

Microbial Find, Inform, and Test Model for Identifying Spatially Distributed Contamination Sources: Framework Foundation and Demonstration of Ruminant *Bacteroides* Abundance in River Sediments

Corinne Wiesner-Friedman, Rachelle E. Beattie, Jill R. Stewart, Krassimira R. Hristova, and Marc L. Serre*



Cite This: *Environ. Sci. Technol.* 2021, 55, 10451–10461



Read Online

ACCESS |



Metrics & More



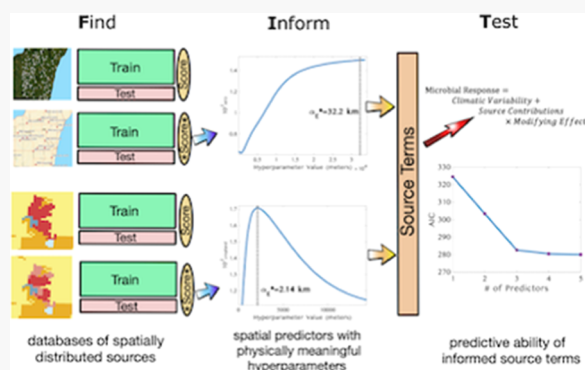
Article Recommendations



Supporting Information

ABSTRACT: Microbial pollution in rivers poses known ecological and health risks, yet causal and mechanistic linkages to sources remain difficult to establish. Host-associated microbial source tracking (MST) markers help to assess the microbial risks by linking hosts to contamination but do not identify the source locations. Land-use regression (LUR) models have been used to screen the source locations using spatial predictors but could be improved by characterizing transport (i.e., hauling, decay overland, and downstream). We introduce the microbial Find, Inform, and Test (FIT) framework, which expands previous LUR approaches and develops novel spatial predictor models to characterize the transported contributions. We applied FIT to characterize the sources of *BoBac*, a ruminant *Bacteroides* MST marker, quantified in riverbed sediment samples from Kewaunee County, Wisconsin. A 1 standard deviation increase in contributions from land-applied manure hauled 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) in the sediment. This is the first work finding an association between the upstream land-applied manure and the offsite bovine-associated fecal markers. These findings have implications for the sediment as a reservoir for microbial pollution associated with AFOs (e.g., pathogens and antibiotic-resistant bacteria). This framework and application advance statistical analysis in MST and water quality modeling more broadly.

KEYWORDS: land-use regression, molecular microbial source tracking, ground hauling manure, river networks



1. 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. Advances in host-associated microbial source tracking (MST) approaches help to directly identify the hosts contributing to pollution.^{1–3} To more directly characterize the health risks, pollution contributions from sources must be modeled. Mechanistic approaches are the gold standard when sources and flow processes (e.g., adsorption, advection, and diffusion) are accurately identified. However, the sources of emerging microbial pollutants^{4,5} and delivery parameters to characterize the microbial fate and transport have not been well characterized.⁶ 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 the microbial fate and transport from sources, as described by Pruden et al.⁷ in a conceptual mass-balance model for antimicrobial resistance responses. While this type of regression is frequently used for nutrient and chemical pollutants,^{8–11} 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.^{7,12–21} Many of these studies do not distinguish

Received: March 10, 2021

Revised: July 4, 2021

Accepted: July 6, 2021

Published: July 22, 2021



ACS Publications

© 2021 American Chemical Society

10451

<https://doi.org/10.1021/acs.est.1c01602>
Environ. Sci. Technol. 2021, 55, 10451–10461

between statistical exploration and physically meaningful models.^{14–17,20,21} The physically meaningful statistical models distinguish between sources and modifiers (i.e., attenuating or amplifying effects).⁷ 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,^{14–16,20} depictions of concentrations of contaminants would increase moving away from sources [e.g., animal feeding operations (AFOs) reduce the 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 a modifier of microbial contamination. Exploratory spatial-statistical models often cannot provide the evidence required for recommendations to environmental and public health agencies.

Physically meaningful LUR models (i.e., distinguish sources from modifiers) have been developed and implemented for chemical pollutants in groundwater, such as nitrate, where observed concentrations, y_i , for sample i were expressed as the following non-linear multivariate regression equation²²

$$y_i = \beta_0 + \left\{ \sum_{u=1}^U \beta_u s_i^{(u)}(\alpha^{(u)}) \right\} \exp \left\{ \sum_{l=1}^L \beta_l m_i^{(l)}(\alpha^{(l)}) \right\} + \varepsilon_i \quad (1)$$

where β_0 is the intercept, β_u 's are the positive linear regression coefficients for the *source terms*, $s_i^{(u)}(\alpha^{(u)})$ (e.g., manure application fields), that increase microbial contamination, and β_l 's are the non-linear regression coefficients for the *modifier terms* $m_i^{(l)}(\alpha^{(l)})$ (e.g., wetlands, forested areas, and buffers) that amplify or attenuate pollution. The *modifier terms* are exponential, so when β_l 's are positive or negative, the sum of contributions from sources is multiplied or fractionated, respectively, corresponding to amplification or attenuation. $\alpha^{(u)}$'s and $\alpha^{(l)}$'s are hyperparameters that express the spatial scales at which a source or a modifier affects the 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 *Escherichia coli* have utilized databases derived from remote sensing,¹⁹ national land-cover databases,¹⁶ and a database derived from interviews with local agricultural farmers and managers.¹² 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 technique to select between comparable databases.

After choosing the databases, *spatial predictor models* (SPMs) are used to construct the source terms. Microbial LUR models have captured the source density with studies that use lumped-source terms (e.g., percent cropland-cover-per-watershed and cattle density-per-county).^{13,15–18,20,23} The upstream capacities of sources were first captured with an inverse-distance-

weighted interpolation method,⁷ which constructs source terms at a higher resolution on the river network, but contributions from sources can increase without a source's influence (see [Supporting Information S1](#) for a detailed explanation of this issue). The sum of exponentially decaying contributions (SEDCs) from the upstream sources addresses this issue.^{19,21} This class of SPM utilizes interpretable hyperparameters,^{11,22,24} which helps to depict the gradual effect of proximity and density of sources.⁹

Additionally, the SEDC models do not require mechanistic understanding of transport or initial loadings at sources. Such is the case for the current tools^{6,12,25–29} that predict bacterial and nutrient pollution in rivers, such as the Soil and Water Assessment Tool (SWAT) or the SPATIally Referenced Regressions on Watershed attributes (SPARROW) model.^{26–29} Additionally, the finest spatial resolution these water quality models offer is at the sub-watershed level. However, these models show that dilution and climatic processes are critical in the variability of microbial contamination,^{7,12,19,21,23,25,27,30–35} but these processes have not been incorporated into a unique statistical approach that estimates at a fine resolution (eq 1).

Another critical component in the fate and transport of microbial pollution is ground hauling of waste in trucks from its origin (e.g., AFOs) to release points into the environment (e.g., manure application fields).^{36,37} To our knowledge, no existing SPM (e.g., SEDC) accounts for hauling for estimating microbial responses. Additionally, no published work has compared the predictive power of different SEDC SPMs for microbial contamination responses. See [Table S1](#) 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 1) for microbial responses in river networks. We develop new SPMs for source terms that do not require prior knowledge of physical processes by using hyperparameters to define the spatial scale associated with ground hauling, overland, and river distance decay and microbial pollution flow. We also formalize the 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.⁷ have used regression models to quantify the strength of the association between livestock source terms and the relative abundance of *antibiotic resistance genes* in river sediments. To our knowledge, no work has used regression models to quantify the association of source terms and the relative abundance of a *ruminant-specific MST marker* in river sediments. Given sediments' capacity to store chemical pollutants³⁸ and bacterial genes^{39,40} for a long term, the magnitude of this association has implications to environmental and human health.^{41–43} Additionally, water column samples may not capture intermittent pollution from sources,

while sediment samples provide time-integrated information.⁴⁴ Here, we applied FIT to learn about bovine sources of fecal contamination in river sediments from Kewaunee County, Wisconsin. A novel SPM that accounts for the gradual effect of proximity and density of upstream 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 [Supporting Information S3](#) for details) of ruminant *Bacteroides* abundance in river sediment.

2. MATERIALS AND METHODS

2.1. Materials and Methods for the Demonstration of the Novel Microbial LUR Framework.

2.1.1. Study Area and Sampling Data. Kewaunee County, Wisconsin, was selected as the study area due to the local water quality concerns⁴⁵ and the high density of AFOs compared to the nearby counties. 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 were further described in Beattie et al.⁴⁶

2.1.2. 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 coordinates were projected into "Wisconsin Central" State Plane.

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 the six closest weather stations (Brussels, Denmark WWTP, Forestville, Green Bay, Kewaunee, and New Franken). See [Supporting Information S4](#) for river network and climatic data details.

2.1.3. Spatial Databases of Potential Microbial Contamination Sources. There were two types of potential microbial contamination sources for which spatial databases were obtained. 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 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 Wisland-2 land-cover database.⁴⁷ See [Supporting Information S5](#) and Table S3 for details.

2.1.4. Microbial Analysis of Samples: Sample Processing, DNA Extraction, and Quantitative PCR. Sediment grab samples were collected ($n = 90$) as described in Beattie et al.,⁴⁶ immediately placed on ice following collection, and stored at $-20\text{ }^{\circ}\text{C}$ until DNA extraction. The sediment samples were homogenized and subsampled for two replicate DNA extractions per sample. DNA was extracted from 0.5 g of sediment as described in Beattie et al.⁴⁶

FAM TaqMan probe chemistry was used for quantitative polymerase chain reaction (qPCR) assays of bovine-specific *Bacteroides* 16S rRNA genes,⁴⁸ 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). The TaqMan qPCR reaction mixtures contained 10 μL of TaqMan Gene Expression Master Mix (Applied Biosystems, Foster City, CA), 1 μL each of 10 nM F/R primers, 0.4 μL of 5 nM probe, 2.6 μL of H_2O , and 5 μL 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. The PCR amplification protocols consisted of $50\text{ }^{\circ}\text{C}$ for 2 min, followed by $95\text{ }^{\circ}\text{C}$ for 10 min, and 40 cycles of $95\text{ }^{\circ}\text{C}$ for 30 s, and an annealing temperature of $57\text{ }^{\circ}\text{C}$ for 45 s. SYBR qPCR reaction mixtures and cycling parameters are as previously described.⁴⁶ Below detect values were set to one-half of the detection limit.⁴⁸ Primers, annealing temperatures, and inhibition assay details are given in [Supporting Information S6](#).

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

$$\begin{cases} z_i = \frac{\text{The absolute abundance of BoBac (gene copies)}}{16\text{S rRNA abundance (gene copies)}} \\ y_i = \log_{10}(z_i) \end{cases} \quad (2)$$

See [Supporting Information S7](#) for details on absolute versus relative abundance.

2.2. Developing the Novel, Microbial LUR Framework. **2.2.1. Physically Meaningful Model for Contamination of Spatially Distributed Sources.** We introduce a microbial LUR model that formalizes previous microbial models and expands the non-linear LUR from [eq 1](#) to include climatic terms previously developed for a range of microbiological responses^{10,18,49–51}

$$y_i = \beta_0 + P1_i(\beta_1 + \beta_2 P2_i) + \beta_3 \text{freezing}_i + \left\{ \sum_{u=1}^U \beta_u s_i^{(u)}(\alpha^{(u)}) \right\} \exp \left\{ \sum_{l=1}^L \beta_l m_i^{(l)}(\alpha^{(l)}) \right\} + \varepsilon_i \quad (3)$$

The observed response value y_i ([eq 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](#) becomes linear. In subsequent papers, we will add more sources and modifiers.

The source terms, $s_i^{(u)}(\alpha^{(u)})$, are standardized to a mean of 0 and a standard deviation of 1, so that a 1 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, $\text{RAR}^{(u)}$, as the ratio of relative abundances for a 1 standard deviation increase in source u , which is given by $\text{RAR}^{(u)} = z(s^{(u)} + 1)/z(s^{(u)}) = 10^{\beta_u}$ (see [Supporting Information S8](#)). It follows that a 1 standard deviation increase in the source term $s^{(u)}$ corresponds to a $(10^{\beta_u} - 1) \times 100$ 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

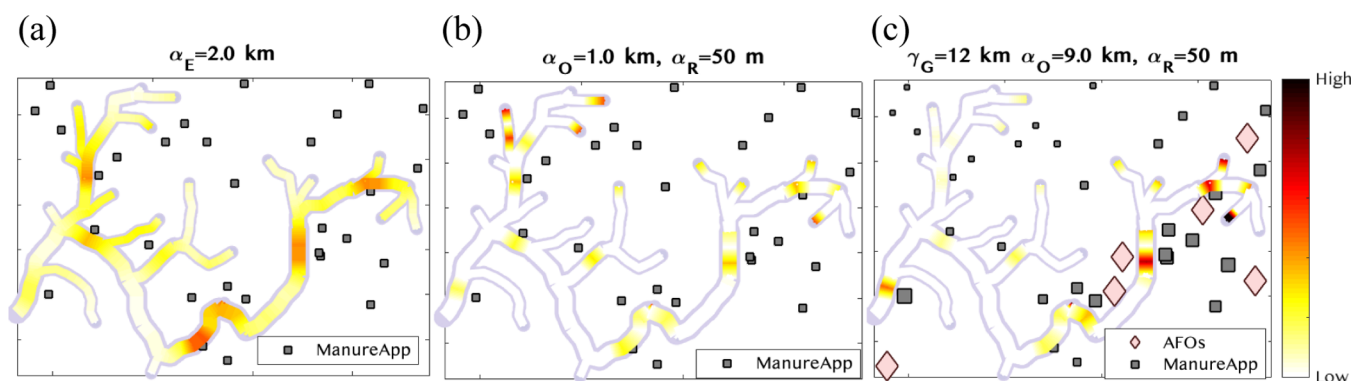


Figure 1. Diagrams that demonstrate (a) Euclidean, (b) ORF, and (c) GORF SEDC in a downstream combining the river network with examples of hyperparameter values.

antecedent precipitation, P2, on P1.⁵⁰ P1 and P2 at time t are calculated as SEDCs from the precipitations $P_{it'}$ at times $t' < t$ and location i as

$$P1_i = \sum_{t' < t} P_{it'} \exp\left(\frac{-3|t - t'|}{\alpha_{p1}}\right) \quad (4)$$

$$P2_i = \sum_{t' < t} P_{it'} \exp\left(\frac{-3|t - t'|}{\alpha_{p2}}\right) \quad (5)$$

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 washout effects (see Supporting Information S9 for details).

2.2.2. Conventional and Novel SPMs Describing Spatially Distributed Contamination Source Contributions. We construct the source predictors $s_i^{(u)}(\alpha^{(u)})$ using three SPMs of increasing sophistication: (1) a Euclidean model, (2) an overland and river distance flow (ORF) model, and (3) a ground hauling, overland, and river distance flow (GORF) model. The first approach uses the Euclidean distance.²² The value of $s_i^{(u)}(\alpha^{(u)})$ at location i is calculated as the standardized SEDC²² from the surrounding sources j

$$s_i^{(u)}(\alpha^{(u)}) = s_i^{(u)}(\alpha_E^{(u)}) = z - \text{score} \left(\sum_{j=0}^N C_{0j} e^{-3D_{ij}^{(E)}/\alpha_E} \right) \quad (6)$$

where $D_{ij}^{(E)}$ is the Euclidean distance between the 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, $s_i^{(u)}(\alpha_E^{(u)})$, is shown in Figure 1a using this Euclidean approach when the range of influence, $\alpha_E^{(u)}$, is 2.0 km.

The sources in Figure 1a 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, sources located next to the river network contribute to the nearby river segments both downstream and upstream, which oppose first-order principles and studies indicating higher levels of microbial contamination downstream of sources than upstream.^{52–60}

Accordingly, in the second approach, we use the 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,⁶¹ it is diluted within the flow Q . The Strahler stream order is used as a proxy for flow in sediment.⁶² The standardized ORF SPM is

$$\begin{aligned} s_i^{(u)}(\alpha^{(u)}) &= s_i^{(u)}(\alpha_O^{(u)}, \alpha_R^{(u)}) \\ &= z - \text{score} \left(\sum_{j=0}^N \frac{M_{0j}}{Q_i} e^{-3D_{ij}^{(O)}/\alpha_O - 3D_{ij}^{(R)}/\alpha_R} \delta_{ij} \right) \end{aligned} \quad (7)$$

where δ_{ij} is a hydrologic connectivity indicator equal to 1 if sampling location i receives the 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.^{19,61,63–66} An example of ORF-modeled contributions is shown in Figure 1b 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 the hydrologic processes involved in transporting microbial contaminants from where they are applied to downstream sampling locations. However, it does not capture how the mass M_{0j} applied at location j is hauled (i.e., via trucking or irrigation) from AFOs.^{37,67,68}

To address this latter point, we introduce the GORF model, where the mass applied at application field j is calculated based on the amounts that are ground-hauled from the 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 the size of that AFO) and ω_{kj} is the proportion of M_k ground-transported to the application field j . We assume that ω_{kj} is proportional to an exponential decrease with the distance D_{kj} between the AFO k and the application field j , that is, $\omega_{kj} \propto \exp\left(\frac{-3D_{kj}}{\gamma_G}\right)$, where the distance hyperparameter γ_G is a transportation range reflecting how far the manure is hauled. The ground hauling distances, D_{kj} , are expressed with Euclidean distances, which are good proxies for road network path distances (see Supporting Information S10). The model assumes that the application of manure onto the fields decreases exponentially at a Euclidean distance closer to AFOs because of the exponentially decreasing hauling capacity and increasing costs to AFOs over hauling distances estimated by Hadrach et al.³⁷ We also assume that there is no

loss of manure during hauling, so that mass is conserved, that is, $\sum_{j=1}^J \omega_{kj} = 1$. Hence, the full equation of the standardized GORF SPM is

$$s_i^{(u)}(\alpha^{(u)}) = s_i^{(u)}(\gamma_G^{(u)}, \alpha_O^{(u)}, \alpha_R^{(u)}) \\ = z - \text{score} \left(\frac{1}{Q_i} \sum_{j=0}^N \sum_{k=1}^K \omega_{kj} M_k e^{-3D_{ij}^{(O)}/\alpha_O + -3D_{ij}^{(R)}/\alpha_R \delta_{ij}^-} \right), \\ \text{where } \omega_{kj} = \frac{\exp\left(\frac{-3D_{kj}}{\gamma_G}\right)}{\sum_{j=1}^J \exp\left(\frac{-3D_{kj}}{\gamma_G}\right)} \quad (8)$$

An example of the GORF model is shown in Figure 1c 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 concentrations downstream of manure application fields that are near AFOs.

2.2.3. Optimizing Source Term Hyperparameters in the FIT Framework. An approach is needed to obtain hyperparameters, $\alpha^{(u)}$, of the source terms $s_i^{(u)}(\alpha^{(u)})$ in the microbial regression eq 3. In the Find and Inform steps of the FIT framework, we consider each source term individually and construct the source term by obtaining the hyperparameters that maximize the RAR (10^{β_u}) for some training set of observed outcomes y_i (see Supporting Information S11 for details). In the test step of the FIT framework, we use a variable selection of source and climatic terms to define a model and test whether terms are statistically significant.

Previous works have selected hyperparameters that maximize the R^2 .²² However, this can lead to physically invalid results (e.g., a source term may reduce pollution). Our FIT framework addresses this issue by emphasizing the physical meaningfulness and reliability over statistical prediction. The FIT steps of the FIT framework are detailed in the next three sections. See Supporting Information S12 for graphical summary of FIT and visual representation of this application.

2.3. Stages of the Novel FIT Framework. **2.3.1. Find Reliable Databases of Spatially Distributed Sources.** We develop a reliability score as a criterion for finding a database d of a given source u that is selected based on which most reliably produces a positive $\beta_u > 0$ for the source term $s_i^{(u)}(\alpha^{(u)})$ defined in eq 3. We start with model a corresponding to the microbial regression (eq 3) reduced to an intercept and the source term $s_i^{(u)}(\alpha^{(u)})$. 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 $\hat{\alpha}_{\text{training}_k}^{(u)}$ which maximizes $\beta_u^{(u)}$ (see Supporting Information S11 for details). We adopt $\hat{\alpha}_{\text{training}_k}^{(u)}$ for the test set regressions to obtain $\beta_u^{(\text{test}_k)}$. We define the sign stability score SSS as a number between 1 and $k + 1$ ($k + 1 = 6$), which quantifies how many of the $\beta_u^{(\text{test}_k)}$ values are positive. We assess the components of reliability through taking the sum of $\beta_u^{(\text{test}_k)}$ across the folds to represent an average of test set coefficients, M . Last, we define $\sigma_{\beta_u^{(\text{test}_k)}}$ as the standard deviation of the $\beta_u^{(\text{test}_k)}$ values. We reward high SSS, high M , and low $\sigma_{\beta_u^{(\text{test}_k)}}$ values by setting the reliability score RS equal to the ratio of SSS multiplied by M over σ_{β_u} (see Supporting Information S13 for details). Finally, we select the database d of source type u with the highest reliability score.

2.3.2. Inform Spatial Predictors with Hyperparameters. We inform each u^{th} source term individually by obtaining the

hyperparameter $\hat{\alpha}(u)$ which maximizes β_u using 100% of the observations y_i in model a . This ensures that for each SPM (Euclidean, ORF, and GORF), we inform the source term with all the data in hand. The maximization procedure is the same as in the “Find” stage and described in detail in Supporting Information S11. For the “Inform” stage, a penalty (eq S13) was added to the objective function so as not to obtain a combination of γ_G and α_O yielding poor regression or mapping qualities (i.e., non-normal residuals or $\hat{y}_i > 0$).

2.3.3. 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). An Akaike information criterion exploits a stepwise algorithm to select the informed source and climatic terms—a standard practice in microbial LUR studies.^{12,13,15,17,19,21} However, when datasets have a high dimension of climatic and source terms, LASSO may be preferred to limit collinearity, selection bias, and overfitting. Overall, a possible outcome of the selection procedure is that no source terms are selected. To compare modeling options (i.e., Euclidean, ORF, and GORF), a 6-fold cross-validation of the “Test” step of the FIT framework was performed. See Supporting Information S14 for details.

3. RESULTS AND DISCUSSION

3.1. 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 databases representing source locations. The criterion is based on the SSS, M , and σ_{β_u} respectively, capturing the consistency, effect, and variability of contributing source locations across the 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 imagery insufficiently captured sources and could not replace permitting information. Last, compared to the crop rotation land-cover database, the dairy rotation subclass more often represented a source of bovine fecal contamination reliably.

We also report on the reliability of SPM approaches. GORF terms 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 hauling of manure was essential to consistently capturing the contributing effects of AFOs to sediment bovine markers. See Supporting Information S15 for details.

3.2. Informing of Spatial Predictor Variables. At the Inform stage of the FIT framework using the GORF SPM, 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). Potential explanations include possible long-range overland contamination due to subsurface tile drainage⁶⁹ and karst geography.^{70,71} Due to this long-range hauling and overland contamination, we find that all 15 Kewaunee County CAFOs are implicated with fecal contamination in sediments. Finally, downstream contamination of riverbed sediments from overland discharge points occurs over sub-kilometer distances (i.e., $\alpha_R < 1$ km),

Table 1. Regression Model Results of Bovine Fecal Contamination in Sediments (log 10 BoBac-Copies-per-16S-rRNA-Copies) Shown for Each SPM Used for Source Terms (i.e., Euclidean, ORF, and GORF) and the Stepwise-Selected Model Performance Expressed as R^2 and Adjusted R^2 ^a

model performance	standardized explanatory variables	hyperparameters α	regression coefficient β	relative abundance ratio 10^β (95% CI)
source terms SPM: Euclidean				
$R^2 = 0.415$	P1	$\alpha_{P1} = 0.177$ days	0.995 ^b	9.90 (4.19, 23.4)
	P1 $\times$ P2	$\alpha_{P2} = 2.04$ days	−0.221	0.602 (0.311, 1.16)
	freezing	n/a	−3.31 ^b	4.94×10^{-4} (7.06×10^{-5} , 0.00347)
adjusted $R^2 = 0.395$	AFOs	$\alpha_E = 32.4$ km	not selected	not selected
	manure fields	$\alpha_E = 1.51$ km	not selected	not selected
source terms SPM: overland and river distance—flow (ORF)				
$R^2 = 0.428$	P1	$\alpha_{P1} = 0.177$ days	1.00 ^b	10.1 (4.28, 23.6)
	P1 $\times$ P2	$\alpha_{P2} = 2.04$ days	−0.222	0.600 (0.312, 1.16)
	freezing	n/a	−3.29 ^b	5.04×10^{-4} (7.27×10^{-5} , 0.00349)
adjusted $R^2 = 0.401$	AFOs	$\alpha_O = 2.58$ km, $\alpha_R = 308$ m	not selected	not selected
	manure fields	$\alpha_O = 935$ m, $\alpha_R = 54.0$ m	0.168	1.47 (0.858, 2.53)
source terms SPM: ground hauling, overland and river distance—flow (GORF)				
$R^2 = 0.442$	P1	$\alpha_{P1} = 0.177$ days	0.988 ^b	9.90 (4.18, 22.6)
	P1 $\times$ P2	$\alpha_{P2} = 2.04$ days	−0.220	0.602 (0.360, 1.01)
	freezing	n/a	−3.25 ^b	5.62×10^{-4} (8.30×10^{-5} , 0.00380)
adjusted $R^2 = 0.417$	AFOs (via ground hauling to manure fields)	$\gamma_G = 11.3$ km, $\alpha_O = 8.94$ km, $\alpha_R = 50.5$ m	0.248 ^b	1.77 (1.04, 3.02)

^aFor each SPM, standardized explanatory variables, their physically meaningful hyperparameters, α , their regression coefficients, β s, and resulting relative abundance ratio, 10^β , are provided row-wise. ^b p -value < 0.05.

suggesting localized sediment contamination. See [Supporting Information S16](#) for details.

3.3. Testing the Predictive Ability of Source Terms.

Results of the Test stage of the FIT framework are shown in [Table 1](#). The first column defines the SPM used to construct source terms and summarizes model performance with R^2 . The results of the inform stage of FIT are shown in the column-labeled hyperparameters. All climatic terms were selected (see [Supporting Information S17](#)). Each source term has hyperparameters even if they were not selected. The test stage results appear in the last two columns, which show the standardized regression coefficients and the corresponding RAR.

Among the models resulting from different spatial predictors, only the GORF source term indicated that fecal contamination in sediment was *significantly* associated with AFOs via manure application. The selection of this GORF predictor, AFO-ground hauling to manure fields, suggests that flow and hauling are transport mechanisms for bovine fecal contamination in sediment. The GORF approach performed slightly better in a 6-fold cross-validation (see [Supporting Information S18](#)). Some microbial LUR and geostatistical studies have found that river distance measures are more predictive and generalizable than Euclidean for modeling contamination in a river network.^{19,66,72,73} However, this work is the first to suggest hauling as a key transport process. Hauling, which is a critical aspect of nutrient management from CAFOs,^{37,68,74} determines where manure is applied and subsequently determines locations on the river network that are more impacted by AFOs.

We focus on the Test stage results resulting from the GORF SPM due to its statistical significance. The finding is that a 1 standard deviation increase in the contributions of AFOs via land application of manure hauled to fields is 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 sediments. No study of sediment exists to which this value can be compared. However, in a surface water study of absolute abundance of fecal contamination (log 10 gene copies not normalized by 16S rRNA), Dila et al. found that a 1 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.¹⁸ For comparison, we also applied FIT to the absolute abundance of bovine fecal contamination (log 10 *BoBac*-copies-per-gram-sediment). A 1 standard deviation increase in manure application from AFOs was associated with a 79% (p -value < 0.05) increase in bovine gene (*BoBac*) copies (see [Supporting Information S19](#) 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 association between bovine sources and the relative abundance of a bovine-specific fecal marker in the downstream sediment. Though a positive association has been observed between bovine sources and fecal contamination *in water* by Dila,¹⁸ researchers should be careful in comparing the strength of these associations. First, the two MST markers (i.e., *BoBac* in our study, *BacR* in Dila's study)¹⁸ may degrade at different rates.^{75,76} Additionally, inhibition or other factors may affect true-positive rates in the sediment. While the *BacR* marker has been reported to have a higher source specificity, *BoBac* has offered a higher source sensitivity⁷⁷—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.

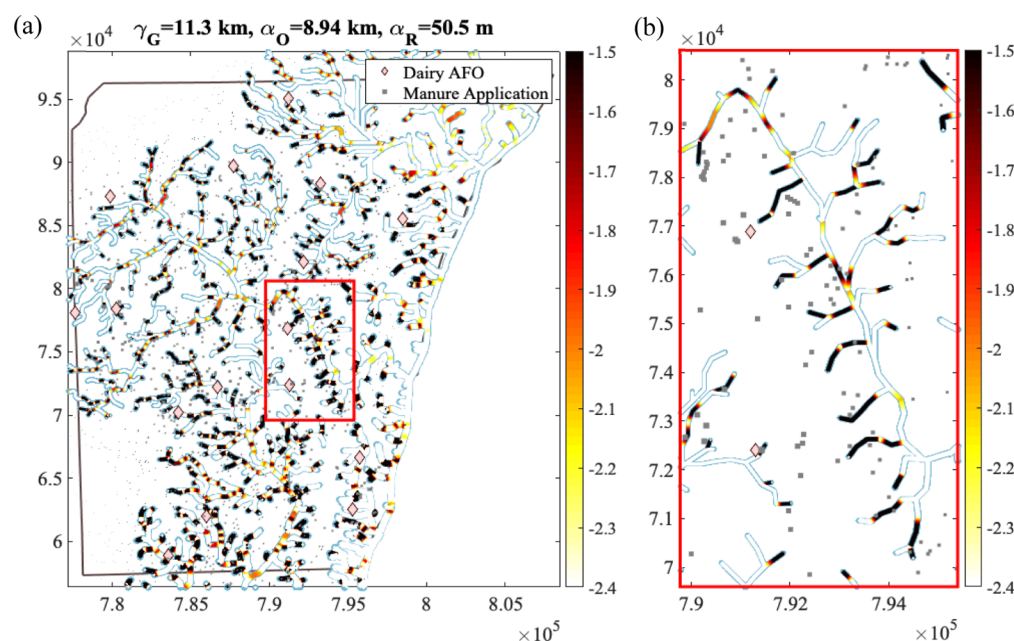


Figure 2. Modeled relative abundance ($\log_{10}$ *BoBac*-copies-per-16S-rRNA-copies) plotted on the river network for Kewaunee County given average recent and antecedent precipitation and freezing temperature. (a) View for all of Kewaunee County rivers and streams associated with the Kewaunee, Ahnapee, and East Twin Rivers. (b) Zoom into the portion of (a) highlighted by the red rectangle. The Find stage results for AFO and manure application fields are depicted by the red diamonds (WPDES CAFOs that are unweighted) and the gray squares (land cover database of dairy rotation).

Other factors contributing to variability in the estimate of this association include sediment depth, size, composition, and proximity to geological and other anthropogenic factors.^{78–82} For example, fine-course, silty sediments have been correlated with higher microbial concentrations and bacterial growth.^{79–81} In one study, while host-associated fecal markers observed in sediments 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 the sediment and below.⁸² These factors indicate opportunities for well-maintained spatial databases of sediment types and precise sampling approaches that would not be possible for surface water due to its more transient nature.

3.4. Implications, Limitations, and Future Work. To our knowledge, this is the first LUR modeling of the abundance of a bovine-associated MST marker in river sediments. Our primary finding is that ground hauling of bovine manure from AFOs and applying onto fields contribute to riverbed sediment fecal microbial abundance. This study's fecal abundance response, specific to ruminants and highly associated with bovine hosts,⁴⁸ 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.⁴² Sediment as a reservoir of AFO-associated fecal contamination is particularly concerning because AFOs frequently use antibiotics to treat and prevent disease, increasing the risk of elevated antimicrobial resistance in these contaminated sediments.⁸³ A wider application of FIT to contaminants from AFOs (e.g., antimicrobials, antibiotic resistant bacteria, and pathogens) may help untangle the impacts of AFOs on antibiotic resistance in rivers.

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 sediments in a river network (Figure 2). Figure 2a was created by applying the model resulting from the Test stage to all flow-connected points of the Kewaunee, Ahnapee, or East Twin Rivers using 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 2 depicts Inform stage results, which combined and expanded previous SPMs, such that it was the first to account for the gradual effect of proximity and density of upstream sources, dilution due to flow, and ground hauling (i.e., GORF SPM) of microbial contamination from sources. Each of the GORF processes can be seen in greater detail in Figure 2b. The proximity of the AFOs (viz., red diamonds) to manure field centroids (viz., gray squares) increases the amount applied on manure fields. Similarly, the proximity and density of highly applied manure fields to the river network increase the fecal contamination levels in sediments (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 microbial LUR models. This model explained 44.2% of the variability, which is within the range of previous LUR models of fecal contamination.^{7,13,15–18,20,21} Unstudied source types may cause some unexplained variability. We observed that the measured values at the river sampling site abbreviated BPKR⁴⁶ 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.³⁸ Sediment reservoirs around dams, therefore, have the potential to act as amplifiers of microbial contamination in sediments. Values at 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,^{22,84} and constructed wetlands can attenuate microbial pollution.^{85,86} 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 fine-scale modeling approaches, increased sampling size to inform and validate the models, and well-maintained spatial databases that characterize geology, land use and management, and land cover. In the future, the FIT framework could help identify key sources of emerging microbial pollutants and be instructive to assessing the microbial risk associated with river water use.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.1c01602>.

Literature review of SPMs used in LUR for microbial responses, review of performance evaluation for the existing models of microbial contamination, gaps in modeling non-human sources of microbial contamination, processing river network and climatic data, processing spatial data to create candidate databases of microbial contamination sources, bovine *Bacteroides* primer sets and inhibition assay, relative versus absolute abundance, relative abundance ratio, selecting the hyperparameters of the precipitation predictors P1 and P2, details on the ground hauling ranges from the GORF SPM, selecting the hyperparameters for the contamination source terms using constrained optimization, graphic summary and overview of each stage of the FIT framework and application, details of the reliability score calculation, details on performance statistics and cross-validation, details of the reliability score results, details of the hyperparameter results and statistics for source terms, details on the predictive ability of climatic variables, details on cross-validation results, and results of the FIT analysis for the absolute abundance of sediment *BoBac* (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marc L. Serre – Gillings School of Global Public Health, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7400, United States; Phone: (919) 966-7014; Email: marc_serre@unc.edu; https://sph.unc.edu/adv_profile/marc-serre-phd

Authors

Corinne Wiesner-Friedman – Gillings School of Global Public Health, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7400, United States;

orcid.org/0000-0002-5845-1969

Rachelle E. Beattie – Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin 53233, United States

Jill R. Stewart – Gillings School of Global Public Health, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7400, United States; orcid.org/0000-0002-3474-5233

Krassimira R. Hristova – Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin 53233, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.est.1c01602>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

C.W.-F. was supported by a grant from the National Institute of Environmental Health Sciences (NIEHS) T32ES007018. This research was funded in part by the Marquette University Innovation Grant (KH) and by NSF grant 1316318 as part of the joint NSF-NIH-USA Ecology and Evolution of Infectious Diseases program.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the Kewaunee County Department of Land and Water Conservation and the Wisconsin Department of Natural Resources for their assistance in obtaining useful spatial databases.

■ ABBREVIATIONS

AFO animal feeding operation
BoBac Ruminant *Bacteroides* host marker
CAFO concentrated animal feeding operation
ORF overland and river distance flow
GORF ground hauling, overland, and river distance flow
LUR land-use regression
RAR relative abundance ratio
SSS sign stability score

■ REFERENCES

- (1) 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.
- (2) *Microbial Source Tracking: Methods, Applications, and Case Studies*; Hagedorn, C., Blanch, A. R., Harwood, V. J., Eds.; Springer: New York, NY, 2011.
- (3) Stewart, J. R.; Boehm, A. B.; Dubinsky, E. A.; Fong, T.-T.; Goodwin, K. D.; Griffith, J. F.; Noble, R. T.; Shanks, O. C.; Vijayavel, K.; Weisberg, S. B. Recommendations following a multi-laboratory comparison of microbial source tracking methods. *Water Res.* **2013**, *47*, 6829–6838.

- (4) 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.
- (5) 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.
- (6) 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* **2020**, *12*, 1876.
- (7) Pruden, A.; Arabi, M.; Storteboom, H. N. Correlation between upstream human activities and riverine antibiotic resistance genes. *Environ. Sci. Technol.* **2012**, *46*, 11541–11549.
- (8) 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.
- (9) 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.
- (10) 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.
- (11) 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.
- (12) 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.
- (13) 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.
- (14) Li, X.; Atwill, E. R.; Antaki, E.; Applegate, O.; Bergamaschi, B.; Bond, R. F.; Chase, J.; Ransom, K. M.; Samuels, W.; Watanabe, N.; Harter, T. 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.
- (15) Brendel, C.; Soupier, M. Relating Watershed Characteristics to Elevated Stream *Escherichia coli* Levels in Agriculturally Dominated Landscapes: An Iowa Case Study. *Water* **2017**, *9*, 154.
- (16) 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.
- (17) Saeidi, N.; Gu, X.; Tran, N. H.; Goh, S. G.; Kitajima, M.; Kushmaro, A.; Schmitz, B. W.; Gin, K. Y. Occurrence of Traditional and Alternative Fecal Indicators in Tropical Urban Environments under Different Land Use Patterns. *Appl. Environ. Microbiol.* **2018**, *84*, No. e00287-18.
- (18) 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.
- (19) Christenson, E. 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.
- (20) 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.
- (21) Amos, G. C.; 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.
- (22) 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.
- (23) 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.
- (24) 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.
- (25) 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.
- (26) Wang, Y.; Jiang, R.; Xie, J.; Zhao, Y.; Yan, D.; Yang, S. Soil and water assessment tool (SWAT) model: A systemic review. *J. Coast Res.* **2019**, *93*, 22.
- (27) 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.
- (28) 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.
- (29) Costa, C. M. d. 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.
- (30) 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.
- (31) Ballesté, E.; Belanche-Muñoz, L. A.; Farnleitner, A. H.; Linke, R.; Sommer, R.; Santos, R.; Monteiro, S.; Maunula, L.; Oristo, S.; Tiehm, A. A. 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.* **2020**, *171*, 115392.
- (32) 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.
- (33) Zivkovich, B. R.; Mays, D. C. Predicting nonpoint stormwater runoff quality from land use. *PLoS One* **2018**, *13*, No. e0196782.
- (34) Hohl, A.; Václavík, T.; Meentemeyer, R. K. Go with the flow: geospatial analytics to quantify hydrologic landscape connectivity for passively dispersed microorganisms. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1626–1641.
- (35) 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.
- (36) Seltentrich, N. Manure irrigation: environmental benefits, potential human health risks. *Environ. Health Perspect.* **2017**, *125*, 124004.
- (37) Hadrich, J. C.; Harrigan, T. M.; Wolf, C. A. Economic comparison of liquid manure transport and land application. *Appl. Eng. Agric.* **2010**, *26*, 743–758.
- (38) Wohl, E. Legacy effects on sediments in river corridors. *Earth-Sci. Rev.* **2015**, *147*, 30–53.
- (39) 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.

- (40) 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.
- (41) 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.
- (42) 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.
- (43) 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.
- (44) Kasich, J.; Taylor, M.; Nally, S. J. *Sediment Sampling Guide and Methodologies*; Ohio Environmental Protection Agency, 2012.
- (45) Wright, K.; Peck, J. E.; Kewaunee County Citizens. *Industrial Dairies in Kewaunee County, Wisconsin; Socially Responsible Agriculture Project*, 2015.
- (46) 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*, fty131.
- (47) 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).
- (48) Layton, A.; McKay, L.; Williams, D.; Garrett, V.; Gentry, R.; Saylor, 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.
- (49) 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.
- (50) Holcomb, D. A.; Messier, K. P.; Serre, M. L.; Rowney, 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.
- (51) 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.
- (52) 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.
- (53) 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 Neglected Trop. Dis.* **2014**, *8*, No. e3186.
- (54) 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.
- (55) 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.
- (56) 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.
- (57) 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. *Agric. Water Manag.* **2018**, *200*, 10–18.
- (58) Amos, G. C. A.; Ploumaki, 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.
- (59) 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.
- (60) 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.
- (61) 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*, 2537–2549.
- (62) Hughes, R. M.; Kaufmann, P. R.; Weber, M. H. National and regional comparisons between Strahler order and stream size. *J. North Am. Benthol. Soc.* **2011**, *30*, 103–121.
- (63) 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.
- (64) 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.
- (65) 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.
- (66) 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.
- (67) Powell, J. M.; McCrory, D. F.; Jackson-Smith, D. B.; Saam, H. Manure collection and distribution on Wisconsin dairy farms. *J. Environ. Qual.* **2005**, *34*, 2036–2044.
- (68) Paudel, K. P.; Bhattarai, K.; Gauthier, W. M.; Hall, L. M. Geographic information systems (GIS) based model of dairy manure transportation and application with environmental quality consideration. *Waste Manag.* **2009**, *29*, 1634–1643.
- (69) 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.; Topp, E.; Lapen, D. R. 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.
- (70) 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.
- (71) Diston, D.; Robbi, R.; Baumgartner, A.; Felleisen, R. Microbial source tracking in highly vulnerable karst drinking water resources. *J. Water Health* **2017**, *16*, 138.
- (72) Curriero, F. C. On the Use of Non-Euclidean Distance Measures in Geostatistics. *Math. Geol.* **2007**, *38*, 907–926.
- (73) Little, L. S.; Edwards, D.; Porter, D. E. Kriging in estuaries: as the crow flies, or as the fish swims? *J. Exp. Mar. Biol. Ecol.* **1997**, *213*, 1–11.
- (74) Powell, J. M.; Jackson-Smith, D. B.; McCrory, D. F.; Saam, H.; Mariola, M. Nutrient management behavior on wisconsin dairy farms. *Agron. J.* **2007**, *99*, 211–219.
- (75) 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.

(76) 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.

(77) Reischer, G. H.; Ebdon, J. E.; Bauer, J. M.; Schuster, N.; Ahmed, W.; Åström, J.; Blanch, A. R.; Blöschl, G.; Byamukama, D.; Coakley, T.; Ferguson, C.; Goshu, G.; Ko, G.; de Roda Husman, A. M.; Mushi, D.; Poma, R.; Pradhan, B.; Rajal, V.; Schade, M. A.; Sommer, R.; Taylor, H.; Toth, E. M.; Vrajmasu, V.; Wuertz, S.; Mach, R. L.; Farnleitner, A. H. 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.

(78) 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.

(79) 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.

(80) 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*, 4.

(81) 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*, No. e2019JG005213.

(82) Marti, R.; Ribun, S.; Aubin, J.-B.; Colinet, C.; Petit, S.; Marjolet, L.; Gourmelon, M.; Schmitt, L.; Breil, P.; Cottet, M.; Cournoyer, B. 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.

(83) 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. U.S.A.* **2015**, *112*, 5649–5654.

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

(85) 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.

(86) 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.