Environmental Science & Technology

pubs.acs.org/est Article

Characterizing Differences in Sources of and Contributions to Fecal Contamination of Sediment and Surface Water with the Microbial FIT Framework

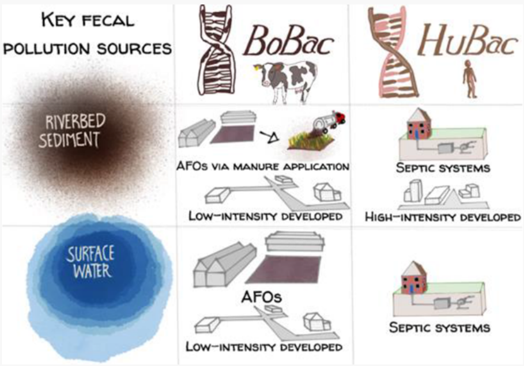
Corinne Wiesner-Friedman, Rachelle E. Beattie, Jill R. Stewart, Krassimira R. Hristova, and Marc L. Serre*

Cite This: *Environ. Sci. Technol.* 2022, 56, 4231−4240

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Surface water monitoring and microbial source tracking (MST) are used to identify host sources of fecal pollution and protect public health. However, knowledge of the locations of spatial sources and their relative impacts on the environment is needed to effectively mitigate health risks. Additionally, sediment samples may offer time-integrated information compared to transient surface water. Thus, we implemented the newly developed microbial find, inform, and test framework to identify spatial sources and their impacts on human (*HuBac*) and bovine (*BoBac*) MST markers, quantified from both riverbed sediment and surface water in a bovine-dense region. Dairy feeding operations and low-intensity developed land-cover were associated with 99% (p-value < 0.05) and 108% (p-value < 0.05) increases, respectively, in the relative abundance of *BoBac* in sediment, and with 79% (p-value < 0.05) and 39% increases in surface water. Septic systems were associated with a 48% increase in the relative abundance of *HuBac* in sediment and a 56% increase in surface water. Stronger source signals were observed for sediment responses compared to water. By defining source locations, predicting river impacts, and estimating source influence ranges in a Great Lakes region, this work informs pollution mitigation strategies of local and global significance.



KEYWORDS: *land-use regression, molecular microbial source tracking, surface water, sediment, river networks, animal feeding operations, septic systems*

1. INTRODUCTION

Fecal pollution in waterways is a public health hazard, especially in regions that economically depend on clean water for agriculture, recreation, and other direct uses. Effective remediation strategies and accurate risk analyses depend on identifying the human and animal hosts contributing to fecal pollution.[1−3] Microbial source tracking (MST) methods are under development to identify animal hosts of fecal contamination, usually by quantifying host-associated genetic markers. However, quantifying the concentration of host-associated markers does not characterize their spatial sources (hereafter referred to as sources) upstream. To fully characterize and provide valuable information for mitigating public health risks, knowledge of concentration, source locations, and transport of host-associated markers are needed. Spatial modeling studies on example watersheds can fulfill this need by revealing source associations at a fine scale and generalizing knowledge of their impacts on fecal contamination.

Researchers have begun to use land-use regression (LUR) frameworks to find associations between spatial predictors describing sources and MST markers.[4,5] In particular, the spatial predictor model (SPM) known as the sum of exponentially decaying contributions (SEDCs) constructs source terms by utilizing interpretable hyperparameters that characterize transport with models of distance decay ranges around sources.[6] Our recent work developed SEDC spatial predictors to characterize overland and downstream distance decay from sources and account for dilution due to flow.[7]

Animal feeding operations (AFOs),[5,8−10] septic systems,[11−16] sewer lines to wastewater treatment plants, and combined sewer overflow events frequently contribute to fecal contamination in rivers.[4,17−19] Previous research has emphasized these sources more than other potential sources and land-cover. Fewer studies have examined the capacity for fate and transport of fecal contaminants from land-applied biosolids and wastewater from

Received: January 10, 2022
Revised: February 25, 2022
Accepted: March 1, 2022
Published: March 17, 2022

https://doi.org/10.1021/acs.est.2c00224

nonhuman waste.[20] Several studies have found significant correlations between fecal contamination, impervious surfaces, and developed land-cover.[4,21−25]

Most of these studies have sampled either surface water or sediment. To our knowledge, no study has modeled associations with sources and fecal pollution in both sediment and surface water (see Supporting Information, S1 for details). Furthermore, no studies have quantified associations between multiple sources and fecal contamination in riverbed sediment using LUR or other approaches. However, recent studies suggest that long-term storage of pollutants in sediments adds to microbial risks.[26,27] Additionally, this long-term storage may offer time-integrated information about how sources contribute to microbial contamination. Thus, we expect to see stronger associations between fecal responses and modeled sources in sediment versus surface water.

Our work aimed to understand how quantifying microbial responses from sediment in addition to surface water contributes to the knowledge of how sources impact fecal pollution in rivers. With bovine and human-associated fecal markers quantified from surface water and riverbed sediment samples, we modeled impacts from upstream sources by estimating relative abundance ratios (RARs) for these responses with the newly developed microbial Find, Inform, and Test (FIT) framework.[7]

The samples were previously obtained for a 2018 study of a dense-agriculture Great Lakes region where microbial contamination could be attributed to many types of putative sources (e.g., animal agriculture, septic systems, wastewater treatment plants, and different densities of imperviousness).[28] Previous research in this region has modeled spatial associations at a coarse spatial resolution for bovine and human-associated markers in surface water.[5] Here, we aim to bring attention to sediment by modeling differences in fecal pollution source impacts for time-integrated sediment versus transient surface water samples. We can inform risk mitigation strategies by estimating overland influence ranges around polluters and downstream decay with the Overland and River distance with Flow (ORF) SPM,[7] which allows us to predict downstream source impacts at unsampled sediment and surface water locations with a fine spatial resolution. Lastly, we better define databases representing spatially distributed fecal pollution sources in this region.

Knowledge of how sources impact sediment differently than surface water can inform fecal pollution mitigation strategies that reduce environmental health risks and improve river ecosystem functioning. Estimating influence ranges around sources in a watershed with many putative sources can inform these strategies locally and globally. Protecting the Great Lakes water quality is environmentally, economically, and healthfully beneficial due to water use for drinking, recreation, and fisheries.[5]

2. METHODS

2.1. Microbial Contamination Responses Sample Collection, Processing, DNA Extraction, and Quantitative PCR.

Sediment ($n = 90$) and river water (n=98) samples were collected from 20 river locations on five dates between July 2016 and May 2017 in Kewaunee County. 1 L of grab samples was collected using sterile 1 L bottles for surface water. In-depth details on sampling site selection and riverbed sediment sampling methods are described in Beattie et al.[28]

Here, we describe unpublished sampling and processing methods (i.e., surface water analysis for *BoBac*, sediment and surface water analysis for *HuBac*). Sediment and surface water grab samples were processed as described in Beattie et al.[28] For additional sampling details, see Supporting Information, S2. DNA was quantified using QuBit (Thermo Fisher Scientific, Waltham, MA) fluorometric quantitation for each sample. Quality control details can be found in Supporting Information, S3.1.

FAM TaqMan probe chemistry was used to quantitate bovine (*BoBac*) and human-specific (*HuBac*) 16S rRNA genes. The V3 region of the 16S rRNA gene was quantified using SYBR Green chemistry as a proxy for total bacteria. All quantifications were performed in duplicate on a real-time PCR system (Bio Rad CFX Connect) from two replicate DNA extractions per sample and averaged for total gene abundance (four reactions total per sample). Primers and annealing temperatures used in this study originally published in Layton et al.[29] are listed in Table S3. Extraction details in Supporting Information, 3.2. For each qPCR assay, plasmid DNA containing the cloned gene of interest was used to generate a six-point standard curve for each gene analyzed. Standard curves quantified in duplicate were included with each 96-well assay plate. Standard curves and resulting gene quantifications for samples were only accepted if the standard curve r^2 value was >0.980 and the amplification efficiency was between 90 and 110%. For accepted standard curves across all genes, slopes ranged from −3.22 to −3.41 and intercepts ranged from 39.524 to 42.637. The entire sample plate and standard curve were re-assayed when standard curves deviated from these ranges. 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. 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.[28]

In total, there were $r = 4$ types of responses representing two gene markers measured in the sediment and surface water. The absolute abundance of the bovine and human genes, $z_i^{(r)}$, 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, to obtain the relative abundance of the bovine and human genes. Before statistical analysis, these values were log10 transformed to obtain the response, $y_i^{(r)}$, representing the log10 transform of a value greater than 0, but less than 1 $\left(y_i^{(r)} < 0\right)$.

2.2. River Network and Climatic Data.

River network and climatic data were processed identically to Wiesner-Friedman et al.,[7] where a digital elevation model obtained through the Wisconsin Department of Natural Resources (WIDNR) created flow lines to form a directed tree network, and daily precipitation and monthly average temperature data were obtained from the National Oceanic and Atmospheric Administration.[30] Climatic values at sampling sites were inverse-distance-weighted interpolated from station data.[7]

2.3. Databases Representing Sources of Microbial Contamination.

The study area has many potential bovine and human fecal contamination sources and is further vulnerable by its Karst geology.[31,32] For the implementation of FIT, we considered a total of seven types of potential bovine sources with different database options available: AFOs (four options), manure application fields (two options), land-applied sludge (two options), land-applied sludge with industrial waste (three options), wastewater treatment plants treating industrial wastes (one option), low-intensity developed land-cover (two options), and ground hauling of manure from AFO to application fields (eight options).

We considered a total of six types of human sources with different database options available: septic systems (one option), wastewater treatment plants (two options), land-applied sludge (two options), land-applied sludge from municipal wastewater or septage (three options), high-intensity developed land-cover (two options), and the ground hauling of treated wastewater or septage from septic systems locations to land-applied sludge sites (three options). These databases were obtained from personal communication with the Kewaunee County Land and Water Conservation Department, open records requests with the WIDNR, and 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 S4.

2.4. Physically Meaningful Model for Contamination from Spatially Distributed Sources. For associations between responses and source terms, a physically meaningful LUR model[6,7] was considered to characterize source contributions without modifying effects on those contributions.[6] This model was recently expanded to incorporate contributions from climatic terms, which are essential to environmental microbiological response variability.[7] Climatic factors, including precipitation and temperature, have been related to increases in fecal contamination.[4,5,15,24,33,34]

$$y_i^{(r)} = \beta_0 + P1_i(\beta_1 + \beta_2 P2_i) + \beta_3 \text{freezing}_i + \sum_{u=1}^{U} \beta_u s_i^{(u)}(\alpha^{(u)}) + \varepsilon_i \tag{1}$$

The observed response value y_i for sample i is a function of climatic terms and source terms. β_1, β_2, and β_3 are linear regression coefficients associated with the climatic variables $P1_i$ (recent precipitation), $P2_i$ (antecedent precipitation), and Freezingi. 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.[35] The source terms, $s_i^{(u)}(\alpha^{(u)})$, are standardized to a mean of 0 and standard deviation of 1, so that a one-standard-deviation increase in the uth source term leads to a β_u increase in the relative abundance response, y_i. We constructed each source term with the ORF SEDC SPM described further in Wiesner-Friedman et al.[7]

$$s_i^{(u)}(\alpha^{(u)}) = s_i^{(u)}(\alpha_{\mathrm{O}}^{(u)}, \alpha_{\mathrm{R}}^{(u)}) = z\text{ - score}\left(\sum_{j=1}^{J} \frac{M_{0j}}{Q_i} \mathrm{e}^{-3D_{ij}^{(\mathrm{O})}/\alpha_{\mathrm{O}} + -3D_{ij}^{(\mathrm{R})}/\alpha_{\mathrm{R}}} \delta_{\overleftarrow{ij}}\right) \tag{2}$$

An ORF source term for each ith observation is the z-scored SEDC value that is the sum of the mass of contaminants from each jth source, M_{0j}, after decaying overland distances, $D_{ij}^{(\mathrm{O})}$, and river distances, $D_{ij}^{(\mathrm{R})}$, and accounting for flow, Q_i. Hyperparameters, α_{O} and α_{R}, statistically characterize overland and downstream distance decay, respectively. $\delta_{\overleftarrow{ij}}$ indicates the flow connectivity between each jth source location and each ith observation location (i.e., an adjacency matrix denoting the downstream connectivity of any point on the directed river network).

Due to our previous work suggesting that hauling of manure is a crucial process in fecal contamination,[7] we modeled some source terms with the Ground hauling, Overland and River distance with Flow (GORF) SEDC SPM.

$$s_i^{(u)}(\alpha^{(u)}) = s_i^{(u)}(\gamma_{\mathrm{G}}^{(u)}, \alpha_{\mathrm{O}}^{(u)}, \alpha_{\mathrm{R}}^{(u)}) = z\text{ - score}\left(\frac{1}{Q_i}\sum_{j=1}^{J}\sum_{k=1}^{K} \omega_{kj} M_k \mathrm{e}^{-3D_{ij}^{(\mathrm{O})}/\alpha_{\mathrm{O}} + -3D_{ij}^{(\mathrm{R})}/\alpha_{\mathrm{R}}} \delta_{\overleftarrow{ij}}\right)$$

where

$$\omega_{kj} = \exp\left(\frac{-3D_{kj}}{\gamma_{\mathrm{G}}}\right) \Big/ \sum_{j=1}^{J} \exp\left(\frac{-3D_{kj}}{\gamma_{\mathrm{G}}}\right) \tag{3}$$

A GORF source term represents the z-scored SEDC value from land application sources (i.e., manure application fields, land-applied waste) where each ith value is ORF source term from land application sources that are weighted by their proximity to the origin of what is applied to the land (i.e., AFOs, septic systems via treatment plants). The mass of contaminants from each kth origin location, M_k, is distributed across each jth land application source location. This mass depends on the spatial relationships between each kth source location and each jth land application location described by ω_{kj}, and the hyperparameter, γ_{G}, acts as a toggle that determines how far the application occurs from the mass's origin. If γ_{G} is infinite, then the distribution of the mass from the origin becomes homogeneous across application locations.

ORF SPMs were used to construct source terms [see Table S4 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 SPM was used to construct source terms characterizing the contributions that AFOs make to the river network via the ground hauling and manure application 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 nonindustrial waste (u = 6). Additionally, the GORF SPM was used to construct source terms characterizing the contributions that septic systems make to the river network via the ground hauling and application of treated septage at land applied waste sites.

2.5. Implementation of the Microbial FIT Framework. We implemented the FIT framework using the physically meaningful model[7] for contamination from spatially distributed sources (eq 1) to identify and characterize the sources of microbial contamination for each of the four microbial relative abundances (log10 gene-copies-per-16S-rRNA-copies) responses (r = 1, 2, 3, 4): (1) the log10 relative abundance of bovine-associated markers (log10 copies-*BoBac*-per-16S-rRNA-copies) in sediment, (2) the log10 relative abundance of bovine-associated markers (log10 copies-*BoBac*-per-16S-rRNA-copies) in surface water, (3) the log10 relative abundance of human-associated markers (log10 copies-*HuBac*-per-16S-rRNA-copies) in sediment, and (4) the log10 relative abundance of human-associated markers (log10 copies-*HuBac*-per-16S-rRNA-copies) in surface water. FIT was implemented for each of the four

https://doi.org/10.1021/acs.est.2c00224
Environ. Sci. Technol. 2022, 56, 4231−4240

microbial contamination responses. The microbial FIT framework is described in further detail in Wiesner-Friedman et al.[7]

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 represented by regression coefficient is positive) for the contaminant for unseen (i.e., a test set) response data. The reliability score has three components. The sign stability score (SSS) indicates the number of models of test set data where the source term, constructed from training data assumptions, is positively associated (i.e., the database of source locations appears to contribute). The average magnitude, M, indicates the sum of the quantified associations between test set data and the source term constructed from training data assumptions. Lastly, the standard deviation on the test set regression coefficients, σ_β, was obtained. Across each uth source type, FIT selects the dth database with the highest reliability score.

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 $\hat{\alpha}(u)$ that maximizes the β_u using 100% (i.e., all of the data) of the observations $y_i^{(r)}$ using a penalty on a combination hyperparameters that yield poor regression or mapping qualities (i.e., non-normal residuals or $\hat{y}_i^{(r)} > 0$). We visualized the objective function in the hyperparameter space to determine the upper and lower bound values for α_O, α_R, and γ_G that would better capture a more global maximum. We assessed collinearity between source terms with a Pearson correlation coefficient. If source terms were correlated ($\rho \geq 0.7$), then a source term was chosen from among the correlated source terms with a higher univariate R-squared and a positive coefficient.[36]

After source terms were constructed, the predictive ability of the climatic terms and each noncollinear, informed source term were tested with a stepwise regression approach using the Akaike Information Criterion (i.e., *test* stage of FIT). Sixfold cross-validation of the stepwise selection procedure was used to evaluate the inclusion of source terms in the case of influential observations.

The code to obtain the results of this paper and coded examples of the *find* and *inform* stages of FIT can be accessed at github.com/wiesnerfriedman/FIT_FecalContamination. Instructions on running the code (e.g., required MATLAB toolboxes) and the link to this study's data (required for running the code) can be found in the README.md file at this web address.

3. RESULTS AND DISCUSSION

3.1. Overland and River Flows from AFOs Impact Both Sediment and Surface Water. Table 1 provides results from FIT for bovine-associated responses. For details on detecting host-associated responses, the variability of responses across sampling events, and the effects of precipitation and freezing temperature, see Supporting Information, S4. FIT selected the GORF AFO spatial predictor, representing contributions from manure hauled from AFOs and applied onto fields. Results indicate that a one-standard-deviation increase in these contributions was associated with a 99% (p-value < 0.05) increase in the relative abundance of bovine-associated markers (copies-*BoBac*-per-16S-rRNA-copies) in sediment. For surface water, FIT selected the ORF AFO spatial predictor representing the contributions of AFO sites directly to the river (i.e., not hauled to distant manure fields). A one-standard-deviation increase in AFO contributions was associated with a 79% (p-value < 0.05) increase in the relative abundance of bovine-associated markers in surface water.

One other LUR study associated bovine sources with bovine-associated markers in rivers and found that a one-standard-deviation increase in cattle density at the watershed level was associated with a 20.2% (p-value < 0.05) increase in the absolute abundance of bovine-associated markers (*BacR*-copies) in surface water.[5] The difference in the modeling approach (i.e., source terms modeled at the watershed-level in Dila et al.[5]) and the host source sensitivity of the marker (i.e., *BoBac* has a higher source-sensitivity[37]) offer explanations for larger effects in our study.

Differences existed in selecting the AFO spatial predictor (i.e., AFO via ground hauling vs AFO) and the selection of databases representing AFO (i.e., WPDES CAFO vs AFO manure storages). These differences may reflect that where and how manure is applied depends on the scale of the AFO[38] or that varying manure application practices (e.g., broadcasting or injection) may contribute to differences in the detection and persistence of host-associated markers and fecal indicators when comparing soil sediments and surface waters.[39−41] Additionally, sediment samples were unavailable for this dataset during high flow conditions due to site inaccessibility. Results suggest that under higher flow conditions, AFOs may be better represented as sources by AFO manure storages, which are known to leak under extreme precipitation events,[42] compared to manure application from CAFOs.

3.2. Overland and River Flows from Low-Intensity Developed Land-Cover Impact Both Sediment and Surface Water. FIT selected the ORF low-intensity developed land-cover predictor, representing contributions from land classified as a low density of impervious/manmade features.[43] A one-standard-deviation increase in these contributions was associated with a 108% (p-value < 0.05) increase in the relative abundance of bovine-associated markers in the sediment.

Few studies have explored relationships between bovine fecal markers and low-intensity land-cover or impervious surfaces. None have found any statistically significant associations.[4,25,44] However, studies have explored and revealed relationships between low-intensity land-cover and fecal indicator bacteria in freshwater.[4,21,22,25,44] In an area strongly affected 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.[21] Therefore, this land cover may act as an important source of fecal contamination, but its behavior as a source may be tied to the presence of other sources.

Previous work investigating the relationship between bovine fecal markers and low-intensity developed land-cover did not take samples from livestock-dense regions.[4,25,44] Thus, an essential consideration for defining source terms with this low-intensity developed land-cover is that the land-cover may capture different phenomena for different regions. In Kewaunee County, low-intensity developed land-cover captures an intermediate manure application rate due to two phenomena unique to dense livestock farming in populated areas (Figure 1). First, we have found that on average, low-intensity developed land-cover areas are located 6.48 km from CAFOs, 442 m from AFOs, and 292 m from manure fields. These distances support the application of manure on nearby large farms that migrate off-site, specifically when these large farms are near low-intensity developed land-cover. An alternative second scenario is contributions from small farms located in the low-intensity

Table 1. Two-by-Two Table of Regression Results for Predicting the Relative Abundance of Bovine in Sediment (log10 *BoBac*-Copies-per-16S-rRNA-Copies), the Relative Abundance of Bovine in Surface Water (log10 *BoBac*-Copies-per-16S-rRNA-Copies), the Relative Abundance of Human in Sediment (log10 *HuBac*-Copies-per-16S-rRNA-Copies), and the Relative Abundance of Human in Surface Water (log10 *HuBac*-Copies-per-16S-rRNA-Copies)[a]

response	bovine in sediment ($n = 90$)			bovine in water ($n = 98$)		
source term	hyperparameters $\boldsymbol{\alpha}$	coefficients β	RAR	hyperparameters $\boldsymbol{\alpha}$	coefficients β	RAR
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 hauling to manure fields)	$\gamma_G = 13$ km, $\alpha_O > 1$ km	0.298**	1.99	NS	*NS*	*NS*

response	human in sediment ($n = 90$)			human in water ($n = 79$)		
source term	hyperparameters $\boldsymbol{\alpha}$	coefficients β	RAR	hyperparameters $\boldsymbol{\alpha}$	coefficients β	RAR
septic systems	$\alpha_O > 1$ km	0.170	1.48	$\alpha_O > 1$ km	0.178*	1.51
high-intensity developed	$\alpha_O > 1$ km	0.193	1.56	NS	NS	NS

[a] For each of the responses, columns represent the *inform* and *test* stages of the FIT framework. Hyperparameters, $\boldsymbol{\alpha}$, standardized regression coefficients, β, and the RAR. Rows correspond to stepwise selected source terms. The table only displays consistencies for hyperparameter values, $\boldsymbol{\alpha}$. Below detect values were removed prior to statistical modeling. **p-value < 0.05, *p-value < 0.10. NS: not selected. Bovine source terms never selected: manure fields (no ground hauling from AFO), land-applied sludge (all types or industrial types), industrial wastewater treatment plants. Human source terms never selected: wastewater treatment plants and land-applied sludge (all types or wastewater/septage).

developed land. Due to the pervasiveness of dairy farming in Wisconsin,[45] we cannot exclude that the low-intensity developed land-cover may represent small family farms.

The second phenomenon may account for the contributions of cowpats from small farms, which may be substantial enough to increase the abundance of bovine markers in riverbed sediment significantly. Bovine-associated markers can persist in cowpats past a month, disseminate from cattle grazing, and release manure into aquatic settings over days.[46−48] For the first phenomenon, impervious surfaces near pollution sources may increase the conveyance of pollutants and reduce the soil's ability to filter pollution.

We similarly find that the low-intensity developed land-cover is a source of bovine fecal contamination for surface water. A one-standard-deviation increase in this land cover was associated with a 39% increase (inclusion reduced AIC) in the relative abundance of bovine-associated markers. However, this association was not statistically significant and indicates that the sample size is insufficient for this response.

3.3. Overland and River Flows from Septic Systems Impact Sediment and Surface Water. The lower half of Table 1 provides the results of the human-associated responses. FIT selected the ORF septic system source term, representing contributions from drainfield polygon centroids to explain human-associated fecal pollution in surface water and sediment. 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 (copies-*HuBac*-per-16S-rRNA-copies) in surface water and a 48% (inclusion lowered AIC) increase in this relative abundance in the riverbed sediment.

Two previous studies of human-associated markers have found that an increase in 1 septic system per square 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., copies-, HF183-, or HumM2-per-100 mL-water).[11,15] An increase in distance of 1 km 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-100 mL-water).[11] Similar findings on proximity of septic systems occurred in a groundwater study.[14] These studies suggest that the density or proximity of sources is key to characterize human fecal contamination. The associations we found with the ORF SPM link these studies by finding associations using both the density and proximity of systems and overland and downstream flow.

Below-detects (BD = 19) were only found for *HuBac* in surface water. Removing BDs from the data bias our estimates of RARs upward. We find that BD inclusion (i.e., set them to one-half of the detection limit) yields climatic-only models of the relative abundance of *HuBac* in surface water (i.e., RAR = 1). With BDs removed, the septic system source term was selected and yielded the model presented in Table 1. Due to this exclusion of low values, we are likely overestimating the RAR for septic system impacts on surface water.

3.4. 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 source term, which characterizes a high density of impervious surfaces. A one-standard-deviation increase in high-intensity developed land-cover was associated with a 56% (inclusion lowered AIC) increase in the relative abundance of human fecal markers. High-intensity developed land-cover has been previously associated with human fecal pollution in surface water and groundwater.[4,23−25] To our knowledge, our study is the first to report this association in the riverbed sediment. In this scenario, our study and others may be capturing signals from cross-reactive

Figure 1. Example of a low-intensity developed land-cover area, defined as a low density of impervious surfaces, is highlighted in yellow. This low-density imperviousness is near a large farm with an AFO manure storage. Image generated with ArcGIS 10.5 and Wiscland-2 Land Cover.[43,49]

fecal contamination (i.e., nonhuman hosts), which is expected due to *HuBac's* low estimated specificity to human feces.[2]

3.5. Higher Host-Associated Marker Specificity Leads to Better Signals from Sources. The databases representing bovine sources were more reliable in behaving as sources (i.e., positive test-set regression coefficient from training model in the *find* stage) of bovine marker abundance compared to databases of human sources for human marker abundance. Figure 2 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 the databases options relevant to each microbial response. This rate was calculated by averaging the SSSs calculated in the *find* stage and dividing them by the number folds used. The database-as-source success rate was higher for the bovine-associated responses (57.3%) compared to human-associated responses (50.8 and 33.8%). However, databases representing potential sources of human fecal contamination were small and numerous compared to the databases representing potential sources of bovine fecal contamination (e.g., septic system vs AFO). More sampling sites may be needed to characterize contributions for smaller and more numerous sources.

Additionally, the higher specificity of *BoBac* to bovine feces versus *HuBac* to human feces[29,37,50,51] may lead to better signals from sources. Although *HuBac* has been shown to efficiently detect human fecal contamination (>95% sensitivity),[52] the high rate of false positives for other fecal contamination sources (e.g., swine)[2] may confound the results of this study. Future work should look to more recently developed markers with higher specificities to validate our findings. However, a tradeoff may exist between marker sensitivity and marker specificity. In one study, quantifying the human-associated HF183 marker from sewage-contaminated environmental samples led to more false negatives compared to *HuBac*.[53]

Alternatively, better source signals may come from the ease of detecting *BoBac* versus *HuBac* (p-value < 0.05) in our study (see Supporting Information, S8), likely due to larger loads of bovine versus human waste. This detection pattern has been previously observed in a region with a higher bovine to human ratio.[54] Our findings suggest the value of the *find* stage of FIT in determining database-as-source success rates, which can help evaluate the performance of host-associated markers in LUR studies.

3.6. Sediment Responses Better Capture Signals from Distant Sources than Surface Water. As a result of the *inform* stage of FIT, hyperparameter values were selected and provide the first estimate of overland ranges of influence (α_O) of sources for surface water and sediment contamination in the same study, thereby allowing comparisons of α_O for surface water versus sediment contamination (see Supporting Information, S9 for detailed hyperparameter results). We observed that the overland flow hyperparameter, α_O, was longer for responses measured in the sediment compared to surface water. The hyperparameter values selected to inform the source terms are available in Table 1. The RAR is maximized when overland distance decay ranges, α_O, are greater versus less than 1 km for sediment responses compared to water. The spatial depiction of overland transport would show larger overland ranges of influence around sources for sediment contamination compared to surface water contamination. The longer α_O for responses from sediment versus surface water found in this study support that sediment can provide more time-integrated information than river waters due to their transient nature. Sediment sampling may offer a more complete characterization of microbial risks in surface water, which also increase under conditions that disturb the sediment.[27,55,56]

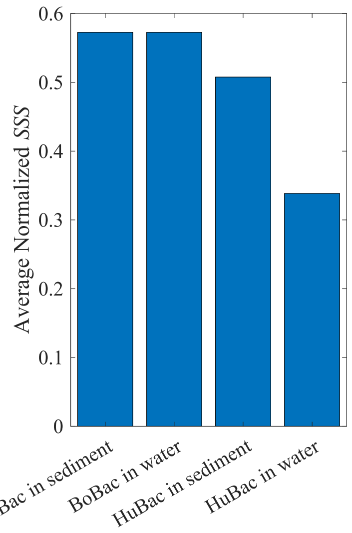


Figure 2. Average normalized SSS across candidate databases for each of the four responses. The average normalized SSS represents a database-as-source success rate. The database-as-source success rate describes the average success across a k-fold cross-validation on the regression coefficient positivity. 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 to determine how well the SPM hyperparameters work to construct spatial predictors that are positively associated (i.e., source terms) with unseen data. Hyperparameters and associations are obtained for $k = 5$ folds (i.e., five sets of already modeled/unseen data). The SSS equals 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.

Mechanisms leading to long-term pollutant storage in sediment may differ from the short-term contamination of surface water.[26,27] Short-term surface water contamination may be affected by high-intensity runoff over a short distance.[57] In contrast, our novel inclusion of long-term sediment contamination may allow us to detect time-integrated contamination over longer distances than short-distance runoff. These mechanisms fall under two explanations.

The first may be differences in physical pathways of transport. They include differences in manure application (e.g., broadcasting, incorporation, or injection), and that the transport occurs over different pathways (e.g., sheet flow or runoff,[40,57,58] or subsurface flow through geological features including karst[59−61] and tile drainage,[39,59,62] or even wind transport[63]). Different attachment to particles may also influence transport.[64]

Another explanation may be differences in sensitivity in measuring an effect using time-integrated sediment concentrations versus transient water concentrations. In general, host-associated markers persist longer in sediments than surface water.[65,66] It is possible that this persistence in sediment represents time-integrated responses to pollution and is, therefore, more sensitive to pollution and detecting effects from more distant sources.[67] This is supported further by our study, where we observe a higher average database-as-source success rate (Figure 2) for the human-associated response in sediment (50.8%) versus water (33.8%).

3.7. Implications, Limitations, and Future Work. This study found that sediment samples were important for detecting

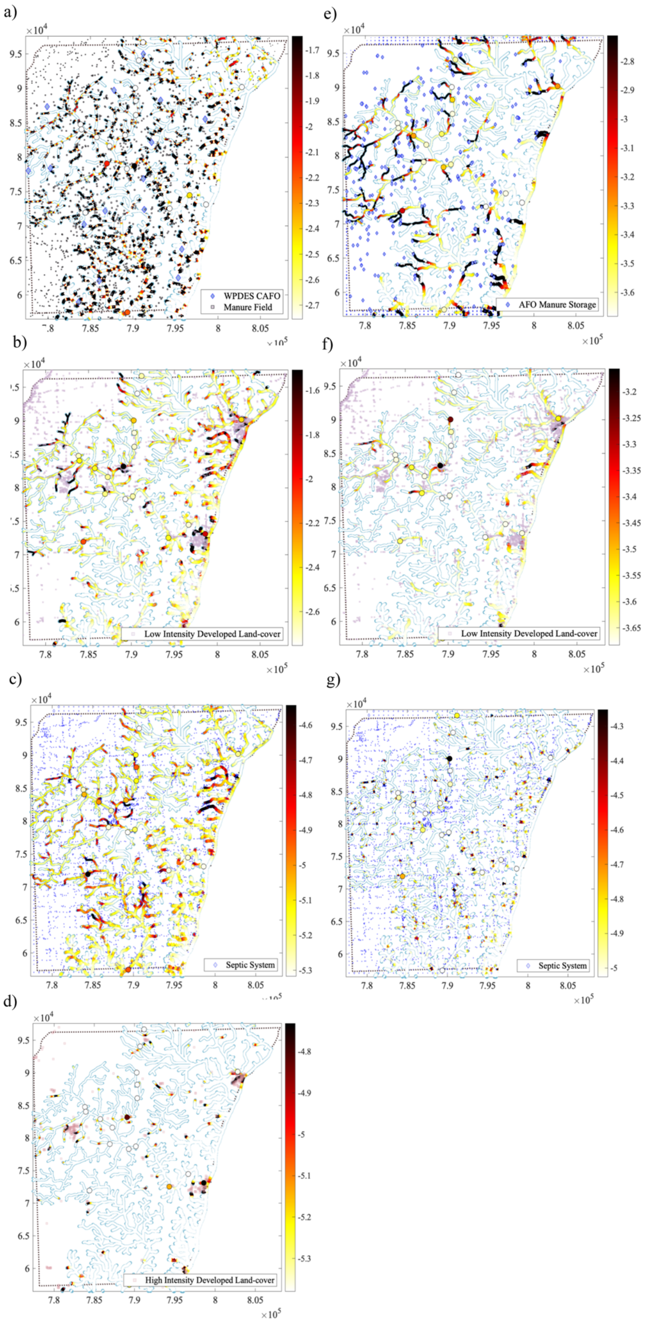


Figure 3. Modeled relative abundance host-associated fecal contamination (logl0 gene-copies-per-16S-rRNA-copies) plotted on the river network for Kewaunee County given average recent and antecedent precipitation and freezing temperature. (a) AFO contributions to the relative abundance of *BoBac* in sediment, (b) low-intensity developed landcover 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, (t) lowintensity developed land-cover contributions to *BoBac* in surface water, and (g) high-intensity developed land-cover contributions to *HuBac* in surface water. The river network is depicted by a blue outline. Kewaunee County sampling sites are depicted as circles with colors corresponding to modeled relative abundance.

signals from AFOs, septic systems, and developed land cover by defining spatially distributed source locations and estimating influence ranges around sources and RARs. One advantage of FIT is that model information can predict impacts from sources at unsampled locations, informing quantitative microbial risk assessment, epidemiological study design, or future monitoring site selection. Using the databases selected from the *find* stage of FIT, the hyperparameter values obtained from the *inform* stage, the regression coefficients obtained from the *test* stage, and averages of the climatic variables, we predict key source impacts to the log10 relative abundance gene marker responses (Figure 3). The modeled associations between source terms and bovine and human marker responses were used to depict the fecal pollution impacts from these sources in the riverbed sediment and surface water from AFOs (Figure 3a,e), low-intensity developed land-cover (Figure 3b,f) septic systems (Figure 3c,g), and high-intensity developed land-cover (Figure 3d) under average precipitation and temperature sampling scenarios (see Supporting Information, S10 for details).

From Figure 3, we see that the spatial predictions of source impacts appear localized around dense clusters of sources, but impacts are not limited to smaller order streams. Our model indicates that higher-order reaches of the river (e.g., the main reaches of the Kewaunee, Ahnapee, and East Twin Rivers) are affected. Because smaller-order streams do not benefit from dilution, they are predicted to be the most affected.

Other spatial factors may contribute to the variability, which is one limitation of our study. Factors such as geology, soil characteristics, or fields with tile drainage, where pipes convey water away from agricultural fields toward nearby rivers and streams,[39,59,68] will vary from region to region. For example, Kewaunee County has karst geology where fractured rock may easily convey pollutants to rivers through groundwater.[31,69] Tile drainage is also present in Kewaunee County and field experiments in other regions report its impact on surface water pollution.[39,68] In future work, we plan to explore how tile drainage, in addition to water table depth, bedrock depth, and soil permeability, may amplify or attenuate microbial pollution in surface and sediment.

This study supports previous work[28,70] by showing how bovine and human host-associated sources impact fecal pollution region's sediment and surface water. Our SPM, which accounts for density and utilizes overland ranges of influence, indicates how buffering sources from rivers and streams by distances greater than 1 km and reducing fecal source density could improve water quality in agriculture-dense regions, such as Kewaunee. This study highlights the importance of sediment sampling when characterizing microbial risks.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c00224.

> Literature review on source-focused studies of microbial responses in sediment and surface water; sampling methods; host-associated MST response data; detection of host-associated responses and the variability across sampling events; spatial distributions of the relative abundance of host-associated responses averaged across sampling events; types of sources and candidate databases of spatially distributed sources; reliability score results; statistical analysis, stepwise-selected models, and description of climatic effects; hyperparameter results for source terms selected by the FIT model; depiction of source

contributions to log10 relative abundance responses; and cross-validation (DOCX)

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

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.2c00224

Author Contributions

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

Funding

Corinne Wiesner-Friedman 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-USDA 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 WIDNR for their assistance in obtaining useful spatial databases.

ABBREVIATIONS

AFO	animal feeding operation
BoBac	ruminant host marker
CAFO	concentrated animal feeding operation
HuBac	human host marker
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) 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.

(2) 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.

(3) 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.

(4) 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.

(5) 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.

(6) 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.

(7) Wiesner-Friedman, C.; Beattie, R. E.; Stewart, J. R.; Hristova, K. R.; Serre, M. L. Microbial find, inform, and test model for identifying spatially distributed contamination sources: framework foundation and demonstration of ruminant bacteroides abundance in river sediments. *Environ. Sci. Technol.* **2021**, *55*, 10451−10461.

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

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) Adegoke, A. A.; Stenstrom, T.-A.. Septic Systems. *Water and Sanitation for the 21st Century: Health and Microbiological Aspects of Excreta and Wastewater Management (Global Water Pathogen Project)*; Michigan State University, 2019.

(14) 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.

(15) 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.

(16) 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. U.S.A.* **2015**, *112*, 10419−10424.

(17) 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* **2018**, *6*, 46.

(18) 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.

(19) Marti, R.; Ribun, S.; Aubin, J.-B.; Colinon, 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.

(20) 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* **2014**, *6*, 3701−3723.

(21) 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. *SE. Geogr.* **2016**, *56*, 428−453.

(22) 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.

(23) 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.

(24) 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.

(25) 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.

(26) Wohl, E. Legacy effects on sediments in river corridors. *Earth-Sci. Rev.* **2015**, *147*, 30−53.

(27) 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.

(28) 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*, fiy131.

(29) 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.

(30) Datasets|Climate Data Online (CDO)|National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/cdo-web/datasets (accessed Dec 15, 2021).

(31) 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.

(32) 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.

(33) 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.

(34) 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.

(35) 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.

(36) Dormann, C. F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J. R. G.; Gruber, B.; Lafourcade, B.; Leitão, P. J.; Münkemüller, T.; McClean, C.; Osborne, P. E.; Reineking, B.; Schröder, B.; Skidmore, A. K.; Zurell, D.; Lautenbach, S. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27−46.

(37) 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.

(38) 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.

(39) 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.

(40) 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.

(41) Sharma, M.; Millner, P. D.; Hashem, F.; Vinyard, B. T.; East, C. L.; Handy, E. T.; White, K.; Stonebraker, R.; Cotton, C. P. Survival of Escherichia coli in Manure-Amended Soils Is Affected by Spatiotemporal, Agricultural, and Weather Factors in the Mid-Atlantic United States. *Appl. Environ. Microbiol.* **2019**, *85*, 02392.

(42) Burch, T. R.; Stokdyk, J. P.; Spencer, S. K.; Kieke, B. A.; Firnstahl, A. D.; Muldoon, M. A.; Borchardt, M. A. Quantitative microbial risk assessment for contaminated private wells in the fractured dolomite aquifer of kewaunee county, wisconsin. *Environ. Health Perspect.* **2021**, *129*, 67003.

(43) 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).

(44) 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.

(45) Rosen, C. J.; Ostergren, R. C.; Vale, T. R. Wisconsin land and life. *Geogr. Rev.* **1998**, *88*, 452.

(46) 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.

(47) 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.

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

(49) Esri. "Imagery" [basemap]. Scale Not Given. "World Imagery," 2021.

(50) Shanks, O. C.; White, K.; Kelty, C. A.; Sivaganesan, M.; Blannon, J.; Meckes, M.; Varma, M.; Haugland, R. A. Performance of PCR-based

assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples. *Environ. Sci. Technol.* **2010**, *44*, 6281−6288.

(51) Shanks, O. C.; White, K.; Kelty, C. A.; Hayes, S.; Sivaganesan, M.; Jenkins, M.; Varma, M.; Haugland, R. A. Performance assessment PCR-based assays targeting bacteroidales genetic markers of bovine fecal pollution. *Appl. Environ. Microbiol.* **2010**, *76*, 1359−1366.

(52) Ahmed, W.; Hughes, B.; Harwood, V. Current status of marker genes of bacteroides and related taxa for identifying sewage pollution in environmental waters. *Water* **2016**, *8*, 231.

(53) Coakley, T.; Brion, G. M.; Fryar, A. E. Prevalence of and Relationship between Two Human-Associated DNA Biomarkers for Bacteroidales in an Urban Watershed. *J. Environ. Qual.* **2015**, *44*, 1694−1698.

(54) 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.

(55) 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.

(56) 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.

(57) 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.

(58) 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.

(59) 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.

(60) 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 Res.: X* **2019**, *5*, 100038.

(61) 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.

(62) 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.

(63) 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.

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

(65) 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.

(66) 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*, No. e00005.

(67) Kasich, J.; Taylor, M.; Nally, S. J.. *Sediment Sampling Guide and Methodologies*; Ohio Environmental Protection Agency, 2012.

(68) 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.

(69) Diston, D.; Robbi, R.; Baumgartner, A.; Felleisen, R. Microbial source tracking in highly vulnerable karst drinking water resources. *J. Water Health* **2017**, *16*, 138.

(70) Socially Responsible Agriculture Project. *RAP Sheets: Industrial Dairies in Kewaunee County, Wisconsin*; 2015. https://sraproject.org/wp-content/uploads/SRAP-Rapsheet-2015.pdf.