FIND, INFORM, AND TEST (FIT): A SPATIAL MODELING FRAMEWORK TO ESTIMATE CONTRIBUTIONS OF SPATIALLY DISTRIBUTED SOURCES TO MICROBIAL CONTAMINANTS IN THE ENVIRONMENT

Corinne E. Wiesner-Friedman

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering.

Chapel Hill
2021

Approved by:

Marc Serre

Jill Stewart

Rebecca Fry

Leena Nylander-French

Ana Rule

ABSTRACT

Corinne E. Wiesner-Friedman: Find, Inform, and Test (Fit): A Spatial Modeling Framework to Estimate Contributions of Spatially Distributed Sources to Microbial Contaminants in the Environment
(Under the direction of Marc Serre)

The dissemination of fecal contamination and antibiotic-resistant bacteria from Concentrated Animal Feeding Operations (CAFOs) is a major public health concern. There is a need for modeling the dissemination of microbial contamination from sources to be able to link public health risks to sources such as CAFOs, especially in aquatic environments, such as rivers and streams. Host-associated microbial source tracking markers are a major advancement to quantifying risks associated with water use but fall short of relating microbial contamination with source locations. Using host-associated markers in conjunction with water quality modeling can help to establish strong linkages between spatial features and fecal contamination. However, existing water quality models do not have desirable qualities. They are often developed to model nutrient pollution, rely on a priori knowledge of delivery parameters, require information about the contaminants at sources, and most often model at the watershed level (i.e., sub-watershed or point-level resolution is preferred). Land-use regression (LUR) models address some of these concerns and have been previously implemented to associate sources and land cover with microbial contamination responses measured in the environment. Few of these microbial land-use regression studies have modeled quantified genes corresponding to the influence of specific hosts or genes encoding antibiotic resistance. Additionally, these land-use regression studies have primarily focused on human sources of microbial contamination and microbial contamination in surface water. This body of work advances previous microbial land-use regression approaches by developing a microbial land-use regression model and new spatial predictor models to characterize source contributions to river networks. This novel microbial modeling framework is applied to host-associated fecal contamination responses and antibiotic resistance responses in surface water and sediment in a Wisconsin spatial stream network for a CAFO-dense region, with many other potential sources. From these applications, this work contributes to the literature on offsite migration of microbial contamination from CAFOs, advances knowledge about other sources of host-associated and antibiotic resistance gene responses in riverbed sediment and surface water and contributes insights into how the characteristics of molecular targets can influence the characterization of source impacts on the environment and microbial risks. This research advances the field of microbial water quality modeling and provides quantifiable associations between sources and microbial contamination responses that can be considered for assessing current regulatory standards for different sources and investing in future wastewater treatment infrastructure in livestock-dense areas.

I would like to dedicate this work to everyone in my life that has surrounded me with their insatiable love of learning and creating.

I would also like to dedicate this to my dear friend Sara Planer, who needlessly encourages my sense of humor:

If the schlemiel is the person who spills, and the schlimazel is the person who gets spilt on, that would make me the… schlimodeler?

ACKNOWLEDGEMENTS

I would first like to thank my co-advisors Dr. Marc Serre and Dr. Jill Stewart, who are the best advisors a graduate student could get.

Marc, thank you so much for all of the time that you have spent in meetings with me. You are a truly amazing mentor and teacher. I have grown so much over the past 5 years and such a large part of that was working with you. You have challenged me in this program, but I would describe the time that I spent working under your mentorship as the best experience of my life. I cannot express how much I will miss it.

Jill, thank you for all that I learned about microbiology and all the friends I made in your lab. They are not only among the best scientists I know, but are also the kindest, quirkiest, and most generous people. Thank you for creating this environment that fuels positive experiences in independent research and collaboration.

I would like to thank my committee members, Dr. Rebecca Fry, Dr. Leena Nylander-French, and Dr. Ana Rule who have guided my coursework, my research, helped me discover different reaches of environmental science and engineering. Thank you to Rebecca, who helped to keep me funded for 4 years of my Ph.D.

I would like to thank Dr. Rachelle Beattie and Dr. Krassimira Hristova from Marquette University for collecting data that I used for this dissertation and for providing expertise about the study area. Thank you especially to Rachelle for your friendship and support.

There are a few other faculty members that have been important to my formation in this community. Thank you to Dr. Courtney woods for the teaching assistant experience with which I was funded my first year, and later for inviting me to join such lovely capoeira events at Earth Seed. Thank you, Will Vizuete for making time to talk about air and jazz and jazz on the air as well as for joining and supporting the Gillings Jazz Band.

Thank you to the Environmental Sciences and Engineering department for its excellence in a research and encouragement of the culture of health and work-life balance. Thank you especially to the chair of our department, Dr. Barbara Turpin for all of the conversations and opportunities that you have offered me and your dedication to ideals that are important in a department like ours, such as sustainability and equity.

I have learned as much from my peers as I have from my advisors and professors. Thank you to Dr. Sarah Rhodes for teaching me microbiology and showing me the best example of what friendship can mean. Thank you Dr. Prahlad Jat and Dr. Dalshad Al-Gaaf for your support and many philosophical conversations. Thank you to Connor LaMontagne for our friendship and for helping to form the Gillings Jazz Band, which I hope continues for many years. Thank you to Alyssa Grube, who entered the program at the same time as me, has walked with me through this program, and who has taught me an incredible amount about antimicrobial resistance and positive thinking. Thank you to Kristen Downs who has always been there for me when I needed someone to work socially with. Thank you to other members of the Stewart Lab who have provided a real sense of scientific community and friendship: Dr. David Holcomb, Dr. Elizabeth Christenson, Lindsay Wickersham, Knox Collin Coleman, Nikhil Gothegal, and Taylor Franklin. Thank you to other members of the BMElab for always making some time to get together, though we were often working in on different things and in different places: Alejandro Valencia, Risa Sayre, Stephanie Cleland, Ricky Strott, Angie Bittner, Claire Dust, and Calvin Arter.

I would also like to thank a number of environmental sciences and engineering staff members without whom the department could not exist, but also without whom I would not have had funding, adhered to UNC deadlines for class registration, eaten bagel breakfasts, had a working computer, and there would not have been a singer for the Gillings Jazz Band. Thank you Robin Whitely, Rhoda Cerny, Doris Dworschak, Jennifer Moore, and Jason Pruszko.

Thank you to my husband for telling me when it is too late to keep working and encouraging me to focus a little longer when it is too soon to stop.

My parents were my first advisors. Thank you to my mom, Dr. Miriam Heller, for teaching me to delight in learning, especially that which is difficult or complex to understand. Thank you to my dad, Dr. Mark Wiesner. You always knew that I would be a scientist. Thank you for introducing me to lab work and showing me how creative and cool scientific research can be.

I would also like to thank some of the other communities that have nurtured me during my time in this Ph.D. program: DurhamISH, the UNC capoeira club, and the Gurley Flynns.

I would like to thank the funding sources that have sustained me financially and contributed to my training as a scientist: The National Institute of Environmental Health Sciences (NIEHS) T32ES007018 and the NSF grant 1316318 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program.

I would also like to acknowledge a grant that provided the resources for the data collected by Marquette University researchers used in this dissertation: The Marquette University Innovation Grant (KH).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFO	animal feeding operation
ARG	antibiotic resistance gene
BoBac	bovine host-associated *Bacteroides* marker
CAFO	concentrated animal feeding operations
E. coli	*Esheriria coli*
ermB	ARG associated with macrolide resistance, specifically erythromycin
GORF	ground transport, overland and river distance with flow
HuBac	human host-associated *Bacteroides* marker
intI1	class 1 integron, integrase gene which is a mobile genetic element
ManureApp	source term for cropland where manure may be applied
MST	microbial source tracking
ORF	overland and river distance with flow
qnrA	ARG associated with quinolone resistance
qPCR	quantitative polymerase chain reaction
S. aureus	*Staphylococcus aureus*
Spa	gene that codes for a S. aureus-specific protein
sul1	ARG associated with sulfonamide resistance
tetW	ARG associated with tetracycline resistance
WWTP	wastewater treatment plant

CHAPTER 1: INTRODUCTION AND RESEARCH OBJECTIVES

1.1 Introduction

In the mid to late 20th century, large-scale livestock production proliferated with technological innovations, increased demand for meat and dairy products,[1] and to accommodate a growing population.[2] This trend in industrial-scale livestock production occurs at operations that house thousands of animals, called animal feeding operations (AFOs). AFOs of a certain size (depending on animal type) are now referred to as concentrated animal feeding operations (CAFOs) by Clean Water Act regulation.[3] Historically and currently, these operations have been linked to community and environmental health concerns.[2] Among the concerns, the co-occurrence of human populations residing in rural areas and CAFOs can result in human exposures to zoonotic diseases (i.e. diseases that can be passed from animals to humans),[4] with a particular concern about exposures to and dissemination of antibiotic resistant pathogens.[5]

Antibiotic resistant pathogens can result from the misuse or overuse of antibiotics.[6] In CAFOs, antibiotics are frequently administered to treat and prevent (i.e., the use of antibiotics at sub-therapeutic doses) disease among concentrated animals.[7] In particular, the administration of preventative antibiotics and disinfectants at subinhibitory concentrations can lead to mutations and horizontal gene transfer that allow the proliferation of antibiotic resistance.[8–11] In previous research, the operations, workers, and outdoor environments nearby (e.g. water, soil, air) have been associated with elevated levels of antimicrobial resistance (AMR).[12–16] To effectively mitigate the impacts that CAFOs have on antibiotic resistant bacteria in the environment, and subsequently reduce public health risks, the following needs to be characterized. 1) The transport of fecal contamination and AMR from all potential sources (e.g. wastewater treatment plants, land applied sludge, septic systems), 2) how contributions can be fractionated or multiplied (i.e. attenuation or amplification) to better understand the impact of sources, without these factors,[17] and 3) for various microbial responses, identification of key contributors, attenuators and amplifiers.

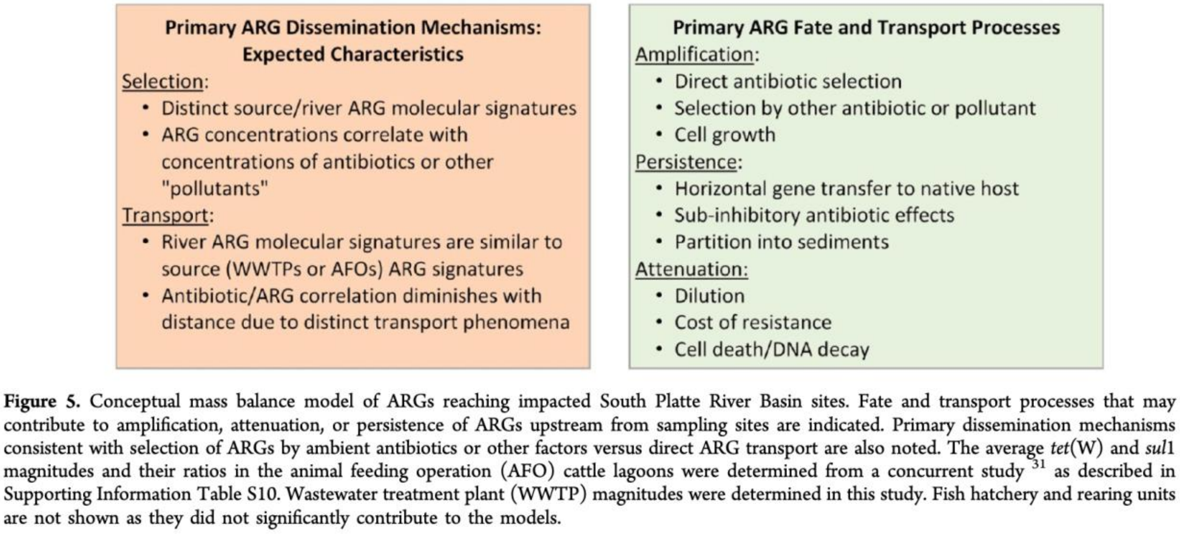
Most existing water quality models are designed with nutrient pollution in mind, though have been applied to fecal indicators. Examples of these models useful to nutrient pollution are the Soil and Water Assessment Tool (SWAT), Spatially Referenced Regression Model on Watershed Attributes (SPARROW), AQUATOX, CE-QUA-W2, and the Water Quality Analysis Simulation Program (WASP).[18] In general, there is no "one-size fits all" model and each of these tools and others offer different advantages and disadvantages.[19] Some limitations of these existing models in their application to microbial responses are that they have primarily been tested for nutrient pollution, they require a priori knowledge of delivery parameters and/or concentrations of contaminant at sources, and estimate or predict water quality at watershed-level resolutions (i.e., point-level resolution is preferred) .

For sources of microbial contamination, research has only begun to characterize microbial communities associated with sources (i.e., microbial community "fingerprints"),[20] the variability of genes encoding specific functions across different sources,[21,22] the persistence, decay, and inhibition of different organisms and genes in different environmental matrices,[23–26] as well as the contaminants that may affect microbial communities onsite and as they migrate offsite (e.g., antibiotics, antimicrobials, metals).[27] Acquiring this knowledge is made more difficult due to restrictions in sampling at some sources. In particular, in the United States, the ability to sample at CAFOs is often restricted due to the legal history of CAFOs with nuisance and other lawsuit types.[2] This limits the types of methods appropriate for estimating contributions from all CAFOs and other sources.

The field of microbial source tracking (MST) has seen recent major advancements in the ability to characterize microbial risks based on the detection of host-associated microbial targets. These targets can help to identify and quantify the extent to which a particular host (e.g., human, bovine, swine) has contributed to fecal contamination, which can in turn be used to better assess microbial risks.[28] From a regulatory standpoint, these methods alone are limited in their ability to connect host-associated targets detected in samples back to spatially distributed microbial contamination sources.

Land-use regression (LUR) approaches can be used to identify and characterize sources of pollution with source terms (i.e., predictors constructed from source locations with spatial predictor models). Recent work has used host-associated microbial sources tracking responses as the measured environmental responses. LUR has been used to describe statistical associations between sources and microbial contamination in rivers and groundwater,[17,29–38] including AMR.[17,38,39] Few of these studies modeled responses that were molecular targets.[17,29,30,37,38,40] Only two have modeled responses that are molecular targets and AMR responses.[17,38] Among these studies, there has been a greater emphasis on microbial contamination from human sources and environmental responses measured in surface water.

In this body of work, a microbial land-use regression model is developed to capture microbial transport concepts previously laid out in study by Pruden *et al.*[17] which conceptualized a mass balance model for contributions, amplification, attenuation, and persistence of ARGs (Figure 1.1). A modeling framework that implements this model will be developed by expanding and developing on previous microbial LUR components such as database selection, spatial predictor models, and predictor selection. This framework and microbial land-use regression model will be applied to molecular responses in sediment and surface water to identify sources of fecal contamination and elevated AMR and add to the knowledge of how sources differently impact sediment versus surface water in a CAFO-dense region of Wisconsin.



Primary ARG Dissemination Mechanisms: Expected Characteristics

Selection:

- Distinct source/river ARG molecular signatures
- ARG concentrations correlate with concentrations of antibiotics or other "pollutants"

Transport:

- River ARG molecular signatures are similar to source (WWTPs or AFOs) ARG signatures
- Antibiotic/ARG correlation diminishes with distance due to distinct transport phenomena

Primary ARG Fate and Transport Processes

Amplification:

- Direct antibiotic selection
- Selection by other antibiotic or pollutant
- Cell growth

Persistence:

- Horizontal gene transfer to native host
- Sub-inhibitory antibiotic effects
- Partition into sediments

Attenuation:

- Dilution
- Cost of resistance
- Cell death/DNA decay

Figure 5. Conceptual mass balance model of ARGs reaching impacted South Platte River Basin sites. Fate and transport processes that may contribute to amplification, attenuation, or persistence of ARGs upstream from sampling sites are indicated. Primary dissemination mechanisms consistent with selection of ARGs by ambient antibiotics or other factors versus direct ARG transport are also noted. The average *tet*(W) and *sul*1 magnitudes and their ratios in the animal feeding operation (AFO) cattle lagoons were determined from a concurrent study [31] as described in Supporting Information Table S10. Wastewater treatment plant (WWTP) magnitudes were determined in this study. Fish hatchery and rearing units are not shown as they did not significantly contribute to the models.

Figure 1.1 From Pruden et al. [17]

1.2 Specific Aims

1.2.1 Aim 1 Development of a physically meaningful microbial land-use regression framework and demonstration on bovine-associated Bacteroides responses in riverbed sediment

Estimate the distribution of bovine fecal contamination in riverbed sediment associated with CAFOs, with AFOs more broadly, and manure application fields by expanding upon previous microbial land-use regression approaches to develop a novel spatial modeling framework that leverages a physically meaningful land-use regression model transport, attenuation, and amplification as conceptualized by Pruden et al[17]. See Figure 1.1. This framework will also implement spatial predictor models that account for the density and proximity of upstream sources, dilution due to flow, and potential ground transportation processes (e.g., the hauling of manure). This framework is applied for demonstration on bovine-associated fecal markers measured from sediment samples in Kewaunee County, Wisconsin. This work

a) Determines which spatial databases most reliably capture the locations of AFOs and cropland areas where manure can be applied.
b) Estimates the relative impacts that bovine sources have on bovine-associated fecal contamination in riverbed sediment using three ways of modeling the contributions from these sources (Euclidean distance decay, overland and river distance decay with flow, and ground transport, overland and river distance decay with flow).
c) Using a novel spatial predictor model, determines if waste generated by AFOs is transported to proximal areas of manure application.
d) Compares the drivers of relative abundance to absolute abundance of bovine markers by expressing the effect of drivers in terms of an "abundance ratio" that parallels in definition to a risk ratio.
e) Contributes to the literature on the advantages of using different approaches for constructing spatial predictors (i.e., Euclidean, overland and river distance- flow, and ground transportation, overland and river distance- flow) of contaminants in rivers

1.2.2 Aim 2 Application and expansion of novel microbial land-use regression framework to human-associated and bovine-associated Bacteroides markers in sediment and surface water

Identify and characterize the sources most associated with human-associated fecal contamination and bovine-associated fecal contamination in sediment and surface water using the framework. In this work, the potential sources of human-associated fecal contamination are limited to septic systems, municipal wastewater treatment plants, land applied sludge from municipal waste or septage, and high-intensity developed land cover. The potential sources of bovine-associated fecal contamination are limited to AFOs, manure application fields, manure from AFOs applied to manure application fields from hauling, low-intensity developed land cover, industrial wastewater treatment plants, and land applied sludge from industrial waste and food processing. This work

a) Determines which spatial databases most reliably capture the locations of all potential sources.
b) Identifies key sources of human- and bovine-associated fecal contamination.
c) Characterizes differences in contributions to human- versus bovine-associated fecal contamination.
d) Characterizes differences in source contributions to riverbed sediment versus surface water.
e) Defines transport characteristics from hyperparameters that portray the ability of a source to pollute sediment or surface water given its distance away from a river network.
f) Contribute to the knowledge around the differences in how sources contribute differently to riverbed sediment and surface water and how bovine sources versus human sources contribute to fecal contamination.

1.2.3 Aim 3 Application and expansion of novel framework microbial land-use regression framework to a panel of antibiotic resistance genes in sediment and surface water

Identify and characterize the sources most associated with a panel of 5 ARGs measured from sediment and surface water samples using the novel framework developed in aim 1. In this work, the potential sources of elevated ARGs are septic systems, wastewater treatment plants, land applied sludge from municipal waste or septage, high-intensity developed land cover, AFOs, manure application fields, manure from AFOs applied to manure application fields from hauling, low-intensity developed land cover, industrial wastewater treatment plants, and land applied sludge from industrial waste and food processing. This work:

a) Determines which spatial databases most reliably capture the locations of all sources of ARGs

b) Identifies key sources of elevated ARGs that encode resistance to erythromycin, tetracyclines, fluoroquinolones, and sulfonamides and one gene, *intI1*, which is a mobile genetic element.

c) Defines transport characteristics from hyperparameters that portray the ability of a source to pollute sediment or surface water given its distance away from a river network.

d) Compares the identified sources of ARG pollution to surface water and sediment

e) Compares the identified sources of pollution across ARG responses.

f) Conducts interviews with Wisconsin dairy cattle veterinarians to better understand how antibiotic use at AFOs may affect levels of antimicrobial resistance.

g) Aggregates transport characteristics across ARGs to define conservative pollution scenarios for a variety of sources.

h) Contributes to the knowledge of ARG transport from sources.

CHAPTER 2: BACKGROUND

2.1 Concentrated Animal Feeding Operations, Antimicrobial Resistance, Fecal Contamination, and Public Health

2.1.1 Concentrated animal feeding operations and public health risks

Relationships have been found between CAFOs and various respiratory health outcomes, gastrointestinal illness, carriage of opportunistic, antibiotic resistant pathogens, and antibiotic resistant infections in different geographical regions.[4,41] Respiratory health outcomes associated with CAFOs are asthma,[42] COPD, [43] decreases in lung function measured by FEV1,[44] pneumonia, atopic eczema, gastro-intestinal illnesses and unspecified infectious diseases,[45] upper respiratory tract infections, such as bronchitis,[46] and peaks in influenza-like-illnesses around the time of the H1N1 outbreak,[47] Additionally, CAFOs have been associated with antibiotic resistance *S. aureus* carriage in community referent groups from nearby residences.[13] In a review across 25 studies, populations exposed to livestock were 7.03 times more associated with methicillin resistant *S. aureus* carriage, which is an opportunistic pathogen (i.e., infection can occur in immunocompromised individuals). This association was strongest for swine livestock compared to poultry or cattle.[41] Links existing between gastrointestinal illnesses and livestock-associated pathogens have been established and manure application migrating offsite into waters used for agriculture/irrigation has been identified as one possible source. However, pathways have not been defined well enough to characterize risks.[48,49] The combined knowledge of these established and putative health risks associated with fecal contamination from CAFOs, as well as the links to AMR, make it vital to better characterize environmental transport of AMR and fecal contamination, especially in aquatic environments.[50–52]

2.1.2 Defining antimicrobial resistance

To characterize the public health concerns around antimicrobial resistance, it is necessary to understand what antimicrobial resistance is and how it occurs. Antimicrobial resistance describes when bacteria, viruses, fungi, or parasites develop the ability to survive in the presence of antimicrobials (e.g. antibiotics, metals, disinfectants) which are designed to destroy or inhibit those microorganisms.[6] Antibiotic resistance is when microorganism's express resistance to clinical doses of antibiotics, where genes encoding this expression can be intrinsic to the organism or acquired through horizontal gene transfer. Genes which are intrinsic to microorganisms can mutate under stress, leading to genes that can encode for new types of resistance.[53] When one antibiotic resistant bacterium and many susceptible bacteria are exposed to antibiotics, the antibiotic resistant bacterium can continue to grow, while susceptible bacteria are inhibited, and the system can become dominated by the antibiotic resistant variety. Additionally, due to antimicrobials utilizing similar mechanisms for destroying or inhibiting microorganisms, one specific antimicrobial might select for bacteria resistant to a different antimicrobial.[6] Due to these mechanisms of antimicrobial resistance, the availability of effective antibiotics for treating infections in humans and animals has increasingly diminished since the discovery of the first antibiotic, penicillin.[54]

The relationship between the natural environment and the prevalence of antibiotic resistance in humans and animals has been frequently highlighted.[5,50,51] For example, Figure 2.1 from Berendonk et al. described this connectivity graphically and identified approaches to fill gaps of knowledge from 2015.[52] Since then some work has been done to characterize resistance for environmental strains. [12,55] Additionally, epidemiological studies have begun to understand relationships between environmental exposures and human carriage of antibiotic resistance strains of bacteria. [12,13,41,56]

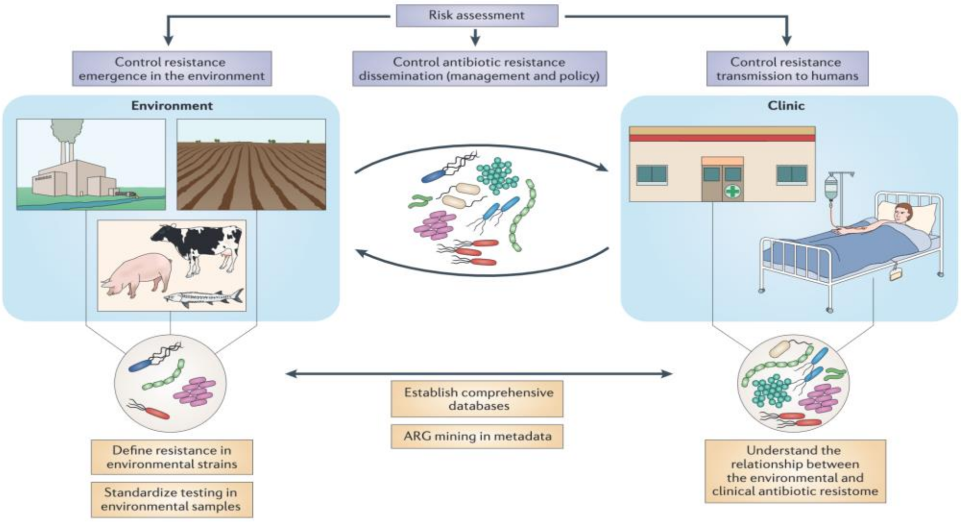


Figure 2.1 Diagram which presents the connectivity between controlling the emergence of AMR in the environment and the transmission to humans from Berendonk et al. [52]

2.1.3 Fecal contamination in the environment and public health

While the environmental transport of antimicrobial resistance has not been well-characterized, research in the field of microbial source tracking has frequently studied fecal contamination to identify sources, describe transport, and mitigate potential health risks.[28] Fecal contamination of the outdoor environment occurs when fecal matter disseminates from hosts (e.g., humans, swine, dogs, seagulls) into the water, soil, or air. Fecal contamination in the environment is often measured by culturing fecal organisms or using molecular approaches to identify phenotypic properties or nucleic acids that have been previously associated with fecal matter.[28] There is a great deal of evidence indicating the correlation of fecal contamination, waterborne illnesses (e.g., gastrointestinal illnesses, respiratory illnesses, and skin, eye and ear infections), and pathogens (e.g., Enteroviruses, *Campylobacter*, *S. aureus*, *Cryptosporidium*) present in surface water. See Figure 1.2 for some of the correlations (positive or negative) found between MST markers and various types of waterborne illness from a review of these relationships from 2014.[57] Additional research has shown that riverbed sediment resuspension can affect the types and quantities of pathogens present in surface water associated with health risks, which could extend the association of health risks to fecal contamination in sediment.[58] Compared to surface water, fewer studies have sought to understand the microbial ecology of sediments and related health risks.

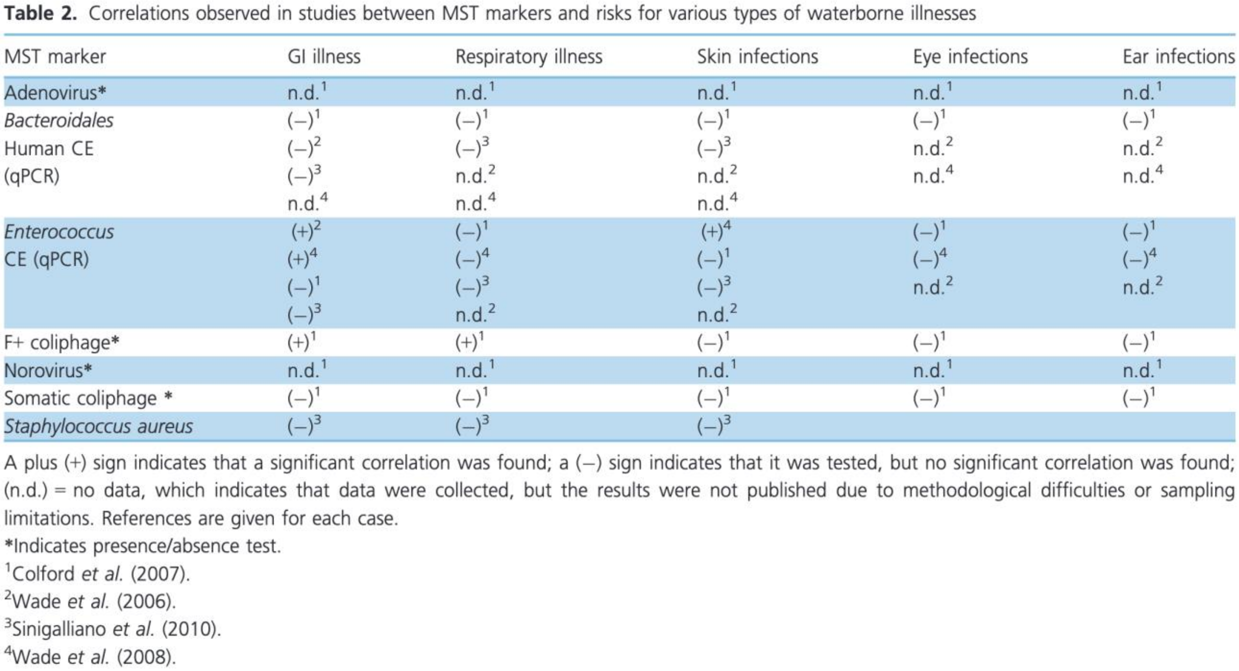


Table 2. Correlations observed in studies between MST markers and risks for various types of waterborne illnesses

MST marker	GI illness	Respiratory illness	Skin infections	Eye infections	Ear infections
Adenovirus*	n.d.[1]	n.d.[1]	n.d.[1]	n.d.[1]	n.d.[1]
Bacteroidales Human CE (qPCR)	(–)[1] (–)[2] (–)[3] n.d.[4]	(–)[1] (–)[3] n.d.[2] n.d.[4]	(–)[1] (–)[3] n.d.[2] n.d.[4]	(–)[1] n.d.[2] n.d.[4]	(–)[1] n.d.[2] n.d.[4]
Enterococcus CE (qPCR)	(+)[2] (+)[4] (–)[1] (–)[3]	(–)[1] (–)[4] (–)[3] n.d.[2]	(+)[4] (–)[1] (–)[3] n.d.[2]	(–)[1] (–)[4] n.d.[2]	(–)[1] (–)[4] n.d.[2]
F+ coliphage*	(+)[1]	(+)[1]	(–)[1]	(–)[1]	(–)[1]
Norovirus*	n.d.[1]	n.d.[1]	n.d.[1]	n.d.[1]	n.d.[1]
Somatic coliphage *	(–)[1]	(–)[1]	(–)[1]	(–)[1]	(–)[1]
Staphylococcus aureus	(–)[3]	(–)[3]	(–)[3]		

A plus (+) sign indicates that a significant correlation was found; a (–) sign indicates that it was tested, but no significant correlation was found; (n.d.) = no data, which indicates that data were collected, but the results were not published due to methodological difficulties or sampling limitations. References are given for each case.

*Indicates presence/absence test.

[1] Colford *et al.* (2007).

[2] Wade *et al.* (2006).

[3] Sinigalliano *et al.* (2010).

[4] Wade *et al.* (2008).

Figure 1.2 Examples of MST markers and example of associations with waterborne illnesses. see Harwood et al. [57]

Given the established and putative relationships between fecal contamination and public health risks, one element that has been suggested to greatly contribute to the field of quantitative microbial assessment (i.e., field dedicated to estimating public health risks associated with exposure to microorganisms in the environment) is understanding of the fate and transport of fecal markers and elevated AMR in the environment.[59]

2.1.4 Sources of fecal contamination and antimicrobial resistance

Fecal contamination, antibiotic resistant bacteria, genes associated with antibiotic resistance (ARGs), and agents (i.e. antibiotics, metals, biocides, microplastics) that co-select for antimicrobial resistance (AMR) [51,60–62] can be disseminated from anthropogenic sources into soil, waterbodies, sediments, the air, and onto food (i.e., from these environments or directly connected to food production) where people can be exposed. Dissemination of pollution from anthropogenic sources may exert selective pressures on microbial communities in the natural environment, which can select for AMR.[63] Furthermore, these ARGs of natural or anthropogenic origin can be acquired and forfeited by bacteria through horizontal gene transfer which can further their spread leading to environmental, human, and animal health consequences.[64] Most primary sources of fecal contamination also contribute to antimicrobial resistance. Animal feeding operations (AFOs), wastewater treatment plants (WWTPs), septic systems, and land applied sludge, have all been suggested as potential sources of fecal contamination and AMR, as will be described further below. Additional sources might exist for AMR due to aforementioned co-selective properties of other kinds of pollutants.[51]

- Fecal contamination and AMR from livestock has been found to be emitted into environmental systems, such as the air [65], from barns, other confined areas, and pits of animal waste like manure storages, lagoons, anaerobic lagoons or poultry litter.[12,39,66–76]
- Wastewater treatment plants (WWTPs) contribute to microbial contamination and AMR in the water[57,77] and air.[55] Hospital and pharmaceutical WWTPs have been associated with higher impacts on AMR compared to municipal plants. [55,78–80]
- Septic systems have been correlated with fecal contamination in surface and groundwater.[37,81,82]. AMR has been studied less frequently with respect to these systems

compared to WWTPs, which are also a source of fecal contamination and AMR, but septic systems have been associated with lesser reductions in ARGs compared to municipal WWTPs.[83]

- Other populations that are reported to use more antibiotics than average is the elderly, the sick, and the dying. Retirement homes, which are closed environments, often with poor ventilation, filtration, and removal of circulated air, can propagate the spread of microorganisms to vulnerable populations. Populations older than 65 are more susceptible to infections due to weakened immune systems from aging and treatments for other late-in-life diseases such as cancer and are more likely to die from infections and complications. Because of these reasons, prescription of antibiotics to geriatric patients is highly prevalent.[84] In the case of one retirement home in France, the effluent corresponded with more strains of antibiotic resistant *E. coli* compared to a hospital, albeit the hospital could not accommodate as many people.[78]
- In a study done in South Africa, phenotypically resistant *E. coli* were isolated from samples collected at two cemeteries. The authors noted that the corpses themselves might contribute to this resistance, but heavy metals from degraded coffin materials could induce co-selection of ARGs. [85,86] To our knowledge, this is the only study of the potential impact of cemeteries on AMR but provides evidence that some anthropogenic sources of AMR are not well-established or may remain unknown.

Challenges exist in finding databases representing these spatially distributed sources. For example, septic system locations are not always available in the United States.[87] However, they might be approximated by centroids of residential parcels that are distant from municipal wastewater treatment plants or sewer lines. In other cases, two or more databases might exist which describe the same spatial feature, especially for those visible through aerial imagery. For instance, sometimes geocoders are used to process addresses of sources into spatial coordinates representing a source node. Geocoders may produce geographic coordinates for a true location or center point of a source with varying locational accuracy. Locational accuracy of these geocoders can vary based on a number of factors. For example, the accuracy may depend on whether the geography is rural or urban.[88] A criterion is needed to help reporting on the selection between databases. A criterion can be helpful not only to finding which databases are most reliable in representing sources, but to help with reporting why a given database was selected, more broadly.

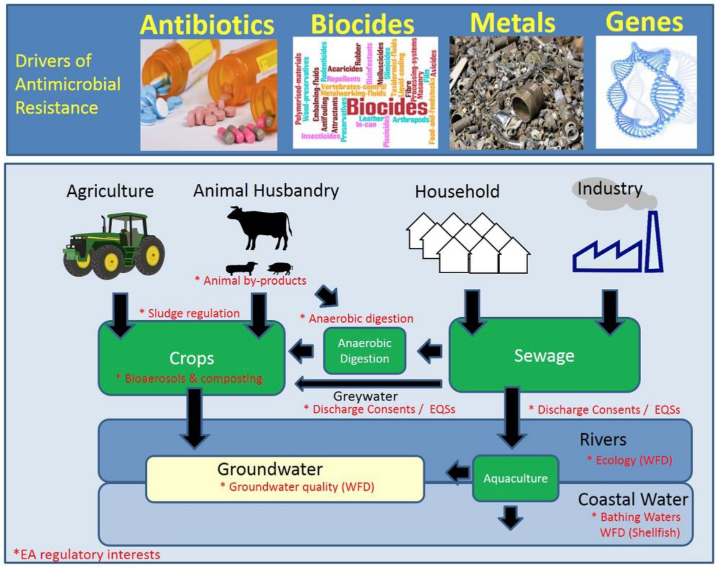


FIGURE 1 | Schematic of the hot-spots and drivers of antimicrobial resistance (AMR). The environmental compartments that are currently monitored or regulated by the Environment Agency (EA; England) are denoted by an asterisk in red. WFD, Water Framework Directive.

Figure 2.3 Schematic from Singer et al. [51]

2.2 Identifying Gaps in Current Approaches for Modeling Contributions to and Modifying Factors of Microbial Contamination and Antimicrobial Resistance

2.2.1 Contributions from sources

Current approaches to identifying sources of microbial contamination fall under two general approaches: microbial source tracking methods that quantify host-associated (e.g., swine, bovine, human) fecal markers, and spatial-statistical methods that rely on databases of spatially distributed sources to construct spatial predictors. More recently, researchers have combined these approaches to understand the extent to which the fecal responses relate to spatially distributed sources.[30,82,87,89,90] Host-associated fecal markers have typically been utilized to identify host origins of microbial contamination in MST. These host-specific and host-associated approaches are desirable over the use of fecal indicator bacteria, such as *E. coli*, which are not particular to any host, in turn they provide more information about the source.[28] Spatial modeling approaches using predictors derived from spatially distributed sources can be used in conjunction with host-associated methods to establish linkages with source locations. In parallel to the specificity of response, as the resolution of spatial predictors derived from spatially distributed sources increases (e.g., point level has greater resolution than watershed level), linkages are better established with source locations. To establish linkages with source locations, spatial predictor models should ideally depict contributions from sources at the finest resolution possible.

2.2.2 Literature review of spatial predictor models used in land-use regression for microbial responses

Various models have been described in the literature to construct spatial predictors of microbial responses. Let y_i be the microbial response (for example a microbial concentration) for some sample collected at location i. We list below three spatial models that have been used to obtain a predictor value for y_i. For the first two models, we denote the predictor value as x_i, as is the usual convention for generic predictors. For the last model we denote the predictor as $s_i(\alpha,..)$ to emphasize that the focus is on capturing the effect of sources of pollution (as opposed to the effect of sources and modifiers) based on a set of model hyper-parameter values $\alpha,\ldots$

2.2.2.1 Lumped

In the lumped model, the predictor value x_i for the i^{th} sampling location is equal to the sum of the sources or sum of target land-cover areas, W_i, divided by the total area, A_i[91]

$$x_i = \frac{W_i}{A_i} \qquad \text{(Eq. 2.1)}$$

For example, A_i can be an area surrounding the sampling location i, and W_i can be the sum of animals in that area, or the sum of areas classified as having an “agricultural” land-cover.

2.2.2.2 Inverse-distance-weighted interpolation

Pruden et al. (2012)[17] used a predictor calculated as the inverse distance weighted (IDW) interpolation of pollution capacities C_{0j} upstream of the sampling location

$$x_i = \frac{\sum_{j=1}^{N} C_{0j}/D_{ij} \times \delta_{\overleftarrow{ij}}}{\sum_{j=1}^{N} 1/D_{ij} \times \delta_{\overleftarrow{ij}}} \qquad \text{(Eq. 2.2)}$$

where N is the number of pollution sources surrounding sampling location i, D_{ij} is the distance between sampling location i and source location j, and $\delta_{\overleftarrow{ij}}$ is a hydrologic connectivity indicator equal to 1 if sampling location i is upstream of source location j, and 0 otherwise. For example, the source locations could be animal farms, and C_{0j} could be the number of animals in the j-th farm.

This spatial predictor is an interpolation of pollution capacities upstream of the sampling point. Therefore, it does not guarantee that pollution decreases away from sources, as would be expected from dilution and degradation processes. A good example of this is shown in Figure 2.4. There are three large farms, with 500 animals each, located on the upstream end of a river reach, followed by a small farm with only 2 animals located further downstream, as depicted by crosses and their corresponding C_{0j} values in Figure 2.4. The IWD interpolated predictor value x_i at two sampling points downstream of the small farm are shown to have values of 111 and 208, respectively. This means that the predictor increases as we move downstream the river and away from the pollution sources. This increase in pollution goes counter to the physical principals of dilution and degradation, and therefore lacks physical meaningfulness.

Figure 2.4 River reach flowing from a (left) to b (right) showing the location of four pollution sources (X) and two sampling points (+). The pollution sources consist of three large farms with 500 animals each on the upstream end, and a small farm with 2 animals on the downstream end. The IDW interpolation predictor calculated at two sampling points downstream of the small farm have values of x_1=111 and x_2=208. These values increase moving downstream and away from the sources, which lacks physical meaningfulness.

2.2.2.3 Sum of decaying contributions

In recent works (one paper and one dissertation), the predictor of microbial contamination at a sampling point has been modeled as the sum of decaying contributions from upstream sources.[38,87] Amos et al. modeled antibiotic resistance gene prevalence using a gravity model applied to upstream sources. In that approach the predictor is given by

$$s_i(\alpha, d_0) = \sum_{j=1}^{N} \frac{C_{0j}}{(D_{ij}+d_0)^{\alpha}} \times \delta_{\overline{ij}} \, , \qquad \text{(Eq. 2.3)}$$

where N, D_{ij} and $\delta_{\overline{ij}}$ are the same as in Eq. 2.2, and α and d_0 are hyperparameters that change the predictor behavior. Without d_0, the predictor at the source is infinity, and therefore Amos et al. (2015) use d_0=1. This model is different from the IDW interpolation in that it accounts for the density of sources, and it is such that the predictor value decreases away from sources, which is physically meaningful.

An improvement on this approach is that used by Christenson[87] in her dissertation, which consists in calculating the predictor as the sum of exponentially decaying contributions from upstream sources using the following equation

$$s_i(\alpha_O, \alpha_R) = \sum_{j=1}^{N} C_{0j} \exp\left(\frac{-3D_{ij}{}^{(O)}}{\alpha_O} + \frac{-3D_{ij}{}^{(R)}}{\alpha_R}\right) \times \delta_{\overline{ij}} \qquad \text{(Eq. 2.4)}$$

where N, D_{ij} and $\delta_{\overline{ij}}$are the same as in Eq. 2.2 and Eq. 2.3 except that sources can now be located on land, $D_{ij}^{(O)}$ is the overland distance from source j to where the contaminant enters the river, $D_{ij}^{(R)}$ is the distance traveled along river from where contamination enters the river to the sampling point, the hyperparameters α_O represents the overland distance over which microbial contamination decays by 95%, and α_R represents the river distance over which microbial contamination decays by 95%. This model has the advantage that it explicitly quantifies the rate at which contamination decays across land and along rivers.

2.2.3 Review of performance evaluation for existing models of microbial contamination

Table 2.1 summarizes information about 17 total models from 11 articles[17,29–38] and 1 dissertation[87] that were used to predict microbial contamination from spatial features. The first column describes the types of sources that were interesting to the researchers, which have been primarily sources of human or animal waste. The second column describes the microbial responses that were observed and the unit of measurement (e.g., Log *E. coli* colony forming units/100mL of surface water). Some have two responses per source type and are distinguished in other columns by the table as headers in italics. The column called, "Resolution of Source Terms", describes whether the spatial variables are lumped, interpolative, or depict gradients. In this column a general description is included of how they are lumped, or the spatial predictor model used for gradient-depicting terms. The description of the source terms is defined in the next column along with greater detail about the spatial predictor model for gradient-depicting source terms. If there are multiple variables of the same source type, they are distinguished by letterings, (a), (b)…. This system continues across the other columns of the table. The next column indicates the spatial scale defined by hyperparameters. These could be a selected radius that defines a path upstream or a circle around the sampling or could be specific values describing a distance decay range overland or downstream. The following column indicates whether the source terms were standardized, which would allow for effect comparisons across studies. The regression coefficient(s) are provided in the seventh column. Occasionally, only the direction of correlation was provided, but not the magnitude. The following column gives the sample size, *n*, the number of total predictors used in the regression model, *k*, the degrees of freedom (i.e., *n*-*k*-1), *df,* and the number of the *k* predictors that have a positive slope, *j*. The number of unique locations sampled in the studies and used in the modeling, *s*, are provided in the following column. Among the *k* predictors, some utilize laboratory or sampling measurements (e.g., human or swine fecal markers, water temperature, or nutrients). The third to last column indicates the number of in-sample predictors, *p*, that are derived from these types of laboratory or sampling measurements. In order to estimate gradients of pollution using models with $p>0$, then these in-sample predictors would need to be estimated themselves and would require nested modeling approaches. The second to last column indicates the model fit expressed as adjusted R-squared, or R-squared if the adjusted value is not available. Some papers did not report their model fit as R-squared, adjusted R-squared, AIC, or using any other model fit measure. In these instances, the column indicates "No Measure of Model Fit Provided". The last column indicates the author and year of publication corresponding to the model(s) for that row.

The model fit information was provided in their main paper or supplementary information for only twelve of the models from ten different authors. [17,29,31–33,35–38,87] Among those, the microbial response was a concentration of fecal indicators for six models. [31–33,35,36,87] a concentration of host-specific fecal markers for 4 models,[29,37] and a relative abundance of antibiotic resistance genes for 2 models.[17,38] The resolution of the source terms was lumped for 8 models,[29,31–33,35,37] could depict gradients of pollution for 3 models,[36,38,87] and used a downstream interpolative approach for 1 model.[17] Across the models explaining fecal indicators and host-associated fecal contamination (i.e. excluding studies of antimicrobial resistance), the average adjusted R-squared was 0.483.[29–37] This average was calculated by using adjusted R-squared when available and otherwise using R-squared, when available. Among these 10 models, one study included 3 predictors ($p=3$) that used in-sample information as predictors.[87] Additionally, there were some models that did not distinguish sources from modifiers.[29,32,33,37] This means that there were source terms with negative regression coefficients corresponding to reductions. Other models benefitted from in-sample measurement predictors and predictors with negative coefficients, k-j, would also contribute to the model fit.

Across models, the average model fit is approximately the same for different numbers of unique sampling locations. The adjusted R-squared in the two models with the greatest number of unique sampling locations (s=395 and s=327) are 0.504 and 0.435.[32,33] In the two models where there are the least number of unique sampling locations ($s=7$ and $s=12$) the adjusted R-squared was 0.552 and 0.510.[36,37] Across the 12 models of microbial contamination reporting model fit, the four studies[17,35,36,38] with the lowest degrees of freedom (df=38, 18, 10, and 10) were the only studies to explain greater than 60% of the variability (Adj. R^2=0.62, 0.816, 0.90, 0.829) using their full models. The degrees of freedom across the other regression models were greater than 68.

Table 2.1 A review of spatial predictors representative of source terms used in regression models to explain microbial responses

Source Type	**Microbial Response**	**Resolution of Source Terms**	**Source Terms**	**Hyperparameters Defining Spatial Scale**	**Were Independent Variables Standardized?**	**Regression Coefficient (Effect, $\boldsymbol{\beta}^{(u)}$)**	**Sample Size (n)** **Predictors (k)** **Degrees of Freedom (*df*)** **Predictors with Positive Slope (*j*)**	**Number of Unique Spatially Distributed Sampling Locations Modeled (*s*)**	**Number Additional Measurements from Samples Used as Predictors (*p*)**	**Model Fit**	**Citation**
Animal Agriculture, Animal Feeding Operations, Animal Manure	**Ruminant Microbial Source Tracking Marker** (Rum2Bac copies/ mL surface water)	Lumped at radius-level	(a) Forest	Radius of 1 km	Unknown	(a) 1.48	n=114 k=3 df=110 j=2	s=14	p=0	0.330 (R^2)	McKee et al. 2020 [29]
	Fecal Indicator Bacteria (Log *E. coli* CFUs/100mL of surface water)	Gradient depiction with the Sum of Exponentially Decaying Contributions (SEDC)	(a)SEDC from swine lagoons	(a)750m overland and 3.05 km downstream	Yes (Expressed as Interquartile Range Ratio without interquartile range provided)	(a)1.52	n=177 k=7 df=169 j=7	s=24	p=3 (Water temperature, Human Fecal Marker, Swine Fecal Marker)	0.411 (Adj. R^2)	Christenson-Diver 2019 [87] *Dissertation, not yet published
	Absolute Abundance of Ruminant *Bacteroides* (Log10 Copies of *BacR* /100mL surface water)	Lumped at watershed-level	Cattle Density	Depends on spatial scale of watershed	Yes	0.08	n=219 k=5 df=213 j=5	s=8	p=0	No Measure of Model Fit Provided	Dila et al. 2018 [30]
	Fecal Indicator Bacteria (*E. coli* MPN/100mL in tropical surface waters)	Lumped with land-use at and around sampling sites	Categorical variables describing the dominant agricultural land-use at the sampling site and “around” the sampling site, potentially adjacent.	Depends on spatial scale of parcels	Unknown	0.039	n=171 k=2 df=167 j=1	s=7	p=0	0.552 (R^2)	Saeidi et al. 2018 [31]

	Fecal Coliforms (Log Colilert MPN/100mL of surface water)	Lumped at county-level	Swine lagoon density	Depends on spatial scale of watershed	Unknown, but mean and standard deviation of predictors are provided	0.614 or 3.07 if dividing by standard deviation to standardize	*n*=327 *k*=18 *df*=308 *j*=7	*s*=327	*p*=0	0.435 (Adj. R^2)	Vitro et al. 2017 [32]
	Fecal Indicator Bacteria (*E. coli* in CFU/100 mL surface water)	Lumped at watershed-level	(a)Feedlot Density (b)# AFO Confinements (c)# AFO Feedlots	Depends on spatial scale of watershed	No	(a)21,289.786 (b)41.705236 (c)–224.3185	*n*=395 *k*=10 *df*=384 *j*=4	*s*=395	*p*=0	0.504 (Adj. R^2)	Brendel and Soupir 2017 [33]
	Fecal Indicator Bacteria in Domestic Wells (*Enterococcus* in log CFU/100 mL groundwater)	Lumped at radius-level	Dairy lagoon or corral located within radius to well	Radius of 2.4 km	Binary (does not provide how many locations within 2.4 km)	1.9	*n*=24 *k*=5 *df*=18 *j*=3	*s*=24	*p*=1 (Potassium)	No Measure of Model Fit Provided	Li et al. 2015 [34]
	Fecal Indicator Bacteria in Groundwater (*E. coli* in log CFU/100 mL groundwater AND *Enterococcus* in log CFU/100 mL groundwater)	Lumped at radius-level	(a) Manure field located within radius to well (b)Manure lagoon located within radius to well	Radius of 2.4 km	Binary (does not provide how many locations within 2.4 km for both)	*E. coli* (a)–2.824 (b)–2.008 *Enterococcus* (a)–1.239 (b)1.836	*n*=214 *k=7* *df*=206 *j*=3 (*E. coli*) *j*=2 (*Enterococcus*)	*s*=190	*p*=0	No Measure of Model Fit Provided	Li et al. 2015 [34]
	Relative Abundance of Antibiotic Resistance Genes (*sul1* copies per gram sediment/copies 16S rRNA)	Interpolative along downstream river distance	Inverse-distance weighted interpolation of upstream AFOs	Undefined spatial scale	No	Strong Positive	*n*=40 *k*=1 *df*=38 *j*=1	*s*=10	*p*=0	0.62 (Adj. R^2) *Total of 40 models were attempted to explain *sul1*, no Bonferroni correction applied and adj. R^2 between 0.08 and 0.62 (average of 0.466)	Pruden et al. 2012 [17]

	Fecal Indicator Bacteria (Presumptive *E. coli* CFU/100 mL surface water)	Lumped at radius-level	Percent Land Cover that is Improved Pasture (Pasture heavily used for animal grazing)	Within band of 2 km during base-flow conditions and within band of 5 km during high-flow conditions	Unknown	Positive	$n=20$ $k=2$* Stratified Dependent Variables $df=18$ $j=2$	$s=20$	$p=0$	0.816 (Adj. R^2)	Crowther et al. 2003 [35]
	Fecal Coliforms (Colonies per 100 mLs of surface water)	Gradient Depiction using the Spatially Explicit Delivery Model (SEDMOD)	Loading rate of pollutant from livestock using locational information from interviews; modeling transport based on land cover, soil moisture, and elevation	Depends on grid-cell size	Unknown	0.25	$n=12$ $k=1$ $df=10$ $j=1$	$s=12$	$p=0$	0.51 (R^2) Full model including turbidity and temperature variables is 0.90 (Adj. R^2)	Fraser et al. 1998 [36]
Human Contributions (e.g., Wastewater treatment plants, septic systems)	**Human Microbial Source Tracking Marker** (*HF183* copies/ mL surface water) **Dog Microbial Source Tracking Marker** (*DogBact* copies/ mL surface water)	Lumped at radius-level or watershed-level	*HF183* (a) WWTP (b) Septic *DogBact* (a)Developed	*HF183* (a)catchment (b)Radius of 1km *DogBact* (a)Radius of 2km	Unknown	*HF183* (a)8.86 (b)-0.007 *DogBact* (a)1.14	$n=114$ $k=4$ $df=109$ $j=1$ $n=114$ $k=3$ $df=110$ $j=1$	$s=14$	$p=0$	HF183 0.549 (R^2) DogBact 0.256 (R^2)	McKee et al. 2020 [29]
	Fecal Indicator Bacteria (Log *E. coli* CFUs/100mL of surface water)	Fine-spatial scale (Contributions from swine lagoons)	Sum of exponentially decaying contributions from human septic	500 m overland and 2.40 km downstream	Yes (Expressed as Interquartile Range Ratio without interquartile range provided)	1.52	$n=177$ $k=7$ $df=169$ $j=7$	$s=24$	$p=3$ (Water temperature, Human Fecal Marker, Swine Fecal Marker)	0.411 (Adj. R^2)	Christenson-Diver 2019 [87] *Dissertation, not yet published

	Absolute Abundance of Human *Lachnospiracaeae* (log 10 copies of Lachno2 per 100 mL surface water) **Absolute Abundance of Human *Bacteroides*** (log 10 copies HB per 100 mL surface water)	Lumped at watershed-level	Population Density	Depends on spatial scale of watershed	Yes	*Lachno2* 0.22 *HB* 0.20	n=219 k=5 df=213 j=5 (*Lachno2*) j=5 (HB)	s=8	p=0	No Measure of Model Fit Provided	Dila et al. 2018 [30]
	Fecal Coliforms (Log Colilert MPN/100mL of surface water)	lumped at county-level	(a)Population density (b)Housing unit density (c)Road network density	Depends on spatial scale of county	Unknown, but mean and standard deviation of predictors are provided	(a)0.0078 or 0.91e-5 if dividing by standard deviation to standardize (b)–0.0243 or -6.88e-4 if dividing by standard deviation to standardize (c)0.139 or 0.0496 if dividing by standard deviation to standardize	n=327 k=18 df=308 j=7	s=327	p=0	0.435 (Adj. R^2)	Vitro et al. 2017 [32]
	Absolute Abundance of Human Bacteroides (log 10 copies of *HF183* per 100mL surface water)	lumped at watershed-level	(a)Septic Density (b)Average distance of septic to stream (c)Impervious Cover (%)	Depends on spatial scale of watershed	No	(a)0.002 (b)–0.003 (c)–0.025	n=72 k=3 df=68 $j=1$	s=24	p=0	0.47 (Adj. R^2)	Sowah et al. 2017 [37]
	Prevalence of Antibiotic	Sum of Decaying	Total Impact of WWTPs	The equation used in this paper was an	No	0.321	n=37	s=13	p=0	0.829 (Adj. R^2)	Amos et al. 2015 [38]

	Resistance Genes (log *intI1* prevalence in surface water)	Contributions Downstream	using the exponentially decaying contributions from wastewater treatment plants from experimental data	inverse distance weighted gravity approach (idwG) $s_i^{(u)idwG}(\alpha, d_0) =$ $s_i^{(u)idwG}(0.3875,1)$ To obtain the spatial scale, we estimate, a_r^*, which is the distance at which a 95% reduction in contributions would occur away from the source, using Eq. S3 and find the spatial scale defined at 2.28 km.		*k*=26 *df*=10 *j=14** *interaction terms were used*

2.2.4 Gaps in modeling non-human sources of microbial contamination

To our knowledge, ten studies focused on modeling non-human sources of microbial contamination (see Table 2.2 for summary and citations). Eight of these studies examined responses in surface water, one looked at groundwater, and another at riverbed sediments. Due to our interest in non-human sources, studies which focus on host-specific responses are more helpful in establishing a connection between non-human sources and microbial contamination. Only eight of the ten studies utilized host-specific or host-associated responses.

For strengthening the relationship between sources and microbial contamination, a clear definition of a source term that contributes to pollution is distinct from factors that might amplify or attenuate pollution by using databases of well-studied sources and that contribute to microbial pollution when modeled as a source term in a regression (i.e., positive slope from regression coefficient). Seven of the ten studies fulfilled this criterion, indicating that this is a minor gap that exists in the modeling approaches. Many of these studies have also incorporated the density of sources at various spatial extents (i.e., watershed, county, parcel, within a radius around, within a radius upstream). However, to strengthen causative relationships between sources and microbial contamination, it is important that source terms not only account for the density of sources, but the gradual effect of proximity to sources, or that the source term should decrease away sources, without the effect of other sources. Only eight out of ten studies accounted for this criterion. There is some room to build upon how six of the ten studies accounted for connectivity of the river by constructing source terms that only draw from upstream source locations. Only two studies constructed source terms with a spatial predictor model that uses flow in order to account for dilution of contaminants coming from the sources in question.

Table 2.2 Summary of gaps in modeling animal sources of microbial contamination in sediment. Boxes with grey indicate a “No” answer to the column’s question.

Citation	Microbial Response	Host-specific?	In Sediment?	Source term delineated from modifier through definition of source and positive slope of regression?	Account for gradual effect of proximity to sources?	Account for density of sources?	Account for upstream-ness of sources?	Account for dilution due to flow?	Degrees of Freedom (*df*)
McKee et al. 2020 [29]	**Ruminant Microbial Source Tracking Marker**	Ruminant	Surface water	Forest land cover term has positive slope, but forest Is not specific	Only within 1 km of sampling sites	Yes- lumped within 1 km radius of sampling sites	No	No	110
Christenson-Diver 2019 [87] *Dissertation, not yet published	**Fecal Indicator Bacteria**	*E. coli*	Surface water	Swine lagoon centroids, positive slope from regression	Using SEDC	Using SEDC	Yes	No	169
Dila et al. 2018 [30]	**Absolute Abundance of Ruminant *Bacteroides***	Ruminant	Surface water	Cattle, positive slope from regression	Only within a given watershed	Yes- lumped at the watershed level	Only to the extent that a whole watershed making contributions	No	213
Saeidi et al. 2018 [31]	**Fecal Indicator Bacteria**	*E. coli*	Surface water	Dominant agricultural land-use, positive slope from regression	No	No	No	No	171
Vitro et al. 2017 [32]	**Fecal Indicator Bacteria**	Coliforms	Surface water	Swine lagoons, positive slope from regression	No	Yes, but depending on the size of the county	No	No	308
Brendel and Soupir 2017 [33]	**Fecal Indicator Bacteria**	*E. coli*	Surface water	AFOs and Feedlots, some have positive slope from regression, other have negative	No	Yes, but depends on the watershed area	No	No	384
Li et al. 2015 [34]	**Fecal Indicator Bacteria**	*Enterococcus and E. coli*	Groundwater	Manure fields and manure lagoons, negative	Only within 2.4 km radius of	No	No	No	206

Citation	Microbial Response	Host-specific?	In Sediment?	Source term delineated from modifier through definition of source and positive slope of regression?	Account for gradual effect of proximity to sources?	Account for density of sources?	Account for upstream-ness of sources?	Account for dilution due to flow?	Degrees of Freedom (*df*)
				slopes from regression	sampling sites				
Pruden et al. 2012 [17]	**Relative Abundance of Antibiotic Resistance Genes**	*sul1*	Yes	AFOs and positive regression coefficient	No. See section on inverse-distance-weighted interpolation for explanation.	No. See section on inverse-distance-weighted interpolation for explanation.	Yes	No	38
Crowther et al. 2003 [35]	**Fecal Indicator Bacteria**	*E. coli*	Surface water	Percent land cover that is improved pasture (heavily used for animal grazing), positive regression coefficient	Only within a given band upstream	Yes	Yes, but only within a given band upstream.	Yes, but only high-flow versus low-flow conditions	18
Fraser et al. 1998 [36]	**Fecal Indicator Bacteria**	Coliforms	Surface water	Loading rate of pollutant from livestock, positive slope from regression	Yes	Yes	Yes	Yes	10

In summary, current methods used to construct source terms in order to estimate microbial responses vary in complexity. Lumped predictors (i.e. count of animals or CAFOs within a watershed or county) have been most frequently used.[30,82,92] More novel approaches include search radius around a sampling location defining a circle, where within this circle's area, the spatially distributed source (land cover, land-use) values are summed,[70] the sum of upstream capacities or average upstream capacities weighted by distance using an interpolation method,[17] the sum of exponentially decaying downstream contributions along an river distance can be used to construct source terms using parameters/hyperparameters that characterizes downstream decay, which is key in describing the transport of microbial contaminants.[38,87]

Amos *et al.* took a more experimental approach to characterizing downstream decay of *intI1* based off of the analysis of wastewater, effluent, and samples taken along a downstream gradient from the facilities. Applying this approach to construct spatial predictors for multiple sources would require a great number of financial resources and long-term dedication to sampling a particular study area. Christenson, in her dissertation, provides an example of a different approach where several different decay parameter values for each of the studied sources are assumed, which produces several options for a given source term to select from. The source term that is selected is one which corresponds to the best model fit for the linear regression. This approach originated from methods used in land-use regression framework applied to nitrate.[93] In these methods, a higher R-squared or a lower p-value was used to determine the fit of the model or predictive ability of the source term. The consequence of such a method is that the decay parameter value that corresponding with the best fit might produce a negative coefficient. Messier and Christenson’s approaches are described in fuller detail in the methods section of aim 1.

In general, a shift away from using p-values has occurred because p-values cannot indicate the actual truth. They are only able to summarize the data based on assumptions that define a null hypothesis. It has been suggested that statistical approaches should move in the direction of emphasizing the size of effects over the p-value.[94]

Another consideration particular to whether it is appropriate to choose a model based on the best fit is the natural occurrence of AMR in the environment. Diverse ARGs have been discovered in 30,000-year-old permafrost which provides firm evidence of widespread indigenous occurrence.[63] Theories suggest that bacteria producing antibiotics to fight against competitors would likely have corresponding antibiotic resistance genes. Naturally occurring metals could provide an additional co-selective mechanism.[6] Models should therefore either include natural occurrence of AMR, as presence does not necessarily imply contamination. Amos et al. accounted for natural occurrence in modeling the prevalence of class 1 integrons (i.e., molecular target was *intI1*) by defining the observed concentration, Y_i, as the sum of indigenous concentrations, C, plus the sum of exponentially decaying contributions (using inverse stream distance) from WWTPs, R_i, multiplied by a term, S, representing the rate at which class 1 integrons would increase from impacts of WWTPs. [38]

Amos' final regression model, where R_i was used as a predictor, was able to explain 82.9% of the variability in *intI1*. However, 26 predictors were used total to explain variability of *intI1* for only 52 samples, which could indicate an overfit model where the predictors are fitting to the noise of the data. Rather than defining the success of a model of AMR in the environment on the ability to explain 100% of the variability, approaches should attempt to identify influential factors and estimate the effect of those factors, reliably.

2.2.5 Modifying effects (attenuation and amplification)

Antibiotics, bacteria, and genetic material, such as ARGs or host-associated fecal markers may be attenuated (i.e., diluted, degraded, adsorbed, absorbed, etc.) by other factors, like land cover or climatic, chemical, and arguably onsite practices such as organic farming. Grassland buffers have been identified as attenuators of *E. coli* associated with livestock production inputs.[95] Additionally, there is some evidence that natural and constructed wetlands may attenuate fecal contamination.[96,97] Moisture, salinity, geochemical processes, temperature, and sunlight have been named as attenuation factors for AMR.[98–101] Not all ARGs persist equally, which could be because of attenuation processes specific to particular antibiotics, microbial communities, and genetic elements. Firm evidence points to *ermB* and *ermF* having shorter half-lives compared to *sul1*, *sul2*, and *intI1*, which may be related to the gene half-life or attenuation factors.[100,101] Geochemical processes such as the sorption of tetracyclines by sediments and soils explains some attenuation of tetracycline resistance genes in surface water. [102] To our knowledge, little is known about the amplification of microbial responses other than those described by Pruden et al. for antibiotic resistance genes: factors which promote cell growth, direct antibiotic selection, or the selection by another antibiotic or pollutant (i.e. co-selection).[17,60,61] Generally, studies have not purposefully investigated attenuation/amplification or happened upon consistent enough results to support definitive claims around what features are possible attenuators or modifiers of microbial contamination. However, the modifying effects of geological features (e.g., water table depth, bedrock depth, soil permeability) in relation to fecal contamination from sources has been suggested by some studies of groundwater.[37,103]

A non-linear regression model has been previously used to characterize modifying effects on contributions and have been primarily applied to air and groundwater for a wide array of non-microbial agents.[93] A more in-detail background of this model approaches can be found in aim 1. This will be the basis of the microbial land-use regression model developed in this work.

2.2.6 Worthwhile note on quantifying antimicrobial resistance

Due to complex microbial relationships, antimicrobial resistance might be difficult to quantify, especially in natural environments. Currently the most widely used approaches to quantify antimicrobial resistance are methods which target particular ARGs using quantitative polymerase reaction chain (qPCR).[61,104–107] Generally, studies have found that gene copies of mobile genetic elements, such as *intI1*, may be good estimators of general antimicrobial resistance and antimicrobial contamination.[38,108–110] However, any one ARG might be insufficient to understand antimicrobial resistance in a large environmental system like water or air.

CHAPTER 3 AIM 1: THE MICROBIAL FIND, INFORM, AND TEST (FIT) MODEL FOR IDENTIFYING SPATIALLY DISTRIBUTED CONTAMINATION SOURCES: FRAMEWORK FOUNDATION, AND DEMONSTRATION FOR RUMINANT *BACTEROIDES* ABUNDANCE IN RIVER SEDIMENT[1]

3.1 Overview

Microbial pollution in rivers poses known ecological and health risks, yet it remains difficult to establish causal and mechanistic linkages to sources. Land-use regression (LUR) models have been used to screen sources using spatial predictors but could be improved by better accounting for the gradual effect of proximity and density of upstream sources. We introduce a novel expansion of LUR, the microbial Find, Inform, and Test (FIT) modeling framework. FIT models the spatial relationships between rivers and sources by characterizing average transport (i.e., ground transportation, decay overland and decay downstream). We applied FIT to characterize sources of a *BoBac*, a ruminant *Bacteroides* fecal marker, quantified in riverbed sediment samples from Kewaunee County, Wisconsin. A one-standard-deviation-increase in manure application from animal feeding operations (AFOs) was associated with a 77% (p-value<0.05) increase in the relative abundance of ruminant *Bacteroides* (*BoBac*-copies-per-16S-rRNA-copies). To our knowledge, this is the first LUR to quantify the association between upstream sources and bovine-associated fecal markers in sediment. These findings have broad implications for sediment as a reservoir for microbial pollution associated with AFOs, such as

---

[1] This chapter was reproduced with permission from Environmental Science & Technology, submitted for publication. Unpublished work copyright 2021 American Chemical Society.

pathogens and antimicrobial-resistant bacteria. This application demonstrates the potential for applying FIT to broader microbial water quality measurements.

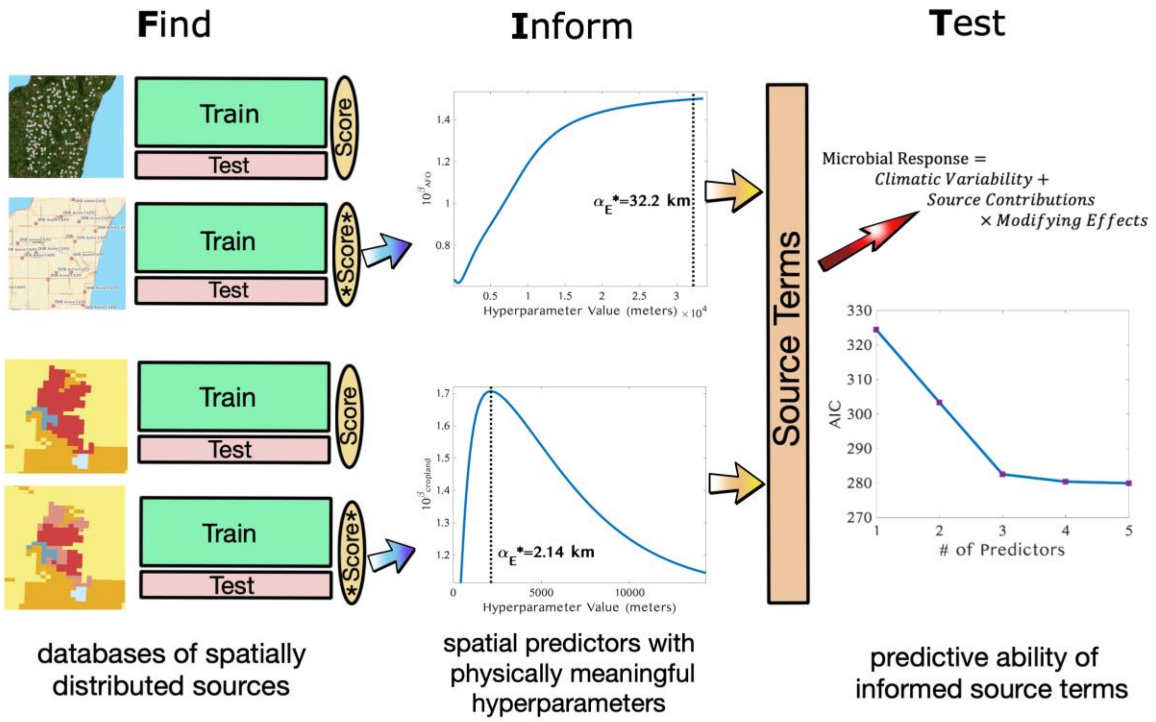


Figure 3.1 Graphical abstract depicting the general stages of the microbial Find, Inform, and Test (FIT) framework

3.2 Introduction

Microbial pollution of rivers and streams is a major public health concern. Mitigation of health risks demands microbial pollution models that depict the influence of key pollution sources. Mechanistic approaches are the gold-standard when sources are identified and well-characterized. However, there is often high uncertainty concerning the sources of emerging microbial pollutants[111,112] and delivery parameters to characterize microbial fate and transport.[113] Screening potential sources and identifying those likely associated with microbial pollution can be accomplished with regression-based approaches that leverage land-use/land-cover databases with transport-oriented models to construct *source terms*. Furthermore, regression approaches are needed that distinguish between source terms and modifying effects (i.e., amplification and attenuation) to address microbial fate and transport from sources, as described by Pruden *et al*.[17] in a conceptual mass-balance model for antimicrobial resistance responses. While this type of regression is frequently used for nutrient and chemical pollants,[114–117] no such land-use regression (LUR) model has been developed or implemented for microbial pollution or antimicrobial resistance more widely.

To our knowledge, few studies have aimed to relate land-use and land-cover with microbial pollution using LUR models.[17,30–38,87] Many of these studies do not distinguish between statistical exploration and physically meaningful models.[31–34,37,38] Physically meaningful statistical models distinguish between sources and modifiers (i.e., attenuating or amplifying effects).[17] This distinction is made in two ways: 1) by citing evidence that a set of spatial locations either produce microbial contamination or are modifiers of that contamination and 2) by ensuring that source terms can only contribute to pollution or not, whereas modifying effects would just scale those contributions. In regression, this implies a positive coefficient for sources. When negative coefficients are estimated for source terms,[32–34,37] depictions of concentrations of contaminants would increase moving away from sources (e.g., AFOs reduce microbial contamination). A model serves an exploratory purpose when no preliminary evidence exists for a land-cover/land-use to act as a source or modifier of microbial contamination. Exploratory spatial-statistical models often cannot provide the evidence required for recommendations to environmental and public health agencies.

LUR models distinguishing between sources and modifiers have been developed and implemented for chemical pollutants in groundwater, such as nitrate, where observed concentrations, y_i, for sample i was expressed as the following non-linear multivariate regression equation: [93]

$$y_i = \beta_0 + \left\{\sum_{u=1}^{U} \beta_u s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)\right\} \exp\left\{\sum_{l=1}^{L} \beta_l m_i^{(l)}\left(\boldsymbol{\alpha}^{(l)}\right)\right\} + \varepsilon_i, \qquad \text{(Eq. 3.1)}$$

where β_0 is the intercept, the β_u's are positive linear regression coefficients for the *source terms*, $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$(e.g., manure application fields) that increase microbial contaminations, and the β_l's are non-linear regression coefficient for the *modifier* terms $m_i^{(l)}\left(\boldsymbol{\alpha}^{(l)}\right)$(e.g., wetlands, forested areas, and buffers) that amplify or attenuate pollution. *Modifier* terms are exponential so that when the β_l's are positive or negative, the sum of contributions from sources is multiplied or fractionated, respectively, corresponding to amplification or attenuation. The $\alpha^{(u)}$'s and $\alpha^{(l)}$'s are hyperparameters that express the spatial scales at which a source or modifier affects microbial contamination. Based on previous approaches, the first step to applying this type of model to microbial contamination in river networks is to select databases of spatially distributed sources and then construct source terms with those databases.

There are opportunities for researchers to optimize selecting candidate databases of spatially distributed sources and standardize reporting of choice. For example, studies examining manure application fields' relationship to *E. coli* have utilized databases derived from remote sensing,[87] national land-cover databases,[32] and a database derived from interviews with local agricultural farmers and managers.[36] Each of these databases differently characterizes the locations of manure application fields. A tool that scores databases of spatially distributed sources by measuring the reliability in producing physically valid source terms (i.e., consistent estimation of positive regression coefficients) could be used as a qualitative technique to select from comparable databases.

After databases have been chosen, *spatial predictor models* are used to construct source terms. Microbial LUR models have captured source density with studies that use lumped-source terms (e.g., percent cropland-cover-per-watershed and cattle density-per-county).[30–33,35,37,91] Upstream capacities of sources were first captured with an inverse-distance-weighted interpolation method,[17] which constructs source terms at a higher resolution on the river network, but contributions from sources can increase without a source's influence (see section 2.2.2.2 for a detailed explanation of this issue). The sum of exponentially decaying contributions (SEDC) from upstream sources addresses this issue.[38,87] This class of spatial predictor model utilizes interpretable hyperparameters,[93,117,118] which results in depictions in the gradual effect of proximity and density of sources.[115]

Additionally, SEDC models do not require *a priori* knowledge of contamination sources and delivery parameters. Such is the case for current tools[36,89,113,119–121] that predict bacterial and nutrient pollution in rivers, such as the Soil and Water Assessment Tool (SWAT).[89,120,121] However, tools such as these have shown that dilution and climatic processes play critical roles in the variability of microbial contamination in rivers,[17,36,38,87,89,91,119,122–127] but have not been incorporated into physically meaningful statistical frameworks (Eq. 3.1).

Another critical component in the fate and transport of microbial pollution is the ground transportation from points of generation to release points into the environment (e.g., manure is produced where cattle are grown but applied to the land for nutrient management or fertilizer via trucking or irrigation).[128,129] To our knowledge, no existing spatial predictor model, such as SEDC, accounts for ground transport for estimating microbial responses. Additionally, no published work has compared the predictive power of different SEDC spatial predictor models for microbial contamination responses. See Table 2.1 for a literature summary.

We introduce the novel Find, Inform, and Test (FIT) framework, which advances previous microbial LUR approaches by distinguishing sources and modifiers. This is done by tailoring the non-linear multivariate regression (Eq. 3.1) for microbial responses in river networks. We develop a new spatial predictor model for source terms that do not require *a priori* knowledge of physical processes by using hyperparameters to define the spatial scale associated with ground transportation, overland and river distance decay, and microbial pollution flow. We also formalize approaches that test the predictive ability of source terms. The framework includes the following steps:

1. *Find databases of spatially distributed sources* using a reportable criterion
2. *Inform spatial predictors* by identifying hyperparameters that maximize physical meaningfulness
3. *Test the statistical significance* of various potential sources

With the novel FIT framework, we will quantify the effect that bovine manure application has on sediment fecal contamination for the first time. Previously, Pruden *et al.*[17] has used regression models to quantify the strength of the association between livestock source terms and the relative abundance of *antibiotic resistance genes* in river sediment. To our knowledge, no work has used regression models to quantify the association of source terms and the relative abundance of a *ruminant-specific fecal marker* in river sediment. Given sediments' capacity to store chemical pollutants[130] and bacterial genes[105,131] long-term, the magnitude of this association has implications to environmental and human health.[57,58,132] Here, we applied FIT to learn about bovine sources of fecal contamination in river sediment from Kewaunee County, Wisconsin. A novel spatial predictor model that accounts for the gradual effect of proximity to sources, source density, upstream-ness of sources, and dilution due to flow was developed to identify sources and characterize contributions. This is the first modeling approach incorporating all of these elements into a spatial predictor (see section 2.2.2.3 for details) of ruminant Bacteroides abundance in river sediment.

3.3 Materials and Methods

3.3.1 Study area and Sampling Data

Kewaunee County, Wisconsin, was selected as a study area due to local water quality concerns[133] and the high density of animal feeding operations (AFOs) compared to nearby counties in Wisconsin. Sediment samples (*n*=90) were taken at 20 river locations on five dates between July 2016 and May 2017. The River sampling methods and sites are further described in Beattie *et al.* [134]

3.3.2 River network and climatic data

The river network was extracted in ArcMap 10.5 from a Digital Elevation Model (10 m) raster file obtained from the Wisconsin Department of Natural Resources (WIDNR). Spatial data were projected into a “Wisconsin Central” State Plane coordinate system.

Daily, site-specific precipitation in centimeters and average monthly temperature were obtained by inverse-distance-weighted interpolation from weather station data from the National Oceanic and Atmospheric Administration’s (NOAA) climate data from April 1, 2016 to August 1, 2017 from 6 weather stations surrounding the study area (Brussels, Denmark WWTP, Forestville, Green Bay, Kewaunee, and New Franken). See A1.1 for detailed information on processing river network and climatic data.

3.3.3 Spatial databases of potential microbial contamination sources

There were two types of potential microbial contamination sources for which spatial databases were obtained. Animal feeding operations (AFOs) were the first type with three database options: 1) a county database of manure storages weighted by log-total-gallons of manure provided by the Kewaunee County Department of Land and Water Conservation 2) the same county database of manure storages but unweighted 3) a Wisconsin Pollution Discharge Elimination System (WPDES) database of CAFOs provided by the WIDNR and 4) the same WPDES database but weighted by animal units. Manure application fields (ManureApp) were the second type of source with two database options: 1) land cover database of crop rotation and 2) land cover database of dairy rotation. Both were obtained from the Wiscland-2 land-cover database.[135] See A1.2 and Table A.1 for details.

3.3.4 Microbial analysis of samples: sample processing, DNA extraction, and quantitative PCR

Sediment grab samples were collected (*n*=90) as described in Beattie *et al.*, immediately placed on ice following collection and stored at -20ºC until DNA extraction. Sediment samples were homogenized and subsampled for two replicate DNA extractions per sample. DNA was extracted from 0.5g of sediment as described in Beattie *et al.*[134]

FAM TaqMan® probe chemistry was used for quantitative polymerase chain reaction (qPCR) assays of bovine specific *Bacteroides* 16S rRNA genes,[136] and SYBR Green chemistry was used for qPCR assays of the V3 region of the 16S rRNA gene as a proxy for total bacteria. Genes were quantified in duplicate on a Real-Time PCR System (Bio Rad CFX Connect) from two separate DNA extractions per sample and averaged for total gene abundance (four reactions total per sample). Primers and annealing temperatures used in this study are listed in A1.3. TaqMan qPCR reaction mixtures contained 10 μL of TaqMan Gene Expression Master Mix (Applied Biosystems, Foster City, CA) 1 μL each of 10nM F/R primers, 0.4uL of 5nM probe, 2.6 μL of H_2O and 5uL of 4 ng/ μL gDNA. Plasmid DNA containing the cloned gene of interest was used to generate seven-point standard curves for each gene, and standard curves were run on every plate for *BoBac* and 16S rRNA genes. PCR amplification protocols consisted of 50°C for 2 min, followed by 95°C for 10 min and 40 cycles of 95°C for 30 s and Annealing Temp 57°C for 45 s. SYBR qPCR reaction mixtures and cycling parameters are as previously described.[134] Below detect values were set to one-half of the detection limit.[136]

The absolute abundance of *BoBac* (gene copies/g) for sample *i* was normalized to 16S rRNA (gene copies/g) to obtain the relative abundance z_i of *BoBac* (copies-per-16S-rRNA-copies), and base10 log-transformed to obtain y_i:

$$\begin{cases} z_i = \dfrac{\text{the absolute abundance of } BoBac\ (gene\ copies)}{16S\ rRNA \text{ abundance (gene copies)}} \\ y_i = \log_{10}(\, z_i) \end{cases}, \qquad \text{(Eq. 3.2)}$$

See A1.4 for details on absolute versus relative abundance.

3.3.5 Physically Meaningful Model for Contamination of Spatially Distributed Sources.

We introduce a microbial spatial model that formalizes previous microbial LUR models and expands the non-linear land-use regression from Eq. 3.1 to include climatic terms previously developed for a range of microbiological responses:[30,40,92,116,137]

$$y_i = \beta_0 + P1_i(\beta_1 + \beta_2 P2_i) + \beta_3 Freezing_i +$$
$$\left\{\sum_{u=1}^{U} \beta_u s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)\right\} \exp\left\{\sum_{l=1}^{L} \beta_l m_i^{(l)}\left(\boldsymbol{\alpha}^{(l)}\right)\right\} + \varepsilon_i, \qquad \text{(Eq. 3.3)}$$

The observed response value y_i (Eq. 3.1) for sample *i* is now a function of *source terms* and *modifier terms* and β_1, β_2 and β_3 are linear regression coefficients for the climatic variables $P1_i$ (recent precipitation), $P2_i$ (antecedent precipitation), and $Freezing_i$. In this work, we focus exclusively on the case where there are just a few independently distributed source terms and no modifiers, so Eq. 3.3 becomes linear. In subsequent papers, we will add more sources and modifiers.

The source terms, $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$, are standardized to a mean of 0 and standard deviation of 1, so that a one-standard-deviation increase in the u^{th} source term, leads to a β_u increase in the response, $y_i = \log_{10}(z_i)$. Conceptually similar to a Risk Ratio, we define the Relative Abundance Ratio, $RAR^{(u)}$, as the ratio of relative abundances for a one-standard-deviation increase in source *u*, which is given by $RAR^{(u)} = z(s^{(u)} + 1)\,/z(s^{(u)}) = 10^{\beta_u}$. (See A1.5 for derivation). It follows that a one-standard-deviation increase in the source term $s^{(u)}$ corresponds to a $(10^{\beta_u} - 1) * 100$ percent increase in *BoBac* relative abundance z.

In Eq. 3, β_1 represents the effect of recent precipitation on the microbial response and β_2 represents the diminished effect of antecedent precipitation, P2, on P1.[40] P1 and P2 at time *t* are calculated as sums of exponentially decaying contributions from the precipitations $P_{it'}$ at times t'<t and location *i* as:

$$P1_i = \sum_{t'<t} P_{it'} \exp\left(\frac{-3\,|t-t'|}{\alpha_{P1}}\right), \qquad \text{(Eq. 3.4)}$$

$$P2_i = \sum_{t'<t} P_{it'} \exp\left(\frac{-3\,|t-t'|}{\alpha_{P2}}\right), \qquad \text{(Eq. 3.5)}$$

where $\alpha_{P1} < \alpha_{P2}$. The exponential time decay parameters are constrained such that α_{P1} is less than α_{P2} and are optimized so that the diminishing effect of P2 on P1 is maximized to characterize a washout effect (see A1.6 for details).

3.3.6 Spatial predictor models describing spatially distributed contamination sources

We construct the *source* predictors $s_i^{(u)}(\alpha^{(u)})$ using three spatial predictor models of increasing sophistication: (1) a Euclidean model, (2) an Overland and River Distance-Flow (ORF) model and (3) a Ground Transport, Overland and River Distance Flow (GORF) model. The first approach uses Euclidean distance.[93] The value of $s_i^{(u)}(\alpha^{(u)})$ at location *i* is calculated as the standardized sum of exponentially decaying contributions (SEDC)[93] from surrounding sources *j* :

$$s_i^{(u)}\left(\alpha^{(u)}\right) = s_i^{(u)}\left(\alpha_E^{(u)}\right) = z - score\left(\sum_{j=0}^{N} C_{0j} \mathrm{e}^{\frac{-3D_{ij}{}^{(E)}}{\alpha_E}}\right), \qquad \text{(Eq. 3.6)}$$

where $D_{ij}^{(E)}$ is the Euclidean distance between observation point *i* and source j, C_{0j} is the initial relative abundance at source j, and $\alpha_E^{(u)}$ is the Euclidean hyperparameter corresponding to the distance away from a source where an initial source abundance of C_{0j} is reduced by 95%. An example of the *source* term, for *u*=manure application fields, $s_i^{(u)}\left(\alpha_E^{(u)}\right)$ is shown in Figure 3.1(a) using this Euclidean approach when the range of influence, $\alpha_E^{(u)}$, is 2.0 km.

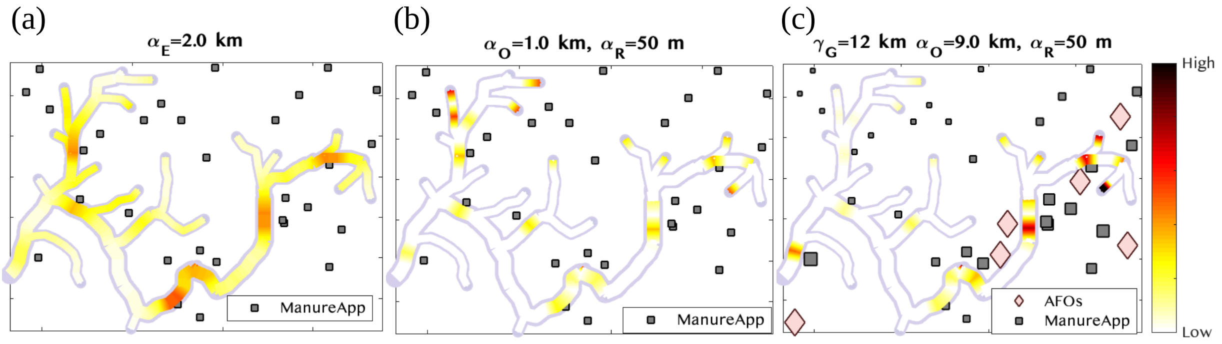


Figure 3.2 Diagrams that demonstrate a) Euclidean b) Overland and River Distance—Flow (ORF) c) Ground Transportation, Overland and River—Flow (GORF) sum of exponentially decaying contributions in a downstream combining river network with examples of hyperparameter values.

The sources in Figure 3.1(a) are manure application field centroids representing areas of equal size. In this Euclidean approach, contributions from sources to abundance appear high when manure application fields are dense and close to the river network. However, a source located next to the river network appears to contribute to nearby river segments both downstream and upstream, which opposes first-order principles and studies indicating a higher prevalence of microbial contamination downstream of sources than upstream.[72,106,138–144]

Accordingly, in a second approach, we use the Overland and River Distance—Flow (ORF) model. The mass M_{0j} of microbial contaminants applied at a source location j is transported to the river over an overland distance $D_{ij}^{(O)}$, and then down the river over a river distance $D_{ij}^{(R)}$ until it reaches the sampling location *i* where, as described in Jat and Serre,[145] it is diluted within the flow Q_i. The Strahler stream order is used as a proxy for flow in sediment.[146] The standardized ORF spatial predictor model is:

$$s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right) = s_i^{(u)}\left(\alpha_O^{(u)}, \alpha_R^{(u)}\right) = z - score\left(\sum_{j=0}^{N} \frac{M_{0j}}{Q_i} \mathrm{e}^{\frac{-3D_{ij}{}^{(O)}}{\alpha_O} + \frac{-3D_{ij}{}^{(R)}}{\alpha_R}} \delta_{\overline{ij}}\right) \qquad \text{(Eq. 3.7)}$$

where $\delta_{\overline{ij}}$ is a hydrologic connectivity indicator equal to 1 if sampling location *i* receives flow coming from application location *j*, 0 otherwise, and the two hyperparameters, α_O and α_R, describe the overland and river exponential decay ranges, respectively.[87,145,147–150] An example of ORF modeled contributions is shown in Figure 3.1(b) using $\alpha_O^{(u)}$=1.0 km, $\alpha_R^{(u)}$= 50 m, and M_{0j}'s proportional to the size of manure application fields. The ORF model captures hydrologic processes involved in transporting microbial contaminants from where they are applied to a downstream sampling location. However, it does not capture how the mass M_{0j} applied at location *j* is influenced by the proximity of AFOs from where manure originates.

To address this latter point, we introduce the Ground Transportation, Overland and River Distance—Flow (GORF) model, where the mass applied at application field *j* is calculated based on amounts that are ground transported (e.g., hauled by trucks and irrigated) from nearby AFOs. Here $M_{0j} = \sum_{k=1}^{K} \omega_{kj} M_k$, where M_k is the mass produced at AFO *k* (proportional to size of that AFO) and ω_{kj} is the proportion of M_k ground transported to application field *j*. We assume that ω_{kj} is proportional to an exponential decrease with the distance D_{kj} between AFO *k* and application field *j* (since it is more costly and time consuming to drive waste far), i.e., $\omega_{kj} \propto$ exp $\left(\frac{-3D_{kj}}{\gamma_G}\right)$ where the distance hyperparameter γ_G is a transportation range reflecting how far is manure being hauled. The ground transport distances, D_{kj}, are expressed with a Euclidean distance metric as transit and irrigation records are not publicly available. However, our model, which exponentially weights the application of manure onto fields at a Euclidean distance closer to AFOs, is not an unreasonable proxy based on the exponentially decreasing hauling capacity over transport distances estimated by Hadrich *et al.*[129] We also assume that there is no loss of manure during hauling, so that mass is conserved, i.e., $\sum_{j=1}^{J} \omega_{kj} = 1$. Hence the full equation of the standardized GORF spatial predictor model is

$$s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right) = s_i^{(u)}\left(\gamma_G^{(u)}, \alpha_O^{(u)}, \alpha_R^{(u)}\right) = z - score\left(\frac{1}{Q_i} \sum_{j=0}^{N} \sum_{k=1}^{K} \omega_{kj} M_k \, \mathrm{e}^{\frac{-3D_{ij}{}^{(O)}}{\alpha_O} + \frac{-3D_{ij}{}^{(R)}}{\alpha_R}} \delta_{\overleftarrow{ij}}\right),$$

$$\text{where } \omega_{kj} = \exp\left(\frac{-3D_{kj}}{\gamma_G}\right) / \sum_{j=1}^{J} \exp\left(\frac{-3D_{kj}}{\gamma_G}\right) \qquad \text{(Eq. 3.8)}$$

An example of GORF model is shown in Figure 3.1(c) using $\gamma_G^{(u)}$=12 km, $\alpha_O^{(u)}$= 9.0 m, $\alpha_R^{(u)}$= 50 m, and M_k's proportional to the size of AFOs. This approach yields a map with greater concentration downstream of manure application fields that are near AFOs.

3.3.7 Selection of source hyperparameters in the FIT framework

An approach is needed to select hyperparameters, $\boldsymbol{\alpha}^{(u)}$, of the source terms $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$ in the microbial regression equation 3. In the Find and Inform steps of the FIT framework, we consider each source term individually, and for that source term we select the hyperparameter that maximizes the RAR (10^{β_u}) for some training set of observed outcomes y_i (see A1.7 for details). In the Test step of the FIT framework, we test whether the source term with its selected hyperparameter is statistically significant.

Previous works have selected hyperparameters that maximize the R-squared.[93] However, this can lead to physically invalid results (e.g., a source term may reduce microbial pollution). Our FIT framework addresses this issue by emphasizing physical meaningfulness over statistical prediction. The Find, Inform, and Test steps of the FIT framework are detailed in the next three sections.

3.3.8 Find reliable databases of spatially distributed sources

We develop a reliability score as a criterion for finding databases *d* of a given source *u*, that is selected based on which most reliably produces a positive β_u>0 for the source term $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$ defined in Eq. 3.3. We start with model *a* corresponding to the microbial regression (Eq. 3.3) reduced to an intercept and the source term $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$. We create k-folds (k=5) of training and test sets from the data. We use model *a* with each training set to obtain hyperparameter $\widehat{\boldsymbol{\alpha}}_{training_k}^{(u)}$which maximizes β_u (see A1.7 for details). We adopt $\widehat{\boldsymbol{\alpha}}_{training\_k}^{(u)}$, for the test set regressions to obtain $\beta_u^{(test\_k)}$. We define the Sign Stability Score *SSS* as a number between 1 and k+1 (k+1=6) which quantifies how many of the $\beta_u^{(test\_k)}$ values are positive. We assess component of reliability through taking the sum of the $\beta_u^{(test\_k)}$ across the folds to represent an average magnitude, *M*. Lastly, we define $\sigma_{\beta_u^{(test\_k)}}$ as the standard deviation of the $\beta_u^{(test\_k)}$ values. We reward high *SSS*, high *M*, and low $\sigma_{\beta_u^{(test\_k)}}$ values by setting the reliability score *RS* equal approximately to the ratio of *SSS* multiplied by *M* over σ_{β_u} (see A1.8 for details). Finally, we select for each source term the corresponding database *d* with the highest reliability score.

3.3.9 Inform Spatial Predictors with Hyperparameters

We inform each source term individually by selecting the hyperparameter $\widehat{\boldsymbol{\alpha}}^{(u)}$ which maximizes the β_u using 100% of the observations y_i in model *a*. This insures that for each spatial predictor model (Euclidean, ORF, and GORF) we inform the source term with all the data at hand. The maximization procedure is the same as in the “Find” stage and described in detail in

A1.9. For the “Inform” stage, a penalty was added to the objective function such that a combination of γ_G and α_O yielding poor regression or mapping qualities (i.e., non-normal residuals or $\hat{y}_i > 0$) would not be selected.

3.3.10 Test the Statistical Significance of potential sources

After we individually inform each source term with hyperparameters that maximize physical meaningfulness, we test their statistical significance by considering standardized climatic and source terms in the microbial regression (Eq. 3.3). An Akaike Information Criterion (AIC) exploits a stepwise algorithm to select the informed source terms — a standard practice in microbial LUR studies.[31,33,35,36,38,87] A possible outcome is that none of the source terms are selected. To compare various modeling options (i.e., Euclidean, ORF, GORF), a 6-fold cross validation of the “Test” step of the FIT framework was performed. See A1.9 for details.

3.4 Results and Discussion

3.4.1 Finding reliable databases

A criterion was successfully developed at the Find stage of the FIT framework where a higher score corresponds to greater reliability of database locations representing source locations. The criterion is based on the *SSS*, *M*, and σ_{β_u}, respectively, capturing the consistency, magnitude, and variability of contributing effects across test sets. Across candidate databases, we report that the WPDES database of CAFOs representing AFOs was more reliable than the county database of manure storages (weighted or unweighted) obtained through imagery, indicating that using imagery insufficiently captured sources and could not replace permitting information. Lastly, compared to the crop rotation land-cover database, the dairy rotation subclass more often appeared to represent a source of bovine fecal contamination reliably.

We also report here on the reliability of spatial predictor modeling approaches. GORF-modeled predictors more consistently represented sources (*SSS* ranging from 3 to 5*)* than ORF *(SSS* ranging from 1 to 5) or Euclidean (*SSS* ranging from *2 to 5*). The higher *SSS* of GORF-modeled predictors suggest that modeling ground transport of manure was essential to consistently capturing the contributing effects of AFOs to sediment bovine markers. See A1.10 for details.

3.4.2 Informing of spatial predictor variables

At the Inform stage of the FIT framework using the GORF spatial predictor model we find that hauling of manure from large AFOs occurs over distances of kilometers (i.e., $\gamma_G > 1\ km$). We found that fecal contamination also reduced from manure fields over distances of kilometers overland ($\alpha_O > 1\ km$). There are several possible explanations including possible long-range overland contamination due to subsurface tile drainage[151] and karst geography.[152,153] Finally, downstream contamination of riverbed sediment from overland discharge points occurs over sub-kilometer distances (i.e., $\alpha_R < 1\ km$), suggesting localized sediment contamination. See A1.11 for details.

3.4.3 Testing the predictive ability of source terms

Results of the Test stage of the FIT framework are shown in Table 3.1. The first column defines the spatial predictor model used to construct source terms and summarizes model performance with R-squared. The results of the inform stage of FIT are shown in the column labeled hyperparameters. All climatic terms were selected (see A1.12 for details). Each source term has hyperparameters even if they were not selected. The Test Stage results are shown in the last two columns, which show the standardized regression coefficient value and the corresponding RAR value.

Among the models resulting from different spatial predictors, only the GORF source term indicated that fecal contamination in sediment was associated with AFOs via manure application at a statistically significant level. The selection of this GORF predictor, AFO ground transport to ManureApp, suggests that flow and ground transport from AFO to manure fields are transport mechanisms for bovine fecal contamination in sediment. The GORF approach also performed slightly better in a 6-fold cross validation (see A1.13 for details). Some microbial LUR and geostatistical studies have found that river distance measures are more predictive and generalizable than Euclidean approaches for modeling contamination in a river network.[87,150,154,155] However, this is the first modeling approach to suggest hauling as a key transport process.

Table 3.1 Regression model results of bovine fecal contamination in sediment (log10 *BoBac*-copies-per-16S-rRNA-copies) shown for each spatial predictor model (SPM) used for source terms (i.e., Euclidean, ORF, and GORF) and the stepwise-selected model performance expressed as R-squared and adjusted R-squared. For each SPM, standardized explanatory variables, their physically meaningful hyperparameters, $\boldsymbol{\alpha}$, their regression coefficients, βs, and resulting relative abundance ratio, 10^{β}, are provided in row-wise.

SPM for Source Terms Stepwise-selected Model Performance	Standardized Explanatory Variables	Hyperparameters $\boldsymbol{\alpha}$	Regression Coefficient β	Relative Abundance Ratio 10^{β} (95% CI)
Euclidean R^2=0.415 Adjusted R^2=0.395	P1	$\alpha_{P1=}$0.177 days	0.995**	9.90 (4.19, 23.4)
	P1×P2	α_{P2}=2.04 days	-0.221	0.602 (0.311, 1.16)
	Freezing	n/a	-3.31**	4.94e-4 (7.06e-5, 0.00347)
	AFO	α_E = 32.4 km	NOT SELECTED	NOT SELECTED
	ManureApp	α_E = 1.51 km	NOT SELECTED	NOT SELECTED
Overland and River Distance -Flow (ORF) R^2=0.428 Adjusted R^2=0.401	P1	α_{P1} =0.177 days	1.00**	10.1 (4.28, 23.6)
	P1×P2	α_{P2} =2.04 days	-0.222	0.600 (0.312, 1.16)
	Freezing	n/a	-3.29**	5.04e-4 (7.27e-5, 0.00349)
	AFO	α_O=2.58 km , α_R= 308 m	NOT SELECTED	NOT SELECTED
	ManureApp	α_O=935 m, α_R= 54.0 m	0.168	1.47 (0.858, 2.53)
Ground Transport, Overland and River Distance -Flow (GORF) R^2=0.442 Adjusted R^2=0.417	P1	α_{P1} =0.177 days	0.988**	9.90 (4.18, 22.6)
	P1×P2	α_{P2} =2.04 days	-0.220	0.602 (0.360, 1.01)
	Freezing	n/a	-3.25**	5.62e-4 (8.30e-5,0.00380)
	AFO ground transport to ManureApp	γ_G=11.3 km, α_O= 8.94 km, α_R= 50.5 m	0.248**	1.77 (1.04, 3.02)

**indicates a p-value <0.05

We focus on the test stage results resulting from the GORF spatial predictor model due to its statistical significance. The finding is that a one-standard-deviation increase in manure application from AFOs was associated with a 77% (p-value<0.05) increase in the relative abundance of bovine fecal contamination (*BoBac*-copies-per-16S-rRNA-copies) in riverbed sediment. No study of sediment exists with which to compare this value. However, in a study of absolute abundance of fecal contamination (log10 gene copies not normalized by 16S rRNA) in surface water, Dila *et al*. found that a one-standard-deviation increase in cattle density per watershed was associated with 20% (p-value<0.05) increase in the absolute abundance of bovine gene (*BacR*) copies.[30] For comparison, we also applied FIT to the absolute abundance of bovine fecal contamination (log10 gene copies-*BoBac* per gram sediment). A one-standard-deviation increase in manure application from AFOs was associated with a 79% (p-value<0.05) increase in bovine gene (*BoBac)* copies (see A1.14 for details), suggesting a stronger signal from sources in sediment than in water, though for different markers.

Our work is the first to use a regression model to quantify the strength of the association between a bovine source predictor and the relative abundance of a bovine-*specific fecal marker* in downstream river sediment. Though a positive association has been observed between bovine sources and fecal contamination *in water* by Dila,[30] researchers should be careful in comparing the strength of these associations. First, the two bovine markers (i.e., *BoBac* in our study, *BacR* in Dila's study)[30] may degrade at different rates.[156,157] Additionally, inhibition or other factors may affect true-positive rates in sediment. While the *BacR* marker has been reported to have a higher source-specificity, *BoBac* has offered a higher source-sensitivity[158]–one rationale for a higher effect in our study. Overall, there is still a need for medium-scale experiments to compare sensitivity, specificity, and inhibition of these genetic markers from complex environmental samples using PCR assays. Alternatively, if both markers are comparable, transport processes may also contribute to differences in effect. There is a need for experiments that measure degradation, deposition, and resuspension of host-associated markers into and between surface water and sediment.

Other factors contributing to variability in the estimate of the association of sources on ruminant *Bacteroides* in sediment include sediment depth, size, composition, and proximity to geological and other anthropogenic factors.[159–163] For example, fine-course, silty sediments have been correlated with higher microbial concentrations and bacterial growth.[160–162] In one study, while host-associated fecal markers observed in sediment along a river gradient corresponded with respective upstream animal and human hosts, notable differences in the concentrations and upstream correlations existed between the first centimeter of sediment and below.[163] These factors indicate the need for well-maintained spatial databases of sediment types and precise sampling approaches.

3.4.4 Implications, limitations, and future work

To our knowledge, this is the first LUR modeling of the abundance of a bovine-associated fecal marker in river sediment. Our primary finding is that ground transport of bovine manure from AFOs to application fields contributes to riverbed sediment fecal microbial abundance. This study's fecal abundance response, specific to ruminants and highly-associated with bovine hosts,[136] provides strong evidence of offsite migration from AFOs into the environment. This effect may have long-term consequences as microorganisms can remain in sediment for long periods or be resuspended into the surface water.[58] Sediment as a reservoir of AFO-associated fecal contamination is particularly concerning because AFO operations frequently use antibiotics to treat and prevent disease, increasing the risk of elevated antimicrobial resistance in these contaminated sediments.[11] The finding in previous work of a positive association between sediment ARGs and upstream capacities of AFOs[17] created a need to understand how ARGs may be present in sediment downstream of AFOs. Our findings provide the strongest support to date that a mechanism at play may be ARG *transport* (i.e., ground transport, overland and downstream) through the dissemination of fecal *matter* from manure application. An additional mechanism for elevated ARGs in sediment near AFOs could be the selection pressure on low levels of naturally occurring antibiotic resistance[164,165] from antimicrobials that disseminate into river sediments. A wider application of LUR approaches to antimicrobials, ARGs, and microbial contamination (fecal bacteria and pathogens), may help untangle the impacts of microbial contamination versus antimicrobials from AFOs on antibiotic resistance in the riverbed sediments and the environment more generally.

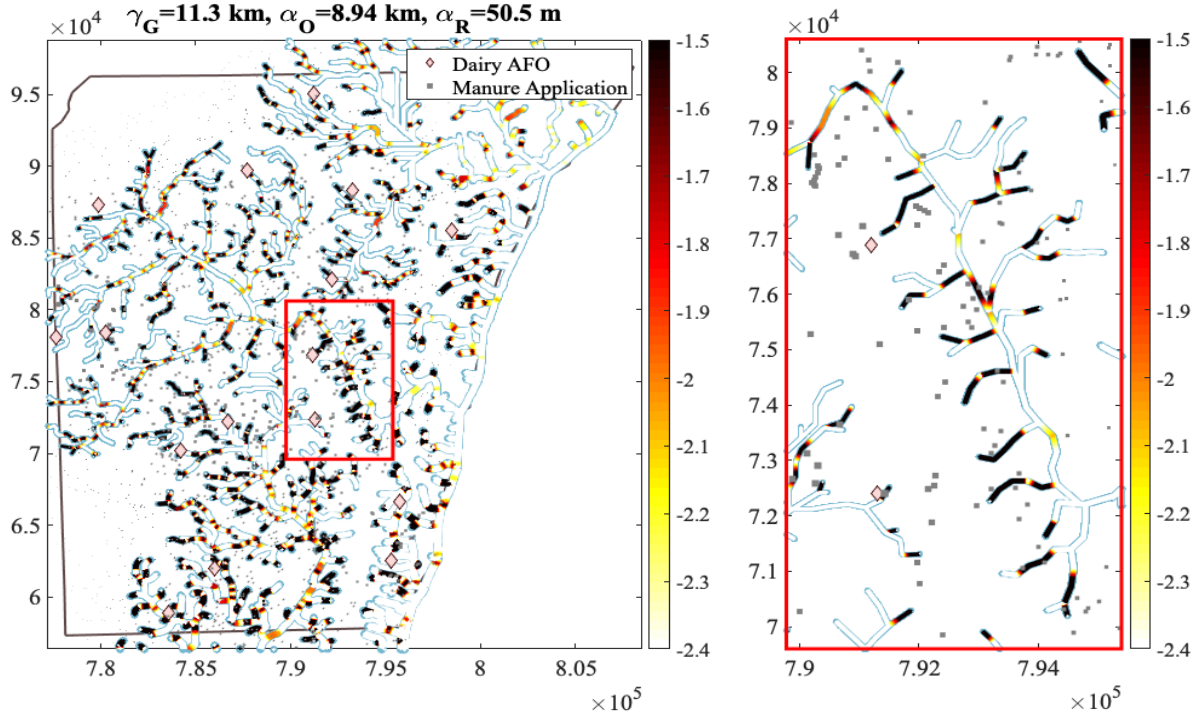


Figure 3.3 Modeled relative abundance (log10 *BoBac*-copies-per-16S-rRNA-copies) plotted on the river network for Kewaunee County given average recent and antecedent precipitation and freezing temperature. The Find stage results for AFO and manure application fields are depicted by the red diamonds (WPDES CAFOs that are unweighted) and the grey squares (land cover database of dairy rotation).

Our novel LUR framework successfully found reliable databases of spatially distributed sources, informed source terms, and tested those source terms' predictive ability, resulting in the first-ever depiction of bovine fecal markers in sediment in a river network (Figure 3.2). Figure 3.2 was created by applying the model resulting from the Test stage for all river network points, given the databases found to reliably represent spatially distributed sources from the Find stage. Overall, this depiction can help water quality by suggesting key locations for monitoring sediment contamination in Kewaunee County rivers.

Figure 3.2 depicts results from the Inform stage, which combined and expanded previous spatial predictor models, such that it was the first to account for the gradual effect of proximity to sources, the density and upstream-ness of sources, dilution due to flow, and ground transportation of microbial contamination before its dissemination into the environment. Each of the GORF processes can be seen in Figure 3.2. The proximity of the AFOs (viz. red diamonds) to manure field centroids (viz. grey squares) increases the amount applied on manure fields. Similarly, the proximity and density of highly applied manure fields to the river network, increases the fecal contamination levels in sediment (viz. redder in color). The dilution effect can be seen as sharp drops in fecal contamination downstream of confluences of pristine, higher-order streams (viz. depicted as greater thickness) and highly contaminated, lower-order streams.

Overall, the model resulting from the FIT framework performed similarly to other LUR models of microbial contamination. This model explained 44.2% of the variability (adjusted to number of predictors: 41.7%), which is within the range of previous LUR models of fecal contamination.[17,30–33,35,37,38] Unstudied source types may cause some unexplained variability. We observed that measured values at the river sampling site abbreviated BPKR[134] were higher than our predicted values. The site is named BPKR due to its proximity to Bruemmer Park Zoo, home to other ruminants, such as goats and a giraffe. Zoos and other neglected potential sources should be studied as sources of bovine markers to determine their inclusion in future LUR studies of fecal pollution.

Further variability may be explained by amplifying or attenuating factors. For example, the BPKR site was located 500 m downstream of a dam. Dams and levees cause upstream accumulation of fine-grained sediments, and the nutrients required for bacterial survival adsorb to these fine-grained sediments.[130] Sediment reservoirs around dams, therefore, have the potential to act as amplifiers of microbial contamination in sediment. Values at some other sampling sites were lower than expected by the model, possibly due to the influence of wetlands or forested areas. There is some evidence that wetlands or forests can attenuate nutrient loads,[93,166] and constructed wetlands can attenuate microbial pollution.[96,97] Future work will develop approaches for finding reliable databases and informing modifying effects in a physically meaningful way.

The results of this work emphasize the need for more fine-scale modeling approaches, increased sampling size to inform and validate the models, and well-maintained spatial databases that characterize geological, land-use and management, and land-cover information. In the future, the FIT framework application could identify key sources of emerging microbial pollutants and be instructive to assessing the microbial risk associated with river water use.

CHAPTER 4 AIM 2: CHARACTERIZING SOURCES OF BOVINE AND HUMAN-ASSOCIATED FECAL CONTAMINATION IN WISCONSIN SPATIAL STREAM NETWORK WITH THE MICROBIAL FIND, INFORM, AND TEST FRAMEWORK

4.1 Overview

Fecal pollution in surface water and sediment increases public health risks to water users. Microbial source tracking methods that utilize host-associated genetic markers of fecal contamination can be useful to estimating public health risks. However, to mitigate these risks, the primary sources of these host-associated fecal markers need to be identified and average transport processes need to be characterized. This can be accomplished with the microbial Find, Inform, and Test framework. FIT was implemented for bovine and human-associated fecal markers, *BoBac* and *HuBac*, respectively, quantified from riverbed sediment and surface water samples in a dairy livestock-dense region with many potential sources of bovine or human-associated fecal pollution. For the relative abundance of *BoBac* (gene-copies-per-16S-rRNA-copies) in sediment and surface water, one-standard-deviation increases in contributions from dairy animal feeding operations (AFOs) were associated with 99% (p-value<0.05) and 79% (p-value<0.05) increases, respectively, and contributions from low-intensity developed landcover were associated with a 108% (p-value<0.05) and 39% (inclusion reduces model AIC) increases, respectively. For the relative abundance of *HuBac* (gene-copies-per-16S-rRNA-copies) in sediment and surface water, a one-standard deviation increase in septic system contributions was associated with 79% (inclusion reduces model AIC) and 56% (inclusion reduces model AIC) increases, respectively. A one-standard-deviation increase in high-intensity developed land cover was associated with a 47% (inclusion reduces model AIC) increase in the relative abundance of

HuBac (log10 gene-copies-per-16S-rRNA-copies) in sediment. Sediment was found to better capture the contributions to fecal contamination from distant sources, which has broad implications to microbial risk assessment and mitigation. Patterns of *bovine*-associated fecal pollution flowing away from rural areas and toward developed residential areas and *human*-associated pollution flowing away from developed residential areas and toward rural areas was revealed. This was the first comprehensive land-use regression modeling done for two host-associated markers across surface water and sediment. The region is overall impacted by fecal contamination from bovine and human sources and investments in wastewater infrastructure are needed. Careful attention to the proximity of sources to rivers and streams and sources to impervious land cover may help agriculture and tourism stakeholders make decisions about rural and urban planning and land-use.

4.2 Introduction

Fecal pollution in water is a public health hazard, especially in regions that economically depend on clean water for agriculture and tourism. Public health risks can be estimated by identifying the types of hosts contributing to fecal pollution.[26] For example, human exposure to human fecal contamination is traditionally considered risker for humans compared to non-human fecal contamination due to its association with human pathogens.[57] However, livestock-associated fecal contamination carries additional risks compared to other non-human fecal contamination due to the frequent use of antimicrobials and antibiotics for treatment and disease prevention in livestock.[11] One method for identifying particular hosts associated with samples is the quantification of host-specific or host-associated genetic markers of fecal contamination. However, quantified host-associated markers cannot be attributed to source locations without spatially relating contributions from sources to the measured host-associated responses. Because of this missing connection to source locations, researchers have recently begun to use host-associated markers as responses in land-use regression (LUR) models that find associations between spatial predictors describing sources and responses.[29,30] In particular, LUR models that use the *sum of exponentially decaying contributions* (SEDC) as a model for spatial predictors to construct source terms utilize interpretable hyperparameters that characterize distance decay ranges around sources.[93] This can help to determine adequate buffer areas around sources. Only one study has used SEDC spatial predictors to characterize distance decay from sources through overland and downstream distance decay and accounted for dilution due to flow. (Wiesner-Friedman 2021*)

Across many regions, elevated levels of fecal contamination have been measured around the following sources or have been associated with increases or proximity of these sources in *surface water* and/or *groundwater*: AFOs,[30,33,34,75] septic systems,[37,82,103,167–169] and sewer lines to wastewater treatment plants and combined sewer overflow events.[29,163,170,171] Fewer studies have examined the capacity for fate and transport of fecal contaminants from land applied biosolids and wastewater from non-human waste.[172] A number of studies have also found significant correlations between fecal contamination and impervious surfaces and developed land cover.[29,35,90,173–175] No modeling approach has comprehensively compared the effects of these sources to identify key contributing sources among them. To the best of our knowledge, associations have not been made between the locations of these particular sources and fecal contamination in *riverbed sediment* using LUR or other approaches. However, recent studies suggest that long-term storage of pollutants in sediments also contribute to microbial risks.[58,130] Ultimately, to reduce risks from fecal contamination in freshwater, key source locations contributing to fecal contamination need to be identified and characterized for surface water and riverbed sediments.[26]

Kewaunee County, Wisconsin is an area with Karst geography[152,176] that is known for dairy livestock agriculture, dairy product industry, and a tourism industry that thrives around water recreation. This area is home to 17 dairy concentrated animal feeding operations (CAFO) that house approximately 100,000 animals,[177] which would be estimated to produce almost 6 million kilograms (i.e., 6 Olympic-sized swimming pools) of feces and urine *per day* that would eventually be applied to land.[178] The area also has smaller AFOs and family farms. In comparison, there are approximately 20,000 permanent residents and at least tens of thousands of tourists that visit annually.[179]

Fecal contamination has been reported to infiltrate into drinking well water at unsafe levels.[133] Due to the ratio of bovine to humans, fecal contamination is suspected to primarily come from bovine hosts.[180] However, there are other discharge locations that may represent a large mass of fecal bacteria that may disseminate into rivers and streams. In Kewaunee County, the application of municipal and industrial biosolids and wastewater (e.g., industrial wastes, municipal waste/sludge, and septage waste),[181] septic systems, and wastewater treatment plants may represent such areas. Due to the number and availability of databases describing potential sources of fecal contamination by location in this region, a comprehensive land-use regression study is possible.

In addition to spatial factors such as sources, precipitation and snowmelt, in particular, have been related to increases in fecal contamination.[29,30,82,90,182,183] However, studies of climatic, meteorological, or hydrological variability have primarily focused on human-associated markers and microbial responses in water.[29,30,82,103,174,182,183] There are opportunities to learn about the impacts of climatic factors on non-human-associated markers and microbial responses in sediment.

The aim of this study was to implement the first comprehensive land-use regression study for two host-associated fecal markers measured from riverbed sediment and surface water. This study will identify key sources, characterize overland and downstream decay from those sources, and account for important climatic factors associated with fecal pollution in sediment. Multiple spatial databases are available for this region that characterize approximately the same source category. For source terms, spatial predictor models are needed that characterize overland and downstream flow from many different source locations. A method is also needed to screen these potential sources to identify those that primarily contribute to the fecal pollution. The microbial FIT land-use regression framework fulfills these needs by 1) *finding* reliable databases representing spatially distributed sources across comparable candidate databases representing an overarching source category, 2) *informing* source terms by identifying hyperparameters that maximize physical meaningfulness for contributions to a river network, and 3) *testing* the predictive ability of informed source using a physically meaningful LUR model so that key sources can be identified. By identifying primary contributors to bovine and human- associated fecal contamination, this work will help to mitigate public health risks and improve river ecology. The knowledge of ecological differences between sediment and surface water can also help to better model microbial risks to public health.

4.3 Methods

4.3.1 Microbial contamination responses: sample collection, processing, DNA extraction, and quantitative PCR

Sediment and Sediment (*n*=90) and river samples (*n*=98*)* were taken at 20 river locations in Kewaunee County, Wisconsin on five dates between July 2016 and May 2017. For surface water, 1 L grab samples were taken using a sterile bucket. Some samples were not collected due to access. More surface water samples because flooding conditions made it difficult to collect sediment samples, while surface water could still be collected from bridges. The riverbed sediment sampling methods and sites are further described in Beattie *et al.* [134] Across the 20 sampling sites, the minimum Euclidean distance between sampling sites was approximately 1 km and the maximum was approximately 47 km.

Sediment grab samples were filtered as described in Beattie *et al.* [134] For surface water samples, 100 mL was filtered onto 0.22 μm mixed cellulose esters membrane filters (47 mm, type GSWP, Millipore). FAM TaqMan® probe chemistry was used to quantitate bovine and human specific *Bacteroides* 16S rRNA gene qPCR assays, and SYBR Green chemistry was used to quantitate the V3 region of the 16S rRNA gene as a proxy for total bacteria. Genes were quantified in duplicate on a Real-Time PCR System (Bio Rad CFX Connect) from two separate DNA extractions per sample and averaged for total gene abundance (four reactions total per sample). Primers and annealing temperatures used in this study are listed in Table 4.1. TaqMan qPCR reaction mixtures contained 10 μL of TaqMan Gene Expression Master Mix (Applied Biosystems, Foster City, CA) 1 μL each of 10nM F/R primers, 0.4uL of 5nM probe, 2.6 μL of H_2O and 5uL of 4 ng/μL gDNA. Plasmid DNA containing the cloned gene of interest was used to generate seven-point standard curves for each gene. PCR amplification protocols consisted of 50°C for 2 min, followed by 95°C for 10 min and 40 cycles of 95°C for 30 s and Annealing Temp for 45 s with a fluorescence plate read. SYBR qPCR reaction mixtures and cycling parameters are as previously described.[134] Prior to statistical analysis, absolute abundance of the bovine and human *Bacteroides* genes per gram of sediment or 100 mL of water were normalized to 16S rRNA gene abundance per gram of sediment or 100 mL of water, respectively. See Wiesner-Friedman *et al.* (2021: under review) for details. Below detect values were removed prior to spatial-statistical modeling.

Table 4.1 Primers and annealing temperatures from Layton et al. [136]

Names	Primers	Size of Product	Annealing Temp
HuBac	HuBacF-5′-GGGTTTAAAGGGAGCGTAGG-3′ HuBacR- 5′-CTACACCACGAATTCCGCCT-3′ HuBacProbe- (FAM)TAAGTCAGTTGTGAAAGTTTGCGGCTC(BHQ-1)-3′	116bp	60C
BoBac	BoBacF 5'-GAAG(G/A)CTGAACCAGCCAAGTA-3′ BoBacR- 5′-GCTTATTCATACGGTACATACAAG-3′ BoBacProbe-(FAM)TGAAGGATGAAGGTTCTATGGATTGTAAACTT(BHQ-1)-3′	100bp	57C
Universal V3 region primers (16S)	Universal V3 region primers 314F- 5'-CCTACGGGAGGCAGCAG-3' 518R- 5'-ATTACCGCGGCTGCTGG-3'	204bp	58C

4.3.2 River network and climatic data

River network and climatic data were processed identically to what is described in Wiesner-Friedman *et al.* (2021: revision requested to ES&T).

4.3.3 Study area description and databases representing sources of microbial contamination

The study area in Kewaunee County, Wisconsin has been described in Wiesner-Friedman *et al.* (2021: under review) and Beattie *et al.*[22,134] Overall, the study area has many potential sources of bovine and human fecal contamination and is made further vulnerable by its Karst geology.[152,176] The study area is ideal for a land-use regression study due to the availability of spatial databases representing potential microbial contamination sources. There was a total of 7 types of bovine sources with different database options: AFOs (4 options), manure application fields (2 options), land applied sludge (2 options), land applied sludge with industrial waste (3 options), wastewater treatment plants treating industrial wastes (1 option), low-intensity developed land cover (2 options), and ground transport of manure from AFO to application fields (8 options). There was a total of 6 types of bovine sources with different database options: septic systems (1 option), wastewater treatment plants (2 options), land applied sludge (2 options), land applied sludge from municipal wastewater and/or septage (3 options), high-intensity developed land cover (2 options), and the ground transport of treated wastewater or septage from septic systems locations to land applied sludge sites (3 options). These databases were obtained from personal communication with the Kewaunee County Land and Water Conservation Department, open records requests with the Wisconsin Department of Natural Resources (WIDNR), and from the WIDNR open data portal. The source types, candidate databases, database descriptions, and rationale for the source type being a concern for bovine versus human fecal pollution are described in detail in Table A.7.

To summarize the databases in relation to the river network, within a 500 m buffer to the river network are 12 dairy CAFOs, 169 manure storages associated with bovine AFOs, 193,000 km^2 of dairy rotation land-cover, 194 septic systems, 324 land-applied sludge sites, 17,400 km^2 of low-intensity developed land cover, and 3,250 km^2 of high-intensity developed land cover. Additionally, in Kewaunee County there are 3 wastewater treatment plants that treat primarily industrial wastes. The second major industry in Kewaunee County after producing milk is processing milk into dairy products. Additionally, 3 wastewater treatment plants treat waste from municipal systems. The potential sources of bovine-associated fecal contamination in question were AFOs, manure application, low-intensity developed land cover, and land applied sludge and the potential sources of human-associated markers in question were septic systems, wastewater treatment plants, high-intensity developed land cover, and land applied sludge.

4.3.4 Physically meaningful model for contamination from spatially distributed sources

To identify and characterize sources of microbial contamination and effects that modify their contributions, a microbial spatial land-use regression model that can associate contributions from precipitation and source terms and account for modifying effects on those contributions was utilized. Wiesner-Friedman *et al.* (2021: revision requested to ES&T).

$$\mathrm{y}_i = \beta_0 + P1_i(\beta_1 + \beta_2 P2_i) + \beta_3 Freezing_i + \left\{\sum_{u=1}^{U} \beta_u s_i^{(u)}\left(\alpha^{(u)}\right)\right\} \exp\left\{\sum_{l=1}^{L} \beta_l m_i^{(l)}\left(\alpha^{(l)}\right)\right\} + \varepsilon_i, \qquad \text{(Eq. 4.1)}$$

The observed response value y_i for sample *i* is a function of *climatic terms, source terms* and *modifier terms*. β_1, β_2 and β_3 are linear regression coefficients for the climatic variables $P1_i$ (recent precipitation), $P2_i$ (antecedent precipitation), and $Freezing_i$. β_1 represents the effect of recent precipitation on the microbial response and β_2 represents the diminished effect of antecedent precipitation, P2, on P1.[40] The source terms, $s_i^{(u)}\left(\alpha^{(u)}\right)$, are standardized to a mean of 0 and standard deviation of 1, so that a one-standard-deviation increase in the *u*th source term, leads to a β_u increase in the relative abundance response, y_i. Each *source term* was constructed using the Overland and River Distance—Flow (ORF) spatial predictor model described further in Wiesner-Friedman *et al.* (2021*).

$$s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right) = s_i^{(u)}\left(\alpha_O^{(u)}, \alpha_R^{(u)}\right) = z - score\left(\sum_{j=0}^{N} \frac{M_{0j}}{Q_i} \mathrm{e}^{\frac{-3D_{ij}{}^{(O)}}{\alpha_O} + \frac{-3D_{ij}{}^{(R)}}{\alpha_R}} \delta_{\overleftarrow{\overline{ij}}}\right) \qquad \text{(Eq. 4.2)}$$

An ORF source term for each i^{th} observation is the z-scored value that is sum of the mass of contaminants from each j^{th} source, M_{0j}, after decaying overland distances, $D_{ij}{}^{(O)}$, and river distances, $D_{ij}{}^{(R)}$, and accounting for flow, Q_i. Overland and downstream decay are characterized by hyperparameters, α_O, and α_R, respectively. $\delta_{\overleftarrow{ij}}$ indicates the flow connectivity between each *j*th source location and each *i*th observation location, so that it is a matrix that denoting the adjacency of nodes representing locations on a directed network.

Another type of spatial predictor model used to construct source terms was the Ground Transportation, Overland and River Distance—Flow (GORF) spatial predictor model.

$$s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right) = s_i^{(u)}\left(\gamma_G^{(u)}, \alpha_O^{(u)}, \alpha_R^{(u)}\right) = z - score\left(\frac{1}{Q_i}\sum_{j=0}^{N}\sum_{k=1}^{K}\omega_{kj}M_k \,\mathrm{e}^{\frac{-3D_{ij}{}^{(O)}}{\alpha_O}+\frac{-3D_{ij}{}^{(R)}}{\alpha_R}}\delta_{\overleftarrow{ij}}\right),$$

where $\omega_{kj} = \exp\left(\frac{-3D_{kj}}{\gamma_G}\right) / \sum_{j=1}^{J} \exp\left(\frac{-3D_{kj}}{\gamma_G}\right)$ (Eq. 4.3)

A GORF source term each *i*th observation is the z-scored value that is a result of using an ORF term, where the mass of contaminants from each *k*th source location, M_k, is distributed across each *j*th land application location. This depends on the spatial relationships between each *k*th source location and each *j*th land application location described by ω_{kj}, The hyperparameter, γ_G, acts as a toggle that determines how near to the source the contaminant is applied. If γ_G is infinite, then the distribution of the contaminant becomes homogenous across application locations.

ORF spatial predictor models were used to construct source terms (see Table A.7 for detailed descriptions of source types (*u)* and representative databases) for bovine-associated responses representing AFOs (*u*=1), manure application (*u*=2), low-intensity developed land cover (*u*=9), land applied sludge (*u*=5), and land applied sludge with industrial waste (*u*=7). Additionally, the GORF spatial predictor model was used to construct source terms characterizing the contributions that AFOs make to the river network *via* the ground transport and application of manure onto fields. For human-associated responses ORF source terms were used to represent septic systems (*u*=3), wastewater treatment plants (*u*=4), high-intensity developed land cover (*u*=8), land applied sludge (*u*=7), and land applied sludge with non-industrial waste (*u*=6). Additionally, the GORF spatial predictor model was used to construct source terms characterizing the contributions that septic systems make to the river network *via* the ground transport and application of treated septage at land applied sludge sites.

4.3.5 Implementation of the microbial Find, Inform, and Test (FIT) framework

The FIT framework was implemented using the physically meaningful model for contamination from spatially distributed sources (Eq. 4.1) to identify and characterize sources of microbial contamination for *each* of the 4 microbial contamination responses–1) the relative abundance of bovine-associated markers in sediment 2) the relative abundance of bovine-associated markers in surface water 3) the relative abundance of human-associated markers in sediment, and 4) the relative abundance of human-associated markers in surface water. Described below are details of the implementation of FIT that were applied to *each* of the 4 microbial contamination responses. The *microbial Find Inform and Test framework* is described in further detail in Wiesner-Friedman *et al.* (2021).

For the *Find* stage, reliability scores for each *d* database for a given source type *u* were obtained. Reliability scores help to represent databases that reliably represent a source (i.e., association with response is positive) for the contaminant represented by unseen (i.e., a test set) of unseen response data. This is accomplished by selecting decay hyperparameters by maximizing the coefficient for a source term constructed with ORF or GORF spatial predictor models from the regression model defined in Eq. 4.1 using a training set of data. The reliability score has three components. The sign stability score (*SSS*), which indicates the number of models of test set data where the source term constructed from training data assumptions is positively associated. The average magnitude, *M,* indicates the sum of the quantified associations between test set data and source term constructed from training data assumptions. Lastly, the standard deviation on the regression coefficient (i.e., of the source term constructed from training data assumptions), σ_β, was obtained. Across each u^{th} source type, the d^{th} database with the highest reliability score was selected.

After reliable databases of spatially distributed contamination sources were selected for each response, each source term was *informed* individually by using an objective function that selects the hyperparameter $\hat{\boldsymbol{\alpha}}^{(u)}$ that maximizes the β_u using 100% (i.e., all of the data) of the observations y_i, using a penalty on a combination hyperparameters that yield poor regression or mapping qualities (i.e., non-normal residuals or $\hat{y}_i > 0$). To maximize the objective function, the MATLAB function, patternsearch, was used as it is more computationally efficient for selecting hyperparameters with the objective function compared to fmincon, which was used in Wiesner-Friedman *et al.* (2021*). The objective function was also visualized in the hyperparameter space to determine upper and lower bound values for α_O, α_R, and γ_G that would better capture a more global maximum (see A2.5). Collinearity between source terms was assessed with a Pearson correlation coefficient. If source terms were correlated ($\rho \geq 0.7$), then a source term was chosen from among those correlated that had a higher univariate R-squared and a positive coefficient (i.e., a positive coefficient defines a source). [184]

After source terms were constructed, the predictive ability of climatic terms and the non-collinear, informed source terms were *tested* with a stepwise regression approach using the Akaike Information Criterion. A 6-fold cross validation of the stepwise selection procedure was used to evaluate the inclusion of source terms in case of influential observations.

4.4 Results and Discussion

4.4.1 Detection of host-associated markers in sediment and surface water varies significantly across sampling events likely due to precipitation or freezing temperature

Figure 4.1 is a comparison of the relative abundance of human and bovine-associated markers across different times sampled from sediment and surface water. High variability was observed within each sampling time during non-freezing sampling times that will allow for capturing spatial variability. There is also between-day variability for February and May sampling for which there are significant differences in median relative abundances across sample types and host-associated markers. There is also smaller within day variability observed for the February sampling time. The values displayed in Figure 4.1 for the relative abundance of human-markers in February are constructed by setting the below detect values to be equal to 0.5 gene copies prior to normalizing by 16S rRNA and log10 transforming.

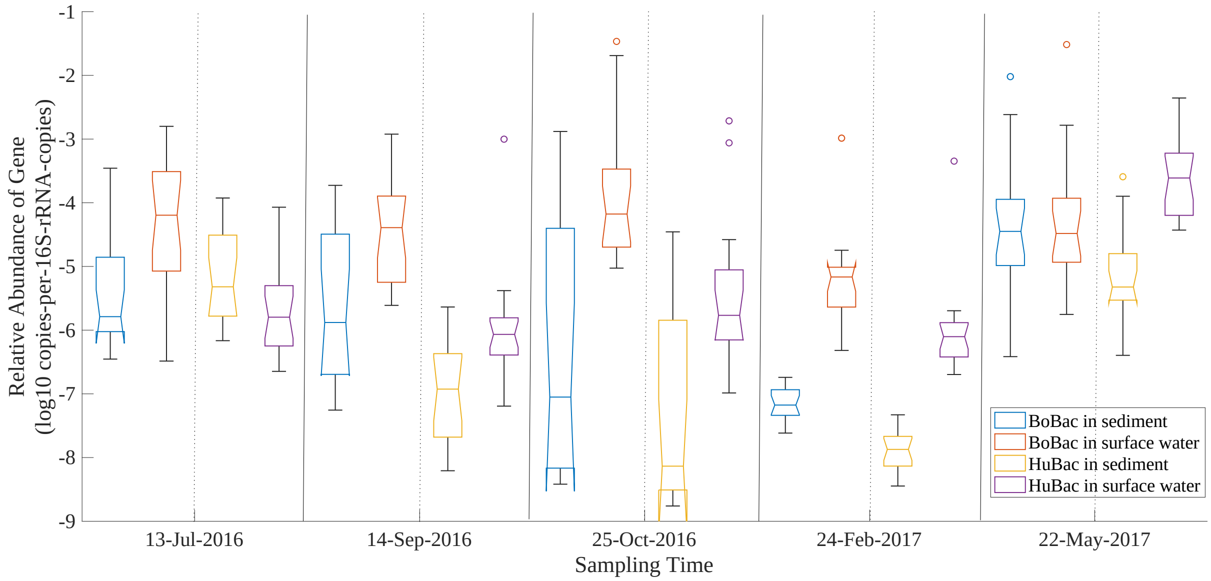


Figure 4.1 Distribution of relative abundance values across sampling time for each response, *BoBac* and *HuBac*, and for each sample type (i.e., sediment and surface water). The boxes show the upper and lower quartiles for each response and sample type. The middle line of the box represents the median. The tapered and shaded area around the middle line shows the 95% confidence interval around the median.

Climatic factors that vary across sampling times are significantly associated with the relative abundance of host-associated markers. The results of the *Test* stage of FIT for the relative abundance of bovine and human associated markers can be found in Tables 4.2 and 4.3, respectively. These tables show the results from the stepwise-selected climatic and source terms. When the average monthly temperature was freezing, the relative abundances of bovine-associated markers (log10 copies-*BoBac*-per-16S-rRNA-copies) in sediment and surface water were reduced by 3.26 (p-value<0.05) and 1.02 (p-value<0.05), respectively. For human-associated markers (log10 copies-*HuBac*-per-16S-rRNA-copies) in sediment, freezing temperature was associated with 0.835 reduction (inclusion in model reduced the AIC).

A one-standard-deviation increase in recent precipitation, P1, was associated with a 0.996 (p-value<0.05) increase the relative abundance of bovine-associated markers (log10 copies *BoBac*-per-16S-rRNA-copies) in sediment. For a one-standard-deviation increase in antecedent precipitation, P2, the effect of recent precipitation was diminished by 0.259 (p-value<0.10). This pattern of contributions from precipitation was also found for both human-associated responses (Table 4.3). A one-standard-deviation increase in recent precipitation was associated with increases of 0.580 (p-value<0.05) and 1.26 (p-value<0.05) in the relative abundance of human-associated markers (log10 copies *HuBac*-per-16S-rRNA-copies) in sediment and surface water, respectively. For a one-standard-deviation increase in antecedent precipitation the effect of recent precipitation on the relative abundance of human associated markers in sediment and surface water was diminished by 0.539 (p-value<0.05) and 0.273 (p-value<0.05).

Our findings add to existing evidence that precipitation is strongly associated with increases of host-associated markers in rivers and streams[29,30,82,103,174,182,183] One study of *HuBac* markers in inland waters did not find an association with precipitation.[174] However, one difference in our study is that the diminishing effect of antecedent precipitation on recent precipitation was modeled. The effect of this modeling is that fecal pollution contributions from recent precipitation can be diminished by either short periods of intense rainfall preceding the recent precipitation event or long periods of moderately wet weather. This process has been previously designated as a *flush effect*.[40,185–188] To the best of our knowledge this is the first work to report a flush effect across human- and animal-associated fecal responses, as well as across sediment and surface water.

Table 4.2 Regression results for predicting the relative abundance of bovine *Bacteroides* in sediment (log10 *BoBac*-copies-per-16S-rRNA-copies) and the relative abundance of bovine *Bacteroides* in surface water (log10 *BoBac*-copies-per-16S-rRNA-copies). For each of the responses there are columns representing the Inform and Test stages of the FIT framework. Hyperparameters, **α,** standardized regression coefficients, β**,** and the relative abundance ratio (RAR). Rows correspond to climatic and source terms. The last two rows provide the model fit as expressed by R^2 and adjusted R^2. The table only displays consistencies for hyperparameter values, **α**.

Response	Bovine *Bacteroides* in Sediment (*n=90*)			Bovine *Bacteroides* in Water (*n=98*)		
Variables	Hyperparameters **α**	Coefficients β	RAR	Hyperparameters **α**	Coefficients β	RAR
Climatic						
P1	$\alpha_{P1} < \alpha_{P2}$	0.996**	9.90	$\alpha_{P1} < \alpha_{P2}$	NS	NS
P1 x P2		-0.259*	0.551		NS	NS
Freezing	NA	-3.26**	5.50e-4	NA	-1.02**	0.0955
Source						
Low Intensity Developed	$\alpha_O > 1$ km	0.318**	2.08	$\alpha_O < 1\,km$	0.142	1.39
AFO	NS	NS	NS	$\alpha_O < 1\,km$	0.253**	1.79
AFO (via ground transport to Manure Fields)	γ_G=13 km, $\alpha_O > 1$ km	0.298**	1.99	NS	NS	NS
Adjusted R^2	0.455			0.204		
R^2	0.485			0.228		

Also not selected: Manure fields (no ground transport from AFO), Land-applied sludge (all types or industrial types), industrial wastewater treatment plants
**p-value<0.05, *p-value<0.10

4.4.2 AFOs are a source of bovine fecal contamination.

Table 4.2 provides results from FIT for bovine-associated responses. The FIT model selected the GORF AFO spatial predictor representing the dissemination of manure generated at AFOs and transported to application fields into riverbed sediment. Results indicate that a one-standard-deviation increase in contributions from AFOs via the ground transport and application of manure on fields was associated with a 99% ($p\text{-value}<0.05$) increase in the relative abundance of bovine-associated markers (copies-*BoBac*-per-16S-rRNA-copies) in *sediment*. For surface water, the FIT model selected the ORF AFO spatial predictor representing the contributions of AFO site directly to the river (i.e., not transported to distant manure fields). A one-standard-deviation increase in AFO contributions was associated with a 79% ($p\text{-value}<0.05$) increase in the relative abundance of bovine-associated markers (copies-*BoBac*-per-16S-rRNA-copies) in *surface water*.

We find that this is a higher value than previously estimated in Wiesner-Friedman *et al.* (2021: under review by ES&T). The greater magnitude of this association in sediment here compared to Wiesner-Friedman *et al.* is due to the inclusion of low-intensity developed land cover in the model. The only land-use regression study to associate bovine sources to bovine-associated markers in rivers is Dila *et al.*, who found a one-standard-deviation increase in cattle density at the watershed level was associated with a 20.2% ($p\text{-value}<0.05$) increase in the absolute abundance of bovine associated markers (*BacR*-copies) in *surface water*.[30] Factors to consider in comparing our estimates to Dila *et al*. are outlined in Wiesner-Friedman *et al.* (2021: under review).

Differences in the selection of the AFO spatial predictor may reflect where and how manure is applied depending on the scale of the AFO. Manure application practices may be affected by the scale of an AFO based on different costs.[129] Additionally, manure application practices, which may vary based on the size of the AFO, may contribute to differences in the detection and persistence of host-associated markers and fecal indicators in soil versus sediments versus surface waters.[189–191]

4.4.3 Low-intensity developed land-cover is another source of bovine fecal contamination

The FIT model selected the ORF low-intensity developed land cover predictor which represents contributions from land classified as low density of impervious/manmade features. A one-standard-deviation increase in this land cover was associated with a 108% ($p\text{-value}<0.05$) increase in the relative abundance of bovine-associated markers (copies-*BoBac*-per-16S-rRNA-copies) in *sediment.* To our knowledge, this is the first study find a positive association between low-intensity land-cover and bovine-associated markers in sediment. A few studies have explored and revealed relationships between low-intensity land-cover and *fecal indicator bacteria* in freshwater.[29,35,173,175,192] Some have found differences in associations between low-intensity land-cover and comparable land cover such as built-up areas like parking lots or farm yards and fecal indicator bacteria.[35,173] In an area strongly impacted by industrial livestock operations, Alford *et al.* found that a percentage point increase in low-intensity developed land cover increased fecal coliforms by 5% in rivers.[173]

Other studies have also explored relationships between bovine fecal markers and low-intensity land-cover or impervious surfaces, but not found any statistically significant associations.[29,175,192] However, these study areas were not located in livestock dense regions. An important consideration for modeling using this land cover is that the land cover may not capture the same phenomena for different regions. In Kewaunee County, low-intensity developed land cover captures an intermediate rate of manure application that are due to two phenomena unique to dense livestock farming in populated areas (see Figure 4.2), which are 1) the application from small farms located in low-intensity developed land and 2) the application on nearby large farms that migrate off site due to small imperviousness when these large farms are near low-intensity developed land. Bovine-associated markers have been shown to persist in cow pats past a month, disseminate from cattle grazing areas, and more recently, be released from manure into aquatic settings over days.[193–195] From these findings, it is possible that the sum of the contributions from small farms may lead to substantial increases in fecal contamination in riverbed sediment. Additionally, impervious surfaces would reasonably increase conveyance to the river network and reduce the capacity for soil filtration for contaminants from nearby manure application from small or large farms.

For *surface water*, we similarly find that low-intensity developed land cover is a source of bovine fecal contamination based on AIC (i.e., low-intensity developed land cover was selected based on the fact that its inclusion reduced the AIC). A one-standard-deviation increase in this land cover was associated with a 39% increase in the relative abundance of bovine-associated markers (copies-*BoBac*-per-16S-rRNA-copies). However, this association was not statistically significant and indicates that the sample size is not sufficient for this response.

Figure 4.2 Example of low-intensity developed landcover highlighted in yellow overlaid on a small farm with some impervious surfaces located next to an AFO manure storage (not highlighted).

4.4.4 Septic systems are a likely source of human-associated fecal contamination.

Surface water contamination associated with septic systems was expected due to the large body of research that indicates that septic systems commonly pollute groundwater and streams, using a variety of methods and in similar and different regions.[37,82,103,168,169,175,192] In this study the ORF septic system predictor representing drainfield polygon centroids was selected to model human-associated fecal pollution in surface water. A one-standard-deviation increase in contributions from septic systems was associated with a 79% (p-value<0.10) increase in the relative abundance of human fecal markers (*HuBac*-copies-per-16S-rRNA-copies). For fecal contamination in *riverbed sediment*, the FIT model also selected the ORF septic system predictor. A one-standard-deviation increase was associated with a 56% (inclusion lowered AIC) increase in the relative abundance of human-associated markers (*HuBac*-copies-per-16S-rRNA-copies).

Two studies of other human-associated markers have found that an increase in 1 septic system per kilometer was significantly (p-value<0.05) associated with a 1% or 18.6% increase in the absolute abundance of human fecal markers (e.g., HF183 or HumM2 copies-per-100mL-water).[37,82] Proximity to the nearest septic system has also been found to be an important factor for human fecal pollution in surface water.[37,168] A distance *increase* of 1 kilometer for a septic system to a stream sampling site was associated with a decrease of 0.003 (p-value<0.05) in the absolute abundance of human-associated markers (log10 HF183-copies-per-100mL-water).[37] This may imply that a distance decrease of 1 kilometer corresponded to a 1% increase, approximately.

To the best of our knowledge, ours is the first study to model the associations between septic system contributions and human fecal pollution in *sediment.* Furthermore, this is also the first study to characterize contributions from septic systems with an ORF spatial predictor model, which shows how overland and downstream flow transport are involved in how septic systems contaminate the environment. This work adds to the large body of work that has found associations between septic systems and human-associated markers in surface water and points to widespread septic system failure,[37,82,103,169,175,192] especially in regions (e.g., Kewaunee County) with fractured rock and Karst geology.[152,153,175,176,196–198] The ability for septic tanks to pollute is based on a variety of factors ranging from geological features of the region to the age and type of system.[167] This calls for innovation in small-scale or decentralized wastewater treatment for rural areas, especially those with Karst geology.

Table 4.3 Regression results for predicting the relative abundance of human *Bacteroides* in sediment (log10 *HuBac*-copies-per-16S-rRNA-copies) and the relative abundance of human *Bacteroides* in surface water (log10 *HuBac*-copies-per-16S-rRNA-copies). For each of the responses there are columns representing the Inform and Test stages of the FIT framework. Hyperparameters, **α,** standardized regression coefficients, β, and the relative abundance ratio (RAR). Rows correspond to climatic and source terms. The last two rows provide the model fit as expressed by R^2 and adjusted R^2.

Response	Human *Bacteroides* in Sediment (*n=90*)			Human *Bacteroides* in Water (*n=79*)		
Variables	Hyperparameters **α**	Coefficients β	RAR	Hyperparameters **α**	Coefficients β	RAR
Climatic						
P1	$\alpha_{P1} < \alpha_{P2}$	0.580**	3.80	$\alpha_{P1} < \alpha_{P2}$	1.26**	18.2
P1 x P2		-0.539**	0.290		-0.273**	0.533
Freezing	NA	-0.835	0.146	NA	NA	NA
Source						
Septic Systems	$\alpha_O > 1\ km$	0.170	1.56	$\alpha_O <$ 1 km	0.178*	1.51
High Intensity Developed	$\alpha_O > 1\ km$	0.193	1.47	NS	NS	NS
Adjusted R^2	0.358			0.500		
R^2	0.394			0.519		

**p-value<0.05, *p-value<0.10
Also not selected: Wastewater treatment plants, Land-applied sludge (all types or wastewater/septage)

4.4.5 High-intensity developed land-cover as a source of human fecal pollution in sediment.

For the model of human-associated pollution in sediment, FIT selected the ORF high-intensity developed land cover spatial predictor which characterizes a high density of impervious surfaces. A one-standard-deviation increase in high-intensity developed land-cover was associated with a 47% (inclusion lowered AIC) increase in the relative abundance of human fecal markers (*HuBac*-copies-per-16S-rRNA-copies). High-intensity developed land-cover has been previously associated with human fecal pollution in groundwater or surface water.[29,90,174,175]

Two studies have found associations between high-intensity developed/high density of impervious land cover and human-associated fecal contamination in surface water. One study also associated developed land-cover with *HuBac* markers and found that percent impervious and percent developed land-cover were positively and significantly associated with raw copies numbers of *HuBac* in inland waters using a nonparametric Spearman rank test.[174] In addition to this, an increase of 1% in developed land-cover within radii of 1 to 2km, but not 0.5km, was associated with approximately a 770% (credible interval suggests significance) increase in the detection rate of human-associated markers (mean presence of HF183 and HumM2) in *surface water* from the Rio Grande.[90] To our knowledge, this is the first study to report this association in *riverbed sediment.*

Because our first finding for human-associated markers implies the failures of septic systems, the association with high-intensity developed land cover suggests that highly impervious surfaces in proximity to polluting septic systems may convey fecal pollution to the river network without the benefit of microbial reductions through soil infiltration. This warrants careful attention to the maintenance of septic systems and the placement of impervious surfaces and septic systems relative to each other.

4.4.6 Higher host-associated marker specificity leads to better signals from sources

From results of the *find* stage of FIT, available databases for bovine sources more sufficiently represented reliable databases of spatially distributed sources compared to available databases of human sources. Figure 4.3 summarizes the success rate of a candidate database representing a source of contamination for unseen response data (i.e., database-as-source success rate) across all databases options. This was calculated by averaging the sign stability scores calculated in the *Find* stage and dividing them by the number folds that were used. The database-as-source success rate was greater for the bovine-associated responses (57.3%) compared to human-associated responses (50.8% and 33.8%). Several factors may contribute to this difference. Databases representing potential sources of human fecal contamination were usually of a smaller scale and more numerous compared to the databases representing potential sources of bovine fecal contamination (e.g., septic system versus AFO). More sampling sites may be needed to characterize contributions from smaller and more abundant sources. On the response-side of the matter, the higher specificity of the *BoBac* compared to *HuBac*[136] may lead to better signals from sources. Alternatively, the pathways of fecal contamination from sources may create scenarios where detection of bovine feces is greater than detection of human feces, especially in areas where there are more bovine than humans.[199]

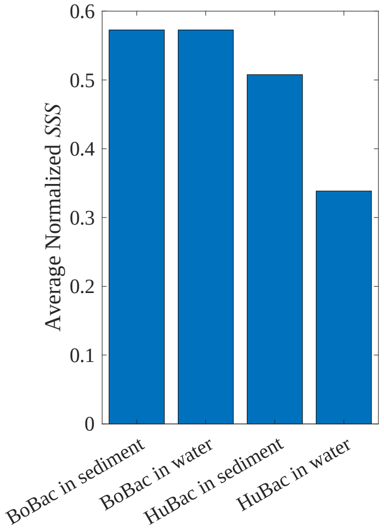


Figure 4.3 The average normalized sign stability score (*SSS)* across candidate databases for each of the four responses. The average normalized *SSS* represents a database-as-source success rate. The database-as-source success rate captures one element of the reliability score for *Finding* reliable databases of spatially distributed sources. The goal of the reliability score is determine how well the spatial predictor model hyperparameters work to construct spatial predictors that are positively associated (i.e., source terms) with unseen data. This is done for k=5 folds (i.e., sets of already modeled/unseen data). The *SSS* is equal to the number of times that the spatial predictor corresponded to a source term. The average normalized *SSS* is the *SSS* normalized by the number of folds. This process can be described as a k-fold cross-validation on the regression coefficient positivity.

4.4.7 Sediment responses better capture signals from distant sources than surface water

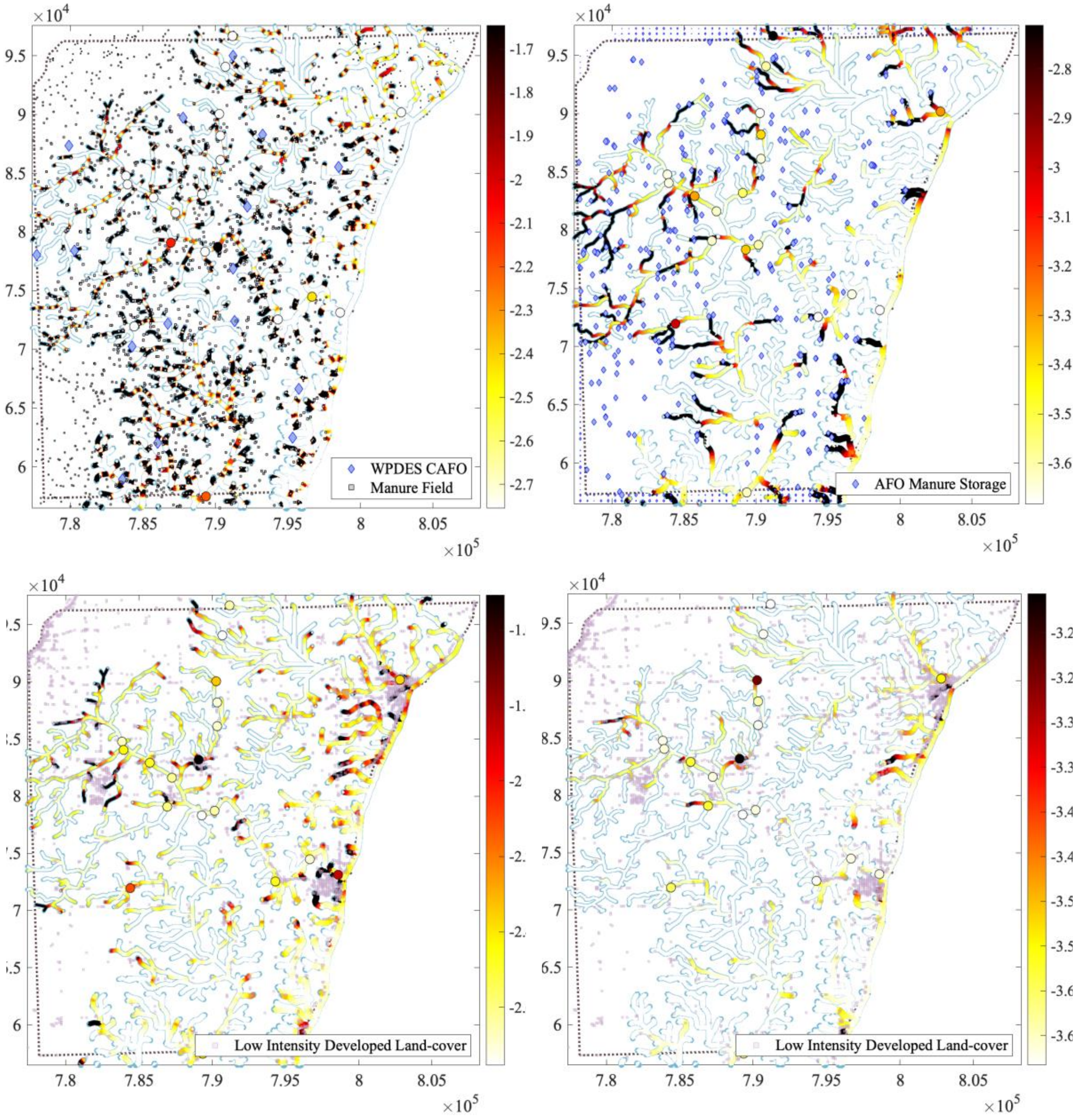
As a result of the *Inform* stage of FIT, hyperparameter values were selected to reveal some information about pathways of contamination from sources. The hyperparameter values selected to inform the source terms are available in Table 4.2 (bovine-associated responses) and Table 4.3 (human-associated responses). One consistency observed in the hyperparameter values was that the overland flow hyperparameter, α_O, was longer for responses measured in sediment compared to surface water. From the process of the *inform* stage, which selects based on maximum RAR, we find that the RAR is greatest for pollution responses in sediment when overland distance decay ranges, α_O, are greater than 1 km. Compared to pollution responses in surface water, the greatest RAR occurs when where α_O are all less than 1 km. The spatial depiction of this process would show sediment contamination reaching the river network from sources that are far away versus surface water contamination reaching the river network from sources that are close to the river.

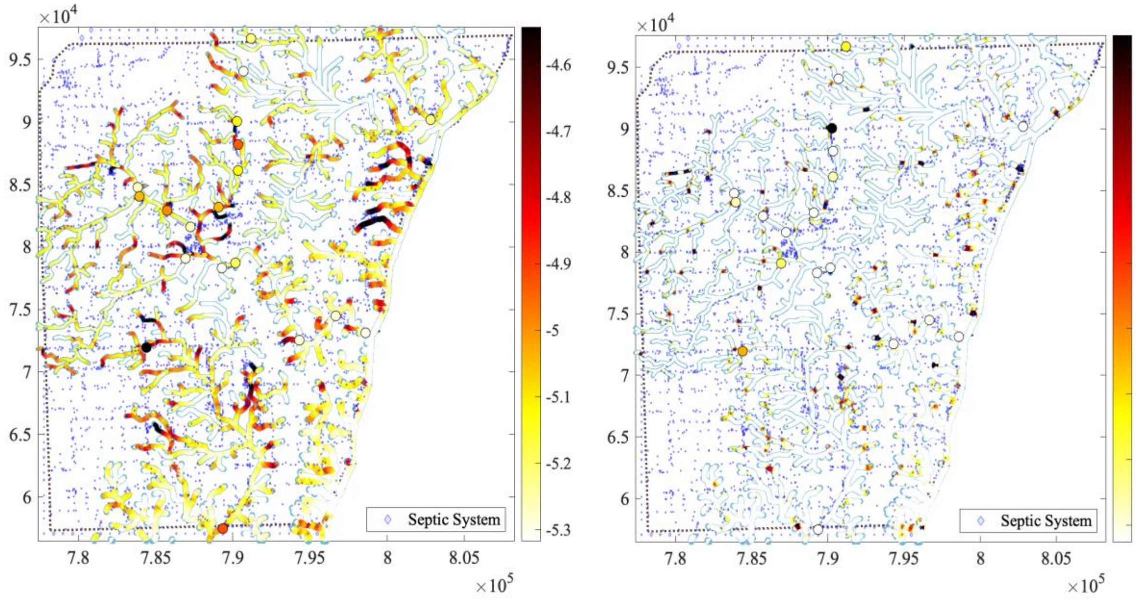
However, the hyperparameter, α_O, represents average overland transport processes. There are many transport mechanisms that could explain why what ends up in the water *travels* over much shorter distances overland than what ends up in the sediment. Possible explanations are that how manure is applied on land differs (e.g., broadcasting, incorporation, or injection), and/or that the transport occurs over different pathways (e.g., sheet flow or runoff,[190,200,201] or alternatively sub surface flow through various geological features including karst,[202–204] or tile drainage,[189,202,205] or even wind[206]). Tyrrel and Quinton *et al.*[207] also hypothesized that different attachment or non-attachment to soil or waste particles might influence transport. Information about manure application trends at sources, precise sampling methods, and larger samples sizes are needed to differentiate transport pathways overland.

Additionally, the hyperparameter, α_O, acts as a toggle for the number of upstream sources included in the estimation of source contributions to the river network. In this sense, α_O not only characterizes average transport overland but represents the extent to which a signal is received from distant sources when using responses that are host-specific and highly associated with sources. In this sense, the relationship to the source depends on the response data and the extent to which source-signals are received can be influenced by differential degradation of the response in environmental media (i.e., water, soil, or sediment). Differences in gene degradation have also been observed for host-associated markers for different sunlight, nutrient, and vegetation conditions.[23,193,208–211] In general, host-associated markers are known to persist longer in sediments compared to surface water.[23,208] This is supported by our study by the average database-as-source success rate for the human-associated response in sediment versus water, which were 50.8% and 33.8%, respectively. Lastly, as mentioned in the last section, the specificity of the response to the type of source may also lead to better signals from sources.

The longer α_O for responses from sediment versus surface water found in this study may therefore indicate a better ability to capture a signal from sources in sediment, rather than water samples. Possible explanations may have to do with the settling of microbes into sediments or the characteristics of the molecular target.[212] Host-associated markers often represent obligate anaerobes (e.g., *BoBac* and *HuBac* represent *Bacteroides*) and would not be expected to survive long in higher oxygen levels, like water. This implies that α_O captures much more than average overland transport distances. This transport hyperparameter represents both an exponential influence range around sources that portrays both average overland transport of pollution from sources and how specific characteristics of the molecular target may be affected during this transport.

Consequently, quantitative microbial risk assessments based on the measurement of genes representing obligate anaerobes or quantifying responses from only surface water may not fully capture the extent to which microbial communities and function are associated with sources. This work indicates that molecular targets used to quantify risks in surface water should be carefully selected or sediment sampling should be considered alongside water samples to better estimate and mitigate public health risks. Additionally, settled sediments can be resuspended into surface water and microbial risks have been estimated to increase under conditions that disturb the sediment.[58,213,214] Therefore a better understanding of factors contributing to variability in sediment microbial communities can help to better estimate risks from sediment resuspension.


×10^4
9.5
9
8.5
8
7.5
7
6.5
6
7.8
7.85
7.9
7.95
8
8.05
×10^5
-1.7
-1.8
-1.9
-2
-2.1
-2.2
-2.3
-2.4
-2.5
-2.6
-2.7
WPDES CAFO
Manure Field
-2.8
-2.9
-3
-3.1
-3.2
-3.3
-3.4
-3.5
-3.6
AFO Manure Storage
-1.
-1.
-2
-2.
-2.
-2.
Low Intensity Developed Land-cover
-3.2
-3.2
-3.3
-3.3
-3.4
-3.4
-3.5
-3.5
-3.6
-3.6
Low Intensity Developed Land-cover



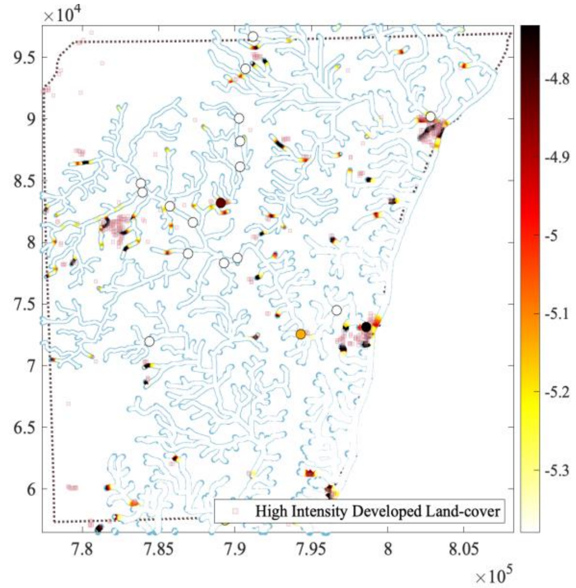


Figure 4.4 Modeled relative abundance host-associated fecal contamination (log10 gene-copies-per-16S-rRNA-copies) plotted on the river network for Kewaunee County given average recent and antecedent precipitation and freezing temperature.

a) AFO contributions the relative abundance of *BoBac* in sediment b) low intensity developed land-cover contributions to the relative abundance of *BoBac* in sediment c) septic system contributions to the relative abundance of *HuBac* in sediment d) high intensity developed land-cover contributions to the relative abundance of *HuBac* in sediment e) AFO contributions to the relative abundance of *BoBac* in surface water f) low intensity developed land-cover contributions to *BoBac* in surface water g) high intensity developed land-cover contributions to *HuBac* in surface water

4.4.8 Implications, limitations, and future work

This study is the first to find that AFOs contribute to fecal contamination in both sediments and surface waters by characterizing dilution processes from flow, the density of upstream AFOs, and the proximity of upstream AFOs. The observed differences in the portrayal of sources from FIT (i.e., WPDES CAFO ground transport to manure fields versus manure storages representing AFOs of any size) suggests that differences in manure application practices may be key to understanding how sediment and surface water are differently impacted by AFOs.

This study also indicates that when low-intensity land cover is accounted for, AFOs appear to have a greater effect on fecal contamination.

Low intensity developed land cover was considerably associated with fecal pollution. This highlights that small farms and/or low impervious surfaces may represent moderate bovine fecal applications and/or accumulations. Alford *et al.* has previously suggested that land cover gradients between primarily human residences and livestock-dense areas are associated with increased fecal indicators in surface water.[173] This work which used GORF/ORF spatial predictor models to describe contributions from low-intensity developed areas indicates that flow processes may be key to this relationship.

In Figure 4.4, the modeled associations between source terms and responses were applied to the whole river network under average precipitation and temperature scenarios to depict an estimate of contributions to the relative abundance of bovine fecal pollution (log10 *BoBac*-copies-per-16S-rRNA-copies) in riverbed sediment and surface water from AFOs (Figure 4.4a,e) and low-intensity developed land cover (Figure 4.4b,f). The depictions of bovine fecal contamination in sediment and surface water associated with *AFOs* show contributions to bovine fecal contamination away from semi-urban and urban areas (roughly depicted by higher densities of low-intensity and high-intensity land cover centroids). In contrast, the depictions of bovine fecal contamination in sediment and surface water associated with *low-intensity developed land cover* show contributions to bovine fecal contamination that tend to flow from rural areas toward more semi-urban and urban areas. Overall, the combination of contributions from AFOs and low-intensity developed land cover describes a flow pattern of fecal contamination in sediments and surface water from rural areas in the direction of more residentially developed areas.

In Figure 4.4, the modeled associations between source terms and responses were applied on the whole river network under average precipitation and temperature scenarios to depict an estimate of contributions to the relative abundance of human fecal pollution (log10 *HuBac*-copies-per-16S-rRNA-copies) in riverbed sediment and surface water from septic systems (Figure 4.4c,g) and in surface water from high-intensity developed land cover (Figure 4.4d). The depictions of contributions from high-intensity developed areas are localized to areas around towns and cities (i.e., Casco, Luxemburg, Kewaunee, and Algoma) as well as residential developments and shopping centers. In surface water, contributions from septic systems are also highly localized around dense clusters of rural residences. However, contributions from septic systems to sediment flows away from residential developments towards primarily rural areas. Largely, this constitutes a flow pattern of human-associated fecal contamination from semi-urban and urban areas towards rural areas.

Since antibiotic resistance can be passed from livestock-associated bacteria to human pathogens through a variety of cellular mechanisms,[215] the prevalence of antibiotic resistant human pathogens may increase around livestock associated fecal contamination. Given the associations between fecal contamination and pathogens, the confluence of fecal contamination from bovine and human sources in locations of agricultural and recreational water use is a major public health concern and a One Health issue.[5] One Health research frameworks (i.e., research frameworks that focus on human, animal, and environmental health) are needed to better understand the risks associated with surface water and sediment for buffers between dense industrial livestock agriculture and semi-urban/urban areas.

Flow patterns (i.e., river connectivity) should be considered along with microbial risk and river ecology under the scenarios of population growth and the expansion of residential developments to rural areas.[216] This warrants thoughtful design and planning in both urban *and* rural settings, especially as it relates to the placement of impervious surfaces and on-site rural wastewater management systems such as septic systems, manure storages, and manure application fields. As a whole, this study has demonstrated that this region is impacted by fecal contamination. This necessitates improvements in current rural wastewater infrastructure as well as futuristic designs for land-use and wastewater infrastructure in rural areas, especially those with livestock agriculture and Karst geography.

CHAPTER 5 AIM 3: BOVINE SOURCE AND LAND APPLIED SLUDGE FROM SEPTAGE AND MUNICIPAL AND INDUSTRIAL WASTE ARE SOURCES OF ELEVATED ANTIBIOTIC RESISTANCE GENES IN RIVERBED SEDIMENTS AND SURFACE WATER

5.1 Overview

Antimicrobial resistance (AMR) is a major public health concern. A large body of work has identified potential sources of elevated AMR in the environment due to the dissemination of antibiotic resistant bacteria, antibiotic resistance genes, antibiotics, and other antimicrobials. Culture-independent quantitation such as the quantifying of antibiotic resistance genes has been increasingly used to detect elevated AMR and evaluate microbial risks. Few studies exist that have use ARG levels in the environment to identify key sources contributing to elevated AMR. The *Find, Inform,* and *Test* (FIT) framework which leverages a physically meaningful land-use regression model was applied to the relative abundance of 5 antibiotic resistance genes (gene-copies-per-16S-rRNA) from sediment and surface samples to *Find* reliable databases representing spatially distributed sources of ARGs, *Inform* source terms with hyperparameters that characterize average transport processes and exponential ranges of influence around sources, and *Test* the informed source terms through a stepwise selection of regression terms. A one-standard-deviation-increase in contributions from bovine sources was associated with 58% (p-value<0.10), 80% (inclusion lowers AIC), and 34% (inclusion lowers AIC) increases in the relative abundances of *erm(B), tet(W),* and *sul1* in riverbed sediment, respectively. A one-standard-deviation-increase in contributions from bovine sources was associated with 33% (p-value<0.10), 77% (p-value<0.05), 49% (p-value<0.05), 41% (p-value<0.10), and 36% (inclusion lowers AIC) increases *erm(B), tet(W), qnrA, sul1,* and *intI1* in surface water, respectively. A one-standard-deviation increase in contributions from land applied wastes (e.g., septage, and municipal and industrial waste) was associated with 47% (p-value<0.10), 34% (inclusion lowers AIC), 36% (p-value<0.10), and 41% (p-value<0.10) increases in the relative abundance of *sul1* and *intI 1* in sediment and *tet(W)* and *intI1* in surface water, respectively. This is the first study to identify and characterize source contributions for a panel of more than 2 ARGs. The work is the first to find an impact from land applied wastes on elevated ARGs measured from the environment. For the first time, influence ranges around sources were characterized. These produced the first-ever depictions of scenarios of AFO and land applied waste contributions to elevated AMR. Overall, sources of elevated ARGs are characterized by the application of treated or untreated biosolids and semi-solids from bovine, human, and industrial wastewater. This work demonstrates the ability for FIT to derive potential exposure variables in studies associating antibiotic resistant infections and colonization to agriculture or recreational water use. Ecologically, this work suggests that ARG levels vary in proximity to dense upstream sources and are affect by flow in both riverbed sediment and surface water. The results of this study reinforce the need for investments in both human and animal wastewater infrastructure in rural areas.

5.2 Introduction

The widespread use of both natural and artificial antibiotics has contributed to the substantial reduction in human disease over the past centuries. The overuse and uncontrolled release of antibiotics has led to the proliferation of antimicrobial resistance (AMR), microbes unaffected by antibiotics.[217] These microorganisms and associated antibiotic resistance genes (ARGs) can be dispersed via runoff into rivers and streams, ultimately exposing humans to the threat of antibiotic resistant bacteria.[218] Furthermore, ARGs can be acquired and forfeited by bacteria through horizontal gene transfer, which introduces the potential for non-resistant pathogens to acquire resistance. A large body of work has outlined potential sources where antibiotic resistant bacteria and genes have been detected in elevated levels compared to controls.[16,17,34,64,70,72,77,81,109,141,219–222] However, little work has been done to quantify the association of AMR levels in the environment and those potential sources. Land-use regression has been used to identify sources of contaminants in the environment. The application of land-use regression studies can increase ecological understanding of how sources influence levels of antimicrobial resistance and can be informative to assessing microbial risk associated with using water with elevated levels of antimicrobial resistant bacteria and genes.

AFOs have been found to a particularly important source of antimicrobial resistance, due to practices of using antibiotics for disease treatment and prevention, often at sub-therapeutic doses.[16,17,34,70,72,109] Other potential sources of antibiotic resistant bacteria (ARB) or sources of antibiotics, antimicrobials, and other contaminants that exert a resistance-oriented selective pressure on microbial communities include wastewater treatment plants,[64,77,141] and treated and untreated biosolids from agricultural, industrial, or municipal wastes.[219–222] Additionally elevated levels of ARB have been detected in groundwater near to septic systems.[81] Apart from point sources, other factors determined to be important to AMR in surface waters and sediments are season, rainfall, and land cover.[165]

AMR responses can be quantified with phenotypic testing from recovered isolates grown in media (i.e., culture-dependent methods) but can be seasonally biased when cold temperatures and stress lead bacteria into viable but non-culturable states (VNCS). Alternatively, culture-independent, molecular methods, such as qPCR, offer a quantification of AMR relative to antibiotic resistance genes (ARG) copies which will be less influenced by VNCS. [223] These culture-independent methods for gathering microbial responses data are also considered faster alternatives to culture-based methods.[212]

Few studies have aimed to associate sources with antibiotic resistance responses derived from culture-independent methods. As far as is known, only two studies have previously modeled the association between antibiotic resistance genes and potential sources of antimicrobial resistance. Pruden *et al.*[17] found that averaged upstream capacities of AFOs and contributions from upstream capacities of wastewater treatment plants were associated with the relative abundance of *sul1* (log10 gene-copies-per-16S-rRNA). Amos *et al.*[38] associated the locations and types of wastewater treatment plants, land cover, and season to the relative abundance of *intI1* genes associated with antibiotic resistance and gene mobility known as class 1 integrons in sediment.

For the association found by Pruden *et al.*, the terms describing sources were constructed spatial predictor models that do not capture the gradual effect of density and proximity to sources. Wastewater treatment plant source terms were constructed by taking the sum of upstream capacities of a source and constructed AFO source terms with interpolated upstream capacities.[17] This type of spatial predictor characterizes average upstream capacities and proximity to large upstream sources but fails to capture the influence of proximity to small upstream sources, unless there is only one source. Later work by Amos *et al.*[38] models the decaying contributions from upstream sources which fully characterizes the gradual effect of density and proximity of upstream sources but does not account for dilution due to flow.

This work expands on Pruden and Amos' spatial predictor models by using novel models that represent the sum of exponentially decaying contributions from spatially distributed sources overland and downstream, and that are diluted due to flow. Additionally, ground transport processes are modeled to characterize the hauling of manure from AFOs to manure application fields or the land application of septage in proximity to denser residential areas. Additionally, this work will adopt the recently developed microbial *Find*, *Inform,* and *Test* (FIT) framework to identify sources, quantify associations between sources and ARG responses, and characterize average transport processes. The microbial FIT framework 1) *finds* reliable databases representing spatially distributed sources of antibiotic resistances genes from a selection of candidate databases, 2) *informs* source terms constructed from selected databases with interpretable hyperparameters that help to characterize exponential influence ranges around sources 3) *tests* the predictive ability of informed source terms to identify key sources of elevated ARGs in rivers. We expect that the FIT model will select sources that have been previously found to contribute to fecal contamination in this region–bovine sources (e.g., AFOs, manure application fields), low-intensity developed land cover (i.e., land cover representing moderate application of manure from small farms and nearby large farms that migrate off site due to imperviousness when large farms are close to this land cover), and septic systems. (Wiesner-Friedman *et al.* 2021: under coauthor review) Additionally, we expect to find an association with wastewater treatment plants which were associated with the relative abundance of *sul1* in sediment by Pruden *et al.* and the relative abundance of *intI1* in sediment by Amos *et al.*[17,38]

5.3 Methods

5.3.1 Sampling, sample analysis of antibiotic resistance genes, data for spatially distributed sources, application of the fit approach to arg responses

Please see Beattie et al. for sampling methods, ARG selection, DNA extraction, and qPCR protocols.[134] See Wiesner-Friedman *et al.* (2021: under co-author review) for details on databases representing spatially distributed sources of fecal contamination. The same databases were utilized in this study of antimicrobial resistance.

The FIT framework was applied as it was in Wiesner-Friedman *et al.* (2021: under co-author review) for each of the 10 ARG responses (i.e., the relative abundance of *erm(B), tet(W), qnrA, sul1,* and *intI1* from sediment samples and surface water). The FIT framework utilizes a physically meaningful land-use regression model for identifying sources of microbial contamination from spatially distributed sources:

$$\mathrm{y}_i = \beta_0 + P1_i(\beta_1 + \beta_2 P2_i) + \beta_3 Freezing_i + \left\{\sum_{u=1}^{U} \beta_u s_i^{(u)}\left(\alpha^{(u)}\right)\right\} \exp\left\{\sum_{l=1}^{L} \beta_l m_i^{(l)}\left(\alpha^{(l)}\right)\right\} + \varepsilon_i, \qquad \text{(Eq. 5.1)}$$

Here the first three variables, $P1_i$, $P2_i$, and *Freezing* represent recent precipitation, antecedent precipitation, and a monthly average freezing temperature for each i^{th} sample. β_1 represents the association with the response, y_i, and recent precipitation. β_2 represents the effect of antecedent precipitation on recent precipitation. β_3 represents a change of mean in the response given freezing temperature. There is a sum of contributions from spatially distributed sources, $\sum_{u=1}^{U} \beta_u s_i^{(u)}\left(\alpha^{(u)}\right)$, that are either fractionated or multiplied by the modifying factors in the exponent, $\sum_{l=1}^{L} \beta_l m_i^{(l)}\left(\alpha^{(l)}\right)$, which defines a non-linear regression. Future work will address modifying factors, so in this work this equation remains linear. Each source term, $s_i^{(u)}\left(\alpha^{(u)}\right)$, is defined as a function of the spatial locations of sources of type *u*, exponential decay hyperparameters that characterize overland and downstream transport and potentially ground transport, and the sampling locations for each i^{th} sample.

The three stages of the FIT framework, *Find, Inform,* and *Test* were applied to the 10 ARG responses (i.e., sediment and water measurements of 5 ARGs) in the same manner described in Wiesner-Friedman *et al.* (2021: under co-author review) with one difference. No assumptions were made about which of the potential sources would be associated with elevated ARGS, so all potential sources were considered for each of the 10 ARG responses. In summary, these potential sources were AFOs, manure application, AFOs via the ground transport or hauling of manure, septic systems, land applied waste from industrial origins, land applied waste from municipal wastewater or septage, land applied waste from municipal wastewater or septage that is applied more greatly in proximity to denser residential areas, municipal wastewater treatment plants, industrial wastewater treatment plants, low-intensity developed land cover, and high-intensity developed land cover. Source terms were constructed using the Overland and River Distance—Flow (ORF) spatial predictor model described in detail in Wiesner-Friedman *et al.* (2021: under review). The Ground Transportation, Overland and River Distance—Flow (GORF) spatial predictor model was used to characterize AFOs via the hauling of manure and land applied waste from municipal wastewater or septage that is applied more greatly in proximity to denser residential areas. If the correlation of source terms were large ($\rho \geq 0.7$), then the term yielding the highest univariate R-squared was picked.[184] Climatic and source terms were stepwise-selected with AIC and a 6-fold cross validation was performed on the stepwise selection procedure.

5.3.2 Conducting interviews/surveys with wisconsin dairy cattle veterinarians

Rogers *et al.* were able to leverage local knowledge from dairy operators in order to inform their findings on ARGs shed by wild deer.[70] A 2007 study exists that outlined and quantified antibiotic usage on Wisconsin dairies,[224] but attitudes and policies around antibiotic usage may have changed. To gain knowledge of how practices may have changed and help to contextualize our results, we reached out to Wisconsin dairy cattle veterinarians. Google maps was used to search for veterinarians, veterinary services, and “large animal” services. Through this search 12 veterinary offices were identified as serving dairy cattle, representing over 30 veterinarians. An IRB was written for conducting interviews, and an exempt status was obtained for conducting interviews due to the study area encompassing a whole state and interviews being anonymous. Emails were sent to veterinarian email addresses with the survey. Two phone interviews were conducted due to preference and 2 surveys were completed online through Qualtrics. The survey is outlined in Table A.11 and interviews were conducted in the same format as the survey.

5.4 Results and Discussion

5.4.1 Patterns of associations with recent and antecedent precipitation and freezing temperature

Table 5.1 shows the results from the FIT framework applied to the log10 antibiotic resistance gene relative abundance (log10 gene-copies-per-16S-rRNA-copies) for each of the 5 responses *erm(B), tet(W), qnrA, sul1,* and *intI1.* The associations with precipitation terms are shown in blue. The associations with freezing temperature are shown in white. Association with source terms are discussed later. Here, we find that there are three patterns of association between recent precipitation and the effect of antecedent precipitation on recent precipitation.

The first pattern resembles increases in the log10 gene relative abundance given an increase in recent precipitation and a diminishing effect of antecedent precipitation on recent precipitation (i.e., wet antecedent conditions protect against contributions from a recent precipitation event). This was characterized for log10 relative abundance of *erm(B), tet(W),* and *qnrA* in surface water. A one-standard-deviation increase in recent precipitation was associated with a 0.417 (p-value<0.05) increase in the log10 relative abundance of *erm(B)* in surface water. A one-standard-deviation increase in antecedent precipitation diminished the effect of recent precipitation by 0.383 (p-value<0.05). Similarly, a one-standard-deviation-increase in recent precipitation was associated with increases of 0.635 (p-value<0.05) and 0.213 (p-value<0.05) in the log10 relative abundances of *tet(W)* and *qnrA* in surface water. For a one-standard-deviation increase in antecedent precipitation, these effects were diminished by 0.300 (p-value<0.05) and

1.10 (p-value<0.05), respectively. Additionally, a one-standard-deviation increase in recent precipitation was associated with an increase of 0.637 (p-value<0.05) in the log10 relative abundance of *sul1*in surface water without any diminishing effect from antecedent precipitation.

The second pattern represents increases from recent precipitation with further increases due to antecedent precipitation (i.e., wet antecedent conditions exacerbate the contributing effects of a recent precipitation event). A one-standard-deviation increase in recent precipitation was associated with an increase of 0.654 (p-value<0.05) and 0.277 (p-value<0.05) in the log10 relative abundance of *tet(W)* in sediment and *intI1* in surface water, respectively. A one-standard-deviation increase in antecedent precipitation increased the effect of recent precipitation by 0.259 and 0.637, respective to *tet(W)* in sediment and *intI1* in surface water.

The last pattern represents a decrease in baseline levels of the log10 gene relative abundance from recent precipitation where this decreasing effect is diminished by antecedent precipitation (i.e., wet antecedent conditions followed by a recent precipitation event increases the levels of ARG responses). A one-standard-deviation increase in recent precipitation was associated with a decrease of 0.723 (p-value<0.05) and 0.337 (p-value<0.05) in the log10 relative abundance of *erm(B)* and *intI1* in sediment, respectively. The effect of this decrease was diminished by 0.561 (p-value<0.05) and 0.259 (p-value<0.05) for *erm(B)* and *intI1* in sediment, respectively.

The first pattern describing the effect of antecedent precipitation on recent precipitation contributions was only occurred for ARG responses in surface water. This is consistent with findings around the flush effect for fecal indicators and host-associated responses in surface water.[40,137,185–188] For surface water, a second pattern was found for the relative abundance of *intI1*. This second pattern was also found for the *tet(W)* response in sediment. The third pattern only occurred for sediment responses and has not been reported before. Previously, wet events have been associated with elevated ARGs and ARG levels have been found to increase during and after precipitation events.[225–227]

Lastly, freezing temperature was not frequently selected by FIT across the 5 ARG responses in surface water and sediment (i.e., 10 responses). When selected for sediment, freezing temperature characterized decreases to the ARG responses. Freezing temperature was associated with decreases of 0.621 (p-value<0.10) and 0.812 (inclusion lowers AIC) in the log10 relative abundance of *erm(B)* and *intI1* in sediment. Freezing temperature was only selected once for surface water and characterized increases to the log10 relative abundance of *erm(B).* Freezing temperature was associated with a 1.93 (p-value<0.05) increase in the relative abundance of *erm(B)* in surface water. Overall, precipitation and freezing temperature were primarily helpful to explaining between-sampling event variability, but studies are needed with a finer temporal resolution to untangle the associations between recent and antecedent precipitation on ARGs and ambient temperatures.

Table 5.1 Regression results for predicting the relative abundance of *erm(B), tet(W), qnrA, sul1, and inI1* (log10 gene-copies-per-16S-rRNA-copies) in sediment (columns toward the right) and surface water (left-most 5 columns). The sample size is indicated for each of the responses in each column. For each of the climatic and source terms the standardized regression coefficient, **β**, is provided resulting from the *Test* stage of FIT. For each source term, there are two additional rows resulting from *Find* and *Inform* stages of the FIT framework. For each source term category (i.e., bovine or land applied waste), the source description, the relative abundance ration ($RAR = 10^{\beta}$) and hyperparameters indicating the influence range around sources, **α,** are summarized.

		Log10 relative abundance (log10 gene-copies-per-16S-rRNA-copies)								
Environmental Media		Sediment				Surface water				
ARG (*n*=sample size)		***erm(B)*** (*n*=91)	***tet(W)*** (*n*=91)	***sul1*** (*n*=91)	***intI1*** (*n*=91)	***erm(B)*** (*n*=98)	***tet(W)*** (*n*=98)	***qnrA*** (*n*=98)	***sul1*** (*n*=98)	***intI1*** (*n*=98)
Climatic Terms										
Recent Precip.	Std. Regression Coefficient ($\boldsymbol{\beta}_{P1}$)	-0.723**	0.654**	*NS*	-0.337**	0.417**	0.635**	0.213**	0.637**	0.277**
Recent x Antecedent Precip.	Std. Regression Coefficient ($\boldsymbol{\beta}_{P1xP2}$)	0.561**	0.923**	*NS*	0.259**	-0.383**	-0.300**	-1.10**	*NS*	0.637**
Freezing	Regression Coefficient ($\boldsymbol{\beta}_{freezing}$)	-0.621*	*NS*	*NS*	-0.812	1.93**	*NS*	*NS*	*NS*	*NS*
Source Terms										
Bovine Sources	Bovine Source Description	GORF AFO (via ground transport of manure to application fields)	GORF AFO (via ground transport of manure to application fields)	GORF AFO (via ground transport of manure to application fields)	*NS*	ORF AFO (via ground transport of manure to application fields)	ORF AFO	ORF AFO	ORF Manure App. Fields	ORF Manure App. Fields
	Std. Regression Coefficient ($\boldsymbol{\beta}_{source}$)	0.199*	0.255	0.128	*NS*	0.123*	0.247**	0.173**	0.148*	0.134
	RAR ($\mathbf{10}^{\boldsymbol{\beta}_{source}}$)	1.58	1.80	1.34	*NS*	1.33	1.77	1.49	1.41	1.36
	Influence Range ($\boldsymbol{\alpha_o}$)	$< 13\ km$				$< 10\ km$				
Land App. Waste Sources	Land applied waste Source Description	*NS*	*NS*	Land applied waste-*residential*	Septage ground transport to land app. sludge-*residential*	*NS*	Land applied waste-*industrial*	*NS*	*NS*	Land applied waste-*industrial*
	Std. Regression Coefficient ($\boldsymbol{\beta}_{source}$)	*NS*	*NS*	0.166*	0.126	*NS*	0.134*	*NS*	*NS*	0.148*
	RAR ($\mathbf{10}^{\boldsymbol{\beta}_{source}}$)	*NS*	*NS*	1.47	1.34	*NS*	1.36	*NS*	*NS*	1.41
	Influence Range ($\boldsymbol{\alpha_o}$)	$< 8\ km$				$< 10\ km$				

**indicates p-value<0.05, * indicates p-value<0.10
NS indicates that no terms were selected for the source category.
Other source categories not selected: Wastewater treatment plants, septic systems, and developed land cover
No terms were selected by AIC for the log10 relative abundance of *qnrA* in sediment

5.4.2 Bovine sources consistently contribute to elevated ARGs

After implementing the FIT framework across 5 ARGs in sediment and surface water, we find that all 5 of the ARG responses (see Table 5.1) are positively associated with bovine sources in at least surface water or sediment. For ARG responses in *sediment,* FIT selected the GORF AFO source term which represents the contributions from the application of manure on fields from AFOs. For these sediment responses, the contributions from AFOs were characterized specifically by contributions from CAFOs. For *erm(B)* and *tet(W),* CAFOs were weighted by animal units, but for *sul1,* CAFOs were equally weighted. For *erm(B)* and *sul1*, manure application fields were characterized by crop rotation, but for *tet(W)* they were characterized by dairy rotation. A one-standard-deviation increase in GORF AFO contributions was associated with a 58% (p-value<0.10) increase in the relative abundance of *erm(B)* in sediment, an 80% (inclusion lowers AIC) increase in the relative abundance of *tet(W)* in sediment, and a 34% (inclusion lowers AIC) increase in the relative abundance of *sul1* in sediment. No sources were associated with *qnrA* in sediment.

Across ARG responses in *surface water,* AFOs were characterized specifically by locations of manure storages. Manure application was characterized by crop rotation land cover. For *erm(B),* FIT selected the GORF AFO source term representing the contributions from the application of manure on fields from AFOs. A one-standard-deviation increase in AFO contributions was associated with a 33% (p-value<0.10) in the relative abundance of *erm(B)* in surface water. Then, for *tet(W)* and *qnrA* responses, FIT selected the ORF AFO source term representing the contributions from manure storage locations weighted by the log-total-gallons of manure. A one-standard-deviation increase in AFO contributions was associated with 77% (p-value<0.05) and 49% (p-value<0.05) increases in the relative abundance of *tet(W)* and *qnrA*, respectively. For *sul1* and *intI1,* FIT selected the ORF manure application field source term representing the contributions from manure applied to crop rotation land cover. A one-standard-deviation increase in manure application field contributions was associated with 41% (p-value<0.10) and 36% (inclusion lowers AIC) increases in the relative abundance of *sul1* and *intI1*, respectively.

Bovine sources related to microbial responses in sediment were characterized by the GORF source term representing the hauling and application of manure from AFOs onto fields, and AFOs were characterized by WPDES CAFOs, whereas for surface water, GORF or ORF AFO source terms were characterized by AFO manure storages. Previously, a similar pattern was revealed for bovine-associated fecal contamination from these samples. (Wiesner-Friedman *et al.* 2021: under coauthor review) This may reflect different overland pathways of contamination to sediment or surface water, different detection attributes in sediment versus surface water, or this may reflect the higher impact of AFO manure storages on microbial contamination in rivers during flooding conditions, when surface water was able to be sampled, but not sediment.

Across all the responses, the magnitude of the association between *tet(W)* and AFOs was greatest across sediment and surface water responses. A greater association may indicate that there is more *tet(W)* genes located at AFOs, that *tet(W)* is more specific to AFOs compared to the environment in general, or that the ability to detect *tet(W)* in sediment and surface water is greater, compared to the other ARGs. Dairy manure-amended soils compared to soils amended with other manure have been reported to have a greater prevalence of tetracycline resistance genes and tetracycline resistant bacteria.[228] In particular, feces from dairy calves and heifers have been reported to have greater relative abundances of *tet(W)* compared to *tet(M)* or *tet(Q)*.[229] This supports the specificity of tetracycline resistance and *tet(W)* to dairy feces and reflects a relationship to the hauling of manure from AFOs.

This is the first time that GORF or ORF spatial predictor models have been used to characterize contributions to elevated ARG responses from AFOs. This implies that hauling of manure, overland and downstream, and dilution from flow are key processes in the dissemination of antimicrobial resistance from AFOs. Our novel spatial predictors and modeling approach is in agreement with the association found by Pruden *et al.* using a different spatial predictor model and for study area in a different region.[17] Pruden *et al* found that the relative abundance of *sul1* in *sediment* was positively correlated ($R^2 = 0.35$, $p < 0.001$) with average upstream capacities of AFOs (i.e., source term was represented by the inverse-distance-weighted *interpolated* upstream AFO capacities). However, Pruden found no significant relationship with the relative abundance of *tet(W)* in sediment.[17] In our study, 7 novel associations were quantified between bovine sources and ARG responses. This is the first study to quantify the association between bovine sources and *sul1* in *surface water* as well as *erm(B), tet(W), qnrA,* and *intI1* in surface water (i.e., 5 novel associations). This is the first study to quantify the association between bovine sources and *erm(B)* and *tet(W)* in sediment (i.e., 2 novel associations).

5.4.3 Land applied septage, municipal and industrial waste are another source of elevated ARGs

After implementing the FIT framework across 5 ARGs in sediment and surface water, we find that 3 of the ARG responses (see Table 5.1) are positively associated with bovine sources in at least surface water or sediment. For ARG responses in *sediment,* the land application of *municipal waste* or *septage* characterized land applied waste sources. For *sul1,* FIT selected the ORF land applied waste from residential use source term representing the land application of municipal waste or septage consisting of solid or semi-solid residue generated during the treatment of domestic sewage via primary, secondary, or advanced wastewater treatment *and* the wastewater contents of septic or holding tanks, dosing chambers, grease interceptors, seepage beds/pits/trenches, privies, or portable restrooms.[181] A one-standard-deviation increase in contributions from land applied waste from residential use was associated with a 47% (p-value<0.10) increase in the relative abundance of *sul1*. For *intI1*, FIT selected the GORF land applied waste from residential use source term representing increased land applied septage in proximity to residences using septic systems. A one-standard-deviation increase in land applied waste from residential use was associated with a 34% (inclusion lowers AIC) increase in the relative abundance of *intI1*.

For ARG responses in *surface water*, the land application of *industrial waste* characterized these secondary sources. For *tet(W)*, FIT selected the ORF land applied waste from industrial use source term representing by-product solids from the animal product or food processing industry (i.e., remains of butchered animals, paunch manure, and vegetable waste materials). A one-standard-deviation increase in land applied waste from industrial use was associated with a 36% (p-value<0.10) increase in the relative abundance of *tet(W)*. Then for *intI1*, FIT selected the ORF land applied waste from industrial use representing both the by-product solids from animal product or food processing and liquid waste such as silage, leachate, whey, whey permeate, whey filtrate, contact cooling water, cooling or boiler water containing water treatment additives, and wash water generated in industrial, commercial, and agricultural operations.[181] A one-standard-deviation increase in land applied waste from industrial use was associated with a 41% (p-value<0.10) increase in the relative abundance of *intI1*.

Our finding that the land application of municipal waste and septage is associated with ARGs in sediment is consistent with current knowledge that ARGs are enriched in biosolids from treatment processes (i.e., primary, secondary, or advanced wastewater treatment).[83,108,219] This is the first study to report an association between contributions from land applied waste and

ARGs recovered from the environment. This work is also the first to show that the disposal of municipal wastes and septage on land does pollute and corresponds with a quantifiable environmental impact to antibiotic resistance levels. Our study is also the first to show that land application of industrial waste is associated with elevated ARGs. The WPDES database lists the permit holders for these land application sites. The majority of permittees applying industrial waste are companies that produce dairy products. Our work therefore indicates that the disposal of wastes from the processing of dairy is an additional way that the livestock industry can impact the environment.

5.4.4 Fecal contamination may disseminate from small farms, but not contribute to elevated ARGS

Other sources that were *tested* in the FIT model but were not selected were municipal and industrial wastewater treatment plants, low-intensity developed land cover, high-intensity developed land cover, and septic systems. The non-selection of low-intensity is interesting because low-intensity developed land cover was found to be key source of fecal contamination in previous work but did not represent a source of elevated ARGs in our work. In this region, low-intensity developed land cover captures a moderate rate of manure application from small farms. Our work suggests that the frequency and quantity of antibiotics used at large farms represented by AFO manure storages or permitted CAFOs are much greater compared to small farms that are represented by low-intensity developed land cover.

Additionally, wastewater treatment plants were expected to show associations with *sul1* and *intI1*,[17,38] but neither municipal or industrial wastewater treatment plants were selected by the FIT model to show associations with any of the 5 relative abundance responses across surface water or sediment. Both *sul1* and *intI1* were associated with land applied municipal waste and septage. This indicates that some ARGs may originate from wastewater treatment plants, but that the persistent application of treated biosolids on land may represent a more significant source. The land applied municipal waste and septage does represent the aggregation of treated septage and wastewater. The greater density of wastewater treatment plants in the South Platte River Basin in Colorado, USA and the Thames Watershed in Oxfordshire, UK compared to this study, as well as the larger bovine to human ratio in this study may also contribute to the ability to detect the impacts of wastewater treatment plants. AMR quantitation methods with low limits of detection may be needed to detect subtler impacts.

5.4.5 The exponential influence range around sources extends up to 13 km

Bovine source terms associated with elevated ARGs in *sediment* had influence ranges of α_O<13 km to the river network indicating that exponentially decaying contributions would still be detected when manure fields were up to 13 km away from the river network. For all bovine source terms associated with ARG responses in *surface water*, exponentially decaying contributions would still be detected when sources were up to 10 km from the river network. For all land applied waste source terms associated with ARGs, the influence range was up to 8 km for *sediment* responses and up to 10 km for *surface water* responses.

These exponential influence ranges, α_O, were determined as a result of the *Inform* stage of FIT. Wiesner-Friedman *et al.* (2021: under coauthor review) remarked that this hyperparameter captures much more than average overland transport. This hyperparameter characterizes the extent to which a microbial response can capture a signal from sources. Two explanations for larger exponential influence ranges in sediment were proposed by Wiesner-Friedman *et al.* The first was that the detection of obligate anaerobes may be decreased under aerobic conditions such as in surface water, which is also an explanation for increased probability of detecting host associated *Bacteroides* markers in sediments compared to surface water.[23,208,212] This first case pertains to characteristics of the molecular target. Second, the settling of microbes from surface water to sediments may increase the probability of detecting molecular targets in sediments from surface water. This second case pertains to a plausible physical phenomenon. However, the longer exponential influence ranges (i.e., up to 10 km) around bovine sources for ARG responses in surface water compared to the short exponential influence ranges (i.e., less than 1 km) around bovine sources host associated *Bacteroides* genes in surface water in previous work reinforces the first case (i.e., the characteristics of the molecular target or the bacteria carrying the genes such as oxygen levels). ARGs can be carried by either aerobic or anaerobic bacteria and have found to be more greatly enriched under aerobic versus anaerobic conditions.[230] This may increase the ability to detect ARGs in surface waters compared to *Bacteroides*, which affects the extent to which a signal may be detected from any source (i.e., longer exponential influence range for surface water responses would be expected). One implication of longer influence ranges around bovine sources for ARGs compared to host-associated *Bacteroides* genes are that the risks associated fecal contamination from bovine sources in surface water may be underrepresented by host-associated markers that decay quickly in surface water. Work to quantify the risks associated with sources should carefully consider the characteristics of molecular targets to model particular risks.

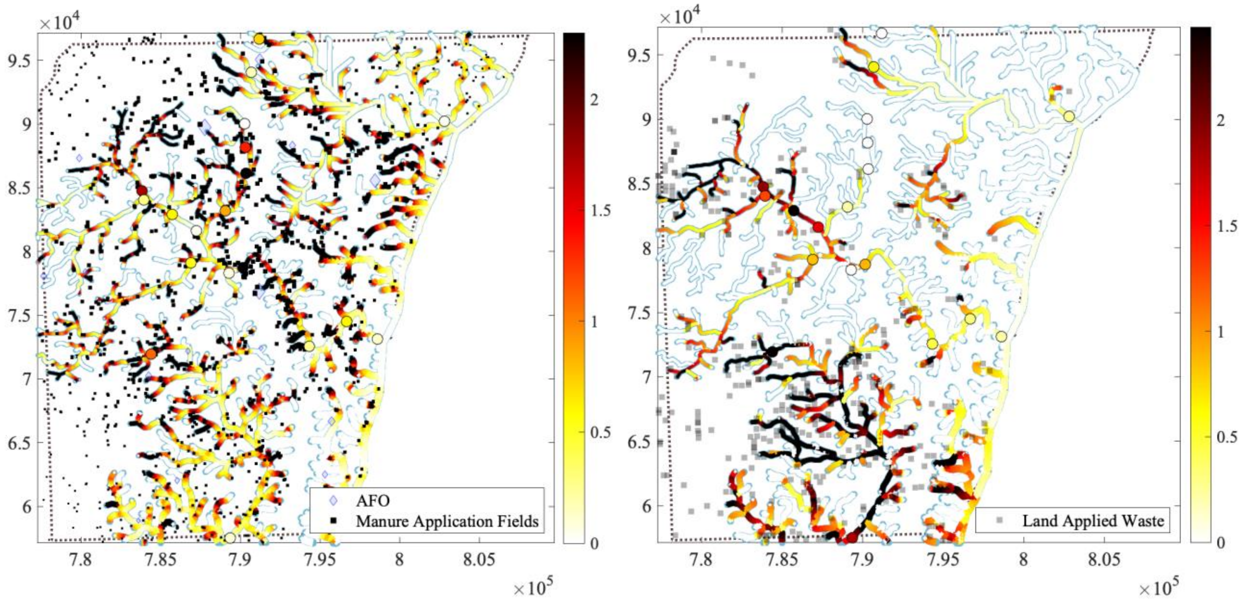


Figure 5.1 Conservative source contribution scenarios for ARG responses. The depicted source term is the standardized (z-scored according to sampling scheme) contributions selected by the FIT model due to associations with the responses (Table 5.1). An increase of 1 on the color-scale corresponds to a one-standard-deviation increase. This standard deviation increase is approximately associated with the percent increases in the relative abundance responses from which the depiction was characterized. These are considered very conservative scenarios because they characterize contributions with 1) upper bound values for hyperparameters that characterize both transport ranges and detection probabilities and 2) contributions from databases representing an overarching class (e.g., rather than dairy rotation, crop rotation is represented) Circles represent the sampling locations associated with this study. (a) AFO contributions via manure application to riverbed sediment scenario that was characterized from three relative abundance responses (*erm(B)*, *tet(W)*, and *sul1*) in sediment. (b) Land applied waste contributions to surface water scenario characterized from two relative abundance responses (*tet(W)* and *intI1*) in surface water.

5.4.6 ARG levels are influenced by the proximity and density of upstream sources and dilution from flow

For different ARG responses, various pollution scenarios are possible. Figure 5.1 depicts two source contribution scenarios where the depiction is standardized based on the sampling scheme used in this study. The depictions are conservative because they are determined by the upper-bounds of the hyperparameters that describe transport or the probability of detection for the ARGs in question, and the databases used to represent release points are from larger classes of data (i.e., crop rotation is used instead of dairy rotation, land applied industrial wastewater and sludge is used instead of sites only associated with food processing). For example, Figure 5.1a depicts AFO contributions to riverbed sediments via the application of manure on fields. The orange color depicts a one-standard-deviation increase away from 0 contributions (depicted in white). The locations on the river network where the standardized AFO contributions are equal to 1 therefore conservatively represent 58%, 80%, and 34% associated increases in the relative abundances of *erm(B), tet(W),* and *sul1* in sediment, respectively. Figure 5.1b depicts contributions from the land application of industrial wastes. The locations on the river network where the standardized contributions from land applied industrial wastes are equal to 1 conservatively represent 36% and 41% associated increases in the relative abundances of *tet(W)* and *intI1* in surface water.

When sources are located directly next to the river network, notches in color appear along reaches of the river network, demonstrating the effect of proximity of sources. Additionally, where there are many sources, greater contributions contributions have been characterized. When there are no sources upstream of river reaches, the reaches remain white indicating there are no contributions. Lastly, when a river reach that has many contributions (i.e., dark in color) converges with a river reach that has little contributions from sources (i.e., white in color), a sudden drop in the contributions to the reach downstream of the two occurs. This characterizes the effect of dilution due to flow. Ecologically, this work is the first to model associated source contributions to ARGs that depend on the proximity and density of upstream sources, as well as dilution from increased flow. Previously, seasonal variation in flow has been related to changes in ARG levels.[106] Here, we observe that the flow related to upstream watershed area dilutes the levels of ARGs.

5.4.7 Patterns of antibiotic use at bovine sources described by Wisconsin dairy veterinarians

Two of the veterinarians interviewed had greater than 20 years of experience in the field and indicated that antibiotic use has decreased over the years, especially for treatments of mastitis, of which they mentioned newer treatments, but did not identify them specifically. A common use of antibiotics is for treatment of infection from calving. One veterinarian discussed the use of antibiotics to prevent illness. “Tetracycline flushes” were described as when tetracyclines are administered to the calving mother for three days after calving to avoid infection. Currently guidelines have sought to decrease this practice, but veterinarians have noted that if operations are not monitoring cows well, the cow can “become febrile”. Due to ethical reasons, some veterinarians do not like the “modern way” because cows may be poorly monitored and can suffer. In the past, antibiotics were primarily used for the treatment of mastitis, but veterinarians indicated that newer treatments now exist. Our interviews suggest that veterinarians may be resistant to reducing antibiotic use if it leads to more animal suffering. Policies aimed at reducing the impacts of AFOs on antimicrobial resistance should work closely with veterinarians to identify other ways to reduce animal suffering.

According to these veterinarians, Ceftiofur was the frequently prescribed antibiotic, which is a Beta-lactam.[231] Other antibiotics that were identified as were enrofloxacin, florfenicol, tulathromycin, and oxytetracycline which belong to the broader classes of fluoroquinolones, sulfonamides, macrolides, and tetracyclines. These interviews reflected information that was consistent with the findings of a 2007 study on antibiotics usage at conventional dairies in Wisconsin.[224] The ARGs measured by Beattie *et al.*[134] and used for this study*, qnrA, sul1, erm(B),* and *tet(W)* encode resistance to each of these broader classes (i.e., fluoroquinolones, sulfonamides, macrolides, and tetracyclines) of antibiotics, respectively. Additionally, *sul1* and *intI1* have been found to co-occur genomicaly and on plasmids.[232]. The common use of these classes of antibiotics for dairy cattle infection treatment or prevention supports that the conditions of antimicrobial use at AFOs lead to elevated levels the panel of ARGs found in this study. This is important information due to barriers to sampling at dairy operations.

No study of commonly prescribed antibiotics in clinical settings exists for Wisconsin. However, in the United States, fluoroquinolones are among one of the most commonly clinically prescribed classes of antibiotics.[233] The dissemination of fluoroquinolone-associated resistance encoded by both chromosomal and plasmid derived *qnrA* may increase the risks of quinolone-resistant human pathogens in surface water from plasmid-mediated horizontal gene transfer.[234] Our findings that connect AFOs to levels of *qnrA* in surface water is concerning and efforts should be made to prevent the use of antibiotics that are clinically relevant to humans in livestock settings.

5.4.8 Implications, limitations, and future work

This is the first comprehensive land-use regression study of sources of elevated ARGs and is the first study of more than 2 antibiotic resistance gene responses in surface water or sediment samples. The primary finding of this study is that bovine sources were consistent sources of elevated antibiotic resistance genes. This finding was expected due to known practices at AFOs where antibiotics are used for prevention and treatment of disease.[235,236] Additionally, a large body of work has detected ARGs in livestock manure and slurry, on soil where the manure or slurry is applied, and downstream of livestock operations.[16,101,201,237–241]

In a similar capacity, previous work has revealed the land application of treated biosolids from septage, and municipal and industrial wastes as a potential source of elevated ARGs.[219,242] However, no work has found associations with measured levels of ARGs in the environment and the sum of decaying contributions from all land applied waste sites. A large body of work suggests potential offsite migration of other microbial contaminants from land applied sludge into soils, groundwater, surface water, and the air.[243–251] This work adds to this body of work but is the first to suggest offsite migration from land applied sludge sites associated with antibiotic resistance.

Under current guidelines provided by The U.S. Environmental Protection Agency's (EPA's) Part 503 rule, there are treatment standards for two classes of biosolids. Class A biosolids are treated so that pathogens are reduced to levels that make the biosolids safe to apply to crops, lawns, and home gardens. Class B biosolids must meet a requirement of 2 million colony forming units (CFU) or MPN for fecal coliforms. Because of the high pathogen levels in class B biosolids, there are restrictions on the application of Class B biosolids to crops and public access to the land to protect public health.[252] However, this work and previous work suggest that microbial contamination does disseminate from these sites to places where people may be exposed to contaminants that are detrimental to public health. The offsite migration of microbial contaminants from land applied sludge should be considered by regulators in determining treatment standards, land application restrictions, and environmental monitoring of potential offsite migration.

Few regulations exist around treatment of animal manures and its application on land, despite similarities with class B biosolids. Regulations of livestock wastewater, which only apply to permitted CAFOs, have focused on the management of nutrients from animal manures to fertilize soil for agriculture, while minimizing runoff. However, the presence antibiotic resistant bacteria, endotoxins, prions, pathogenic bacteria and protozoa have been reported in both class B biosolids and livestock waste.[249] An explanation for the increased consistency of association between elevated ARGs and bovine sources compared to land applied waste sources is that these land applied wastes are more regulated than land applied manure, and that this regulation has been beneficial.

However, communities neighboring land applied sludge sites and manure application fields have reported negative changes to the environment and their health following the application of sludge.[253] Negative health outcomes have been measured for residents living in regions of with dense industrial livestock agriculture.[45,254] To protect the health of the communities living near to these sites, overall health impacts need to be evaluated with epidemiological studies related to health outcomes and the contaminants present in biosolids from human, animal, and industrial wastes (e.g., pathogens, antimicrobial resistant bacteria and genes, heavy metals, disinfectants, fire retardants, prescription and non-prescription pharmaceuticals, and polycyclic aromatic hydrocarbons).[219,242,255,256] An important note is that many co-occurring contaminants in biosolids have been reported to exert selective pressures on microbial communities inducing higher levels of antimicrobial resistance.[60,232] Monitoring elevated antimicrobial resistance may therefore be complementary to quantifying the impacts of other contaminants from the land application of biosolids on the environment. Furthermore, the conservative source pollution scenarios derived from the FIT framework's application to ARGs may be useful in defining source-associated exposure variables for future epidemiological studies.

Though only two of the associations found in this work were significant at the 95% confidence level, out of all the types of source terms (e.g., septic systems, developed land-cover, wastewater treatment plants, bovine sources, land applied waste sources), bovine or land applied waste source terms were selected to be in the model for each ARG response. The extent of this reproducibility across 5 different ARG responses in sediment and surface water strengthens the evidence that elevated ARGs in the environment are linked to land-application of biosolids and sludges. The dissemination of ARGs from the land application of biosolids and semi-solids from human, animal, and industrial waste calls for more waste treatment strategies that remove or reduce ARGs, policies to reduce antimicrobial use in livestock and humans, and investments in alternative treatments for infections (e.g., phage-therapy).

In the intermediary, recent work suggests that the careful placement of constructed wetlands around sources may result in some attenuation of ARG levels and other contaminants.[96,257] Little work has been done to distinguish *natural* wetlands, other land cover, and geological features as attenuators or amplifiers of antibiotic resistance genes. Since the FIT framework leverages a physically meaningful land-use regression model, one strength or limitation is that prior evidence is needed about whether a source will contribute, or whether a feature will attenuate or amplify contributions. (Wiesner-Friedman *et al.* 2021:ES&T invited to resubmit) While potential sources have been well characterized, more work is needed to distinguish land cover and geological features as potential attenuators or amplifiers of ARGs in the environment. These measures would not resolve workplace and airborne exposures to antimicrobial resistant pathogens.

This is the first time that influence ranges around these sources have been defined for elevated ARG responses in rivers and streams. These exponential influence ranges indicate that sources that are very close to the river will have a great impact, but that distant sources up to 13 km away from the river network can also impact ARG levels. For the land application of manure and slurry, a setback distance of 34-67 m from surface water has previously been recommended from rainfall simulation tests under experimental conditions.[237] We expect that when there is more chaos in the environment, that differences in the influence ranges around sources for ARGs in rivers would exist. Chaos resulting from different precipitation patterns, application methods, and different environmental pathways (e.g., subsurface flow and tile drainage), may affect flow velocities that would not be represented under experimental conditions. For example, a factor that may lead to long influence ranges in this region is the Karst geology, where fractures, sinkholes, caves, disappearing streams, and springs may provide direct pathways for antibiotics and ARGs and other contaminants to reach ground and surface waters.[258,259] This would provide rapid decay of ARGs levels via sheet-flow and in soil, but levels could remain high within groundwater that flows to rivers. Our findings on the influence ranges around sources are consistent with influence range from a study in a Karst region in Germany, where elevated ARGs and human-specific fecal markers were detected in a spring 9 km away from the suspected source.[259]

This work identified and characterized sources contributing to elevated antibiotic resistance genes in surface water and sediment characterized by a panel of 5 ARGs. Overall, this works suggests that the land application of biosolids represents the most import sources of elevated antimicrobial resistance in this region. Sources were found to have influence ranges of up to 13 km, which indicates that some increases in ARG levels can be detected due to the land application of biosolids anywhere in Kewaunee County. This work calls for innovation in wastewater treatment to consider reductions in ARGs from wastewater destined for land application, especially for livestock wastewater. Epidemiological studies are needed to evaluate the impacts of higher ARG levels in surface water on health outcomes to quantify public health risks.

CHAPTER 6: CONCLUSIONS

6.1 Summary of Findings

A novel microbial land-use regression framework was developed to identify and characterize sources of fecal contamination and elevated antibiotic resistance. Spatial predictor models were developed that characterize contributions from a limited set of sources in different ways (Euclidean, ORF, GORF). Additionally, a criterion was developed to help with database selection. Three stages were developed to accommodate these developments. The *Find* stage consisted of using a criterion to choose reliable databases representing spatially distributed sources. The *Inform* stage consisted of selecting hyperparameters that maximize physical meaningfulness and characterize overland and downstream transport, as well as ground transportation. The *Test* stage consisted of testing the statistical significance of various potential source.

This microbial *Find, Inform,* and *Test* (FIT) framework was first applied for demonstration to the relative abundance of bovine-associated fecal contamination in sediment for a limited number of potential sources linked with bovine hosts. From the results of the *Find* stage of FIT, permitting data was found to be crucial to capturing an association between bovine-associated fecal contamination and AFOs. From the results of the *Inform* stage, hyperparameters were selected that would help with creating depictions of source contributions and constructing source terms. From the results of the *Test* stage, a statistically significant and positive association was found between the GORF manure application source term representing the ground transport of manure from AFOs to application fields where average transport overland and downstream would occur. This was the first study to suggest the hauling of manure as a key transport process. It was also among the first studies to suggest dilution from flow as a key transport process for microbial contamination in sediment. Overall, this work provides strong evidence of offsite migration from AFOs into the environment.

In the next applications to the relative abundance of human and bovine-associated fecal contamination in surface water and sediment, there were more potential sources considered. Potential sources of human fecal contamination were considered to be septic systems, municipal wastewater treatment plants, high-intensity developed land cover, and land applied sludge from municipal wastewater or septage. Potential sources of bovine fecal contamination were considered to be AFOs, manure application fields, low-intensity developed land cover, industrial wastewater treatment plants, low-intensity developed land cover, and land applied sludge from industrial wastewater or food processing. AFOs were found to be a source of bovine fecal contamination. Low-intensity developed land cover was also found to be a source bovine fecal contamination, likely because it captures moderate rate of manure application from small farms or a moderate rate of manure application from large farms in proximity to impervious surfaces that may convey pollution to rivers or reduce the capacity of the land to filter pollutants. Septic systems are a likely source of human-associated fecal contamination. From this work it was suggested that sediment responses may better capture signals from distant sources when using molecular targets derived from obligate anaerobes. This work suggests that the selected molecular target may result in different characterization of source impacts and different estimates of microbial risks. Additionally, this work suggests that surface water measurements alone may not be enough to fully characterize some microbial risks. The patterns of microbial contaminant flow from sources to the river depicted in this research can help to shape creative solutions around rural wastewater management in consideration of river connectivity. This work suggests that this area is impacted by fecal contamination and calls for innovation in rural wastewater treatment and infrastructure.

In the last application, the FIT framework was applied to a panel of 5 ARGs quantified from riverbed sediment and surface water samples. All potential sources of human and bovine fecal contamination were considered as potential sources of elevated AMR. Among these potential sources, municipal or industrial wastewater treatment plants, septic systems, low-intensity developed land cover, and high-intensity developed land cover were never selected by FIT. Bovine sources were found to consistently contribute (i.e., 5 out of 5 of the ARGs from either surface water or sediment were associated with bovine sources) to levels of antimicrobial resistance as characterized by this panel of ARGs. This work found particularly high associations with *tet(W)* which encodes tetracycline resistance, consistent with knowledge of tetracycline use at dairy operations and sampling of dairy manure amended soils. Land application of waste sources were sometimes (i.e., 3 out of 5 ARGs from either surface water or sediment) found to contribute. To the best of our knowledge, this work is the first to show that the disposal of municipal wastes and septage on land corresponds with a quantifiable environmental impact to antibiotic resistance levels.

In characterizing source contributions, longer influence ranges were characterized sources of ARG responses compared to host-associated fecal responses. An explanation that follows from previous work is the ranges of influence characterized by the *Inform* stage capture much more than transport. They may also describe characteristics of the molecular target or the bacteria carrying the genes, which disconnects associations between host-associated fecal markers that represent anaerobic bacteria and molecular targets that can also represent aerobic bacteria (e.g., ARGs). This may complicate the estimation of risks using molecular targets that only capture a subset of the microbial risk associated with sources. Overall, this work indicates that elevated antimicrobial resistance is associated with the land application of biosolids of human, animal, or industrial origin. Despite similarities between these biosolids in terms of their makeup (e.g., types of pathogens and other contaminants), there are no treatment standards and few restrictions imposed on the application of animal wastes.

6.2 Significance

This work has advanced microbial water quality modeling and land-use regression approaches. These models have helped to identify key sources of fecal contamination and elevated antimicrobial resistance to rivers in a CAFO-dense region. The characterization of the contributions around sources of microbial contamination in riverbed sediments versus surface water, contributed knowledge about the characteristics of molecular targets and how these characteristics may influence the extent to which a signal can be detected from a source. This work provides useful insights for quantitative microbial risk assessment for selecting molecular targets to help estimate microbial risks. This work confirms that CAFOs, as well as AFOs, are significantly contributing to microbial contamination in this region and calls for higher standards for treating animal wastes, innovation in rural wastewater management and reduction of ARGs from wastewater, and/or more restrictions on manure application and antimicrobial use. Lastly, this framework allows for the depiction of contributions from sources which may be useful for monitoring, to regulators, and as exposure-variables for epidemiological studies that are needed to define the risks associated with these contributions.

APPENDIX 1: SUPPLEMENTAL INFORMATION FOR AIM 1: THE MICROBIAL FIND, INFORM, AND TEST (FIT) MODELING FRAMEWORK FOR IDENTIFYING SPATIALLY DISTRIBUTED CONTAMINATION SOURCES

A1.1 Processing River Network and Climatic Data

The river network was extracted in ArcMap 10.5 from a Digital Elevation Model (10 meter) raster file obtained from the Wisconsin Department of Natural Resources (WIDNR) Open Data website. Flow accumulation functions were used to convert to polyline shapefile. This data was projected into the "Wisconsin Central" State Plane Projected Coordinate System. This newly created river network was loaded into MATLAB 2018b as a shapefile. The Bayesian Maximum Entropy River Library, BME-riverlib [149] was used to process the polyline shapefile into MATLAB. Using this library, the river segment distances, the river topology, flow ratios, and flow connectivity were obtained for all points along the river network. Flow in sediment was approximated by using the Strahler order for each reach as a proxy.

Daily precipitation in centimeters and average monthly temperature in Celsius was obtained from the National Oceanic and Atmospheric Administration's (NOAA) climate data from April 1st, 2016 to August 1st, 2017 from across weather stations surrounding the study Kewaunee County, Wisconsin area (Brussels, Denmark WWTP, Forestville, Green Bay, Kewaunee, and New Franken). Inverse distance (D) weighted averages of daily precipitation (P) and monthly average temperature (T) were applied to each i^{th} study site to estimate values between April 1st, 2016 to August 1st, 2017 for each available time period (t'= day or month) and account for spatial variability. The precipitation $P_{it'}$ and the temperature $T_{it'}$ at location i and time t' are calculated as

$$P_{it'} = \frac{\sum_{w=1}^{6} \frac{P_{wt'}}{D_{iw}}}{\sum_{w=1}^{6} 1/D_{iw}} \text{, and} \qquad \text{(Eq. A1)}$$

$$T_{it'} = \frac{\sum_{w=1}^{6} \frac{T_{wt'}}{D_{iw}}}{\sum_{w=1}^{6} {}^{1}/_{D_{iw}}}\,, \qquad \text{(Eq. A2)}$$

respectively, where $P_{wt'}$ and $T_{wt'}$ are the precipitation and temperature, respectively, at weather station w and time t', and D_{iw} is the distance between locations i and w.

A1.2 Processing Spatial Data to Create Candidate Databases of Microbial Contamination Sources

A1.2.1 Description of each spatial databases

Here we consider two potential sources of microbial contamination in rivers of the Kewaunee watershed: AFOs, where manure is stored, and manure application fields, where manure is land applied. For each of these two potential contamination sources we obtained various candidate databases of the spatial locations of that source (Table A1).

For AFOs, these databases consisted of county databases of manure storages for Kewaunee County that were remotely sensed from aerial imagery for Kewaunee County either weighted by the log-total gallons of manure (AFO database option 1) or unweighted (AFO database option 2). Alternatively, the Wisconsin Pollution Discharge Elimination System (WPDES) database of CAFOs, which are AFOs with greater than 999 animals with unweighted points were used (AFO database option 3). These candidate databases for AFOs differ in the year the database was completed, information provided in the database, approach to obtaining geographic location (i.e., remote sensing versus geocoding), and that CAFOs (AFOs with greater than 999 animal units) can be considered a subclass of AFO.

For manure application fields, there was the option of a land cover database of crop rotation (manure application fields option 1) and of a land cover database of dairy rotation (manure application fields option 2) which both were represented as centroids of land cover polygons weighted by the area of the polygon corresponding to crop rotation (option 1) and dairy rotation (option 2), where dairy rotation is a subclass of crop rotation from the Wiscland-2 land cover database. [135]

Table A.1 Description of the candidate databases associated with each potential contamination source of bovine Bacteroides (BoBac) in the river sediment in Kewaunee County.

Potential Contamin. Source	**Candidate Spatial Database**	**Database description**
Animal Feeding Operations (AFO)	Option 1: County Database of Manure Storages (weighted) **Figure A.1(a)**	Database providing centroids of each manure storage in Kewaunee County weighted by the storage capacity of the manure storages in log total gallons of manure as determined by the estimated dimensions of the manure storages. When this information was unavailable, an average of the available data on log total gallons was applied. The database was made in 2014. Data was provided through personal communication with the Kewaunee County Land and Water Conservation Department
	Option 2: County Database of Manure Storages (unweighted) **Figure A.1(b)**	Same provider as above but the centroids are equally weighted. In other words, they were not weighted by storage capacity in log total gallons.
	Option 3: WPDES Database of CAFOs (unweighted) **Figure A.1(c)**	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer
	Option 4: WPDES Database of CAFOs (weighted by animal units)	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer
Manure Application Fields (ManureApp)	Option 1: Land Cover Database of Crop Rotation **Figure A.1(d)**	Database of crop rotation land cover polygon centroids weighted by the area of the land cover polygons. "The 30- meter raster layer combines land cover information from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) with land management information to map five-year crop rotations over Common Land Units (CLUs)… The crop rotation dataset was validated independently from Wiscland 2 using data from Department of Agriculture, Trade and Consumer Protection (DATCP) cattle inventory records and dairy producer data, NASS crop acreage estimates and transect surveys. The CDL datasets on which the crop rotation layer is primarily based also provide accuracy information by state and by class. Reported accuracies for the Principal Crops (e.g. Corn, Soybeans, Alfalfa, Vegetables, etc.) range from 87-93% over the 2008-2012 period; the accuracy of the crop rotations will differ due to the spatiotemporal aggregation of classes and the integration of the crop classification with CLU boundaries" [135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover

	Option 2: Land Cover Database of Dairy Rotation **Figure A.1(e)**	Database is of dairy rotation land cover polygon centroids weighted by the area of the land cover polygons. Dairy rotation land cover is a subclass of crop rotation land cover from the Wiscland-2 land cover database mentioned above [135]. Dairy rotation land cover is described by “corn grain, corn silage, and alfalfa in a 6-year rotation, with typically 2 years of corn, 4 years of alfalfa, and a cover crop between corn and alfalfa years. Occasionally, soybeans or a grain (e.g., wheat, barley, or oats) are planted in place of corn” [135]. Dairy rotation areas are typically associated with use by dairy farms and operations for manure application and feed.

A1.2.2 Visual representation of each spatial databases

Figure A.1(a) shows AFO option 1, which are centroids of manure storages with weights corresponding to the log total gallons of manure estimated based on the size of the manure storage. These weights are depicted on a color scale. Figure A.1(b) shows AFO option 2, which comes from the same database as above, but equally weights each manure storage. Figure A.1(c) shows AFO option 3, which are points corresponding to geocoded locations from WPDES permits, which is why the basemap displayed is that showing streets, rather than imagery. Figure A.1(d) depicts manure application fields option 1, which are from the Wiscland-2 land cover database showing crop rotation centroids that are proportional to the area of the land cover polygons, from which the centroids were derived. Figure A.1(e) depicts dairy rotation centroids from the same land cover database where the centroids are weighted proportionally to the size of the dairy rotation land cover polygons.

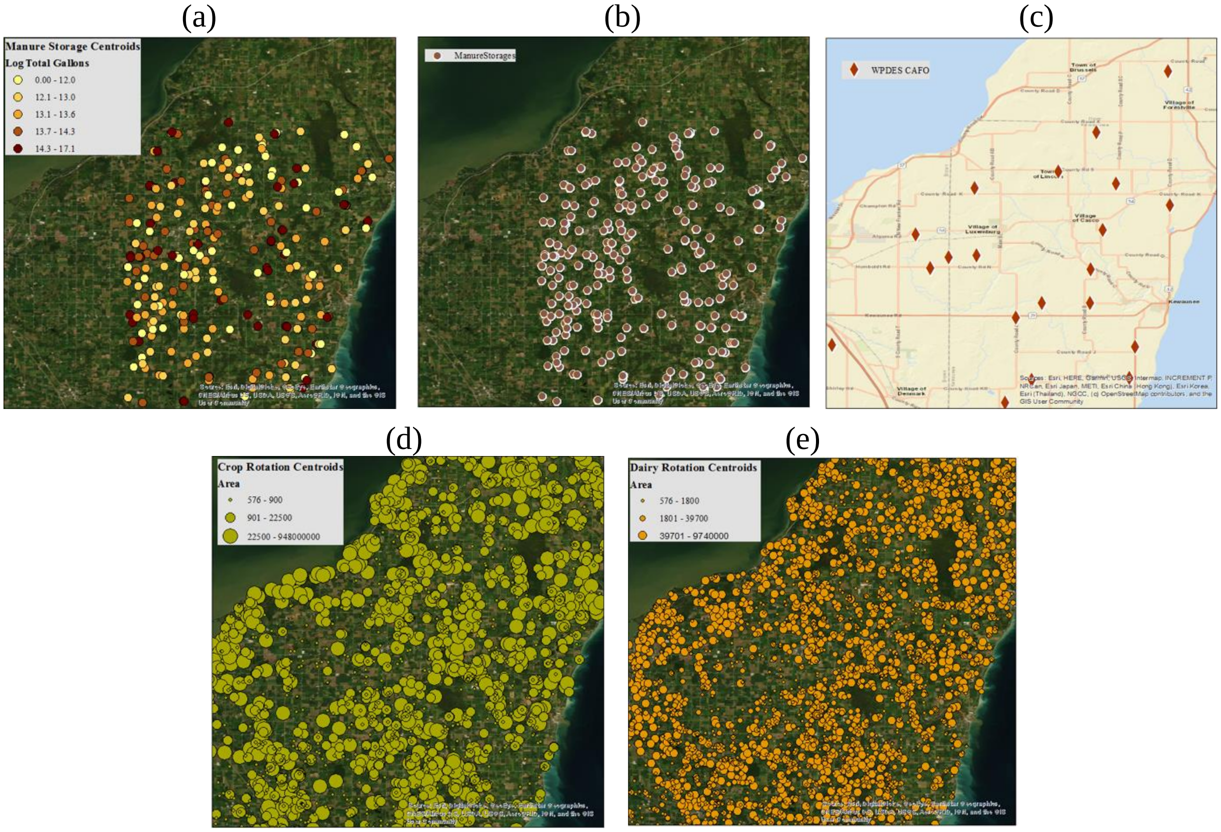


Figure A.1 (a) County database of manure storages weighted by log total gallons of manure (b) County database of manure storages (unweighted) (c) WPDES database of CAFOs (AFOs with >999 dairy cattle) (d) Land cover database of crop rotation (e) Land cover database of dairy rotation

A1.2.3 Details on augmenting the county database of manure storages outside of Kewaunee county, WI

No manure storage locations were remotely sensed outside of Kewaunee County, though one of the 20 sampling sites is located in Door County, rather than Kewaunee. Rather than omitting Door County data, manure storage information was simulated in bordering counties. Gridded points located approximately 483 m apart were applied outside of Kewaunee County. Values were assigned to the gridded points to represent the average density of manure storage points within Kewaunee County. To account for the non-homogenous point process, the manure storage density at each grid-point inside Kewaunee County was regressed using percent land-cover and road density. This model was applied to the density at grid-points outside of

Kewaunee, where manure storage data was missing, but land cover and road data were available (road shapefiles were obtained from county open data GIS sites for Kewaunee, Manitowoc, Door, and Brown). An example of estimating the AFO option 1 data outside of Kewaunee County is displayed in Figure A.2. Using this model, manure storage points were modelled as centroids per 610 m x 610 m quadrant outside of Kewaunee county indicating a density of manure storage points within the quadrants adjusted for high-density developed areas and dairy rotation. The simulated density outside of Kewaunee county was set to the actual density inside of Kewaunee county.

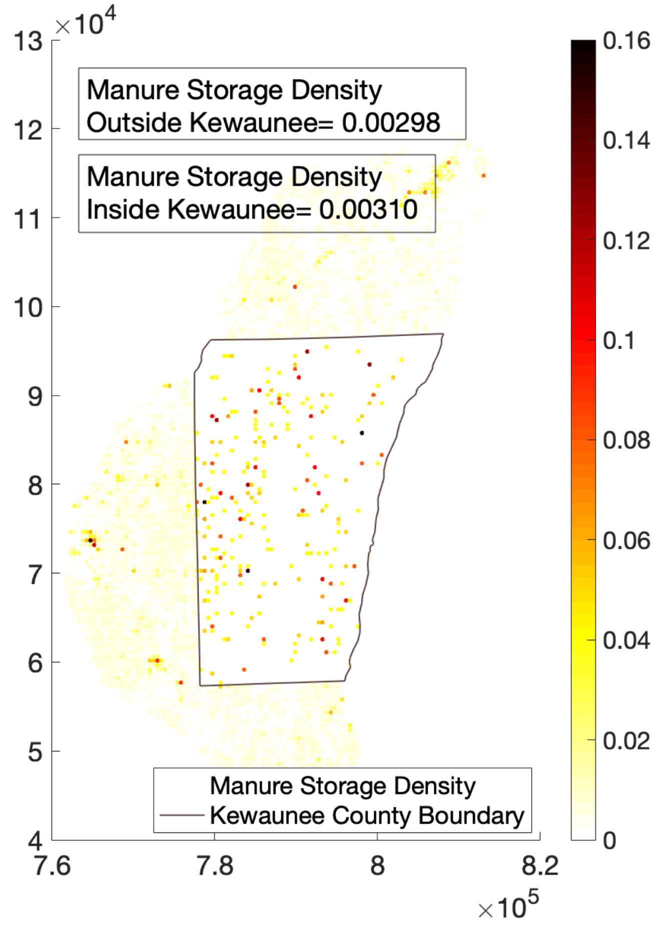


Figure A.2 Observed manure storage density per each 610 m x 610 m quadrant in Kewaunee County and simulated manure storage density per each 610 m x 610 m quadrant outside Kewaunee County.

A1.3 Bovine Bacteroides Primer Sets

To provide estimates of fecal contamination from specific hosts, copies of genes can be quantified. These are genes that have been verified to be specific or associated with certain animal hosts. The observation of these in the environment can be used to understand which animal hosts might be contributing to fecal contamination. The field which focuses on this is called microbial source tracking [28].

In our study, our objective is to understand the impact AFOs and manure application fields have downstream in riverbed sediment. We have chosen to utilize a ruminant-specific fecal marker, *BoBac* [136], that is highly associated with cattle, in particular. Because measured copies of this gene can vary due to things like biomass or extraction rates, we normalize by a measure of biomass done by measuring the copies of 16S rRNA, which is present in all bacterial species.

To measure these gene copies, we have used the methodology described in Layton et al., who use quantitative Polymerase Reaction Chain (qPCR). Essentially, this method relies on primer sets binding to target genes and measuring a fluorescent light that is produced due to a probe when this binding occurs. In order to make this method work the following information is needed: the primer sets, product size, and annealing temperature. These are provided in Table A.2 for *BoBac* and *16S rRNA.* For more details see Layton et al. [136]

Table A.2 Primer sets (forward, reverse, probe), product sizes, and annealing temperature for *BoBac* and *16S rRNA*

Target Gene	**Primer Sets**	**Size of Product (Base Pairs)**	**Annealing Temperature**
BoBac	BoBacF 5'-GAAG(G/A)CTGAACCAGCCAAGTA-3′ BoBacR- 5′-GCTTATTCATACGGTACATACAAG-3′ BoBacProbe- (FAM)TGAAGGATGAAGGTTCTATGGATTGTAAACTT(BHQ-1)-3′	100bp	57 ºC
16S rRNA	Universal V3 region primers 314F- 5'-CCTACGGGAGGCAGCAG-3' 518R- 5'-ATTACCGCGGCTGCTGG-3'	204bp	58 ºC

A1.4 Relative versus Absolute Abundance

The absolute abundance of *BoBac* in a sample (gene copies/g) has variability that is influenced by factors driving the variability of *BoBac* in the environment, as well as extraneous factors explaining the variability of the overall microbial biomass in the sample (as quantified by 16S rRNA gene copies), including laboratory factors such as biomass extraction rates factors, and climatic factors influencing the microbial community as a whole. By normalizing the absolute abundance of *BoBac* with 16S rRNA, we obtain a relative abundance (*BoBac* copies per 16S copies) that captures the proportion of *BoBac* in the microbial biomass. This variable offers two potential attractive features: It lessens the effect of laboratory variability on extraction rate of biomass, and it has variability that is influenced by environmental factors that are more specific to *Bovine* Bacteroides, a bovine marker, rather than the larger microbial community.

Temporal averages of the relative abundance are shown across sites in Figure A.3 with relation to some of the available spatial databases such as Wiscland2 land cover and a county database of AFO manure storages within the boundaries of Kewaunee County. One site is not within this frame and is located in Door County just north of Kewaunee.

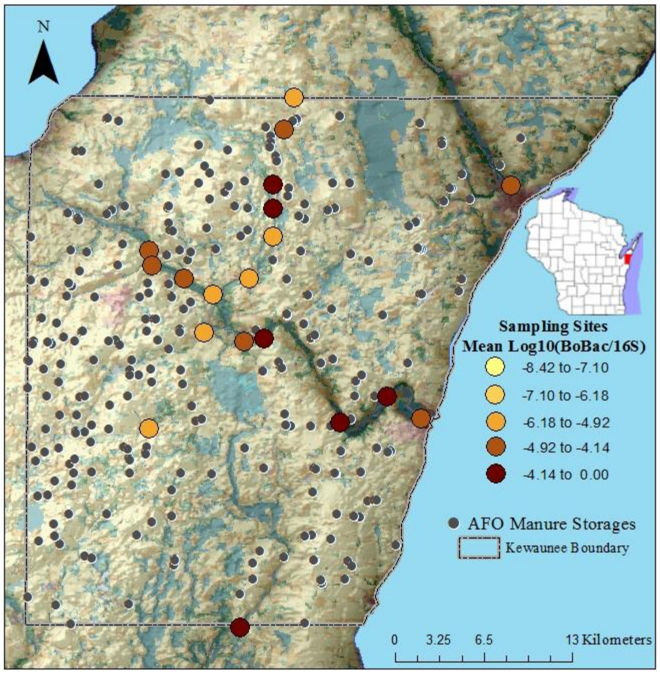


Figure A.3 Spatial distribution of sampling sites with measure of the temporally averaged relative abundance of ruminant *Bacteroides* from lowest values (yellow) to highest (maroon) in relation to AFO manure storages from a county database, land cover categories, overlaid on Digital Elevation Model. Blue shaded areas indicates wetlands land-cover, red is impervious surfaces, green is forest, and yellow is cropland from the Wiscland2 land-cover database. Lake Michigan is the light blue background.

A1.5 Relative Abundance Ratio

A1.5.1 Classical definition of risk ratio

Let’s consider the case where z_i is the risk of an adverse health outcome for the *i*–th observation of some study, and we use the natural base log to define the log risk $y_i = \log(z_i)$.

Without loss of generality, let’s consider the case where y_i is expressed in terms of a linear regression with respect to two predictors

$$y_i = \log(z_i) = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \varepsilon_i, \qquad \text{(Eq. A3)}$$

where β_1 and β_2 are the linear regression coefficients for predictors $x_i^{(1)}$ and $x_i^{(2)}$, respectively. The risk ratio for predictor $x^{(1)}$ is defined as the ratio of risks for a one unit increase in $x^{(1)}$:

$$\begin{aligned} RR^{(1)} = \frac{z(x^{(1)}+1)}{z(x^{(1)})} &= \exp\left(\log\left(\frac{z\left(x^{(1)}+1\right)}{z(x^{(1)})}\right)\right) = \exp\left(\log\left(z\left(x^{(1)}+1\right)\right) - \log\left(z\left(x^{(1)}\right)\right)\right) \\ &= \exp\left(\left(\beta_0 + \beta_1(x^{(1)}+1) + \beta_2 x^{(2)}\right) - \left(\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)}\right)\right) \\ &= \exp\left(\left(\beta_0 + \beta_1 x^{(1)} + \beta_1 + \beta_2 x^{(2)}\right) - \left(\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)}\right)\right) = \exp(\beta_1) \\ &= e^{\beta_1} \end{aligned}$$

Hence, the *RR* for predictor 1 is obtained by raising *e* to the power of β_1

A1.5.2 Definition of relative abundance ratio for log10 transformed data

We now consider the case where z_i is the relative abundance of *BoBac* (*BoBac* gene copies per 16S rRNA gene copies) in the *i*–th sample, and the response variable in the regression is the log-base-10 of the relative abundance, $y_i = \log_{10}(z_i)$. Using Equation 3.1 when there are no modifier, and focusing on source *u*, we get

$$y_i = \log(z_i) = \beta_0 + \cdots + \beta_u s_i^{(u)} + \cdots + \varepsilon_i,$$

where β_0 is the intercept and β_u is the linear regression coefficient for $s_i^{(u)}$. Similar to a RR, we can define the Relative Abundance Ratio $RAR^{(u)}$ as the ratio of relative abundance for a one standard deviation increase in source *u*, i.e. $RAR^{(u)} = z(s^{(u)}+1)\,/z(s^{(u)})$. It follows that

$$\begin{aligned} \log_{10}(RAR^{(u)}) = \log_{10}\left(\frac{z\left(s^{(u)}+1\right)}{z(s^{(u)})}\right) &= \log_{10}\left(z\left(s^{(u)}+1\right)\right) - \log_{10}\left(z\left(s^{(u)}\right)\right) \\ &= \left(\beta_0 + \cdots + \beta_u\left(s^{(u)}+1\right) + \cdots\right) - \left(\beta_0 + \cdots + \beta_u s^{(u)} + \cdots\right) \\ &= \left(\beta_0 + \cdots + \beta_u s^{(u)} + \beta_u + \cdots\right) - \left(\beta_0 + \cdots + \beta_u s^{(u)} + \cdots\right) + \cdots) = \beta_u \end{aligned}$$

which yields

$$RAR^{(u)} = 10^{\beta_u}$$

Hence, the *RAR* for source *u* is obtained by raising 10 to the power of β_u.

Similarly, where no normalization occurs on the response, but the response is still log10 transformed, an absolute abundance ratio (AAR) can be calculated as

$$AAR^{(u)} = 10^{\beta_u}$$

A1.6 Selecting the Hyperparameters of the Precipitation Predictors P1 and P2

The more recent precipitation variable $P1$ is calculated, for a given time *t*, as the weighted average of precipitations prior to *t*, with a weight that decreases over a time range α_{P1} (Eq. 3.4). The less recent precipitation variable $P2$ is calculated for a time range α_{P2} (Eq. 3.5) extending further back in time, i.e., α_{P2}>α_{P1}. Recent precipitations generally increase microbial contamination because rain transport microbial contaminants that have deposited on land, while less recent precipitations may wash off that deposit and therefore have a diminishing effect on microbial contamination. Hence, the regression coefficient $\beta_1 > 0$ represents the main effect that recent precipitation has on microbial response, while $\beta_2 < 0$ represents the diminished effect due to washout from less recent precipitation. We obtain the hyperparameters α_1 and α_2 by minimizing an objective function $f_P(\alpha_{P1}, \alpha_{P2})$ that maximizes the difference between β_1 and β_2 when β_1 is positive and statistically significant, or otherwise simply maximizes the main effect β_1, i.e.,:

$$f_P(\alpha_{P1}, \alpha_{P2}) = \begin{cases} -(\hat{\beta}_1(\alpha_{P1}) - \hat{\beta}_2(\alpha_{P2})) \text{ if } \hat{\beta}_1(\alpha_{P1}) > 0 \text{ and } p - value(\hat{\beta}_1(\alpha_{P1})) < 0.05 \\ 1/10^{(\hat{\beta}_1(\alpha_{P1}))} \qquad\qquad \text{otherwise} \end{cases} \quad \text{(Eq. A4)}$$

where $\hat{\beta}_1(\alpha_{P1})$ and $\hat{\beta}_2(\alpha_{P2})$ are the regression coefficients in the microbial regression equation (Eq. 3.1) restricted to the climatic variables (*P1*, *P2*, and *Freezing*). The values for α_{P1} and α_{P2} are selected by minimizing the objective function subject to the constraint α_{P2}>α_{P1}.

This objective function has two distinct regions in the $(\alpha_{P1}, \alpha_{P2})$ space constrained to α_{P2}>α_{P1}, corresponding to line 1 and 2, respectively, of Eq. S4. The first region consists of the

$(\alpha_{P1}, \alpha_{P2})$ values such that $\alpha_{P2} > \alpha_{P1}$, and for which the conditions $\hat{\beta}_1(\alpha_{P1}) > 0$ and $p-value(\hat{\beta}_1(\alpha_{P1})) < 0.05$ are met. In that region, minimizing the objective function is done by maximizing the difference between $\hat{\beta}_1(\alpha_{P1})$ and $\hat{\beta}_2(\alpha_{P2})$, which is our goal, and translates with a negative objective function. The other region is where the condition is not met. In that region the exponentiated form of the objective function guarantees that the objective function is positive. Hence if both regions exist, then the minimum will be in the first region where β_1 is positive and statistically significant and the difference between β_1 and β_2 is maximized. If the first region does not exist, then only the second region exist. Minimizing the objective function in the second region is effectively the same as finding the value of α_1 that maximizes β_1.

The minimization with constraint was implemented using the constrained minimization function fmincon.m in MATLAB version 2018b.

A1.7 Selecting the Hyperparameters for the Contamination Source Terms Using Constrained Optimization

A1.7.1 Tutorials and examples using simulated data

A1.7.1.1 The set up

In the FIT framework we select the hyperparameters $\boldsymbol{\alpha}^{(u)}$ of the source terms $s_i^{(u)}(\boldsymbol{\alpha}^{(u)})$ in the microbial regression equation (Eq. 3.1) by choosing those that maximize the RAR (10^{β_u}) for some training set of observed outcomes y_i, where $y_i = \log_{10}(z_i)$, and z_i (copies of BoBac RNA per copies of 16S RNA) is the relative abundance of BoBac.

In order to illustrate how the approach works without loss of generality, we present here a tutorial using simulated data for the case where there is only one source term in the microbial regression equation (Eq. 3.1), and when this source term follows the Euclidean model, i.e., it corresponds to the sum of exponentially decaying contributions (SEDC) from source locations. Hence, in this case, the observation y_i at sampling location *i* is expressed as

$$y_i = \beta_0 + \beta_u s_i^{(u)}\left(\alpha_E^{(u)}\right) + \varepsilon_i \qquad \text{(Eq. A5)}$$

where β_0 is the intercept, β_u is the linear regression coefficient, ε_i is a random measurement error, and the standardized predictor $s_i^{(u)}\left(\alpha^{(u)}\right)$ is expressed as the z-score of the sum of exponentially decayed contributions from surrounding sources *j*:

$$s_i^{(u)}\left(\alpha_E^{(u)}\right) = z - score\left(\sum_{j=0}^{N} C_{0j} \mathrm{e}^{\frac{-3D_{ij}{}^{(E)}}{\alpha_E}}\right) \qquad \text{(Eq. A6)}$$

where $D_{ij}^{(E)}$ is the Euclidean distance between sampling point *i* and source j, C_{0j} is the initial value of *y* at source j, and $\alpha_E^{(u)}$ is the Euclidean hyperparameter corresponding to the distance away from a source where an initial source concentration of C_{0j} is reduced by 95%.

The problem at hand consists in estimating the hyperparameter $\alpha_E^{(u)}$ given some observational data y_i. In this tutorial we illustrate how the hyperparameter is estimated by creating a simulated truth where the hyperparameter value $\alpha_E^{(u)}$ is known, from this simulated truth we create a training dataset of observed values y_i, using these observed values we apply our method to obtain an estimate $\hat{\alpha}_E^{(u)}$ of the hyperparameter value, and we compare the estimated hyperparameter value $\hat{\alpha}_E^{(u)}$with its simulated truth $\alpha_E^{(u)}$.

A1.7.2 Simulating the truth

The simulated truth is shown in Figures A.4(a) to A.4(c). We set the simulated truth as having the 100 source locations *j* shown in Figure A.4(a). Each source location *j* generates an initial value C_{0j} that decreases exponentially away from the source. For this simulation, all C_{0j} are set equal to a relative abundance of 1 (*l*og10(*BoBac* copies/16S rRNA copies)). We set the true exponential decay range to $\alpha_E^{(u)}$=2 km. Using this hyperparameter value, we may determine the standardized predictor, $s_i^{(1)}\left(\alpha_E^{(1)}\right)$, across our study area as shown in Figure A.4(b). Additionally, the true response, y, is modeled in Figure A.4(c) as

$$y = 1.5 + 0.25 s_i^{(1)}\left(\alpha_E^{(1)}\right) \qquad \text{(Eq. A7)}$$

where Eq. (A7) was modeled using the hyperparameter value, $\alpha_E^{(1)}$ as 2 km, β_0 is set to 1.5 *l*og10(*BoBac* copies/16S rRNA copies), and β_1 is set to a 0.25 (*l*og10(*BoBac* copies/16S rRNA copies)) increase in the response per one standard deviation-increase in the SEDC. This corresponds to an RAR or 1.778 (unitless).

Then we simulate an observed response, y, as modeled in Figure A.4(d) as

$$y = 1.5 + 0.25 s_i^{(1)}\left(\alpha_E^{(1)}\right) + \varepsilon \qquad \text{(Eq. A8)}$$

where some measurement error, ε, is added as 10% of a randomly generated, normally distributed variable with a mean of 0 and a standard deviation of 1.

We randomly select 90 points from our study area in order to simulate 90 sampling sites where there are observed responses determined by Eq. (A8) shown in Figure A.4(d).

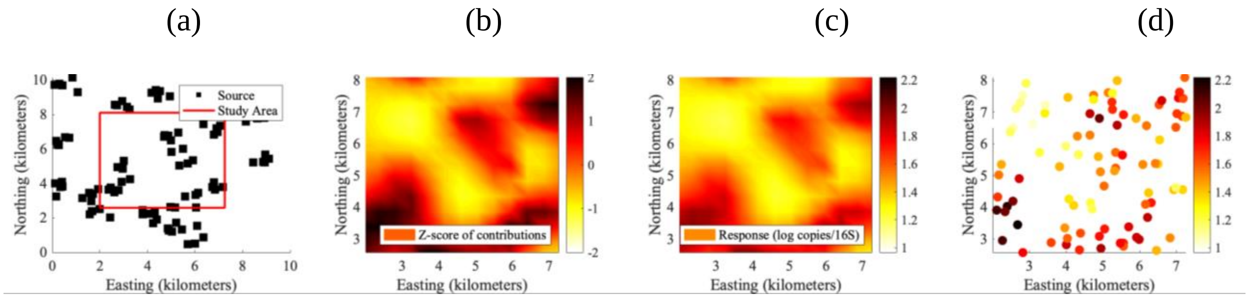


Figure A.4 From left to right: a) source locations in a study area which is outlined in red, b) true contributions to polluting agent with the knowledge of the exponential decay range of pollution emitted from sources, $s_k^{(u)}(\alpha_E = 2km)$, across the k points of the study area, which is not known, c) true concentrations of the polluting agent across the study area, $y_k = \beta_0 + \beta_u s_k^{(u)}(\alpha_E = 2km)$, which is unknown d) observed concentrations of the polluting agent at each ith sampling site, $y_i = \beta_0 + \beta_u s_i^{(u)}(\alpha_E = 2km) + \epsilon_i$

A1.7.2.1 Performing the experiment

In order to model the true response (Figure A.4(c)) as best as possible, the true exponential decay range $\alpha_E^{(u)}$, needs to be estimated. The challenge is how to estimate this hyperparameter. A numerical approach to this estimation is to select different hyperparameter values to construct the standardized predictor, $s_i^{(1)}\left(\alpha_E^{(1)}\right)$, compare the standardized predictor to the observed response (Figure A.4(d)) using a regression, and then observe the estimated coefficients and model fit from the regression. In Figure A.5(1a) the standardized predictor, $s_i^{(1)}\left(\alpha_E^{(1)}\right)$, has been modeled using an experimental exponential decay range, $\alpha_E^{(u)}$, equal to 0.815 km. From the regression shown in in Figure A.5(1b), we obtain an R-squared of 0.62 and a standardized regression coefficient indicating an increase of 0.218 in the relative abundance (log10 copies/16S rRNA) for a one standard deviation increase in the sum of exponentially decaying contributions from sources. Alternatively stated, the RAR is 1.65. The resulting model from this first example's experimental decay range can be seen in the background of Figure A.5(1c) and compared to the observed values shown in the points, which are the same as shown in Figure A.4(d). We see that in Figure A.5(2a) looks nearly identical to Figure A.4(b). Figure

A.5(2b) shows a regression where the residuals are relatively small compared to Figures A.5(1b) and A.5(3b). This is reflected in its having the highest R-squared. Additionally, the standardized regression coefficient in the second example is closest to that of the simulated model of 0.25. These are only three examples of experimental exponential decay ranges. Among these three, the second example would be the best selection as an estimate of the true exponential decay range due to the model fit or the estimate of the true standardized regression coefficient from Eq. (A7).

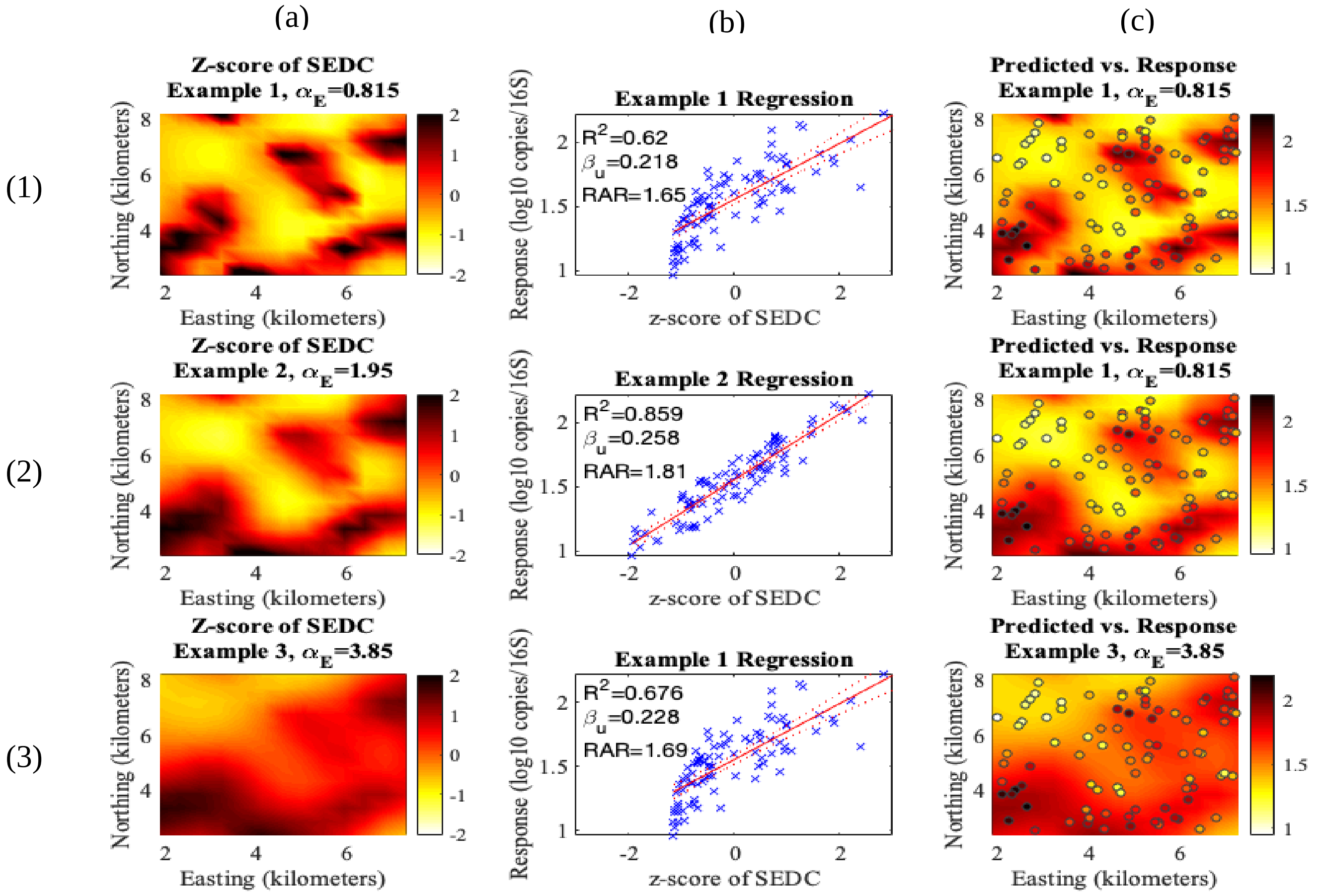


Figure A.5 Demonstration of estimating the hyperparameter value, α_E, numerically. A different predictor variable can be constructed with different hyperparameter values to create a map predicting the response that is observed (here as relative abundance of a gene) for sampling locations. In examples 1-3 (distinguished by rows), column (a) shows the standardized predictor or the z-score of the SEDC, $s_i^{(1)}\left(\alpha_E^{(1)}\right)$, for different values of the hyperparameter value, α_E. In examples 1-3, the hyperparameter values are 0.815 km, 1.95 km, and 3.85 km. Column (b) shows a regression of the standardized predictor (unitless) against the response, which in this example is a relative abundance (log10 copies /16s). Column (c) compares the predicted values using a regression model with experimental hyperparameter in the background and the observed values at the 90 sites as dots.

Experimental exponential decay ranges could be attempted for a range of values from 0 to 6 kilometers. The R-squared and standardized coefficient values could be obtained from each of these regressions to plot what these values look like as a function of the experimental exponential decay ranges. Figure A.6 shows what these plots look like and demonstrates that a peak occurs for the model fit and RAR (function of the standardized regression coefficient) at approximately $\alpha_E = 1.95\ km$. This value rounds to 2 kilometers, which was the true exponential decay range.

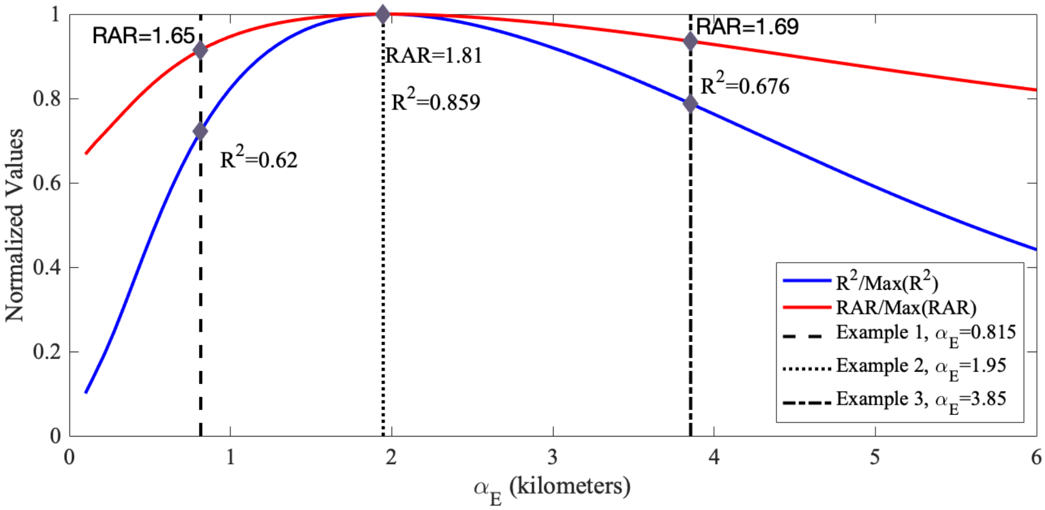


Figure A.6 Normalized R-squared and RAR values as a function of experimental exponential decay range hyperparameter values. The three examples from Figure A.5 are shown in black lines from left to right. Example 2 happens to exist at the peak of both R-squared and RAR curves.

An in-depth tutorial and interactive demonstration of the SEDC and selecting hyperparameters can be found at the following link: https://mserre.sph.unc.edu/BMElab_web/SEDCtutorial/. In this example there are 7 sources (or source nodes of the same source type) and 20 sampling locations. Additionally, the truth has been simulated twice. In the first example the true exponential decay range is also 2 km. However, in the second example, the true exponential decay range has been set to 4 km.

A1.7.2 Implementation details for BoBac in Kewaunee County

For both the "Find" and "Inform" stages, the RAR was derived from the coefficient resulting from a weighted regression, where each spatial location was equally weighted to account for samples missing at random. For the "Find" stage, to find the hyperparameter values which maximize the RAR from a weighted regression, the function, fmincon.m, was used in MATLAB 2018b for constrained minimization. The interior-point algorithm [260], which is the default option, was used and the objective function was set to 1/RAR in order to find which hyperparameters correspond to the maximum RAR. For the "Inform" stage, a penalty, $\phi(\gamma_G, \alpha_O)$ is added to the objective function. This penalty is based on the shape function, $\varphi(\gamma_G, \alpha_O)$ (Eq. A9, Figure A.7 which penalizes combinations of γ_G and α_O yielding poor regression and mapping qualities (i.e., non-normal residuals, estimated values of the log10 relative abundance greater than 0).

$$\phi(\gamma_G, \alpha_O) = \left(\frac{\varphi(\gamma_G, \alpha_O)}{1}\right)^{100}, \text{ where } \varphi(\gamma_G, \alpha_O) = \left(\frac{10000-\gamma_G}{10000}\right)^{8} + \left(\frac{10000-\alpha_O}{10000}\right)^{8} \quad \text{(Eq. A9)}$$

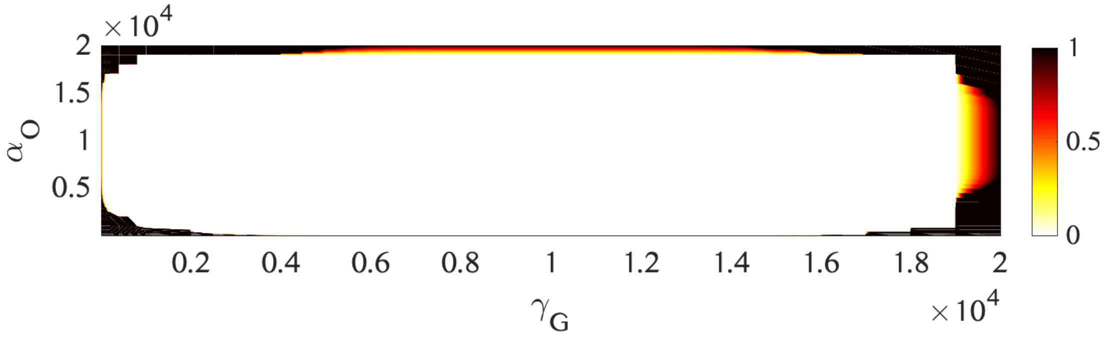


Figure A.7 Penalty added to RAR across different values of α_O and γ_G.

Upper and lower bounds were used as constraints in the minimization search. The lower bounds were all set to 50 m. Upper bounds for the Euclidean and ground transportation distance parameters were set to be extremely long at 34.0 km. Due to tortuosity, the river distance parameter upper bound was set to 340 km, a value greater than the Euclidean distance decay hyperparameter upper bound. The overland distance upper bound parameter was set to 10 km, which is greater than the maximum distance from a watershed's edge to any of the points along the river network.

To be more likely to capture the global rather than local minimum, fifteen random restarts were used with seeding for reproducible results. When the initialization point, $\boldsymbol{\alpha}_{\boldsymbol{init}}^{(\boldsymbol{u})}$, corresponds to a slope of 0 for the objective function, it was considered a locally flat minimum and the associated optimal hyperparameter value was set equal to the initialization point, $\boldsymbol{\alpha}_{\boldsymbol{init}}^{(\boldsymbol{u})}$. For each random restart, we got an optimal hyperparameter, $\widehat{\boldsymbol{\alpha}}^{(\boldsymbol{u})}$, and objective function value $f(\widehat{\boldsymbol{\alpha}}^{(\boldsymbol{u})})$. The hyperparameter that was selected was that which corresponded to the most minimum objective function value from all the restarts. If there were ties between the minimum objective function values, $f(\widehat{\boldsymbol{\alpha}}^{(\boldsymbol{u})})$, resulting from different initialization points, then the hyperparameters, $\widehat{\boldsymbol{\alpha}}^{(\boldsymbol{u})}$**,** corresponding to those tied minimum objective function values were averaged.

A1.8 Details of the Reliability Score Calculation

For each potential contamination source *u*, we have various candidate databases *d* that provide the locations of that source. We want to calculate a reliability score for each of the candidate database, so we can rank them and select the one that most reliably represents a source. For instance, for *u*=AFO, we may have four candidate databases denoted by index *d*=1, 2, 3 and 4, and we may want to select only one of these three candidate databases that most reliably captures the contamination effect that AFOs have on *BoBac* in sediment. Here we define reliability as the degree to which the effect that is calculated remains the positive when we change from the training set to the validation set. First, we consider the microbial regression equation (Eq. 2) without climatic variables and the source term *u*, i.e., the following equation, which we refer to as model *a*:

Model a: $y_i = \beta_0 + \beta_u s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$ (Eq. A10)

where y_i is the observed $\log_{10}$ relative abundance of *BoBac* at sample *i* and β_u is the regression coefficient measuring the effect of the source *u* where the source predictor values $s_i^{(u)}\left(\boldsymbol{\alpha}^{(u)}\right)$ are calculating with one of the source models (Eqs. 3.6-3.8) with hyperparameter $\boldsymbol{\alpha}^{(u)}$. We divide the observations y_i into k=5 test sets of approximately equal size (5-folds). A training set is defined by all of the other sets. We select the hyperparameter $\boldsymbol{\alpha}^{(u)}$ by finding that which maximizes β_u in model *a* for each training set, and we denote that selected hyperparameter as $\widehat{\boldsymbol{\alpha}}_{training\_k}^{(u)}$. Using $\widehat{\boldsymbol{\alpha}}_{training\_k}^{(u)}$ we re-calculate the regression coefficient using model *a* with the test set, and we denote it as $\beta_u^{(test\_k)}$. A database that reliably represents a source *u* should yield similar positive values of $\beta_u^{(test\_k)}$ for each k$^{\text{th}}$ fold.

An important component of the reliability of the database is if it produces similar signs for the regression coefficients. We assess this through the Sign Stability Score *SSS* calculated as

$$SSS = \left(\sum_{test=1}^{5} I\left(\beta_u^{(test\_k)} > 0 \right) \right) + 1$$

(Eq. A11)

The higher the *SSS*, the more often that a hyperparameter, $\widehat{\boldsymbol{\alpha}}_{training\_k}^{(u)}$, selected using the training data produced a source term for the test data that was positively associated with the microbial response. If a database for source *u* yields an *SSS* of 1 it means that this database indicates that source *u* does not increase contamination. The maximum value of the *SSS* is a value of k+1, means that for unseen data, a hyperparameter can be selected that can be used to construct a predictor such that the slope for that term from a regression consistently indicates that source *u* increases contamination.

Another important component of reliability of the database is if the magnitude of the kth effect, $\beta_u^{(test\_k)}$, is strong. We assess this through taking the sum of the $\beta_u^{(test\_k)}$ across the folds to represent an average magnitude, *M*.

$$M = \sum_{test=1}^{5} \beta_u^{(test\_k)} \quad \text{(Eq. A12)}$$

Lastly, those databases that produce consistently positive coefficients, $\beta_u^{(test\_k)} > 0$, have a small standard deviation of the coefficients, $\sigma_{\beta_u^{(test\_k)}}$.

$$\sigma_{\beta_u^{(test\_k)}} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\beta_u^{(test\_k)} - \overline{\beta_u^{(test\_k)}})^2} \quad \text{(Eq. A13)}$$

,where $\overline{\beta_u^{(test\_k)}}$ is the mean of the coefficients $\beta_u^{(test\_k)}$

Do to all of these components, our overall reliability score *RS* rewards high SSS, high M, and low $\sigma_{\beta_u^{(test\_k)}}$ values. This is done using the following formula:

$$RS = SSS \times M/(\sigma_{\beta_u^{(test\_k)}}) \quad \text{(Eq. A14)}$$

A1.9 Details on Performance Statistics and Cross-Validation

A1.9.1 Performance statistics

Stepwise selection of the predictors in the linear regression model (Eq. 3.1) was performed using the Akaike Information Criterion (AIC) which rewards fit but includes a penalty that increases with the number *k* of predictors. The AIC is given by

$$\text{AIC}=2k - 2l \quad \text{(Eq. A15)}$$

where *l* is the log-likelihood function for the linear regression. For example, if, without loss of generality, we let β be the regression coefficient value for the first predictor considered, σ^2 be the variance of the error terms, $\boldsymbol{y}$ be the vector of the response values, and $\boldsymbol{x}$ be the vector of values for the predictor considered, then the log-likelihood function *l* is expressed as

$$l(\beta, \sigma^2; \boldsymbol{y}, \boldsymbol{x}) = \ln\big(L(\beta, \sigma^2; \boldsymbol{y}, \boldsymbol{x})\big) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2 \quad \text{(Eq. A16)}$$

Stepwise selection consists in adding predictors one at a time until the AIC fails to decrease. Since the AIC includes a penalty for the number of predictors, then overfitting is discouraged, which is desirable because increasing the number of predictors in the model almost always improves the goodness of the fit.

Similarly, we used the adjusted R-squared, R^2, to assess the fit of the final models. R-squared is a measure of fit and the adjusted R-squared, $\bar{R}^2$, penalizes when there are additional variables. The inputs of these metrics include the observed response values, y_i, the predicted values, $\hat{y}_i$, the mean of the response, $\bar{y}$, the degrees of freedom of the estimate (df_e= *n*-1) and the degrees of freedom of the estimate of the underlying population error variance (df_t= *n*-*k*-1), where *n* and *k* are the sample size and the number of predictors, respectively. Using these inputs, R^2 and $\bar{R}^2$ are expressed as

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad \text{(Eq. A17)}$$

$$\bar{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2/df_e}{\sum(y_i - \bar{y})^2/df_t} \qquad \text{(Eq. A18)}$$

A1.9.2 Cross-validation for “Test” step of the FIT framework

In order to understand how our selections from the “Test” step would work for unseen data, we used a 6-fold cross validation on the stepwise selection of predictors for our 90 observed response values. For this 6-fold cross validation, the 90 observed response values were stratified randomly into 6 distinct sets of 15 observed values each. Each of these set (or fold) was used one at a time as the validation set (of size 15) and held out. The remaining 5 sets were combined as the corresponding training set (of size 75) to select the predictors using step wise AIC selection, and the selected predictors were used to predict the response values for the validation set that was held out. The result of this 6-fold cross validation is shown in Table A.5. The selected predictors for each fold are shown in shaded cells, along with the average root mean square error (rMSE) quantifying how the 15 predicted values for that fold match the corresponding 15 validation values that were held out.

A1.10 Details of the Reliability Score Results

Detailed reliability scores are given in Table A.3. In comparing the reliability of the databases to represent sources of bovine *Bacteroides* in the sediment, we find that reliability scores are higher for GORF-modeled source predictors, followed by ORF, and then Euclidean. The GORF, ORF, and Euclidean models utilize 3, 2, and 1 hyperparameters, respectively. Because of this, there is a higher risk of overfitting for GORF-modeled predictors, which would yield less consistently positive coefficients describing the effect of the source locations on the response. Instead, we find that for this study the GORF-modeled predictors more consistently represent sources (*SSS* ranging between 3 and 5) than the ORF (*SSS* ranging between 1 and 5) or and Euclidean (*SSS* ranging between 2 and 5) predictors, and that this conclusion holds, for both AFO and manure application field predictors. This indicates that, regardless of the spatial database used, the GORF-modeled source predictors are more consistently being associated with contributions versus reductions. Additionally, the weighted standard deviation on the coefficient (Eq. A13) is higher for Euclidean versus ORF-modeled AFO and manure application field source predictors. Overall, this indicates that source predictors are more reliable when the modeling approach accounts for processes relevant to a river network, such as overland and river distance and flow and ground transportation to manure fields.

Across all modeling approaches, the databases most reliably representing AFOs as sources of bovine fecal contamination was the state database of geocoded CAFO locations from WPDES permits. CAFOs are AFOs with >999 animals, representing somewhat of a subclass to AFOs. It is possible that only AFOs of a certain scale are consistent sources of bovine fecal contamination, but it is also possible that the data quality associated with WPDES permits leads to a more reliable database compared to the county manure storage databases. The databases most reliably representing manure application fields as sources of contamination was dairy rotation land cover, which is a subclass of the crop rotation land cover, which receive daily haul manure.

Table A.3 Reliability score for the candidate database options associated with each potential contamination source of bovine Bacteroides (BoBac) in river sediment. Also shown is the sign stability scores (*SSS*, Eq. A11) and the weighted standard deviation on the coefficient (σ_{β_u}), which are used in the calculation of the reliability score. The database with the highest reliability score for each potential source is selected and shown in bold face. The databases were obtained from the following providers: County Database (Kewaunee County Land and Water Conservation), DNR *WPDES Database (Wisconsin DNR), Wiscland-2 Land Cover Database (Wisconsin State Cartographer's Office)*

Modelling Approach	Potential Contamination Source (*)	Candidate Spatial Database	Sign Stability Score *SSS* (Eq. A11)	Average Magnitude $SSS \times M$ (Eq. A12)	Reliability Score *RS* (Eq. S14)
Euclidean	AFO	Option 1: County Database of Manure Storages (weighted)	2	-2.95	-7.18
		Option 2: County Database of Manure Storages (unweighted)	2	-3.16	-7.92
		Option 3: WPDES Database of CAFOs (unweighted)	4	3.11	5.92
		Option 4: WPDES Database of CAFOs (weighted) *	**5**	**5.56**	**14.3**
	Manure Application Fields (ManureApp)	Option 1: Land Cover Database of Crop Rotation	4	1.03	2.63
		Option 2: Land Cover Database of Dairy Rotation *	**4**	**1.99**	**5.28**
ORF	AFO	Option 1: County Database of Manure Storages (weighted)	1	-1.98	-6.17
		Option 2: County Database of Manure Storages (unweighted)	2	-3.60	-9.54
		Option 3: WPDES Database of CAFOs (unweighted)	1	-1.19	-4.29
		Option 4: WPDES Database of CAFOs (weighted) *	**1**	**-1.10**	**-3.84**
	Manure Application Fields (ManureApp)	**Option 1: Land Cover Database of Crop Rotation ***	**5**	**1.74**	**3.39**
		Option 2: Land Cover Database of Dairy Rotation	4	-1.36	-2.41
GORF	AFO manure ground transported to Manure Application Fields	AFO Option 1 Ground Transported to ManureApp Option 1	3	-2.04	-3.67
		AFO Option 1 Ground Transported to ManureApp Option 2	3	-1.59	-3.04
		AFO Option 2 Ground Transported to ManureApp Option 1	3	-1.90	-3.48
		AFO Option 2 Ground Transported to ManureApp Option 2	3	-2.20	-3.82
		AFO Option 3 Ground Transported to ManureApp Option 1	5	2.33	4.36
		AFO Option 3 Ground Transported to ManureApp Option 2 *	**4**	**2.25**	**5.28**
		AFO Option 4 Ground Transported to ManureApp Option 1	4	0.0161	0.0305

	AFO Option 4 Ground Transported to ManureApp Option 2	4	1.60	3.45

(*) AFOs are potential contamination sources because they are where manure is stored. Manure Application Fields (ManureApp) are potential sources because they are where manure is land applied for nutrient management or fertilizer

A1.11 Details of the hyperparameter Results and Statistics for Source Terms

Hyperparameters describing Euclidean, overland and river distance, and ground transportation from the various spatial predictor models were successfully selected using the objective function to inform the spatial predictors explaining the relative abundance of bovine fecal contamination. See Figure A.8. Maximum values correspond to objective function minimums. For source predictors constructed with more than one hyperparameter, the figures represent transects of the inverse objective function or RAR surface at the maxima. Figure A.8(a) shows a peaking curve with a defined maximum across the Euclidean exponential decay ranges. Figure A.8(b) shows a clear peak before 50 m across the ORF river distance exponential decay ranges, but the minimum value is constrained to 50 meters. Across the overland exponential decay ranges in Figure A.8(c), there is a slight peak, but it appears relatively flat. For the GORF approach, hyperparameter transects can be seen in Figure A.8(d-f). Here there is a strong peak for the ground transportation hyperparameter γ_G and river distance hyperparameter, α_R. However, a locally flat maximum exists for α_0. The selected hyperparameter is an average of the initialization points that corresponded to the maximum RAR. One possible reason for the flatter curves associated with overland exponential decay ranges seen in Figure A.8 (c) and (e) may be the presence of tile drainage in Wisconsin where clay and concrete pipes have been placed under and around cropland areas in order to allow roots to grow deeper and be less likely submerged in water during wetter seasons. These drain tiles might convey agriculture liquids out to streams and rivers over longer distances than would occur without them. Unfortunately, no spatial databases capture these drain tiles at finer spatial scales than the county level.

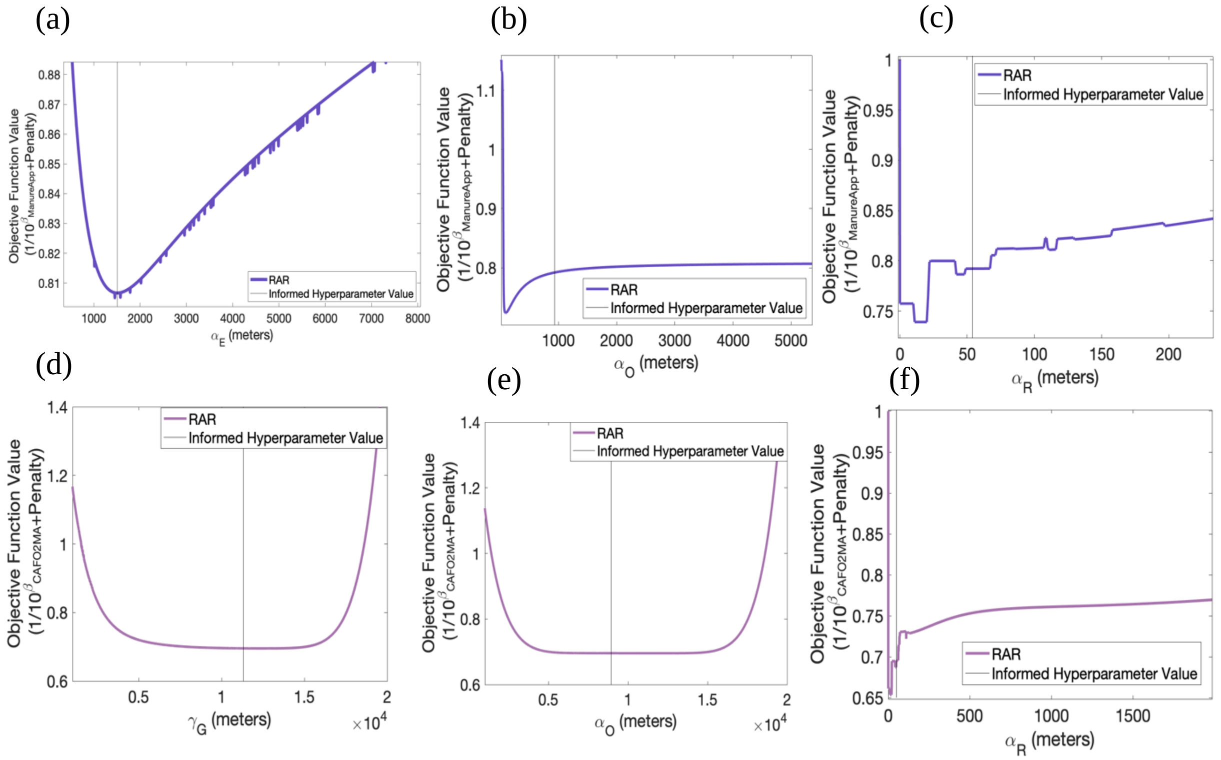


Figure A.8 Maxima obtained from the "Informing the Source Predictors" step of the FIT framework. (a) Euclidean approach for ManureApp (b) Transect across α_R at overland hyperparameter maximum ($\alpha_0 = \alpha_O^*$) for the ORF approach for ManureApp (c) Transect across α_O at river distance hyperparameter maximum ($\alpha_R = \alpha_R^*$) for the ORF approach for ManureApp (d) Transect across γ_G at overland hyperparameter maximum ($\alpha_0 = \alpha_O^*$) and river distance hyperparameter maximum ($\alpha_R = \alpha_R^*$) for the GORF approach for AFO ground transport to ManureApp (e) Transect across α_O at the ground transportation hyperparameter maximum ($\gamma_G = \gamma_G^*$) and river distance hyperparameter maximum ($\alpha_R = \alpha_R^*$) for the GORF approach for AFO ground transport to ManureApp (f) Transect across α_R at the ground transportation hyperparameter maximum ($\gamma_G = \gamma_G^*$) and overland hyperparameter maximum ($\alpha_O = \alpha_O^*$) for the GORF approach for AFO ground transport to ManureApp

Table A.4 Predictor means and variances after using the sum of exponentially decaying contributions models with the selected hyperparameter values for spatial and temporal factors.

Predictor	Mean	Std. Dev.
Euclidean AFO (AFO manure production units/g sediment)	8700	1730
Euclidean Manure Application Fields (manure units/g sediment)	4680	2180
ORF AFO (AFO manure production units/g sediment)	3.84e-19	1.59e-18
ORF Manure Application Fields (manure units/g sediment)	0.729	1.95
GORF AFO manure ground transport to manure application fields (manure units/g sediment)	0.00230	0.00500
More Recent Precipitation (P1) (cm of precipitation over 0.177 days)	7.65e-08	9.55e-08
Less Recent Precipitation (P2) (cm of precipitation over 2.04 days)	0.737	0.459

A1.12 Details on the Predictive Ability of Climatic Variables

We find that all three climatic variables were selected. Freezing temperature exhibited the largest effect on the relative abundance of *BoBac*, likely due to laws restricting liquid manure application in freezing temperatures.[261] The RAR for more recent precipitation, P1, was 9.90 (CI: 4.26, 23.0). However, for a one standard deviation increase in antecedent precipitation, P2, the contributions from precipitation were reduced by a fraction, 0.602 (CI: 0.311, 1.16). These precipitation effects support the diminishing effect of antecedent precipitation on recent precipitation and washout effects.[30,40,87,137,262] Many mechanisms might contribute to the variability explained by precipitation including runoff processes[30,263] and seasonal trends.[261] Greater seasonal variability was observed in September and October, a time when manure is heavily applied following harvest.[134]

A1.13 Details on Cross-Validation Results

The results show that the GORF model always selects the “AFO ground transport to ManureApp” predictor, and that the resulting rMSE is lower than the corresponding ones for the Euclidean and ORF models, which did not use the “AFO ground transport to ManureApp” predictor.

Table A.5 Results of the 6-fold cross-validation of test stage of the FIT framework. The shaded cells show which predictor was selected for each of the fold. The column named rMSE shows the root mean square error for the validation set that was held out.

		Euclidean						ORF						GORF				
Fold	Variable	P1	P1xP2	Freezing	AFO	Manure App	rMSE	P1	P1xP2	Freezing	AFO	Manure App	rMSE	P1	P1xP2	Freezing	AFO ground transport	rMSE
1		🟩	🟩	🟩			0.442	🟩	🟩	🟩			0.442	🟩	🟩	🟩	🟩	0.430
2		🟩		🟩			1.28	🟩		🟩		🟩	1.30	🟩		🟩	🟩	1.29
3		🟩	🟩	🟩		🟩	1.37	🟩	🟩	🟩		🟩	1.66	🟩	🟩	🟩	🟩	1.47
4		🟩		🟩			1.42	🟩		🟩			1.42	🟩		🟩	🟩	1.33
5		🟩		🟩			1.26	🟩		🟩			1.26	🟩		🟩	🟩	1.23
6		🟩	🟩	🟩		🟩	2.41	🟩	🟩	🟩			2.41	🟩	🟩	🟩	🟩	2.25
Mean rMSE		1.36						1.41						1.33				

A1.14 Results of the FIT Analysis for the Absolute Abundance of Sediment Bobac

A.1.14.1 Results from the FIT framework

In the paper we present results of the microbial FIT framework for the relative abundance of *BoBac*. Here, for comparison purposes, we use the same microbial FIT framework to model the absolute abundance, using the same climatic variables and same microbial source models (Euclidean, ORF and GORF). Overall, the results for spatial predictors were similar for absolute abundance (Table A.6) compared to the relative abundance (Table 3.1). For example, the effect for GORF modeled CAFO ground transport to manure application fields was greater than the ORF manure application, and the confidence interval for the GORF approach was narrower than for the ORF. However, the increase in the magnitude and significance of the climatic effects shows that absolute abundance appears to be driven more by climatic factors. This supports the hypothesis that using the absolute abundance, rather than normalizing by 16S, represents variability of *BoBac* as part of a fuller microbial community that is reacting to climatic variability and potentially other unaccounted for factors influencing the microbial community as a whole.

Table A.6 Standardized regression coefficients, βs, absolute abundance ratio, optimized hyperparameters, αs, adjusted R-squared, and cross validation statistics with climatic (shaded in grey) and spatial (not shaded) explanatory variables for the Absolute Abundance of Bovine Bacteroides per gram of sediment.

Model	Standardized Explanatory Variables	Hyperparameters $\boldsymbol{\alpha}$	Regression Coefficient β	Absolute Abundance Ratio 10^{β} (95% CI)
Euclidean Adjusted R^2=0.472	P1	α_{P1} =0.0082 days	1.12(*)	13.2 (6.12, 28.3)
	P1×P2	α_{P2}=2.77 days	-0.483(*)	0.329 (0.178, 0.6061)
	Freezing		-3.77(*)	1.69e-4 (2.62e-5,0.0011)
	AFO	α_E = 50.0 m	NOT SELECTED	NOT SELECTED
	ManureApp	α_E = 864 m	NOT SELECTED	NOT SELECTED
Overland and River Distance - Flow (ORF) Adjusted R^2=0.492	P1	α_{P1} =0.0082 days	1.12(*)	13.1 (6.20, 27.8)
	P1×P2	α_{P2}=2.77 days	-0.475(*)	0.335 (0.184, 0.610)
	Freezing		-3.74(*)	1.82e-4 (1.12e-4,2.96e-4)
	AFO	α_O=51.7 m , α_R= 70.0 m	NOT SELECTED	NOT SELECTED
	ManureApp	α_O=2.40 km , α_R= 69.1 m	0.237(*)	1.72 (1.06, 2.81)
Ground Transport, Overland and River Distance - Flow (GORF) Adjusted R^2=0.496	P1	α_{P1} =0.0082 days	1.11(*)	13.0 (6.16, 27.5)
	P1×P2	α_{P2}=2.77 days	-0.482(*)	0.237 (0.182, 0.560)
	Freezing		-3.71(*)	1.32e-4 (1.20e-4, 3.17e-4)
	AFO ground transport to ManureApp	γ_G=9.74 km, α_O= 7.15 km, α_R= 50.6 m	0.252(*)	1.79 (1.10, 2.91)

*indicates a p-value <0.05

APPENDIX 2: SUPPLEMENTAL INFORMATION FOR AIM 2

A2.1 Details on the Distributions of the Absolute Abundance and 16S-Rrna Values Between and Within Sampling Events

Figure A.9 provides details on the distribution of the absolute abundance of *BoBac* and *HuBac* per gram sediment or per 100 mL of surface water. The least within-sampling event variability is observed for February. The greatest within-sampling event variability for sediment is observed in October, while the greatest within-sampling event variability for surface water may be for the July or September sampling events. There is substantial between-sampling event variability observed for February and July for each of the respective responses.

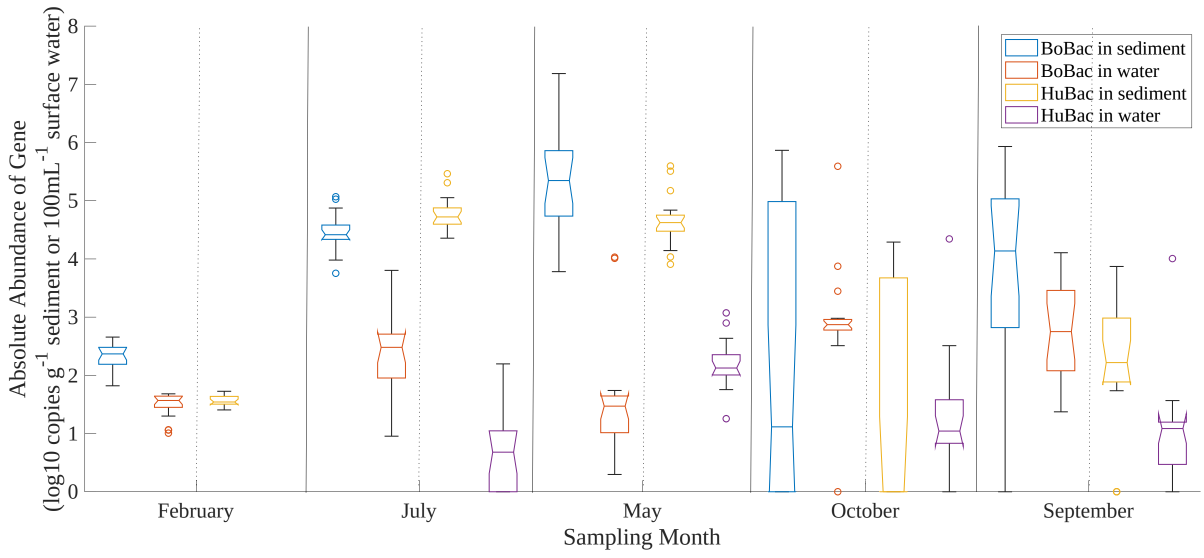


Figure A.9 Distribution of absolute abundance values across sampling time for each response, *BoBac* and *HuBac*, and for each sample type (i.e., sediment and surface water). The boxes show the upper and lower quartiles for each response and sample type. The middle line of the box represents the median. The tapered shaded area around the middle line shows the 95% confidence interval around the median. Sampling times are 13-Jul-2016, 14-Sep-2016, 25-Oct-2016, 24-Feb-2017, and 22-May-2017.

Figure A.10 depicts the between and within event variability for 16S-rRNA copies per gram sediment or per 100 mL. The lowest concentrations in sediment and highest concentrations in surface water occur during the October sampling event. The lowest of the median concentrations between sampling events for surface water occurs for the May sampling event. The highest of the median concentrations between sampling events for sediment occurs for the July sampling event.

Figure A.11 summarizes precipitation and temperature changes occurring throughout the sampling period. Precipitation appears to follow the temperature trend, which has implications for collinearity in the model. This is one reason that freezing temperature was preferred in the modeling approach to Temperature in degrees Celsius.

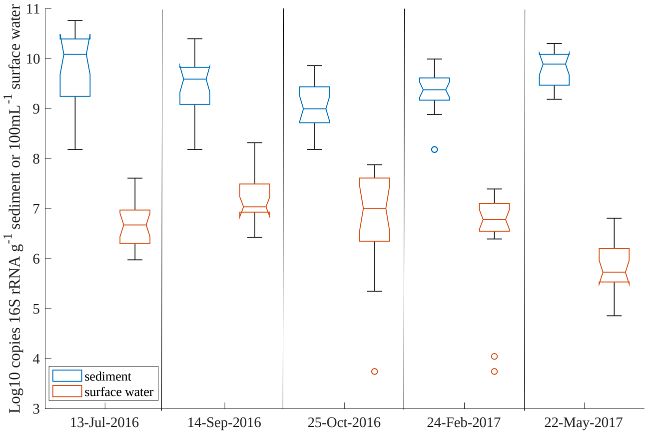


Figure A.10 Distribution of the abundance of 16S rRNA copies in sediment and surface water across different sampling times

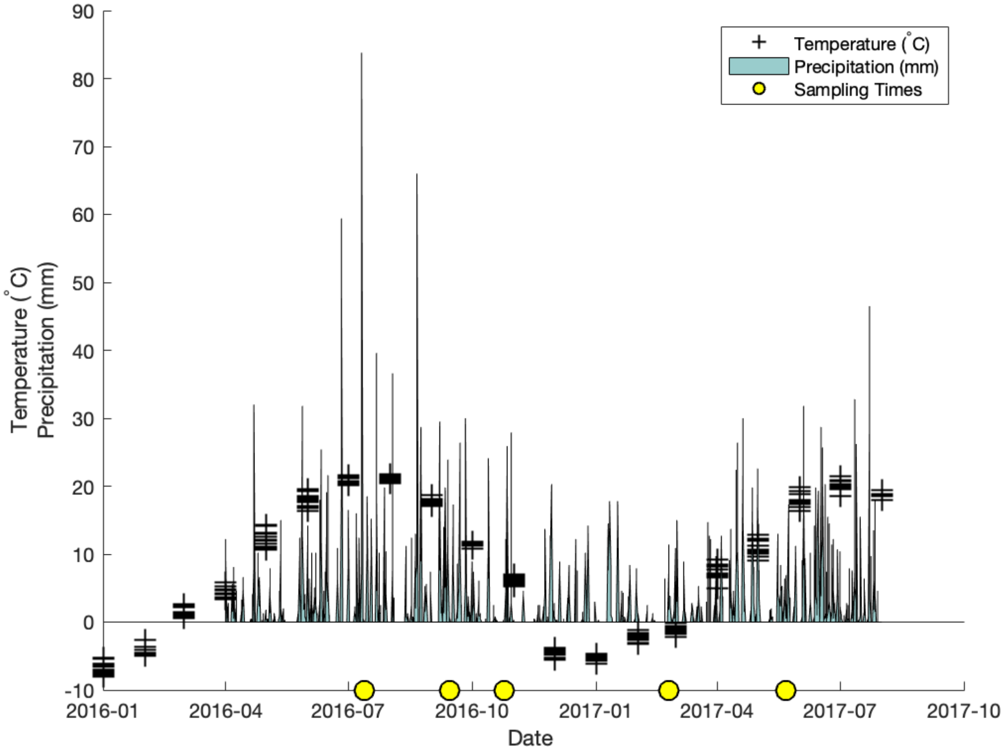


Figure A.11 Monthly average temperature and daily precipitation shown for dates between January 1st, 2016 and August 1st, 2017. The dates of the 5 sampling events are shown in yellow.

A2.2 Details on the Spatial Distributions of the Relative Abundance of Host-Associated Responses Averaged Across Sampling Events

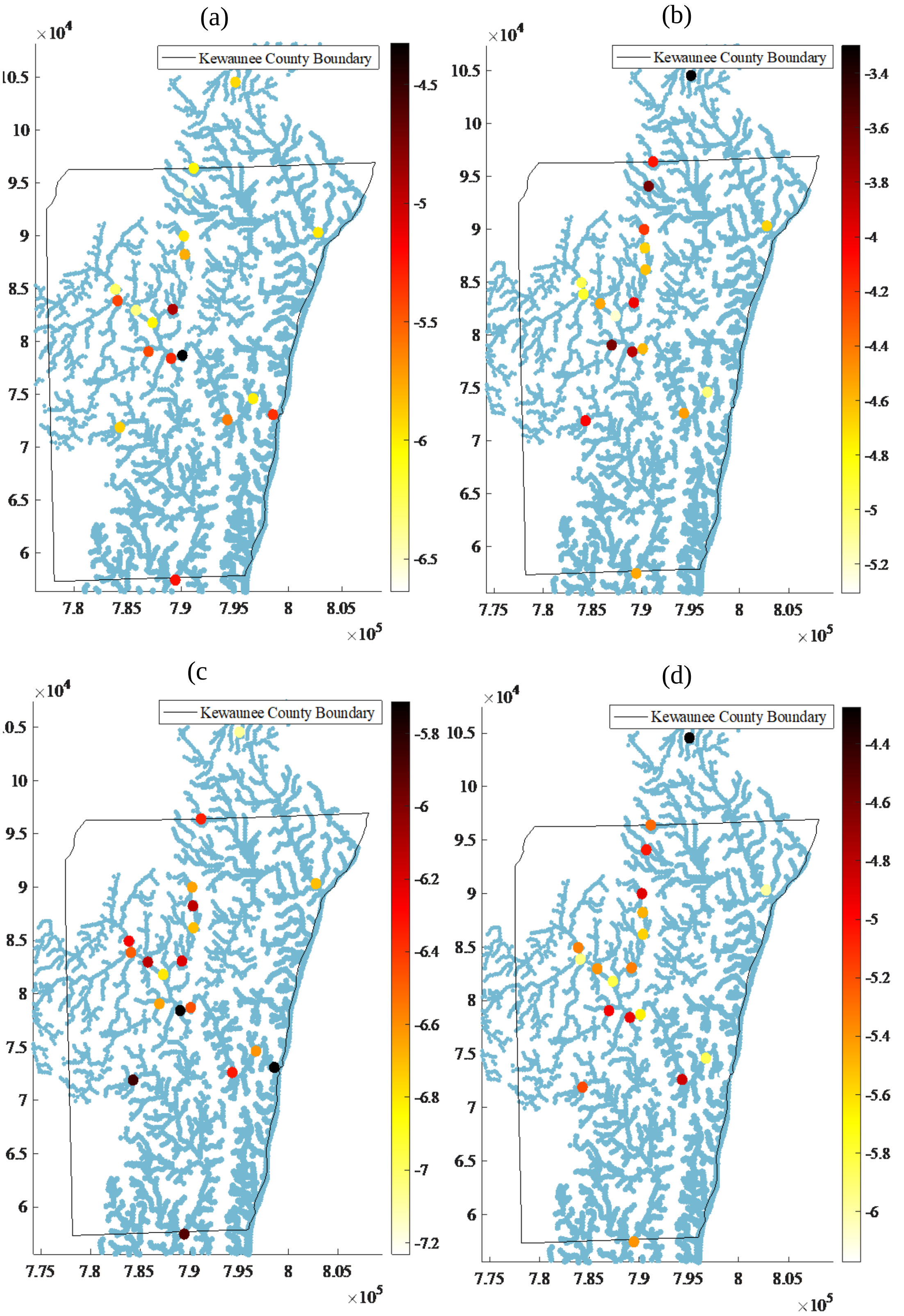


Figure A.12 Spatial average of the relative abundance (log10 gene-copies-per-16S-rRNA) of each host-associated marker measured in surface water and sediment a) *BoBac* in sediment b) *BoBac* in surface water c) *HuBac* in sediment d) *HuBac* in surface water.

A2.3 Details on the Types of Sources and Candidate Databases of Spatially Distributed Sources

Table A.7 Source types, candidate spatial databases, database descriptions, and description of pertinent host-associated pollution associated with the source type and rationale

Source Type, *u*	Potential Contamin. Source (Abbrev.)	Candidate Spatial Database	Database description	Host-associated Pollution Concern
1	**Animal Feeding Operations** **(AFO)**	Option 1: County Database of Manure Storages (weighted)	Database providing centroids of each manure storage in Kewaunee County weighted by the storage capacity of the manure storages in log total gallons of manure as determined by the estimated dimensions of the manure storages. When this information was unavailable, an average of the available data on log total gallons was applied. The database was made in 2014. Data was provided through personal communication with the Kewaunee County Land and Water Conservation Department	**Bovine:** dairy cattle manure
		Option 2: County Database of Manure Storages (unweighted)	Same provider as above but the centroids are equally weighted. In other words, they were not weighted by storage capacity in log total gallons.	**Bovine:** dairy cattle manure
		Option 3: WPDES Database of CAFOs (unweighted)	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer	**Bovine:** dairy cattle CAFOs
		Option 4: WPDES Database of CAFOs (weighted by animal units)	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer	**Bovine:** dairy cattle CAFOs
2	**Manure Application Fields**	Option 1: Land Cover Database of Crop Rotation	Database of crop rotation land cover polygon centroids weighted by the area of the land cover polygons. "The	**Bovine:** cattle manure applied on land for

	(ManureApp)		30- meter raster layer combines land cover information from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) with land management information to map five-year crop rotations over Common Land Units (CLUs)… The crop rotation dataset was validated independently from Wiscland 2 using data from Department of Agriculture, Trade and Consumer Protection (DATCP) cattle inventory records and dairy producer data, NASS crop acreage estimates and transect surveys. The CDL datasets on which the crop rotation layer is primarily based also provide accuracy information by state and by class. Reported accuracies for the Principal Crops (e.g. Corn, Soybeans, Alfalfa, Vegetables, etc.) range from 87-93% over the 2008-2012 period; the accuracy of the crop rotations will differ due to the spatiotemporal aggregation of classes and the integration of the crop classification with CLU boundaries” [135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	nutrient management or as fertilizer
		Option 2: Land Cover Database of Dairy Rotation	Database is of dairy rotation land cover polygon centroids weighted by the area of the land cover polygons. Dairy rotation land cover is a subclass of crop rotation land cover from the Wiscland-2 land cover database mentioned above [135]. Dairy rotation land cover is described by “corn grain, corn silage, and alfalfa in a 6-year rotation, with typically 2 years of corn, 4 years of alfalfa, and a cover crop between corn and alfalfa years. Occasionally, soybeans or a grain (e.g., wheat, barley, or oats) are planted in place of corn” [135]. Dairy rotation areas are typically associated with use by dairy farms and operations for manure application and feed.	**Bovine:** dairy cattle manure applied on land for nutrient management, as fertilizer, and/or for feed production
3	**Septic Systems (Sep)**	Only option: County Database of Drainfield Locations (unweighted)	Database providing the centroids of drainfields in Kewaunee County that are unweighted. When this information was unavailable, an average of the available data was applied. The database was made in 2014. Data was provided through personal communication with the Kewaunee County Land and Water Conservation Department.	**Human:** septic systems primarily located on parcels categorized as “residential”.

4	**Wastewater Treatment Plants** **(WWTP)**	Option 1: State Database of all WPDES wastewater treatment facilities (unweighted)	Database providing the locations of wastewater treatment facilities under the Wisconsin Pollutant Discharge Elimination Program (WPDES) provided by the Wisconsin Department of Natural Resources (WIDNR) through an open records request	**Human:** wastewater facilities are located near to town/city centers
		Option 2: State Database of WPDES Wastewater Treatment Facilities under the municipal category, weighted by population.	Database providing the locations of wastewater treatment facilities under the Wisconsin Pollutant Discharge Elimination Program (WPDES) that have *municipal* waste types provided by the Wisconsin Department of Natural Resources (WIDNR) through an open records request	**Human:** wastewater facilities are located near to town/city centers
		Option 3: State Database of WPDES Wastewater Treatment Facilities under the industrial category (unweighted)	Database providing the locations of wastewater treatment facilities under the Wisconsin Pollutant Discharge Elimination Program (WPDES) that have *municipal* waste types provided by the Wisconsin Department of Natural Resources (WIDNR) through an open records request	**Bovine:** Second to livestock agriculture, the primary industry in this region is dairy production
5	**All Land Applied Sludge Sites** **(LAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage	A state database of land applied sludge sites under the Wisconsin Pollutant Discharge Elimination Program (WPDES) obtained through an open records request from WIDNR. The database includes information regarding the types of waste produced and applied by the permittee. The information is used to more precisely describe land applied sludge sites by defining new source types. For the source type, *All Land Applied Sludge Sites*, all sites were considered. Acreage information was available in this dataset. When the acreage information was missing or listed as 0, the value was set to be the average acreage. Here, the database is weighted by the log-transformed acreage associated with the site.	**Human or Bovine:** Land applied sludge that is produced and applied by permittees come from municipal wastewater or septage (see Sep and WWTP descriptions) or from industrial sources. After livestock production, the second largest industry in Kewaunee County is dairy product production.
		Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)	Same as above, but the sites are unweighted.	**Human or Bovine:** Land applied sludge that is produced and applied by permittees come from municipal wastewater or septage (see Sep and WWTP descriptions) or from

				industrial sources. After livestock production, the second largest industry in Kewaunee County is dairy product production.
6	**Land Applied Sludge with Non-Industrial Waste** **(DomLAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites with wastewater treatment or septage	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified either as “wastewater treatment” or “septage”.	**Human:** See LAS source type
		Option 2: State Database of WPDES Land Applied Sludge Sites with septage	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified as “septage”.	**Human:** See LAS source type
		Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified as “wastewater treatment”.	**Human:** See LAS source type
7	**Land Applied Sludge with Industrial Waste** **(IndLAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites with Industrial Sludge	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified either as “industrial wastewater” or “food processing”.	**Bovine:** See LAS source type
		Option 2: State Database of WPDES Land Applied Sludge Sites with Industrial Wastewater	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified as “industrial wastewater”.	**Bovine:** See LAS source type
		Option 3: State Database of WPDES Land Applied Sludge Sites from Food Processing	Same as LAS: option 1, but the type of waste produced and applied by the permittee is classified as “food processing”.	**Bovine:** See LAS source type
8	**High Intensity Developed Land Cover** **(HID)**	Option 1: Land cover database of high-intensity developed land cover as polygon	Database of high-intensity developed land cover polygon centroids weighted by the area of the land cover polygons. High-intensity developed land cover is described as “areas with 50% or greater solid impervious cover of man-made materials. ” [135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Human:** This type of land cover is primarily collocated with downtown areas and suburban residences

		Option 2: Land cover database of high-intensity developed land cover as gridded points	Database of high-intensity developed land cover with source nodes that are 100 m x 100 m grid points representing locations with that land cover type on a 100 m x 100 m and equally weighted. High-intensity developed land cover is described as “areas with 50% or greater solid impervious cover of man-made materials. ” [135] This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Human:** This type of land cover is primarily collocated with downtown areas and suburban residences
9	**Low Intensity Developed Land Cover** **(LID)**	Option 1: Land cover database of low-intensity developed land cover as polygon	Database of low-intensity developed land cover polygon centroids weighted by the area of the land cover polygons. Low-intensity developed land cover is described as “areas with 25% or greater solid impervious cover of man-made materials, but less than 50%. May have some interspersed vegetation. ” [135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Bovine:** This type of land cover is primarily collocated with infrastructure in more rural areas such as roads, small farms, parking lots, and residential areas at the edge of semi-urban and rural areas
		Option 2: Land cover database of low-intensity developed land cover as gridded points	Database of low-intensity developed land cover with source nodes that are 100 m x 100 m grid points representing locations with that land cover type and equally weighted. Low-intensity developed land cover is described as “areas with 25% or greater solid impervious cover of man-made materials, but less than 50%. May have some interspersed vegetation. ” [135] This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Bovine:** This type of land cover is primarily collocated with infrastructure in more rural areas such as roads, small farms, parking lots, and residential areas at the edge of semi-urban and rural areas
Source Type, *u*	Potential Contamin. Source (Abbrev.)	Candidate Spatial Database	Database description	Host-associated Pollution Concern
1	**Animal Feeding Operations** **(AFO)**	Option 1: County Database of Manure Storages (weighted)	Database providing centroids of each manure storage in Kewaunee County weighted by the storage capacity of the manure storages in log total gallons of manure as determined by the estimated dimensions of the manure storages. When this information was unavailable, an average of the available data on log total gallons was applied. The database was made in 2014. Data was provided through personal communication with the	**Bovine:** dairy cattle manure

			Kewaunee County Land and Water Conservation Department	
		Option 2: County Database of Manure Storages (unweighted)	Same provider as above but the centroids are equally weighted. In other words, they were not weighted by storage capacity in log total gallons.	**Bovine:** dairy cattle manure
		Option 3: WPDES Database of CAFOs (unweighted)	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer	**Bovine:** dairy cattle CAFOs
		Option 4: WPDES Database of CAFOs (weighted by animal units)	Database providing locations of Wisconsin WPDES Permitted Concentrated Feeding Operations (CAFOs), which are AFOs with >999 animals, in effect in Wisconsin as of January 2017 using geocoded addresses from WPDES permit data. https://services.arcgis.com/SXbDpmb7xQkk44JV/arcgis/rest/services/CAFO_WIDNR_2017/FeatureServer	**Bovine:** dairy cattle CAFOs
2	**Manure Application Fields** **(ManureApp)**	Option 1: Land Cover Database of Crop Rotation	Database of crop rotation land cover polygon centroids weighted by the area of the land cover polygons. "The 30- meter raster layer combines land cover information from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) with land management information to map five-year crop rotations over Common Land Units (CLUs)… The crop rotation dataset was validated independently from Wiscland 2 using data from Department of Agriculture, Trade and Consumer Protection (DATCP) cattle inventory records and dairy producer data, NASS crop acreage estimates and transect surveys. The CDL datasets on which the crop rotation layer is primarily based also provide accuracy information by state and by class. Reported accuracies for the Principal Crops (e.g. Corn, Soybeans, Alfalfa, Vegetables, etc.) range from 87-93% over the 2008-2012 period; the accuracy of the crop rotations will differ due to the spatiotemporal aggregation	**Bovine:** cattle manure applied on land for nutrient management or as fertilizer

			of classes and the integration of the crop classification with CLU boundaries” [135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	
		Option 2: Land Cover Database of Dairy Rotation	Database is of dairy rotation land cover polygon centroids weighted by the area of the land cover polygons. Dairy rotation land cover is a subclass of crop rotation land cover from the Wiscland-2 land cover database mentioned above [135]. Dairy rotation land cover is described by “corn grain, corn silage, and alfalfa in a 6-year rotation, with typically 2 years of corn, 4 years of alfalfa, and a cover crop between corn and alfalfa years. Occasionally, soybeans or a grain (e.g., wheat, barley, or oats) are planted in place of corn” [135]. Dairy rotation areas are typically associated with use by dairy farms and operations for manure application and feed.	**Bovine:** dairy cattle manure applied on land for nutrient management, as fertilizer, and/or for feed production
3	**Septic Systems (Sep)**	Only option: County Database of Drainfield Locations (unweighted)	Database providing the centroids of drainfields in Kewaunee County that are unweighted. When this information was unavailable, an average of the available data was applied. The database was made in 2014. Data was provided through personal communication with the Kewaunee County Land and Water Conservation Department.	**Human:** septic systems primarily located on parcels categorized as “residential”.
4	**Wastewater Treatment Plants** **(WWTP)**	Option 1: State Database of all WPDES wastewater treatment facilities	Database providing the locations of wastewater treatment facilities under the Wisconsin Pollutant Discharge Elimination Program (WPDES) provided by the Wisconsin Department of Natural Resources (WIDNR) through an open records request	**Human:** wastewater facilities are located near to town/city centers
		Option 2: State Database of WPDES Wastewater Treatment Facilities under the municipal category, weighted by population.	Database providing the locations of wastewater treatment facilities under the Wisconsin Pollutant Discharge Elimination Program (WPDES) that have *municipal* waste types provided by the Wisconsin Department of Natural Resources (WIDNR) through an open records request	**Human:** wastewater facilities are located near to town/city centers
5	**All Land Applied Sludge Sites** **(LAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage	A state database of land applied sludge sites under the Wisconsin Pollutant Discharge Elimination Program (WPDES) obtained through an open records request from WIDNR. The database includes information regarding the types of waste produced and applied by the permittee. The information is used to more precisely describe land	**Human or Bovine:** Land applied sludge that is produced and applied by permittees come from municipal wastewater or septage

			applied sludge sites by defining new source types. For the source type, *All Land Applied Sludge Sites*, all sites were considered. Acreage information was available in this dataset. When the acreage information was missing or listed as 0, the value was set to be the average acreage. Here, the database is weighted by the log-transformed acreage associated with the site.	(see Sep and WWTP descriptions) or from industrial sources. After livestock production, the second largest industry in Kewaunee County is dairy product production.
		Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)	Same as above, but the sites are unweighted.	**Human or Bovine:** Land applied sludge that is produced and applied by permittees come from municipal wastewater or septage (see Sep and WWTP descriptions) or from industrial sources. After livestock production, the second largest industry in Kewaunee County is dairy product production.
6	**Land Applied Sludge with Non-Industrial Waste** **(DomLAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites with municipal or septage	See LAS: option 1, "**Municipal waste** or 'Sewage sludge' or 'sludge' or "biosolids" means the solid, semi‒solid or liquid residue generated during the treatment of domestic sewage in a treatment works. Sewage sludge includes scum or solids removed in primary, secondary or advanced wastewater treatment processes and material derived from sewage sludge. Sewage sludge does not include ash generated during the firing of a sewage sludge incinerator or grit and screenings generated during preliminary treatment of domestic sewage in a treatment works. Note: All 3 terms defined here are interchangeable and recognized by the department, as they are all in common use… '**Septage**' means the wastewater or contents of septic or holding tanks, dosing chambers, grease interceptors, seepage	**Human:** See LAS source type

			beds, seepage pits, seepage trenches, privies or portable restrooms."[181]	
		Option 2: State Database of WPDES Land Applied Sludge Sites with septage	DomLAS: option 1, but the type of waste produced and applied by the permittee is classified as "septage".	**Human:** See LAS source type
		Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment	DomLAS: option 1, but the type of waste produced and applied by the permittee is classified as "municipal waste".	**Human:** See LAS source type
7	**Land Applied Sludge with Industrial Waste** **(IndLAS)**	Option 1: State Database of WPDES Land Applied Sludge Sites with Industrial Sludge or Wastewater	See LAS: option 1, Industrial waste or sludge includes "'by−product solids' which means waste materials from the animal product or **food processing** industry including, but not limited to: remains of butchered animals, paunch manure and vegetable waste materials such as leaves, cuttings, peelings and actively fermenting sweet corn silage. **'Liquid waste'** means process wastewater and waste liquid products, including silage leachate, whey, whey permeate, whey filtrate, contact cooling water, cooling or boiler water containing water treatment additives, and wash water generated in industrial, commercial and agricultural operations which result in a point source discharge to a land treatment system. 'Sludge' means the accumulated solids generated during the biological, physical or chemical treatment, coagulation or sedimentation of water or wastewater."[181]	**Bovine:** See LAS source type
		Option 2: State Database of WPDES Land Applied Sludge Sites with Industrial Wastewater	IndLAS: option 1, but the type of waste produced and applied by the permittee is classified as "liquid waste".	**Bovine:** See LAS source type
		Option 3: State Database of WPDES Land Applied Sludge Sites from Food Processing	Ind LAS: option 1, but the type of waste produced and applied by the permittee is classified as being from "food processing".	**Bovine:** See LAS source type
8	**High Intensity Developed Land Cover** **(HID)**	Option 1: Land cover database of high-intensity developed land cover as polygon	Database of high-intensity developed land cover polygon centroids weighted by the area of the land cover polygons. High-intensity developed land cover is described as "areas with 50% or greater solid impervious cover of man-made materials. " [135]. This data was made	**Human:** This type of land cover is primarily collocated with downtown areas

			available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	and suburban residences
		Option 2: Land cover database of high-intensity developed land cover as gridded points	Database of high-intensity developed land cover with source nodes that are 100 m x 100 m grid points representing locations with that land cover type on a 100 m x 100 m and equally weighted. High-intensity developed land cover is described as “areas with 50% or greater solid impervious cover of man-made materials. ”[135] This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Human:** This type of land cover is primarily collocated with downtown areas and suburban residences
9	**Low Intensity Developed Land Cover** **(LID)**	Option 1: Land cover database of low-intensity developed land cover as polygon	Database of low-intensity developed land cover polygon centroids weighted by the area of the land cover polygons. Low-intensity developed land cover is described as “areas with 25% or greater solid impervious cover of man-made materials, but less than 50%. May have some interspersed vegetation. ”[135]. This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Bovine:** This type of land cover is primarily collocated with infrastructure in more rural areas such as roads, small farms, parking lots, and residential areas at the edge of semi-urban and rural areas
		Option 2: Land cover database of low-intensity developed land cover as gridded points	Database of low-intensity developed land cover with source nodes that are 100 m x 100 m grid points representing locations with that land cover type and equally weighted. Low-intensity developed land cover is described as “areas with 25% or greater solid impervious cover of man-made materials, but less than 50%. May have some interspersed vegetation. ”[135] This data was made available from the Wiscland-2 Land Cover Database released in 2016. https://p.widencdn.net/lkfpeb/wiscland2_landcover	**Bovine:** This type of land cover is primarily collocated with infrastructure in more rural areas such as roads, small farms, parking lots, and residential areas at the edge of semi-urban and rural areas

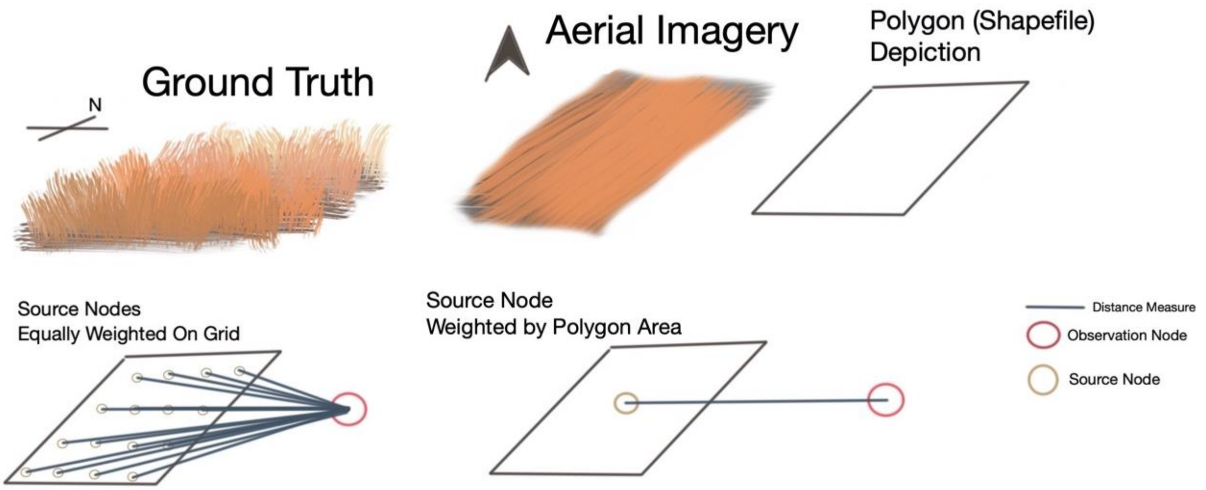


Figure A.13 Visual description of difference between source type LID and HID and candidate database options 1 and 2. The source nodes that are equally weighted on the grid (bottom left) represent an example of how LID or HID option 2 are obtained. The source node that is weighted by the polygon area (bottom center) represent an example of how LID or HID option 1 are obtained. The polygon shapefiles for land cover in our study were derived in ArcGIS 10.5 from the Wiscland-2 Land Cover Database Level 4 Raster File.

A2.4 Details of the Reliability Score Results

Table A.8 Reliability score for the candidate database options associated with each potential contamination source for each of the responses in river sediment or surface water. Also shown is the sign stability scores (*SSS*) and the weighted standard deviation on the coefficient (σ_{β_u}) which are used in the calculation of the reliability score. The database with the highest reliability score for each potential source is selected and shown in bold face.

Relative Abundance of *BoBac* in Sediment

Variable	Description of Option	Reliability Score	Sign Stability Score *SSS*	Magnitude *M*	Std. Dev. of Coefficient σ_{β_u}
AFO	Option 1: County Database of Manure Storages (weighted)	1.59	2	0.0933	0.235
AFO	Option 2: County Database of Manure Storages (unweighted)	1.71	4	0.0993	0.232
AFO	Option 3: WPDES Database of CAFOs (unweighted)	7.73	3	0.764	0.395
AFO	**Option 4: WPDES Database of CAFOs (weighted by animal units)**	**12.1**	**3**	**1.14**	**0.375**
IndLAS	Option 1: State Database of WPDES Land Applied Sludge Sites with Industrial Sludge	-1.18	4	-0.129	0.438
IndLAS	**Option 3: State Database of WPDES Land Applied Sludge Sites from Food Processing**	**1.72**	**3**	**0.233**	**0.544**
IndLAS	Option 2: State Database of WPDES Land Applied Sludge Sites with Industrial Wastewater	-1.17	4	-0.128	0.435
LAS	**Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)**	**-0.331**	**4**	**-0.0314**	**0.379**
LAS	Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage	-0.297	3	-0.0288	0.388
LID	Option 1: Land cover database of low-intensity developed land cover as polygon	7.37	2	0.380	0.206
LID	**Option 2: Land cover database of low-intensity developed land cover as gridded points**	**8.75**	**1**	**0.590**	**0.270**

ManureApp	**Option 1: Land Cover Database of Crop Rotation**		**-0.501**	**2**	**-0.0662**	**0.397**
ManureApp	Option 2: Land Cover Database of Dairy Rotation		-2.12	1	-0.200	0.378
WWTP	**Option 3: State Database of WPDES Wastewater Treatment Facilities under the industrial category (unweighted)**		**3.89**	**2**	**0.476**	**0.488**
AFO2ManureApp	AFO Option 1 ground transport to	ManureApp Option 1	4.70	3	1.57	0.587
AFO2ManureApp	AFO Option 1 ground transport to	ManureApp Option 2	1.38	3	0.460	0.617
AFO2ManureApp	AFO Option 2 ground transport to	ManureApp Option 1	4.56	3	1.52	0.587
AFO2ManureApp	AFO Option 2 ground transport to	ManureApp Option 2	0.75	3	0.249	0.631
AFO2ManureApp	AFO Option 3 ground transport to	ManureApp Option 1	2.53	3	0.845	0.595
AFO2ManureApp	**AFO Option 3 ground transport to**	**ManureApp Option 2**	**11.6**	**4**	**2.89**	**0.413**
AFO2ManureApp	AFO Option 4 ground transport to	ManureApp Option 1	2.06	3	0.688	0.576
AFO2ManureApp	AFO Option 4 ground transport to	ManureApp Option 2	-0.51	3	-0.170	0.651

Relative Abundance of *BoBac* in Surface Water

Variable	Description of Option	Reliability Score	Sign Stability Score	Magnitude	Std. Dev. of Coefficient
AFO	**Option 1: County Database of Manure Storages (weighted)**	**10.8**	**4**	**0.621**	**0.235**
AFO	Option 2: County Database of Manure Storages (unweighted)	10.2	4	0.593	0.232
AFO	Option 3: WPDES Database of CAFOs (unweighted)	-5.32	0	-1.08	0.395
AFO	Option 4: WPDES Database of CAFOs (weighted by animal units)	-6.11	0	-1.45	0.375
IndLAS	Option 1: State Database of WPDES Land Applied Sludge Sites with Industrial Sludge	2.65	3	0.231	0.438
IndLAS	**Option 3: State Database of WPDES Land Applied Sludge Sites from Food Processing**	**38.1**	**4**	**0.658**	**0.544**
IndLAS	Option 2: State Database of WPDES Land Applied Sludge Sites with Industrial Wastewater	2.66	3	0.232	0.435
LAS	Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)	-7.83	3	-0.589	0.379
LAS	**Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage**	**-6.81**	**2**	**-0.674**	**0.388**
LID	Option 1: Land cover database of low-intensity developed land cover as polygon	-7.31	2	-0.613	0.206
LID	**Option 2: Land cover database of low-intensity developed land cover as gridded points**	**3.29**	**2**	**0.127**	**0.270**
ManureApp	**Option 1: Land Cover Database of Crop Rotation**	**44.0**	**5**	**1.04**	**0.397**
ManureApp	Option 2: Land Cover Database of Dairy Rotation	38.6	5	1.29	0.378

WWTP	Option 3: State Database of WPDES Wastewater Treatment Facilities under the industrial category (unweighted)		**-8.21**	**1**	**-0.775**	**0.488**
AFO2ManureApp	AFO Option 1 ground transport to	ManureApp Option 1	6.53	3	0.422	0.258
AFO2ManureApp	AFO Option 1 ground transport to	ManureApp Option 2	6.91	3	0.451	0.261
AFO2ManureApp	AFO Option 2 ground transport to	ManureApp Option 1	6.50	3	0.420	0.258
AFO2ManureApp	AFO Option 2 ground transport to	ManureApp Option 2	6.92	3	0.453	0.262
AFO2ManureApp	**AFO Option 3 ground transport to**	**ManureApp Option 1**	**13.9**	**3**	**0.568**	**0.163**
AFO2ManureApp	AFO Option 3 ground transport to	ManureApp Option 2	12.5	4	0.698	0.224
AFO2ManureApp	AFO Option 4 ground transport to	ManureApp Option 1	1.28	3	0.0941	0.221
AFO2ManureApp	AFO Option 4 ground transport to	ManureApp Option 2	-1.97	3	-0.264	0.268

Relative Abundance of *HuBac* in Sediment

Variable	Description of Option	Reliability Score	Sign Stability Score	Magnitude	Std. Dev. of Coefficient
DomLAS	**Option 2: State Database of WPDES Land Applied Sludge Sites with septage**	**17.7**	**2**	**1.408**	**0.239**
DomLAS	Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment	-8.77	2	-1.440	0.492
DomLAS	Option 1: State Database of WPDES Land Applied Sludge Sites with wastewater treatment or septage	5.70	2	0.783	0.412
HID	Option 1: Land cover database of high-intensity developed land cover as polygon	-4.51	3	-0.500	0.443
HID	**Option 2: Land cover database of high-intensity developed land cover as gridded points**	**-3.17**	**1**	**-1.113**	**0.703**
LAS	Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)	1.89	4	0.157	0.416
LAS	**Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage**	**2.07**	**4**	**0.180**	**0.433**
Sep	*Only option: County Database of Drainfield Locations (unweighted)*	*-5.84*	*4*	*-0.459*	*0.393*
WWTP	**Option 2: State Database of WPDES Wastewater Treatment Facilities under the municipal category, weighted by population.**	**-2.17**	**2**	**-0.446**	**0.616**
WWTP	Option 1: State Database of all WPDES wastewater	-9.60	3	-0.829	0.345

	treatment facilities (unweighted)					
Sep2DomLAS	Only option: County Database of Drainfield Locations (unweighted)	Option 2: State Database of WPDES Land Applied Sludge Sites with septage	-7.37	2	-0.756	0.308
Sep2DomLAS	**Only option: County Database of Drainfield Locations (unweighted)**	**Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment**	**-5.68**	**2**	**-1.034**	**0.546**
Sep2DomLAS	Only option: County Database of Drainfield Locations (unweighted)	Option 1: State Database of WPDES Land Applied Sludge Sites with wastewater treatment or septage	-5.89	2	-1.028	0.523

Relative Abundance of *HuBac* in Surface Water

Variable	Description of Option	Reliability Score	Sign Stability Score	Magnitude	Std. Dev. of Coefficient
DomLAS	**Option 2: State Database of WPDES Land Applied Sludge Sites with septage**	**-4.70**	**2**	**-0.344**	**0.220**
DomLAS	Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment	-6.88	0	-1.487	0.216
DomLAS	Option 1: State Database of WPDES Land Applied Sludge Sites with wastewater treatment or septage	-6.81	2	-0.283	0.125
HID	Option 1: Land cover database of high-intensity developed land cover as polygon	-6.84	1	-0.639	0.187
HID	**Option 2: Land cover database of high-intensity developed land cover as gridded points**	**-4.10**	**1**	**-0.507**	**0.247**
LAS	**Option 2: State Database of WPDES Land Applied Sludge sites (unweighted)**	**-3.48**	**1**	**-0.605**	**0.348**
LAS	Option 1: State Database of WPDES Land Applied Sludge Sites weighted by acreage	-5.48	1	-1.115	0.407
Sep	*Only option: County Database of Drainfield Locations (unweighted)*	*-0.99*	*3*	*-0.107*	*0.433*
WWTP	Option 2: State Database of WPDES Wastewater Treatment Facilities under the municipal category, weighted by population.	-10.41	2	-0.367	0.106
WWTP	**Option 1: State Database of all WPDES wastewater treatment facilities (unweighted)**	**-6.62**	**2**	**-0.600**	**0.272**

Sep2DomLAS	Only option: County Database of Drainfield Locations (unweighted)	Option 2: State Database of WPDES Land Applied Sludge Sites with septage	-7.41	3	-0.738	0.398
Sep2DomLAS	Only option: County Database of Drainfield Locations (unweighted)	Option 3: State Database of WPDES Land Applied Sludge Sites with wastewater treatment	-7.87	2	-0.562	0.214
Sep2DomLAS	**Only option: County Database of Drainfield Locations (unweighted)**	**Option 1: State Database of WPDES Land Applied Sludge Sites with wastewater treatment or septage**	**3.99**	**2**	**0.331**	**0.249**

A2.5 Details on the Hyperparameter Results for Source Terms Selected by the FIT Model

Figure A.14 depicts the RAR in the hyperparameter space. These depictions were used to verify that the hyperparameter values selected automatically in by the MATLAB patternsearch function using an objective function equal to the RAR and a penalty, adequately captured the hyperparameters that correspond with the maximum RAR. For Figure A.14(a), the maximum RAR appears to occur between 5 km $< \alpha_O < 8\ km$ when $\alpha_R = 68$ capturing an average overland influence range and downstream influence range from manure hauled from AFOs and applied on fields (b) capturing an average overland influence range from low-intensity developed land cover that is greater than 10 km, and a downstream influence range between 1.5 and 2 km. (c) capturing an average overland influence range from AFOs that is greater than 3 km, but less than 4 km, and a downstream influence range between 1.5 and 2 km. (d) capturing an average overland influence range from low-intensity land cover that is greater than 100 m, but less than 200 m, and a downstream influence range between 1 and 2 km. (e) capturing an average overland influence range from septic systems that is greater than 3 km, but less than 9 km, and a downstream influence range between 2.5 and 3 km (f) capturing an average overland influence range from high-intensity land cover that is greater than 5 km, and a downstream influence range between 4.5 and 4.5 km and (g) capturing an average overland influence range from septic systems that is greater than 50 m, but less than 200 m, and a downstream influence range between 50 and 200 m.

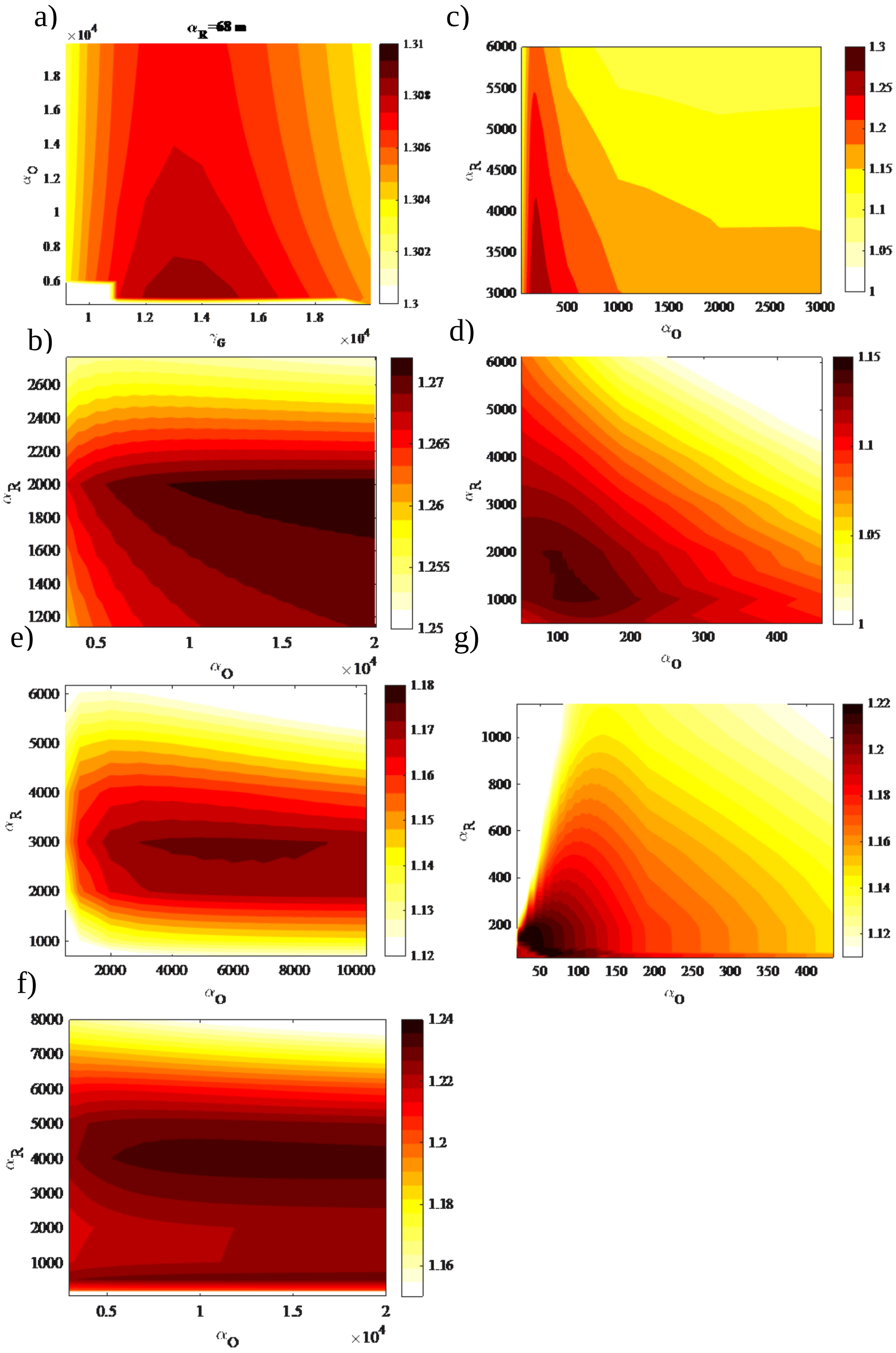


Figure A.14 Relative Abundance Ratios (RARs) across different hyperparameter values (meters) for source terms selected in the Test stage of the FIT framework for a) the GORF spatial predictor model, AFO ground transport to ManureApp, where α_R=68.0 m and b) the ORF spatial predictor model for low-intensity developed land cover to explain the relative abundance of bovine fecal markers in sediment (log10 *BoBac*-copies-per-16S-rRNA-copies). Then c) The ORF spatial predictor model for AFOs and d) the ORF spatial predictor model for low-intensity developed land cover to explain the relative abundance of bovine fecal markers in surface water (log10 *BoBac*-copies-per-16S-rRNA-copies). Then e) the ORF spatial predictor model for septic systems and f) the ORF spatial predictor model for high-intensity developed land-cover to explain human fecal markers in sediment (log10 *HuBac*-copies-per-16S-rRNA-copies). Lastly, f) the ORF spatial predictor model for septic systems to explain human fecal markers in surface water (log10 *HuBac*-copies-per-16S-rRNA-copies).

A2.6 Details on Cross-Validation Results

Table A.9 shows the results of the 6-fold cross validation on the stepwise selection procedure was performed. Overall, this demonstrated the selection was stable for 3 out of the 4 responses. The selection varied for climatic and source terms associated with the relative abundance of human-associated fecal markers in sediment. However, septic systems and high-intensity developed land cover source terms were still selected more frequently compared to other terms. These results do suggest that the selection may be primarily determined by a few influential measurements.

Table A.9 Results of the 6-fold cross-validation of <u>test</u> stage of the FIT framework for *bovine*-associated responses. The shaded cells show which predictor was selected for each of the fold. The column named rMSE shows the root mean square error for the validation set that was held out.

		Sediment											Surface Water										
Fold	Variables	P1	P1xP2	Freezing	AFO	ManureApp	AFO2ManureApp	LID	Industrial WWTP	LAS	IndLAS	rMSE	P1	P1xP2	Freezing	AFO	ManureApp	AFO2ManureApp	LID	Industrial WWTP	LAS	IndLAS	rMSE
1		🟩	🟩	🟩			🟩	🟩				0.439			🟩	🟩			🟩				1.12
2		🟩	🟩	🟩			🟩	🟩				1.24			🟩	🟩							1.68
3		🟩	🟩	🟩			🟩	🟩				1.24			🟩	🟩			🟩				0.450
4		🟩	🟩	🟩			🟩	🟩				1.27			🟩	🟩			🟩				0.886
5		🟩	🟩	🟩			🟩	🟩				1.17			🟩	🟩			🟩				0.532
6		🟩	🟩	🟩			🟩	🟩				1.94			🟩	🟩			🟩				0.659
Mean rMSE		1.22											0.889										

Table A.10 Results of the 6-fold cross-validation of <u>test</u> stage of the FIT framework for *human*-associated responses. The shaded cells show which predictor was selected for each of the fold. The column named rMSE shows the root mean square error for the validation set that was held out.

	Sediment										Surface Water									
	P1	P1xP2	Freezing	Sep	LAS	DomLAS	Sep2DomLAS	WWTP	HID	rMSE	P1	P1xP2	Freezing	Sep	LAS	DomLAS	Sep2DomLAS	WWTP	HID	rMSE
1	🟩	🟩								2.39	🟩	🟩	🟩							0.384
2	🟩	🟩	🟩	🟩					🟩	1.69	🟩	🟩	🟩							1.49
3			🟩		🟩			🟩		1.69	🟩	🟩	🟩	🟩						0.927
4					🟩	🟩				1.59	🟩	🟩	🟩	🟩						0.882
5	🟩	🟩	🟩	🟩			🟩			2.11	🟩	🟩	🟩	🟩						0.479
6	🟩	🟩							🟩	1.78	🟩	🟩	🟩							0.627
Mean rMSE	1.88										0.797									

APPENDIX 3: SUPPLEMENTAL INFORMATION FOR AIM 3

A3.1 Details on Wisconsin Dairy Veterinarian Interviews and Responses

Table A.11 Interview questions and responses from Wisconsin veterinarians

Question	Veterinarian #1	Veterinarian #2	Veterinarian #3	Veterinarian #4
What antibiotic or antibiotics do you think is or are most frequently used on <u>commercial/conventional dairy cattle operations</u> and for what most common treatment or reason?	Beta lactams & ceftiofur for metritis & mastitis	Ceftiofur would be the most common antibiotic used in the adult dairy herd. It is used for the treatment of pneumonia, mastitis, metritis and footrot. In the young stock, ceftiofur is also used, but other common antibiotics are enrofloxacin, florfenicol, tulathromycin, and oxytetracycline used for the treatment of pneumonia.	Exenel (ceftiofur hydrochloride--cephalasporin), Excede (ceftiofur crystalline- cephalasporin), Polyflex (ampicillin), beta-lactams. Used for metritis, pneumonia, and foot problems.	ceftiofur
What antibiotic or antibiotics do you think is or are most frequently used on <u>organic dairy cattle operations</u> and for what most common treatment or reason?	None	I have no experience with organic dairies.	None	None
What antibiotic or antibiotics do you think is or are most frequently used on <u>small family dairy farms</u> and for what most common treatment or reason?	Beta lactams & ceftifur for metritis & mastitis	See above.	Small dairy farms are similar to conventional operation's use.	
Have you seen any changes in the use of antibiotics in livestock over time? Describe to what these changes are attributed to? If no change has occurred, you can write "none" or explain further.	Less antibiotics are being used in general. More culturing to assure appropriate usage.	There has been a decreased use of antibiotics over time as dairy operations have improved their housing and management resulting in decreased cases of pneumonia, diarrhea and mastitis.	Over time, more antibiotics used to be used. A lot less usage of antibiotics for mastitis. Hardly sell antibiotics for mastitis anymore. There are better methods and products for testing and treatment now. "Flushing" with tetracyline used to be common practice. This vet does not like the "modern" way. "Flushing" would be used in	None.

			the past 3 days after the cow passed the placenta to make sure the cow was "clean". Now the operations are supposed to monitor the cow's health after they have the calf to check for potential infection so that they would only use antibiotics if the cow was sick. However, this vet finds that the operations are not monitoring the cows well, so the placenta starts to rot and the cow becomes febrile.	
What do you think the prevalence of mastitis is among dairy cattle per 10,000? If you do not know, please write "not known".	500	Not known.	Not known.	Not known.
How long have you been working as a veterinarian with dairy cattle?	1-5 years	20+ years	20+ years	1-5 years

A3.2 Details of the Databases Representing Sources Terms Selected by FIT

Table A.12 provides the selected databases used to construct the source terms selected by the FIT model. Source types that are under the category of bovine sources are AFOs (alone, or via the ground transport of manure to application fields) and manure application fields. The source types that are under the category of land applied waste are the land applied sludge sites from municipal wastewater or septage, land applied sludge sites from industrial wastewater and food processing.

Table A.13 Databases representing the selected source terms as a result of the *Find* stage of FIT.

	Sediment				Surface Water				
	ermB	tetW	sul1	intI1	ermB	tetW	qnrA	sul1	intI1
Source Terms									
Bovine Sources **Database(s) representing spatially distributed sources**	AFO (via ground transport of manure to application fields) ***From:*** **WPDES CAFO weighted by Animal Units** ***To:*** **Crop rotation**	AFO (via ground transport of manure to application fields) ***From:*** **WPDES CAFO weighted by Animal Units** ***To:*** **Dairy rotation**	AFO (via ground transport of manure to application fields) ***From:*** **WPDES CAFO (unweighted)** ***To:*** **Crop rotation**	NS	AFO (via ground transport of manure to application fields) ***From:*** **Remotely sensed manure storages (unweighted)** ***To:*** **Crop rotation**	AFO **Remotely sensed manure storages weighted by log-total-gallons of manure**	AFO **Remotely sensed manure storages weighted by log-total-gallons of manure**	Manure App. Fields **Crop rot. land cover**	Manure App. Fields **Crop rotation land cover**
Land Application of Waste **Database(s) representing spatially distributed sources**	NS	NS	Land Applied Waste-Residential **WPDES Land applied sludge from treated septage or municipal wastewater**	Septage ground transport to land application sites ***From:*** **Septic Systems** ***To:*** **WPDES Land applied sludge from treated septage**	NS	Land Applied Waste-Industrial **WPDES Land applied sludge from treated food processing**	NS	NS	Land Applied Waste-Industrial **WPDES Land applied sludge from treated industrial sludge**

REFERENCES

(1) Daniel, C. R.; Cross, A. J.; Koebnick, C.; Sinha, R. Trends in meat consumption in the USA. *Public Health Nutr.* **2011**, *14*, 575–583.

(2) Morris, J. “One ought not have so delicate a nose”: CAFOs, agricultural nuisance, and the rise of the right to farm. *Gale In Context: Science*. 2017, p. 261+.

(3) Wannisky, K. E. *Environmental Protection Agency : National Pollutant Discharge Elimination System Permit Regulation and Effluent Limitation Guidelines and Standards for Concentrated Animal Feeding Operations (CAFOs).*; Washington, DC : U.S. General Accounting Office, [2003], 2003.

(4) Hollenbeck, J. E. Interaction of the role of Concentrated Animal Feeding Operations (CAFOs) in Emerging Infectious Diseases (EIDS). *Infect. Genet. Evol.* **2016**, *38*, 44–46.

(5) Robinson, T. P.; Bu, D. P.; Carrique-Mas, J.; Fèvre, E. M.; Gilbert, M.; Grace, D.; Hay, S. I.; Jiwakanon, J.; Kakkar, M.; Kariuki, S.; et al. Antibiotic resistance is the quintessential One Health issue. *Trans. R. Soc. Trop. Med. Hyg.* **2016**, *110*, 377–380.

(6) Amábile-Cuevas, C. *Antibiotics and antibiotic resistance in the environment*; CRC Press, 2015.

(7) Pollock, J.; Muwonge, A.; Hutchings, M. R.; Mainda, G.; Bronsvoort, B. M.; Gally, D. L.; Corbishley, A. Resistance to change: AMR gene dynamics on a commercial pig farm with high antimicrobial usage. *Sci. Rep.* **2020**, *10*, 1708.

(8) Zhang, Y.; Gu, A. Z.; He, M.; Li, D.; Chen, J. Subinhibitory Concentrations of Disinfectants Promote the Horizontal Transfer of Multidrug Resistance Genes within and across Genera. *Environ. Sci. Technol.* **2017**, *51*, 570–580.

(9) Wang, Y.; Lu, J.; Zhang, S.; Li, J.; Mao, L.; Yuan, Z.; Bond, P. L.; Guo, J. Non-antibiotic pharmaceuticals promote the transmission of multidrug resistance plasmids through intra- and intergenera conjugation. *ISME J.* **2021**.

(10) Chow, L. K. M.; Ghaly, T. M.; Gillings, M. R. A survey of sub-inhibitory concentrations of antibiotics in the environment. *J. Environ. Sci. (China)* **2021**, *99*, 21–27.

(11) Van Boeckel, T. P.; Brower, C.; Gilbert, M.; Grenfell, B. T.; Levin, S. A.; Robinson, T. P.; Teillant, A.; Laxminarayan, R. Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5649–5654.

(12) Davis, M. F.; Pisanic, N.; Rhodes, S. M.; Brown, A.; Keller, H.; Nadimpalli, M.; Christ, A.; Ludwig, S.; Ordak, C.; Spicer, K.; et al. Occurrence of Staphylococcus aureus in swine and swine workplace environments on industrial and antibiotic-free hog operations in North Carolina, USA: A One Health pilot study. *Environ. Res.* **2018**, *163*, 88–96.

(13) Hatcher, S. M.; Rhodes, S. M.; Stewart, J. R.; Silbergeld, E.; Pisanic, N.; Larsen, J.; Jiang, S.; Krosche, A.; Hall, D.; Carroll, K. C.; et al. The Prevalence of Antibiotic-Resistant Staphylococcus aureus Nasal Carriage among Industrial Hog Operation Workers, Community Residents, and Children Living in Their Households: North Carolina, USA. *Environ. Health Perspect.* **2017**, *125*, 560–569.

(14) Raff, Z.; Meyer, A. CAFOs and Surface Water Quality: Evidence from the Proliferation of Large Farms in Wisconsin. *SSRN Journal* **2019**.

(15) Ogneva-Himmelberger, Y.; Huang, L.; Xin, H. CALPUFF and CAFOs: Air Pollution Modeling and Environmental Justice Analysis in the North Carolina Hog Industry. *ISPRS Int J Geoinf* **2015**, *4*, 150–171.

(16) Lopatto, E.; Choi, J.; Colina, A.; Ma, L.; Howe, A.; Hinsa-Leasure, S. Characterizing the soil microbiome and quantifying antibiotic resistance gene dynamics in agricultural soil following swine CAFO manure application. *PLoS One* **2019**, *14*, e0220770.

(17) Pruden, A.; Arabi, M.; Storteboom, H. N. Correlation between upstream human activities and riverine antibiotic resistance genes. *Environ. Sci. Technol.* **2012**, *46*, 11541–11549.

(18) Costa, C. M. da S. B.; Leite, I. R.; Almeida, A. K.; de Almeida, I. K. Choosing an appropriate water quality model-a review. *Environ. Monit. Assess.* **2021**, *193*, 38.

(19) Liu, Y.; Li, S.; Wallace, C. W.; Chaubey, I.; Flanagan, D. C.; Theller, L. O.; Engel, B. A. Comparison of computer models for estimating hydrology and water quality in an agricultural watershed. *Water Resour. Manage.* **2017**, *31*, 3641–3665.

(20) McCarthy, D. T.; Jovanovic, D.; Lintern, A.; Teakle, I.; Barnes, M.; Deletic, A.; Coleman, R.; Rooney, G.; Prosser, T.; Coutts, S.; et al. Source tracking using microbial community fingerprints: Method comparison with hydrodynamic modelling. *Water Res.* **2017**, *109*, 253–265.

(21) Hassoun-Kheir, N.; Stabholz, Y.; Kreft, J.-U.; de la Cruz, R.; Romalde, J. L.; Nesme, J.; Sørensen, S. J.; Smets, B. F.; Graham, D.; Paul, M. Comparison of antibiotic-resistant bacteria and antibiotic resistance genes abundance in hospital and community wastewater: A systematic review. *Sci. Total Environ.* **2020**, *743*, 140804.

(22) Beattie, R. E.; Bandla, A.; Swarup, S.; Hristova, K. R. Freshwater sediment microbial communities are not resilient to disturbance from agricultural land runoff. *Front. Microbiol.* **2020**, *11*, 539921.

(23) Korajkic, A.; Wanjugi, P.; Brooks, L.; Cao, Y.; Harwood, V. J. Persistence and decay of fecal microbiota in aquatic habitats. *Microbiol. Mol. Biol. Rev.* **2019**, *83*.

(24) Bell, A.; Layton, A. C.; McKay, L.; Williams, D.; Gentry, R.; Sayler, G. S. Factors influencing the persistence of fecal Bacteroides in stream water. *J Environ Qual* **2009**, *38*, 1224–1232.

(25) Schulz, C. J.; Childers, G. W. Fecal bacteroidales diversity and decay in response to variations in temperature and salinity. *Appl. Environ. Microbiol.* **2011**, *77*, 2563–2572.

(26) Holcomb, D. A.; Stewart, J. R. Microbial indicators of fecal pollution: recent progress and challenges in assessing water quality. *Curr. Environ. Health Rep.* **2020**, *7*, 311–324.

(27) Rodgers, K.; McLellan, I.; Peshkur, T.; Williams, R.; Tonner, R.; Hursthouse, A. S.; Knapp, C. W.; Henriquez, F. L. Can the legacy of industrial pollution influence antimicrobial resistance in estuarine sediments? *Environ Chem Lett* **2018**, *17*, 1–13.

(28) *Microbial source tracking: methods, applications, and case studies*; Hagedorn, C.; Blanch, A. R.; Harwood, V. J., Eds.; Springer New York: New York, NY, 2011.

(29) McKee, B. A.; Molina, M.; Cyterski, M.; Couch, A. Microbial source tracking (MST) in Chattahoochee River National Recreation Area: Seasonal and precipitation trends in MST marker concentrations, and associations with E. coli levels, pathogenic marker presence, and land-use. *Water Res.* **2020**, *171*, 115435.

(30) Dila, D. K.; Corsi, S. R.; Lenaker, P. L.; Baldwin, A. K.; Bootsma, M. J.; McLellan, S. L. Patterns of Host-Associated Fecal Indicators Driven by Hydrology, Precipitation, and Land Use Attributes in Great Lakes Watersheds. *Environ. Sci. Technol.* **2018**, *52*, 11500–11509.

(31) Saeidi, N.; Gu, X.; Tran, N. H.; Goh, S. G.; Kitajima, M.; Kushmaro, A.; Schmitz, B. W.; Gin, K. Y.-H. Occurrence of Traditional and Alternative Fecal Indicators in Tropical Urban Environments under Different Land Use Patterns. *Appl. Environ. Microbiol.* **2018**, *84*.

(32) Vitro, K. A.; BenDor, T. K.; Jordanova, T. V.; Miles, B. A geospatial analysis of land-use and stormwater management on fecal coliform contamination in North Carolina streams. *Sci. Total Environ.* **2017**, *603–604*, 709–727.

(33) Brendel, C.; Soupir, M. Relating Watershed Characteristics to Elevated Stream Escherichia coli Levels in Agriculturally Dominated Landscapes: An Iowa Case Study. *Water (Basel)* **2017**, *9*, 154.

(34) Li, X.; Atwill, E. R.; Antaki, E.; Applegate, O.; Bergamaschi, B.; Bond, R. F.; Chase, J.; Ransom, K. M.; Samuels, W.; Watanabe, N.; et al. Fecal Indicator and Pathogenic Bacteria and Their Antibiotic Resistance in Alluvial Groundwater of an Irrigated Agricultural Region with Dairies. *J Environ Qual* **2015**, *44*, 1435–1447.

(35) Crowther, J.; Wyer, M. D.; Bradford, M.; Kay, D.; Francis, C. A. Modelling faecal indicator concentrations in large rural catchments using land-use and topographic data. *J. Appl. Microbiol.* **2003**, *94*, 962–973.

(36) Fraser, R. H.; Barten, P. K.; Pinney, D. A. K. Predicting Stream Pathogen Loading from Livestock using a Geographical Information System-Based Delivery Model. *J Environ Qual* **1998**, *27*, 935–945.

(37) Sowah, R. A.; Habteselassie, M. Y.; Radcliffe, D. E.; Bauske, E.; Risse, M. Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. *Water Res.* **2017**, *108*, 330–338.

(38) Amos, G. C. A.; Gozzard, E.; Carter, C. E.; Mead, A.; Bowes, M. J.; Hawkey, P. M.; Zhang, L.; Singer, A. C.; Gaze, W. H.; Wellington, E. M. H. Validated predictive modelling of the environmental resistome. *ISME J.* **2015**, *9*, 1467–1476.

(39) McEachran, A. D.; Blackwell, B. R.; Hanson, J. D.; Wooten, K. J.; Mayer, G. D.; Cox, S. B.; Smith, P. N. Antibiotics, bacteria, and antibiotic resistance genes: aerial transport from cattle feed yards via particulate matter. *Environ. Health Perspect.* **2015**, *123*, 337–343.

(40) Holcomb, D. A.; Messier, K. P.; Serre, M. L.; Rowny, J. G.; Stewart, J. R. Geostatistical prediction of microbial water quality throughout a stream network using meteorology, land cover, and spatiotemporal autocorrelation. *Environ. Sci. Technol.* **2018**, *52*, 7775–7784.

(41) Liu, Y.; Han, C.; Chen, Z.; Guo, D.; Ye, X. Relationship between livestock exposure and methicillin-resistant Staphylococcus aureus carriage in humans: A systematic review and dose-response meta-analysis. *Int. J. Antimicrob. Agents* **2019**.

(42) Schultz, A. A.; Peppard, P.; Gangnon, R. E.; Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **2019**, *130*, 104911.

(43) Guillien, A.; Puyraveau, M.; Soumagne, T.; Guillot, S.; Rannou, F.; Marquette, D.; Berger, P.; Jouneau, S.; Monnet, E.; Mauny, F.; et al. Prevalence and risk factors for COPD in farmers: a cross-sectional controlled study. *Eur. Respir. J.* **2016**, *47*, 95–103.

(44) Borlée, F.; Yzermans, C. J.; Aalders, B.; Rooijackers, J.; Krop, E.; Maassen, C. B. M.; Schellevis, F.; Brunekreef, B.; Heederik, D.; Smit, L. A. M. Air Pollution from Livestock Farms Is Associated with Airway Obstruction in Neighboring Residents. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, 1152–1161.

(45) Hooiveld, M.; Smit, L. A. M.; van der Sman-de Beer, F.; Wouters, I. M.; van Dijk, C. E.; Spreeuwenberg, P.; Heederik, D. J. J.; Yzermans, C. J. Doctor-diagnosed health problems in a region with a high density of concentrated animal feeding operations: a cross-sectional study. *Environ Health* **2016**, *15*, 24.

(46) van Dijk, C. E.; Zock, J.-P.; Baliatsas, C.; Smit, L. A. M.; Borlée, F.; Spreeuwenberg, P.; Heederik, D.; Yzermans, C. J. Health conditions in rural areas with high livestock density: Analysis of seven consecutive years. *Environ. Pollut.* **2017**, *222*, 374–382.

(47) Lantos, P. M.; Hoffman, K.; Höhle, M.; Anderson, B.; Gray, G. C. Are people living near modern swine production facilities at increased risk of influenza virus infection? *Clin. Infect. Dis.* **2016**, *63*, 1558–1563.

(48) Pérez-Rodríguez, F.; Mercanoglu Taban, B. A State-of-Art Review on Multi-Drug Resistant Pathogens in Foods of Animal Origin: Risk Factors and Mitigation Strategies. *Front. Microbiol.* **2019**, *10*, 2091.

(49) Heredia, N.; García, S. Animals as sources of food-borne pathogens: A review. *Anim. Nutr.* **2018**, *4*, 250–255.

(50) Ben, Y.; Fu, C.; Hu, M.; Liu, L.; Wong, M. H.; Zheng, C. Human health risk assessment of antibiotic resistance associated with antibiotic residues in the environment: A review. *Environ. Res.* **2019**, *169*, 483–493.

(51) Singer, A.; Shaw, H.; Rhodes, V.; Hart, A. Review of antimicrobial resistance in the environment and its relevance to environmental regulators. *Front. Microbiol.* **2016**, *7*, 1728.

(52) Berendonk, T. U.; Manaia, C. M.; Merlin, C.; Fatta-Kassinos, D.; Cytryn, E.; Walsh, F.; Bürgmann, H.; Sørum, H.; Norström, M.; Pons, M.-N.; et al. Tackling antibiotic resistance: the environmental framework. *Nat. Rev. Microbiol.* **2015**, *13*, 310–317.

(53) Reygaert, W. C. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol.* **2018**, *4*, 482–501.

(54) *Antimicrobial resistance in the 21st century*; Fong, I. W.; Shlaes, D.; Drlica, K., Eds.; Springer International Publishing: Cham, 2018.

(55) Zhang, M.; Zuo, J.; Yu, X.; Shi, X.; Chen, L.; Li, Z. Quantification of multi-antibiotic resistant opportunistic pathogenic bacteria in bioaerosols in and around a pharmaceutical wastewater treatment plant. *J. Environ. Sci. (China)* **2018**, *72*, 53–63.

(56) Liu, C. M.; Stegger, M.; Aziz, M.; Johnson, T. J.; Waits, K.; Nordstrom, L.; Gauld, L.; Weaver, B.; Rolland, D.; Statham, S.; et al. Escherichia coli ST131-H22 as a Foodborne Uropathogen. *MBio* **2018**, *9*.

(57) Harwood, V. J.; Staley, C.; Badgley, B. D.; Borges, K.; Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **2014**, *38*, 1–40.

(58) Abia, A. L. K.; Ubomba-Jaswa, E.; Genthe, B.; Momba, M. N. B. 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.* **2016**, *566–567*, 1143–1151.

(59) Zhang, Q.; Gallard, J.; Wu, B.; Harwood, V. J.; Sadowsky, M. J.; Hamilton, K. A.; Ahmed, W. Synergy between quantitative microbial source tracking (qMST) and quantitative microbial risk assessment (QMRA): A review and prospectus. *Environ. Int.* **2019**, *130*, 104703.

(60) Imran, M.; Das, K. R.; Naik, M. M. Co-selection of multi-antibiotic resistance in bacterial pathogens in metal and microplastic contaminated environments: An emerging health threat. *Chemosphere* **2019**, *215*, 846–857.

(61) Zhao, Y.; Cocerva, T.; Cox, S.; Tardif, S.; Su, J.-Q.; Zhu, Y.-G.; Brandt, K. K. Evidence for co-selection of antibiotic resistance genes and mobile genetic elements in metal polluted urban soils. *Sci. Total Environ.* **2019**, *656*, 512–520.

(62) Ding, J.; An, X. L.; Lassen, S. B.; Wang, H. T.; Zhu, D.; Ke, X. Heavy metal-induced co-selection of antibiotic resistance genes in the gut microbiota of collembolans. *Sci. Total Environ.* **2019**, *683*, 210–215.

(63) D'Costa, V. M.; King, C. E.; Kalan, L.; Morar, M.; Sung, W. W. L.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; et al. Antibiotic resistance is ancient. *Nature* **2011**, *477*, 457–461.

(64) Bueno, I.; Williams-Nguyen, J.; Hwang, H.; Sargeant, J. M.; Nault, A. J.; Singer, R. S. Systematic Review: Impact of point sources on antibiotic-resistant bacteria in the natural environment. *Zoonoses Public Health* **2018**, *65*, e162–e184.

(65) Douglas, P.; Robertson, S.; Gay, R.; Hansell, A. L.; Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int J Hyg Environ Health* **2018**, *221*, 134–173.

(66) Zhang, Y.; Zhang, C.; Parker, D. B.; Snow, D. D.; Zhou, Z.; Li, X. Occurrence of antimicrobials and antimicrobial resistance genes in beef cattle storage ponds and swine treatment lagoons. *Sci. Total Environ.* **2013**, *463–464*, 631–638.

(67) Chapin, A.; Rule, A.; Gibson, K.; Buckley, T.; Schwab, K. Airborne multidrug-resistant bacteria isolated from a concentrated swine feeding operation. *Environ. Health Perspect.* **2005**, *113*, 137–142.

(68) Gibbs, S. G.; Green, C. F.; Tarwater, P. M.; Scarpino, P. V. Airborne antibiotic resistant and nonresistant bacteria and fungi recovered from two swine herd confined animal feeding operations. *J Occup Environ Hyg* **2004**, *1*, 699–706.

(69) Banhazi, T. M.; Rutley, D. L.; Pitchford, W. S. Identification of risk factors for sub-optimal housing conditions in Australian piggeries: Part 4. Emission factors and study recommendations. *J. Agric. Saf. Health* **2008**, *14*, 53–69.

(70) Rogers, S. W.; Shaffer, C. E.; Langen, T. A.; Jahne, M.; Welsh, R. Antibiotic-Resistant Genes and Pathogens Shed by Wild Deer Correlate with Land Application of Residuals. *Ecohealth* **2018**, *15*, 409–425.

(71) Mallin, M. A.; McIver, M. R. Season matters when sampling streams for swine CAFO waste pollution impacts. *J Water Health* **2018**, *16*, 78–86.

(72) Heaney, C. D.; Myers, K.; Wing, S.; Hall, D.; Baron, D.; Stewart, J. R. Source tracking swine fecal waste in surface water proximal to swine concentrated animal feeding operations. *Sci. Total Environ.* **2015**, *511*, 676–683.

(73) Gibbs, S. G.; Green, C. F.; Tarwater, P. M.; Mota, L. C.; Mena, K. D.; Scarpino, P. V. Isolation of antibiotic-resistant bacteria from the air plume downwind of a swine confined or concentrated animal feeding operation. *Environ. Health Perspect.* **2006**, *114*, 1032–1037.

(74) Hu, D.; Wang-Li, L.; Simmons III, O. D.; Classen, J. J.; Osborne, J. A. Spatiotemporal variations of bioaerosols in the vicinity of an animal feeding operation facility in the US. *JEP* **2015**, *06*, 614–627.

(75) Jacobs, K.; Wind, L.; Krometis, L.-A.; Hession, W. C.; Pruden, A. Fecal Indicator Bacteria and Antibiotic Resistance Genes in Storm Runoff from Dairy Manure and Compost-Amended Vegetable Plots. *J Environ Qual* **2019**, *48*, 1038–1046.

(76) Burch, T. R.; Spencer, S. K.; Stokdyk, J. P.; Kieke, B. A.; Larson, R. A.; Firnstahl, A. D.; Rule, A. M.; Borchardt, M. A. Quantitative microbial risk assessment for spray irrigation of dairy manure based on an empirical fate and transport model. *Environ. Health Perspect.* **2017**, *125*, 087009.

(77) Pazda, M.; Kumirska, J.; Stepnowski, P.; Mulkiewicz, E. Antibiotic resistance genes identified in wastewater treatment plant systems - A review. *Sci. Total Environ.* **2019**, *697*, 134023.

(78) Petit, F.; Denamur, E.; Clermont, O.; Leclercq, R.; Deloffre, J.; Cattoir, V.; Oberlé, K.; Budzinski, H.; Berthe, T. Fate of Antibiotics and Antibiotic-Resistant Fecal Bacteria in Water and Sediments from the Contamination Source to the Estuary: Impact and/or Resilience? Resilience to Contamination by Antibiotics. In *Marine Productivity: Perturbations and Resilience of Socio-ecosystems*; Ceccaldi, H.-J.; Hénocque, Y.; Koike, Y.; Komatsu, T.; Stora, G.; Tusseau-Vuillemin, M.-H., Eds.; Springer International Publishing: Cham, 2015; pp. 79–91.

(79) Czekalski, N.; Gascón Díez, E.; Bürgmann, H. Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. *ISME J.* **2014**, *8*, 1381–1390.

(80) Reinthaler, F. F.; Posch, J.; Feierl, G.; Wüst, G.; Haas, D.; Ruckenbauer, G.; Mascher, F.; Marth, E. Antibiotic resistance of E. coli in sewage and sludge. *Water Res.* **2003**, *37*, 1685–1690.

(81) O'Dwyer, J.; Hynds, P.; Pot, M.; Adley, C. C.; Ryan, M. P. Evaluation of levels of antibiotic resistance in groundwater-derived E. coli isolates in the Midwest of Ireland and elucidation of potential predictors of resistance. *Hydrogeol. J.* **2017**, *25*, 939–951.

(82) Peed, L. A.; Nietch, C. T.; Kelty, C. A.; Meckes, M.; Mooney, T.; Sivaganesan, M.; Shanks, O. C. Combining land-use information and small stream sampling with PCR-based methods for better characterization of diffuse sources of human fecal pollution. *Environ. Sci. Technol.* **2011**, *45*, 5652–5659.

(83) Chen, H.; Zhang, M. Occurrence and removal of antibiotic resistance genes in municipal wastewater and rural domestic sewage treatment systems in eastern China. *Environ. Int.* **2013**, *55*, 9–14.

(84) Yoshikawa, T. T. Antimicrobial resistance and aging: beginning of the end of the antibiotic era? *J. Am. Geriatr. Soc.* **2002**, *50*, S226-9.

(85) Abia, A. L. K.; Ubomba-Jaswa, E.; Schmidt, C.; Dippenaar, M. A. Where Did They Come from-Multi-Drug Resistant Pathogenic Escherichia coli in a Cemetery Environment? *Antibiotics (Basel)* **2018**, *7*.

(86) Abia, A. L. K.; Alisoltani, A.; Ubomba-Jaswa, E.; Dippenaar, M. A. Microbial life beyond the grave: 16S rRNA gene-based metagenomic analysis of bacteria diversity and their functional profiles in cemetery environments. *Sci. Total Environ.* **2019**, *655*, 831–841.

(87) Christenson, E. Christenson, Elizabeth. A Longitudinal, Landscape-Scale Field Study Assessing The Effects of Commercial Hog Operations on Microbial Quality of Surface Waters in North Carolina, USA. Doctoral dissertation, The University of North Carolina at Chapel Hill, 2019.

(88) Chow, T. E.; Dede-Bamfo, N.; Dahal, K. R. Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. *Ann. GIS* **2016**, *22*, 29–42.

(89) Jeong, J.; Wagner, K.; Flores, J. J.; Cawthon, T.; Her, Y.; Osorio, J.; Yen, H. Linking watershed modeling and bacterial source tracking to better assess E. coli sources. *Sci. Total Environ.* **2019**, *648*, 164–175.

(90) Wu, J. Linking landscape patterns to sources of water contamination: Implications for tracking fecal contaminants with geospatial and Bayesian approaches. *Sci. Total Environ.* **2019**, *650*, 1149–1157.

(91) Peterson, E. E.; Sheldon, F.; Darnell, R.; Bunn, S. E.; Harch, B. D. A comparison of spatially explicit landscape representation methods and their relationship to stream condition. *Freshw Biol* **2011**, *56*, 590–610.

(92) Chen, Q.; Mei, K.; Dahlgren, R. A.; Wang, T.; Gong, J.; Zhang, M. Impacts of land-use and population density on seasonal surface water quality using a modified geographically weighted regression. *Sci. Total Environ.* **2016**, *572*, 450–466.

(93) Messier, K. P.; Kane, E.; Bolich, R.; Serre, M. L. Nitrate variability in groundwater of North Carolina using monitoring and private well data models. *Environ. Sci. Technol.* **2014**, *48*, 10804–10812.

(94) Nuzzo, R. Scientific method: statistical errors. *Nature* **2014**, *506*, 150–152.

(95) Tate, K. W.; Atwill, E. R.; Bartolome, J. W.; Nader, G. Significant Escherichia coli attenuation by vegetative buffers on annual grasslands. *J Environ Qual* **2006**, *35*, 795–805.

(96) Lamori, J. G.; Xue, J.; Rachmadi, A. T.; Lopez, G. U.; Kitajima, M.; Gerba, C. P.; Pepper, I. L.; Brooks, J. P.; Sherchan, S. Removal of fecal indicator bacteria and antibiotic resistant genes in constructed wetlands. *Environ. Sci. Pollut. Res. Int.* **2019**, *26*, 10188–10197.

(97) Knox, A. K.; Dahlgren, R. A.; Tate, K. W.; Atwill, E. R. Efficacy of natural wetlands to retain nutrient, sediment and microbial pollutants. *J Environ Qual* **2008**, *37*, 1837–1846.

(98) Lu, Z.; Na, G.; Gao, H.; Wang, L.; Bao, C.; Yao, Z. Fate of sulfonamide resistance genes in estuary environment and effect of anthropogenic activities. *Sci. Total Environ.* **2015**, *527–528*, 429–438.

(99) Yao, L.; Wang, Y.; Tong, L.; Li, Y.; Deng, Y.; Guo, W.; Gan, Y. Seasonal variation of antibiotics concentration in the aquatic environment: a case study at Jianghan Plain, central China. *Sci. Total Environ.* **2015**, *527–528*, 56–64.

(100) Wang, M.; Sun, Y.; Liu, P.; Sun, J.; Zhou, Q.; Xiong, W.; Zeng, Z. Fate of antimicrobial resistance genes in response to application of poultry and swine manure in simulated manure-soil microcosms and manure-pond microcosms. *Environ. Sci. Pollut. Res. Int.* **2017**, *24*, 20949–20958.

(101) Fahrenfeld, N.; Knowlton, K.; Krometis, L. A.; Hession, W. C.; Xia, K.; Lipscomb, E.; Libuit, K.; Green, B. L.; Pruden, A. Effect of manure application on abundance of antibiotic resistance genes and their attenuation rates in soil: field-scale mass balance approach. *Environ. Sci. Technol.* **2014**, *48*, 2643–2650.

(102) Rysz, M.; Alvarez, P. J. J. Amplification and attenuation of tetracycline resistance in soil bacteria: aquifer column experiments. *Water Res.* **2004**, *38*, 3705–3712.

(103) Murphy, H. M.; McGinnis, S.; Blunt, R.; Stokdyk, J.; Wu, J.; Cagle, A.; Denno, D. M.; Spencer, S.; Firnstahl, A.; Borchardt, M. A. Septic systems and rainfall influence human fecal marker and indicator organism occurrence in private wells in southeastern pennsylvania. *Environ. Sci. Technol.* **2020**, *54*, 3159–3168.

(104) Zheng, J.; Zhou, Z.; Wei, Y.; Chen, T.; Feng, W.; Chen, H. High-throughput profiling of seasonal variations of antibiotic resistance gene transport in a peri-urban river. *Environ. Int.* **2018**, *114*, 87–94.

(105) Calero-Cáceres, W.; Méndez, J.; Martín-Díaz, J.; Muniesa, M. The occurrence of antibiotic resistance genes in a Mediterranean river and their persistence in the riverbed sediment. *Environ. Pollut.* **2017**, *223*, 384–394.

(106) Knapp, C. W.; Lima, L.; Olivares-Rieumont, S.; Bowen, E.; Werner, D.; Graham, D. W. Seasonal variations in antibiotic resistance gene transport in the almendares river, havana, cuba. *Front. Microbiol.* **2012**, *3*, 396.

(107) Li, J.; Cao, J.; Zhu, Y.-G.; Chen, Q.-L.; Shen, F.; Wu, Y.; Xu, S.; Fan, H.; Da, G.; Huang, R.-J.; et al. Global survey of antibiotic resistance genes in air. *Environ. Sci. Technol.* **2018**, *52*, 10975–10984.

(108) Burch, T. R.; Sadowsky, M. J.; LaPara, T. M. Fate of antibiotic resistance genes and class 1 integrons in soil microcosms following the application of treated residual municipal wastewater solids. *Environ. Sci. Technol.* **2014**, *48*, 5620–5627.

(109) Ling, A. L.; Pace, N. R.; Hernandez, M. T.; LaPara, T. M. Tetracycline resistance and Class 1 integron genes associated with indoor and outdoor aerosols. *Environ. Sci. Technol.* **2013**, *47*, 4046–4052.

(110) Gillings, M. R.; Gaze, W. H.; Pruden, A.; Smalla, K.; Tiedje, J. M.; Zhu, Y.-G. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J.* **2015**, *9*, 1269–1279.

(111) Ahmed, W.; Hamilton, K.; Toze, S.; Cook, S.; Page, D. A review on microbial contaminants in stormwater runoff and outfalls: Potential health risks and mitigation strategies. *Sci. Total Environ.* **2019**, *692*, 1304–1321.

(112) Huijbers, P. M. C.; Blaak, H.; de Jong, M. C. M.; Graat, E. A. M.; Vandenbroucke-Grauls, C. M. J. E.; de Roda Husman, A. M. Role of the environment in the transmission of antimicrobial resistance to humans: A review. *Environ. Sci. Technol.* **2015**, *49*, 11993–12004.

(113) Weiskerger, C. J.; Phanikumar, M. S. Numerical modeling of microbial fate and transport in natural waters: review and implications for normal and extreme storm events. *Water (Basel)* **2020**, *12*, 1876.

(114) Cowie, C. T.; Garden, F.; Jegasothy, E.; Knibbs, L. D.; Hanigan, I.; Morley, D.; Hansell, A.; Hoek, G.; Marks, G. B. Comparison of model estimates from an intra-city land-use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO2 for a birth cohort in Sydney, Australia. *Environ. Res.* **2019**, *174*, 24–34.

(115) Messier, K. P.; Akita, Y.; Serre, M. L. Integrating address geocoding, land-use regression, and spatiotemporal geostatistical estimation for groundwater tetrachloroethylene. *Environ. Sci. Technol.* **2012**, *46*, 2772–2780.

(116) Messier, K. P.; Campbell, T.; Bradley, P. J.; Serre, M. L. Estimation of groundwater radon in north carolina using land-use regression and bayesian maximum entropy. *Environ. Sci. Technol.* **2015**, *49*, 9817–9825.

(117) Su, J. G.; Jerrett, M.; Beckerman, B. A distance-decay variable selection strategy for land-use regression modeling of ambient air pollution exposures. *Sci. Total Environ.* **2009**, *407*, 3890–3898.

(118) Ding, J.; Jiang, Y.; Liu, Q.; Hou, Z.; Liao, J.; Fu, L.; Peng, Q. Influences of the land-use pattern on water quality in low-order streams of the Dongjiang River basin, China: A multi-scale analysis. *Sci. Total Environ.* **2016**, *551–552*, 205–216.

(119) Porter, K. D. H.; Reaney, S. M.; Quilliam, R. S.; Burgess, C.; Oliver, D. M. Predicting diffuse microbial pollution risk across catchments: The performance of SCIMAP and recommendations for future development. *Sci. Total Environ.* **2017**, *609*, 456–465.

(120) Wang, Y.; Jiang, R.; Xie, J.; Zhao, Y.; Yan, D.; Yang, S. Soil and water assessment tool (SWAT) model: A systemic review. *Journal of Coastal Research* **2019**, *93*, 22.

(121) Chang, C.; Sun, D.; Feng, P.; Zhang, M.; Ge, N. Impacts of nonpoint source pollution on water quality in the yuqiao reservoir. *Environ Eng Sci* **2017**, *34*, 418–432.

(122) Cloutier, D. D.; Alm, E. W.; McLellan, S. L. Influence of land-use, nutrients, and geography on microbial communities and fecal indicator abundance at lake michigan beaches. *Appl. Environ. Microbiol.* **2015**, *81*, 4904–4913.

(123) Ballesté, E.; Belanche-Muñoz, L. A.; Farnleitner, A. H.; Linke, R.; Sommer, R.; Santos, R.; Monteiro, S.; Maunula, L.; Oristo, S.; Tiehm A, A.; et al. Improving the identification of the source of faecal pollution in water using a modelling approach: From multi-source to aged and diluted samples. *Water Res.* **2019**, *171*, 115392.

(124) Cho, K. H.; Pachepsky, Y. A.; Oliver, D. M.; Muirhead, R. W.; Park, Y.; Quilliam, R. S.; Shelton, D. R. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. *Water Res.* **2016**, *100*, 38–56.

(125) Zivkovich, B. R.; Mays, D. C. Predicting nonpoint stormwater runoff quality from land-use. *PLoS One* **2018**, *13*, e0196782.

(126) Hohl, A.; Václavík, T.; Meentemeyer, R. K. Go with the flow: geospatial analytics to quantify hydrologic landscape connectivity for passively dispersed microorganisms. *International Journal of Geographical Information Science* **2014**, *28*, 1626–1641.

(127) Neill, A. J.; Tetzlaff, D.; Strachan, N. J. C.; Hough, R. L.; Avery, L. M.; Watson, H.; Soulsby, C. Using spatial-stream-network models and long-term data to understand and predict dynamics of faecal contamination in a mixed land-use catchment. *Sci. Total Environ.* **2018**, *612*, 840–852.

(128) Seltenrich, N. Manure irrigation: environmental benefits, potential human health risks. *Environ. Health Perspect.* **2017**, *125*, 124004.

(129) J. C. Hadrich; T. M. Harrigan; C. A. Wolf. Economic comparison of liquid manure transport and land application. *Appl Eng Agric* **2010**, *26*, 743–758.

(130) Wohl, E. Legacy effects on sediments in river corridors. *Earth-Science Reviews* **2015**, *147*, 30–53.

(131) Mao, D.; Luo, Y.; Mathieu, J.; Wang, Q.; Feng, L.; Mu, Q.; Feng, C.; Alvarez, P. J. J. Persistence of extracellular DNA in river sediment facilitates antibiotic resistance gene propagation. *Environ. Sci. Technol.* **2014**, *48*, 71–78.

(132) Pei, R.; Kim, S.-C.; Carlson, K. H.; Pruden, A. Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Res.* **2006**, *40*, 2427–2435.

(133) Wright, K.; Peck, J. E.; Kewaunee County Citizens. *Industrial Dairies in Kewaunee County, Wisconsin*; Socially Responsible Agriculture Project, 2015.

(134) Beattie, R. E.; Walsh, M.; Cruz, M. C.; McAliley, L. R.; Dodgen, L.; Zheng, W.; Hristova, K. R. Agricultural contamination impacts antibiotic resistance gene abundances in river bed sediment temporally. *FEMS Microbiol. Ecol.* **2018**, *94*.

(135) University of Wisconsin Madison. Wiscland 2 project complete, data now available – State Cartographer’s Office – UW–Madison https://www.sco.wisc.edu/2016/09/23/wiscland-2-project-complete-data-now-available/ (accessed Jul 1, 2020).

(136) Layton, A.; McKay, L.; Williams, D.; Garrett, V.; Gentry, R.; Sayler, G. Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl. Environ. Microbiol.* **2006**, *72*, 4214–4224.

(137) Coulliette, A. D.; Money, E. S.; Serre, M. L.; Noble, R. T. Space/time analysis of fecal pollution and rainfall in an eastern north carolina estuary. *Environ. Sci. Technol.* **2009**, *43*, 3728–3735.

(138) Ponce-Terashima, R.; Koskey, A. M.; Reis, M. G.; McLellan, S. L.; Blanton, R. E. Sources and distribution of surface water fecal contamination and prevalence of schistosomiasis in a Brazilian village. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3186.

(139) Hu, X.; He, K.; Zhou, Q. Occurrence, accumulation, attenuation and priority of typical antibiotics in sediments based on long-term field and modeling studies. *J. Hazard. Mater.* **2012**, *225–226*, 91–98.

(140) Drummond, J. D.; Davies-Colley, R. J.; Stott, R.; Sukias, J. P.; Nagels, J. W.; Sharp, A.; Packman, A. I. Microbial transport, retention, and inactivation in streams: A combined experimental and stochastic modeling approach. *Environ. Sci. Technol.* **2015**, *49*, 7825–7833.

(141) Brown, P. C.; Borowska, E.; Schwartz, T.; Horn, H. Impact of the particulate matter from wastewater discharge on the abundance of antibiotic resistance genes and facultative pathogenic bacteria in downstream river sediments. *Sci. Total Environ.* **2019**, *649*, 1171–1178.

(142) Xue, F.; Tang, J.; Dong, Z.; Shen, D.; Liu, H.; Zhang, X.; Holden, N. M. Tempo-spatial controls of total coliform and E. coli contamination in a subtropical hilly agricultural catchment. *Agricultural Water Management* **2018**, *200*, 10–18.

(143) Amos, G. C. A.; Ploumakis, S.; Zhang, L.; Hawkey, P. M.; Gaze, W. H.; Wellington, E. M. H. The widespread dissemination of integrons throughout bacterial communities in a riverine system. *ISME J.* **2018**, *12*, 681–691.

(144) Guan, Y.; Jia, J.; Wu, L.; Xue, X.; Zhang, G.; Wang, Z. Analysis of bacterial community characteristics, abundance of antibiotics and antibiotic resistance genes along a pollution gradient of ba river in xi'an, china. *Front. Microbiol.* **2018**, *9*, 3191.

(145) Jat, P.; Serre, M. L. A novel geostatistical approach combining Euclidean and gradual-flow covariance models to estimate fecal coliform along the Haw and Deep rivers in North Carolina. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1–13.

(146) Hughes, R. M.; Kaufmann, P. R.; Weber, M. H. National and regional comparisons between Strahler order and stream size. *Journal of the North American Benthological Society* **2011**, *30*, 103–121.

(147) Money, E.; Carter, G. P.; Serre, M. L. Using river distances in the space/time estimation of dissolved oxygen along two impaired river networks in New Jersey. *Water Res.* **2009**, *43*, 1948–1958.

(148) Jat, P.; Serre, M. L. Bayesian Maximum Entropy space/time estimation of surface water chloride in Maryland using river distances. *Environ. Pollut.* **2016**, *219*, 1148–1155.

(149) Money, E. S.; Carter, G. P.; Serre, M. L. 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.* **2009**, *43*, 3736–3742.

(150) Money, E. S.; Sackett, D. K.; Aday, D. D.; Serre, M. L. Using river distance and existing hydrography data can improve the geostatistical estimation of fish tissue mercury at unsampled locations. *Environ. Sci. Technol.* **2011**, *45*, 7746–7753.

(151) Wilkes, G.; Brassard, J.; Edge, T. A.; Gannon, V.; Gottschall, N.; Jokinen, C. C.; Jones, T. H.; Khan, I. U. H.; Marti, R.; Sunohara, M. D.; et al. Long-term monitoring of waterborne pathogens and microbial source tracking markers in paired agricultural watersheds under controlled and conventional tile drainage management. *Appl. Environ. Microbiol.* **2014**, *80*, 3708–3720.

(152) Erb, K.; Ronk, E.; Koundinya, V.; Luczaj, J. Groundwater Quality Changes in a Karst Aquifer of Northeastern Wisconsin, USA: Reduction of Brown Water Incidence and Bacterial Contamination Resulting from Implementation of Regional Task Force Recommendations. *Resources* **2015**, *4*, 655–672.

(153) Diston, D.; Robbi, R.; Baumgartner, A.; Felleisen, R. Microbial source tracking in highly vulnerable karst drinking water resources. *J Water Health* **2017**, wh2017215.

(154) Curriero, F. C. On the Use of Non-Euclidean Distance Measures in Geostatistics. *Math Geol* **2007**, *38*, 907–926.

(155) Little, L. S.; Edwards, D.; Porter, D. E. Kriging in estuaries: as the crow flies, or as the fish swims? *J. Exp. Mar. Bio. Ecol.* **1997**, *213*, 1–11.

(156) Ahmed, W.; Staley, C.; Kaiser, T.; Sadowsky, M. J.; Kozak, S.; Beale, D.; Simpson, S. Decay of sewage-associated bacterial communities in fresh and marine environmental waters and sediment. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 7159–7170.

(157) Brooks, L. E.; Field, K. G. Global model fitting to compare survival curves for faecal indicator bacteria and ruminant-associated genetic markers. *J. Appl. Microbiol.* **2017**, *122*, 1704–1713.

(158) Reischer, G. H.; Ebdon, J. E.; Bauer, J. M.; Schuster, N.; Ahmed, W.; Aström, J.; Blanch, A. R.; Blöschl, G.; Byamukama, D.; Coakley, T.; et al. Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* **2013**, *47*, 8548–8556.

(159) Ichim, I.; Radoane, M. Channel sediment variability along a river: A case study of the Siret River (Romania). *Earth Surf. Process. Landforms* **1990**, *15*, 211–225.

(160) Ann, V.; Freixa, A.; Butturini, A.; Romaní, A. M. Interplay between sediment properties and stream flow conditions influences surface sediment organic matter and microbial biomass in a Mediterranean river. *Hydrobiologia* **2019**, *828*, 199–212.

(161) Fluke, J.; González-Pinzón, R.; Thomson, B. Riverbed Sediments Control the Spatiotemporal Variability of E. coli in a Highly Managed, Arid River. *Front. Water* **2019**, *1*.

(162) Bodmer, P.; Wilkinson, J.; Lorke, A. Sediment properties drive spatial variability of potential methane production and oxidation in small streams. *J. Geophys. Res. Biogeosci.* **2020**, *125*.

(163) Marti, R.; Ribun, S.; Aubin, J.-B.; Colinon, C.; Petit, S.; Marjolet, L.; Gourmelon, M.; Schmitt, L.; Breil, P.; Cottet, M.; et al. Human-Driven Microbiological Contamination of Benthic and Hyporheic Sediments of an Intermittent Peri-Urban River Assessed from MST and 16S rRNA Genetic Structure Analyses. *Front. Microbiol.* **2017**, *8*, 19.

(164) Allen, H. K.; Donato, J.; Wang, H. H.; Cloud-Hansen, K. A.; Davies, J.; Handelsman, J. Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **2010**, *8*, 251–259.

(165) Singer, R. S.; Ward, M. P.; Maldonado, G. Can landscape ecology untangle the complexity of antibiotic resistance? *Nat. Rev. Microbiol.* **2006**, *4*, 943–952.

(166) Uuemaa, E.; Palliser, C.; Hughes, A.; Tanner, C. Effectiveness of a natural headwater wetland for reducing agricultural nitrogen loads. *Water (Basel)* **2018**, *10*, 287.

(167) Adegoke, A. A.; Stenstrom, T.-A. Septic Systems. In *Water and sanitation for the 21st century: health and microbiological aspects of excreta and wastewater management (global water pathogen project)*; Michigan State University, 2019.

(168) Schneeberger, C. L.; O'Driscoll, M.; Humphrey, C.; Henry, K.; Deal, N.; Seiber, K.; Hill, V. R.; Zarate-Bermudez, M. Fate and transport of enteric microbes from septic systems in a coastal watershed. *J Environ Health* **2015**, *77*, 22–30.

(169) Verhougstraete, M. P.; Martin, S. L.; Kendall, A. D.; Hyndman, D. W.; Rose, J. B. Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10419–10424.

(170) McLellan, S. L.; Sauer, E. P.; Corsi, S. R.; Bootsma, M. J.; Boehm, A. B.; Spencer, S. K.; Borchardt, M. A. Sewage loading and microbial risk in urban waters of the Great Lakes. *Elementa (Wash D C)* **2018**, *6*.

(171) Sauer, E. P.; Vandewalle, J. L.; Bootsma, M. J.; McLellan, S. L. Detection of the human specific Bacteroides genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. *Water Res.* **2011**, *45*, 4081–4091.

(172) Oun, A.; Kumar, A.; Harrigan, T.; Angelakis, A.; Xagoraraki, I. Effects of biosolids and manure application on microbial water quality in rural areas in the US. *Water (Basel)* **2014**, *6*, 3701–3723.

(173) Alford, J. B.; Debbage, K. G.; Mallin, M. A.; Liu, Z.-J. Surface water quality and landscape gradients in the north carolina cape fear river basin: the key role of fecal coliform. *Southeast Geogr* **2016**, *56*, 428–453.

(174) Gentry-Shields, J.; Rowny, J. G.; Stewart, J. R. HuBac and nifH source tracking markers display a relationship to land-use but not rainfall. *Water Res.* **2012**, *46*, 6163–6174.

(175) Hinojosa, J.; Green, J.; Estrada, F.; Herrera, J.; Mata, T.; Phan, D.; Pasha, A. B. M. T.; Matta, A.; Johnson, D.; Kapoor, V. Determining the primary sources of fecal pollution using microbial source tracking assays combined with land-use information in the Edwards Aquifer. *Water Res.* **2020**, *184*, 116211.

(176) Zhang, Y.; Kelly, W. R.; Panno, S. V.; Liu, W.-T. Tracing fecal pollution sources in karst groundwater by Bacteroidales genetic biomarkers, bacterial indicators, and environmental variables. *Sci. Total Environ.* **2014**, *490*, 1082–1090.

(177) WDNR Runoff Management - CAFO Permittees https://dnr.wi.gov/topic/AgBusiness/data/CAFO/cafo_sum.asp (accessed Apr 18, 2021).

(178) Van Horn, H. H.; Wilkie, A. C.; Powers, W. J.; Nordstedt, R. A. Components of dairy manure management systems. *J. Dairy Sci.* **1994**, *77*, 2008–2030.

(179) Winter Park Wraps Up Season with High Attendance https://www.visitkewauneecounty.com/news/winter-park-wraps-up-season-with-high-attendance-2021-03-15/ (accessed Apr 18, 2021).

(180) Groundwater Collaboration Workgroup Final Report.

(181) Chapter; TANKS, S.; CHAMBERS, P.; INTERCEPTORS, G.; BEDS, S.; PITS, S.; TRENCHES, S.; PRIVIES; PORTABLE. Summary of Wastewater Land Application in Kewaunee County.

(182) McGinnis, S.; Spencer, S.; Firnstahl, A.; Stokdyk, J.; Borchardt, M.; McCarthy, D. T.; Murphy, H. M. Human Bacteroides and total coliforms as indicators of recent combined sewer overflows and rain events in urban creeks. *Sci. Total Environ.* **2018**, *630*, 967–976.

(183) Hart, J. D.; Blackwood, A. D.; Noble, R. T. Examining coastal dynamics and recreational water quality by quantifying multiple sewage specific markers in a North Carolina estuary. *Sci. Total Environ.* **2020**, *747*, 141124.

(184) Dormann, C. F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J. R. G.; Gruber, B.; Lafourcade, B.; Leitão, P. J.; et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46.

(185) Stumpf, C. H.; Piehler, M. F.; Thompson, S.; Noble, R. T. Loading of fecal indicator bacteria in North Carolina tidal creek headwaters: hydrographic patterns and terrestrial runoff relationships. *Water Res.* **2010**, *44*, 4704–4715.

(186) Riedel, T. E.; Thulsiraj, V.; Zimmer-Faust, A. G.; Dagit, R.; Krug, J.; Hanley, K. T.; Adamek, K.; Ebentier, D. L.; Torres, R.; Cobian, U.; et al. Long-term monitoring of molecular markers can distinguish different seasonal patterns of fecal indicating bacteria sources. *Water Res.* **2015**, *71*, 227–243.

(187) Sansalone, J. J.; Cristina, C. M. First flush concepts for suspended and dissolved solids in small impervious watersheds. *J. Environ. Eng.* **2004**, *130*, 1301–1314.

(188) BERTRAND-KRAJEWSKI, J.-L.; CHEBBO, G.; SAGET, A. DISTRIBUTION OF POLLUTANT MASS VS VOLUME IN STORMWATER DISCHARGES AND THE FIRST FLUSH PHENOMENON. *Water Res.* **1996**.

(189) Frey, S. K.; Topp, E.; Ball, B. R.; Edwards, M.; Gottschall, N.; Sunohara, M.; Zoski, E.; Lapen, D. R. Tile Drainage Management Influences on Surface-Water and Groundwater Quality following Liquid Manure Application. *J Environ Qual* **2013**, *42*, 881–892.

(190) Blaustein, R. A.; Pachepsky, Y. A.; Shelton, D. R.; Hill, R. L. Release and Removal of Microorganisms from Land-Deposited Animal Waste and Animal Manures: A Review of Data and Models. *J Environ Qual* **2015**, *44*, 1338–1354.

(191) Sharma, M.; Millner, P. D.; Hashem, F.; Vinyard, B. T.; East, C. L.; Handy, E. T.; White, K.; Stonebraker, R.; Cotton, C. P. E. coli survival duration in manure-amended soils is affected by spatiotemporal, agricultural, and weather factors: a multi-season, multi-site field study in the Mid-Atlantic U.S. *Appl. Environ. Microbiol.* **2018**.

(192) Bucci, J. P.; Shattuck, M. D.; Aytur, S. A.; Carey, R.; McDowell, W. H. A case study characterizing animal fecal sources in surface water using a mitochondrial DNA marker. *Environ. Monit. Assess.* **2017**, *189*, 406.

(193) Oladeinde, A.; Bohrmann, T.; Wong, K.; Purucker, S. T.; Bradshaw, K.; Brown, R.; Snyder, B.; Molina, M. Decay of fecal indicator bacterial populations and bovine-associated source-tracking markers in freshly deposited cow pats. *Appl. Environ. Microbiol.* **2014**, *80*, 110–118.

(194) Korajkic, A.; McMinn, B. R.; Ashbolt, N. J.; Sivaganesan, M.; Harwood, V. J.; Shanks, O. C. Extended persistence of general and cattle-associated fecal indicators in marine and freshwater environment. *Sci. Total Environ.* **2019**, *650*, 1292–1302.

(195) Wagner, K. L.; Gentry, T. J.; Harmel, R. D.; Pope, E. C.; Redmon, L. A. Grazing Effects on Bovine-Associated and Background Fecal Indicator Bacteria Levels in Edge-of-Field Runoff. *Water (Basel)* **2021**, *13*, 928.

(196) Murphy, S.; Jordan, P.; Mellander, P. E.; O' Flaherty, V. Quantifying faecal indicator organism hydrological transfer pathways and phases in agricultural catchments. *Sci. Total Environ.* **2015**, *520*, 286–299.

(197) Rasmuson, A.; Erickson, B.; Borchardt, M.; Muldoon, M.; Johnson, W. P. Pathogen Prevalence in Fractured versus Granular Aquifers and the Role of Forward Flow Stagnation Zones on Pore-Scale Delivery to Surfaces. *Environ. Sci. Technol.* **2020**, *54*, 137–145.

(198) Pronk, M.; Goldscheider, N.; Zopfi, J. Particle-size distribution as indicator for fecal bacteria contamination of drinking water from karst springs. *Environ. Sci. Technol.* **2007**, *41*, 8400–8405.

(199) Lamendella, R.; Domingo, J. W. S.; Oerther, D. B.; Vogel, J. R.; Stoeckel, D. M. Assessment of fecal pollution sources in a small northern-plains watershed using PCR and phylogenetic analyses of Bacteroidetes 16S rRNA gene. *FEMS Microbiol. Ecol.* **2007**, *59*, 651–660.

(200) Frey, S. K.; Gottschall, N.; Wilkes, G.; Grégoire, D. S.; Topp, E.; Pintar, K. D. M.; Sunohara, M.; Marti, R.; Lapen, D. R. Rainfall-induced runoff from exposed streambed sediments: an important source of water pollution. *J Environ Qual* **2015**, *44*, 236–247.

(201) Joy, S. R.; Bartelt-Hunt, S. L.; Snow, D. D.; Gilley, J. E.; Woodbury, B. L.; Parker, D. B.; Marx, D. B.; Li, X. Fate and transport of antimicrobials and antimicrobial resistance genes in soil and runoff following land application of swine manure slurry. *Environ. Sci. Technol.* **2013**, *47*, 12081–12088.

(202) Schilling, K. E.; Wolter, C. F.; Isenhart, T. M.; Schultz, R. C. Tile drainage density reduces groundwater travel times and compromises riparian buffer effectiveness. *J Environ Qual* **2015**, *44*, 1754–1763.

(203) Buckerfield, S. J.; Quilliam, R. S.; Waldron, S.; Naylor, L. A.; Li, S.; Oliver, D. M. Rainfall-driven E. coli transfer to the stream-conduit network observed through increasing spatial scales in mixed land-use paddy farming karst terrain. *Water Research X* **2019**, *5*, 100038.

(204) de Brauwere, A.; Ouattara, N. K.; Servais, P. Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review. *Crit. Rev. Environ. Sci. Technol.* **2014**, *44*, 2380–2453.

(205) Samarajeewa, A. D.; Glasauer, S. M.; Lauzon, J. D.; O’Halloran, I. P.; Parkin, G. W.; Dunfield, K. E. Bacterial contamination of tile drainage water and shallow groundwater under different application methods of liquid swine manure. *Can. J. Microbiol.* **2012**, *58*, 668–677.

(206) Hatvani, I. G.; Kirschner, A. K. T.; Farnleitner, A. H.; Tanos, P.; Herzig, A. Hotspots and main drivers of fecal pollution in Neusiedler See, a large shallow lake in Central Europe. *Environ. Sci. Pollut. Res. Int.* **2018**, *25*, 28884–28898.

(207) Tyrrel, S. F.; Quinton, J. N. Overland flow transport of pathogens from agricultural land receiving faecal wastes. *J. Appl. Microbiol.* **2003**, *94 Suppl*, 87S–93S.

(208) Ahmed, W.; Zhang, Q.; Kozak, S.; Beale, D.; Gyawali, P.; Sadowsky, M. J.; Simpson, S. Comparative decay of sewage-associated marker genes in beach water and sediment in a subtropical region. *Water Res.* **2019**, *149*, 511–521.

(209) Zimmer-Faust, A. G.; Thulsiraj, V.; Marambio-Jones, C.; Cao, Y.; Griffith, J. F.; Holden, P. A.; Jay, J. A. Effect of freshwater sediment characteristics on the persistence of fecal indicator bacteria and genetic markers within a Southern California watershed. *Water Res.* **2017**, *119*, 1–11.

(210) Dick, L. K.; Stelzer, E. A.; Bertke, E. E.; Fong, D. L.; Stoeckel, D. M. Relative decay of Bacteroidales microbial source tracking markers and cultivated Escherichia coli in freshwater microcosms. *Appl. Environ. Microbiol.* **2010**, *76*, 3255–3262.

(211) Eichmiller, J. J.; Borchert, A. J.; Sadowsky, M. J.; Hicks, R. E. Decay of genetic markers for fecal bacterial indicators and pathogens in sand from Lake Superior. *Water Res.* **2014**, *59*, 99–111.

(212) Rogers, S. W.; Donnelly, M.; Peed, L.; Kelty, C. A.; Mondal, S.; Zhong, Z.; Shanks, O. C. Decay of bacterial pathogens, fecal indicators, and real-time quantitative PCR genetic markers in manure-amended soils. *Appl. Environ. Microbiol.* **2011**, *77*, 4839–4848.

(213) Martín-Díaz, J.; García-Aljaro, C.; Pascual-Benito, M.; Galofré, B.; Blanch, A. R.; Lucena, F. Microcosms for evaluating microbial indicator persistence and mobilization in fluvial sediments during rainfall events. *Water Res.* **2017**, *123*, 623–631.

(214) Droppo, I. G.; Krishnappan, B. G.; Liss, S. N.; Marvin, C.; Biberhofer, J. Modelling sediment-microbial dynamics in the South Nation River, Ontario, Canada: Towards the prediction of aquatic and human health risk. *Water Res.* **2011**, *45*, 3797–3809.

(215) Williams-Nguyen, J.; Sallach, J. B.; Bartelt-Hunt, S.; Boxall, A. B.; Durso, L. M.; McLain, J. E.; Singer, R. S.; Snow, D. D.; Zilles, J. L. Antibiotics and antibiotic resistance in agroecosystems: state of the science. *J Environ Qual* **2016**, *45*, 394–406.

(216) The Characteristics, Causes, and Consequences of Sprawling Development Patterns in the United States | Learn Science at Scitable https://www.nature.com/scitable/knowledge/library/the-characteristics-causes-and-consequences-of-sprawling-103014747/ (accessed Apr 17, 2021).

(217) Davies, J.; Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 417–433.

(218) Amarasiri, M.; Sano, D.; Suzuki, S. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. *Crit. Rev. Environ. Sci. Technol.* **2019**, 1–44.

(219) Pepper, I. L.; Brooks, J. P.; Gerba, C. P. Antibiotic resistant bacteria in municipal wastes: is there reason for concern? *Environ. Sci. Technol.* **2018**, *52*, 3949–3959.

(220) Yang, L.; Liu, W.; Zhu, D.; Hou, J.; Ma, T.; Wu, L.; Zhu, Y.; Christie, P. Application of biosolids drives the diversity of antibiotic resistance genes in soil and lettuce at harvest. *Soil Biol. Biochem.* **2018**, *122*, 131–140.

(221) Munir, M.; Xagoraraki, I. Levels of antibiotic resistance genes in manure, biosolids, and fertilized soil. *Journal of Environment Quality* **2011**, *40*, 248.

(222) Duarte, D. J.; Oldenkamp, R.; Ragas, A. M. J. Modelling environmental antibiotic-resistance gene abundance: A meta-analysis. *Sci. Total Environ.* **2019**, *659*, 335–341.

(223) Amábile-Cuevas, C. *Antibiotics and antibiotic resistance in the environment*; CRC Press, 2015.

(224) Pol, M.; Ruegg, P. L. Treatment practices and quantification of antimicrobial drug usage in conventional and organic dairy farms in Wisconsin. *J. Dairy Sci.* **2007**, *90*, 249–261.

(225) Barkovskii, A. L.; Babb, C. M.; Hurley, D.; Shin, E. Origins and environmental mobility of antibiotic resistance genes, virulence factors and bacteria in a tidal creek's watershed. *J. Appl. Microbiol.* **2015**, *118*, 764–776.

(226) Ahmed, W.; Zhang, Q.; Lobos, A.; Senkbeil, J.; Sadowsky, M. J.; Harwood, V. J.; Saeidi, N.; Marinoni, O.; Ishii, S. Precipitation influences pathogenic bacteria and antibiotic resistance gene abundance in storm drain outfalls in coastal sub-tropical waters. *Environ. Int.* **2018**, *116*, 308–318.

(227) Jang, J.; Kim, M.; Baek, S.; Shin, J.; Shin, J.; Shin, S. G.; Kim, Y. M.; Cho, K. H. Hydrometeorological Influence on Antibiotic-Resistance Genes (ARGs) and Bacterial Community at a Recreational Beach in Korea. *J. Hazard. Mater.* **2021**, *403*, 123599.

(228) Srinivasan, V.; Nam, H.-M.; Sawant, A. A.; Headrick, S. I.; Nguyen, L. T.; Oliver, S. P. Distribution of tetracycline and streptomycin resistance genes and class 1 integrons in Enterobacteriaceae isolated from dairy and nondairy farm soils. *Microb. Ecol.* **2008**, *55*, 184–193.

(229) Kyselková, M.; Jirout, J.; Vrchotová, N.; Schmitt, H.; Elhottová, D. Spread of tetracycline resistance genes at a conventional dairy farm. *Front. Microbiol.* **2015**, *6*, 536.

(230) Xu, M.; Wang, F.; Sheng, H.; Stedtfeld, R. D.; Li, Z.; Hashsham, S. A.; Jiang, X.; Tiedje, J. M. Does anaerobic condition play a more positive role in dissipation of antibiotic resistance genes in soil? *Sci. Total Environ.* **2021**, *757*, 143737.

(231) Dowling, P. M. Pharmacologic Principles. In *Equine Internal Medicine*; Elsevier, 2004; pp. 169–233.

(232) Pal, C.; Bengtsson-Palme, J.; Kristiansson, E.; Larsson, D. G. J. Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* **2015**, *16*, 964.

(233) Antibiotic Use in the United States, 2017: Progress and Opportunities | Antibiotic Use | CDC https://www.cdc.gov/antibiotic-use/stewardship-report/2017.html (accessed Apr 20, 2021).

(234) Cummings, D. E.; Archer, K. F.; Arriola, D. J.; Baker, P. A.; Faucett, K. G.; Laroya, J. B.; Pfeil, K. L.; Ryan, C. R.; Ryan, K. R. U.; Zuill, D. E. Broad dissemination of plasmid-mediated quinolone resistance genes in sediments of two urban coastal wetlands. *Environ. Sci. Technol.* **2011**, *45*, 447–454.

(235) Zalewska, M.; Błażejewska, A.; Czapko, A.; Popowska, M. Antibiotics and antibiotic resistance genes in animal manure – consequences of its application in agriculture. *Front. Microbiol.* **2021**, *12*.

(236) Schrag, N. F. D.; Godden, S. M.; Apley, M. D.; Singer, R. S.; Lubbers, B. V. Antimicrobial use quantification in adult dairy cows - Part 3 - Use measured by standardized regimens and grams on 29 dairies in the United States. *Zoonoses Public Health* **2020**, *67 Suppl 1*, 82–93.

(237) Hall, M. C.; Mware, N. A.; Gilley, J. E.; Bartelt-Hunt, S. L.; Snow, D. D.; Schmidt, A. M.; Eskridge, K. M.; Li, X. Influence of setback distance on antibiotics and antibiotic resistance genes in runoff and soil following the land application of swine manure slurry. *Environ. Sci. Technol.* **2020**, *54*, 4800–4809.

(238) Miller, D. N.; Jurgens, M. E.; Durso, L. M.; Schmidt, A. M. Simulated winter incubation of soil with swine manure differentially affects multiple antimicrobial resistance elements. *Front. Microbiol.* **2020**, *11*, 611912.

(239) Peng, S.; Feng, Y.; Wang, Y.; Guo, X.; Chu, H.; Lin, X. Prevalence of antibiotic resistance genes in soils after continually applied with different manure for 30 years. *J. Hazard. Mater.* **2017**, *340*, 16–25.

(240) Wepking, C.; Avera, B.; Badgley, B.; Barrett, J. E.; Franklin, J.; Knowlton, K. F.; Ray, P. P.; Smitherman, C.; Strickland, M. S. Exposure to dairy manure leads to greater antibiotic resistance and increased mass-specific respiration in soil microbial communities. *Proc. Biol. Sci.* **2017**, *284*.

(241) Guo, T.; Lou, C.; Zhai, W.; Tang, X.; Hashmi, M. Z.; Murtaza, R.; Li, Y.; Liu, X.; Xu, J. Increased occurrence of heavy metals, antibiotics and resistance genes in surface soil after long-term application of manure. *Sci. Total Environ.* **2018**, *635*, 995–1003.

(242) Murray, R.; Tien, Y.-C.; Scott, A.; Topp, E. The impact of municipal sewage sludge stabilization processes on the abundance, field persistence, and transmission of antibiotic resistant bacteria and antibiotic resistance genes to vegetables at harvest. *Sci. Total Environ.* **2019**, *651*, 1680–1687.

(243) Brooks, J. P.; Maxwell, S. L.; Rensing, C.; Gerba, C. P.; Pepper, I. L. Occurrence of antibiotic-resistant bacteria and endotoxin associated with the land application of biosolids. *Can. J. Microbiol.* **2007**, *53*, 616–622.

(244) Mohapatra, D. P.; Cledón, M.; Brar, S. K.; Surampalli, R. Y. Application of wastewater and biosolids in soil: occurrence and fate of emerging contaminants. *Water. Air. Soil Pollut.* **2016**, *227*, 77.

(245) Tanner, B. D.; Brooks, J. P.; Gerba, C. P.; Haas, C. N.; Josephson, K. L.; Pepper, I. L. Estimated occupational risk from bioaerosols generated during land application of class B biosolids. *J Environ Qual* **2008**, *37*, 2311–2321.

(246) Gottschall, N.; Edwards, M.; Topp, E.; Bolton, P.; Payne, M.; Curnoe, W. E.; Coelho, B. B.; Lapen, D. R. Nitrogen, phosphorus, and bacteria tile and groundwater quality following direct injection of dewatered municipal biosolids into soil. *J Environ Qual* **2009**, *38*, 1066–1075.

(247) Edwards, M.; Topp, E.; Metcalfe, C. D.; Li, H.; Gottschall, N.; Bolton, P.; Curnoe, W.; Payne, M.; Beck, A.; Kleywegt, S.; et al. Pharmaceutical and personal care products in tile drainage following surface spreading and injection of dewatered municipal biosolids to an agricultural field. *Sci. Total Environ.* **2009**, *407*, 4220–4230.

(248) Zerzghi, H.; Gerba, C. P.; Brooks, J. P.; Pepper, I. L. Long-term effects of land application of class B biosolids on the soil microbial populations, pathogens, and activity. *J Environ Qual* **2010**, *39*, 402–408.

(249) Pepper, I. L.; Brooks, J. P.; Gerba, C. P. Land application of organic residuals: municipal biosolids and animal manures. In *Environmental and pollution science*; Elsevier, 2019; pp. 419–434.

(250) Esseili, M. A.; Kassem, I. I.; Sigler, V.; Czajkowski, K.; Ames, A. Genetic evidence for the offsite transport of E. coli associated with land application of Class B biosolids on agricultural fields. *Sci. Total Environ.* **2012**, *433*, 273–280.

(251) Lapen, D. R.; Topp, E.; Metcalfe, C. D.; Li, H.; Edwards, M.; Gottschall, N.; Bolton, P.; Curnoe, W.; Payne, M.; Beck, A. Pharmaceutical and personal care products in tile drainage following land application of municipal biosolids. *Sci. Total Environ.* **2008**, *399*, 50–65.

(252) Walker, J.; Knight, L.; Stein, L. Plain english guide to the EPA part 503 biosolids rule. **1994**.

(253) Lowman, A.; McDonald, M. A.; Wing, S.; Muhammad, N. Land application of treated sewage sludge: community health and environmental justice. *Environ. Health Perspect.* **2013**, *121*, 537–542.

(254) Greger, M.; Koneswaran, G. The public health impacts of concentrated animal feeding operations on local communities. *Fam. Community Health* **2010**, *33*, 11–20.

(255) Ma, Y.; Wilson, C. A.; Novak, J. T.; Riffat, R.; Aynur, S.; Murthy, S.; Pruden, A. Effect of various sludge digestion conditions on sulfonamide, macrolide, and tetracycline resistance genes and class I integrons. *Environ. Sci. Technol.* **2011**, *45*, 7855–7861.

(256) Kinney, C. A.; Furlong, E. T.; Zaugg, S. D.; Burkhard, M. R.; Werner, S. L.; Cahill, J. D.; Jorgensen, G. R. Survey of organic wastewater contaminants in biosolids destined for land application. *Environ. Sci. Technol.* **2006**, *40*, 7207–7215.

(257) Fang, H.; Zhang, Q.; Nie, X.; Chen, B.; Xiao, Y.; Zhou, Q.; Liao, W.; Liang, X. Occurrence and elimination of antibiotic resistance genes in a long-term operation integrated surface flow constructed wetland. *Chemosphere* **2017**, *173*, 99–106.

(258) Xiang, S.; Wang, X.; Ma, W.; Liu, X.; Zhang, B.; Huang, F.; Liu, F.; Guan, X. Response of microbial communities of karst river water to antibiotics and microbial source tracking for antibiotics. *Sci. Total Environ.* **2020**, *706*, 135730.

(259) Stange, C.; Tiehm, A. Occurrence of antibiotic resistance genes and microbial source tracking markers in the water of a karst spring in Germany. *Sci. Total Environ.* **2020**, *742*, 140529.

(260) Byrd, R. H.; Gilbert, J. C.; Nocedal, J. A trust region method based on interior point techniques for nonlinear programming. *Math. Program.* **2000**, *89*, 149–185.

(261) NR 243.14. *Wis. Admin. Code, Animal Feeding Operations*; 2007.

(262) Corsi, S. R.; Borchardt, M. A.; Spencer, S. K.; Hughes, P. E.; Baldwin, A. K. Human and bovine viruses in the Milwaukee River watershed: hydrologically relevant representation and relations with environmental variables. *Sci. Total Environ.* **2014**, *490*, 849–860.

(263) Howard, K. J.; Martin, E.; Gentry, T.; Feagley, S.; Karthikeyan, R. Effects of Dairy Manure Management Practices on E. coli Concentration and Diversity. *Water. Air. Soil Pollut.* **2017**, *228*, 4.