factors influencing particle attachment and virulence (Piorkowski et al. 2013); and optimizing microbial source tracking (Belanche-Muñoz and Blanch 2008; Ballestè et al. 2010; Smith et al. 2010; Molina et al. 2014). Some other applications of modeling include using turbidity or rainfall to predict *E. coli* concentrations at unmonitored sites (Coulliete et al. 2009; Money et al. 2009), estimating *E. coli* loads using physical, chemical, and biological factors within a neural network (Dwivedi et al. 2013), and hyporheic-groundwater interactions associated with transport of *E. coli* within sediment porewater (Dwivedi et al. 2016). Modeling can inform decision-makers concerning what drives impairment, addressing some of the shortcomings of a single indicator approach.

Maxent, a commonly used ecological niche model (Phillips et al. 2004; Phillips and Dudík 2008), identified environmental variables associated with probability of *E. coli* stream impairment, making inferences concerning source and mechanisms driving fecal pollution. Although modeling *E. coli* using a machine learning model such as Maxent is not a novel approach, e.g., Dwivedi, Mohanty and Lesikar, (2013), this study is unique in the following ways: it focuses on how the water quality is associated with *E. coli* impairment in lower order streams, uses nonparametric bootstrapping as a probabilistic assessment of model performance based on the area under the curve (AUC) of the receiving operator characteristic (ROC), and uses loss of information as an indicator of sensitive variables. Ecological niche models have been utilized for species distribution (Lozier et al. 2009), conservation of rare species (Guisan et al. 2006), invasive species (Thuiller et al. 2005), and disease vector epidemiology studies (Boeckmann and Joyner 2014), but this is a new application of Maxent to microbial water quality. Additionally, developing models in lower order streams has not been previously reported; this is important because water from low order streams is used for domestic water supply and recreation in many areas of the United States.

The motivation for using Maxent to predict *E. coli* impairment is to investigate how environment, i.e. water quality parameters, shapes the niche of *E. coli* impairment based on a decision boundary; in this case, a water quality standard. A probabilistic procedure for univariate and multivariate model development is presented using nonparametric

bootstrapping cross-validation. Univariate models generated action values, or environmental thresholds of impairment, to indicate potential *E. coli* impairment based on a single parameter. Multivariate models predicted probability of impairment given a suite of environmental variables, and jack-knife sensitivity analysis removed unresponsive variables in multivariate models to elicit a set of the most responsive water quality parameters. Using Maxent to model how water quality influences *E. coli* impairment aids in inferring source and mechanisms of fecal pollution. This approach allows for estimation of both linear and non-linear effects of water quality, demonstrates a probabilistic method for variable selection, and reframes the question from "How much *E. coli* in our watershed?" to "what factors separate *E. coli* impairment from compliance?," which is useful when evaluating watershed decisions.

Methods

Sampling Sites and Data Collection

Sinking Creek is a 1st to 3rd Strahler order mixed-use stream that is noncompliant for State of Tennessee standards for fecal coliform and *E. coli* (Tennessee Department of Environmental and Conservation 2006). Starting in August 2004, samples were collected by hand in August, November, February, and May of each year until August 2011 as a long-term monitoring plan at 14 sites in Sinking Creek, and samples were analyzed for 10 water quality parameters and populations of *E. coli* (Fig. 2.1).

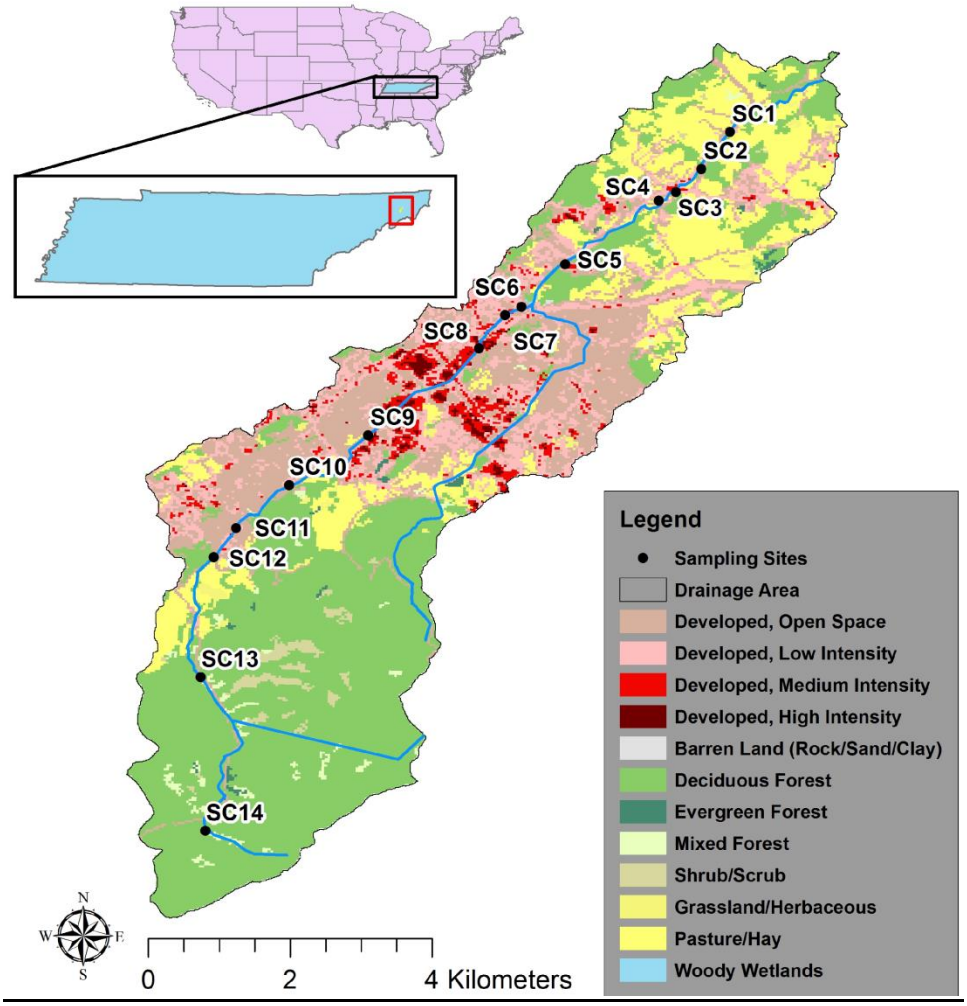

Figure 2.1. Map of sampling sites and watershed of the study area, Sinking Creek. The inset map shows the United States and the state of Tennessee, and the location of Sinking Creek. Samples were taken from August 2004 to August 2011 during the months of August, November, February, and May. The outline represents the watershed boundary of Sinking Creek, and 2006 National Land Cover Dataset (NLCD) has been clipped to the watershed (Fry et al. 2011). Stream flows from its headwaters at SC14 downstream to SC1.

Specific conductance (conductivity) and water temperature were measured using an Orion 115A+ conductivity meter (Thermo Fisher Scientific, Waltham, MA). The pH was measured using an EL2 portable pH meter (Mettler Toledo, Columbus, OH). Dissolved oxygen was collected using an YSI Model 55 dissolved oxygen meter (YSI Inc., Yellow Springs, OH). Samples for nitrates (NO_3^-), phosphates (PO_4^{3-}), biochemical oxygen demand (BOD_5), alkalinity,

and hardness were collected in clean 2 L polyethylene bottles and stored at 4 °C until laboratory analysis. A flow meter (Global Water, FP101) was placed in the center of the channel to measure stream velocity. Stream width was calculated where the stream velocity was measured, and depth was averaged over three points across the stream width. Velocity was multiplied by stream width and average depth to estimate discharge.

NO_3^- and PO_4^{3-} analyses were performed in triplicate using colorimetric HACH™ methods (HACH, Loveland, CO) and reagents. NO_3^- and PO_4^{3-} analyses were conducted by adding 10 mL of water to a vial containing NitraVer5 or PhosVer3 for the respective analyses. Vials were shaken to dissolve the reagent and samples were analyzed with a DR890 colorimeter (HACH, Loveland, CO) (HACH Company 2013). Triplicate sample for alkalinity and hardness were determined using 100 mL sample volumes and a digital titrator (HACH, Loveland, CO) (HACH Company 2013). Phenolphthalein and bromcresol green-methyl red indicators were used, and the sample was titrated with 1.6 N sulfuric acid to a grey-green endpoint (HACH Company 2006). BOD_5 was measured in triplicate using the 5-day BOD_5 test (American Public Health Association 2005). Populations of *E. coli* were determined using the Colilert defined substrate test. Briefly, 97 wells were filled with 100 mL of water sample with the Colilert substrate added. Samples were incubated for 24 hours, and wells that fluoresce under a UV light were considered positive for *E. coli,* and a most probable number estimate was made based on the number of positive wells in both the large and small wells (American Public Health Association 2005). If a sample was in excess of the geometric mean United States recreational water quality criteria, the sample site was considered impaired. Impairment was based on recommendation 1, which is a threshold of 126 $\frac{CFU}{100\ mL}$ that corresponds to an illness rate of $\frac{36}{1{,}000}$ people (United States Environmental Protection Agency 2012).

To get an estimation of land use throughout the Sinking Creek watershed, land cover data were downloaded from the National Land Cover Dataset (NLCD) for 2006 (Fry et al. 2011). Each sampling site's drainage area was delineated using StreamStats version 3 from the United States Geologic Survey (Ries III et al. 2017). Land was grouped into 3 categories; forested, developed, and agricultural. Forested land includes the categories deciduous forest, evergreen forest, and mixed forest. Developed land use includes all developed categories; open space

(less than 20% impervious surface), low intensity (20–49% impervious surface), medium intensity (50–79% impervious surface), and high intensity (80–100% impervious surface). Agricultural land included grassland/herbaceous and pasture/hay. The area of each land use was divided by the total area of the drainage area to get the percentage land use shown in Table 2.1, and sampling sites as well as land cover categories are shown in Fig. 2.1.

Table 2.1. Sampling sites, land use, and *E. coli* concentrations in Sinking Creek. Percentage of each land cover types (Agricultural, Developed, and Forested) as well as *E. coli* geometric means (GM), geometric standard deviations (GSD), and maximum and minimum values for each site used in the study are shown.

Sampling Site	Agricultural Land Use (%)	Developed Land Use (%)	Forested Land Use (%)	E. coli GM (GSD)	Min, Max
SC1	15.6	36.4	47.3	254.5 (3.4)	43.7,2398.8
SC2	14	37.2	48.1	182.3 (6.1)	17.4,39810.7
SC3	9.7	38	51.5	137.1 (4.0)	14.5,1737.8
SC4	9.7	37.9	51.6	169.8 (5.7)	8.5,23988.3
SC5	8.7	38.1	52.4	140.0 (7.2)	4.1,30903.0
SC6	7.1	30.2	61.6	50.2 (8.3)	0.5,8709.6
SC7	7.1	30	61.8	36.7 (9.4)	0.5,10232.9
SC8	7.7	24.3	66.8	73.9 (5.3)	10.7,8709.6
SC9	7.4	19.9	71.4	110.3 (5.8)	14.5,3981.1
SC10	5.2	6.6	86.5	70.6 (5.2)	6.2, 1995.3
SC11	5.6	3.8	89	17.2 (9.9)	0.5,1202.3
SC12	5.8	2.1	90.3	91.3 (3.8)	5.2,812.8
SC13	0	1.1	96.5	7.8 (5.5)	0.5,102.3
SC14	0	0	100	5.0 (6.1)	0.5, 245.5

Modeling Background

Maxent is an iterative machine learning model commonly used for mapping species distributions (Phillips, S., Dudík, M., Schapire 2010). Within the sample space, x, and given a set of environmental features (parameters), $f_1(x)$, $f_2(x)$, …, $f_n(x)$, the Maxent distribution estimates a vector of feature weights, $\beta = (\beta_1\beta_2, \ldots, \beta_n)$, that maximizes the entropy of the raw distribution of impairments, $q_\beta(x)$, using a Gibbs distribution,

$$q_\beta(x) = \frac{\exp(\sum_{j=1}^{n} \beta_j f_j(x))}{Z_\beta} \quad (2.1)$$

where Z_β is the normalization constant that ensures that $q_\beta(x)$ integrates to one over the study area (Phillips, S., Dudík, M., Schapire 2010). This modeling approach is justified because it provides the maximum information concerning impairment. From a water quality management standpoint, this approach is beneficial because decision-makers and stake-holders are more concerned with factors associated with impairment rather than compliance when approaching fecal pollution monitoring and management.

Original features (parameters) can be transformed into quadratic, product, hinge, and threshold feature classes so that complex multivariate responses can be modelled (Phillips and Dudík 2008), but Maxent incorporates L1 regularization to balance satisfying the constraints on the features while minimizing overfitting. L1 regularization is not unique to Maxent, and is used in many general linear models (Elith et al. 2011). A regularization parameter λ_j smooths probability distributions, generating sparse solutions and removing unnecessary features; this shrinks weights to balance fit and complexity (Elith et al. 2011). Because of regularization, Maxent fits a penalized maximum likelihood model equivalent to minimizing the relative entropy dependent on the error-bound constraints,

$$max_\beta \ \frac{1}{m}\sum_{i=1}^{m} \ln(q_\beta(x_i)) - \sum_{j=1}^{n} \lambda_j|\beta_j| \quad (2.2)$$
$$subject\ to\ \int q_\beta(x)dx = 1$$

where m is the number of positive samples, n is the number of features, and x is the feature vector for occurrence point i. Eq. (2.2) provides insight into how Maxent uses background data: the first term is larger for models that distinguish between impairment states the best. The second term represents the regularization, which gets larger as the weights β_j

increase, indicating a complex model more likely to over fit. The output of $q_\beta(x)$, is termed the raw distribution, but it is difficult to interpret due to its scale dependence. More background points result in smaller raw values because their sum cannot exceed 1 over a large amount of points (Phillips and Dudík 2008; Elith et al. 2011). For this reason, the logistical output of the Maxent model, $P(x)$, will be used because it represents the probability of impairment given the sample space, x. This is a logistic model using the same set of weights β with the intercept of the model determined by the entropy of $q_\beta(x)$, H. The model is shown in Eq. (2.3) below.

$$P(x) = \frac{e^H q_\beta(x)}{1 + e^H q_\beta(x)} \quad (2.3)$$

Univariate Models

Data were processed using a list-wise deletion process, where individual samples from a site were removed if they were missing a parameter measurement due to laboratory errors, equipment malfunctions, calibration issues, or sites being dry at the time of sampling. A sample of 100 bootstrapped models was developed, and 20% was subsampled for testing validation. Bootstrapping is a nonparametric resampling technique to make inferences about a population based on resampling from a set population, generating population level statistics, while providing an estimate of uncertainty of those statistics (Campolongo and Saltelli 1997). For this modeling approach, all background points are used in the development of the null model, and the impaired samples are bootstrapped. Although Maxent can incorporate a wide variety of feature classes, only linear and quadratic feature classes were used to develop action values, or thresholds of impairment. The rationale for using these types of feature classes is for ease of generating action values as well as to assess both linear and non-linear effects of single parameters.

The AUC was calculated for the training and testing datasets. The AUC is a metric of performance for binary classification. The true positive prediction rate (sensitivity) and false positive prediction rates (1–specificity) of each sample are plotted as a ROC for different decision boundaries, and the area under that ROC is integrated. An AUC of 0.5 indicates that the model is no better than random chance, and a value of 1.0 indicates perfect model performance (Zweig and Campbell 1993; Zou et al. 2007).

The decision boundary (logistic threshold) between impaired and unimpaired samples should maximize accurately predicting impairment (exceedance of the *E. coli* criteria) while balancing correct negative predictions (Bean et al. 2012). Therefore, maximum test sensitivity and specificity was defined as the appropriate decision boundary. A low sensitivity would indicate poor performance in identifying impairment, while a low specificity would indicate an overcautious model in which resources might be wasted in remediation of false positives. Accuracy for Maxent models was calculated as follows: $\frac{TP+TN}{TP+FP+TN+FN}$, where TP are true positives, TN are true negatives, FP are false positives and FN are false negatives. Significance of the univariate model was determined by calculating the χ^2 statistic for each confusion matrix, with the null hypothesis being that the classifier was no better than random chance.

Action values (environmental thresholds) are conditions in which a parameter (variable) is at the threshold of impairment, indicating potential exceedance of the *E. coli* standard. Action values were calculated for significant ($p < 0.01$) univariate models by averaging bootstrapped weights and estimating the parameter value at which the probability of impairment equals the logistic threshold. Fig. 2-2 demonstrates the concepts of the AUC performance metric, selected decision boundary, and the concept of the action value in relation to the selected decision boundary and Maxent model function (Eq. (2.3)).

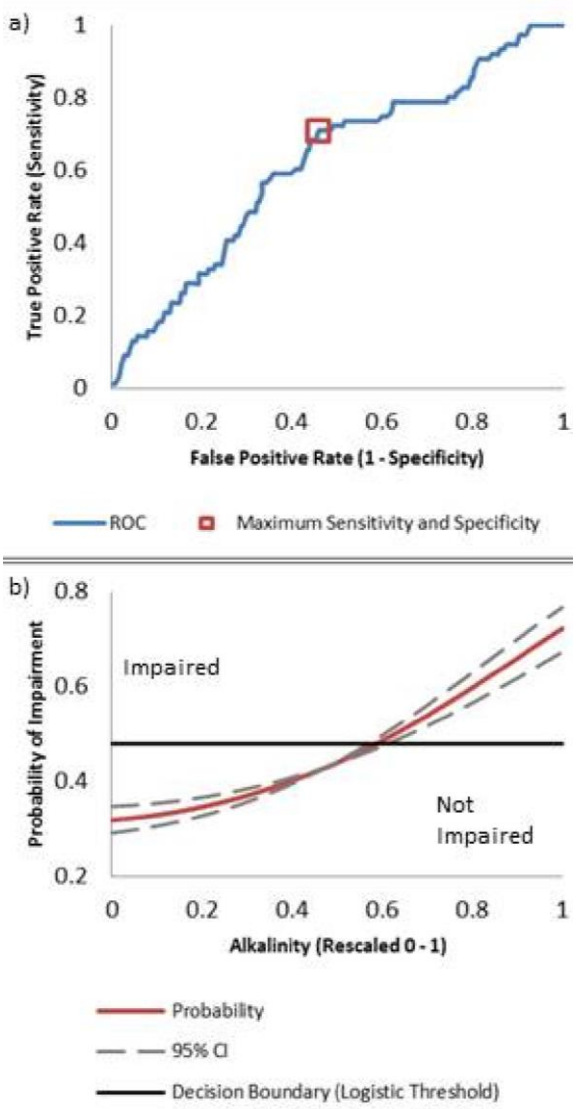

Figure 2.2. Theoretical plots to illustrate the concept of the ROC, decision boundaries, and action values. (a) Plot of an ROC curve, where the x-axis represents the false positive rate, or the compliment of the specificity, and the y-axis represents the true positive rate, the sensitivity. The curve is integrated to obtain the AUC, the performance metric for each of the models. The box represents the point at the decision boundary. (b) Theoretical plot of a univariate Maxent model function (Eq. (2.3)) with values for alkalinity rescaled from 0 to 1. The solid red line represents Eq. (2.3), the dotted lines represent the upper and lower 95% confidence intervals, and the horizontal black line represents the decision boundary. The action values, or environmental thresholds, and associated confidence intervals are the intersections between the results of Eq. (2.3) and the decision boundary.

Multivariate Models and Sensitivity Analysis

Although some authors state that collinearity is not as problematic in Maxent compared to traditional regression approaches, collinearity was explored and subsequently removed using Pearson correlation coefficients (Elith et al. 2011). Variables that were highly correlated ($r > 0.8$) were evaluated to determine which variable to include based on expertise, connections to previous models, and accuracy metrics within the analysis. The initial multivariate models included all noncollinear variables and were developed using 100 bootstrapped samples like the univariate models, with the addition of product feature classes to incorporate variable interaction. Average variable contribution was determined by calculating the increase in information gain associated with a change in each feature for each iteration of the model algorithm, normalized to percentages. The permutation importance of a feature is an indicator of variable sensitivity. In each model run, the feature training presence and background data are randomly permutated, and the resulting drop in training AUC is normalized to percentages for each variable and averaged over the 100 bootstrapped runs.

A jack-knife sensitivity analysis was used to determine the best subset of covariates to include in a trimmed model. Each variable was removed from the analysis, and a comparison was made to determine if the removal of a variable caused a significant ($p < .05$) change in training or testing information gain. Student's *t*-tests were performed on each jack-knifed model to evaluate significance, and variables were included if the information gain from either the testing or training sets decreased; decrease in information would correspond to a significant loss of information, providing criterion for inclusion of the variable in final models.

Results

Univariate model performance

The sampling program resulted in 29 sampling trips over 14 sites, allowing for a potential of 406 samples for analysis. 127 samples were removed due to missing information in the dataset, leaving 279 total samples for model development. This included 95 impairments, identified by exceedance of the *E. coli* recreational water quality standard. Table 2.1 presents the summary statistics for *E. coli* and the associated land use in each sampling site's drainage

area. Each training set included 279 background points, 76 points for training, and 19 points to evaluate performance on testing data.

Table 2.2 summarizes the training and testing performance of the univariate Maxent models used to identify *E. coli* impairment based on environmental variables. Water temperature performed best based on AUCs, but had lower accuracy than conductivity, dissolved oxygen, and alkalinity. The plausible explanation of these differences is the latter variables had higher specificity rather than sensitivity at the chosen decision boundary (Table 2.2). Accuracy was found to be highest for alkalinity and lowest for pH.

Table 2.2. Summary of training and testing performance of Maxent models based on AUC metrics, accuracy based on maximum test sensitivity and specificity decision boundary (logistic threshold), and action values with 95% confidence intervals. If an upper bound of a confidence interval exceeds the maximum sampling value for a set of data, the maximum value is given.

Variables	**Training AUC (+/-SE)**	**Testing AUC (+/-SE)**	**Accuracy**	**Sensitivity**	**Specificity**	**Action Values (x) ¥ (95% CI)**
Alkalinity ($\frac{mg}{L}$)	0.616 (0.003)	0.620 (0.006)	68.5	0.537	0.761	x>129 $\frac{mg}{L}$ (125, 134)
BOD_5 ($\frac{mg}{L}$)	0.572 (0.004)	0.554 (0.008)	60.6	0.621	0.598	x<0.976 $\frac{mg}{L}$ (0.825,1.09) x>6.19 $\frac{mg}{L}$ (4.51, 6.43)
Conductivity (μS)	0.628 (0.003)	0.638 (0.006)	65.6	0.568	0.701	x>306 μS (287,315)
Dissolved Oxygen($\frac{mg}{L}$)	0.635 (0.003)	0.640 (0.007)	67.7	0.642	0.696	x<9.39 $\frac{mg}{L}$ (8.68, 10.6)
Discharge($\frac{m^3}{s}$)	0.556 (0.004)	0.553 (0.006)	63.8	0.242	0.842	*

Hardness($\frac{mg}{L}$)	0.632 (0.003)	0.627 (0.006)	59.9	0.684	0.554	x>132 $\frac{mg}{L}$ (122, 152)
NO_3^- ($\frac{mg}{L}$)	0.581 (0.004)	0.579 (0.007)	63.4	0.453	0.728	x>1.78 $\frac{mg}{L}$ (1.63, 1.84)
pH	0.571 (0.003)	0.562 (0.006)	55.6	0.537	0.565	*
PO_4^{3-} ($\frac{mg}{L}$)	0.581 (0.004)	0.580 (0.008)	63.8	0.421	0.750	0.0642$\frac{mg}{L}$<x<7.80 $\frac{mg}{L}$ (0.0873,0.766) ᴗ (6.27,9.01) ᴖ
Water Temperature (^{O}C)	0.666 (0.003)	0.670 (0.005)	65.2	0.674	0.641	x>12.4 ^{O}C (11.3, 15.5<x<20.0)
8-variable model	0.770 (0.002)	0.709 (0.005)	78.5	0.789	0.783	
5-variable model	0.753 (0.002)	0.723 (0.006)	77.8	0.684	0.826	
4 variable model	0.750 (0.002)	0.726 (0.005)	77.8	0.695	0.821	

*Model was not significant.

¥ Values of the variables that corresponded to impairment

ˇ95% CI for the lower bound of the action value

ˆ95% CI for the upper bound of the action value.

Action values were developed for 8 significant univariate models by solving for the value of the variable when probability of impairment equals the logistic threshold. For example, the action value for alkalinity is 128 mg/L. This means that *E. coli* impairment is likely to occur when alkalinity is observed to be higher than this threshold. Action values and 95% confidence

intervals are included in Table 2.2. Action function graphs for each significant univariate model are presented in Fig. 2.S1 to aid in interpretation of Table 2.2, and summary statistics for each variable are given in Table 2.S1.

Multivariate model performance

Pearson correlations ranged from −0.269 to 0.834, with three variables identified as collinear; alkalinity, conductivity, and hardness. Conductivity was selected because of its use in previously developed fecal indicator models (Wilkes et al. 2011; Gonzalez et al. 2012; Piorkowski et al. 2013; Gonzalez and Noble 2014). The 8-variable model displayed improved accuracy on all univariate models. Variable contribution was dominated by water temperature, conductivity, and discharge in the 8-variable model, with water temperature contributing 36.4% of the information, and conductivity and discharge accounting for 22.6% and 12.1% of the information, respectively. The permutation importance for the 8-variable model demonstrated a similar pattern. A summary of accuracy metrics is shown in Table 2.2 and Table 2.3 illustrates the contribution of each variable in the multivariate models.

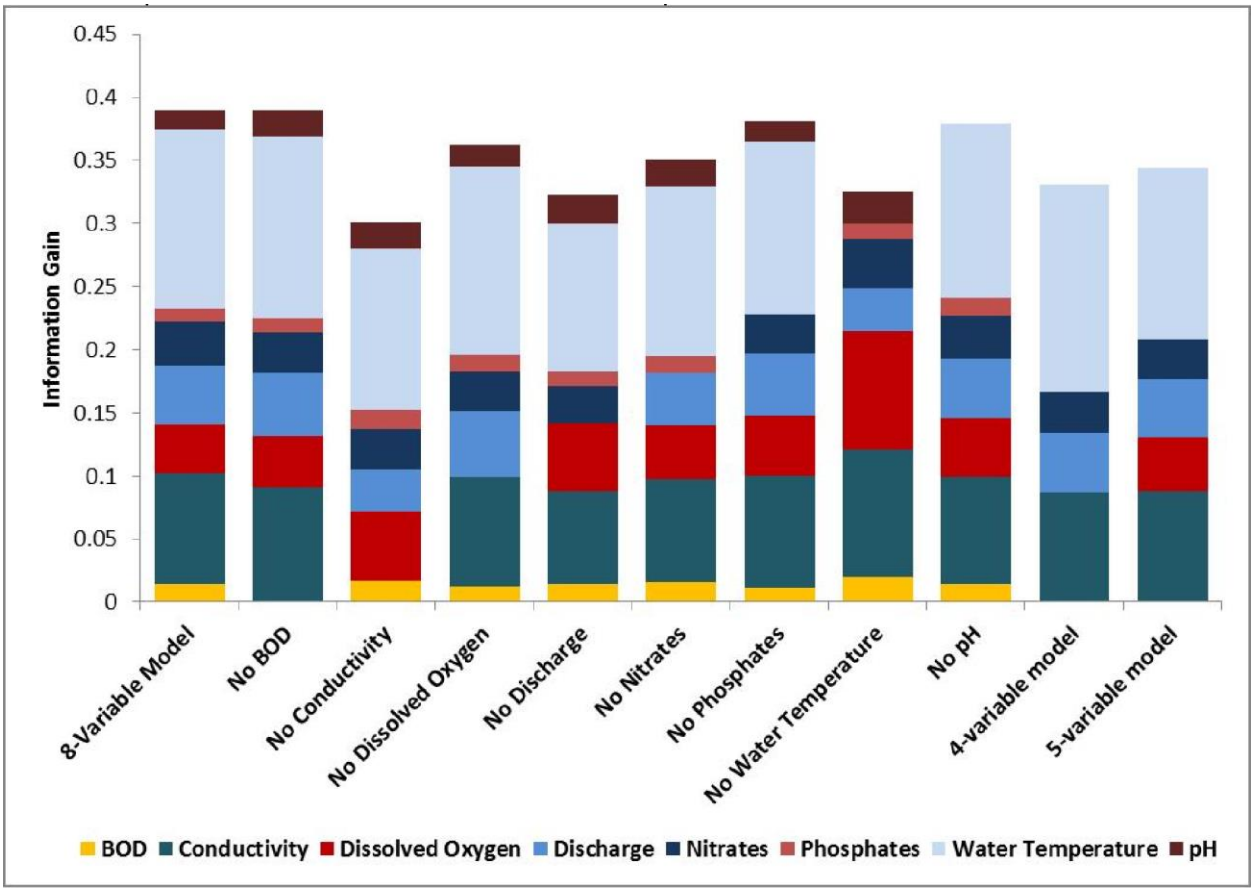

Figure 2.3. Bar graph displaying results of jack-knife sensitivity analysis. Each color represents the information gain contributed for each parameter in the model, and features are removed one at a time to assess their importance in the trimmed model.

Table 2.3. Variable contribution and permutation importance for the multivariate models, normalized to percentages.

Variable	4-variable model		5-variable model		8-variable model	
	Percent Contribution	Permutation Importance	Percent Contribution	Permutation Importance	Percent Contribution	Permutation Importance
BOD_5					3.6	5.9
Conductivity	26.2	23.0	22.6	22.3	25.6	27.5
Discharge	14.5	22.0	12.1	20.1	13.4	21.6
Dissolved Oxygen			9.9	5.2	12.3	6.6
NO_3^-	9.5	8.5	8.9	8.6	8.9	10.3
pH					3.9	1.7
PO_4^{3-}					2.7	2.5
Water Temperature	49.9	46.5	36.4	33.7	39.7	34.0

Conductivity was the most sensitive parameter based on sensitivity analysis, with other sensitive parameters including water temperature, dissolved oxygen, discharge, and NO_3^-. The removal of dissolved oxygen improved model performance based on testing AUC, justifying development of two optimized multivariate models; a 5-variable model including all sensitive parameters, and a 4-variable model that excluded dissolved oxygen.

Accuracy of the 5- and 4- variable optimized model was 77.8%. The patterns of variable contribution were consistent in each model, with water temperature accounting for most of the information gain in each model. Fig. 2.3 shows the variable contribution for the initial multivariate models, each model run during the sensitivity analysis, and the final 4-variable and 5-variable models produced. The information gain for each model is also shown within this figure.

Response surfaces were developed for each of the model runs to assess spatiotemporal trends. Each grid within the surface represents a single sample, with each sampling site representing a single column. The columns are oriented in a downstream fashion, with headwaters sites starting on the left (SC14) and sites further downstream existing on the right

(SC1). The temporal scale is represented by the rows, with each row indicating a specific sampling trip. Although the data resolution is coarse, the goal is to demonstrate the potential of visualizing trends in the probability of impairment over space and time. Fig. 2.4 displays the response surface for the estimated probability of impairment for the 4-variable model and the 5- and 8-variable models are shown in Fig. 2.S2. Classification performance for the univariate models and multivariate models is shown in Table 2.S2. Mean probabilities for the 8-, 5-, and 4-variable model were 0.338 (95% CI: 0.319, 0.358), 0.353 (0.334, 0.373), and 0.359 (0.340, 0.378). Generally, the sites influenced by the greatest amount of developed or agricultural land use (SC5–SC1) had the highest probability of impairment. August had the highest probability of impairment, followed by May, November, and February. Mean probability of impairment and associated 95% confidence intervals are shown in Table 2.S3.

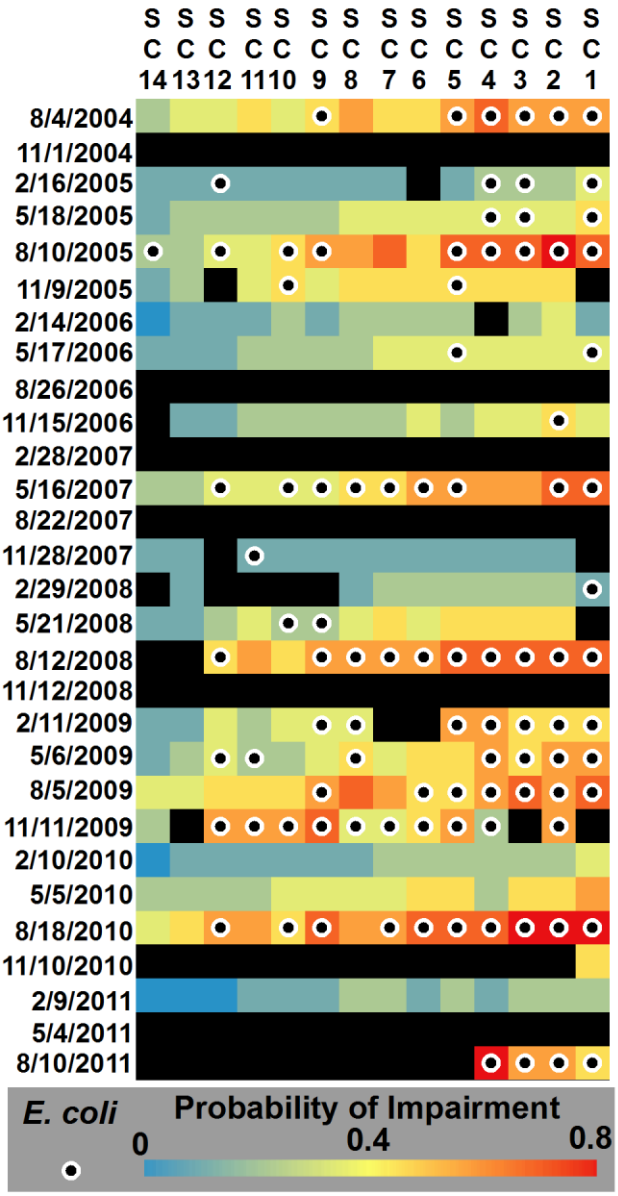

Figure 2.4. Response surface for the 4-variable Maxent model. Surface shows the probability of impairment for each sample for the monitoring program. This represents the mean probability of 100 bootstrapped runs. Rows are oriented by each sampling period, while columns represent each sampling site over the length of the stream; left to right indicates flow direction. Black cells denote samples in which a parameter was missing and were excluded from analysis, while circles with black centers represent samples in which a stream would be identified as impaired in the study.

Discussion

Over 170,000 miles of US rivers and streams are listed as pathogen impaired based on FIOs. To address these impairments, characterization of sources and transport mechanisms is necessary (United States Environmental Protection Agency 2017), and statistical models can be used as an inferential tool to overcome these issues. We applied Maxent to identify individual and interacting factors influencing *E. coli* fate and transport that resulted in impairments using univariate and multivariate approaches. In this particular stream, water temperature, conductivity, discharge, and NO_3^- were found to be the most influential group of factors driving fecal pollution. The results indicate that seasonality and agricultural runoff are the suggested causes of impairment in this watershed. Seasonality is demonstrated by influence of temperature in the models, whereas the influence of agricultural runoff is suggested by the other variables and the association between land use and *E. coli* in the watershed. Even small increases in agricultural land cause substantial increases in *E. coli* concentrations (Table 2.1), whereas similar increases in developed land do not have the same pronounced effect. This study highlights the need for multi-month sampling across a stream continuum to truly estimate spatiotemporal variability associated with impairment.

The fact that water temperature dominated the information in this model suggests that seasonality plays an important role in *E. coli* survival. Although fecal indicators and pathogens have been found to possess diverse temperature-survival relationship (Hofstra 2011; Sterk et al. 2013), the high August probability for *E. coli* impairment indicates favorable conditions for long-term survival in the summer. Warming due to climate change could exacerbate this condition by increasing those favorable conditions (Weniger et al. 1983; Atherholt et al. 1998; Patz et al. 2000; Guzman Herrador et al. 2015). However, August was not the only month with numerous *E. coli* impairments. Therefore, monitoring for FIOs only in the summer months could distort estimates of impairment in watersheds with year-round users.

Although discharge was not predictive of *E. coli* impairment alone, its interactive effect stresses the importance of both flow dependent and independent processes associated with *E. coli* impairment. Dissolved solutes such as NO_3^- and ions measured through conductivity are largely discharge-dependent; however, FIOs are not as strongly dependent on discharge. This

flow independence is due to additional ecological mechanisms such as nutrient limitation and competition (Surbeck et al. 2006; Drummond et al. 2015). Various forms of nitrogen are associated with increased concentration of FIOs in certain environments (Carrillo et al. 1985; Herrig et al. 2015), and results of the Maxent models suggest that nutrient loading in the form of NO_3^- contributes to *E. coli* impairment in Sinking Creek. Other studies have found that dissolved organic carbon can affect magnitude and extent of fecal indicators (Surbeck et al. 2010; Blazewicz et al. 2013; Cloutier et al. 2015), but this was not collected during this sampling program and was found to be insignificant using BOD_5 as a surrogate for organic pollution. This interaction between nutrient levels and fecal pollution highlights the potential for synergistic effects of different sources of pollution, suggesting a limitation of TMDL development when only considering one pollutant at a time.

Although machine learning application to microbial water quality problems is not unique, this study presents some beneficial techniques in this area of research. First, it demonstrates the ability to open the black box of Maxent, using action values to predict threshold of impairment based on a single variable. Multivariate action functions can be developed as well, but are not presented in this manuscript. The probabilistic approach to model validation and variable selection allows for inclusion of uncertainty, improving on deterministic methods traditionally used for validation and criteria for variable inclusion. Probabilistic methods have been used in TMDLs (Borsuk et al. 2002), frequency of water quality posting errors (Kim and Grant 2004), and uncertainty of different fecal indicator methodologies (Gronewold et al. 2008); this paper adds to this framework through identifying the probability of stream impairment given a set of environmental variables. This improves confidence in decision making for implementation of monitoring, management, and remediation strategies. Modeling microbial water quality is a challenge no matter the method used, but this study demonstrates that Maxent provides a valid approach to understand the factors driving impairment.

Streams are dynamic systems with multiple flow regimes, confounding an already difficult modeling process. Understanding how models behave in extreme situations is useful for regulation, monitoring, and management of these ecosystems. Over the long-term study

periods, samples from both drought and high water conditions were captured. Maxent has been suggested as a strong prediction of extreme values (Petrov et al. 2013), and this study found that Maxent sufficiently predicted impairment during the high flow sampling date of November 11, 2009. Depending on which multivariate model was used, accuracy ranged from 72.8% to 90.9% for this sampling date. Five sampling dates resulted in at least one site being dry, indicating drought-like conditions. Maxent correctly predicted impairment in these situations 62.2% to 73.0% of the time. This suggests that Maxent can be useful for certain extreme situations, but is highly dependent on the environmental variables used for prediction.

While this study presents proof of concept of using Maxent to infer source and mechanisms of impairment, there are some limitations to this study. Although the dataset has a large time scale (8 years), only collecting from 4 months makes the resolution coarse, reducing the scale at which inferences can be made. The list-wise deletion of samples before univariate modeling removed some data that could inform each of those models; however, using the same series of data in the multivariate models and list-wise deletion are commonly used procedures in statistical models. Future applications of Maxent will improve on the coarse resolution of the data by using monthly and potentially weekly sampling approaches, and research will be developed as to the best approach for handling missing data in Maxent. While AUC scores above 0.70 indicate good model fit, only considering physiochemical water quality parameters limits the potential to accurately predict impairment; however, this study demonstrates that these parameters are informative as a proof of concept for using Maxent as a modeling approach. Future areas of research include using Maxent to optimize water quality monitoring to identify causes of impairment with FIOs and specific pathogens in the most cost-effective way using a variety of microbial, chemical, and physical parameters.

It is a difficult task to develop and implement remediation strategies in watersheds with many diffuse causes of fecal impairment, but modeling can increase confidence in decision making through inferring mechanisms and sources of fecal pollution. Incorporating environmental variables into models allows for insights into the ecology of fecal indicators, identifying causes of chronic FIO impairment. Although one indicator cannot mimic the plethora of existing pathogens in water, incorporating modeling can fine tune an indicator's utility,

ultimately informing the public concerning health risks, and aiding in overcoming the shortcomings of a single indicator monitoring strategy.

Conclusions

Characterizing *E. coli* impairment is essential because of the plethora of streams polluted with fecal wastes. This study used Maxent to identify water quality parameters associated with *E. coli* impairment in a low-order, mixed-use watershed. Univariate models generated action values, or thresholds of impairment, based on single parameters, while multivariate models extracted information concerning multivariate interaction. We presented a probabilistic approach to sensitivity analysis, improving confidence in variable selection. Maxent presents a flexible machine learning approach to aid in understanding mechanisms and sources of fecal pollution as well as a host of other complex decision boundary problems. We demonstrated that:

- Models using alkalinity and water temperature were found to be either the most accurate or best performing univariate models; this stresses the importance of discharge composition and seasonality in *E. coli* impairment. Discharge, however, was not an influential univariate parameters by itself, stressing the importance of flow-independent processes that correlate with impairment.
- Sensitivity analysis indicated that the most information was lost when conductivity was removed from the multivariate models, and water temperature, discharge, dissolved oxygen, and NO_3^- represent other sensitive parameters sensitive to *E. coli* impairment in this watershed.
- Results suggest that *E. coli* impairment in this stream is driven by seasonality and agricultural runoff. This suggests that multi-month sampling along a stream continuum is essential to characterize spatiotemporal variability, importance of flow in relation to other water quality parameters, and the potential synergistic effect of nutrient and fecal pollution.
- Incorporating modeling can fine tune an indicator's utility, informing the public concerning human health risks, enhancing our understanding of FIOs, assisting in water

quality decision making, and providing input variables for quantitative microbial risk assessment.

References

Alm EW, Burke J, Spain A. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res. 37(16):3978–3982. doi:https://doi.org/10.1016/S0043-1354(03)00301-4.

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. 21st ed. Washington, DC: American Public Health Association.

Anderson KL, Whitlock JE, Harwood VJ. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl Environ Microbiol. 71(6):3041–3048. doi:10.1128/AEM.71.6.3041.

Atherholt TB, Lechevallier MW, Norton WD, Rosen JS. 1998. Effect of rainfall on Giardia and Cryptosporidium. Am Water Work Assoc. 90(9):66–80. doi:10.1002/j.1551-8833.1998.tb08499.x.

Ballestè E, Bonjoch X, Belanche LA, Blanch AR. 2010. Molecular indicators used in the development of predictive models for microbial source tracking. Appl Env Microbiol. 76(6):1789–1795. doi:10.1128/AEM.02350-09.

Bean WT, Stafford R, Brashares JS. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. (May 2011):250–258. doi:10.1111/j.1600-0587.2011.06545.x.

Belanche-Muñoz L, Blanch AR. 2008. Machine learning methods for microbial source tracking. Environ Model Softw. 23(6):741–750. doi:10.1016/j.envsoft.2007.09.013.

Benham BL, Baffaut C, Zeckoski RW, Mankin KR, Pachepsky YA, Sadeghi AM, Brannan KM, Soupir ML, Habersack MJ. 2006. Modeling bacteria fate and transport in watersheds to support TMDLs. Trans ASABE. 49(4):987–1002.

Black LE, Brion GM, Freitas SJ. 2007. Multivariate logistic regression for predicting total culturable virus presence at the intake of a potable-water treatment plant : novel application of the atypical coliform / total coliform ratio. Appl Envir Microbiol. 73(12):3965–3974. doi:10.1128/AEM.02780-06.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. 7(11):2061–2068. doi:10.1038/ismej.2013.102.

Blount ZD. 2015. The unexhausted potential of E. coli. Elife. 4:e05826. doi:10.7554/eLife.05826.

Boeckmann M, Joyner TA. 2014. Old health risks in new places? An ecological niche model for I. ricinus tick distribution in Europe under a changing climate. Health Place. 30:70–77. doi:10.1016/j.healthplace.2014.08.004.

Borsuk ME, Stow CA, Reckhow KH. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. Environ Sci Technol. 36(10):2109–2115.

de Brauwere A, Ouattara NK, Servais P. 2014. Modeling fecal indicator bacteria concentrations in natural surface waters: a review. Crit Rev Environ Sci Technol. 44(21):2380–2453.

Brion GM, Lingireddy S. 1999. A neural network approach to identifying non-point sources of microbial contamination. Water Res. 33(14):3099–3106.

Brion GM, Neelakantan TR, Lingireddy S. 2002. A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water. Water Res. 36(15):3765–3774.

Byappanahalli M, Shively D, Nevers M, Sadowsky M, Whitman R. 2003. Growth and survival of Escherichia coli and enterococci populations in the macro-alga Cladophora (Chlorophyta). FEMS Microbiol Ecol. 46(2):203–11. doi:10.1016/S0168-6496(03)00214-9.

Campolongo F, Saltelli A. 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliab Eng Syst Saf. 57(1):49–69. doi:10.1016/S0951-8320(97)00021-5.

Carrillo M, Estrada E, Hazen TC. 1985. Survival and enumeration of the fecal indicators Bifidobacterium adolescentis and Escherichia coli in a tropical rain forest watershed. Appl Environ Microbiol. 50(2):468–476.

Characklis GW, Dilts MJ, Simmons OD, Likirdopulos C a., Krometis LAH, Sobsey MD. 2005. Microbial partitioning to settleable particles in stormwater. Water Res. 39(9):1773–1782. doi:10.1016/j.watres.2005.03.004.

Cloutier DD, Alm EW, McLellan SL. 2015. Influence of land use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Appl Environ Microbiol. 81(15):4904–4913. doi:10.1128/AEM.00233-15.

Coulliete A, Money ES, Serre ML, Noble RT. 2009. Space/time analysis of fecal pollution and rainfall in an eastern north carolina estuary. Environ Sci Technol. 43(10):3728–3735.

Drummond JD, Davies-Colley RJ, Stott R, Sukias JP, Nagels JW, Sharp A, Packman AI. 2015. Microbial transport, retention, and inactivation in streams: a combined experimental and stochastic modeling approach. Environ Sci Technol. 49(13):7825–33. doi:10.1021/acs.est.5b01414.

Dwivedi D, Mohanty BP, Lesikar BJ. 2013. Estimating Escherichia coli loads in streams based on various physical, chemical, and biological factors. Water Resour Res. 49(5):2896–2906. doi:10.1002/wrcr.20265.

Dwivedi D, Mohanty BP, Lesikar BJ. 2016. Impact of the Linked Surface Water-Soil Water-Groundwater System on Transport of E. coli in the Subsurface. Water, Air, Soil Pollut. 227(9):351. doi:10.1007/s11270-016-3053-2.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Divers Distrib. 17(1):43–57. doi:10.1111/j.1472-4642.2010.00725.x.

Ferguson C, Husman AM de R, Altavilla N, Deere D, Ashbolt N. 2003. fate and transport of surface water pathogens in watersheds. Crit Rev Environ Sci Technol. 33(3):299–361. doi:10.1080/10643380390814497.

Field KG, Samadpour M. 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 41(16):3517–3538. doi:10.1016/j.watres.2007.06.056.

Fry JA, Xian G, Jin S, Dewitz JA, Homer CG, Limin Y, Barnes CA, Herold ND, Wickham JD. 2011. Completion of the 2006 national land cover database for the conterminous United States. Photogramm Eng Remote Sensing. 77(9):858–864.

Gonzalez RA, Conn KE, Crosswell JR, Noble RT. 2012. Application of empirical predictive modeling using conventional and alternative fecal indicator bacteria in eastern North Carolina waters. Water Res. 46(18):5871–5882. doi:http://dx.doi.org/10.1016/j.watres.2012.07.050.

Gonzalez RA, Noble RT. 2014. Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods. Water Res. 48:296–305. doi:http://dx.doi.org/10.1016/j.watres.2013.09.038.

Gronewold AD, Borsuk ME, Wolpert RL, Reckhow KH. 2008. An assessment of fecal indicator bacteria-based water quality standards. Environ Sci Technol. 42(13):4676–4682.

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE. 2006. Using niche-based models to improve the sampling of rare species. Conserv Biol. 20(2):501–511.

Guzman Herrador BR, de Blasio BF, MacDonald E, Nichols G, Sudre B, Vold L, Semenza JC, Nygård K. 2015. Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. Environ Heal. 14(1):29. doi:10.1186/s12940-015-0014-y.

HACH Company. 2006. Digital Titrator - Model 16900: Procedure Manual.

HACH Company. 2013. DR/890 Colorimeter Procedures Manual. :616.

Hall KK, Evanshen BG, Maier KJ, Scheuerman PR. 2014. Application of multivariate statistical methodology to model factors influencing fate and transport of fecal pollution in surface waters. J Environ Qual. 43(1):358–370. doi:10.2134/jeq2013.05.0190.

Herrig IM, Böer SI, Brennholt N, Manz W. 2015. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. Water Res. 85:148–157. doi:http://dx.doi.org/10.1016/j.watres.2015.08.006.

Hofstra N. 2011. Quantifying the impact of climate change on enteric waterborne pathogen concentrations in surface water. Curr Opin Environ Sustain. 3(6):471–479. doi:http://dx.doi.org/10.1016/j.cosust.2011.10.006.

Ishii S, Ksoll WB, Hicks RE, Sadowsky MJ. 2006. Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds. Appl Environ Microbiol. 72(1):612–621. doi:10.1128/AEM.72.1.612-621.2006.

Jamieson RC, Gordon R, Joy D, Lee H. 2004. Assessing microbial pollution of rural surface waters: A review of current watershed scale modeling approaches. Agric Water Manag. 70(1):1–17. doi:10.1016/j.agwat.2004.05.006.

Kim JH, Grant SB. 2004. Public Mis-Notification of Coastal Water Quality: A Probabilistic Evaluation of Posting Errors at Huntington Beach, California. Environ Sci Technol. 38(9):2497–2504. doi:10.1021/es034382v.

Kistemann T, Claßen T, Koch C, Dangendorf F, Fischeder R, Gebel J, Vacata V, Exner M. 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl Environ Microbiol. 68(5):2188–2197.

LaLiberte P, Grimes DJ. 1982. Survival of Escherichia coli in lake bottom sediment. Appl Environ Microbiol. 43(3):623–628.

Lasalde C, Rodriguez R, Toranzos G a, Smith HH. 2005. Heterogeneity of uidA gene in environmental Escherichia coli populations. J Water Health. 3(3):297–304.

Lipp EK, Kurz R, Vincent R, Rodriguez-Palacios C, Farrah SR, Rose JB. 2001. The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. Estuaries. 24(2):266–276. doi:10.2307/1352950.

Lozier JD, Aniello P, Hickerson MJ. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. J Biogeogr. 36(9):1623–1627. doi:10.1111/j.1365-2699.2009.02152.x.

Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM. 2011. Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. In: Proceedings of the National Academy of Sciences. Vol. 108.

Lušić DV, Kranjčević L, Maćešić S, Lušić D, Jozić S, Linšak Ž, Bilajac L, Grbčić L, Bilajac N. 2017. Temporal variations analyses and predictive modeling of microbiological seawater quality. Water Res. 119:160–170. doi:https://doi.org/10.1016/j.watres.2017.04.046.

Maier RM, Pepper IL, Gerba CP. 2009. Environmental Microbiology. 2nd ed. New York: Elsevier.

McCambridge J, McMeekin TA. 1980. Relative effects of bacterial and protozoan predators on survival of Escherichia coli in estuarine water samples. Appl Environ Microbiol. 40(5):907–911.

McKergow LA, Davies-Colley RJ. 2010. Stormflow dynamics and loads of Escherichia coli in a large mixed land use catchment. Hydrol Process An Int J. 24(3):276–289.

McLellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol. 22(12):697–706. doi:10.1016/j.tim.2014.08.002.

Molina M, Hunter S, Cyterski M, Peed LA, Kelty CA, Sivaganesan M, Mooney T, Prieto L, Shanks OC. 2014. Factors affecting the presence of human-associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. Water Res. 64:196–208. doi:https://doi.org/10.1016/j.watres.2014.06.036.

Money ES, Carter GP, Serre ML. 2009. Modern space/time geostatistics using river distances: data integration of turbidity and E.coli measurements to assess fecal contamination along the Raritan River in New Jersey. Environ Sci Technol. 43(10):3736–3742.

Patz JA, Mcgeehin MA, Bernard SM, Ebi KL, Epstein PR, Gubler DJ, Reiter P, Romieu I, Rose JB, Samet JM. 2000. The potential health impacts of climate variability and change for the United States: Executive summary of the report of the health sector of the U.S. National Assessment. Env Heal Perspect. 108(4):367–376.

Perkins TL, Perrow K, Rajko-Nenow P, Jago CF, Jones DL, Malham SK, McDonald JE. 2016. Decay rates of faecal indicator bacteria from sewage and ovine faeces in brackish and freshwater microcosms with contrasting suspended particulate matter concentrations. Sci Total Environ. 572:1645–1652. doi:https://doi.org/10.1016/j.scitotenv.2016.03.076.

Petrov V, Guedes Soares C, Gotovac H. 2013. Prediction of extreme significant wave heights using maximum entropy. Coast Eng. 74:1–10. doi:https://doi.org/10.1016/j.coastaleng.2012.11.009.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent : new extensions and a comprehensive evaluation. Ecography (Cop). 31(2):161–175. doi:10.1111/j.2007.0906-7590.05203.x.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on Machine learning. Association for Computing Machinery.

Phillips, S., Dudík, M., Schapire R. 2010. Maxent Software, version 3.3.3k.

Piorkowski G, Jamieson R, Bezanson G, Truelstrup L, Yost C. 2013. Evaluation of statistical models for predicting Escherichia coli particle attachment in fluvial systems. Water Res. 47(17):6701–6711. doi:10.1016/j.watres.2013.09.003.

Ries III KG, Newson JK, Smith MJ, Guthrie JD, Steeves PA, Haluska TL, Kolb KR, Thompson RF, Santoro RD, Vraga HW. 2017. StreamStats, version 4. Reston, VA.

Savichtcheva O, Okabe S. 2006. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res. 40(13):2463–2476. doi:10.1016/j.watres.2006.04.040.

Sinton LW, Hall CH, Lynch P a, Davies-Colley RJ. 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. Appl Environ Microbiol. 68(3):1122–1131. doi:10.1128/AEM.68.3.1122.

Smith A, Sterba-Boatwright B, Mott J. 2010. Novel application of a statistical technique, Random Forests, in a bacterial source tracking study. Water Res. 44(14):4067–4076. doi:https://doi.org/10.1016/j.watres.2010.05.019.

Sterk A, Schijven J, de Nijs T, de Roda Husman AM. 2013. Direct and indirect effects of climate change on the risk of infection by water-transmitted pathogens. Environ Sci Technol. 47(22):12648–12660.

Stott R, Davies-Colley R, Nagels J, Donnison A, Ross C, Muirhead R. 2011. Differential behaviour of Escherichia coli and Campylobacter spp. in a stream draining dairy pasture. J Water Health. 9(1):59–69. doi:10.2166/wh.2010.061.

Surbeck CQ, Jiang SC, Ahn JH, Grant SB. 2006. Flow fingerprinting fecal pollution and suspended solids in stormwater runoff from an urban coastal watershed. Environ Sci Technol. 40:4435–4441. doi:10.1021/es060701h.

Surbeck CQ, Jiang SC, Grant SB. 2010. Ecological control of fecal indicator bacteria in an urban stream. Environ Sci Technol. 44(2):631–637. doi:10.1021/es903496m.

Tennessee Department of Environmental and Conservation. 2006. Proposed total maximum daily load (TMDL) for E. Coli in the Watauga River Watershed (HUC 06010103).

Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M. 2005. Niche based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Glob Chang Biol. 11(12):2234–2250.

Tyrrel SF, Quinton JN. 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. J Appl Microbiol. 94:87–93. doi:10.1046/j.1365-2672.94.s1.10.x.

United States Environmental Protection Agency. 2012. Recreational Water Quality Criteria.

United States Environmental Protection Agency. 2017. National summary of impaired waters and TMDL information. [accessed 2017 Aug 1]. http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T.

Viau EJ, Goodwin KD, Yamahara KM, Layton BA, Sassoubre LM, Burns SL, Tong H-I, Wong SHC, Lu Y, Boehm AB. 2011. Bacterial pathogens in Hawaiian coastal streams—Associations with fecal indicators, land cover, and water quality. Water Res. 45(11):3279–3290. doi:https://doi.org/10.1016/j.watres.2011.03.033.

Vidon P, Campbell MA, Gray M. 2008. Unrestricted cattle access to streams and water quality in till landscape of the Midwest. Agric water Manag. 95(3):322–330.

Wade TJ, Pai N, Eisenberg JNS, Colford JM. 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Perspect. 111(8):1102–1109.

Weniger BG, Blaser MJ, Gedrose J, Lippy EC, Juranek DD. 1983. An outbreak of waterborne giardiasis associated with heavy water runoff due to warm weather and volcanic ashfall. Am J Public Health. 73(8):868–872. doi:10.2105/AJPH.73.8.868.

Wilkes G, Edge TA, Gannon VPJ, Jokinen C, Lyautey E, Neumann NF, Ruecker N, Scott A, Sunohara M, Topp E, et al. 2011. Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. Water Res. 45(18):5807–5825. doi:10.1016/j.watres.2011.06.021.

Winfield MD, Groisman EA. 2003. Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. Appl Environ Microbiol. 69(7):3687–3694. doi:10.1128/AEM.69.7.3687.

Yates M V. 2007. Classical indicators in the 21st century—far and beyond the coliform. Water Environ Res. 79(3):279–286.

Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 115(5):654–657.

Zweig MH, Campbell G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 39(4):561–577.

CHAPTER 3

CANONICAL VARIABLE SELECTION FOR ECOLOGICAL MODELING OF FECAL INDICATORS

DENNIS GILFILLAN, KIMBERLEE HALL, ANDREW JOYNER, PHILLIP SCHEUERMAN

Abstract

More than 270,000 km of rivers and streams are impaired due to fecal pathogens, creating an economic and public health burden. Fecal indicator organisms such as *Escherichia coli* are used to determine if surface waters are pathogen impaired, but they fail to identify human health risks, provide source information, or have unique fate and transport processes. Statistical and machine learning models can be used to overcome some of these weaknesses, including identifying ecological mechanisms influencing fecal pollution. In this study, canonical correlation analysis (CCorA) was performed to select parameters for the machine learning model, Maxent, to identify how chemical and microbial parameters can predict *E. coli* impairment and F+ somatic bacteriophage detections. Models were validated using a bootstrapping cross-validation. Three suites of models were developed; initial models using all parameters, models using parameters identified in CCorA, and optimized models after further sensitivity analysis. Canonical correlation analysis reduced the number of parameters needed to achieve the same degree of accuracy in the initial *E. coli* model (84.7%), and sensitivity analysis improved accuracy to 86.1%. Bacteriophage model accuracies were 79.2, 70.8, and 69.4% for the initial, CCorA, and optimized models, respectively; this suggests complex ecological interactions of bacteriophages are not captured by CCorA. Results indicate distinct ecological drivers of impairment depending on the fecal indicator organism used. *E. coli* impairment is driven by increased hardness and microbial activity, whereas bacteriophage detection is inhibited by high levels of coliforms in sediment. Both indicators were influenced by organic pollution and phosphorus limitation.

Introduction

More than 270,000 km of rivers and streams are impaired by fecal pathogens, creating an economic and public health burden (United States Environmental Protection Agency 2017). Because pathogens are sporadically distributed, are costly to identify, and present health risks to laboratory workers, fecal indicator organisms are used to assess public health risk and evaluate impairment (Field and Samadpour 2007). Fecal indicator organisms (FIOs) should exist whenever pathogens are present, be versatile in their use, not reproduce in the water column, and have an enteric origin (Savichtcheva and Okabe 2006). Such FIOs as fecal coliform bacteria and *Escherichia coli* are traditionally used for determining surface water pathogen impairment. Although FIOs have ultimately reduced public health risk, three key weaknesses must be improved: (i) poor quantification of public health risk, (ii) issues related to pathogen–FIO correlation, and (iii) the inability to identify sources of impairment.

Identifying public health risks based on FIOs is difficult because of geographic variability and results from epidemiologic studies are often site-specific. Although significant illness rates in children could be identified from water contaminated with >1000 *E. coli* 100 mL^{-1}, the illness rate and risk vary based on local climate and cultural conditions (Moe et al. 1991). Colford et al. (2007) found that fecal indicator bacteria did not predict health effects at a marine bathing beach (Colford et al. 2007). They suggested, however, that their results may be site-specific and influenced by the lack of human sources and negative detection of enteric viruses. Fujioka et al. (2015) noted that the 2012 Recreational Water Quality Criteria did not improve strategies to assess bathers' health risks in all types for recreational waters (Fujioka et al. 2015). Additionally, questions remain concerning human health risks associated with nonpoint sources of fecal pollution (Field and Samadpour 2007; Yates 2007).

Because there are many aquatic pathogens that express different responses to environmental conditions, identifying a single indicator is difficult and has stimulated development of several indicators. Harwood et al. (2005) evaluated the use of total coliform bacteria, fecal coliform bacteria, *Clostridium perfringens*, and F-specific coliphages to predict presence of pathogens at wastewater reclamation facilities (Harwood et al. 2005). Although pathogens were detected, no strong correlations were identified between the pathogens and

fecal indicators. Buckalew et al. (2006) conducted a 3-yr study comparing the Idexx Laboratories, Inc. Colilert assay to membrane filtration and concluded that the Colilert assay was suitable for assessing fecal pollution in surface waters (Buckalew et al. 2006). The search for an indicator that reliably predicts human health risks and indicates potential sources is an ongoing effort.

One of the primary limitations of the single indicator paradigm is limited or no source tracking potential. Microbial source tracking using genetic, chemical, or phenotypic methods are considered the gold standard for identifying human and nonhuman sources.(Shanks et al. 2016) However, there are issues with geographic stability, environmental persistence, and reproducibility with many of these techniques (Field and Samadpour 2007; Yates 2007). McLellan and Eren (2014) suggested next generation sequencing, microbiome arrays, and better understanding of gut microbiomes may improve source indicators (McLellan and Eren 2014). Pharmaceuticals and other chemical indicators such as caffeine and carbamazepine have been suggested, indicating both recent and persistent sources of human fecal pollution (Daneshvar et al. 2012).

Another challenge in using a single indicator to evaluate human exposure to fecal pathogens and develop microbial risk assessments is the influence of environmental conditions on the fate and transport of FIOs and the pathogens they represent. Fate and transport mechanisms may be affected by rainfall and stormwater runoff (Lipp et al. 2001; Kistemann et al. 2002), remobilization from sediments (Crabill et al. 1999), particle attachment (Lemarchand and Lebaron 2003), and other complex mechanisms. Ecological conditions including ultraviolet light exposure (Sinton et al. 2002), the presence of organic matter (Perkins et al. 2016), and microbial predators (McCambridge and McMeekin 1980) have also been shown to influence the survival, fate, and transport of FIOs. Savichtcheva and Okabe (2006) suggested that better understanding of factors influencing pathogen fate and transport from source to receiving streams can result in better characterization of human health risks (Savichtcheva and Okabe 2006).

Models can be used with FIOs to infer sources of fecal pollution and how environmental conditions influence fate and transport mechanisms. Atypical coliform ratios were incorporated

into multivariate logistic regression models to represent relative age of fecal pollution to identify virus presence (Black et al. 2007). Decision trees and other predictive models have been used to discriminate between human and nonhuman sources using suites of molecular indicators (Ballestè et al. 2010). Classification and regression trees were used to identify environmental and land use factors associated with pathogens, finding distinct indicators of their sporadic distribution (Wilkes et al. 2011). Classification and regression trees, regularized regression, and multivariate adaptive splines were used to investigate factors driving *E. coli* attachment to particles and virulence (Piorkowski et al. 2013). Models provide information to overcome the difficulties and deficiencies associated with using FIOs to assess pathogen impairment, providing a flexible approach that can be implemented in diverse watersheds.

In this study, we present a way to incorporate traditional water quality monitoring data into models of two fecal indicators, *E. coli* and F+ specific bacteriophages, to infer what ecological factors drive impairment in a stream. Canonical correlations analysis (CCorA) was used as an initial variable selection procedure for the machine learning model, Maxent. The multivariate technique CCorA is useful when response and explanatory variables are difficult to define, maximizing correlations between two datasets for determining dominant variables in observed water quality (Noori et al. 2012; Hall et al. 2014) or to select variables for machine learning models on ungauged monitoring sites (Khalil et al. 2011). Maxent is typically applied to ecological niche modeling, fitting a log-linear model that incorporates a least absolute shrinkage and selection operator (LASSO), or L1 regularization, to reduce unnecessary parameters in probability estimation (Phillips et al. 2004; Phillips and Dudík 2008; Elith et al. 2011). This approach incorporates mixed datasets and extracts information concerning mechanisms and sources of fecal pollution, providing a probabilistic approach decision-makers.

The goal of this study was to demonstrate this technique using data collected from a mixed-use watershed in East Tennessee. Two methods of variable selection are introduced in this model paradigm; the use of CCorA to maximize interactions between microbial and chemical datasets, and further sensitivity analysis using a leave one variable out jack-knife approach. Models were validated using nonparametric bootstrapping, giving an estimate of model performance uncertainty and rationale for additional variable reduction. These models

can identify factors driving *E. coli* impairment and bacteriophage detection, inferring ecological mechanisms, sources, and processes unique to each of these fecal indicators.

Materials and Methods

Study Area and Data Collection

Water samples were collected monthly for 12 months during 2011 from six sites on Sinking Creek, a 303(d) listed stream for *E. coli* impairment (*n* = 72 samples). Sinking Creek flows through national forest lands in its headwaters, urbanized areas in Johnson City, TN, and eventually through agricultural land before it seeps underground and enters the Watauga River. Sites were selected using a targeted sampling approach to represent a characteristic view of the watershed in terms of land use patterns, likely sources of contamination, and the influences of urbanization.

Water samples were collected in sterile 1-L bottles in triplicate for total and fecal coliform bacteria in water (TCW/FCW), and in duplicate for heterotrophic plate counts (HPC). Water samples for *E. coli* were collected in sterile 100-mL bottles (IDEXX Laboratories). Water samples for nitrates (NO_3^-), phosphates (PO_4^{3-}), ammonia (NH_3), 5-d biochemical oxygen demand (BOD_5), alkalinity, and hardness were collected in sterile 2-L bottles. Sediment samples for total and fecal coliform bacteria in sediment (TCS/FCS), acridine orange direct counts, and microbial enzyme activity (MEA) analyses were collected in 59 ml (2 oz) sterile Whirl-pak bags. All samples were transported to the laboratory on ice and analyzed within appropriate holding times. Parameters, abbreviations, and indications of fecal pollution are all shown in Table 3.1.

Microbial Analyses

The TCW, FCW, and HPC analyses were conducted according to the Standard Methods for the Examination of Water and Wastewater using membrane filtration for coliforms and R2A agar for HPC (American Public Health Association 2005). For TCS/FCS sediment analyses, 0.5 g of sediment was added to 25 mL of sterile water + 1% (v/v) Tween 80. The samples were vortexed, allowed to settle for 30 min, and filtered according to Standard Methods for the Examination of Water and Wastewater (American Public Health Association 2005). *E. coli*

concentrations were determined using the Colilert Quanti-Tray method (American Public Health Association 2005).

Samples for bacteriophage analysis were collected and analyzed in triplicate using the double-layer agar procedure described in USEPA Method 1601 using *E. coli* C3000 as the host strain (ATCC Number 15597) (United States Environmental Protection Agency 2001b). The host strain was cultured using ATCC 271 broth (10 g L^{-1} tryptone, 1 g L^{-1} yeast extract, 8 g NaCl, 10 mL L^{-1} of 10% (w/v) glucose solution, 2 mL L^{-1} of 1 M $CaCl_2$, 1 mL L^{-1} of 10 mg mL^{-1} thiamine) at 37°C. The MEA analyses included dehydrogenase (DHA), acid and alkaline phosphatases (AcidP/AlkP), galactosidase (Gal), and glucosidase (Glu); these procedures were followed as outlined by Hall et al. (2014) (Hall et al. 2014). AcidP/AlkP, Gal, and Glu activity were determined using a spectrophotometer at an absorbance of 418 nm, while DHA activity was determined at an absorbance of 460 nm. All MEA analyses were completed in triplicate. Acridine orange direct counts were performed as described by Wilson et al. (1983) (Wilson et al. 1983). Filters were mounted and fixed on slides for enumeration at 1000× using the Olympus BH2 epifluorescent microscope. One sediment sample was processed per site, and three microscopic fields were enumerated on each slide.

Chemical Analyses

NO_3^- , PO_4^{3-}, NH_3, alkalinity, and hardness analyses were performed in triplicate using colorimetric HACH™ methods and reagents as described by the manufacturer (HACH Company 2006; HACH Company 2013). The BOD_5 analyses were conducted in triplicate according to APHA (2005) (American Public Health Association 2005).

Table 3.1. Summary of parameters, abbreviations, units of measurement, and indicator role in Maxent models.

Parameter	**Abbreviation**	**Units†**	**Indication**
Fecal coliform in water	FCW	CFU 100 mL^{-1}	fecal pollution
Total coliform in water	TCW	CFU 100 mL^{-1}	heterotrophic activity
Fecal coliform in sediment	FCS	CFU 100 mL^{-1}	fecal deposition
Total coliform in sediment	TCS	CFU 100 mL^{-1}	heterotrophic activity
Colilert (*E. coli*)	*E. coli*	MPN 100 mL^{-1}	*E. coli* impairment
F+- specific bacteriophage	bacteriophage	PFU mL^{-1}	presence of viruses
Heterotrophic plate count	HPC	CFU mL^{-1}	heterotrophic activity
Acridine orange direct counts	AODC	cells g $sediment^{-1}$	heterotrophic activity
Acid phosphatase	AcidP	µg g $sediment^{-1}$	P-cycling
Alkaline phosphatase	AlkP	µg g $sediment^{-1}$	P-cycling
Dehydrogenase	DHA	µg g $sediment^{-1}$	C-cycling
Galactosidase	Gal	µg g $sediment^{-1}$	C-cycling
Glucosidase	Glu	µg g $sediment^{-1}$	C-cycling
Nitrates	NO_3^-	mg L^{-1}	nutrient runoff
Phosphates	PO_4^{3-}	mg L^{-1}	nutrient runoff
Ammonia	NH_3	mg L^{-1}	nutrient runoff
Biochemical oxygen demand	BOD_5	mg L^{-1}	organic pollution
Hardness	Hard	mg L^{-1}	runoff
Alkalinity	Alk	mg L^{-1}	runoff

† CFU, colony-forming unit; MPN, most probable number; PFU, plaque-forming unit.

Canonical Correlation Analysis

Canonical correlation analysis (CCorA) was conducted to capture between group variation within chemical and microbial parameters, describing potential relationships between

biochemical gradients within Sinking Creek. Using CCorA allows simultaneous analysis of several predictor and explanatory variables by determining the largest correlations within each dataset and between the two datasets. Linear combinations of variables within each dataset are created (canonical variates) followed by determination of the largest correlation between the two datasets, which are referred to as canonical correlations. This is repeated, producing additional combinations of canonical variates that have the next highest correlation of all possible linear combinations, but uncorrelated with the previous combinations. Canonical Variates are created for the number of variables in the smallest of the two datasets.

Canonical loadings are used to interpret the canonical structure by assessing the contribution of each variable to the structure. These loadings measure the correlation between the original variables and the sets of canonical variates (Dillon and Goldstein 1984). These strong associations were used as a variable selection procedure, and only canonical loadings >0.3 were considered to be valuable, given that this is the threshold at which approximately 10% of the variance is explained by a given coefficient (Hair et al. 1998). For detailed explanation of CCorA, readers are referred to Hall et al. (2014) (Hall et al. 2014).

Maxent

Maxent is a log-linear model commonly used for ecological niche models. Maxent uses an iterative machine learning approach that can incorporate linear, quadratic, hinge, product, and threshold feature classes to minimize the relative entropy between the distributions positive sites (contain a species) compared with the null distribution (Phillips et al. 2004). In our case, the positive sites represent detection of bacteriophages or levels of *E. coli* above the Recreational Water Quality Criteria standard of 126 MPN 100 mL^{-1}, but in a single sample rather than a geometric mean (United States Environmental Protection Agency 2012).

The Maxent model fits a Gibbs distribution, which produces a model that is maximally informative at the impairment sites and minimally informative elsewhere (Elith et al. 2011). Parameters can be transformed using a variety of transformations mentioned above, but models in this study only used linear, quadratic, and product transformations. Maxent incorporates regularization, or a LASSO penalty, to minimize overfitting; this smooths the

probability distribution, giving sparse solutions, and removing many unnecessary parameters (Phillips et al. 2004; Phillips and Dudík 2008; Elith et al. 2011). Because of regularization, Maxent fits a penalized maximum likelihood model, minimizing the relative entropy dependent on the error-bound constraints. Multiple outputs exist for the Maxent models, but for the purposes of this study, the logistic output is used, which provides a clear interpretation of probability of impairment. Readers are referred to Phillips et al. (2004), Phillips and Dudík (2008), and Elith et al. (2011) for additional information concerning Maxent (Phillips et al. 2004; Phillips and Dudík 2008; Elith et al. 2011).

Maxent Estimation

Three suites of models were developed: an initial model including all parameters, a model using biochemical gradients identified in CCorA, and a final model after further jack-knife sensitivity analyses. All significant variables (canonical loadings >0.3) were included in the bacteriophage model; since *E. coli* concentrations were used to define impairment, this parameter was not included in the *E. coli* models because it would add redundancy to the model. Maxent performance was assessed using bootstrapping with cross-validation. Bootstrapping is a nonparametric resampling technique to make inferences about a population, identifying uncertainty in model performance and providing a rationale for variable removal in sensitivity analysis (Campolongo and Saltelli 1997). One thousand bootstrapped realizations were created, and a 20% subset was selected as a testing set within each realization. For this modeling approach, all background points are used in the development of the null model, and the impairment samples were bootstrapped. Model performance was evaluated using the area under the curve (AUC) of the receiving operating characteristic (ROC). The AUC is a metric of performance for binary classification. For a series of decision boundaries between two states, the true positive prediction rate (sensitivity) and false positive prediction rates (1 – specificity) of each sample are plotted as a ROC, and the area under that ROC is then integrated. An AUC of 0.5 would indicate that the model is no better than random chance, and a value of 1.0 would indicate perfect model performance.

Many options exist for defining the logistic threshold, or decision boundary, but a threshold that maximizes test sensitivity and specificity was most appropriate for this analysis(Bean et al. 2012). Since the goal is a model that performs well on new datasets, low sensitivity sacrifices ability to recognize impairment, but low specificity indicates an overcautious model resulting in wasted resources. Accuracy was calculated using the following formula (Eq. 3.1):

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

where TP is the true positive predictions, TN is the true negative predictions, FP is the false positive predictions, and FN is the false negative predictions. Significance of the classifier was determined by calculating the χ^2 statistic for each confusion matrix, with the null hypothesis being that the classifier was no better than random chance. To assess relative contribution of individual features in the models, variable contribution was determined in each model by calculating the increase in information gain associated with a parameter during an iteration of the machine learning algorithm, averaged over 1000 model runs, and normalized to percentages.

Sensitivity Analysis

Sensitivity of parameters was assessed using a leave one variable out jack-knife procedure. Sensitive parameters were determined by re-running the models without a given parameter, calculating the training and testing information gain, and determining how much the information gain changed with exclusion of the variable. Bootstrapping cross-validation was the same in the sensitivity analysis. Parameters in which information gain decreased demonstrated a loss of information, and significance ($p < 0.01$) was determined using a one-tailed Student's *t* test comparing the CCorA model and each parameter's jack-knife model. Parameters that were found to be sensitive were included in a final optimized model using the same bootstrapping cross-validation procedures mentioned above.

Results

Summary Statistics of Fecal Indicators Organisms

Overall and site-specific summary statistics for the 17 covariates and 2 response variables are shown in Supplemental Table 3.S1. The geometric mean for *E. coli* at the sites selected for analysis was 34.0 most probable number (MPN) 100 mL^{-1} with a geometric standard deviation of 5.5. However, variation was high between the six sites. The *E. coli* concentrations were highest in the downstream Sites 2 and 4, with geometric means of 246.3 MPN 100 mL^{-1} (geometric standard deviation [GSD] = 2.4) and 152.1 MPN 100 mL^{-1} (GSD = 2.0), respectively. Site 7 was found to have the lowest *E. coli* concentrations, with a geometric mean of 6.5 MPN 100 mL^{-1} (GSD = 3.4). Exceedance of the geometric mean standard for *E. coli* occurred in nine samples for Site 2, five for Site 4, none for Site 7, two for Site 10, and one exceedance for Sites 13 and 14. Overall, 18 samples were found to exceed the geometric mean standard. All months had at least one site in exceedance of the geometric mean standard except for February, September, and December.

Although Site 2 had the highest geometric mean of 1.0 (GSD = 8.6), bacteriophages were only detected in January and November. Bacteriophages were detected 18 times, and at all sites at least once, with four detections at Sites 4, 7, and 10. Bacteriophages were not detected in March, April, July, September, and December. Supplemental Fig. 3.S1 displays the location of the sampling sites, with *E. coli* geometric mean criteria status and number of times bacteriophages were detected.

Canonical Correlation Analysis Parameter Selection

Canonical correlation analysis was performed using 18 parameters that included *E. coli* concentration, which was used as a response variable in the Maxent models. The rationale to include *E. coli* in the CCorA was because *E. coli* is an accepted FIO, and it might predict bacteriophage detections; however, bacteriophages were not considered an explanatory variable for *E. coli* impairment. The first squared canonical correlation coefficient is 0.59, and the correlation coefficient for the second axis was 0.25. Significance of canonical variate pairs was determined to be $p = 0.05$, and the first two canonical variate pairs were used to select

variables for the bacteriophage and *E. coli* models ($p < 0.001$ and $p = 0.035$, respectively). Alkalinity (0.85) and hardness (0.86) were most influential in the canonical structure of the first canonical variate pair, whereas BOD_5 (-0.62) was found to be the most influential in the second canonical variate pair (Fig. 3.1). Overall, 12 covariates were found to be influential to the canonical structure and were selected as variables for the CCorA bacteriophage models and 11 covariates were selected for the CCorA *E. coli* models.

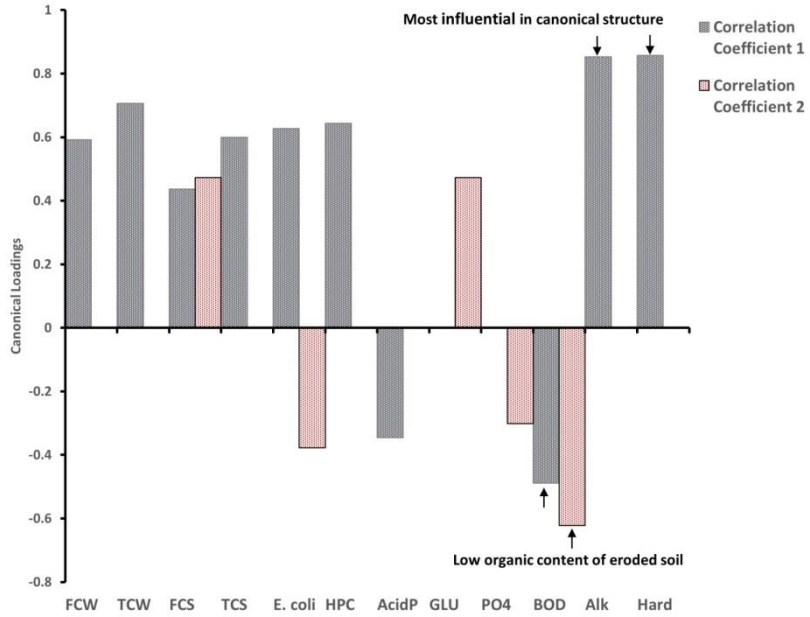

Figure 3.1. Canonical correlation analysis canonical loadings plot using the 17 parameters and 1 response variable, *E. coli*. Only significant canonical correlations are shown, with the squared canonical coefficient of the first two canonical correlations being 0.59 and 0.25. Significant loadings are only shown, with the threshold being 0.3. AcidP, acid phosphatase; Alk, alkalinity phosphatase; BOD_5, biochemical oxygen demand; FCS, fecal coliform in sediment; FCW, fecal coliforms in water; Glu, glucosidase; Hard, hardness; HPC, heterotrophic plate count; TCS, total coliform in sediment; TCW, total coliform in water

Maxent Models and Sensitivity Analysis

Table 3.2 summarizes the performance metrics for the initial, CCorA, and optimized models for both *E. coli* and bacteriophages. The initial *E. coli* model was 84.7% accurate, achieving a training AUC of 0.885 and a testing AUC of 0.720 using 16 covariates. The initial

bacteriophage model was 79.2% accurate, with a training AUC of 0.869 and testing AUC of 0.693, and using 17 covariates.

Training AUC was decreased to 0.851 in the *E. coli* CCorA model, but testing AUC was improved to 0.735. Accuracy was not affected, but sensitivity improved from 0.500 to 0.556 (data not shown). For the bacteriophage CCorA model, both training and testing AUC were lowered to 0.811 and 0.642 respectively, and the model was 70.8% accurate. Both initial and CCorA models for *E. coli* were highly significant ($p < 0.001$ for both), and the bacteriophage models were also significant for the initial and CCorA models ($p = 0.001$ and $p = 0.028$, respectively).

Sensitivity analysis indicated that AcidP, BOD_5, Glu, hardness, PO_4^{3-}, HPC, and TCW ($p < 0.001$ for all variables) were responsive parameters for the *E. coli* models. The parameters AcidP, BOD_5, Glu, TCS, and TCW were found to be significantly sensitive variables for bacteriophage models ($p < 0.001$, $p < 0.001$, $p < 0.001$, $p < 0.001$, and $p = 0.007$, respectively). The most sensitive parameter was hardness for *E. coli* models and TCS for bacteriophage models. Optimized models were developed from the seven parameters identified in the *E. coli* sensitivity analysis, while the bacteriophage sensitivity analysis revealed five parameters to be used in optimized models.

Optimized models produced lower training AUC, with values of 0.827 and 0.738 for the *E. coli* and bacteriophage models, respectively. Testing AUC were similar to the results from the CCorA, with the *E. coli* optimized model having a testing AUC of 0.731 and the bacteriophage optimized model having a testing AUC of 0.641. Accuracy improved for the *E. coli* optimized model, achieving 86.1% correct predictions. The optimized model for bacteriophage was 69.4% accurate. The optimized model for *E. coli* was strongly significant ($p < 0.001$), and the optimized model for bacteriophage was significant ($p = 0.012$).

Table 3.2. Summary of performance metrics for each of the six Maxent models.

Maxent model	Training AUC (SD)†	Testing AUC (SD)	ΔAUC (SD)	Accuracy (LT)‡
E. coli initial model	0.885 (0.030)	0.720 (0.161)	0.165 (0.167)	84.7 (0.419)
E. coli CCorA§ model	0.851 (0.037)	0.735 (0.154)	0.117 (0.161)	84.7 (0.442)
E. coli optimized model	0.827 (0.040)	0.731 (0.144)	0.096 (0.151)	86.1 (0.449)
Bacteriophage initial model	0.869 (0.038)	0.693 (0.163)	0.176 (0.165)	79.2 (0.458)
Bacteriophage CCorA model	0.811 (0.054)	0.642 (0.165)	0.170 (0.170)	70.8 (0.455)
Bacteriophage optimized model	0.738 (0.070)	0.641 (0.148)	0.097 (0.154)	69.4 (0.471)

† AUC, area under the curve; SD = Standard deviation for 1000 bootstrapped runs.

‡ LT = logistic threshold of the probability that maximizes test sensitivity and specificity.

§ CCorA, canonical correlation analysis.

Variable (Parameter) Contribution

Table 3.3 shows the variable contribution, based on increases in information gain, for each covariate included in the six Maxent models. Hardness was the dominant contributor in the initial *E. coli* model, averaging 16.2% contribution over the 1000 bootstrapped runs. Heterotrophic plate count, BOD_5, and NO_3^- were also dominant contributors, increasing the average information gain by 13.4, 8.3, and 8.3%, respectively. For the bacteriophage initial model, TCS was the dominant contributor, with 21.7% of the information gain being attributed to TCS. Other strong contributors in the full model were DHA, BOD_5, and Glu; each of these represents 9.9, 8.7, and 7.3% of the information gain in the model runs.

Hardness was the dominant contributor in both the CCorA and optimized *E. coli* models, providing 28.2 and 33.2% of the information gain. Heterotrophic plate count, BOD_5, AcidP, and

TCW contributed 13.8, 13.5, 11.5, and 10.9% of the information for the CCorA model and 17.5, 14.7, 12, and 12.7% of the information in the optimized model. Analysis of response curves shows increased probability of impairment corresponding to increased HPC, hardness, and TCW, while BOD_5 and AcidP demonstrated a negative relationship, with increased BOD_5 and AcidP corresponding to lower probability of impairment. Response curves are shown for the optimized model for *E. coli* in Supplemental Fig. 3.S2.

Total coliform in sediment was the dominant contributor in the CCorA and optimized bacteriophage models, contributing 33.5 and 53.2% of the information. Other dominant contributors for the CCorA models were Glu (13.1%), BOD_5 (10.3%), and AcidP (7.5%), and contributors to the optimized model were Glu (16.9%), AcidP (11.2%), BOD_5 (11%), and TCW (7.7%). Response curves showed decreased probability of detection from increased AcidP, Glu, and TCS, while increased BOD_5 and TCW corresponded to increased probability of detection (Supplemental Fig. 3.S3).

Table 3.3. Variable contribution for each of the six models developed, averaged over 1000 runs and normalized to percentages. Only significant variables from the canonical correlation analysis (CCorA) are included in the table (canonical loadings > 0.3).

	E. coli			**Bacteriophages**		
Variable†	**Initial model**	**CCorA model**	**Optimized model**	**Initial model**	**CCorA model**	**Optimized model**
FCW	3.8	4.6		4.4	5.7	
TCW	7.1	10.9	12.7	2.6	4	7.7
FCS	1.9	2.8		1.8	2.4	
TCS	2.8	3.6		21.7	33.5	53.2
E. coli				3.6	4.6	
HPC	13.4	13.8	17.5	4.2	5.8	
AODC	3.3			5.7		
AcidP	7.4	11.5	12	5.7	7.5	11.2
AlkP	1.3			3.3		

DHA	5.6			9.9		
Gal	2.3			3.9		
Glu	6	6	6.5	7.3	13.1	16.9
NO_3^-	8.3			5.4		
PO_4^{3-}	2.6	3.1	3.4	3	4.2	
NH_3	4.8			4.2		
BOD_5	8.3	13.5	14.7	8.7	10.3	11
Alkalinity	1.3	2		3.7	6	
Hardness	16.2	28.2	33.2	1	2.8	

† See Table 1 for variable definitions.

Probability of Impairment

Mean probability of impairment for *E. coli* models ranged from 0.282 to 0.369 for all three models, while the bacteriophage models' mean probabilities ranged from 0.322 to 0.416. Generally, *E. coli* probabilities were higher in the summer months than in the other seasons. August had the highest mean probability of impairment for the initial (0.434) and optimized (0.504) models, whereas July showed the highest probability of impairment in the CCorA model (0.473). Bacteriophage models displayed a different pattern, with November having the highest probability of impairment in the initial model at 0.466 but April having the highest mean probability in the CCorA (0.509) and optimized (0.524) bacteriophage model. Figure 3.2 shows the probability of impairment and associated 95% confidence interval for each of the six models, stratified by month.

All three *E. coli* models found the highest mean probability at Site 4, with values of 0.435 (initial), 0.438 (CCorA), and 0.553 (optimized). Using the optimized model, the estimated probability of impairment was similar at Site 2 (0.552) to Site 4. However, Site 2 had greater variability throughout the year, and the upper limit of the 95% confidence interval was 0.650 compared with 0.632 for Site 4. Using the bacteriophage models, each model predicted the mean probability of impairment to be highest at Site 14, with values of 0.427 (initial), 0.425 (CCorA), and 0.455 (optimized). However, based on the upper limits of the 95% confidence

interval Site 4 exhibited the highest probability of impairment. Mean probability of impairment and 95% confidence intervals for each model are presented in Table 3.4; these are presented for the overall stream model and stratified by month and site. To assist in visualizing the spatiotemporal trends associated with probability of impairment, response surfaces were developed for the optimized models (Fig. 3.3), and the other models are presented in Supplemental Fig. 3.S4. To read these plots, the sites are represented by each row, and the months are represented by each column. Each sampling trip represents a grid in the surface. These plots show a distinct *E. coli* hot spot in Sites 2 and 4 during the months of June, July, and August, and the standard deviation can help infer uncertainty associated with each probability. For the bacteriophage optimized model, increased probability of detection during the late fall and early winter months is shown, which corresponds to most of the bacteriophage detections. Although conditions were characteristic for bacteriophage detection in April, as indicated by the higher probability, it is unusual that no detections of bacteriophages occurred during that month.

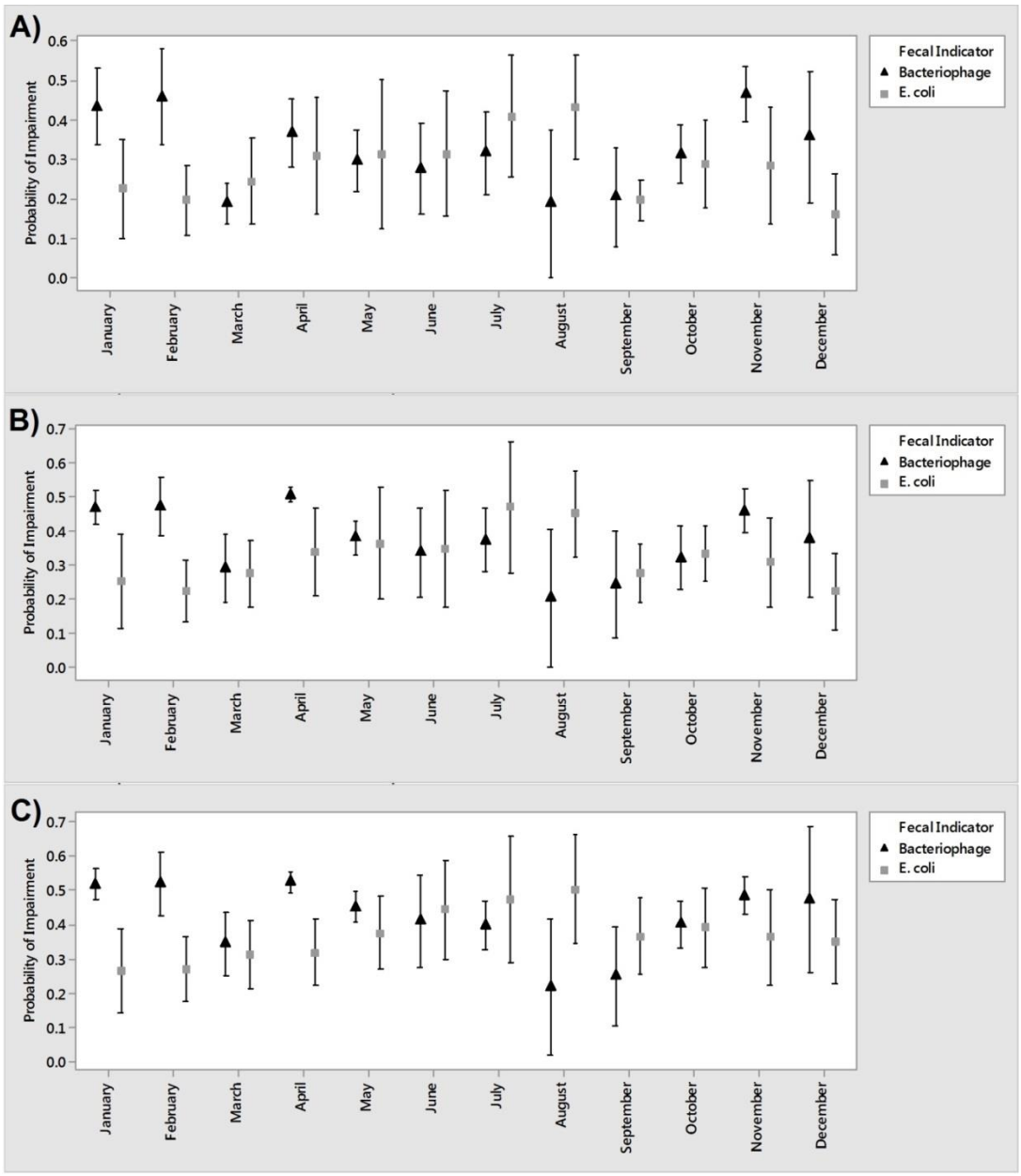

Figure 3.2. Plot of probability of impairment for the (A) initial, (B) canonical correlation analysis, and (C) optimized model with 95% confidence intervals based on each month of sampling. Triangles represent mean bacteriophage probability of detection, and circles represent mean *E. coli* probability of exceedance of the 2012 recreational water quality geometric mean criteria.(United States Environmental Protection Agency 2012)

Table 3.4. Summary of mean probability of impairment with associated 95% confidence intervals for each *E. coli* and bacteriophage model. Results are also stratified by month and site to show differences in spatiotemporal trends of impairment.

	E. coli			**Bacteriophage**		
	Initial model mean probability (95% CI)	**CCorA† model mean probability (95% CI)**	**Optimized model mean probability (95% CI)**	**Initial model mean probability (95% CI)**	**CCorA model mean probability (95% CI)**	**Optimized model mean probability (95% CI)**
Overall	0.282 (0.249,0.315)	0.324 (0.291,0.356)	0.369 (0.338,0.401)	0.322 (0.291, 0.354)	0.371 (0.339, 0.403)	0.416 (0.383, 0.449)
January	0.225 (0.099,0.351)	0.253 (0.115,0.391)	0.265 (0.140,0.390)	0.435 (0.339,0.531)	0.471 (0.421,0.520)	0.518 (0.472,0.563)
February	0.197 (0.109,0.286)	0.225 (0.134,0.317)	0.270 (0.176,0.363)	0.459 (0.339,0.580)	0.476 (0.390,0.561)	0.521 (0.429,0.612)
March	0.245 (0.137,0.354)	0.277 (0.178,0.375)	0.311 (0.212,0.411)	0.188 (0.137,0.240)	0.291 (0.192,0.390)	0.344 (0.251,0.436)
April	0.309 (0.159,0.459)	0.341 (0.213,0.469)	0.319 (0.222,0.416)	0.368 (0.280,0.455)	0.509 (0.486,0.533)	0.524 (0.493,0.554)
May	0.313 (0.123,0.504)	0.366 (0.202,0.530)	0.375 (0.268,0.482)	0.297 (0.217,0.376)	0.382 (0.333,0.432)	0.452 (0.407,0.498)
June	0.315 (0.155,0.475)	0.349 (0.179,0.520)	0.444 (0.299,0.588)	0.276 (0.159,0.393)	0.339 (0.207,0.471)	0.411 (0.276,0.547)
July	0.410 (0.254,0.566)	0.473 (0.279,0.667)	0.472 (0.287,0.657)	0.316 (0.209,0.422)	0.374 (0.281,0.468)	0.397 (0.325,0.468)
August	0.434 (0.302,0.567)	0.454 (0.328,0.580)	0.504 (0.344,0.665)	0.189 (0.002,0.376)	0.204 (0.002,0.406)	0.216 (0.165,0.416)
September	0.198 (0.146,0.249)	0.278 (0.193,0.363)	0.367 (0.257,0.478)	0.205 (0.079,0.331)	0.244 (0.087,0.400)	0.249 (0.103,0.395)
October	0.289	0.335	0.392	0.312	0.323	0.401

	(0.178,0.401)	(0.193,0.415)	(0.275,0.478)	(0.238,0.386)	(0.228,0.418)	(0.331,0.470)
November	0.285	0.310	0.363	0.466	0.461	0.485
	(0.138,0.432)	(0.178,0.443)	(0.222,0.504)	(0.396,0.537)	(0.396,0.526)	(0.431,0.539)
December	0.161	0.224	0.350	0.357	0.378	0.476
	(0.060,0.262)	(0.113,0.334)	(0.228,0.472)	(0.191,0.524)	(0.204,0.552)	(0.262,0.690)
Site 2	0.393	0.438	0.552	0.198	0.214	0.304
	(0.217,0.569)	(0.260,0.615)	(0.453,0.650)	(0.059,0.338)	(0.057,0.371)	(0.102,0.505)
Site 4	0.435	0.483	0.553	0.309	0.341	0.388
	(0.338,0.532)	(0.382,0.583)	(0.473,0.632)	(0.140,0.478)	(0.163,0.519)	(0.209,0.567)
Site 7	0.250	0.340	0.444	0.280	0.324	0.359
	(0.122,0.377)	(0.222,0.458)	(0.377,0.512)	(0.155,0.405)	(0.188,0.461)	(0.212,0.506)
Site 10	0.279	0.344	0.360	0.310	0.334	0.347
	(0.193,0.365)	(0.290,0.399)	(0.297,0.424)	(0.155,0.465)	(0.179,0.488)	(0.187,0.506)
Site 13	0.170	0.210	0.278	0.320	0.346	0.371
	(0.101,0.238)	(0.145,0.274)	(0.232,0.324)	(0.228,0.412)	(0.262,0.429)	(0.246,0.496)
Site 14	0.251	0.259	0.262	0.427	0.425	0.455
	(0.166,0.336)	(0.192,0.326)	(0.232,0.292)	(0.382,0.473)	(0.384,0.466)	(0.405,0.506)

† CCorA, canonical correlation analysis.

Discussion

The fate, transport, and source of FIOs and pathogens are challenging to untangle. Their ecology allows them to grow and persist, confounding the monitoring process(Byappanahalli, Fowler, et al. 2003; Surbeck et al. 2010). Flexible strategies are needed to overcome these difficulties. Statistical and machine learning models are potential solutions to infer mechanisms of fecal pollution, providing information concerning environmental controls and source. This paper presents a probabilistic approach blending traditional multivariate statistics and machine learning, to understand pathogen impairment. This approach is flexible, can use a variety of datasets, and can extract information in two phases: first through maximizing the correlations between two datasets, and second through Maxent modeling with sensitivity analysis to determine drivers of impairment. This technique provides an adequate approach to infer how environment shapes an indicator's niche, aiding in understanding source, fate, and transport of fecal pollution. Ecological niche models can provide important information about the behavior and stability of populations within the microbial community. Better understanding about the integration and interaction of FIOs and enteric pathogens in microbial communities can aide in understanding sources and community interactions that influence fate and can provide insight to improve the efficacy of fecal indicators to predict human health risks (Kay et al. 2008; Cloutier et al. 2015).

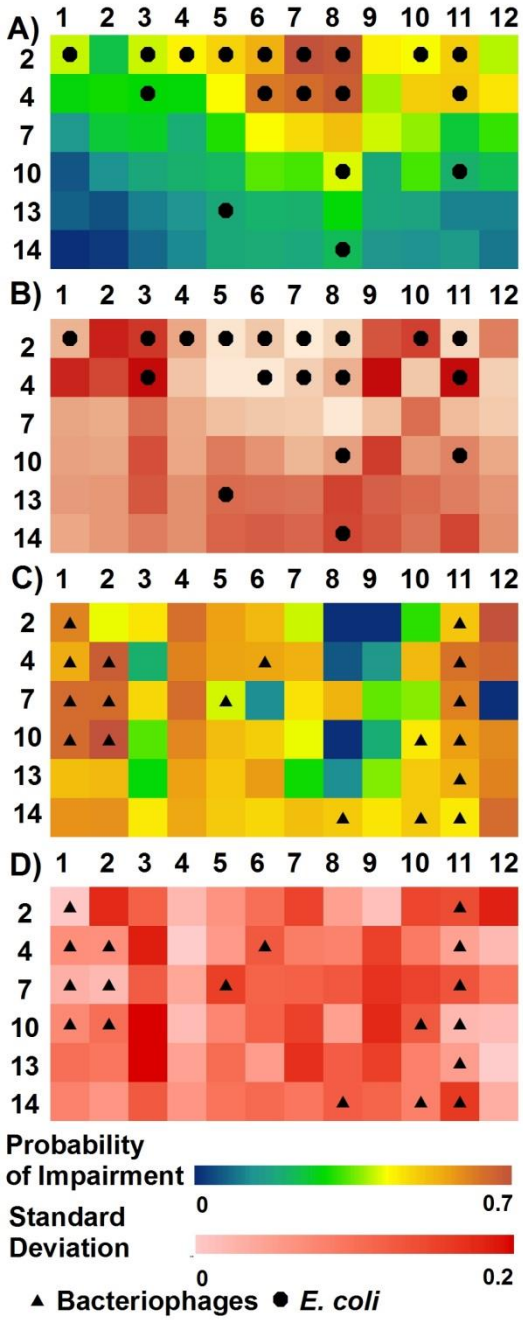

Figure 3.3. Responses surface plots (A, C) and associated standard deviation (B, D) for the optimized of (A) *E. coli* and (B) bacteriophage models. Rows in the plots represent each sampling site, and columns represent each month that a sample was taken from the site. Response surface plots and standard deviations for the initial and canonical correlation analysis models for *E. coli* and bacteriophage are presented in Supplemental Fig. S3.

Canonical correlation analysis has been used previously to identify factors influencing fate and transport of fecal pollution, and this approach can identify dominant trends within heterogeneous watersheds (Hall et al. 2014). Continuous monitoring of this system for more than 10 years has established that alkalinity, hardness, and BOD_5 are typically low with low variability during dry periods and areas are not influenced by urban and industrial point sources. This suggests that variation in these parameters is influenced by runoff entering Sinking Creek. There is a single identified permitted point source on Sinking Creek but a mix of urban, residential, forested, and agricultural areas adjacent to Sinking Creek. Although point sources, leaking wastewater lines, and failed septic tanks are possible sources, our efforts and efforts by the city wastewater department have not identified any large failures that would explain the trends in these water quality parameters. Previous use of multivariate statistical models to identify clusters of sites with similar water quality trends suggested that runoff is a large input (Hall et al. 2014). This suggestion is also supported by the local geology, terrain, and soil type. The local soils are predominantly Alfisols and Ultisols, which are moderately to heavily leached mineral soils with low organic content that experience intense weathering and leaching of calcium, magnesium, and potassium. In combination, this seems to support our conclusion that the alkalinity, hardness, and BOD_5 changes support the dominant influence of runoff. Heterotrophic activity is also influential in this stream, based on loadings of coliform species, indicating that coliform processing of organic matter influences water chemistry greatly; fecal coliform survival is enhanced with moderate amounts of organic material (Whitman et al. 2006). However, the negative loading of HPC indicates that the presence of coliform species is negatively impacted by competition. Canonical correlation analysis allows for inferences into interactions between water chemistry and microbial activity, aiding in understanding processes associated with impairment.

Probabilistic approaches to modeling water quality allow for inclusion of uncertainty, improving on deterministic methods traditionally used. These have been suggested for use in total maximum daily load development (Borsuk et al. 2002), frequency of water quality posting errors (Kim and Grant 2004), and uncertainty of different fecal indicator methodologies

(Gronewold et al. 2008). These approaches allow decision-makers to assess confidence in model predictions, improving confidence in implementation of monitoring, management, and remediation strategies. This study adds to the probabilistic approaches by estimating probability of impairment based on environmental factors, enhancing appropriation of source and processes driving fecal pollution.

E. coli impairment was driven largely by runoff, heterotrophic activity, and both P and C cycling. Hardness was dominant in both the CCorA and Maxent models, indicating that erosion of soils and geologic formations influence the presence of elevated levels of *E. coli.* Depending on oxic conditions, calcium can be used for biofilm formation, suggesting a reservoir of *E. coli* to be exchanged between water column and sediment (Mugnai et al. 2015). Microbial activity can also influence the ecology of *E. coli*, which are influenced by competition and predation. Indigenous bacteria in water, sediments, and on surfaces compete with *E. coli* for resources and space. Protozoa control *E. coli* persistence and numbers through predation (Cooley et al. 2006; Korajkic et al. 2013). High HPC indicates microbial runoff from soil erosion or a favorable niche within the suspended material from continuous inputs. Although PO_4^{3-} was not a strong contributor to impairment in the models, the influence of AcidP indicates that deposition of P in sediments exerts an influence on *E. coli* ecology. Microcosm experiments have shown that *E. coli* can grow and survive at very low concentrations of P (0.07 mg L^{-1}), below the minimum values for PO_4^{3-} found in this study (Surbeck et al. 2010). Increased AcidP and AlkP indicate phosphate limitation because microbial populations increase production of these enzymes to satisfy their need for phosphate (Vadstein et al. 1988; Hill et al. 2010; Hill et al. 2012).

Acid phosphatase had a similar influence on bacteriophage detection and *E. coli*, suggesting that Sinking Creek is P limited. Sediment coliforms exerted a strong influence on bacteriophage detections, where high levels of TCS inhibited bacteriophage detection. This suggests that high bacterial activity in sediments leads to the inability of bacteriophage to persist in the water column. One of the striking issues about variable contribution is that while BOD_5 demonstrated a net positive influence on bacteriophages, the response was inverse for *E. coli* (Supplemental Fig. 3.S1 and 3.S2). This is likely due to seasonal influences on organic matter processing, where higher levels of BOD_5 correspond to leaf litter processing in the fall and

winter. Seasonal patterns of survival were divergent for both species; *E. coli* impairment more readily occurred in the summer months, while bacteriophage detections were more readily occurring in the spring and colder months. Rainfall and flow vary greatly with seasons in this region, and their influence is part of what we capture in the seasonal variation that we have described. The role of flow conditions in total maximum daily load development is an important consideration that warrants attention in future studies, but sufficient stream flow samples were not collected in this study to evaluate with any level of statistical confidence the role of flow conditions on model development. However, these slight differences highlight a key point to consider when developing policy regarding pathogen impairment. Because of the diversity of pathogens with diverse environmental and ecological responses, multiple indicators are necessary to predict all pathogens associated public health risks.

Requiring one bacterial species, such as *E. coli,* to mimic all pathogens limits its effectiveness. Even with the plethora of alternate indicators suggested for use, strong correlations with pathogens are inconsistent, questions arise concerning quantifying human health risk, and universal source-tracking methods are still needed. It is recommended that policy shift in two ways: (i) advocating for the use of alternate indicators for watersheds to better characterize pathogen distribution, and (ii) encouraging cooperation between modelers, molecular biologists, spatial scientists, chemists, and epidemiologists to develop a geographically flexible framework to evaluate source and risk. The use of common water quality parameters and estimations of microbial activity improves understanding of the ecology behind bacteriophage detection and elevated *E. coli* levels, using modeling to inform decision making concerning fecal pollution and guiding management strategies to reduce impairment. Also, as suggested above, better understanding of the ecological behavior of fecal indicators and pathogens can inform development and implementation of more effective indicators. Pathogens in both surface and groundwater present an economic and public health burden, but the use of models alongside alternate indicators can improve response and thus reduce the negative impacts associated with fecal pollution.

Supplemental Material

The supplemental material includes a map of the stream and with sampling sites. With sites violating 2012 standards and sites with bacteriophage detections noted (Fig. 3.S1). The response curves for probability of impairment vs individual parameters for the *E. coli* models (Fig. 3.S2) and the bacteriophage models (Fig. 3.S3) are included in the supplemental material. A response surface plot for the models is included (Fig. 3.S4). Also included is a table providing the minimum and maximum values for the parameters used in the models (Table 3.S1).

References

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. 21st ed. Washington, DC: American Public Health Association.

Ballestè E, Bonjoch X, Belanche LA, Blanch AR. 2010. Molecular indicators used in the development of predictive models for microbial source tracking. Appl Env Microbiol. 76(6):1789–1795. doi:10.1128/AEM.02350-09.

Bean WT, Stafford R, Brashares JS. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. (May 2011):250–258. doi:10.1111/j.1600-0587.2011.06545.x.

Black LE, Brion GM, Freitas SJ. 2007. Multivariate logistic regression for predicting total culturable virus presence at the intake of a potable-water treatment plant : novel application of the atypical coliform / total coliform ratio. Appl Envir Microbiol. 73(12):3965–3974. doi:10.1128/AEM.02780-06.

Borsuk ME, Stow CA, Reckhow KH. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. Environ Sci Technol. 36(10):2109–2115.

Buckalew DW, Hartman LJ, Grimsley GA, Martin AE, Register KM. 2006. A long-term study comparing membrane filtration with Colilert defined substrates in detecting fecal coliforms and Escherichia coli in natural waters. J Environ Manage. 80(3):191–197. doi:10.1016/j.jenvman.2005.08.024.

Byappanahalli M, Fowler M, Shively D, Whitman R. 2003. Ubiquity and persistence of Escherichia coli in a Midwestern Coastal Stream. Appl Environ Microbiol. 69(8):4549–4555. doi:10.1128/AEM.69.8.4549-4555.2003.

Campolongo F, Saltelli A. 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliab Eng Syst Saf. 57(1):49–69. doi:10.1016/S0951-8320(97)00021-5.

Cloutier DD, Alm EW, McLellan SL. 2015. Influence of land use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Appl Environ Microbiol. 81(15):4904–4913. doi:10.1128/AEM.00233-15.

Colford JM, Wade TJ, Schiff KC, Wright CC, Griffith JF, Sandhu SK, Burns S, Sobsey M, Lovelace G, Weisberg SB, et al. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology. 18(1):27–35. doi:10.1097/01.ede.0000249425.32990.b9.

Cooley MB, Chao D, Mandrell RE. 2006. Escherichia coli O157: H7 survival and growth on lettuce is altered by the presence of epiphytic bacteria. J Food Prot. 69(10):2329–2335.

Crabill C, Donald R, Snelling J, Foust R, Southam G. 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. Water Res. 33(9):2163–2171. doi: 10.1016/S0043-1354(98)00437-0.

Daneshvar A, Aboulfadl K, Viglino L, Broséus R, Sauvé S, Madoux-Humery A-S, Weyhenmeyer GA, Prévost M. 2012. Evaluating pharmaceuticals and caffeine as indicators of fecal contamination in drinking water sources of the Greater Montreal region. Chemosphere. 88(1):131–139. doi: 10.1016/j.chemosphere.2012.03.016.

Dillon WR, Goldstein M. 1984. Multivariate analysis: methods and applications. John Wiley & Sons.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Divers Distrib. 17(1):43–57. doi:10.1111/j.1472-4642.2010.00725.x.

Field KG, Samadpour M. 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 41(16):3517–3538. doi:10.1016/j.watres.2007.06.056.

Fujioka SR, Solo-Gabriele MH, Byappanahalli NM, Kirs M. 2015. U.S. Recreational Water Quality Criteria: A Vision for the Future. Int J Environ Res Public Health. 12(7):7752–7776. doi:10.3390/ijerph120707752.

Gronewold AD, Borsuk ME, Wolpert RL, Reckhow KH. 2008. An assessment of fecal indicator bacteria-based water quality standards. Environ Sci Technol. 42(13):4676–4682.

HACH Company. 2006. Digital Titrator - Model 16900: Procedure Manual.

HACH Company. 2013. DR/890 Colorimeter Procedures Manual. :616.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. 1998. Multivariate data analysis. Upper Saddle River, NJ: Prentice Hall.

Hall KK, Evanshen BG, Maier KJ, Scheuerman PR. 2014. Application of multivariate statistical methodology to model factors influencing fate and transport of fecal pollution in surface waters. J Environ Qual. 43(1):358–370. doi:10.2134/jeq2013.05.0190.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, Rose JB. 2005. Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. Appl Environ Microbiol . 71 (6):3163–3170. doi:10.1128/AEM.71.6.3163-3170.2005.

Hill BH, Elonen CM, Seifert LR, May AA, Tarquinio E. 2012. Microbial enzyme stoichiometry and nutrient limitation in US streams and rivers. Ecol Indic. 18:540–551.

Hill BH, McCormick FH, Harvey BC, Johnson SL, Warren ML, Elonen CM. 2010. Microbial enzyme activity, nutrient uptake and nutrient limitation in forested streams. Freshw Biol. 55(5):1005–1019.

Kay D, Crowther J, Fewtrell L, Francis C a., Hopkins M, Kay C, McDonald AT, Stapleton CM, Watkins J, Wilkinson J, et al. 2008. Quantification and control of microbial pollution from agriculture: a new policy challenge? Environ Sci Policy. 11(2):171–184. doi:10.1016/j.envsci.2007.10.009.

Khalil B, Ouarda T, St-Hilaire A. 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J Hydrol. 405(3-4):277–287.

Kim JH, Grant SB. 2004. Public Mis-Notification of Coastal Water Quality: A Probabilistic Evaluation of Posting Errors at Huntington Beach, California. Environ Sci Technol. 38(9):2497–2504. doi:10.1021/es034382v.

Kistemann T, Claßen T, Koch C, Dangendorf F, Fischeder R, Gebel J, Vacata V, Exner M. 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl Environ Microbiol. 68(5):2188–2197.

Korajkic A, Wanjugi P, Harwood VJ. 2013. Indigenous Microbiota and Habitat Influence Escherichia coli Survival More than Sunlight in Simulated Aquatic Environments. Appl Environ Microbiol. 79(17):5329–5337.

Lemarchand K, Lebaron P. 2003. Occurrence of Salmonella spp. and Cryptosporidium spp. in a French coastal watershed: relationship with fecal indicators. FEMS Microbiol Lett. 218(1):203–209.

Lipp EK, Kurz R, Vincent R, Rodriguez-Palacios C, Farrah SR, Rose JB. 2001. The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. Estuaries. 24(2):266–276. doi:10.2307/1352950.

McCambridge J, McMeekin TA. 1980. Relative effects of bacterial and protozoan predators on survival of Escherichia coli in estuarine water samples. Appl Environ Microbiol. 40(5):907–911.

McLellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol. 22(12):697–706. doi:10.1016/j.tim.2014.08.002.

Moe CL, Sobsey MD, Samsa GP, Mesolo V. 1991. Bacterial indicators of risk of diarrhoeal disease from drinking-water in the Philippines. Bull World Health Organ. 69(3):305.

Mugnai R, Sattamini A, dos Santos JAA, Regua-Mangia AH. 2015. A survey of Escherichia coli and Salmonella in the Hyporheic Zone of a subtropical stream: their bacteriological, physicochemical and environmental relationships. PLoS One. 10(6):e0129382.

Noori R, Karbassi A, Khakpour A, Shahbazbegian M, Badam HMK, Vesali-Naseh M. 2012. Chemometric Analysis of Surface Water Quality Data: Case Study of the Gorganrud River Basin, Iran. Environ Model {&} Assess. 17(4):411–420. doi:10.1007/s10666-011-9302-2.

Perkins TL, Perrow K, Rajko-Nenow P, Jago CF, Jones DL, Malham SK, McDonald JE. 2016. Decay rates of faecal indicator bacteria from sewage and ovine faeces in brackish and freshwater microcosms with contrasting suspended particulate matter concentrations. Sci Total Environ. 572:1645–1652. doi:https://doi.org/10.1016/j.scitotenv.2016.03.076.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent : new extensions and a comprehensive evaluation. Ecography (Cop). 31(2):161–175. doi:10.1111/j.2007.0906-7590.05203.x.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on Machine learning. Association for Computing Machinery.

Piorkowski G, Jamieson R, Bezanson G, Truelstrup L, Yost C. 2013. Evaluation of statistical models for predicting Escherichia coli particle attachment in fluvial systems. Water Res. 47(17):6701–6711. doi:10.1016/j.watres.2013.09.003.

Savichtcheva O, Okabe S. 2006. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res. 40(13):2463–2476. doi:10.1016/j.watres.2006.04.040.

Shanks OC, Kelty CA, Oshiro R, Haugland RA, Madi T, Brooks L, Field KG, Sivaganesan M. 2016. Data Acceptance Criteria for Standardized Human-Associated Fecal Source Identification Quantitative Real-Time PCR Methods. Besser TE, editor. Appl Environ Microbiol. 82(9):2773–2782. doi:10.1128/AEM.03661-15.

Sinton LW, Hall CH, Lynch P a, Davies-Colley RJ. 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. Appl Environ Microbiol. 68(3):1122–1131. doi:10.1128/AEM.68.3.1122.

Surbeck CQ, Jiang SC, Grant SB. 2010. Ecological control of fecal indicator bacteria in an urban stream. Environ Sci Technol. 44(2):631–637. doi:10.1021/es903496m.

United States Environmental Protection Agency. 2001. Method 1601: Male-specific (F +) and Somatic Coliphage in Water by Two-step Enrichment Procedure.

United States Environmental Protection Agency. 2012. Recreational Water Quality Criteria.

United States Environmental Protection Agency. 2017. National summary of impaired waters and TMDL information. [accessed 2017 Aug 1]. http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T.

Vadstein O, Jensen A, Olsen Y, Reinertsen H. 1988. Growth and phosphorus status of limnetic phytoplankton and bacteria. Limnol Oceanogr. 33(4):489–503.

Whitman R, Nevers MB, Byappanahalli M. 2006. Examination of the watershed-wide distribution of Escherichia coli along southern Lake Michigan: An integrated approach. Appl Environ Microbiol. 72(11):7301–7310. doi:10.1128/AEM.00454-06.

Wilkes G, Edge TA, Gannon VPJ, Jokinen C, Lyautey E, Neumann NF, Ruecker N, Scott A, Sunohara M, Topp E, et al. 2011. Associations among pathogenic bacteria, parasites, and

environmental and land use factors in multiple mixed-use watersheds. Water Res. 45(18):5807–5825. doi:10.1016/j.watres.2011.06.021.

Wilson JT, McNabb JF, Balkwill DL, Ghiorse WC. 1983. Enumeration and Characterization of Bacteria Indigenous to a Shallow Water-Table Aquifer. Ground Water. 21(2):134–142. doi:10.1111/j.1745-6584.1983.tb00710.x.

Yates M V. 2007. Classical indicators in the 21st century—far and beyond the coliform. Water Environ Res. 79(3):279–286.

CHAPTER 4

MICROBIAL COMMUNITY METABOLISM ASSOCIATED WITH POLLUTION ALONG A STREAM CONTINUUM

DENNIS GILFILLAN, PHILLIP SCHEUERMAN

Abstract

Microbial activity is essential to stream ecosystems because of secondary production of organic matter and nutrient cycling. Microbial community metabolism can be evaluated through monitoring single source carbon degradation using a plethora of carbon substrates allows for the creation of metabolic fingerprints of a community. Using phenotypic techniques can provide insights into functional shifts in a community, indicating ecosystem change and potential anthropogenic disturbances. This study incorporates microbial metabolism, in the form of Biolog EcoPlates, and multivariate statistical techniques to characterize microbial metabolism along a stream continuum. Water and sediment samples were collected on a monthly basis at eight sites from November 2016 to October 2017 on Sinking Creek, a tributary of the Watauga River in Tennessee. Carbon substrate utilization was measured for 31 substrates to develop metabolic fingerprints, and water samples were analyzed for *Escherichia coli*, nitrates, and biochemical oxygen demand. Sediment-water interactions were analyzed using canonical correlation analysis to summarize between group variations. Dominant substrates were used to inform multiple linear regression models for three different types of pollution; fecal (*E. coli*), nutrient (NO_3^-), and organic matter (BOD_5). Results indicate both site-specific and seasonal differences in overall metabolism, substrate groupings, and individual substrates. Sediment-water interactions were summarized by mostly metabolic inhibition, especially in sediments. Metabolic fingerprints for degree of pollution were developed using four substrates for fecal and nutrient, while only two substrates were identified to model organic matter pollution. Microbial activity shifts along the river continuum, and characterizing these shifts can assist in identifying anthropogenic stressors on water quality not readily seen in traditional monitoring strategies.

Introduction

The river continuum concept (RCC) is an important perspective for characterizing relationships between abiotic and biotic components of stream ecology. The RCC posits that organic matter and nutrient processing is governed by physical gradients, and biota are strategically organized to maximize energy efficiency longitudinally along this continuum (Vannote et al. 1980; Creed et al. 2015; Tornwall et al. 2015). Anthropogenic activity also influences stream ecosystem dynamics, and identifying these disturbances is essential for ecological risk assessment. Agricultural intensity was found to be associated with shifts in the macroinvertebrate community composition in the Pomahaka river in New Zealand, and monitoring this intensity may provide a useful tool for identifying conditions for decline in stream health (Harding et al. 1999). Ecological fish guilds have been suggested as good indicators of stream integrity; a successional gradient of fish community structure matches many predictions of the RCC (Aarts et al. 2003). Studies surrounding the RCC have largely focused on macroinvertebrates (Grubaugh et al. 1997; Harding et al. 1999; Tomanova et al. 2007; Rosi-Marshall et al. 2016) and fish (Aarts et al. 2003; Chick et al. 2006; Tornwall et al. 2015), but less importance has been focused on the role of microorganisms within the RCC (Savio et al. 2015).

Microbial activity is essential to stream ecosystems because of secondary production of organic matter and nutrient cycling (Garland and Mills 1991; Garland 1997; Christian and Lind 2007; Tiquia 2010). While microbes are the most abundant and diverse organisms on Earth, information is limited concerning patterns governing their spatial distribution (Whitman et al. 1998). One common theory suggests that microbes hold tremendous dispersal potential, leading to a cosmopolitan distribution governed by environmental stressors (Beijerinck 1913; Fierer and Jackson 2006; Fierer et al. 2007; Fierer 2008; Nemergut et al. 2011). Although modern phylogenetic techniques have enhanced understanding of complex microbial community structures (Alban and Tiedje 2006), phenotypic approaches can assess traits such as microbial metabolism, identifying traits integral to ecosystem function(Green et al. 2008; Krause et al. 2014). Microbial communities respond subtly to environmental shifts, making community metabolism a valuable indicator of ecosystem degradation (Garland and Mills 1991;

Boivin et al. 2002; Maier et al. 2009; Tiquia 2010). Prokaryotes are especially important in headwaters streams because low nutrient concentrations and high proportions of dissolved nutrients in organic form exist in these oligotrophic environments, and these conditions favor heterotrophic bacteria over phytoplankton and bacterial predators (Cotner and Biddanda 2002).

Microbial community metabolism can be evaluated through monitoring single source carbon degradation; using a plethora of carbon substrates allows for community level physiologic profiles (CLPPs) to be produced, creating metabolic fingerprints of each sample (Garland 1997; Preston-Mafham et al. 2002). CLPPs are a low cost method, providing insights into community physiology, with applications in a wide variety of research areas, including water quality (Choi and Dobbs 1999; Christian and Lind 2007; Tiquia 2010), dairy waste activated sludge (Gryta et al. 2014), fecal source tracking (Hagedorn et al. 2003), soil functionality (Acosta-Martinez et al. 2007; Rutgers et al. 2016), and constructed wetlands (Zhang et al. 2010). Initial interest in metabolic fingerprinting came about because there is no inoculation phase; samples can be added directly to wells either as a water sample or a suspension (Garland and Mills 1991; Garland 1997; Preston-Mafham et al. 2002). While weaknesses such as the effects of inoculum density, appropriate incubation temperature, and culture-dependence exist, suggestions to overcome these limitations have been proposed (Preston-Mafham et al. 2002; Christian and Lind 2006; Stefanowicz 2006; Christian and Lind 2007). Despite these limitations, carbon degradation is a cost-effective method to investigate metabolic potential of the cultivatable portion of the community capable of utilizing given substrates, serving as a proxy to characterize various patterns of preferential substrate degradation based on spatiotemporal and environmental gradients (Grover and Chrzanowski 2000; Preston-Mafham et al. 2002; Stefanowicz 2006).

Using phenotypic techniques such as CLPPs provides insights into functional shifts in a community, indicating ecosystem change and potential anthropogenic disturbances. Industrial pollution has been shown to increase antibiotic resistance in fecal bacteria, with mercury in sediments and proximity to the source associated with higher resistance to antibiotics (McArthur and Tuckfield 2000). Chlorophyll concentrations have been negatively correlated

with metabolic diversity in heterotrophic bacteria, suggesting that increased exposure to autotrophs reduces plasticity of microbial communities (Sala and Estrada 2006; Tiquia 2010). Additionally, aquatic ecosystems provide multiple habitats for microorganisms to colonize, and the exchange of surface and groundwater within these environments creates potential for community transport (Storey et al. 1999; Maier et al. 2009). By considering metabolic interactions between these habitats, complex dynamics of organic matter processing can be better characterized.

The hyporheic zone is defined as the stream habitat in which the sediments are hydrologically linked to the stream channel, and hyporheic exchange describes the interactions between surface and ground water in this zone. Ground and surface water ecosystems are highly dependent on one another; surface water provides energy to groundwater systems through deposition of dissolved and particulate matter, while hyporheic exchange inputs nutrients from groundwater systems to surface water (Ghiorse and Wilson 1988; Ford and Naiman 1989). As hyporheic zone bacteria incorporate and oxidize organic carbon, electron acceptors are used in order of decreasing free energy yield, and dissolved oxygen is readily depleted because it is the energetically favored electron acceptor (Christian and Lind 2007). As dissolved oxygen is depleted, consortia of bacteria shift assemblages to utilize less energetically favorable acceptors; this lowers redox potential and increases concentrations of chemically reduced nutrients (Liikanen and Martikainen 2003). Understanding sediment-water dynamics between microbial communities is essential to characterize the flow of organic matter within the river continuum, and can assist in identifying processes linked to sources of pollution.

As a consequence of human activities, pollutants are discharged into streams via point and nonpoint sources, altering the natural state of the river continuum through the deposition of fecal wastes, excess nutrients, organic matter, and other types of pollution. Fecal pollution monitored through indicator organisms such as *Escherichia coli* presents a public health risk due to pathogens in human and animal wastes, representing the primary cause of noncompliance in the United States' waterbodies (Field and Samadpour 2007; McLellan and Eren 2014; United States Environmental Protection Agency 2017). The well documented ability of *E.coli* to reside in sediments and beach environments is due to rich organic matter supplies, reduced predation

and protection from light inactivation (LaLiberte and Grimes 1982; Alm et al. 2003; Byappanahalli, Fowler, et al. 2003; Jamieson, Douglas M Joy, et al. 2005; Whitman et al. 2006; Maier et al. 2009). Pyrosequencing has revealed that sediment housed more pathogens than the corresponding water column from mixed use sites (Ibekwe et al. 2013). Many of these pathogens are introduced into the sediments through runoff, so characterizing the interactions between aquatic and benthic microbial communities can improve identifying patterns of fecal pollution, deposition, and persistence. Nutrient pollution degrades ecosystem services through eutrophication, reducing the utility of surface waters (Carpenter et al. 1998). Nutrients can be introduced into ecosystems through runoff from land surfaces as well as through microbially-mediated cycling within water and sediments (Song et al. 2004; Qu et al. 2005). Bacteria in the hyporheic zone often input nutrients in greater quantities than allochthonous sources such as agricultural and urban runoff (Heinen and McManus 2004; Song et al. 2004). Untangling these natural and anthropogenic drivers of nutrient pollution can assist in remediation efforts and identifying sources in rivers and streams.

Multivariate statistical analyses are common practice in microbial ecology, and CLPPs are typically analyzed using some form of unconstrained analysis such as principal component analysis or cluster analysis (Zak et al. 1994; Preston-Mafham et al. 2002). Hypothesis-based constrained analysis are useful for many ecologically relevant questions concerning the RCC (Buttigieg and Ramette 2014). For example, canonical correlation analysis (CCorA) can assess the interactions between two different types of samples, i.e., water or sediment, to summarize and extract information concerning between group variation (Khalil et al. 2011; Hall et al. 2014). Multiple linear regression (MLR) can be used to assess what metabolic fingerprints correspond to degree of pollution in a watershed. While minimal studies exist concerning the links between pollution gradients and metabolism (Harbott and Grace 2005; Walsh et al. 2005), microbial activity presents an alternative tool to assess these relationships, aiding in source identification and strategizing remediation.

This study incorporates CLPPs and multivariate statistical techniques to characterize microbial metabolism along a stream continuum in the Southern Appalachian Mountains. Spatiotemporal analysis identified unique patterns of substrate utilization; these were further

investigated using CCorA and MLR. CCorA was used to summarize interactions between aquatic and benthic microbial communities to utilize various carbon substrates to better characterize metabolic exchange between these environments. Dominant degraded substrate patterns were then used to inform MLR models on three pollutants; fecal pollution as measured by *E. coli* concentrations, nutrient pollution in the form of nitrates (NO_3^-), and organic pollution in the form of biochemical oxygen demand (BOD_5).

Methods

Data Collection

Samples were collected on a monthly basis at eight sites from November 2016 to October 2017 on Sinking Creek, a tributary of the Watauga River in Tennessee, USA (Fig. 4.S1). Water samples were collected by hand in sterile 2-oz Whirl-pak bags for heterotrophic plate counts (HPC) and CLPP analysis; HPC were collected in duplicate while CLPP was collected in triplicate. Water samples for Colilert® were collected in sterile 100 mL bottles (IDEXX Laboratories, Westbrook, Maine). Sediment samples were collected in 2-oz Whirl-paks and stored in a similar fashion to the water samples for analysis for acridine orange direct counts (AODC) and sediment CLPPs. Triplicate 2 L samples of water were collected for the analysis of NO_3^- and BOD_5 within the water column. Samples were stored on ice and delivered to the laboratory and sample preparation was performed within 6 hours for each analysis.

Microbial and Chemical Analysis

HPCs were conducted according to Standard Methods for the Examination of Water and Wastewater (American Public Health Association 2005). *E. coli* concentrations were determined using the Colilert® Quanti-Tray method (American Public Health Association 2005). One Colilert® sample was collected and processed per site. AODCs were performed as described by Wilson et al. (1983) (Wilson et al. 1983; Hall et al. 2014).. Water samples for NO_3^- were analyzed by ion chromatography (American Public Health Association 2005). BOD_5 was determined using the 5-day BOD_5 test (American Public Health Association 2005).

Biolog EcoPlates™ were used to develop the CLPPs. These are microtiter plates with 31 ecologically relevant carbon substrates and a control well used to correct for effects of background color (Garland and Mills 1991; Garland 1997). As the microbial communities degrade the respective carbon sources, wells change from clear to purple due to a reaction with the tetrazolium dye included in each well (Garland and Mills 1991). The benefit of these plates is that they allow for samples to be inoculated in triplicate, allowing in-plate replication. This replication increases the probability that the CLPP developed indeed represents the microbial community studied (Stefanowicz 2006). For water samples, each well was inoculated with 150 µL of water sample. For sediment samples, 0.3 g of sediment was vortexed with 30 mL of Phosphate Buffer Saline solution and allowed to settle before inoculation. Each well was then inoculated with 150 µL of the suspension. Both water and sediment sample were incubated at 25^oC. Plates were read immediately after inoculation and every 24 hours for 120 hours with a MicroSkan MCC plate (Thermo-Scientific, Waltham, MA USA)reader using a 595 nm filter.

Constructing the CLPPs

Optical density (OD) for each individual well were corrected by subtracting the OD from the control within each replicate, and those wells in which the OD was less than the control well were recorded as zero. Mean values of OD for each substrate were calculated, and average well color development (AWCD) was calculated. AWCD is the mean of the OD for all 31 substrates and is calculated using Eq. (4.1), where OD_i is the corrected optical density for substrate i and n is the number of substrates utilized within the CLPP.

$$AWCD = \frac{\sum_{i=1}^{n} OD_i}{n} \quad (4.1)$$

In addition to determining the OD for each substrate and overall AWCD, substrates were grouped into the following chemical structures: amines, amino acids, carbohydrates, carboxylic acids, phenolic compounds, phosphorylated compounds, and polymers. ODs were averaged within each group to obtain a group AWCD (GAWCD) for the respective chemical structures.

Data Analysis

Pearson's correlation analysis was performed first to determine if there were any effects due to inoculum density. This was calculated by comparing the AWCD for each incubation period (24, 48, 72, 96 and 120 hours of incubation), the HPC for the aquatic samples, and the AODC for the sediment samples. AWCDs were compared for spatiotemporal differences using separate one-way analysis of variance (ANOVA) for each time of incubation. One-way ANOVA was also performed on individual substrate OD and the GAWCD of each chemical group using 120 hours of incubation as the time for comparison. Seasons are defined using the astronomical definition, where winter starts on the winter solstice, spring the spring equinox, summer the summer solstice, and autumn (fall) the autumnal equinox.

Initial multivariate data analysis included all 62 substrate ODs (31 for water, 31 for sediment). the selected time point for analysis was 120 hours of incubation. Variance inflation factors (VIF) were calculated for each substrate using separate MLRs to reduce collinearity of the CLPPs. MLR is an extension of ordinary least squares regression with multiple explanatory variables for a single response variable. The assumption in MLR is that there are gradients associated with changes in the explanatory variables that can be used to predict changes in the

response variable (Buttigieg and Ramette 2014). VIFs are used to assess the collinearity within a set of independent variables, and this is calculated by performing separate MLRs in which each independent variable is selected as a new independent variable, and all other independent variables are used as covariates (Marquardt 1970). An explanation of variance (R^2) is calculated, which represents the amount of variance is explained by the other independent variables. The VIF is then calculated as follows (Eq. (4.2)):

$$\text{VIF} = \frac{1}{1-R^2} \quad (4.2)$$

An VIF of 1 would indicate there is no collinearity present in the independent variables, and the higher the values, the more collinear the independent variables possess. Variables were removed with a VIF above 5, indicating that the other variables can explain 80% of the variance of a given variable. VIF were recalculated, and if the significantly associated variables were still showing collinearity, they were removed from analysis. Two multivariate analyses were applied after reduction of collinearity; CCorA and MLR.

CCorA was used to explore the possible interactions of the aquatic and benthic microbial metabolism. CCorA is a useful multivariate technique when interactions between two datasets exists, maximizing correlations between them for determining dominant variables in observed water quality (Noori et al. 2012; Hall et al. 2014) or as a variable selection technique (Shu and Ouarda 2007; Khalil et al. 2011). In this study, CCorA captured between group variations of the sediment and water CLPPs, describing functional interaction between aquatic and benthic communities. CCorA allows simultaneous analysis of several predictor and explanatory variables by determining the largest correlations within each data set and between the two data sets. Linear combinations of variables within each data set are created (canonical variables) followed by determination of the largest correlation between the two data sets, which are referred to as canonical correlations. These canonical correlations are a measure of the strength of association between the two data sets (Johnson and Wichern 1992).

The process results in the successive extraction of canonical variables so the second canonical variable pair (CVP) is the second most highly correlated pair out of all possible linear combinations that are uncorrelated with the first CVP, resulting in the generation of CVPs of gradually decreasing explanatory power. Canonical loadings are used to interpret the canonical

structure by assessing the contribution of each variable to the structure. These loadings measure the correlation between the original variables and the corresponding canonical variable. These loadings reflect the variance shared between the canonical variables and the original variables, with higher absolute values of loadings demonstrating stronger associations (Dillon and Goldstein 1984). These strong associations were used as a variable selection procedure, and only canonical loadings greater than 0.3 were considered to be valuable, given that this is the threshold at which approximately 10% of the variance is explained by a given coefficient (Hair et al. 1998; Hall et al. 2014).

MLR models were developed to predict three different measures of pollution; Fecal pollution as measured by *E. coli* concentrations, nutrient pollution based on nitrates concentrations in water, and organic pollution as measured by BOD_5. *E. coli* concentrations were log transformed before analysis. Explanatory variables were selected through the previously mentioned CCorA procedure, and a feed-forward regression technique whereby the substrate with the largest correlation with the response variable is considered first for entry into the model. If the coefficients of the model are significantly different from zero ($p<0.05$), the variable is included in the model. This is repeated until no more variables can be included based on the significance criteria.

Results

Effects of inoculum density and spatiotemporal metabolic patterns

Variation in inoculum density can influence interpretation of CLPPs (Garland 1997; Christian and Lind 2007); however, no significant correlations were found between inoculum density and color development in water samples ($p>0.06$) or sediment samples ($p>0.20$) at each of the readings. This suggests that cell density in this case was responsible for less of the observed variation than were other factors (Choi and Dobbs 1999). AWCD in water samples were significantly different between sites for each incubation period ($p<0.002$), except for the first 24-hour readings ($p = 0.097$) (Table 4.1). Generally, SC12 and SC13 showed the highest AWCD at each of the incubation periods (Figure 4.1a). Significant differences were found at all incubation periods ($p<0.015$) for sediment sampling sites, with SC1 having the highest AWCD at each of the time periods (Table 4.1).

Seasonal differences were also found in both water and sediment AWCD. In water samples, only seasonal differences were detected at the 24 hours incubation period ($p<0.001$), with 2 distinct groups of seasons; spring and summer AWCD comprise one group while autumn and winter AWCD comprise a second group (Table 4.1). Sediment samples showed pronounced seasonal variation at all incubation periods ($p<0.001$); the same distinct groupings were found within the sediment samples. Generally, AWCD was higher in the spring and summer for both the water and sediment samples (Figure 4.1b).

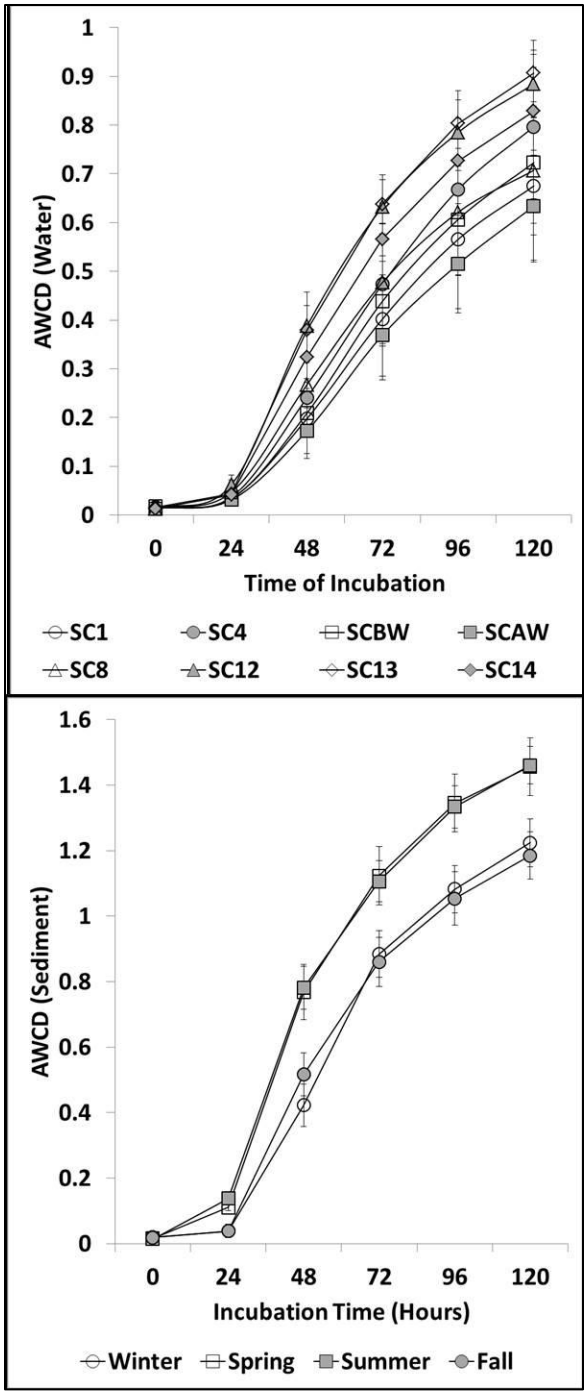

Figure 4.1. Average well color development (AWCD) over the incubation period for water (a) and sediment samples (b). Water AWCD are grouped by site while sediment AWCD are group by season to display the more significant trends. Error bars represent the 95 % confidence intervals of the mean AWCD.

GAWCD at 120 hours of incubation was found to be different between sites for amino acids ($p < 0.001$) (Figure 2a), carboxylic acids ($p < 0.001$) (Figure 2b), phosphorylated compounds ($p = 0.008$), and polymers ($p = 0.013$), while only phosphorylated compounds were found to have seasonal differences ($p = 0.001$). Sediment GAWCD site differences were found in carbohydrates ($p = 0.031$), amino acids ($p = 0.015$), amines ($p < 0.001$) (Figure 2c), and phenolic compounds ($p = 0.005$) (Figure 2d), while all GAWCD were found to be significantly difference based on season ($p = 0.008$).

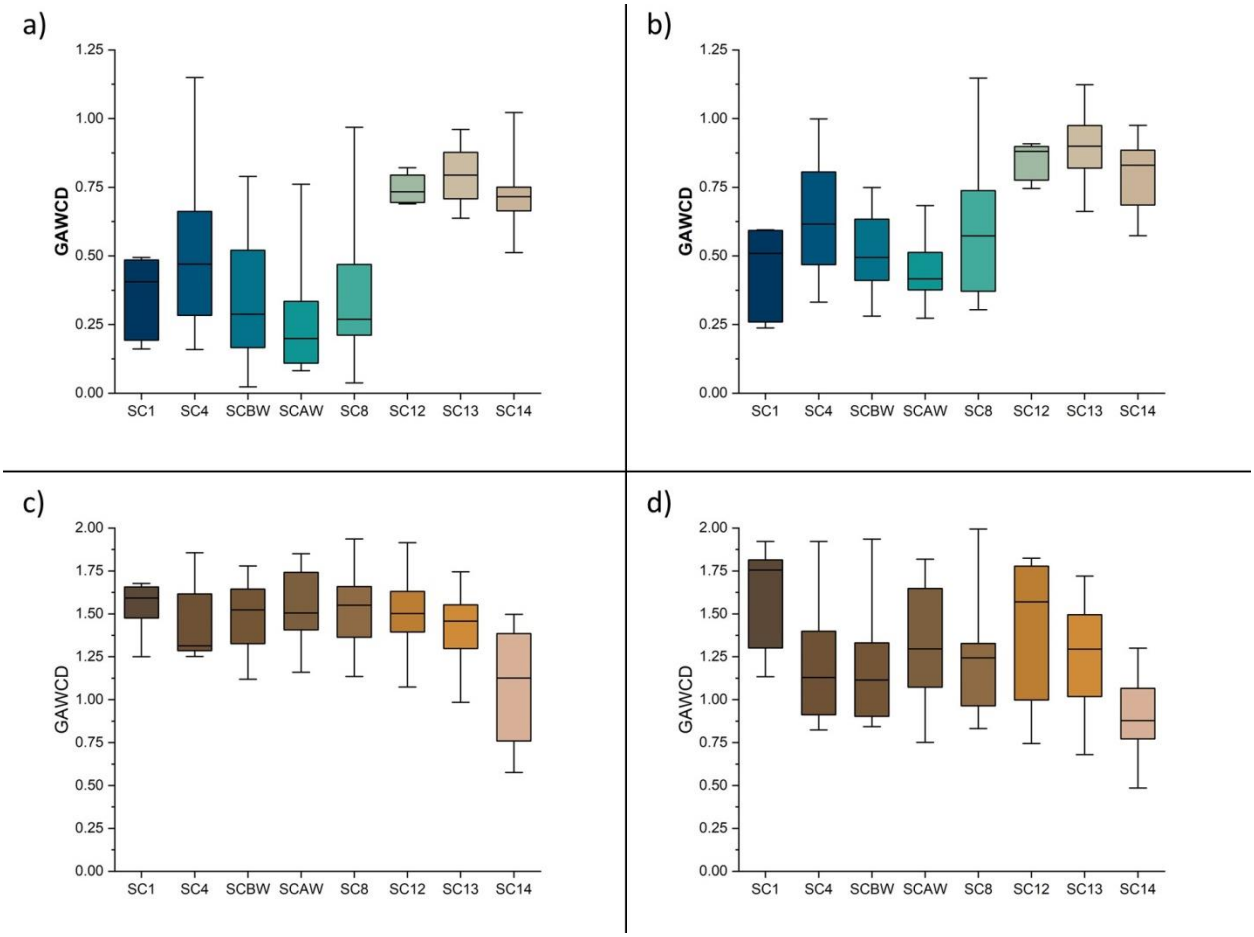

Figure 4.2. Box and whisker plots of the group average well color development (GAWCD) for amino acids in water (a), carboxylic acids (b) in water, amines in sediment (c), and phenolic compounds (d). Boxes represent the 25th and 75th percentile, with the lines denoting the median. Whiskers represent the 5th and 95th percentiles.

Nineteen substrates in water showed site specific differences that were significant, while 11 substrates in water showed significant seasonal differences. Twelve substrates in sediment showed significant site specific differences, while 6 substrates demonstrated significant season variation. Degree of significance is shown in Fig. 4.3 as well as how the substrates were grouped based on chemical structure.

Table 4.1. Mean AWCD of each site and season and the associated standard error (SE) for water and sediment samples during the incubation period. Geometric means and geometric standard deviation (GSD) for the standard plate counts and AODC are shown for water samples and sediment samples, respectively. ^{W} Water samples, ^{S} Sediment samples

	AWCD (SE) 24hrs	AWCD (SE) 48 hrs	AWCD (SE) 72 hrs	AWCD (SE) 96 hrs	AWCD (SE) 120 hrs	HPC (GSD) (CFU/mL)	AODC (GSD) (10^8 cells/g)
SC1^{W}	0.034 (0.005)	0.198 (0.033)	0.402 (0.048)	0.566 (0.058)	0.675 (0.062)	480 (1.53)	
SC4^{W}	0.037 (0.004)	0.241 (0.033)	0.474 (0.056)	0.668 (0.067)	0.797 (0.067)	469 (1.24)	
SCBW^{W}	0.033 (0.004)	0.209 (0.023)	0.439 (0.042)	0.606 (0.052)	0.723 (0.057)	421 (1.51)	
SCAW^{W}	0.031 (0.003)	0.173 (0.021)	0.369 (0.042)	0.515 (0.045)	0.634 (0.052)	429 (1.50)	
SC8^{W}	0.047 (0.011)	0.267 (0.048)	0.477 (0.055)	0.621 (0.059)	0.707 (0.060)	455 (1.44)	
SC12^{W}	0.062 (0.009)	0.388 (0.030)	0.633 (0.024)	0.785 (0.029)	0.884 (0.030)	308 (1.41)	
SC13^{W}	0.043 (0.007)	0.379 (0.023)	0.638 (0.027)	0.803 (0.031)	0.907 (0.030)	199 (1.45)	
SC14^{W}	0.041	0.324	0.566	0.726	0.829	248	

	(0.007)	(0.023)	(0.033)	(0.040)	(0.041)	(1.57)	
SC1[S]	0.147	0.801	1.144	1.370	1.482		
	(0.025)	(0.089)	(0.066)	(0.068)	(0.054)		
SC4[S]	0.098	0.677	1.019	1.218	1.355		1.96
	(0.015)	(0.055)	(0.046)	(0.052)	(0.053)		(2.31)
SCBW[S]	0.096	0.705	1.056	1.259	1.383		1.36
	(0.013)	(0.057)	(0.052)	(0.053)	(0.046)		(2.42)
SCAW[S]	0.102	0.693	1.062	1.273	1.403		1.97
	(0.019)	(0.061)	(0.053)	(0.059)	(0.046)		(1.99)
SC8[S]	0.096	0.710	1.063	1.275	1.399		1.52
	(0.015)	(0.056)	(0.067)	(0.069)	(0.062)		(2.39)
SC12[S]	0.092	0.713	1.120	1.348	1.463		
	(0.016)	(0.081)	(0.068)	(0.069)	(0.062)		
SC13[S]	0.060	0.575	0.943	1.172	1.298		1.97
	(0.011)	(0.051)	(0.046)	(0.051)	(0.051)		(1.69)
SC14[S]	0.051	0.466	0.839	1.047	1.169		1.72
	(0.011)	(0.057)	(0.059)	(0.066)	(0.067)		(2.00)
Winter[W]	0.025	0.224	0.474	0.621	0.726	646	
	(0.003)	(0.022)	(0.048)	(0.055)	(0.055)	(1.32)	
Spring[W]	0.050	0.299	0.511	0.676	0.773	487	
	(0.005)	(0.026)	(0.030)	(0.032)	(0.030)	(2.22)	
Summer[W]	0.048	0.283	0.521	0.709	0.831	203	
	(0.005)	(0.024)	(0.034)	(0.040)	(0.043)	(2.48)	
Fall[W]	0.030	0.254	0.481	0.620	0.733	300	
	(0.003)	(0.027)	(0.038)	(0.040)	(0.044)	(1.65)	
Winter[S]	0.038	0.423	0.884	1.083	1.224		
	(0.004)	(0.040)	(0.042)	(0.041)	(0.041)		
Spring[S]	0.111	0.768	1.123	1.345	1.457		

	(0.007)	(0.032)	(0.031)	(0.032)	(0.028)	
Summer[S]	0.138	0.781	1.106	1.333	1.461	2.18
	(0.010)	(0.031)	(0.036)	(0.039)	(0.035)	(3.14)
Fall[S]	0.038	0.517	0.860	1.054	1.185	1.38
	(0.006)	(0.031)	(0.034)	(0.034)	(0.035)	(5.62)

Sediment-water interactions

VIF indicated that 12 substrates exhibited high collinearity (VIF>5) within the aquatic CLPPs and these were removed from subsequent analysis; only four substrates were found to be highly collinear within the benthic CLPPs. This left 19 substrates for the water dataset and 27 substrates for the sediment dataset. 19 CVPs were developed as linear combinations of the original variables. Of these pairs, only two CVPs were found to be significant ($p = 0.001$ and $p = 0.014$, for CVP 1 and CVP 2, respectively). The squared canonical correlation coefficient for CVP 1 was 0.753 and the canonical correlations for CVP 2 was 0.699. This can be interpreted in the following ways; in the first canonical variable pair, the 75.3 % of the variation in the water canonical variable 1 can be explained by the variation in the sediment canonical variable 1, and while 69.9% of the variation in water canonical variable 2 can be explained by the variation in sediment canonical variable 2.

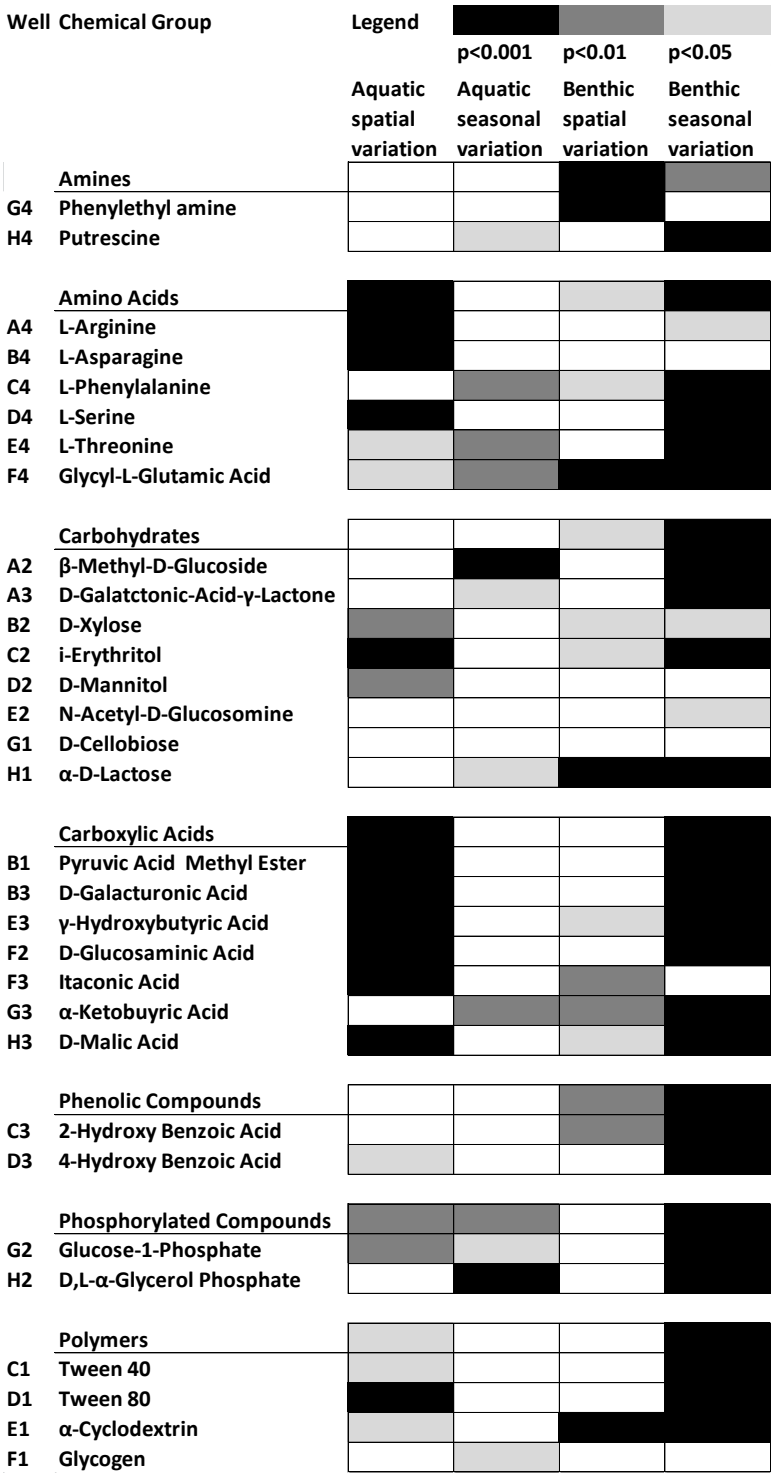

Figure 4.3. Plot of individual substrates, their corresponding chemical groupings, and the patterns of spatial and seasonal variation. Degree of significance is denoted by color of the corresponding cell, with darker colors denote lower p-values. All substrates showed some significant spatial and seasonal variation except for D-Cellobiose.

Loadings measure the strength of associations between the original variables and their canonical variables, and substrates with loadings greater than 0.3 were retained for future analyses. Overall, metabolic inhibition influenced the canonical structure more than substrate utilization (Figure 4.4). Substrates identified with the greatest influence on the canonical structure were metabolic inhibition of α-Ketobuyric Acid, Glycyl-L-Glutamic Acid, γ-Hydroxybutyric Acid, and α-D-Lactose in sediment; metabolic inhibition of Pyruvic Acid Methyl Ester and D-Mannitol in water were also influential (Table 4.2). The ability to degrade D-Malic acid in water was found to be weakly influential as well as the ability to degrade N-Acetyl-D-Glucosomine, D-Cellobiose, and L-Serine in sediment.

Table 4.2. Canonical loadings of each of the substrates with at least 1 loading above the 0.3 threshold.

Well	**Substrate**	**Water Canonical Loadings (CVP 1)**	**Water Canonical Loadings (CVP 2)**	**Sediment Canonical Loadings (CVP 1)**	**Sediment Canonical Loadings (CVP 2)**
A2	β-Methyl-D-Glucoside	-0.491			
A3	D-Galactonic-Acid-γ-Lactone	-0.376			
A4	L-Arginine			-0.386	
B1	Pyruvic Acid Methyl Ester	-0.632			
B2	D-Xylose	-0.522		-0.404	
C3	2-Hydroxy Benzoic Acid			-0.313	
D2	D-Mannitol	-0.588			
D4	L-Serine				0.319
E1	α-Cyclodextrin			-0.512	
E2	N-Acetyl-D-Glucosomine			0.464	-0.316

E3	γ-Hydroxybutyric Acid			-0.586
E4	L-Threonine			-0.481
F1	Glycogen	-0.443	-0.393	
F3	Itaconic Acid			-0.346
F4	Glycyl-L-Glutamic Acid			-0.617
G1	D-Cellobiose		-0.385	0.323
G2	Glucose-1-Phosphate			-0.386
G3	α-Ketobuyric Acid			-0.681
G4	Phenylethyl amine			-0.454
H1	α-D-Lactose			-0.583
H3	D-Malic Acid	0.353		-0.486
H4	Putrescine	-0.314		-0.540

Relationships with pollution

Models were used to assess potential relationships between microbial activity and degree of pollution, and significant predictive models were developed for degree of fecal ($p=0.003$), nutrient ($p<0.001$), and organic pollution ($p=0.001$) (Table 4.3). Results for the fecal pollution model identified four substrates that predict levels of log transformed *E. coli* values: pyruvic acid methyl ester in water, α-D Lactose in sediment, D-Xylose in water, and Putrescine in water. These substrates explained 40.6% of the variance associated with log transformed *E. coli* values. Nutrient pollution models identified D-Malic Acid in water, L-Threonine in sediment, L-Arginine in sediment, and Pyrivic Acid Methyl Ester in water as significant predictor of nitrates, explaining 58.3% of the variance. BOD_5 was found to be predicted by the patterns of two substrates; Putrescine in sediment and D-Malic Acid in water. This explained only 13.9% of the total variance. The degradation of D-Xylose affected the degree of fecal pollution the strongest and the degradation of D-Malic acid in water affected the amount of organic pollution strongest; however, the inhibition of D-Malic acid in water affected the amount of nutrient pollution the strongest.

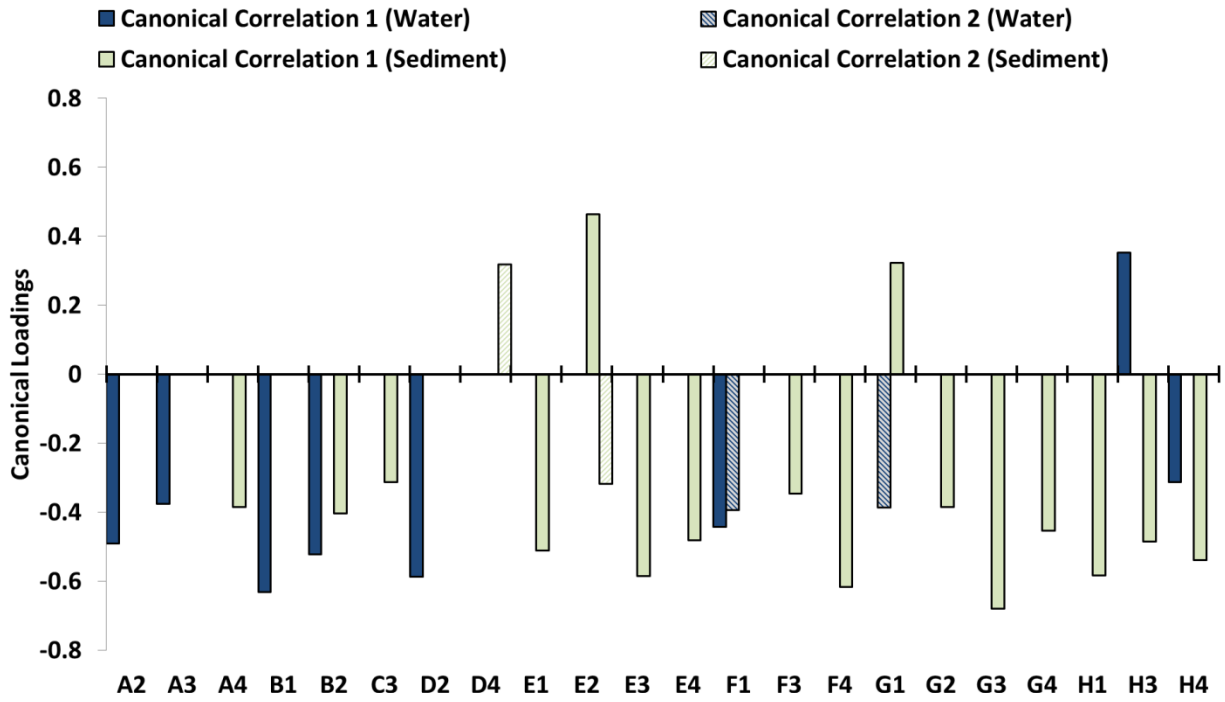

Figure 4.4. Canonical loadings for the sediment-water interactions. Only influential (loadings > 0.3) are shown in the figure, and only significant canonical axes are shown. Labels for each substrate correspond to the well in which the substrate is present, and these abbreviations can be found in figure 3 or table 2.

Table 4.3. Performance metrics, included substrates, and standardized coefficients for the three pollution models developed. Asterisks indicate p-values of less than 0.05 (*), 0.01 (**), and 0.001 (***). ^{W} water samples ^{S} sediment samples

Model	**Adjusted R^2**	**Substrate**	**Standardized coefficients**
Fecal (Log *E. coli*)	0.406**	$B1^{w}$	0.252*
		$B2^{w}$	0.350**
		$H1^{s}$	0.294**
		$H4^{w}$	-0.235*
Nutrient (NO_3^-)	0.583***	$A4^{s}$	-0.264**
		$B1^{w}$	0.224*
		$E4^{s}$	0.324***
		$H3^{w}$	-0.529***
Organic (BOD_5)	0.139**	$H3^{w}$	0.368*
		$H4^{s}$	0.222**

Discussion

Riverine systems transport anthropogenic and natural sources of nutrients and organic matter downstream, linking the water column to the landscape through surface runoff. These terrestrial inputs cause shifts in microbial community activity, structure, and function; due to their role in nutrient cycling (Christian and Lind 2007), characterizing these complex relationships can aid in identifying and remediating human-driven ecologic degradation. This study used EcoPlates to assess changes in metabolism along a stream continuum within aquatic and benthic assemblages of microorganisms, identifying relationships between these habitats and different types of pollutants. Phenotypic approaches such as EcoPlates are low cost proxies to characterize microbial metabolism. While the substrates existent in the EcoPlates might not represent the exact organic matter present *in situ*, EcoPlates provide as set of environmentally relevant substrates that can discriminate between community metabolism effectively (Grover and Chrzanowski 2000).

Generally the most upstream sites (SC14, SC13, and SC12) had higher rates of aquatic metabolism, suggesting these assemblages were more equipped to utilize a wide variety of allochthonous substrates as described in the RCC (Vannote et al. 1980; Rosi-Marshall et al. 2016). However, the heterogeneity of inputs from mixed land use after SC12 do not allow for a clear established gradient in terms of overall metabolism, as measured through AWCD. Sinking Creek is an intermittent stream because some sites (SC1 and SC12) experienced dry periods. Water availability in soils stimulates microbial activity because of increases in nutrients and organic matter (Belnap et al. 2004; Williams 2006). The assemblages at SC1 demonstrated pronounced increase in overall metabolism, consistent with Timoner et al. (2014) who reported that rehydration events lead to increased functional diversity in sediment biofilms (Timoner et al. 2014); a similar but not as drastic difference was also seen in SC12. These desiccation events cause physiologic adaptations and community sorting based on drought tolerance (Fierer et al. 2003; Schimel et al. 2007; Timoner et al. 2014).

Although explanations for seasonal preference of organic substrates are limited (Pettine et al. 1999), this study identifies two temporal regimes for community metabolism; one consisting of spring and summer, and the second consisting of fall and winter, similar to Urakawa et al. (2013) (Urakawa et al. 2013). These regimes were more pronounced in the sediments. Oest et al. 2018 found a different seasonal pattern in sediment; Summer and Fall sediment communities possessed more versatile substrate profiles than Spring and Winter sediment communities (Oest et al. 2018). Duarte et al. 2016 found that seasonal changes in water chemistry influenced microbial activity and diversity more so than warming alone (Duarte et al. 2016). Also, seasonal changes in anthropogenic inputs such as fertilization, landscaping activities, or overflows of municipal and industrial wastewater discharge could alter the composition of allochthonous organic matter, shaping bacterial assemblages based on these patterns (Sala et al. 2006; Sala et al. 2008; Tiquia 2010).

Hyporheic sediments are known to have heterogeneous physiochemical conditions, affecting nitrogen cycling due to complex processes requiring diverse environmental conditions (Storey et al. 1999). Amino acids and amines are both nitrogen-rich, with available ammonium side chains for assimilation by heterotrophic bacteria (Pettine et al. 1999), and strong

preferences for these substrates suggests that nitrogen limitation is a strong influence on community function. High use of amino acids or amines suggests nitrogen deprivation (Oest et al. 2018), while low use would suggest the presence more energetically favorable sources of nitrogen, and could indicate inputs of fertilizers, agricultural wastes, and potential sewage leakages (Carpenter et al. 1998).

Bacterial communities processed phosphorylated organic compounds uniquely between sites, suggesting different conditions for assimilation of another essential nutrient, phosphorus. Low utilization of these substrates would indicate more freely available forms of phosphorus, and potential sources of pollution from urban, industrial, or agricultural sources (Carpenter et al. 1998). Phosphorylated compounds were the only substrate group which exhibited aquatic seasonal variation, suggesting either changes in terrestrial inputs, in-stream availability of phosphorus, or environmental factoring governing nutrient cycling. Productivity in aquatic ecosystems increases in the warmer months, and in lake ecosystems it is found to correlate with total phosphorus, suggesting that the production-phosphorus relationship stimulate changes in microbial metabolism (Hanson et al. 2003). This could explain the increased utilization of phosphorylated compounds in comparisons to cooler, less productive months.

Carboxylic acids are frequently found in aquatic systems; they are a product of bacterial fatty acid degradation (Christian and Lind 2007), photochemical degradation of high molecular weight dissolved organic carbon (Bertilsson and Tranvik 2000), or as the end product of fermentative metabolism (Ding and Sun 2005). Headwaters sites (SC12, SC13, SC14) processed these compounds more rapidly than downstream sites, in line with Berggren et al. (2010) who found bacterial carbon demand is preferential to carboxylic acids in forested streams (Berggren et al. 2010). Organic acids such as carboxylic acids degrade more quickly in aerobic settings based on marine studies, independent of bacterial abundance (Ding and Sun 2005); however, Christian and Lind (2007) found that anaerobic conditions induced the highest activity of carboxylic acid degradation in the sediment-water interface (Christian and Lind 2007). The reduced ability to process carboxylic acids as the stream flows downstream could be due to increased productivity as hypothesized by the RCC, creating higher molecular weight dissolved

organic carbon that not as easily degraded by microbial assemblages (Cotner and Biddanda 2002).

The influence of water stress from desiccation and rewetting of sediments at two sites (SC1 and SC12) is further supported by the abilities of these assemblages to utilize phenolic compounds. Leflaive et al. 2008 found that phenolic compounds were the preferred substrates in nutrient depleted aquatic microcosms (Leflaive et al. 2008). The rewetting and desiccation experienced by these assemblages foster communities that rapidly respond to water availability, affording opportunistic species to capitalize on a wide variety of complex organic compounds such as phenolic compounds (Timoner et al. 2014).

Hyporheic zones provide an environment for rapidly exchanging nutrients and organic matter along steep oxygen gradients, creating a reservoir for diverse microbial activity (Gantzer and Stefan 2003; Vreča 2003; Qu et al. 2005). Results from this study revealed that in this stream substrate inhibition was a strong influence on sediment-water interactions, and these patterns connect to certain degradation pathways. α-Ketobutyric Acid is a degradation product of threonine, an amino acid, and the strong canonical loading supports that metabolic inhibition of α-Ketobutyric Acid in sediment, and to a lesser degree inhibition of threonine in sediment, is important to understanding heterogeneity of metabolism within this stream (Bell and Turner 1976). This also supports the influence of nitrogen on metabolic activity in microbial communities. Another dominant inhibitory substrate is pyruvic acid methyl ester, whose conjugate base pyruvate is a precursor to the tricarboxylic acid cycle; α-Ketobuyric Acid is also a key intermediate within this cycle (Burton and Krebs 1953). The conjugate base of Glycl-L-Glutamic acid, glutamate, is another key organic compound associated with cellular metabolism, and sediment metabolic inhibition of Glycl-L-Glutamic acid further supports the dominance of inhibition of sediment respiration in capturing heterogeneity of this aquatic system.

Models identified substrates utilization patterns that were indicative of degree of fecal, nutrient, and organic pollution. The dominant substrate in fecal pollution models was D-xylose; strains of *E. coli* are known to use D-xylose as a complete source of carbon and energy (Blum 2008). The ability to ferment lactose is another important characteristic of *E. coli*, helping

differentiate *E. coli* from *Shigella* and *Salmonella* (de Sousa 2006). The degradation of D-lactose in sediments also suggests potential colonization and resuspension into the water column. Hagehorn et al. (2003) found that fecal bacteria utilized D-lactose more often in human rather than non-human isolates (Hagedorn et al. 2003). Putrescine is a polyamine, a class of amines widely distributed in living organisms, where they are deeply involved in the regulation of cellular functions (Igarashi and Kashiwagi 2000). Although *E. coli* can utilize putrescine as a carbon and nitrogen source using two alternate degradation pathways (Shaibe et al. 1985; Kurihara et al. 2005; Schneider and Reitzer 2012), the negative coefficient suggests that environmental strains of *E. coli* are less efficient at utilizing this substrate. The most influential substrate in nutrient models was metabolic inhibition of D-Malic acid. Malates have been shown to inhibit nitrogen fixation at high concentrations (Bergersen 1997), and microbial production of malates occurs in nitrogen starvation conditions (Chi et al. 2016). The two amino acids identified as influential substrates displayed contrasting patterns; the ability to degrade L-threonine is offset by the inhibition of metabolizing L-Arginine. This suggest that the ability to degrade more complex dissolved organic nitrogen sources diminishes as increased amounts of inorganic nitrogen are available in the environments. The interlinking between these types of pollution and the degradation of organic matter is further supported by the inclusion of D-malic acid and Putrescine from the fecal and nutrient models within the organic matter pollution model.

The use of CLPPs and as metabolic fingerprints represents an attractive low-cost alternative to evaluate microbial functional diversity, and these CLPPs provide key information concerning changes in metabolism along a stream continuum. In this investigation, spatiotemporal gradients were found in both aquatic and benthic communities; aquatic communities differed strongest in their ability to degrade amino acids and carboxylic acids along the stream continuum, while sediment communities had the strongest differences in their ability to degrade amines and phenolic compounds. Seasonal differences were pronounced in the sediments for all substrate groupings. Metabolic inhibition primarily captured the between group variation, suggesting that interruptions in degradation pathways provide greater insights into community dynamics. The importance of differences in nutrient

cycling, especially for nitrogen, was identified using these metabolic fingerprints, and certain substrates were associated with degree of different types of pollution. Water stress was also identified as a potential driver of functional changes in certain sites. Overall, CLPPs were a useful tool to identify factors most influential to community function; this method has strong potential to be used as an effective ecological indicator to identify changes to river continuums attributable to urbanization and different sources of pollution.

References

Aarts BG, Van Den Brink FW, Nienhuis PH. 2003. Habitat loss as the main cause of the slow recovery of fish faunas of regulated large rivers in Europe: the transversal floodplain gradient. River Res Appl. 20(1):3–23. doi:10.1002/rra.720.

Acosta-Martinez V, Cruz L, Sotomayor-Ramirez D, Periz-Alegria L. 2007. Enzyme activities as affected by soil properties and land use in a tropical watershed. Appl Soil Ecol. 35(1):35–45. doi:10.1016/j.apsoil.2006.05.012.

Alban R, Tiedje JM. 2006. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. Microb Ecol. 53(2):197–207. doi:10.1007/s00248-005-5010-2.

Alm EW, Burke J, Spain A. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res. 37(16):3978–3982. doi: 10.1016/S0043-1354(03)00301-4.

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. 21st ed. Washington, DC: American Public Health Association.

Beijerinck MW. 1913. De infusies en de ontdekking der bakterien. Jaarb van K Akad van Wet.:1–28.

Bell SC, Turner JM. 1976. Bacterial catabolism of threonine. Threonine degradation initiated by L-threonine-NAD+ oxidoreductase. Biochem J. 156(2):449–458.

Belnap J, Phillips SL, Miller ME. 2004. Response of desert biological soil crusts to alterations in precipitation frequency. Oecologia. 141(2):306–316. doi:10.1007/s00442-003-1438-6.

Bergersen FJ. 1997. Regulation of nitrogen fixation in infected cells of leguminous root nodules in relation to O^2 supply. Plant Soil. 191(2):189–203.

Berggren M, Ström L, Laudon H, Karlsson J, Jonsson A, Giesler R, Bergström A, Jansson M. 2010. Lake secondary production fueled by rapid transfer of low molecular weight organic carbon from terrestrial sources to aquatic consumers. Ecol Lett. 13(7):870–880.

Bertilsson S, Tranvik LJ. 2000. Photochemical transformation of dissolved organic matter in lakes. Limnol Oceanogr. 45(4):753–762.

Blum P. 2008. Archaea: new models for prokaryotic biology. Horizon Scientific Press.

Boivin M-EY, Breure AM, Posthuma L, Rutgers M. 2002. Determination of field effects of contaminants—significance of pollution-induced community tolerance. Hum Ecol Risk Assess. 8(5):1035–1055.

Burton K, Krebs HA. 1953. The free-energy changes associated with the individual steps of the tricarboxylic acid cycle, glycolysis and alcoholic fermentation and with the hydrolysis of the pyrophosphate groups of adenosinetriphosphate. Biochem J. 54(1):94–107.

Buttigieg PL, Ramette A. 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol. 90(3):543–550.

Byappanahalli M, Fowler M, Shively D, Whitman R. 2003. Ubiquity and persistence of Escherichia coli in a Midwestern Coastal Stream. Appl Environ Microbiol. 69(8):4549–4555. doi:10.1128/AEM.69.8.4549-4555.2003.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl. 8(3):559–568.

Chi Z, Wang Z-P, Wang G-Y, Khan I, Chi Z-M. 2016. Microbial biosynthesis and secretion of l-malic acid and its applications. Crit Rev Biotechnol. 36(1):99–107. doi:10.3109/07388551.2014.924474.

Chick JH, Pegg MA, Koel TM. 2006. Spatial patterns of fish communities in the Upper Mississippi River System: assessing fragmentation by low-head dams. River Res Appl. 22(4):413–427.

Choi KH, Dobbs FC. 1999. Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. J Microbiol Methods. 36(3):203–213.

Christian BW, Lind OT. 2006. Key Issues Concerning Biolog Use for Aerobic and Anaerobic Freshwater Bacterial Community-Level Physiological Profiling. Int Rev Hydrobiol. 91(3):257–268. doi:10.1002/iroh.200510838.

Christian BW, Lind OT. 2007. Multiple carbon substrate utilization by bacteria at the sediment–water interface: seasonal patterns in a stratified eutrophic reservoir. Hydrobiologia. 586(1):43–56.

Cotner JB, Biddanda BA. 2002. Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems. 5(2):105–121. doi:10.1007/s10021-001-0059-3.

Creed IF, McKnight DM, Pellerin BA, Green MB, Bergamaschi BA, Aiken GR, Burns DA, Findlay SEG, Shanley JB, Striegl RG, et al. 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. Can J Fish Aquat Sci. 72(8):1272–1285. doi:10.1139/cjfas-2014-0400.

Dillon WR, Goldstein M. 1984. Multivariate analysis: methods and applications. John Wiley & Sons.

Ding H, Sun M-Y. 2005. Biochemical degradation of algal fatty acids in oxic and anoxic sediment–seawater interface systems: effects of structural association and relative roles of aerobic and anaerobic bacteria. Mar Chem. 93(1):1–19.

Duarte S, Cássio F, Ferreira V, Canhoto C, Pascoal C. 2016. Seasonal variability may affect microbial decomposers and leaf decomposition more than warming in streams. Microb Ecol. 72(2):263–276.

Field KG, Samadpour M. 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 41(16):3517–3538. doi:10.1016/j.watres.2007.06.056.

Fierer N. 2008. Microbial biogeography: patterns in microbial diversity across space and time. In: Accessing uncultivated microorganisms. American Society of Microbiology. p. 95–115.

Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. In: Proceedings of the National Academy of Sciences. Vol. 103. p. 626–631.

Fierer N, Morse JL, Berthrong ST, Bernhardt ES, Jackson RB, de Sousa CP. 2007. Environmental Controls on the Landscape-Scale Biogeography of Stream Bacterial Communities. Ecology. 88(9):2162–2173.

Fierer N, Schimel JP, Holden PA. 2003. Influence of drying-rewetting frequency on soil bacterial community structure. Microb Ecol. 45(1):63–71. doi:10.1007/s00248-002-1007-2.

Ford TE, Naiman RJ. 1989. Groundwater–Surface Water Relationships in Boreal Forest Watersheds: Dissolved Organic Carbon and Inorganic Nutrient Dynamics. Can J Fish Aquat Sci. 46(1):41–49. doi:10.1139/f89-006.

Gantzer CJ, Stefan HG. 2003. A model of microbial activity in lake sediments in response to periodic water-column mixing. Water Res. 37(12):2833–2846.

Garland JL. 1997. Analysis and interpretation of community-level physiological profiles in microbial ecology. FEMS Microbiol Ecol. 24(4):289–300.

Garland JL, Mills AL. 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. Appl Environ Microbiol. 57(8):2351–2359.

Ghiorse WC, Wilson JT. 1988. Microbial Ecology of the Terrestrial Subsurface. In: Advances in Applied Microbiology. Vol. 33. Academic Press. p. 107–172.

Green JL, Bohannan BJM, Whitaker RJ. 2008. Microbial Biogeography: From Taxonomy to Traits. Science. 320(5879):1039 – 1043. doi:10.1126/science.1153475.

Grover JP, Chrzanowski TH. 2000. Seasonal patterns of substrate utilization by bacterioplankton: case studies in four temperate lakes of different latitudes. Aquat Microb Ecol. 23(1):41–54.

Grubaugh J, Wallace B, Houston E. 1997. Production of benthic macroinvertebrate communities along a southern Appalachian river continuum. Freshw Biol. 37(3):581–596.

Gryta A, Frąc M, Oszust K. 2014. The Application of the Biolog EcoPlate Approach in Ecotoxicological Evaluation of Dairy Sewage Sludge. Appl Biochem Biotechnol. 174(4):1434–1443. doi:10.1007/s12010-014-1131-8.

Hagedorn C, Crozier JB, Mentz K a, Booth a M, Graves a K, Nelson NJ, Reneau Jr. RB. 2003. Carbon source utilizations profile as a method to identify sources of faecal pollution in water. J Appl Microbiol. 94:792–799. doi:10.1046/j.1365-2672.2003.01804.x.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. 1998. Multivariate data analysis. Upper Saddle River, NJ: Prentice Hall.

Hall KK, Evanshen BG, Maier KJ, Scheuerman PR. 2014. Application of multivariate statistical methodology to model factors influencing fate and transport of fecal pollution in surface waters. J Environ Qual. 43(1):358–370. doi:10.2134/jeq2013.05.0190.

Hanson PC, Bade DL, Carpenter SR, Kratz TK. 2003. Lake metabolism: relationships with dissolved organic carbon and phosphorus. Limnol Oceanogr. 48(3):1112–1119.

Harbott EL, Grace MR. 2005. Extracellular enzyme response to bioavailability of dissolved organic C in streams of varying catchment urbanization. J North Am Benthol Soc. 24(3):588–601.

Harding JS, Young RG, Hayes JW, Shearer KA, Stark JD. 1999. Changes in agricultural intensity and river health along a river continuum. Freshw Biol. 42(2):345–347.

Heinen EA, McManus J. 2004. Carbon and Nutrient Cycling at the Sediment-water Boundary in Western Lake Superior. J Great Lakes Res. 30:113–132. doi: 10.1016/S0380-1330(04)70381-0.

Ibekwe AM, Leddy M, Murinda SE. 2013. Potential Human Pathogenic Bacteria in a Mixed Urban Watershed as Revealed by Pyrosequencing. PLoS One. 8(11):e79490. doi:10.1371/journal.pone.0079490.

Igarashi K, Kashiwagi K. 2000. Polyamines: mysterious modulators of cellular functions. Biochem Biophys Res Commun. 271(3):559–564.

Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon RJ. 2005. Resuspension of sediment-associated Escherichia coli in a natural stream. J Environ Qual. 34:581–589. doi:10.2134/jeq2005.0581.

Johnson R, Wichern D. 1992. Applied multivariate statistical methods. Prentice Hall, Englewood Cliffs, NJ.

Khalil B, Ouarda T, St-Hilaire A. 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J Hydrol. 405(3-4):277–287.

Krause S, Le Roux X, Niklaus PA, Van Bodegom PM, Lennon JT, Bertilsson S, Grossart H-P, Philippot L, Bodelier PLE. 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Front Microbiol. 5:251.

Kurihara S, Oda S, Kato K, Kim HG, Koyanagi T, Kumagai H, Suzuki H. 2005. A novel putrescine utilization pathway involves γ-glutamylated intermediates of Escherichia coli K-12. J Biol Chem. 280(6):4602–4608.

LaLiberte P, Grimes DJ. 1982. Survival of Escherichia coli in lake bottom sediment. Appl Environ Microbiol. 43(3):623–628.

Leflaive J, Danger M, Lacroix G, Lyautey E, Oumarou C, Ten-Hage L. 2008. Nutrient effects on the genetic and functional diversity of aquatic bacterial communities. FEMS Microbiol Ecol. 66(2):379–390.

Liikanen A, Martikainen PJ. 2003. Effect of ammonium and oxygen on methane and nitrous oxide fluxes across sediment–water interface in a eutrophic lake. Chemosphere. 52(8):1287–1293. doi: 10.1016/S0045-6535(03)00224-8.

Maier RM, Pepper IL, Gerba CP. 2009. Environmental Microbiology. 2nd ed. New York: Elsevier.

Marquardt DW. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics. 12(3):591–612. doi:10.2307/1267205.

McArthur J, Tuckfield RC. 2000. Spatial Patterns in antibiotic resistance among stream bacteria : effects of industrial pollution. Appl Environ Microbiol. 66(9):3722–3726.

McLellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol. 22(12):697–706. doi:10.1016/j.tim.2014.08.002.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, et al. 2011. Global patterns in the biogeography of bacterial taxa. Environ Microbiol. 13(1):135–144. doi:10.1111/j.1462-2920.2010.02315.x.

Noori R, Karbassi A, Khakpour A, Shahbazbegian M, Badam HMK, Vesali-Naseh M. 2012. Chemometric Analysis of Surface Water Quality Data: Case Study of the Gorganrud River Basin, Iran. Environ Model {&} Assess. 17(4):411–420. doi:10.1007/s10666-011-9302-2.

Oest A, Alsaffar A, Fenner M, Azzopardi D, Tiquia-Arashiro SM. 2018. Patterns of change in metabolic capabilities of sediment microbial communities in river and lake ecosystems. Int J Microbiol. 2018.

Pettine M, Patrolecco L, Manganelli M, Capri S, Farrace MG. 1999. Seasonal variations of dissolved organic matter in the northern Adriatic Sea. Mar Chem. 64(3):153–169. doi: 10.1016/S0304-4203(98)00071-1.

Preston-Mafham J, Boddy L, Randerson PF. 2002. Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - A critique. FEMS Microbiol Ecol. 42(1):1–14. doi:10.1016/S0168-6496(02)00324-0.

Qu W, Morrison RJ, West RJ, Su C. 2005. Diagenetic stoichiometry and benthic nutrient fluxes at the sediment–water interface of Lake Illawarra, Australia. Hydrobiologia. 537(1-3):249–264.

Rosi-Marshall EJ, Vallis KL, Baxter C V, Davis JM. 2016. Retesting a prediction of the river continuum concept: autochthonous versus allochthonous resources in the diets of invertebrates. Freshw Sci. 35(2):534–543. doi:10.1086/686302.

Rutgers M, Wouterse M, Drost SM, Breure AM, Mulder C, Stone D, Creamer RE, Winding A, Bloem J. 2016. Monitoring soil bacteria with community-level physiological profiles using Biolog ECO-plates in the Netherlands and Europe. Appl Soil Ecol. 97:23–35.

Sala MM, Estrada M. 2006. Seasonal changes in the functional diversity of bacterioplankton in contrasting coastal environments of the NW Mediterranean . Aquat Microb Ecol. 44(1):1–9.

Sala MM, Pinhassi J, Gasol JM. 2006. Estimation of bacterial use of dissolved organic nitrogen compounds in aquatic ecosystems using Biolog plates. Aquat Microb Ecol. 42(1):1–5.

Sala MM, Terrado R, Lovejoy C, Unrein F, Pedros-Alio C. 2008. Metabolic diversity of heterotrophic bacterioplankton over winter and spring in the coastal Arctic Ocean. Environ Microbiol. 10(4):942–949. doi:10.1111/j.1462-2920.2007.01513.x.

Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, Blaschke AP, Blöschl G, Mach RL, Kirschner AKT, et al. 2015. Bacterial diversity along a 2600 km river continuum. 17:4994–5007. doi:10.1111/1462-2920.12886.

Schimel J, Balser TC, Wallenstein M. 2007. Microbial stress-response physiology and its implications for ecosystem function. Ecology. 88(6):1386–1394. doi:10.1890/06-0219.

Schneider BL, Reitzer L. 2012. Pathway and Enzyme Redundancy in Putrescine Catabolism in Escherichia coli. J Bacteriol. 194(15):4080–4088.

Shaibe E, Metzer E, Halpern YS. 1985. Metabolic pathway for the utilization of L-arginine, L-ornithine, agmatine, and putrescine as nitrogen sources in Escherichia coli K-12. J Bacteriol. 163(3):933–937.

Shu C, Ouarda T. 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour Res. 43(7).

Song J, Luo Y, Zhao Q, Christie P. 2004. Microcosm studies on anaerobic phosphate flux and mineralization of lake sediment organic carbon. J Environ Qual. 33(6):2353–2356.

de Sousa CP. 2006. Escherichia coli as a specialized bacterial pathogen. Rev biol cienc Terra. 2(2):341–352.

Stefanowicz A. 2006. The biolog plates technique as a tool in ecological studies of microbial communities. Polish J Environ Stud. 15(5):669–676.

Storey RG, Fulthorpe RR, Williams DD. 1999. Perspectives and predictions on the microbial ecology of the hyporheic zone. Freshw Biol. 41(1):119–130.

Timoner X, Borrego CM, Acuña V, Sabater S. 2014. The dynamics of biofilm bacterial communities is driven by flow wax and wane in a temporary stream. Limnol Oceanogr. 59(6):2057–2067. doi:10.4319/lo.2014.59.6.2057.

Tiquia SM. 2010. Metabolic diversity of the heterotrophic microorganisms and potential link to pollution of the Rouge River. Environ Pollut. 158(5):1435–1443. doi:http://dx.doi.org/10.1016/j.envpol.2009.12.035.

Tomanova S, Tedesco PA, Campero M, Van Damme PA, Moya N, Oberdorff T. 2007. Longitudinal and altitudinal changes of macroinvertebrate functional feeding groups in neotropical streams: a test of the River Continuum Concept. Fundam Appl Limnol. 170(3):233–241.

Tornwall B, Sokol E, Skelton J, Brown BL. 2015. Trends in stream biodiversity research since the river continuum concept. Diversity. 7(1):16–35. doi:10.3390/d7010016.

United States Environmental Protection Agency. 2017. National summary of impaired waters and TMDL information. [accessed 2017 Aug 1]. http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T.

Urakawa H, Ali J, Ketover RDJ, Talmage SD, Garcia JC, Campbell IS, Loh AN, Parsons ML. 2013. Shifts of bacterioplankton metabolic profiles along the salinity gradient in a subtropical estuary. ISRN Oceanogr. 2013.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. 1980. The river continuum concept. Can J Fish Aquat Sci. 37(1):130–137. doi:10.1139/f80-017.

Vreča P. 2003. Carbon cycling at the sediment–water interface in a eutrophic mountain lake (Jezero na Planini pri Jezeru, Slovenia). Org Geochem. 34(5):671–680.

Walsh CJ, Roy AH, Feminella JW, Cottingham PD, Groffman PM, Morgan II, RP. 2005. The urban stream syndrome : current knowledge and the search for a cure The urban stream syndrome : current knowledge and. J North Am Benthol Soc. 24(3):706–723.

Whitman R, Nevers MB, Byappanahalli M. 2006. Examination of the watershed-wide distribution of Escherichia coli along southern Lake Michigan: An integrated approach. Appl Environ Microbiol. 72(11):7301–7310. doi:10.1128/AEM.00454-06.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes : The unseen majority. In: Proceedings of the National Academy of Sciences. Vol. 95. p. 6578–6583.

Williams DD. 2006. The biology of temporary waters. Oxford University Press.

Wilson JT, McNabb JF, Balkwill DL, Ghiorse WC. 1983. Enumeration and Characterization of Bacteria Indigenous to a Shallow Water-Table Aquifer. Ground Water. 21(2):134–142. doi:10.1111/j.1745-6584.1983.tb00710.x.

Zak JC, Willig MR, Moorhead DL, Wildman HG. 1994. Functional diversity of microbial communities: a quantitative approach. Soil Biol Biochem. 26(9):1101–1108.

Zhang C-B, Wang J, Liu W-L, Zhu S-X, Ge H-L, Chang SX, Chang J, Ge Y. 2010. Effects of plant diversity on microbial biomass and community metabolic profiles in a full-scale constructed wetland. Ecol Eng. 36(1):62–68.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE DIRECTIONS

Fecal pollution is best viewed as an ecological phenomenon, requiring creative strategies to monitor its spatiotemporal variability, appropriate sources of contamination, and identify effective remediation strategies. This dissertation incorporates machine learning models, long-term monitoring data, and microbial metabolism to better understand how environmental factors shape the niche of fecal indicators in a secondary habitat, a lower order oligotrophic stream in Northeast Tennessee. Chapters 2 and 3 reshape the question into a decision making approach, addressing factors separating compliance from impairment based on water quality parameters. This is important to inform load allocation models, evaluate multiple sources of impairment, and assess management alternatives. While the single indicator paradigm is a useful sentinel for point sources, utilizing water quality data and other microbial indicators can enhance monitoring programs in areas impacted by nonpoint sources of pollution, accelerating restoration of impaired watersheds. Chapter 4 demonstrates the potential of Biolog EcoPlates as an alternative microbial indicator in aquatic systems. This chapter evaluated spatiotemporal variability of microbial metabolism along a stream continuum, identifying interactions between aquatic and benthic communities and connecting these to the degree of pollution. Together, these present a microbial ecology approach for understanding human disruption to aquatic ecosystems, and identifying environmental conditions and community dynamics associated with fecal and other types of pollution.

Stream impairment seasonality was exhibited in two different indicators over multiple years and different resolutions (quarterly versus monthly sampling programs). This stresses the need for multiple year and month sampling to capture heterogeneity in fecal indicator concentrations. A five-sample 30-day geometric mean concentration once every five years does not capture this variability, and sampling in the summer months distorts exposure assessments for management of impaired watersheds. Metabolic patterns of seasonal variability were also identified in microbial communities, suggesting that functional diversity shifts over time, potentially fostering conditions for the formation of a fecal indicator niche. Including

seasonality in monitoring programs improves characterization of the temporal variability of fecal indicators, refining estimations of yearly risk and identifying appropriate management strategies.

Knowledge of watershed hydrology is essential to characterize fate and transport processes associated with introduction and persistence of fecal indicators, but flow independent processes confound modeling and monitoring programs. This research identified runoff, hyporheic exchange, and desiccation as dominant processes in shaping microbial niches. Although discharge is strongly associated with dissolved solutes such as NO_3^-, PO_4^{3-}, and ions measured through conductivity, bacterial community structure is also governed by other ecological factors as well, such as population of heterotrophic bacteria, enzyme activity, and other metabolic indicators. Sediment-water interactions were strongly associated with bacteriophage detections and *E. coli* impairment, with glucosidase and acid phosphatase enzyme activity contributing to the overall information gain of Maxent models. Coliforms in sediments were responsible for over half of the information in bacteriophage models, further supporting sediment-water interactions as important mechanisms of fecal pollution. Desiccation influenced community function, increasing the metabolic potential of this community to degrade complex substrates, i.e., phenolic compounds. Each of these aspects of hydrology warrant further attention to better characterize microbial water quality for appropriate exposure assessment and more accurate loading estimates.

While the use of Maxent has been well demonstrated at the macroscale, this study introduces its application to ecological niches at the microscale. Our research also contributes to the Maxent literature through presenting procedures for use of coefficients, i.e., action values in Chapter 2, probabilistic evaluation of model accuracy, and alternative approaches to variable selection using information gain and nonparametric bootstrapping. Action values were generated to predict a threshold of impairment given a single water quality parameter to demonstrate a coefficient extraction technique, opening the black box of Maxent. Probabilistic methods have been used in a variety of water quality monitoring and modeling projects, but this paper adds to this framework through using a probabilistic sensitivity analysis as a variable selection technique and for the generation of model validation metrics. Although modeling

microorganisms in the environment is a headache on the best of days, to paraphrase one of our reviewers for Chapter 2, this modeling approach extracts information concerning mechanisms associated with the formation of an ecologic niche for fecal indicators, guiding decision making and optimizing water quality monitoring strategies.

Chapter 3 stressed the difficulty of the single indicator paradigm through revealing different ecologies of two fecal indicator organisms. *E. coli* impairment seems to be dominated by runoff, identified by hardness as the dominant contributor. Ions that are measured through hardness and alkalinity may be introduced to soil and geologic formation common to East Tennessee, and runoff is further supported by increased counts of heterotrophic and coliform bacteria. Microbial activity and BOD_5 seemed to inhibit impairment. Bacteriophage detections were strongly inhibited by coliform bacteria in sediment, suggesting competitive exclusion as a strong deterrent for detection. Enzyme activity seemed to follow a similar trend as *E. coli* impairment, but BOD_5 was found to increase detections. It is myopic to consider one indicator to mimic all pathogens, and policy needs to shift to a multiple indicator approach. This can overcome some of the difficulties of source tracking and the differential ecologies of pathogens. Modeling can also help with optimization of key water quality monitoring parameters and best indicator(s) for a given watershed based on inputs and ecological activities.

This dissertation also incorporated microbial ecology, in the form of metabolic potential of microbial communities, to characterize spatiotemporal variability in community function to connect this to pollution gradients. Communities differed in their ability to use a wide variety of substrates, including amino and carboxylic acids in water, amines in sediments, and phenolic compounds in sediments. Sediments were identified as a substantial contributor to group variation in aquatic and benthic communities, especially in metabolic inhibition of single sources of carbon. This suggests that sediment microbial activity could be a substantial contributor to instream water quality and needs to be addressed in future research as a potential source of inorganic nutrients and fecal indicators. Sediment substrate activity was also indicative of degrees of pollution, with increased utilization of α-D lactose being associated with higher *E. coli* concentrations. This suggests that sediment is a potential reservoir for *E. coli* in

this stream, and there is possibility for resuspension, extended residence times, and increased duration for exposure. The ability to degrade lower molecular weight amino acids in sediment was positively correlated with degree of nutrient pollution, while higher molecular weight amino acid degradation had an inhibitory effect, suggesting a two-fold association; ability to degrade amino acids contributing to loading of inorganic nutrients, while ability to degrade more complex amino acids causes competitive exclusion because of the uptake of more nitrogen.

Although monitoring for fecal indicators has protected human health, the question remains whether we have lost sight of the original purpose of fecal indicators, i.e., the indirect monitoring of pathogens to reduce human health risk, and instead focused on the indicator rather than the "disease." A single indicator approach cannot be effective to mimic all pathogens, and our paradigm needs to shift to focus on reliably identifying human versus nonhuman sources, connecting this information to predict distribution of pathogens, and finally characterizing the ecology of these pathogens to design programs to prevent exposure. This research highlights, however, the benefit of using models and other microbial indicators, i.e., metabolic activity of communities, to better understand how environment shapes the niche of fecal indicators but could be easily transferred to understanding ecology of multiple pathogens. Other directions of research include:

- Incorporation and standardization of Next Generation Sequencing high throughput techniques to develop consortium indicators for source identification, identify pathogens within a watershed, and connect to currently used monitoring and modeling techniques
- Incorporation of Maxent or other multivariate models to identify key processes associated with fecal pollution, and extract information for inclusion in TMDL process models to include such characteristics as nutrient conditions, heterotrophic competition, and predation.
- Opening the discussion concerning creation of mixed TMDL models for watersheds dealing with multiple synergistic impairments, i.e. nutrient and fecal pollution from nonpoint agricultural sources

- Connecting microbial metabolism in a stream continuum to the community structure using geospatial analysis and next generation sequencing techniques
- Developing programs that educate and engage citizen to stimulate involvement and procurement of funds to finance nonpoint sources of pollution management programs

The problem of fecal pollution is complex, requiring an interdisciplinary approach to reduce this issue. Even with the plethora of alternate indicators suggested for use, strong correlations with pathogens are inconsistent, quantifying human health risk is highly uncertain, and universal source-tracking methods are still needed. It is recommended that policies shift in two ways: (i) advocating for the use of multiple indicators to better characterize pathogen distribution, and (ii) encouraging cooperation between modelers, molecular biologists, spatial scientists, chemists, and epidemiologists to develop a geographically and ecologically flexible framework for source identification, exposure assessment, and risk characterization. The use of common water quality parameters, estimations of microbial activity, and flexible modeling approaches improves understanding of the ecology behind fecal indicators. Pathogens in both surface and groundwater present an economic and public health burden, but the use of models alongside multiple indicators can improve decision making, reducing the negative impacts associated with fecal pollution.

REFERENCES

Aarts BG, Van Den Brink FW, Nienhuis PH. 2003. Habitat loss as the main cause of the slow recovery of fish faunas of regulated large rivers in Europe: the transversal floodplain gradient. River Res Appl. 20(1):3–23. doi:10.1002/rra.720.

Abia ALK, Ubomba-Jaswa E, Genthe B, Momba MNB. 2016. Quantitative microbial risk assessment (QMRA) shows increased public health risk associated with exposure to river water under conditions of riverbed sediment resuspension. Sci Total Environ. 566:1143–1151.

Acosta-Martinez V, Cruz L, Sotomayor-Ramirez D, Periz-Alegria L. 2007. Enzyme activities as affected by soil properties and land use in a tropical watershed. Appl Soil Ecol. 35(1):35–45. doi:10.1016/j.apsoil.2006.05.012.

Adler RW, Landman JC, Cameron DM. 1993. The clean water act 20 years later. Island Press.

Alban R, Tiedje JM. 2006. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. Microb Ecol. 53(2):197–207. doi:10.1007/s00248-005-5010-2.

Alm EW, Burke J, Spain A. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res. 37(16):3978–3982. doi: 10.1016/S0043-1354(03)00301-4.

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. 21st ed. Washington, DC: American Public Health Association.

Anderson KL, Whitlock JE, Harwood VJ. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl Environ Microbiol. 71(6):3041–3048. doi:10.1128/AEM.71.6.3041.

Arnone RD, Walling JP. 2007. Waterborne pathogens in urban watersheds. J Water Health. 5(1):149–162. doi:10.2166/wh.2006.001.

Ashbolt NJ, Schoen ME, Soller JA, Roser DJ. 2010. Predicting pathogen risks to aid beach management: The real value of quantitative microbial risk assessment (QMRA). Water Res. 44(16):4692–4703. doi:https://doi.org/10.1016/j.watres.2010.06.048.

Atherholt TB, Lechevallier MW, Norton WD, Rosen JS. 1998. Effect of rainfall on Giardia and Cryptosporidium. Am Water Work Assoc. 90(9):66–80. doi:10.1002/j.1551-8833.1998.tb08499.x.

Ballestè E, Bonjoch X, Belanche LA, Blanch AR. 2010. Molecular indicators used in the development of predictive models for microbial source tracking. Appl Env Microbiol. 76(6):1789–1795. doi:10.1128/AEM.02350-09.

Basheer IA, Hajmeer M. 2000. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods. 43(1):3–31.

Bean WT, Stafford R, Brashares JS. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. (May 2011):250–258. doi:10.1111/j.1600-0587.2011.06545.x.

Beijerinck MW. 1913. De infusies en de ontdekking der bakterien. Jaarb van K Akad van Wet.:1–28.

Belanche-Muñoz L, Blanch AR. 2008. Machine learning methods for microbial source tracking. Environ Model Softw. 23(6):741–750. doi:10.1016/j.envsoft.2007.09.013.

Bell SC, Turner JM. 1976. Bacterial catabolism of threonine. Threonine degradation initiated by L-threonine-NAD+ oxidoreductase. Biochem J. 156(2):449–458.

Belnap J, Phillips SL, Miller ME. 2004. Response of desert biological soil crusts to alterations in precipitation frequency. Oecologia. 141(2):306–316. doi:10.1007/s00442-003-1438-6.

Benham BL, Baffaut C, Zeckoski RW, Mankin KR, Pachepsky YA, Sadeghi AM, Brannan KM, Soupir ML, Habersack MJ. 2006. Modeling bacteria fate and transport in watersheds to support TMDLs. Trans ASABE. 49(4):987–1002.

Bergersen FJ. 1997. Regulation of nitrogen fixation in infected cells of leguminous root nodules in relation to O^2 supply. Plant Soil. 191(2):189–203.

Berggren M, Ström L, Laudon H, Karlsson J, Jonsson A, Giesler R, Bergström A, Jansson M. 2010. Lake secondary production fueled by rapid transfer of low molecular weight organic carbon from terrestrial sources to aquatic consumers. Ecol Lett. 13(7):870–880.

Bernhard AE, Field KG. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Appl Environ Microbiol. 66(10):4571–4574.

Berthe T, Ratajczak M, Clermont O, Denamur E, Petit F. 2013. Evidence for coexistence of distinct Escherichia coli Populations in various aquatic environments and their survival in estuary water. Appl Environ Microbiol. 79(15):4684–4693. doi:10.1128/AEM.00698-13.

Bertilsson S, Tranvik LJ. 2000. Photochemical transformation of dissolved organic matter in lakes. Limnol Oceanogr. 45(4):753–762.

Betancourt WQ, Fujioka RS. 2006. Bacteroides spp. as reliable marker of sewage contamination in Hawaii's environmental waters using molecular techniques. Water Sci Technol. 54(3):101–107.

Bisson JW, Cabelli VJ. 1980. Clostridium perfringens as a water pollution indicator. J -Water Pollut Control Fed. 52(2):241–248.

Black LE, Brion GM, Freitas SJ. 2007. Multivariate logistic regression for predicting total culturable virus presence at the intake of a potable-water treatment plant : novel application of the atypical coliform / total coliform ratio. Appl Envir Microbiol. 73(12):3965–3974. doi:10.1128/AEM.02780-06.

Blanch AR, Belanche-Muñoz L, Bonjoch X, Ebdon J, Gantzer C, Lucena F, Ottoson J, Kourtis C, Iversen A, Kühn I. 2006. Integrated analysis of established and novel microbial and chemical methods for microbial source tracking. Appl Environ Microbiol. 72(9):5915–5926.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J. 7(11):2061–2068. doi:10.1038/ismej.2013.102.

Blount ZD. 2015. The unexhausted potential of E. coli. Elife. 4:e05826. doi:10.7554/eLife.05826.

Blum P. 2008. Archaea: new models for prokaryotic biology. Horizon Scientific Press.

Boeckmann M, Joyner TA. 2014. Old health risks in new places? An ecological niche model for I. ricinus tick distribution in Europe under a changing climate. Health Place. 30:70–77. doi:10.1016/j.healthplace.2014.08.004.

Boehm AB, Ashbolt NJ, Colford JMJ, Dunbar LE, Fleming LE, Gold MA, Hansel JA, Hunter PR, Ichida AM, McGee CD, et al. 2009. A sea change ahead for recreational water quality criteria. J Water Health. 7(1):9–20. doi:10.2166/wh.2009.122.

van den Bogaard AE, Willems R, London N, Top J, Stobberingh EE. 2002. Antibiotic resistance of faecal enterococci in poultry, poultry farmers and poultry slaughterers. J Antimicrob Chemother. 49(3):497–505.

Boivin M-EY, Breure AM, Posthuma L, Rutgers M. 2002. Determination of field effects of contaminants—significance of pollution-induced community tolerance. Hum Ecol Risk Assess. 8(5):1035–1055.

Bonjoch X, Balleste E, Blanch AR. 2004. Multiplex PCR with 16S rRNA gene-targeted primers of bifidobacterium spp. to identify sources of fecal pollution. Appl Environ Microbiol. 70(5):3171–3175.

Booth AM, Hagedorn C, Graves AK, Hagedorn SC, Mentz KH. 2003. Sources of fecal pollution in Virginia's Blackwater River. J Environ Eng. 129(6):547–552.

Borah DK, Bera M. 2003. Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. Trans ASAE. 46(6):1553.

Borrego JJ, Córnax R, Morinigo MA, Martínez-Manzanares E, Romero P. 1990. Coliphages as an indicator of faecal pollution in water. Their survival and productive infectivity in natural aquatic environments. Water Res. 24(1):111–116.

Borrego JJ, Moriñigo MA, de Vicente A, Córnax R, Romero P. 1987. Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms. Water Res. 21(12):1473–1480.

Borsuk ME, Stow CA, Reckhow KH. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. Environ Sci Technol. 36(10):2109–2115.

Boxall ABA, Fogg LA, Blackwell PA, Blackwell P, Kay P, Pemberton EJ, Croxford A. 2004. Veterinary medicines in the environment. In: Reviews of environmental contamination and toxicology. New York, NY: Springer. p. 1–91.

Brabec E. 2002. Impervious Surfaces and Water Quality: A Review of Current Literature and Its Implications for Watershed Planning. J Plan Lit. 16(4):499–514.

de Brauwere A, Ouattara NK, Servais P. 2014. Modeling fecal indicator bacteria concentrations in natural surface waters: a review. Crit Rev Environ Sci Technol. 44(21):2380–2453.

Brion GM, Lingireddy S. 1999. A neural network approach to identifying non-point sources of microbial contamination. Water Res. 33(14):3099–3106.

Brion GM, Neelakantan TR, Lingireddy S. 2002. A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water. Water Res. 36(15):3765–3774.

Brookes JD, Hipsey MR, Burch MD, Regel RH, Linden LG, Ferguson CM, Antenucci JP. 2005. Relative value of surrogate indicators for detecting pathogens in lakes and reservoirs. Environ Sci Technol. 39(22):8614–8621.

Buckalew DW, Hartman LJ, Grimsley GA, Martin AE, Register KM. 2006. A long-term study comparing membrane filtration with Colilert defined substrates in detecting fecal coliforms and Escherichia coli in natural waters. J Environ Manage. 80(3):191–197. doi:10.1016/j.jenvman.2005.08.024.

Burton K, Krebs HA. 1953. The free-energy changes associated with the individual steps of the tricarboxylic acid cycle, glycolysis and alcoholic fermentation and with the hydrolysis of the pyrophosphate groups of adenosinetriphosphate. Biochem J. 54(1):94–107.

Buttigieg PL, Ramette A. 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol. 90(3):543–550.

Byappanahalli M, Fowler M, Shively D, Whitman R. 2003. Ubiquity and persistence of Escherichia coli in a Midwestern Coastal Stream. Appl Environ Microbiol. 69(8):4549–4555. doi:10.1128/AEM.69.8.4549-4555.2003.

Byappanahalli M, Shively D, Nevers M, Sadowsky M, Whitman R. 2003. Growth and survival of Escherichia coli and enterococci populations in the macro-alga Cladophora (Chlorophyta). FEMS Microbiol Ecol. 46(2):203–11. doi:10.1016/S0168-6496(03)00214-9.

Cai L, Zhang T. 2013. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. Environ Sci Technol. 47(10):5433–5441.

Calci KR, Burkhardt W, Watkins WD, Rippey SR. 1998. Occurrence of Male-Specific Bacteriophage in Feral and Domestic Animal Wastes, Human Feces, and Human-Associated Wastewaters. Appl Environ Microbiol. 64(12):5027–5029.

Caldwell JM, Raley ME, Levine JF. 2007. Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. Environ Sci Technol. 41(9):3277–3283.

Campolongo F, Saltelli A. 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliab Eng Syst Saf. 57(1):49–69. doi:10.1016/S0951-8320(97)00021-5.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl. 8(3):559–568.

Carrillo M, Estrada E, Hazen TC. 1985. Survival and enumeration of the fecal indicators Bifidobacterium adolescentis and Escherichia coli in a tropical rain forest watershed. Appl Environ Microbiol. 50(2):468–476.

Carson CA, Shear BL, Ellersieck MR, Asfaw A. 2001. Identification of fecal Escherichia coli from humans and animals by ribotyping. Appl Environ Microbiol. 67(4):1503–1507.

Chao Y, Ma L, Yang Y, Ju F, Zhang X-X, Wu W-M, Zhang T. 2013. Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. Sci Rep. 3:3550.

Characklis GW, Dilts MJ, Simmons OD, Likirdopulos C a., Krometis LAH, Sobsey MD. 2005. Microbial partitioning to settleable particles in stormwater. Water Res. 39(9):1773–1782. doi:10.1016/j.watres.2005.03.004.

Chi Z, Wang Z-P, Wang G-Y, Khan I, Chi Z-M. 2016. Microbial biosynthesis and secretion of l-malic acid and its applications. Crit Rev Biotechnol. 36(1):99–107. doi:10.3109/07388551.2014.924474.

Chick JH, Pegg MA, Koel TM. 2006. Spatial patterns of fish communities in the Upper Mississippi River System: assessing fragmentation by low-head dams. River Res Appl. 22(4):413–427.

Choi KH, Dobbs FC. 1999. Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. J Microbiol Methods. 36(3):203–213.

Christian BW, Lind OT. 2006. Key Issues Concerning Biolog Use for Aerobic and Anaerobic Freshwater Bacterial Community-Level Physiological Profiling. Int Rev Hydrobiol. 91(3):257–268. doi:10.1002/iroh.200510838.

Christian BW, Lind OT. 2007. Multiple carbon substrate utilization by bacteria at the sediment–water interface: seasonal patterns in a stratified eutrophic reservoir. Hydrobiologia. 586(1):43–56.

Chung H, Sobsey MD. 1993. Comparative survival of indicator viruses and enteric viruses in seawater and sediment. Water Sci Technol. 27(3-4):425–428.

Cimenti M, Hubberstey a., Bewtra JK, Biswas N. 2007. Alternative methods in tracking sources of microbial contamination in waters. Water SA. 33(2):183–194. doi:10.4314/wsa.v33i2.49059.

Cloutier DD, Alm EW, McLellan SL. 2015. Influence of land use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Appl Environ Microbiol. 81(15):4904–4913. doi:10.1128/AEM.00233-15.

Cohen A, Davidson S. 2011. The watershed approach: Challenges, antecedents, and the transition from technical tool to governance unit. Water Altern. 4(1):1–14.

Cole D, Long SC, Sobsey MD. 2003. Evaluation of F+ RNA and DNA coliphages as source-specific indicators of fecal contamination in surface waters. Appl Environ Microbiol. 69(11):6507–6514.

Colford JM, Wade TJ, Schiff KC, Wright CC, Griffith JF, Sandhu SK, Burns S, Sobsey M, Lovelace G, Weisberg SB, et al. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology. 18(1):27–35. doi:10.1097/01.ede.0000249425.32990.b9.

Cooley MB, Chao D, Mandrell RE. 2006. Escherichia coli O157: H7 survival and growth on lettuce is altered by the presence of epiphytic bacteria. J Food Prot. 69(10):2329–2335.

Copeland C. 1999. Clean Water Act: a summary of the law. Congressional Research Service, Library of Congress Washington, DC.

Corso PS, Kramer MH, Blair KA, Addiss DG, Davis JP, Haddix AC. 2003. Costs of Illness in the 1993 Waterborne Cryptosporidium Outbreak, Milwaukee, Wisconsin. Emerg Infect Dis J. 9(4):426–431. doi:10.3201/eid0904.020417.

Cotner JB, Biddanda BA. 2002. Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems. 5(2):105–121. doi:10.1007/s10021-001-0059-3.

Coulliete A, Money ES, Serre ML, Noble RT. 2009. Space/time analysis of fecal pollution and rainfall in an eastern north carolina estuary. Environ Sci Technol. 43(10):3728–3735.

Crabill C, Donald R, Snelling J, Foust R, Southam G. 1999. The impact of sediment fecal coliform reservoirs on seasonal water quality in Oak Creek, Arizona. Water Res. 33(9):2163–2171. doi:https://doi.org/10.1016/S0043-1354(98)00437-0.

Creed IF, McKnight DM, Pellerin BA, Green MB, Bergamaschi BA, Aiken GR, Burns DA, Findlay SEG, Shanley JB, Striegl RG, et al. 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. Can J Fish Aquat Sci. 72(8):1272–1285. doi:10.1139/cjfas-2014-0400.

Curriero FC, Patz J a., Rose JB, Lele S. 2001. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. Am J Public Heal. 91(8):1194–1199. doi:10.2105/AJPH.91.8.1194.

Daneshvar A, Aboulfadl K, Viglino L, Broséus R, Sauvé S, Madoux-Humery A-S, Weyhenmeyer GA, Prévost M. 2012. Evaluating pharmaceuticals and caffeine as indicators of fecal contamination in drinking water sources of the Greater Montreal region. Chemosphere. 88(1):131–139. doi:https://doi.org/10.1016/j.chemosphere.2012.03.016.

Davies CM, Long JA, Donald M, Ashbolt NJ. 1995. Survival of fecal microorganisms in marine and freshwater sediments. Appl Environ Microbiol . 61 (5):1888–1896.

Desmarais TR, Solo-Gabriele HM, Palmer CJ. 2002. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. Appl Environ Microbiol. 68(3):1165–1172.

Dick LK, Bernhard AE, Brodeur TJ, Santo Domingo JW, Simpson JM, Walters SP, Field KG. 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. Appl Environ Microbiol. 71(6):3184–3191.

Dick LK, Field KG. 2004. Rapid Estimation of Numbers of Fecal Bacteroidetes by Use of a Quantitative PCR Assay for 16S rRNA Genes. Appl Environ Microbiol . 70 (9):5695–5697.

Dick LK, Simonich MT, Field KG. 2005. Microplate subtractive hybridization to enrich for Bacteroidales genetic markers for fecal source identification. Appl Environ Microbiol. 71(6):3179–3183.

Dillon WR, Goldstein M. 1984. Multivariate analysis: methods and applications. John Wiley & Sons.

Ding H, Sun M-Y. 2005. Biochemical degradation of algal fatty acids in oxic and anoxic sediment–seawater interface systems: effects of structural association and relative roles of aerobic and anaerobic bacteria. Mar Chem. 93(1):1–19.

Donald M, Mengersen K, Toze S, Sidhu JPS, Cook A. 2011. Incorporating parameter uncertainty into Quantitative Microbial Risk Assessment (QMRA). J Water Health. 9(1):10–26. doi:10.2166/wh.2010.073.

Drummond JD, Davies-Colley RJ, Stott R, Sukias JP, Nagels JW, Sharp A, Packman AI. 2015. Microbial transport, retention, and inactivation in streams: a combined experimental and stochastic modeling approach. Environ Sci Technol. 49(13):7825–33. doi:10.1021/acs.est.5b01414.

Duarte S, Cássio F, Ferreira V, Canhoto C, Pascoal C. 2016. Seasonal variability may affect microbial decomposers and leaf decomposition more than warming in streams. Microb Ecol. 72(2):263–276.

Duda AM. 1993. Addressing nonpoint sources of water pollution must become an international priority. Water Sci Technol. 28(3-5):1–11.

Duda AM, Lenat DR, Penrose DL. 1982. Water quality in urban streams: what we can expect. J - Water Pollut Control Fed. 54(7):1139–1147.

Düreth S, Herrmann R, Pecher K. 1986. Tracing faecal pollution by coprostanol and intestinal bacteria in an ice-covered Finnish lake loaded with both industrial and domestic sewage. Water Air Soil Pollut. 28(1-2):131–149.

Dutka BJ, Chau ASY, Coburn J. 1974. Relationship between bacterial indicators of water pollution and fecal sterols. Water Res. 8(12):1047–1055.

Dwivedi D, Mohanty BP, Lesikar BJ. 2013. Estimating Escherichia coli loads in streams based on various physical, chemical, and biological factors. Water Resour Res. 49(5):2896–2906. doi:10.1002/wrcr.20265.

Dwivedi D, Mohanty BP, Lesikar BJ. 2016. Impact of the Linked Surface Water-Soil Water-Groundwater System on Transport of E. coli in the Subsurface. Water, Air, Soil Pollut. 227(9):351. doi:10.1007/s11270-016-3053-2.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. Science. 308(5728):1635–1638.

Eleria A, Vogel RM. 2005. Predicting fecal coliform bacteria levels in the Charles River, Massachusetts, USA. JAWRA J Am Water Resour Assoc. 41(5):1195–1209.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Divers Distrib. 17(1):43–57. doi:10.1111/j.1472-4642.2010.00725.x.

Elshorbagy A, Teegavarapu RS V, Ormsbee L. 2005. Total maximum daily load (TMDL) approach to surface water quality management : concepts , issues , and applications. 448:442–448. doi:10.1139/L04-107.

Fahrenfeld NL, Del Monaco N, Coates JT, Elzerman AW. 2016. Fecal sterol and runoff analysis for nonpoint source tracking. J Environ Qual. 45(1):315–322.

Ferguson C, Husman AM de R, Altavilla N, Deere D, Ashbolt N. 2003. fate and transport of surface water pathogens in watersheds. Crit Rev Environ Sci Technol. 33(3):299–361. doi:10.1080/10643380390814497.

Field KG, Chern EC, Dick LK, Fuhrman J, Griffith J, Holden PA, LaMontagne MG, Olson B, Simonich MT. 2003. A comparative study of culture-independent, library-independent genotypic methods of fecal source tracking. J Water Health. 1(4):181–194.

Field KG, Samadpour M. 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Res. 41(16):3517–3538. doi:10.1016/j.watres.2007.06.056.

Fierer N. 2008. Microbial biogeography: patterns in microbial diversity across space and time. In: Accessing uncultivated microorganisms. American Society of Microbiology. p. 95–115.

Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. In: Proceedings of the National Academy of Sciences. Vol. 103. p. 626–631.

Fierer N, Morse JL, Berthrong ST, Bernhardt ES, Jackson RB, de Sousa CP. 2007. Environmental Controls on the Landscape-Scale Biogeography of Stream Bacterial Communities. Ecology. 88(9):2162–2173.

Fierer N, Schimel JP, Holden PA. 2003. Influence of drying-rewetting frequency on soil bacterial community structure. Microb Ecol. 45(1):63–71. doi:10.1007/s00248-002-1007-2.

Fong T-T, Lipp EK. 2005. Enteric viruses of humans and animals in aquatic environments: health risks, detection, and potential water quality assessment tools. Microbiol Mol Biol Rev. 69(2):357–371. doi:10.1128/MMBR.69.2.357-371.2005.

Ford TE, Naiman RJ. 1989. Groundwater–Surface Water Relationships in Boreal Forest Watersheds: Dissolved Organic Carbon and Inorganic Nutrient Dynamics. Can J Fish Aquat Sci. 46(1):41–49. doi:10.1139/f89-006.

Francy DS, Gifford AM, Darner RA. 2003. Escherichia coli at Ohio bathing beaches--Distribution, sources, wastewater indicators, and predictive modeling.

Frey SK, Topp E, Edge T, Fall C, Gannon V, Jokinen C, Marti R, Neumann N, Ruecker N, Wilkes G, et al. 2013. Using SWAT, Bacteroidales microbial source tracking markers, and fecal indicator bacteria to predict waterborne pathogen occurrence in an agricultural watershed. Water Res. 47(16):6326–6337. doi:10.1016/j.watres.2013.08.010.

Friedman JH. 1991. Multivariate adaptive regression splines. Ann Stat.:1–67.

Fries JS, Characklis GW, Noble RT. 2006. Attachment of Fecal Indicator Bacteria to Particles in the Neuse River Estuary, N.C. J Environ Eng. 132(10):1338–1345. doi:10.1061/(ASCE)0733-9372(2006)132:10(1338).

Fry JA, Xian G, Jin S, Dewitz JA, Homer CG, Limin Y, Barnes CA, Herold ND, Wickham JD. 2011. Completion of the 2006 national land cover database for the conterminous United States. Photogramm Eng Remote Sensing. 77(9):858–864.

Fujioka SR, Solo-Gabriele MH, Byappanahalli NM, Kirs M. 2015. U.S. Recreational Water Quality Criteria: A Vision for the Future. Int J Environ Res Public Health. 12(7):7752–7776. doi:10.3390/ijerph120707752.

Gantzer CJ, Stefan HG. 2003. A model of microbial activity in lake sediments in response to periodic water-column mixing. Water Res. 37(12):2833–2846.

Garcia-Armisen T, Servais P. 2009. Partitioning and fate of particle-associated E. coli in river waters. Water Environ Res. 81(1):21–28.

Garland JL. 1997. Analysis and interpretation of community-level physiological profiles in microbial ecology. FEMS Microbiol Ecol. 24(4):289–300.

Garland JL, Mills AL. 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. Appl Environ Microbiol. 57(8):2351–2359.

Ghimire B, Deng Z. 2013. Hydrograph-based approach to modeling bacterial fate and transport in rivers. Water Res. 47(3):1329–1343. doi:10.1016/j.watres.2012.11.051.

Ghiorse WC, Wilson JT. 1988. Microbial Ecology of the Terrestrial Subsurface. In: Advances in Applied Microbiology. Vol. 33. Academic Press. p. 107–172.

Gilpin B, James T, Nourozi F, Saunders D, Scholes P, Savill M. 2003. The use of chemical and molecular microbial indicators for faecal source identification. Water Sci Technol. 47(3):39–43.

Gonzalez RA, Conn KE, Crosswell JR, Noble RT. 2012. Application of empirical predictive modeling using conventional and alternative fecal indicator bacteria in eastern North Carolina waters. Water Res. 46(18):5871–5882. doi:http://dx.doi.org/10.1016/j.watres.2012.07.050.

Gonzalez RA, Noble RT. 2014. Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods. Water Res. 48:296–305. doi:http://dx.doi.org/10.1016/j.watres.2013.09.038.

Graves AK, Hagedorn C, Teetor A, Mahal M, Booth AM, Reneau RB. 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. J Environ Qual. 31(4):1300–1308.

Green JL, Bohannan BJM, Whitaker RJ. 2008. Microbial Biogeography: From Taxonomy to Traits. Science. 320(5879):1039 – 1043. doi:10.1126/science.1153475.

Gronewold AD, Borsuk ME, Wolpert RL, Reckhow KH. 2008. An assessment of fecal indicator bacteria-based water quality standards. Environ Sci Technol. 42(13):4676–4682.

Grover JP, Chrzanowski TH. 2000. Seasonal patterns of substrate utilization by bacterioplankton: case studies in four temperate lakes of different latitudes. Aquat Microb Ecol. 23(1):41–54.

Grubaugh J, Wallace B, Houston E. 1997. Production of benthic macroinvertebrate communities along a southern Appalachian river continuum. Freshw Biol. 37(3):581–596.

Gryta A, Frąc M, Oszust K. 2014. The Application of the Biolog EcoPlate Approach in Ecotoxicological Evaluation of Dairy Sewage Sludge. Appl Biochem Biotechnol. 174(4):1434–1443. doi:10.1007/s12010-014-1131-8.

Guber AK, Pachepsky YA, Shelton DR, Yu O. 2007. Effect of bovine manure on fecal coliform attachment to soil and soil particles of different sizes. Appl Environ Microbiol. 73(10):3363–3370.

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE. 2006. Using niche-based models to improve the sampling of rare species. Conserv Biol. 20(2):501–511.

Guzman Herrador BR, de Blasio BF, MacDonald E, Nichols G, Sudre B, Vold L, Semenza JC, Nygård K. 2015. Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. Environ Heal. 14(1):29. doi:10.1186/s12940-015-0014-y.

Haas CN, Rose JB, Gerba CP. 1999. Quantitative microbial risk assessment. John Wiley & Sons.

HACH Company. 2006. Digital Titrator - Model 16900: Procedure Manual.

HACH Company. 2013. DR/890 Colorimeter Procedures Manual. :616.

Hagedorn C, Crozier JB, Mentz K a, Booth a M, Graves a K, Nelson NJ, Reneau Jr. RB. 2003. Carbon source utilizations profile as a method to identify sources of faecal pollution in water. J Appl Microbiol. 94:792–799. doi:10.1046/j.1365-2672.2003.01804.x.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. 1998. Multivariate data analysis. Upper Saddle River, NJ: Prentice Hall.

Hall KK, Evanshen BG, Maier KJ, Scheuerman PR. 2014. Application of multivariate statistical methodology to model factors influencing fate and transport of fecal pollution in surface waters. J Environ Qual. 43(1):358–370. doi:10.2134/jeq2013.05.0190.

Hanson PC, Bade DL, Carpenter SR, Kratz TK. 2003. Lake metabolism: relationships with dissolved organic carbon and phosphorus. Limnol Oceanogr. 48(3):1112–1119.

Harbott EL, Grace MR. 2005. Extracellular enzyme response to bioavailability of dissolved organic C in streams of varying catchment urbanization. J North Am Benthol Soc. 24(3):588–601.

Harding JS, Young RG, Hayes JW, Shearer KA, Stark JD. 1999. Changes in agricultural intensity and river health along a river continuum. Freshw Biol. 42(2):345–347.

Hartel PG, Summer JD, Hill JL, Collins J, Entry JA, Segars WI. 2002. Geographic variability of Escherichia coli ribotypes from animals in Idaho and Georgia. J Environ Qual. 31(4):1273–1278.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, Rose JB. 2005. Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. Appl Environ Microbiol . 71 (6):3163–3170. doi:10.1128/AEM.71.6.3163-3170.2005.

Harwood VJ, Wiggins B, Hagedorn C, Ellender RD, Gooch J, Kern J, Samadpour M, Chapman ACH, Robinson BJ, Thompson BC. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J Water Health. 1(4):153–166.

Havelaar AH, Pot-Hogeboom WM. 1988. F-specific RNA-bacteriophages as model viruses in water hygiene: ecological aspects. Water Sci Technol. 20(11-12):399–407.

Hayashi Y, Managaki S, Takada H. 2002. Fluorescent Whitening Agents in Tokyo Bay and Adjacent Rivers: Their Application as Anthropogenic Molecular Markers in Coastal Environments. Environ Sci Technol. 36(16):3556–3563. doi:10.1021/es011352o.

Heinen EA, McManus J. 2004. Carbon and Nutrient Cycling at the Sediment-water Boundary in Western Lake Superior. J Great Lakes Res. 30:113–132. doi: 10.1016/S0380-1330(04)70381-0.

Herrig IM, Böer SI, Brennholt N, Manz W. 2015. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. Water Res. 85:148–157. doi:http://dx.doi.org/10.1016/j.watres.2015.08.006.

Hill BH, Elonen CM, Seifert LR, May AA, Tarquinio E. 2012. Microbial enzyme stoichiometry and nutrient limitation in US streams and rivers. Ecol Indic. 18:540–551.

Hill BH, McCormick FH, Harvey BC, Johnson SL, Warren ML, Elonen CM. 2010. Microbial enzyme activity, nutrient uptake and nutrient limitation in forested streams. Freshw Biol. 55(5):1005–1019.

Hofstra N. 2011. Quantifying the impact of climate change on enteric waterborne pathogen concentrations in surface water. Curr Opin Environ Sustain. 3(6):471–479. doi:http://dx.doi.org/10.1016/j.cosust.2011.10.006.

Huang K, Zhang X-X, Shi P, Wu B, Ren H. 2014. A comprehensive insight into bacterial virulence in drinking water using 454 pyrosequencing and Illumina high-throughput sequencing. Ecotoxicol Environ Saf. 109:15–21.

Hundesa A, Maluquer de Motes C, Bofill-Mas S, Albinana-Gimenez N, Girones R. 2006. Identification of Human and Animal Adenoviruses and Polyomaviruses for Determination of Sources of Fecal Contamination in the Environment. Appl Environ Microbiol . 72 (12):7886–7893. doi:10.1128/AEM.01090-06.

Hunter C, McDonald A, Beven K. 1992. Input of fecal coliform bacteria to an upland stream channel in the Yorkshire Dales. Water Resour Res. 28(7):1869–1876. doi:10.1029/92WR00771. [accessed 2016 Feb 9]. http://doi.wiley.com/10.1029/92WR00771.

Hunter PR. 2003. Climate change and waterborne and vector-borne disease. J Appl Microbiol. 94:37–46.

Huysman F, Van Renterghem B, Verstraete W. 1993. Antibiotic resistant sulphite-reducing clostridia in soil and groundwater as indicator of manuring practices. Water Air Soil Pollut. 69(3):243–255. doi:10.1007/BF00478161.

Ibekwe AM, Leddy M, Murinda SE. 2013. Potential Human Pathogenic Bacteria in a Mixed Urban Watershed as Revealed by Pyrosequencing. PLoS One. 8(11):e79490. doi:10.1371/journal.pone.0079490.

Igarashi K, Kashiwagi K. 2000. Polyamines: mysterious modulators of cellular functions. Biochem Biophys Res Commun. 271(3):559–564.

Ishii S, Ksoll WB, Hicks RE, Sadowsky MJ. 2006. Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds. Appl Environ Microbiol. 72(1):612–621. doi:10.1128/AEM.72.1.612-621.2006.

Isobe KO, Tarao M, Chiem NH, Minh LY, Takada H. 2004. Effect of environmental factors on the relationship between concentrations of coprostanol and fecal indicator bacteria in tropical (Mekong Delta) and temperate (Tokyo) freshwaters. Appl Environ Microbiol. 70(2):814–821.

Isobe KO, Tarao M, Zakaria MP, Chiem NH, Minh LY, Takada H. 2002. Quantitative application of fecal sterols using gas chromatography− mass spectrometry to investigate fecal pollution in tropical waters: Western Malaysia and Mekong Delta, Vietnam. Environ Sci Technol. 36(21):4497–4507.

Jamieson RC, Gordon R, Joy D, Lee H. 2004. Assessing microbial pollution of rural surface waters: A review of current watershed scale modeling approaches. Agric Water Manag. 70(1):1–17. doi:10.1016/j.agwat.2004.05.006.

Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon R. 2005. Transport and deposition of sediment-associated Escherichia coli in natural streams. Water Res. 39(12):2665–2675. doi:10.1016/j.watres.2005.04.040.

Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon RJ. 2005. Resuspension of sediment-associated Escherichia coli in a natural stream. J Environ Qual. 34:581–589. doi:10.2134/jeq2005.0581.

Jiang S, Noble R, Chu W. 2001. Human Adenoviruses and Coliphages in Urban Runoff-Impacted Coastal Waters of Southern California. Appl Environ Microbiol. 67(1):179–184. doi:10.1128/AEM.67.1.179-184.2001.

Jiménez-Clavero MA, Escribano-Romero E, Mansilla C, Gómez N, Córdoba L, Roblas N, Ponz F, Ley V, Sáiz J-C. 2005. Survey of Bovine Enterovirus in Biological and Environmental Samples by a Highly Sensitive Real-Time Reverse Transcription-PCR. Appl Environ Microbiol . 71 (7):3536–3543. doi:10.1128/AEM.71.7.3536-3543.2005.

Jiménez-Clavero MA, Fernández C, Ortiz JA, Pro J, Carbonell G, Tarazona JV, Roblas N, Ley V. 2003. Teschoviruses as Indicators of Porcine Fecal Contamination of Surface Water. Appl Environ Microbiol . 69 (10):6311–6315. doi:10.1128/AEM.69.10.6311-6315.2003.

Johnson R, Wichern D. 1992. Applied multivariate statistical methods. Prentice Hall, Englewood Cliffs, NJ.

Kay D, Crowther J, Fewtrell L, Francis C a., Hopkins M, Kay C, McDonald AT, Stapleton CM, Watkins J, Wilkinson J, et al. 2008. Quantification and control of microbial pollution from agriculture: a new policy challenge? Environ Sci Policy. 11(2):171–184. doi:10.1016/j.envsci.2007.10.009.

Khalil B, Ouarda T, St-Hilaire A. 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J Hydrol. 405(3-4):277–287.

Khatib L, Tsai Y, Olson B. 2002. A biomarker for the identification of cattle fecal pollution in water using the LTIIa toxin gene from enterotoxigenic Escherichia coli. Appl Microbiol Biotechnol. 59(1):97–104. doi:10.1007/s00253-002-0959-y.

Khatib LA, Tsai YL, Olson BH. 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic Escherichia coli. Appl Microbiol Biotechnol. 63(2):231–238. doi:10.1007/s00253-003-1373-9.

Kim JH, Grant SB. 2004. Public Mis-Notification of Coastal Water Quality: A Probabilistic Evaluation of Posting Errors at Huntington Beach, California. Environ Sci Technol. 38(9):2497–2504. doi:10.1021/es034382v.

Kim M, Choi CY, Gerba CP. 2008. Source tracking of microbial intrusion in water systems using artificial neural networks. Water Res. 42(4):1308–1314. doi:https://doi.org/10.1016/j.watres.2007.09.032.

Kirschner AKT, Reischer GH, Jakwerth S, Savio D, Ixenmaier S, Toth E, Sommer R, Mach RL, Linke R, Eiler A, et al. 2017. Multiparametric monitoring of microbial faecal pollution reveals the dominance of human contamination along the whole Danube River. Water Res. 124:543–555. doi:https://doi.org/10.1016/j.watres.2017.07.052.

Kistemann T, Claßen T, Koch C, Dangendorf F, Fischeder R, Gebel J, Vacata V, Exner M. 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl Environ Microbiol. 68(5):2188–2197.

Korajkic A, Wanjugi P, Harwood VJ. 2013. Indigenous Microbiota and Habitat Influence Escherichia coli Survival More than Sunlight in Simulated Aquatic Environments. Appl Environ Microbiol. 79(17):5329–5337.

Krause S, Le Roux X, Niklaus PA, Van Bodegom PM, Lennon JT, Bertilsson S, Grossart H-P, Philippot L, Bodelier PLE. 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Front Microbiol. 5:251.

Kreader CA. 1995. Design and evaluation of Bacteroides DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol. 61(4):1171–1179.

Kreader CA. 1998. Persistence of PCR-detectable Bacteroides distasonis from human feces in river water. Appl Environ Microbiol. 64(10):4103–4105.

Kurihara S, Oda S, Kato K, Kim HG, Koyanagi T, Kumagai H, Suzuki H. 2005. A novel putrescine utilization pathway involves γ-glutamylated intermediates of Escherichia coli K-12. J Biol Chem. 280(6):4602–4608.

LaLiberte P, Grimes DJ. 1982. Survival of Escherichia coli in lake bottom sediment. Appl Environ Microbiol. 43(3):623–628.

Lasalde C, Rodriguez R, Toranzos G a, Smith HH. 2005. Heterogeneity of uidA gene in environmental Escherichia coli populations. J Water Health. 3(3):297–304.

Leeming R, Ball A, Ashbolt N, Nichols P. 1996. Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. Water Res. 30(12):2893–2900.

Leflaive J, Danger M, Lacroix G, Lyautey E, Oumarou C, Ten-Hage L. 2008. Nutrient effects on the genetic and functional diversity of aquatic bacterial communities. FEMS Microbiol Ecol. 66(2):379–390.

Lemarchand K, Lebaron P. 2003. Occurrence of Salmonella spp. and Cryptosporidium spp. in a French coastal watershed: relationship with fecal indicators. FEMS Microbiol Lett. 218(1):203–209.

Liikanen A, Martikainen PJ. 2003. Effect of ammonium and oxygen on methane and nitrous oxide fluxes across sediment–water interface in a eutrophic lake. Chemosphere. 52(8):1287–1293. doi: 10.1016/S0045-6535(03)00224-8.

Lipp EK, Kurz R, Vincent R, Rodriguez-Palacios C, Farrah SR, Rose JB. 2001. The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. Estuaries. 24(2):266–276. doi:10.2307/1352950.

Lozier JD, Aniello P, Hickerson MJ. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. J Biogeogr. 36(9):1623–1627. doi:10.1111/j.1365-2699.2009.02152.x.

Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM. 2011. Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. In: Proceedings of the National Academy of Sciences. Vol. 108.

Lušić DV, Kranjčević L, Maćešić S, Lušić D, Jozić S, Linšak Ž, Bilajac L, Grbčić L, Bilajac N. 2017. Temporal variations analyses and predictive modeling of microbiological seawater quality. Water Res. 119:160–170. doi:https://doi.org/10.1016/j.watres.2017.04.046.

MacKenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, Kazmierczak JJ, Addiss DG, Fox KR, Rose JB, et al. 1994. A Massive Outbreak in Milwaukee of Cryptosporidium

Infection Transmitted through the Public Water Supply. N Engl J Med. 331(3):161–167. doi:10.1056/NEJM199407213310304.

Maguire LA. 2003. Interplay of Science and Stakeholder Values in Neuse River Total Maximum Daily Load Process. J Water Resour Plan Manag. 129(4):261–271.

Maier RM, Pepper IL, Gerba CP. 2009. Environmental Microbiology. 2nd ed. New York: Elsevier.

Maluquer de Motes C, Clemente-Casares P, Hundesa A, Martín M, Girones R. 2004. Detection of Bovine and Porcine Adenoviruses for Tracing the Source of Fecal Contamination. Appl Environ Microbiol . 70 (3):1448–1454.

de Man H, van den Berg HHJL, Leenen EJTM, Schijven JF, Schets FM, van der Vliet JC, van Knapen F, de Roda Husman AM. 2014. Quantitative assessment of infection risk from exposure to waterborne pathogens in urban floodwater. Water Res. 48:90–99. doi:https://doi.org/10.1016/j.watres.2013.09.022.

Marquardt DW. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics. 12(3):591–612. doi:10.2307/1267205.

Martellini A, Payment P, Villemur R. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. Water Res. 39(4):541–548.

Martinez G, Pachepsky YA, Whelan G, Yakirevich AM, Guber A, Gish TJ. 2014. Rainfall-induced fecal indicator organisms transport from manured fields: Model sensitivity analysis. Environ Int. 63:121–129. doi:10.1016/j.envint.2013.11.003.

Martins CC, Venkatesan MI, Montone RC. 2002. Sterols and linear alkylbenzenes in marine sediments from Admiralty Bay, King George Island, South Shetland Islands. Antarct Sci. 14(3):244–252.

Mawdsley JL, Bardgett RD, Merry RJ, Pain BF, Theodorou MK. 1995. Pathogens in livestock waste, their potential for movement through soil and environmental pollution. Appl Soil Ecol. 2(1):1–15. doi:10.1016/0929-1393(94)00039-A.

McArthur J, Tuckfield RC. 2000. Spatial Patterns in antibiotic resistance among stream bacteria : effects of industrial pollution. Appl Environ Microbiol. 66(9):3722–3726.

McBride GB, Stott R, Miller W, Bambic D, Wuertz S. 2013. Discharge-based QMRA for estimation of public health risks from exposure to stormwater-borne pathogens in recreational waters in the United States. Water Res. 47(14):5282–5297.

McCambridge J, McMeekin TA. 1980. Relative effects of bacterial and protozoan predators on survival of Escherichia coli in estuarine water samples. Appl Environ Microbiol. 40(5):907–911.

McKergow LA, Davies-Colley RJ. 2010. Stormflow dynamics and loads of Escherichia coli in a large mixed land use catchment. Hydrol Process An Int J. 24(3):276–289.

McLellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol. 22(12):697–706. doi:10.1016/j.tim.2014.08.002.

McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ Microbiol. 12(2):378–392.

Meays CL, Broersma K, Nordin R, Mazumder A. 2004. Source tracking fecal bacteria in water: a critical review of current methods. J Environ Manage. 73(1):71–79.

Moe CL, Sobsey MD, Samsa GP, Mesolo V. 1991. Bacterial indicators of risk of diarrhoeal disease from drinking-water in the Philippines. Bull World Health Organ. 69(3):305.

Molina M, Hunter S, Cyterski M, Peed LA, Kelty CA, Sivaganesan M, Mooney T, Prieto L, Shanks OC. 2014. Factors affecting the presence of human-associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. Water Res. 64:196–208. doi:https://doi.org/10.1016/j.watres.2014.06.036.

Money ES, Carter GP, Serre ML. 2009. Modern space/time geostatistics using river distances: data integration of turbidity and E.coli measurements to assess fecal contamination along the Raritan River in New Jersey. Environ Sci Technol. 43(10):3736–3742.

Moore DF, Harwood VJ, Ferguson DM, Lukasik J, Hannah P, Getrich M, Brownell M. 2005. Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed. J Appl Microbiol. 99(3):618–628.

Morincigo MA, Martinez-Manzanares E, Muncoz A, Cornax R, Romero P, Borrego JJ. 1989. Evaluation of different plating media used in the isolation of salmonellas from

environmental samples. J Appl Bacteriol. 66(4):353–360. doi:10.1111/j.1365-2672.1989.tb02488.x.

Mugnai R, Sattamini A, dos Santos JAA, Regua-Mangia AH. 2015. A survey of Escherichia coli and Salmonella in the Hyporheic Zone of a subtropical stream: their bacteriological, physicochemical and environmental relationships. PLoS One. 10(6):e0129382.

Myoda SP, Carson CA, Fuhrmann JJ, Hahm B-K, Hartel PG, Yampara-Iquise H, Johnson L, Kuntz RL, Nakatsu CH, Sadowsky MJ. 2003. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. J Water Health. 1(4):167–180.

Neilson BT, Stevens DK. 2002. Issues Related to the Success of the TMDL Program. J Contemp Water Res Educ. 122(1):8.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, et al. 2011. Global patterns in the biogeography of bacterial taxa. Environ Microbiol. 13(1):135–144. doi:10.1111/j.1462-2920.2010.02315.x.

Nevers MB, Byappanahalli MN, Phanikumar MS, Whitman RL. 2016. Fecal Indicator Organism Modeling and Microbial Source Tracking in Environmental Waters. In: Manual of Environmental Microbiology, Fourth Edition. American Society of Microbiology.

Noble RT, Allen SM, Blackwood AD, Chu W, Jiang SC, Lovelace GL, Sobsey MD, Stewart JR, Wait DA. 2003. Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study. J Water Health. 1(4):195–207.

Noori R, Karbassi A, Khakpour A, Shahbazbegian M, Badam HMK, Vesali-Naseh M. 2012. Chemometric Analysis of Surface Water Quality Data: Case Study of the Gorganrud River Basin, Iran. Environ Model {&} Assess. 17(4):411–420. doi:10.1007/s10666-011-9302-2.

Nwachuku N, Gerba CP. 2004. Microbial risk assessment: don’t forget the children. Curr Opin Microbiol. 7(3):206–209. doi:https://doi.org/10.1016/j.mib.2004.04.011.

Oest A, Alsaffar A, Fenner M, Azzopardi D, Tiquia-Arashiro SM. 2018. Patterns of change in metabolic capabilities of sediment microbial communities in river and lake ecosystems. Int J Microbiol. 2018.

Okabe S, Okayama N, Savichtcheva O, Ito T. 2007. Quantification of host-specific Bacteroides–Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl Microbiol Biotechnol. 74(4):890–901.

Oshiro R, Olson B. 1998. Coliforms and E. coli: Problem or Solution? In: Kay D, Fricher C, editors. Cambridge: Royal Society of Chemistry.

Ottoson J, Stenström TA. 2003. Faecal contamination of greywater and associated microbial risks. Water Res. 37(3):645–655. doi:10.1016/S0043-1354(02)00352-4.

Pandey PK, Kass PH, Soupir ML, Biswas S, Singh VP. 2014. Contamination of water resources by pathogenic bacteria. AMB Express. 4(1):1–16. doi:10.1186/s13568-014-0051-x.

Park S-J, Lee C-G, Kim S-B. 2008. Quantification of bacterial attachment-related parameters in porous media. Environ Eng Res. 13(3):141–146.

Parkhurst DF, Brenner KP, Dufour AP, Wymer LJ. 2005. Indicator bacteria at five swimming beaches—analysis using random forests. Water Res. 39(7):1354–1360.

Parveen S, Hodge NC, Stall RE, Farrah SR, Tamplin ML. 2001. Phenotypic and genotypic characterization of human and nonhuman Escherichia coli. Water Res. 35(2):379–386.

Parveen S, Portier KM, Robinson K, Edmiston L, Tamplin ML. 1999. Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. Appl Environ Microbiol. 65(7):3142–3147.

Patz JA, Mcgeehin MA, Bernard SM, Ebi KL, Epstein PR, Gubler DJ, Reiter P, Romieu I, Rose JB, Samet JM. 2000. The potential health impacts of climate variability and change for the United States: Executive summary of the report of the health sector of the U.S. National Assessment. Env Heal Perspect. 108(4):367–376.

Patz JA, Vavrus SJ, Uejio CK, McLellan SL. 2008. Climate change and waterborne disease risk in the Great Lakes region of the US. Am J Prev Med. 35(5):451–458.

Payment P, Berte A, Prévost M, Ménard B, Barbeau B. 2000. Occurrence of pathogenic microorganisms in the Saint Lawrence River (Canada) and comparison of health risks for populations using it as their source of drinking water. Can J Microbiol. 46(6):565–576.

Payment P, Franco E. 1993. Clostridium perfringens and somatic coliphages as indicators of the efficiency of drinking water treatment for viruses and protozoan cysts. Appl Environ Microbiol. 59(8):2418–2424.

Peck AM, Hornbuckle KC. 2004. Synthetic musk fragrances in Lake Michigan. Environ Sci Technol. 38(2):367–372.

Perkins TL, Perrow K, Rajko-Nenow P, Jago CF, Jones DL, Malham SK, McDonald JE. 2016. Decay rates of faecal indicator bacteria from sewage and ovine faeces in brackish and freshwater microcosms with contrasting suspended particulate matter concentrations. Sci Total Environ. 572:1645–1652. doi:https://doi.org/10.1016/j.scitotenv.2016.03.076.

Petrov V, Guedes Soares C, Gotovac H. 2013. Prediction of extreme significant wave heights using maximum entropy. Coast Eng. 74:1–10. doi:https://doi.org/10.1016/j.coastaleng.2012.11.009.

Pettine M, Patrolecco L, Manganelli M, Capri S, Farrace MG. 1999. Seasonal variations of dissolved organic matter in the northern Adriatic Sea. Mar Chem. 64(3):153–169. doi: 10.1016/S0304-4203(98)00071-1.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent : new extensions and a comprehensive evaluation. Ecography (Cop). 31(2):161–175. doi:10.1111/j.2007.0906-7590.05203.x.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on Machine learning. Association for Computing Machinery.

Phillips, S., Dudík, M., Schapire R. 2010. Maxent Software, version 3.3.3k.

Piorkowski G, Jamieson R, Bezanson G, Truelstrup L, Yost C. 2013. Evaluation of statistical models for predicting Escherichia coli particle attachment in fluvial systems. Water Res. 47(17):6701–6711. doi:10.1016/j.watres.2013.09.003.

Preston-Mafham J, Boddy L, Randerson PF. 2002. Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - A critique. FEMS Microbiol Ecol. 42(1):1–14. doi:10.1016/S0168-6496(02)00324-0.

Qu W, Morrison RJ, West RJ, Su C. 2005. Diagenetic stoichiometry and benthic nutrient fluxes at the sediment–water interface of Lake Illawarra, Australia. Hydrobiologia. 537(1-3):249–264.

Ries III KG, Newson JK, Smith MJ, Guthrie JD, Steeves PA, Haluska TL, Kolb KR, Thompson RF, Santoro RD, Vraga HW. 2017. StreamStats, version 4. Reston, VA.

Rosi-Marshall EJ, Vallis KL, Baxter C V, Davis JM. 2016. Retesting a prediction of the river continuum concept: autochthonous versus allochthonous resources in the diets of invertebrates. Freshw Sci. 35(2):534–543. doi:10.1086/686302.

Rutgers M, Wouterse M, Drost SM, Breure AM, Mulder C, Stone D, Creamer RE, Winding A, Bloem J. 2016. Monitoring soil bacteria with community-level physiological profiles using Biolog ECO-plates in the Netherlands and Europe. Appl Soil Ecol. 97:23–35.

Sala MM, Estrada M. 2006. Seasonal changes in the functional diversity of bacterioplankton in contrasting coastal environments of the NW Mediterranean . Aquat Microb Ecol. 44(1):1–9.

Sala MM, Pinhassi J, Gasol JM. 2006. Estimation of bacterial use of dissolved organic nitrogen compounds in aquatic ecosystems using Biolog plates. Aquat Microb Ecol. 42(1):1–5.

Sala MM, Terrado R, Lovejoy C, Unrein F, Pedros-Alio C. 2008. Metabolic diversity of heterotrophic bacterioplankton over winter and spring in the coastal Arctic Ocean. Environ Microbiol. 10(4):942–949. doi:10.1111/j.1462-2920.2007.01513.x.

Samadpour M, Roberts MC, Kitts C, Mulugeta W, Alfi D. 2005. The use of ribotyping and antibiotic resistance patterns for identification of host sources of Escherichia coli strains. Lett Appl Microbiol. 40(1):63–68.

Savage DC. 2001. Microbial biota of the human intestine: a tribute to some pioneering scientists. Curr Issues Intest Microbiol. 2(1):1–15.

Savichtcheva O, Okabe S. 2006. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res. 40(13):2463–2476. doi:10.1016/j.watres.2006.04.040.

Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, Blaschke AP, Blöschl G, Mach RL, Kirschner AKT, et al. 2015. Bacterial diversity along a 2600 km river continuum. 17:4994–5007. doi:10.1111/1462-2920.12886.

Schimel J, Balser TC, Wallenstein M. 2007. Microbial stress-response physiology and its implications for ecosystem function. Ecology. 88(6):1386–1394. doi:10.1890/06-0219.

Schneider BL, Reitzer L. 2012. Pathway and Enzyme Redundancy in Putrescine Catabolism in Escherichia coli. J Bacteriol. 194(15):4080–4088.

Scott TM, Jenkins TM, Lukasik J, Rose JB. 2005. Potential use of a host associated molecular marker in Enterococcus faecium as an index of human fecal pollution. Environ Sci Technol. 39(1):283–287.

Scott TM, Parveen S, Portier KM, Rose JB, Tamplin ML, Farrah SR, Koo A, Lukasik J. 2003. Geographical variation in ribotype profiles of Escherichia coli isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl Environ Microbiol. 69(2):1089–1092.

Scott TM, Rose JB, Jenkins TM, Farrah SR, Lukasik J. 2002. Microbial source tracking: current methodology and future directions. Appl Environ Microbiol. 68(12):5796–5803.

Seurinck S, Defoirdt T, Verstraete W, Siciliano SD. 2005. Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. Environ Microbiol. 7(2):249–259. doi:10.1111/j.1462-2920.2004.00702.x.

Shaibe E, Metzer E, Halpern YS. 1985. Metabolic pathway for the utilization of L-arginine, L-ornithine, agmatine, and putrescine as nitrogen sources in Escherichia coli K-12. J Bacteriol. 163(3):933–937.

Shanks OC, Kelty CA, Oshiro R, Haugland RA, Madi T, Brooks L, Field KG, Sivaganesan M. 2016. Data Acceptance Criteria for Standardized Human-Associated Fecal Source Identification Quantitative Real-Time PCR Methods. Besser TE, editor. Appl Environ Microbiol. 82(9):2773–2782. doi:10.1128/AEM.03661-15.

Shanks OC, Santo Domingo JW, Lamendella R, Kelty CA, Graham JE. 2006. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Appl Environ Microbiol. 72(6):4054–4060.

Shanks OC, Santo Domingo JW, Lu J, Kelty CA, Graham JE. 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. Appl Environ Microbiol. 73(8):2416–2422.

Shi P, Jia S, Zhang X-X, Zhang T, Cheng S, Li A. 2013. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res. 47(1):111–120.

Shirmohammadi A. 2006. Uncertainty in TMDL Models. Trans ASABE. 49(4):301–314.

Shu C, Ouarda T. 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour Res. 43(7).

Simpson JM, Santo Domingo JW, Reasoner DJ. 2002. Microbial source tracking: state of the science. Environ Sci Technol. 36(24):5279–5288.

Simpson JM, Santo Domingo JW, Reasoner DJ. 2004. Assessment of equine fecal contamination: the search for alternative bacterial source-tracking targets. FEMS Microbiol Ecol. 47(1):65–75.

Sinton LW, Finlay RK, Hannah DJ. 1998. Distinguishing human from animal faecal contamination in water: a review. New Zeal J Mar Freshw Res. 32(2):323–348.

Sinton LW, Hall CH, Lynch P a, Davies-Colley RJ. 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. Appl Environ Microbiol. 68(3):1122–1131. doi:10.1128/AEM.68.3.1122.

Smith A, Sterba-Boatwright B, Mott J. 2010. Novel application of a statistical technique, Random Forests, in a bacterial source tracking study. Water Res. 44(14):4067–4076. doi:https://doi.org/10.1016/j.watres.2010.05.019.

Smith JH, Wickham JD, Norton D, Wade TG, Jones KB. 2001. Utilization of landscape indicators to model potential pathogen impaired waters. JAWRA J Am Water Resour Assoc. 37(4):805–814.

Soller JA, Bartrand T, Ashbolt NJ, Ravenscroft J, Wade TJ. 2010. Estimating the primary etiologic agents in recreational freshwaters impacted by human sources of faecal contamination. Water Res. 44(16):4736–4747. doi:10.1016/j.watres.2010.07.064.

Soller JA, Bartrand T, Ravenscroft J, Molina M, Whelan G, Schoen M, Ashbolt N. 2015. Estimated human health risks from recreational exposures to stormwater runoff containing animal faecal material. Environ Model Softw. 72:21–32. doi:10.1016/j.envsoft.2015.05.018.

Soller JA, Schoen ME, Bartrand T, Ravenscroft JE, Ashbolt NJ. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. Water Res. 44(16):4674–4691. doi:https://doi.org/10.1016/j.watres.2010.06.049.

Soller JA, Schoen ME, Varghese A, Ichida AM, Boehm AB, Eftim S, Ashbolt NJ, Ravenscroft JE. 2014. Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. Water Res. 66:254–264. doi:10.1016/j.watres.2014.08.026.

Song J, Luo Y, Zhao Q, Christie P. 2004. Microcosm studies on anaerobic phosphate flux and mineralization of lake sediment organic carbon. J Environ Qual. 33(6):2353–2356.

Sorensen DL, Eberl SG, Dicksa RA. 1989. Clostridium perfringens as a point source indicator in non-point polluted streams. Water Res. 23(2):191–197.

Soupir ML, Mostaghimi S. 2011. Escherichia coli and enterococci attachment to particles in runoff from highly and sparsely vegetated grassland. Water, Air, Soil Pollut. 216(1-4):167–178.

de Sousa CP. 2006. Escherichia coli as a specialized bacterial pathogen. Rev biol cienc Terra. 2(2):341–352.

Souza V, Castillo A, Eguiarte LE. 2002. The Evolutionary Ecology of Escherichia coli. Am Sci. 90(4):332–341. doi:10.1511/2002.4.332.

Standley LJ, Kaplan LA, Smith D. 2000. Molecular tracers of organic matter sources to surface water resources. Environ Sci Technol. 34(15):3124–3130.

Stefanowicz A. 2006. The biolog plates technique as a tool in ecological studies of microbial communities. Polish J Environ Stud. 15(5):669–676.

Sterk A, Schijven J, de Nijs T, de Roda Husman AM. 2013. Direct and indirect effects of climate change on the risk of infection by water-transmitted pathogens. Environ Sci Technol. 47(22):12648–12660.

Stoeckel DM, Harwood VJ. 2007. Performance, design, and analysis in microbial source tracking studies. Appl Environ Microbiol. 73(8):2405–2415.

Stoeckel DM, Mathes M V, Hyer KE, Hagedorn C, Kator H, Lukasik J, O'Brien TL, Fenger TW, Samadpour M, Strickler KM. 2004. Comparison of seven protocols to identify fecal contamination sources using Escherichia coli. Environ Sci Technol. 38(22):6109–6117.

Storey RG, Fulthorpe RR, Williams DD. 1999. Perspectives and predictions on the microbial ecology of the hyporheic zone. Freshw Biol. 41(1):119–130.

Stott R, Davies-Colley R, Nagels J, Donnison A, Ross C, Muirhead R. 2011. Differential behaviour of Escherichia coli and Campylobacter spp. in a stream draining dairy pasture. J Water Health. 9(1):59–69. doi:10.2166/wh.2010.061.

Stow CA, Roessler C, Borsuk ME, Bowen JD, Reckhow KH. 2003. Comparison of Estuarine Water Quality Models for Total Maximum Daily Load Development in Neuse River Estuary. J Water Resour Plan Manag. 129(4):307–314.

Suprihatin I, Fallowfield H, Bentham R, Cromar N. 2003. Determination of faecal pollutants in Torrens and Patawalonga catchment waters in South Australia using faecal sterols. Water Sci Technol. 47(7-8):283–289.

Surbeck CQ, Jiang SC, Ahn JH, Grant SB. 2006. Flow fingerprinting fecal pollution and suspended solids in stormwater runoff from an urban coastal watershed. Environ Sci Technol. 40:4435–4441. doi:10.1021/es060701h.

Surbeck CQ, Jiang SC, Grant SB. 2010. Ecological control of fecal indicator bacteria in an urban stream. Environ Sci Technol. 44(2):631–637. doi:10.1021/es903496m.

Tan B, Ng C, Nshimyimana JP, Loh LL, Gin KY-H, Thompson JR. 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Front Microbiol. 6:1027. doi:10.3389/fmicb.2015.01027.

Tartera C, Jofre J. 1987. Bacteriophages active against Bacteroides fragilis in sewage-polluted waters. Appl Environ Microbiol. 53(7):1632–1637.

Tartera C, Jofre J, Lucena F. 1988. Relationship between numbers of enteroviruses and bacteriophages infecting bacterowes fragilis in different environmental samples. Environ Technol. 9(5):407–410.

Tennessee Department of Environmental and Conservation. 2006. Proposed total maximum daily load (TMDL) for E. Coli in the Watauga River Watershed (HUC 06010103).

Tennessee Department of Environmental and Conservation. 2015a. Chapter 0400-40-03 General Water Quality Criteria. In: Rules of the Tennessee Department of Environment and Conservation.

Tennessee Department of Environmental and Conservation. 2015b. How Tennessee Implements the Priority Goal for TMDL Development Under the CWA 303(d) Long-Term Vision.

Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M. 2005. Niche based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Glob Chang Biol. 11(12):2234–2250.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B. Series B(Methodological):267–288.

Timoner X, Borrego CM, Acuña V, Sabater S. 2014. The dynamics of biofilm bacterial communities is driven by flow wax and wane in a temporary stream. Limnol Oceanogr. 59(6):2057–2067. doi:10.4319/lo.2014.59.6.2057.

Tiquia SM. 2010. Metabolic diversity of the heterotrophic microorganisms and potential link to pollution of the Rouge River. Environ Pollut. 158(5):1435–1443. doi:http://dx.doi.org/10.1016/j.envpol.2009.12.035.

Tomanova S, Tedesco PA, Campero M, Van Damme PA, Moya N, Oberdorff T. 2007. Longitudinal and altitudinal changes of macroinvertebrate functional feeding groups in neotropical streams: a test of the River Continuum Concept. Fundam Appl Limnol. 170(3):233–241.

Tornwall B, Sokol E, Skelton J, Brown BL. 2015. Trends in stream biodiversity research since the river continuum concept. Diversity. 7(1):16–35. doi:10.3390/d7010016.

Tsai Y-L, Le JY, Olson BH. 2003. Magnetic bead hybridization to detect enterotoxigenic Escherichia coli strains associated with cattle in environmental water sources. Can J Microbiol. 49(6):391–398.

Tyrrel SF, Quinton JN. 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. J Appl Microbiol. 94:87–93. doi:10.1046/j.1365-2672.94.s1.10.x.

United States Environmental Protection Agency. 2001a. Protocol for Developing Pathogen TMDLs First Edition.

United States Environmental Protection Agency. 2001b. Method 1601: Male-specific (F +) and Somatic Coliphage in Water by Two-step Enrichment Procedure.

United States Environmental Protection Agency. 2012. Recreational Water Quality Criteria.

United States Environmental Protection Agency. 2017. National summary of impaired waters and TMDL information. [accessed 2017 Aug 1]. http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T.

Unno T, Jang J, Han D, Kim JH, Sadowsky MJ, Kim O-S, Chun J, Hur H-G. 2010. Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environ Sci Technol. 44(20):7777–7782.

Urakawa H, Ali J, Ketover RDJ, Talmage SD, Garcia JC, Campbell IS, Loh AN, Parsons ML. 2013. Shifts of bacterioplankton metabolic profiles along the salinity gradient in a subtropical estuary. ISRN Oceanogr. 2013.

Vadstein O, Jensen A, Olsen Y, Reinertsen H. 1988. Growth and phosphorus status of limnetic phytoplankton and bacteria. Limnol Oceanogr. 33(4):489–503.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. 1980. The river continuum concept. Can J Fish Aquat Sci. 37(1):130–137. doi:10.1139/f80-017.

Viau EJ, Goodwin KD, Yamahara KM, Layton BA, Sassoubre LM, Burns SL, Tong H-I, Wong SHC, Lu Y, Boehm AB. 2011. Bacterial pathogens in Hawaiian coastal streams—Associations with fecal indicators, land cover, and water quality. Water Res. 45(11):3279–3290. doi:https://doi.org/10.1016/j.watres.2011.03.033.

Vidon P, Campbell MA, Gray M. 2008. Unrestricted cattle access to streams and water quality in till landscape of the Midwest. Agric water Manag. 95(3):322–330.

Vital M, Hammes F, Egli T. 2008. Escherichia coli O157 can grow in natural freshwater at low carbon concentrations. Environ Microbiol. 10(9):2387–2396. doi:10.1111/j.1462-2920.2008.01664.x.

Vitro KA, BenDor TK, Jordanova T V, Miles B. 2017. A geospatial analysis of land use and stormwater management on fecal coliform contamination in North Carolina streams. Sci Total Environ. 603-604:709–727. doi:10.1016/j.scitotenv.2017.02.093.

Vreča P. 2003. Carbon cycling at the sediment–water interface in a eutrophic mountain lake (Jezero na Planini pri Jezeru, Slovenia). Org Geochem. 34(5):671–680.

Wade TJ, Calderon RL, Sams E, Beach M, Brenner KP, Williams AH, Dufour AP. 2006. Rapidly Measured Indicators of Recreational Water Quality Are Predictive of Swimming-Associated Gastrointestinal Illness. Environ Health Perspect. 114(1):24–28. doi:10.1289/ehp.8273.

Wade TJ, Pai N, Eisenberg JNS, Colford JM. 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Perspect. 111(8):1102–1109.

Walsh CJ, Roy AH, Feminella JW, Cottingham PD, Groffman PM, Morgan II, RP. 2005. The urban stream syndrome : current knowledge and the search for a cure The urban stream syndrome : current knowledge and. J North Am Benthol Soc. 24(3):706–723.

Walters SP, Gannon VPJ, Field KG. 2007. Detection of Bacteroidales fecal indicators and the zoonotic pathogens E. coli 0157:H7, salmonella, and campylobacter in river water. Environ Sci Technol. 41(6):1856–1862.

Weniger BG, Blaser MJ, Gedrose J, Lippy EC, Juranek DD. 1983. An outbreak of waterborne giardiasis associated with heavy water runoff due to warm weather and volcanic ashfall. Am J Public Health. 73(8):868–872. doi:10.2105/AJPH.73.8.868.

Whitlock JE, Jones DT, Harwood VJ. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Res. 36(17):4273–4282.

Whitman R, Nevers MB, Byappanahalli M. 2006. Examination of the watershed-wide distribution of Escherichia coli along southern Lake Michigan: An integrated approach. Appl Environ Microbiol. 72(11):7301–7310. doi:10.1128/AEM.00454-06.

Whitman RL, Nevers MB. 2003. Foreshore Sand as a Source of Escherichia coli in Nearshore Water of a Lake Michigan Beach. Appl Environ Microbiol. 69(9):5555–5562. doi:10.1128/AEM.69.9.5555.

Whitman RL, Nevers MB, Korinek GC, Byappanahalli MN. 2004. Solar and Temporal Effects on Escherichia coli Concentration at a Lake Michigan Swimming Beach Solar and Temporal Effects on Escherichia coli Concentration at a Lake Michigan Swimming Beach †. Appl Environ Microbiol. 70(7):4276. doi:10.1128/AEM.70.7.4276.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes : The unseen majority. In: Proceedings of the National Academy of Sciences. Vol. 95. p. 6578–6583.

Wiggins BA. 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl Environ Microbiol. 62(11):3997–4002.

Wiggins BA, Cash PW, Creamer WS, Dart SE, Garcia PP, Gerecke TM, Han J, Henry BL, Hoover KB, Johnson EL, et al. 2003. Use of Antibiotic Resistance Analysis for Representativeness Testing of Multiwatershed Libraries. Appl Environ Microbiol. 69(6):3399–3405. doi:10.1128/AEM.69.6.3399-3405.2003.

Wilkes G, Brassard J, Edge TA, Gannon V, Jokinen CC, Jones TH, Marti R, Neumann NF, Ruecker NJ, Sunohara M, et al. 2013. Coherence among different microbial source tracking markers in a small agricultural stream with or without livestock exclusion practices. Appl Environ Microbiol. 79(20):6207–6219. doi:10.1128/AEM.01626-13.

Wilkes G, Edge TA, Gannon VPJ, Jokinen C, Lyautey E, Neumann NF, Ruecker N, Scott A, Sunohara M, Topp E, et al. 2011. Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. Water Res. 45(18):5807–5825. doi:10.1016/j.watres.2011.06.021.

Williams AP, Quilliam RS, Thorn CE, Cooper D, Reynolds B, Jones DL. 2012. Influence of Land Use and Nutrient Flux on Metabolic Activity of E. coli O157 in River Water. Water, Air, Soil Pollut. 223(6):3077–3083. doi:10.1007/s11270-012-1090-z.

Williams DD. 2006. The biology of temporary waters. Oxford University Press.

Wilson JT, McNabb JF, Balkwill DL, Ghiorse WC. 1983. Enumeration and Characterization of Bacteria Indigenous to a Shallow Water-Table Aquifer. Ground Water. 21(2):134–142. doi:10.1111/j.1745-6584.1983.tb00710.x.

Winfield MD, Groisman EA. 2003. Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. Appl Environ Microbiol. 69(7):3687–3694. doi:10.1128/AEM.69.7.3687.

Wu J, Long SC, Das D, Dorner SM. 2011. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. J Water Health. 9(2):265–278.

Yates M V. 2007. Classical indicators in the 21st century—far and beyond the coliform. Water Environ Res. 79(3):279–286.

Zak JC, Willig MR, Moorhead DL, Wildman HG. 1994. Functional diversity of microbial communities: a quantitative approach. Soil Biol Biochem. 26(9):1101–1108.

Zhang C-B, Wang J, Liu W-L, Zhu S-X, Ge H-L, Chang SX, Chang J, Ge Y. 2010. Effects of plant diversity on microbial biomass and community metabolic profiles in a full-scale constructed wetland. Ecol Eng. 36(1):62–68.

Zhou L-J, Ying G-G, Zhang R-Q, Liu S, Lai H-J, Chen Z-F, Yang B, Zhao J-L. 2013. Use patterns, excretion masses and contamination profiles of antibiotics in a typical swine farm, south China. Environ Sci Process Impacts. 15(4):802–813.

Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 115(5):654–657.

Zweig MH, Campbell G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 39(4):561–577.

APPENDIX

Supplemental Files for each Chapter

Supplemental Files for Chapter 2

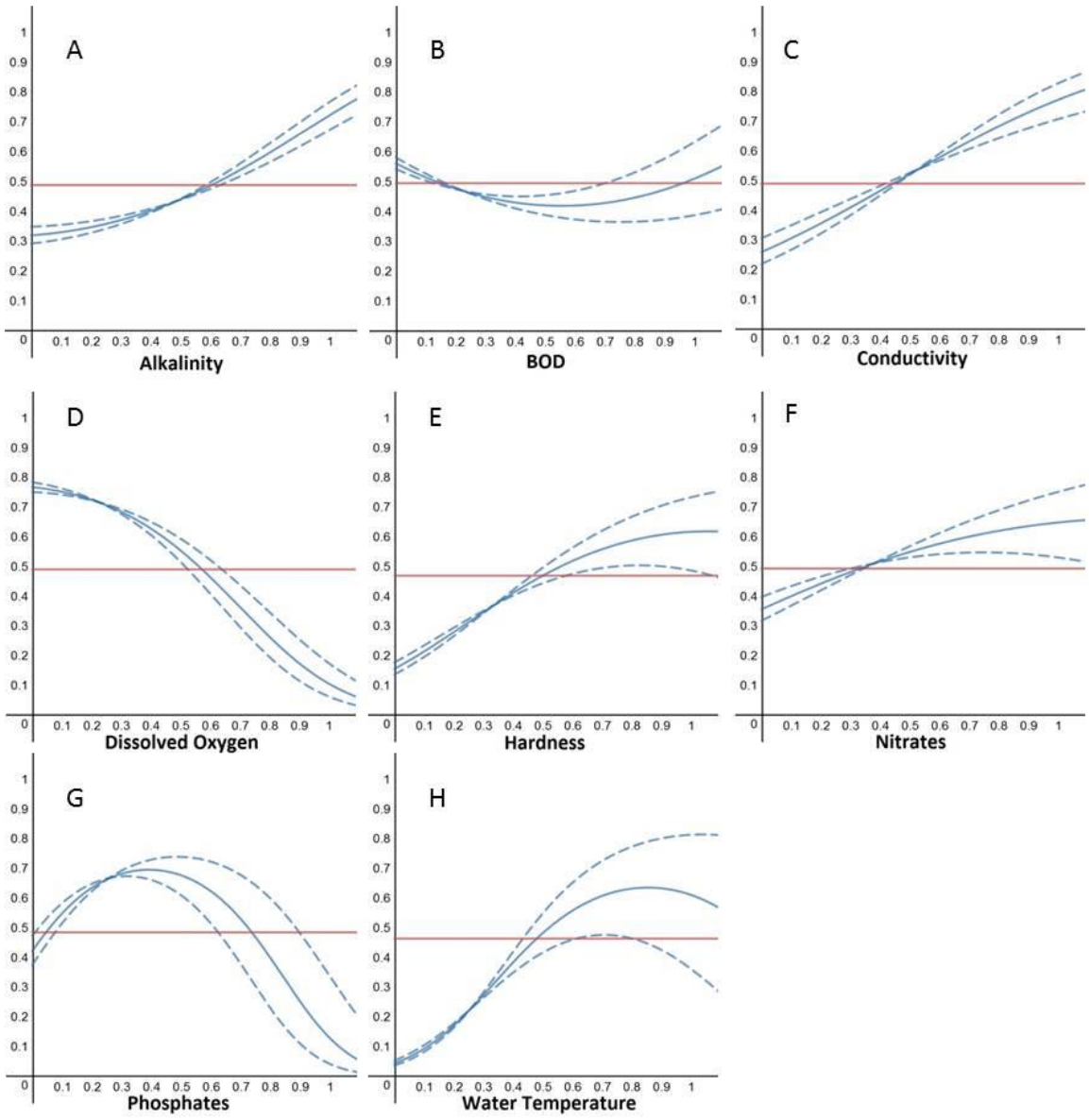

Figure 2.S1. Action functions for each of the univariate models with each variable scaled from 0 to 1 and the vertical axis representing the probability of impairment. Red line represents the logistic threshold, and dotted lines represent the 95% CI for the parameters estimated. Action

function for (A) alkalinity, (B) BOD_5, (C) conductivity, (D) dissolved oxygen, (E) hardness, (F) NO_3^-, (G) PO_4^{3-}, and (H) water temperature are shown.

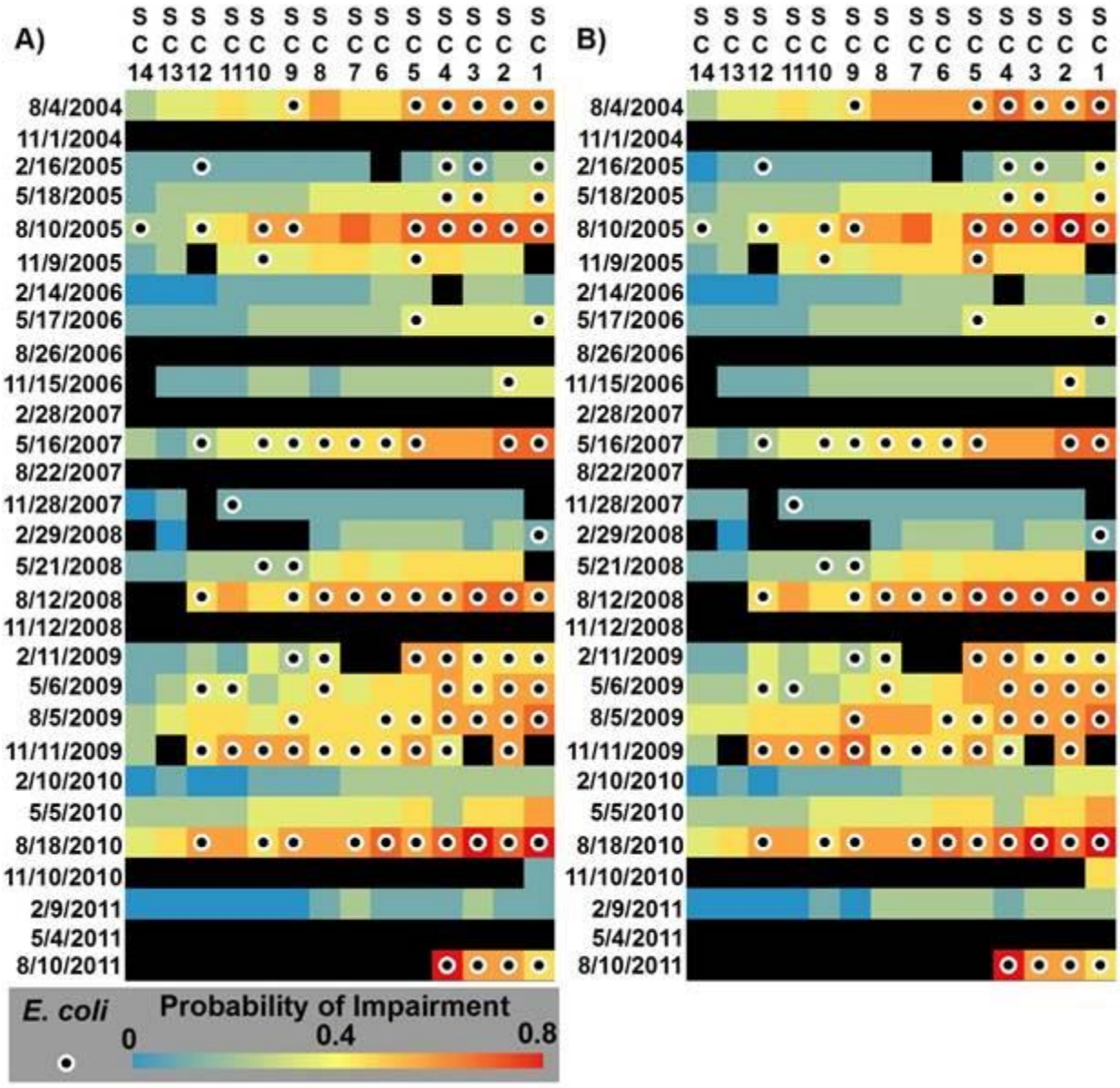

Figure 2.S2. Response surfaces for the 8-variable (A) and 5-variable model (B) showing the probability of each sample for the monitoring program. This represents the mean probability of 100 bootstrapped runs. Rows are oriented by month of sampling, while columns represent each sampling site.

Table 2.S1. Summary statistics of data used in univariate and multivariate Maxent models of *Escherichia coli* impairment.

	Mean	Standard Deviation	Minimum	Maximum
Alkalinity (mg/L)	99.62	52.25	4.00	210.00
BOD_5 (mg/L)	1.66	1.42	0.02	6.43
Conductivity (µS)	218.98	123.06	11.00	676.00
Discharge (m/s^2)	0.77	2.74	0.00	32.75
Dissolved Oxygen (mg/L)	10.19	1.96	0.79	15.90
Hardness (mg/L)	126.15	57.03	7.30	256.70
Nitrates (mg/L)	1.44	0.89	0.00	5.37
pH	7.56	0.45	6.25	8.74
Phosphates (mg/L)	0.36	0.72	0.00	10.04
Water Temperature (^oC)	12.39	4.40	1.30	24.50

Table 2.S2. Classification performance for all models run based on maximum test sensitivity and specificity as the logistic threshold (decision boundary).

Variables	**Logistic Threshold**	**TN**	**FN**	**FP**	**TP**	**χ^2**
Alkalinity	0.488	140	44	44	51	24.61
BOD_5	0.495	110	36	74	59	12.04
Conductivity	0.490	129	41	55	54	19.05
Dissolved Oxygen	0.493	128	34	56	61	29.30
Discharge	0.482	155	72	29	23	2.86
Hardness	0.469	102	30	82	65	14.34
Nitrates	0.493	134	52	50	43	9.16
pH	0.491	104	44	80	51	2.63
Phosphates	0.484	138	55	46	40	8.54
Water Temperature	0.463	118	31	66	64	24.58

8 variables	0.383	144	20	40	75	82.94
4 variables	0.424	152	30	32	65	70.52
5 variables	0.430	151	29	33	66	71.26

Table 2.S3. Probability of Impairment and associated 95% confidence intervals for the total creek, each site, and each month that was sampled.

	8-variable Mean Probability (95% CI)	5-variable Mean Probability (95% CI)	4-variable Mean Probability (95% CI)
Sinking Creek	0.338	0.353	0.359
	(0.319, 0.358)	(0.334, 0.373)	(0.340, 0.378)
SC1	0.427	0.456	0.455
	(0.343, 0.512)	(0.379, 0.534)	(0.376, 0.535)
SC2	0.434	0.453	0.456
	(0.364, 0.503)	(0.384, 0.522)	(0.384, 0.527)
SC3	0.410	0.430	0.434
	(0.334, 0.485)	(0.355, 0.505)	(0.359, 0.508)
SC4	0.409	0.418	0.430
	(0.332, 0.486)	(0.351, 0.508)	(0.349, 0.510)
SC5	0.403	0.418	0.413
	(0.337, 0.470)	(0.353, 0.483)	(0.348, 0.477)
SC6	0.373	0.385	0.386
	(0.314,0.433)	(0.326, 0.444)	(0.327,0.445)
SC7	0.359	0.372	0.371
	(0.295,0.423)	(0.308, 0.435)	(0.307, 0.434)

SC8	0.336	0.351	0.360
	(0.269,0.403)	(0.285, 0.418)	(0.295, 0.425)
SC9	0.317	0.331	0.343
	(0.242,0.391)	(0.255, 0.406)	(0.271, 0.416)
SC10	0.289	0.305	0.315
	(0.232,0.347)	(0.247, 0.364)	(0.258, 0.372)
SC11	0.292	0.303	0.316
	(0.221, 0.363)	(0.233, 0.373)	(0.249, 0.382)
SC12	0.276	0.288	0.303
	(0.202, 0.350)	(0.214, 0.362)	(0.233, 0.374)
SC13	0.189	0.195	0.211
	(0.140, 0.237)	(0.146, 0.245)	(0.167, 0.255)
SC14	0.176	0.184	0.187
	(0.132, 0.221)	(0.139, 0.230)	(0.148, 0.226)
February	0.225	0.211	0.191
	(0.223, 0.228)	(0.209, 0.214)	(0.188, 0.194)
May	0.353	0.352	0.344
	(0.351, 0.356)	(0.349, 0.355)	(0.342, 0.347)
August	0.535	0.529	0.511
	(0.531, 0.538)	(0.526, 0.532)	(0.508, 0.514)
November	0.317	0.315	0.299
	(0.312, 0.323)	(0.309, 0.321)	(0.294, 0.305)

Supplemental Files for Chapter 3

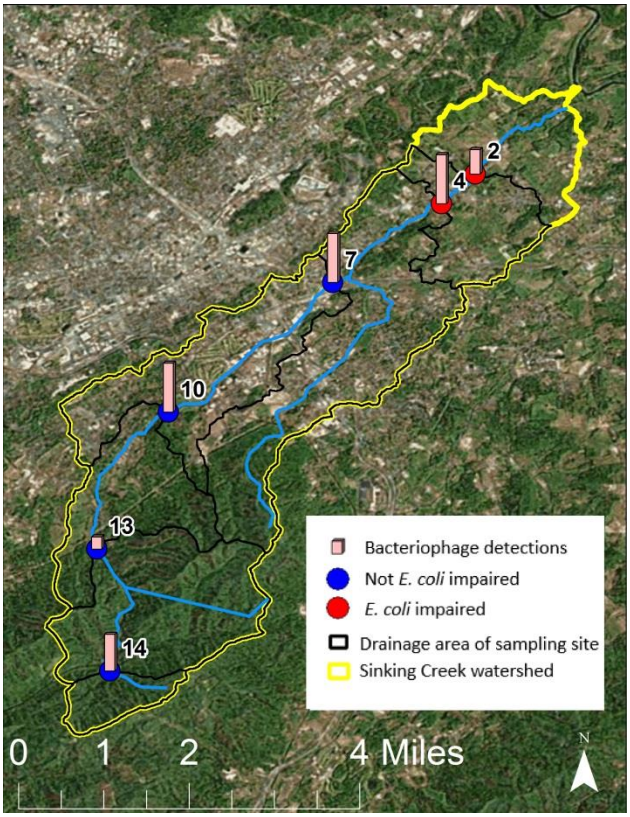

Figure 3.S1. Map of sampling sites, total watershed area, and individual sampling site drainage areas. Sampling sites that are in violation of the 2012 recreational water quality criteria are represented by red circles and those that do not violate one of the criteria are represented by blue circles. Bacteriophage detections are represented by pink bar graphs. For reference, Site 13 is equal to one bacteriophage detection.

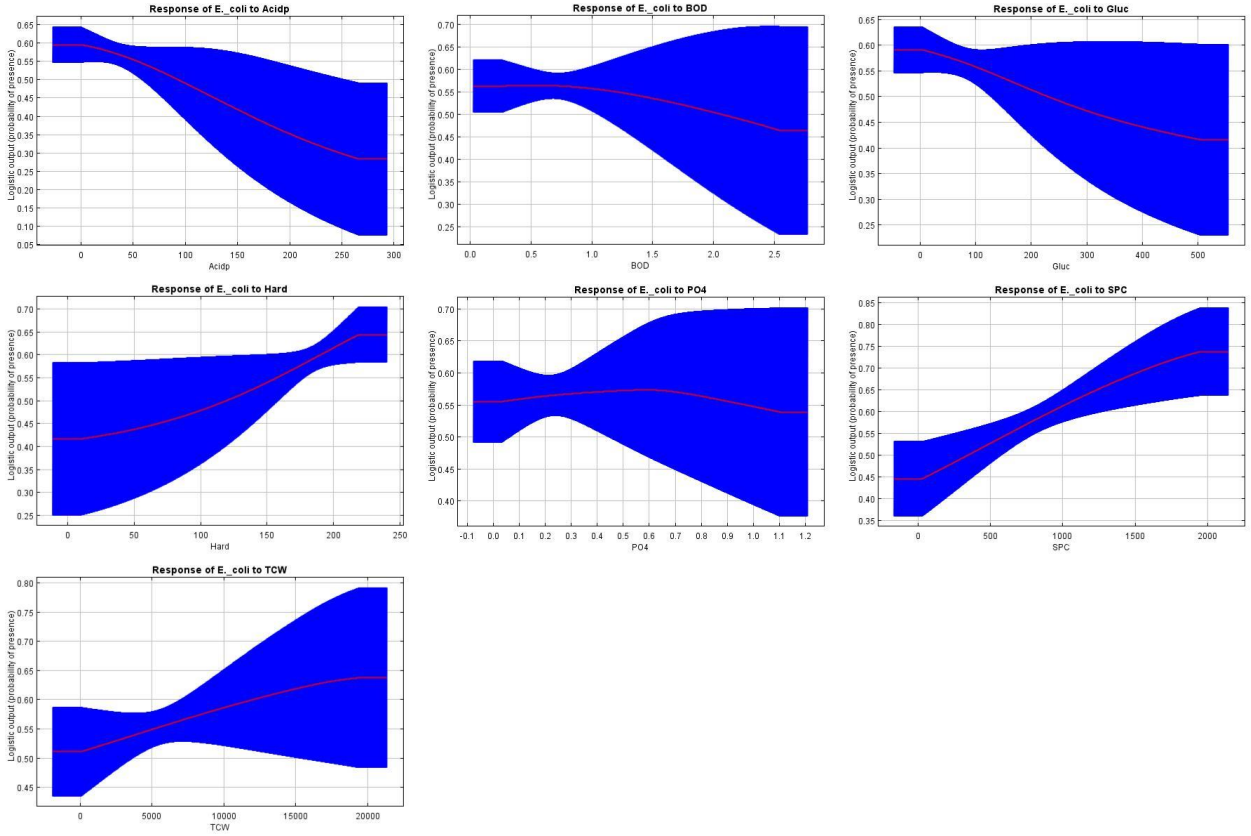

Figure 3.S2. Response curves for each variable included in the optimized *E. coli* model. These curves show how probability of impairment changes as each parameter is changed, keeping all other environmental variables at their mean value. The curves show the mean response (red) +/- one standard deviation (blue).

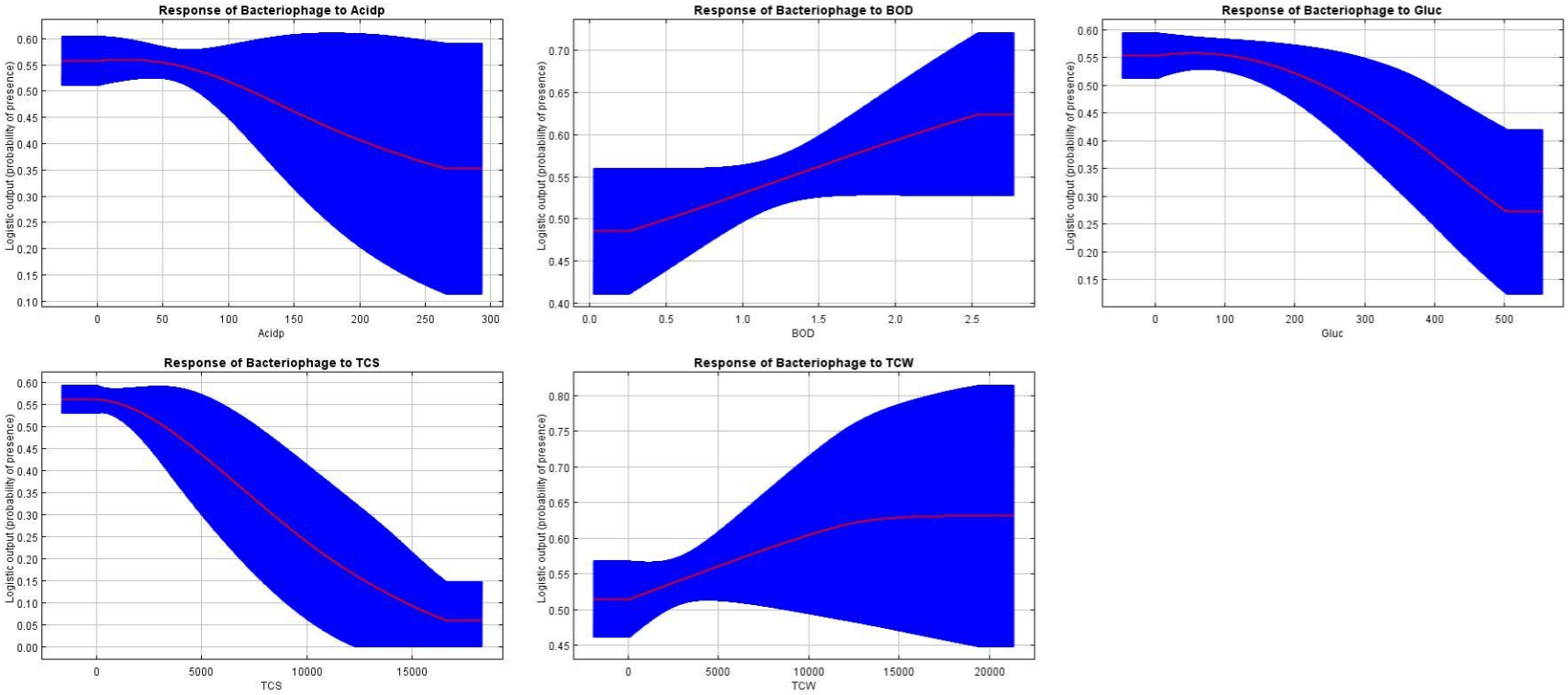

Figure 3.S3. Response curves for each variable included in the optimized bacteriophage model. These curves show how probability of impairment changes as each parameter is changed, keeping all other environmental variables at their mean value. The curves show the mean response (red) +/- one standard deviation (blue).

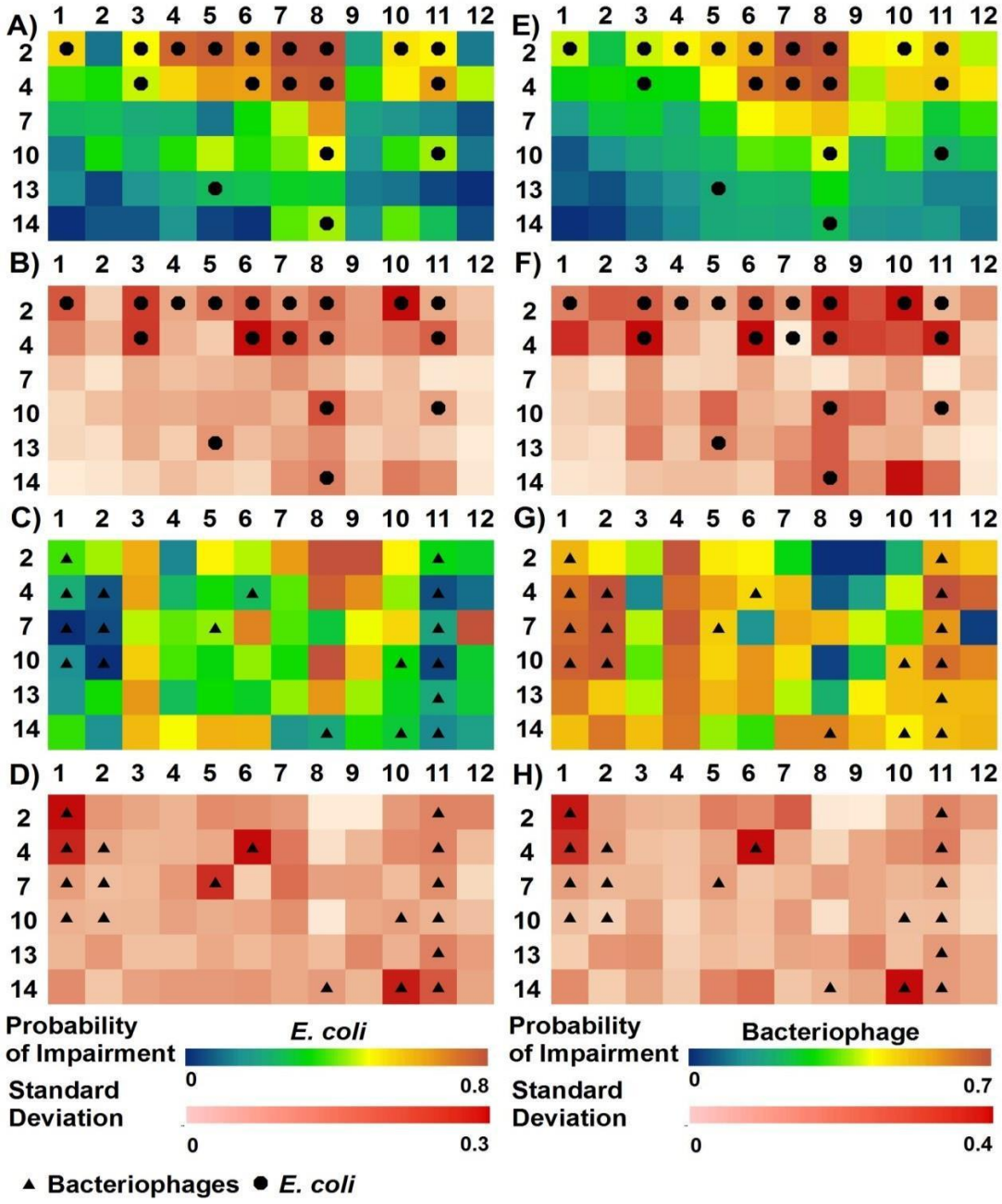

Figure 3.S4. Responses surface plots for the initial (A,C) and CCorA (E, G) models and associated standard deviations (B, D, F, H, respectively). A and E represent the initial and CCorA *E. coli* models, while C and G represent the initial and CCorA bacteriophage models. Rows in the plots represent each sampling site and columns represent each month that a sample was taken from the site.

Table 3.S1. Mean, standard deviation, and maximum and minimum values for the parameters used in the Maxent models. All means and standard deviations (SD) are arithmetic means, unless noted otherwise.

Parameter	**Mean (SD)**	**Max, Min**
FCW	304.5 (4.2) †	3387.7, 50.0
TCW	1151.3 (4.8) †	3508.8, 50.0
FCS	88.2 (3.2) †	1168.3, 25.0
TCS	712.7 (6.5) †	16655.3, 25
Colilert	34.0 (5.5) †	1299.7, 1.0
Bacteriophage	0.7 (3.0) †	1212.0, 0.5
SPC	309.0 (2.3) †	1984.3, 28.3
AODC	$1.26x10^8$(1.9) †	$6.8x10^8$, $284x10^7$
AcidP	61.7 (45.9) †	266.2, 0.1
AlkP	211.7 (207.9)	858.6, 3.3
DHA	27.6 (15.6)	84.4, 5.0
Gal	21.8 (26.7)	128, 0.5
Gluc	118.8 (119.47)	504.2, 3.4
NO_3^-	1.2 (0.5)	2.7, 0.3
PO_4^{3-}	0.3 (0.2)	1.1, 0.0
NH_4	0.1 (0.05)	0.3, 0.0
BOD_5	1.1 (0.6)	2.5, 0.3
Hard	111.4 (67.8)	219.0, 10.0
Alk	97.9 (61.7)	196.3, 8.0

†*Geometric mean and geometric standard deviation*

Supplemental Files for Chapter 4

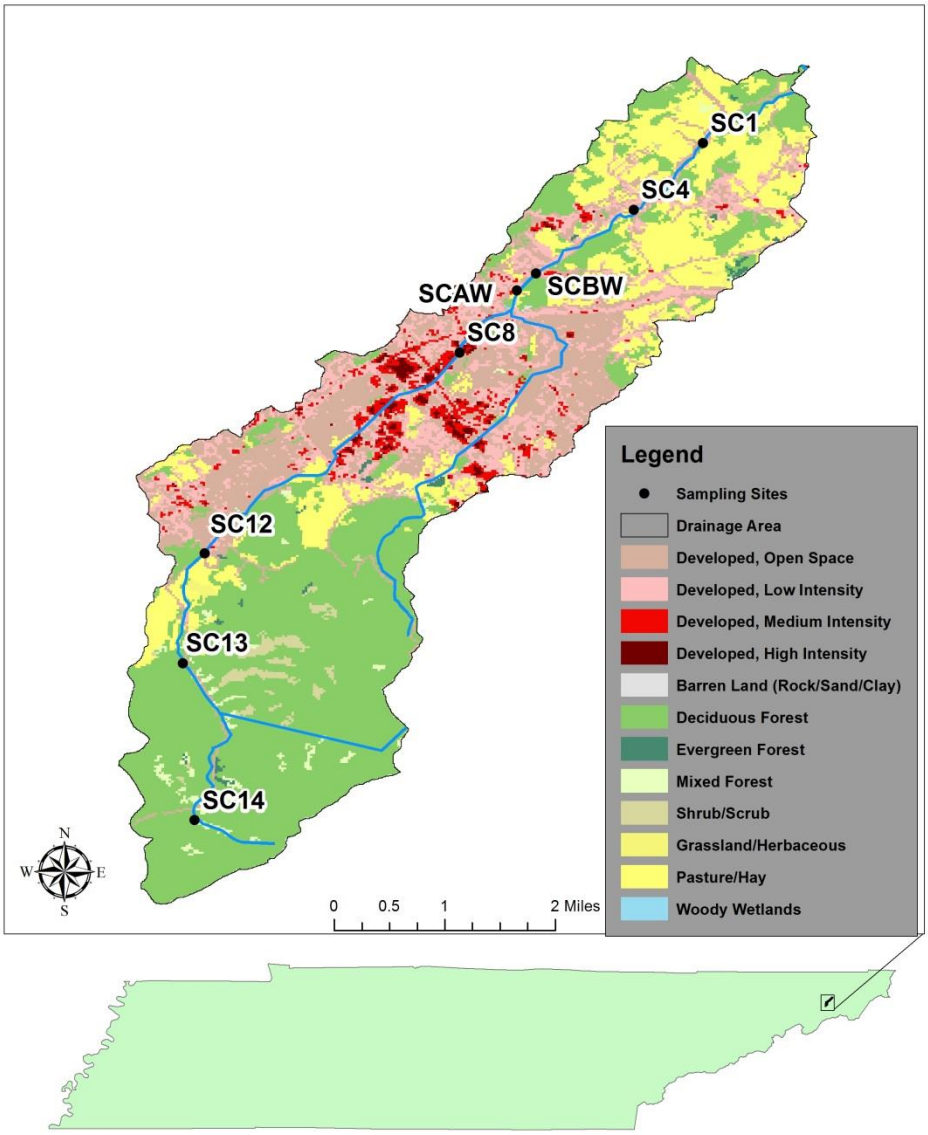

Figure 4.S1. Map of sampling locations used for Chapter 4. The inset map shows the state of Tennessee, and the location of Sinking Creek. Samples were taken from November 2016 to October 2017 on an monthly basis. The outline represents the watershed boundary of Sinking Creek, and 2006 NLCD has been clipped to the watershed.(Fry et al. 2011) Stream flows from its headwaters at SC14 downstream to SC1. Note that two sites are not in Figure 2.1, SCAW and SCBW.

VITA

DENNIS ANDREW GILFILLAN

Personal Information:	Date of Birth: October 11th, 1985
	Place of Birth: Vienna, West Virginia
Education:	Ph.D. Environmental Health Sciences. East Tennessee State University. Johnson City, TN 2018
	Master of Environmental Assessment. North Carolina State University. Raleigh, NC 2014
	Graduate Certificate. Geographic Information Systems. North Carolina State University. Raleigh, NC 2013
	B.S. Physics-Secondary Education, Minor in Mathematics. Appalachian State University. Boone, NC 2008
Professional Experience:	Instructor of Record. ENVH 3040 – Environmental Sanitation. East Tennessee State University. Johnson City, TN 2018
	Instructor of Record. ENVH 3010 – Human Ecology. East Tennessee State University. Johnson City, TN 2016-2017
	Lab Instructor. ENVH 4387 –Biological Analysis in Environmental Health. East Tennessee State University. Johnson City, TN 2015-2018
	Graduate Research Assistant. Environmental Health Sciences Laboratory. East Tennessee State University. Johnson City, TN, 2014 – 2018
	Advanced Placement Physics Instructor. Mitchell High School. Bakersville, NC 2012 – 2014
	Lead Math Instructor. Camp Spring Creek. Bakersville, NC, 2012 – 2014
	Secondary Math and Physics Instructor. Mitchell High School. Bakersville, NC, 2009 – 2014

Secondary Math Instructor. Rosman High School. Rosman, NC, 2008-2009

University Tutor-Physics. Department of Physics and Astronomy. Appalachian State University. Boone, NC 2005 - 2007

Publications:

Gilfillan D. Hall, K. Joyner, T.A. Scheuerman P. Canonical variable selection for ecological modeling of fecal indicators. Journal of Environmental Quality. 2018

Gilfillan D. Joyner T. A. Scheuerman P. Maxent estimation of aquatic Escherichia coli impairment. PeerJ 6:e5610. 2018

Conference Presentations:

Gilfillan D. Scheuerman P. Metabolic fingerprinting for impaired watersheds. Tennessee Environmental Conference. Kingsport, TN. Oral presentation. 2018

Gilfillan D. Alexander R. Scheuerman P. Community level physiologic profiles as a source tracking mechanism for surface water pollution. Appalachian Student Research Forum. Johnson City, TN. Poster Presentation. 2017

Gilfillan D. Scheuerman P. Joyner. T. A. Seasonal and spatial effects on the probability of pathogenic stream impairment. American Association of Geographers Annual Meeting. San Francisco, CA. Illustrated Paper Session. 2016

Honors and Awards:

2nd Place poster, Appalachian Student Research Forum – Society, Behavior, and Learning, 2017

ETSU Graduate Professional Student Association travel awards, $500, 2016, 2017

Coach of fastest 3200m runner in NCHSAA 1A history, 2014

North Carolina Teaching Fellow, $26,000, 2004 – 2008