Title: Scale of analysis drives the observed ratio of spatial to non-spatial variance in microbial water quality: insights from two decades of citizen science data

Target Journal: Applied and Environmental Microbiology

Authors: Daniel L. Weller^{ab}, Donald E. Weller^c, Laura K. Strawn^d, Tanzy M. T. Love^a

^a Department of Biostatistics and Computational Biology, University of Rochester Medical Center

^b Department of Environmental and Forest Biology, SUNY College of Environmental Sciences and Forestry

^c Smithsonian Environmental Research Center, Edgewater MD

^d Department of Food Science, Virginia Polytechnic University

Abstract:

While fecal indicator bacteria (FIB) testing is used to monitor surface waters for potential health hazards, recent studies report substantial variation in FIB levels and that observed variation appeared dependent on scale of analysis (SOA). Citizen science data and random effects models were used to quantify variance in FIB levels attributable to spatial versus temporal factors. Separately, Bayesian models were used to quantify the ratio of spatial to nonspatial variance in FIB levels, and identify associations between environmental factors and FIB levels. Separate analyses were performed for each stream, lake, and watershed as well as at the state-level (three SOAs). As SOA increased (from waterway to watershed to statewide models), variance attributable to spatial sources generally increased and variance attributable to temporal sources generally decreased. While relationships between FIB levels and environmental factors, such as flow conditions, were constant across SOA, the effect of land cover was both highly dependent on SOA and consistently smaller than the effect of outfalls. This study demonstrates the importance of scale when designing monitoring programs or using FIB monitoring to inform management. Moreover, these data represent a comprehensive survey of water quality in Upstate NY, and the study highlights a valuable publicly available database.

Keywords: Spatial Variation, E. coli, Turbidity, Total Suspended Solids, Bayesian Analysis, Seasonality, Water Monitoring

Fecal contamination of surface waterways that supply agricultural (e.g., for irrigation, fertigation, dust abatement), drinking, and recreational water has been linked to an increased risk of gastrointestinal and respiratory illness (1–6). For example, multiple *E. coli* O157:H7 outbreaks associated with romaine lettuce between 2017 and 2019 were traced to the probable use of irrigation water contaminated by cattle feces (7–9). Despite this disease burden, pathogen monitoring of agricultural surface water remains impractical due to the number of potential pathogens, the cost of sampling, access to testing laboratories, and methodological limitations amongst other difficulties (10–14). Instead, waterways are monitored for fecal indicator bacteria (FIB) to identify when a presumptive fecal contamination event may have occurred. Regulations and voluntary industry agreements have codified the use of FIB monitoring for assessing surface water suitability for irrigation and recreation as well as for determining if corrective actions are needed [e.g., (15–24)]. For example, Australia and New Zealand, and the United States (US) have established threshold values for coliforms and *E. coli*, respectively, in irrigation water (15, 16), while many countries, including the US, Canada, the European Union, South Africa, and the United Kingdom, have established similar standards for recreational water (17–21).

While FIB-based water standards assume that the proscribed sampling frequency is sufficient to capture temporal variation in water quality, recent studies suggest that the sampling frequencies in several US regulations do not (13, 25–27). As a result, concerns have been raised by scientists and stakeholders about the utility of *E. coli* monitoring programs for assessing microbial hazards in agricultural water (13, 25, 28, 29). Effective implementation of FIB monitoring is further limited by the lack of guidance on the spatial scale at which sampling should be performed. For example, stakeholders have requested guidance on how *E. coli* monitoring outlined in the US Produce Safety Rule should be implemented to adequately capture spatial variation within a water source (29). Additional information on spatial and temporal variation in microbial surface water quality is thus needed to revise current monitoring guidance, especially to determine if changes to monitoring efforts should focus on increased sampling frequency or a larger number of sampling sites. Since recent studies have raised questions about the reliability and efficacy of FIB-based standards for assessing the suitability of waterways for agricultural uses (13, 25, 28), there may be a need for supplementary or alternative indicators. For example, a survey of New York streams proposed using sediment levels (e.g., turbidity) as a supplementary indicator of microbial water quality (13). The repeated identification of a positive association between sediment levels and FIB levels, and between sediment levels and enteric pathogen presence suggests that these associations are reproducible and robust to study design (e.g., location of study, sampling method, laboratory methods (13, 26, 30–33)). However, before incorporating sediment monitoring into existing FIB monitoring programs, an improved understanding of spatial and temporal variation in sediment and FIB levels is also needed (e.g., to inform sampling frequency).

Many surveys have reported considerable variability in FIB and sediment levels over space and time (e.g., (13, 32–38)). For example, Rafi et al. (34) found that *E. coli* levels, and their relationship with watershed size and stream order, varied among three US ecoregions along an East-West gradient (34). Such variability makes sense because surface waterways are hierarchical systems in which water quality is affected by scale-dependent and spatially heterogeneous processes and landscapes. Several studies have shown that the strength of associations between FIB levels and land use were dependent on the spatial and temporal scales considered (37, 38). Applied papers have shown that such variability complicates the interpretation of FIB test results and the development of science-based monitoring programs (39). In fact, there are repeated reports that FIB levels differ between samples collected at the same (i) time but from different locations in a waterway (e.g., banks of a river, depths in the water column) or (ii) location but at different times (e.g., within a day, on consecutive days) (13, 40). Given such variability, how should stakeholders use monitoring data to guide decision-making? The real-world implications of these knowledge gaps create a clear need for an improved understanding of temporal versus spatial contributions to variance in microbial water quality. Indeed, while a few papers have discussed how scale of analysis affects observed associations between FIB levels and environmental factors (37, 38), they did

not consider the impact of scale on the percent of variance attributable to spatial versus temporal sources, or differences in the impact of scale on observed variability for different water types (e.g., between lakes and streams). These knowledge gaps were highlighted by a review that separately ranked sources of variation in FIB levels in streams, lakes, and coastal systems, but did not make comparisons among water types (39), presumably due to lack of data. The main aims of this study were to address these knowledge gaps by quantifying the impact of spatial scale on the variance in FIB and sediment levels attributable to specific spatial or temporal factors, and on the observed relationship between *E. coli* levels and environmental factors (e.g., land use, weather). A secondary aim was to highlight a publicly-available dataset that can serve as a resource for future studies. Specifically, this study used data from a citizen science initiative run by an NYS certified laboratory that represents one of the more comprehensive publicly-available water quality datasets in Upstate New York State due to the large geographic area sampled (Fig. 1), and the number of water quality parameters measured (Table S1).

Methods:

Study Overview: Data were collected by the Community Science Institute (CSI) and are available at their website <http://www.communityscience.org/>. While this is a publically-available dataset, this is the first publication to summarize all water quality parameters analyzed (Table S1) and to assess spatiotemporal variation in key parameters. CSI runs a NY Department of Health (NYDOH) certified lab, and the methods used to enumerate each of the parameters reported here are detailed on their website, in their certifications (available from the NYDOH), and in Table S1. Data were collected between 2002 and 2020 (N=11,351 samples) and represent 58 waterways (Fig 1).

Metadata Collection. For each sample collected, CSI characterized flow conditions as baseflow or stormflow. The presence of outfalls at or immediately upstream of the sampling site, if the site was at the stream mouth (i.e., where the stream entered a lake), and water type (e.g., canals, lake, stream) were also noted. For streams, the inverse-distance weighted proportion of each stream's watershed under cropland, developed, forest-wetland, and pasture cover was assessed as previously described (41). Upstream land use was not characterized for canals because they lack consistent, unidirectional flow, and delineating upstream watersheds was not possible. Nor was land use characterized for the sampled lakes (e.g., Cayuga Lake, Seneca Lake), which were sufficiently large that we did not expect land cover in the watershed to be associated with water quality at individual sampling points, particularly when samples were collected away from shore. The county and HUC-4 watershed where each sample was collected were also recorded. However, due to the number of samples collected in the Finger Lakes HUC-4 watershed, samples collected there were further subdivided into the areas draining into Cayuga Lake, Keuka Lake, and Seneca Lake, respectively. We call these subdivisions "watersheds", and refer to individual streams and lakes as "waterways" to avoid confusion. Additionally, the season, day of the week, and week of the year (i.e., number of weeks since Jan 1st) for each sampling event was recorded.

Descriptive Analysis. Watershed delineation was performed in ArcMap (42), while all other analyses were performed using R version 3.5.3 (43). While 11,351 samples were collected, not all analytes were enumerated in all samples resulting in substantial missingness (Table S1). Four water quality parameters [*E. coli*, total coliforms, total suspended solids (TSS), and turbidity] were selected as the foci of analyses. These parameters were selected because they are relevant to environmental, human, and/or economic health, and they were measured in more than 1,000 samples collected from multiple watersheds. In comparison, fecal coliform levels, another FIB used in monitoring programs, were only enumerated in 155 samples representing three streams, and as such was not included among the focal parameters in the present study. Spatial and temporal patterns for each of the four focal parameters were visualized, as were differences in water quality between baseflow and stormflow conditions, watersheds, and water types. Spearman's correlation between all continuous parameters (e.g., water temperature, week of the year, year, *E. coli* levels) was assessed and visualized using the corrplot package (44). For the correlation analysis and all other analyses described below the four parameters were log₁₀-transformed.

Statistical Analysis. To assess the effect of scale on observed sources of variance, random effects models and intercept-only Bayesian models of coregionalization (MCOR) were implemented in the following five ways:

- separately for each stream and lake (the “waterway models”)
- separately for each watershed (the “watershed models”)
- using all stream data combined (the statewide “all-stream model”)
- using all lake data combined (the statewide “all-lake model”),
- using all the data (the “statewide model”).

As a result, water type-specific and water type-agnostic comparisons could be made across three scales of analysis (waterway, watershed, and statewide). Multivariable MCOR were implemented in the same five ways to assess the effect of scale on the observed relationship between FIB levels and environmental factors and allow similar comparisons to be made between spatial scales and water types. Five streams (referred to as the direct streams), Deans Creek, Lake Ridge Creek, Mill Creek, Paines Creek, and Townline Creek, were treated as the same waterway for sampling purposes by CSI; as a result, separate waterway models were implemented for each of the five streams as well as all direct streams combined.

In addition to streams and lakes, CSI collected samples from canals/flood control channels (henceforth called canals), springs, ponds, and wetlands. Wetland samples were not used in any analyses here. Spring data was only used in the statewide multivariable MCOR to allow comparison between water types. All sampled canals linked two lakes, drew water from a lake, served as a flood control channel for a lake, or connected a lake to another waterway. Therefore, canal data were grouped with the corresponding lake for the individual lake and all-lake models but kept separate in the descriptive analyses to facilitate visualization of differences between canals and lakes. Results from individual pond and lake models are visualized together in the graphical summaries due to the small number of ponds (N=2). Similarly, pond data were pooled with lake data for the descriptive analysis and all-lake models. As a result, dummy variables for water type were included as covariates to quantify differences among canals, lakes, ponds, streams, and/or springs in the statewide multivariable MCOR.

Random Effects Models. Separate random effects models were built to assess the percent of variance in each water quality parameter that could be attributed to specific spatial and temporal factors. Since the aim of the random effects models was to compare variance attributable to spatial versus temporal sources, individual waterway models were not implemented for waterways where all samples were collected from a single site: Chapel Pond, Lake Ridge Creek, Glen Eldridge Creek, or Townline Creek. Data for Lake Ridge and Townline Creek were used in the Direct Streams waterway model, while data for all four waterways were used in the appropriate watershed and statewide models. Each random effects model included one of the following factors as a random effect: month, season, week day (e.g., Monday, Tuesday), week of the year (i.e., weeks since Jan. 1st), year, sampling site, and county. For waterways and watersheds where different water types were sampled (e.g., impoundments within streams) or where stream mouths were sampled, random effects models were implemented with a random effect of water type. Models were fit using the lme4 package (45), and R^2 was calculated using the MuMin package (46) as previously described (47). Results for each water type (stream versus lake) and each scale of analysis were visualized using ggplot (48).

Bayesian Models of Coregionalization (MCOR). Intercept-only MCOR were fit to characterize the ratio of spatial to non-spatial variance for each water quality parameter (*E. coli*, total coliforms, TSS, and turbidity), and to quantify the effective range, which is the distance at which FIB and sediment levels became independent. Since the aim of the intercept-only models was to quantify the ratio of spatial to non-spatial variance, individual waterway models were not implemented for the four waterways where all samples were collected from a single site as discussed above. Separately, multivariable MCOR were fit to identify drivers of elevated or reduced *E. coli* levels, while explicitly accounting for spatial

autocorrelation among sites. Multivariable models were only developed for *E. coli* because *E. coli* monitoring is commonly used to identify when a presumptive fecal contamination event has occurred in agricultural and recreational waterways, and multiple governments have established *E. coli*-based standards for agricultural and recreational waterways (11, 16, 17, 29, 49). As such, understanding drivers of elevated *E. coli* levels is of economic, environmental, and human health importance (16, 17, 29, 49).

Before developing the univariable and the multivariable MCOR, semivariograms were fit to help generate informative priors for ϕ (3/effective range), τ (nugget, or non-spatial variance), and ρ (partial sill, or spatial variance). If the semivariogram could not be fit using an exponential covariance model then a Gaussian model was attempted; if a semivariogram could still not be fit then non-informative priors were set ($\phi=3/330$, nugget=semivariance, partial sill=semivariance/5). The MCOR were fit and estimates were quantified using the spBayes and coda packages (50–52). Ten thousand iterations were performed for each MCOR, but the first 7,000 were discarded as burn-in. For all MCOR, the 25th, 50th, and 75th percentiles for ϕ , τ , ρ , the ratio of spatial to non-spatial variance, and the percent of variance attributable to spatial sources were calculated. For the multivariable MCOR the 25th, 50th, and 75th percentiles for the effect estimates were also quantified.

In all multivariable MCOR, the following factors were included as explanatory factors when applicable: $\log_{10}$ TSS; flow conditions (baseflow=reference-level); month (April= reference-level); if the sampling site was at the stream mouth (No= reference-level); if the site was at or immediately downstream of an outfall (No= reference-level); and year. In models that included stream but not lake data, between one and three land cover parameters (IDW proportion of area under crop, developed, or pasture cover) were also included. The number of land cover factors depended on the correlations among the three cover parameters for the subset of data used in the model. Developed cover was only included if the correlation between developed and pasture cover was <0.7 , while cropland was only included if the correlation of developed with crop and pasture cover, and of pasture with crop cover were <0.7 . Forest-wetland cover was not included in any models because the correlation between forest-wetland cover and one or more of the other land covers was >0.7 for all data subsets. In the statewide model and all-lake models, water type was also included as a covariate to see if *E. coli* levels varied substantially among water types. Similarly, watershed was included in the statewide model to see if *E. coli* levels varied substantially among watersheds.

Results

Between 2002 and April 2020, the Community Science Institute (CSI) and its volunteers collected and tested 11,351 surface water samples (Table S1). The waterbodies sampled spanned Upstate New York from Lake Erie to the Capital Region, and Lake Ontario to the Pennsylvania border (Fig 1). The number of samples increased from 18 samples collected in 2002 to a peak of 1,310 samples in 2014 (Fig S1). Following sample collection, microbe (e.g., *E. coli*, total coliforms), nutrient (e.g., nitrate, total phosphorous), heavy metal (e.g., arsenic), and/or sediment (e.g., TSS, turbidity) concentrations as well standard physicochemical parameters (e.g., conductivity, pH) were enumerated. However, each parameter was only measured in a subset of the 11,351 samples collected (Table S1). This paper focused on four parameters, *E. coli*, total coliforms, total suspended solids (TSS), and turbidity (Table 1; Table S1; Fig 1).

Visualizing Spatial and Temporal Patterns. Temporal patterns were assessed visually using loess-smoothed trendlines at three temporal scales: interannual, intra-annual (or seasonal), and event-scale (i.e., stormflow versus baseflow). Because sampling was not conducted at the same site or on the same waterway on consecutive days or weeks, day-to-day and week-to-week variation were not assessed. Based on the loess-smoothed trendlines, there was no evidence of substantial interannual variation in TSS and turbidity levels (Fig S3). Conversely, *E. coli* levels appeared to increase, on average, between 2015 and 2020 in streams and lakes but not canals, while total coliform levels appeared to decrease in streams and canals over the same period (Fig S3). Microbial levels in lakes followed a sinusoidal pattern with levels generally increasing from 2015 to 2020 (Fig S3). *E. coli* and total coliform levels also showed

evidence of seasonality (Figs S2-S3) with levels peaking between May and October; this trend was more pronounced in canal and lake samples compared to stream samples. Conversely, TSS and turbidity peaked in the colder months (Fig S2-S3). Both microbial and sediment levels were substantially higher under stormflow than baseflow conditions (Fig 2).

Spatial patterns were examined visually by comparing the distribution of each parameter (*E. coli*, total coliforms, TSS, and turbidity) between water sources and watersheds (Fig 2), and by plotting parameter concentration against latitude, longitude, and land cover (Fig S4). On average, all parameters decreased from North-to-South possibly because of North-South land-use patterns since trends in land cover mirrored trends in water quality (Fig S4). Specifically, southern regions were heavily forested while northern regions were largely agricultural (Fig S4). While sediment levels did not appear to differ substantially between watersheds *E. coli* levels were substantially higher in the Keuka and Seneca Lake watersheds compared to the Cayuga Lake and Susquehanna River watersheds (Fig 2). Both *E. coli* and total coliforms levels were substantially higher in canals and streams compared to lakes (Fig 2). Sediment levels were similar in lakes and streams but higher in canals; in fact, the interquartile range (IQR) for turbidity levels in canals did not overlap the IQR for turbidity levels in either lakes or streams (Fig 2).

The percent of variance attributable to spatial versus nonspatial sources was dependent on the scale of analysis, water type, and waterway/watershed. Random effects models were used to characterize the percent of variance in each water quality parameter accounted for by different spatial (e.g., site, waterway, watershed) and temporal (e.g., week day, week of the year, year) factors as well as water type (Fig. 4, Table S2-S3). Separately, intercept-only MCOR models were used to quantify the percent of variance attributable to spatial sources for each parameter (Fig. S5, Table S2). Comparisons between scales of analysis in the variance attributable to a given source were visualized using boxplots. Briefly, when the point representing the variance attributable to the statewide, all-lake or all-stream models fell outside the interquartile range (IQR) for the waterway or watershed models or when the IQRs did not overlap for the waterway and watershed models it indicated a substantial difference in the variance accounted for by the given factor between model types. Similarly, there was evidence of a potential difference if the median of one group was outside the IQR for the another. Lastly, the difference between the first and third quartile (bottom and top of the box) provides insight into the data distribution, with shorter boxes meaning data hovers around the median values and larger boxes indicating greater variability.

In general, there was considerable variation in the percent of variance attributable to spatial versus non-spatial sources and in the variance attributable to any given factor among the waterway and watershed models (Fig. 4, S5; Table S2, S3) suggesting differences in sources of variance between waterways and between watersheds. However, the variance in all four water quality parameters attributable to spatial factors, such as site, was generally higher for the statewide and watershed models than the waterway models (Fig 4, Table S2-S3). The variance attributable to site in the statewide, and the all-lake and/or all-stream models fell outside the IQR for the waterway models for all four parameters, and, for turbidity and TSS, the IQR for the watershed models did not overlap the IQR for the waterway models (Fig 4). Conversely, the variance attributable to temporal factors largely decreased as the spatial scale of analysis increased (Fig. 4, Fig S5, Table S3). A similar pattern was evident in the intercept-only MCOR models for *E. coli* and total coliforms but not TSS or turbidity. For total coliforms, the IQR for the percent of variance attributable to spatial factors in the watershed MCOR models was higher and did not overlap the IQRs for the lake and/or stream models (Fig S5; Table S2). While a similar pattern was observed for *E. coli*, the IQR, but not the median, for the watershed models overlapped the IQR for the stream models (Fig S5).

While the variance in all four parameters attributable to spatial factors increased as the scale of analysis increased, variance attributable to temporal factors decreased as scale increased and was generally lower in models built using the lake as opposed to stream data. The IQR for the variance

attributable to month, week day, and week of the year in the waterway and watershed models were generally above the variance attributable to the given factor in the statewide, all-lake, and all-stream models (Fig 4; Table S5). For example, the IQR for the variance attributable to week of the year in the stream models did not overlap the IQR for the watershed or lake models for *E. coli*, TSS, and turbidity (Fig. 4; Table S5). Similar differences across scales and/or between water types were observed for *E. coli* in the variance attributable to year, for total coliforms in the variance attributable to month and year, for TSS in the variance attributable to month, week day, and year, and for turbidity in the variance attributable to week day.

Across all four parameters, less variance was attributable to year than to temporal factors that account for seasonality (month, season, and week of the year; Fig. 4, Table S3). Indeed, the temporal factors that accounted for the greatest variance in microbial water quality were seasonal (month, season, or week of the year; Fig 3; Table S3) for all models except one watershed and two stream *E. coli* models for as well as one stream and two lake coliforms models. Overall, variance attributable to season appeared to decrease as scale of analysis increased and was lower for lake compared to stream models (Fig 3; Table S3). Although substantial variance in microbial water quality was attributable to season, little variance was attributable to year for all scales of analysis and water types. While year accounted for 0% of variance in one watershed, three stream, and two lake *E. coli* models, the smallest variance accounted for by week of the year was 8% (Table S3). When the ratio of variance attributable to week of year to variance attributable to year was calculated for each model (replacing 0% with 1%), the ratio was between 1.4 and 25.0 (median=5.9) for watershed models, 1.1 and 77.0 (median=4.5) for stream models, and 2.0 and 23.0 (median=10.9) for lake models. Additionally, 3.7, 3.7, and 3.4 times as much variance was attributable to week of the year compared to year in the statewide, all-stream, and all-lake models. It is also interesting to note that the amount of variance attributable to spatial factors, like site, compared to temporal factors, such as week of the year, also differed by water type. For instance, site and week of the year accounted for 50% and 17%, respectively, of variance in *E. coli* levels in the all-lake model, but for 23% and 22%, respectively, of variance in *E. coli* levels in the all-stream model (Fig. 4, Table S3).

The strength and direction of associations between environmental parameters and log₁₀ *E. coli* concentration were dependent on scale of analysis, water type, and waterway/watershed. Associations between environmental factors and log₁₀ *E. coli* levels were assessed using multivariable Bayesian models of coregionalization (MCOR). Effect estimates in the MCOR can be interpreted as the change in *E. coli* levels (log₁₀ CFU/100-mL) associated with a one-unit change in a continuous factor (such as log₁₀ TSS) or with shifting from the reference-level to a different level of a categorical factor (e.g., from baseflow to stormflow conditions). While there was strong evidence of seasonal patterns in *E. coli* levels for both streams and lakes, the seasonal effect was more pronounced in the stream models compared to the lake models (Fig 4; Table S5). Regardless of waterway, watershed, water type, and scale of analysis, *E. coli* levels were highest in the summer and lowest in the winter (Fig 4; Table S5).

There also appeared to be a strong association between *E. coli* levels and TSS levels, and *E. coli* levels and flow conditions (Fig 4; Table S4). Specifically, *E. coli* levels were positively associated with TSS levels for all water types and scales of analysis as evidenced by the fact that the IQRs for the lake, stream, and watershed models did not include zero but did overlap each other and the effect estimates for the all-stream, all-lake and statewide models (Figure 4). According to the lake, watershed, all-lake, all-stream, and statewide models, *E. coli* levels were also higher under stormflow compared to baseflow conditions (Fig 4; Table S4), which may be evidence of event-scale effects (i.e., elevated *E. coli* levels during rain events). There was also evidence of an association between *E. coli* levels and flow conditions for most individual stream models, but the magnitude and direction of this effect were more variable for the stream models compared to the lake models or models with a larger scale of analysis (e.g., watershed, all-stream, and statewide models; Figure 4; Table S4).

While there was a small, non-negligible effect of land cover in the statewide models (effect estimates ranged between 0.01 and 0.03), some watershed and waterway models showed strong land-use effects (effect estimates ranged between -8.2 and 0.45; Figure 4; Table S4). Overall, the effect of land cover was dependent on scale of analysis and the waterway or watershed under consideration. Specifically, the absolute value of the effect estimates for land cover variables, particularly pasture, were generally but not always greater in the watershed compared to stream models. For example, a 1% increase in pasture cover was associated with a 0.00 (50% credible interval [CI]=0.00, 0.01) log₁₀ CFU/100-mL increase in *E. coli* for the Seneca Lake watershed but a -0.61 (CI=-0.85, -0.40) log₁₀ CFU/100-mL decrease for the Keuka Lake watershed. The effect estimate of land cover factors in many stream models was negligible, but there was a non-negligible effect for a subset of streams (e.g., Canoga Creek, Fall Creek, Kashong Creek, Keuka Outlet, Yawger Creek; Table S4). For example, a 1% increase in pasture cover was associated with a -0.10 (CI=-0.13, -0.06) decrease for Kashong Creek, a stream in the Seneca Lake watershed.

In contrast to land cover, there was a strong relationship between *E. coli* levels and outfalls across scales of analysis. However, the direction of this relationship varied between waterways (e.g., Big Stream versus Fall Creek) and watersheds (e.g., Cayuga Lake versus Seneca Lake), and between scales of analysis (e.g., statewide versus watershed models; Table S4). For example, when an outfall was present, the log₁₀ CFUs/100-mL of *E. coli* increased by 0.39 (CI=0.30, 0.50), 0.85 (CI=0.66, 1.04), and 1.70 (IQR CI 1.57, 1.84) in the statewide, Seneca Lake watershed, and Keuka Outlet waterway models, respectively, but decreased by 0.17 (CI=-0.30, -0.06) and 0.77 (CI=-0.91, -0.62) in the Cayuga Lake watershed and Fall Creek waterway models, respectively (Table S4). The effect of sampling at the stream mouth (as opposed to upstream of the mouth) also varied with scale of analysis, the waterway/watershed under consideration, and water type. For example, *E. coli* levels were, on average, 0.91 (CI=0.79, 1.02) log₁₀ CFU/100-mL higher at the Keuka Outlet mouth (compared to upstream sites) but 0.76 (CI=-0.99, -0.48) log₁₀ CFU/100-mL lower at the Milliken Creek mouth (Table S4). Except for the all-lakes model, sampling at the stream mouth was associated with a decrease in *E. coli* levels at larger scales of analysis (e.g., statewide, all-stream, and watershed models); this may be due to the dilution effect as water moves from streams to stream mouths to lakes.

In models that included waterway type as a covariate, *E. coli* levels were substantially higher in streams and canals, and substantially lower in springs compared to lakes (Table S4). According to the statewide model and the Cayuga Lake watershed model, log₁₀ *E. coli* levels were 0.43 CFU/100-mL (CI=0.29, 0.57), and 0.37 CFU/100-mL (CI=0.24, 0.49), respectively, higher in streams compared to lakes. Similarly, log₁₀ *E. coli* levels were 0.27 CFU/100-mL (CI=0.10, 0.44), 0.19 CFU/100-mL (CI=0.03, 0.36), and 0.74 CFU/100-mL (CI=0.54, 0.95) higher in canals/flood control channels compared to lakes according to the statewide, all-lake, and Cayuga Lake watershed models, respectively. Lastly, in the statewide model, which included watershed as a covariate, *E. coli* levels were 0.69 log₁₀ CFU/100-mL (CI=0.54, 0.84) higher in samples collected from the Keuka Lake Watershed, 0.25 log₁₀ CFU/100-mL (CI=0.17, 0.33) higher in samples collected from the Seneca Lake Watershed, and 0.55 log₁₀ CFU/100-mL (CI=0.73, 0.38) lower in samples collected from the Susquehanna River watershed compared to samples collected from the Cayuga Lake watershed, which was the reference-level.

The distance at which microbial water quality became spatially independent increased as the scale of analysis increased but was also differed between models at the same scale. The effective range in the intercept-only MCOR models (the range at which *E. coli*, total coliform, TSS, or turbidity levels became spatially independent) varied substantially among models and scales of analysis (Table S2). According to the intercept-only MCOR, the effective range for *E. coli* was 778 m (N=385 samples; IQR=39, 1,511 m), 358 m (N=4,294 samples; IQR=259, 1,006 m), and 4 m (N=44 samples; IQR=3, 6 m) for the Seneca Lake, Cayuga Lake, and Susquehanna River watershed models (Table S2). The small effective range for the Susquehanna watershed may be an effect of the small number of Susquehanna watershed samples tested for *E. coli*; however, a similar number of Susquehanna watershed samples were

tested for TSS (N=47 samples; Median=59 m; IQR=41, 141 m) and turbidity (N=60 samples; Median=797 m; IQR=351, 2,799 m). The effective range for *E. coli* varied substantially between waterways as well. For instance, the effective range for *E. coli* in Cayuga Lake was 669 m (N=630 samples; IQR=498, 1,027 m) but was 117 m (N=65 samples; IQR=69, 279 m) in Keuka Lake and 6 m (N=84 samples; IQR = 4, 10 m) in Seneca Lake. While intensively sampled waterways with well-dispersed sampling sites, such as Cayuga Inlet (N=1,060 samples from 38 sites; Median = 96; IQR=57, 231 m) and Six Mile Creek (N=938 samples from 17 sites; Median=68 m; IQR=39, 120 m), often had small effective ranges for *E. coli*, several intensively sampled waterways with well-dispersed sampling sites had quite large effective ranges for *E. coli*, including Fall Creek (N=994 samples; Median = 721; IQR=464, 1,034) and Taughannock Creek (N=258 samples from 9 sites; Median=567 m; IQR=429, 843 m). Although the effective range for total coliforms, TSS, and turbidity similarly varied between waterways and watersheds (Table S2), the effective range for *E. coli* did not appear related to the effective range for total coliforms, TSS, or turbidity in a given waterway/watershed. For instance, the effective range for total coliforms, TSS and turbidity in Cayuga Inlet were 162 m (N=924 samples; IQR=110, 356 m), 7,754 m (N=989 samples; IQR=429, 843 m), and 3,931 m (N=1,053 samples; IQR=3,658, 3,989 m), respectively (Table S2). Despite this variability in effective range for a given parameter between waterways and watersheds, and between parameters for a given waterway/watershed it is important to note that for most waterways the effective range was consistently low (i.e., measured in the 10s or 100s of m as opposed to 1000s of m). Indeed, 64% (14/22) of *E. coli* stream models, 58% (11/19) of total coliform stream models, 57% (13/23) of TSS stream models, and 67% (14/21) of turbidity stream models had an effective range less than 100 m, while approx. 40% had an effective range <10 m, and 14% had an effective range >500 m. It is also important to note that in both the intercept-only and multivariable MCOR, effective range generally increased as spatial scale increased and, for total coliforms, TSS and turbidity, was generally larger for the lake as opposed to the stream models. For example, the median effective range for all *E. coli* intercept-only watershed MCOR was approx. 4.5 times greater (261 m) than the median effective range for all 22 *E. coli* stream MCOR (58 m).

Discussion

The scale of analysis strongly affected the observed sources of variance. The variance attributable to any given factor in the waterway and watershed models ranged widely with large differences among waterways. This finding, that sources of variation in microbial water quality were waterway-specific is supported by previous studies that reported waterway-specific findings and attributed this to the heterogeneity inherent to freshwater environments (e.g., (13, 26, 30, 53)). Despite these differences, as spatial scale increased (from the waterway to watershed to statewide models), the variance attributable to spatial sources generally increased and the variance attributable to temporal sources generally decreased in the present study. This finding is consistent with past studies that associated FIB levels with stream order/watershed size and concluded that observations at coarser scales could reveal patterns (and drivers of these patterns) not detectable at finer scales (34, 54–56). Conversely, the dilution of local effects when data are aggregated to coarser scales is well-established (55, 57). For example, a study of spatial autocorrelation in *E. coli* levels in the Norwalk River, CT, USA found that signals from point sources were rapidly diluted as water moved downstream, with *E. coli* levels being $\geq 90\%$ lower 1,000 m downstream of a potential source than at the source (36). Based on the effective ranges (the distance at which *E. coli* levels became spatially independent) in the present study, the distance at which dilution occurred varied substantially between waterways and across scales of analysis (again increasing with increasing scale), but was consistently below the 1,000 m reported previously (36). Overall, these findings highlight the importance of considering scale when designing monitoring programs. For example, the Norwalk River study and this study, both suggest that monitoring programs designed to identify point sources of contamination should place sampling sites $\leq 1,000$ m apart. However, such intensive sampling may be cost-prohibitive and should be balanced against affordability.

Composite sampling may provide an alternative way to account for spatial variability while reducing the cost of testing multiple samples (58–62). Indeed, multiple studies determined that compositing samples did not produce statistically different FIB counts or result in different management decisions compared to analyzing each sample individually (58, 59, 61, 62). Adapting the intensity of sampling based on the probability of contamination (more sites near point sources and fewer sites, spaced farther apart further from the sources) might also work.

In addition to considering the scale of analysis, monitoring programs should also be adapted to different types of waterways (e.g., lakes, streams) and sampling site (e.g., stream mouth, areas near outfalls). In the present study, more variance in microbial water quality was attributable to spatial factors in the lake compared to the streams models, possibly because the sampled lakes (e.g., Cayuga Lake, Seneca Lake) were considerably larger than the drainage basins of the sampled streams. However, more variance was attributable to spatial sources in the all-lake model than the all-stream model even though the all-stream model represents a larger area than the all-lake model, which may indicate that this finding is not solely an artifact of differences in drainage basin size. Indeed, these findings are supported by Partyka et al. (53), who compared FIB levels in six California reservoirs along a transect from the reservoirs' inflows to their outflows and found significant differences in *E. coli* coliform levels between inflow, outflow, and reservoir samples. These findings also make sense when microbial fate and transport and waterway dynamics are considered. Drivers of microbial contamination and factors that facilitate microbial survival are likely to differ between standing bodies of water (e.g., lakes) and flowing waterways (e.g., streams). For example, while sediments are important in-channel stores for microbial contaminants (63, 64), the influence of sediments on the water column is affected by waterway-specific characteristics, including particle size, flow rate, and depth (65, 66). Thus, the influence of sediments is likely less in the deeper, slower-moving lakes than the shallower, fast-flowing streams sampled here. For example, where Taughannock Creek enters Cayuga Lake depth changes from approx. 0.30-1.2 m at the stream-mouth to approx. 3-15 m in the near-shore area, and approx. 100 m at the lake center (based on topographic maps, usa.fishermapper.org/depth-map/cayuga-lake-ny/). As a result, resuspension of sediments can be expected to have a greater effect on microbial and sediment levels in Taughannock Creek compared to Cayuga Lake purely due to depth and dilution. Overall, our study suggests that different monitoring strategies may be needed to fully capture spatiotemporal variation and assess microbial water quality for different types of surface waterways, particularly if monitoring will be used to guide decisions with public health implications. Our results also suggest that one-size-fits-all regulatory standards may not be appropriate. Instead, standards tailored to specific water types may be appropriate; however, this requires further investigation to assess the efficacy of water type-specific standards.

Intra-annual variability accounted for more of the temporal signals observed here than interannual variability. This study assessed variation during events (baseflow versus stormflow sampling), over a year (represented by week-of-the-year, month, and season), and between years. The descriptive data visualizations and the random effects models showed limited interannual but considerable interannual variation in both FIB and sediment levels. Moreover, substantial variance was attributable to seasonal and event-scale variables. Indeed, in the multivariable lake, stream, and watershed models, the largest magnitudes of effect were attributed to month and flow event, while the effect of year was negligible. Past studies also reported strong seasonal (39, 67, 68), event-scale (39), and weekly patterns in microbial water quality. For instance, a review on sources of temporal variability in microbial water quality reported that event-scale variability was the greatest source with FIB levels changing significantly in response to rain events (39). Since seasonal, event-scale, and weekly patterns were observed by studies conducted in different regions and waterways using different sampling designs, microbial targets, and laboratory methods suggests that the temporal patterns observed here are reproducible.

Because of the dominance of event-scale and seasonal variation in FIB levels, multiple studies have suggested that compliance with the proposed *E. coli*-based FDA agricultural water quality standard (which is calculated using just 20 samples over a 2 to 4 year period) is highly dependent on when the

samples were collected (13, 25, 28). Thus, compliance with the proposed standard may not be indicative of water quality at time of water use, and compliance with the standard may not be an appropriate basis for guiding risk management decisions (e.g., when and where water should be used for irrigation or recreation). Instead, standards and associated monitoring programs, that account for seasonal and event-scale variability may be more useful for guiding these decisions. For instance, sampling could be targeted to conditions known to be associated with increased FIB levels (e.g., summer months, stormflow conditions) to provide a more conservative estimate of microbial contamination risks. Monitoring programs should also consider when and how the water is used. For example, agricultural water use in the Northeastern US, where the present study was conducted, is likely to peak in summer and fall. Monitoring programs that more intensely sample during summer and fall due to seasonal patterns in FIB levels (i.e., increased levels in these seasons), would also be targeting periods with an increased chance of human exposure (e.g., through irrigation). In considering how to adapt standards and monitoring programs, it is also important to note that, a single surface water source may be used in multiple ways that are weather-independent and unrelated to crop water needs, such as pesticide application, dust abatement, frost protection, and fertigation.

Our results also suggest an interesting connection to recreational activities since a non-negligible amount of variance was attributable to day of the week across all scales of analysis. FIB levels tended to be lower during the work week and higher over the weekend with the highest levels on Friday and in the Spring. Previous studies concluded that recreational activity (e.g., camping, hiking, swimming) on the weekends can affect the microbial quality of downstream surface water sources (69–73). Two studies conducted in Washington State found that fecal coliform and fecal streptococci concentrations in surface water were higher on weekends (when approx. 90% of recreational activities occurred) compared to weekdays (69, 70). All waterways sampled in the study reported here are in Upstate New York, which has large tourism and outdoor recreation industries. The authors personally observed campgrounds, hiking pull-offs, swimming areas, public fishing sites, and vacation rentals upstream of or near many of the sampling sites. However, weekly patterns in other human activities, such as agriculture, could also explain the variance attributable to weekday in the present study.

While the effect of land cover on *E. coli* levels was highly dependent on scale of analysis, the magnitude of the land cover effect was consistently smaller than the effect of outfall presence. The effect of land cover was largely dependent on scale of analysis and the waterway or watershed under consideration. The effect estimates for land cover factors in waterway models were mostly negligible, although there was a non-negligible effect for a subset of streams. Conversely, the magnitude of effect for wastewater outfall presence and site type (i.e., if the sampling site was a stream mouth or not) was consistently larger than the magnitude of effect for land cover factors. Overall, this is consistent with past studies, which also found that the effect of land cover was dependent on the spatial scale considered (37, 38, 74). One study of 40 Canadian rivers conducted to determine the scale at which land cover was associated with *E. coli* levels found that land cover and *E. coli* levels were most strongly associated at smaller spatial scales, between 5 and 10 km (38). A New Zealand study found that the relationship between land cover and FIB levels was dependent on stream size and the area considered when calculating land cover with local land cover being more strongly associated with FIB levels in small streams and watershed-scale land cover having a stronger influence on FIB levels in large streams (37). Based on these findings, the New Zealand study concluded that FIB monitoring programs need to be scale-sensitive; a conclusion supported by our findings and the Canadian study (38).

The effect estimate for outfall presence (a point feature) was greater than the effect estimates for all three land-use factors (cropland, pasture, and forest-wetland cover) at all scales of analysis considered in the present study. Other studies also found that impaired water quality was associated with wastewater and other outfalls presence (75–77), including a study that sampled several of the streams sampled in our study (12). A survey of the Seine watershed in France found that microbial inputs from point sources (e.g., wastewater outfalls) were substantially greater than non-point source inputs associated with land use

patterns under baseflow conditions, and recommended prioritization of remediation efforts to mitigate microbial and fecal inputs from point sources as opposed to overland run-off from non-point sources (77). The present study largely supports this conclusion. However, the direction of the relationship between point sources and *E. coli* levels was negative for a minority of models, possibly because outfall presence was confounded with sampling at the stream mouth or impoundments and pools. The aforementioned dilution effect may have obfuscated the point source effect at such sites (39). Indeed, in the present study sampling at a stream mouth or pond was associated with reduced *E. coli* levels compared to streams.

Throughout the 2010s, produce industry stakeholders have highlighted concerns about the agricultural water standard proposed as part of the Produce Safety Rule (29). One concern repeatedly raised at topical summits, conferences, and other venues was the lack of guidance on sampling, including how close together irrigation pumps that drew from the same source needed to be considered the same source for monitoring and compliance purposes (29). Thus, in considering the implications of our findings for monitoring programs, it is important to consider the distance at which microbial water quality became spatially independent. Although this distance increased as scale of analysis increased in the present study, it was variable and consistently low for most waterways (i.e., measurable in the 10s of m as opposed to 1000s of m). This indicates that FIB levels in samples collected relatively close together (e.g., < 1 km), would be spatially independent. Therefore, sampling at individual irrigation intakes, as opposed to treating a single sampling site as representative of water quality at multiple intakes, may be needed to ensure compliance with the proposed standard. Thus, the findings reported here not only highlight the importance of scale when designing monitoring studies, they also bring into question the utility of one-size-fits all standards and instead suggest that standards tailored to specific water types may be appropriate. However, this requires further investigation in follow-on studies and/or meta-analyses specifically aimed at assessing the efficacy of water type-specific standards.

- 536
- 537
- 538 1. Ecological Criteria Division UE. 2009. Review of Published Studies to Characterize
539 Relative Risks from Different Sources of Fecal Contamination in Recreational Water,
540 EPA 822-R-09-001, February 2009.
- 541 2. Barton Behravesh C, Mody RK, Jungk J, Gaul L, Redd JT, Chen S, Cosgrove S, Hedican
542 E, Sweat D, Chávez-Hauser L, Snow SL, Hanson H, Nguyen T-A, Sodha S V, Boore AL,
543 Russo E, Mikoleit M, Theobald L, Gerner-Smidt P, Hoekstra RM, Angulo FJ, Swerdlow
544 DL, Tauxe R V, Griffin PM, Williams IT. 2011. 2008 outbreak of Salmonella Saintpaul
545 infections associated with raw produce. N Engl J Med 364:918–927.
- 546 3. Ackers M, Mahon B, Leahy E, Goode B, Damrow T, Hayes P, Bibb W, Rice D, Barrett T,
547 Hutwagner L, Griffin P, Slutsker L. 1998. An outbreak of *Escherichia coli* O157:H7
548 infections associated with leaf lettuce consumption. J Infect Dis 177:1588–1593.
- 549 4. Lee S, Levy D, Craun G, Beach M, Calderon R. 2002. Surveillance for Waterborne-
550 Disease Outbreaks --- United States, 1999--2000.
- 551 5. Licence K, Oates KR, Synge BA, Reid TMS. 2001. An outbreak of *E. coli* O157 infection
552 with evidence of spread from animals to man through contamination of a private water
553 supply. Epidemiol Infect 126:135–138.
- 554 6. Isaac-Renton JL, Lewis LF, Ong CSL, Nulsen MF. 1994. A second community outbreak
555 of waterborne giardiasis in Canada and serological investigation of patients. Trans R Soc
556 Trop Med Hyg 88:395–399.
- 557 7. US FDA. Factors Potentially Contributing to the Contamination of Romaine Lettuce
558 Implicated in the Three Outbreaks of *E. coli* O157:H7 During the Fall of 2019 | FDA.
- 559 8. Food and Drug Administratio. 2018. FDA Investigated Multistate Outbreak of *E. coli*
560 O157:H7 Infections Linked to Romaine Lettuce from Yuma Growing Region. Center for
561 Food Safety and Applied Nutrition, Washington, D.C.
- 562 9. Bottichio L, Keaton A, Thomas D, Fulton T, Tiffany A, Frick A, Mattioli M, Kahler A,
563 Murphy J, Otto M, Tesfai A, Fields A, Kline K, Fiddner J, Higa J, Barnes A, Arroyo F,
564 Salvatierra A, Holland A, Taylor W, Nash J, Morawski BM, Correll S, Hinnenkamp R,
565 Havens J, Patel K, Schroeder MN, Gladney L, Martin H, Whitlock L, Dowell N, Newhart
566 C, Watkins LF, Hill V, Lance S, Harris S, Wise M, Williams I, Basler C, Gieraltowski L.
567 2019. Shiga Toxin–Producing *Escherichia coli* Infections Associated With Romaine
568 Lettuce—United States, 2018. Clin Infect Dis.
- 569 10. Gorski L, Rivadeneira P, Cooley MB. 2019. New strategies for the enumeration of enteric
570 pathogens in water. Environ Microbiol Rep 1758-2229.12786.
- 571 11. Uyttendaele M, Jaykus L-A, Amoah P, Chiodini A, Cunliffe D, Jacxsens L, Holvoet K,
572 Korsten L, Lau M, McClure P, Medema G, Samplers I, Rao Jasti P. 2015. Microbial
573 Hazards in Irrigation Water: Standards, Norms, and Testing to Manage Use of Water in
574 Fresh Produce Primary Production. Compr Rev Food Sci Food Saf 14:336–356.

- 575 12. Weller D, Belias A, Green H, Roof S, Wiedmann M. 2020. Landscape, Water Quality, and
576 Weather Factors Associated With an Increased Likelihood of Foodborne Pathogen
577 Contamination of New York Streams Used to Source Water for Produce Production. *Front*
578 *Sustain Food Syst* 3:124.
- 579 13. Weller D, Brassill N, Rock C, Ivanek R, Mudrak E, Roof S, Ganda E, Wiedmann M.
580 2020. Complex Interactions Between Weather, and Microbial and Physicochemical Water
581 Quality Impact the Likelihood of Detecting Foodborne Pathogens in Agricultural Water.
582 *Front Microbiol* 11.
- 583 14. Bell RL, Kase JA, Harrison LM, Balan K V., Babu U, Chen Y, Macarasin D, Kwon HJ,
584 Zheng J, Stevens EL, Meng J, Brown EW. 2021. The Persistence of Bacterial Pathogens
585 in Surface Water and Its Impact on Global Food Safety. *Pathog* 2021, Vol 10, Page 1391
586 10:1391.
- 587 15. ANZECC. 2000. Guidelines for Fresh and Marine Water Quality, 1st ed. Australian and
588 New Zealand Environment and Conservation Council, Agriculture and Resource
589 Management Council of Australia and New Zealand, Auckland, New Zealand, Australia
590 and New Zealand.
- 591 16. US FDA. 2015. Standards for the Growing, Harvesting, Packing, and Holding of Produce
592 for Human Consumption, Food Safety Modernization Act. United States.
- 593 17. Environmental Protection Agency. 2012. Recreational water quality criteria. Washington,
594 D.C., United States.
- 595 18. Health Canada. 2012. Guidelines for Canadian Recreational Water Quality Third Edition,
596 3rd ed. Ottawa, Canada, Canada.
- 597 19. UK EA. Bathing Water Quality.
- 598 20. EU Parliament. 2006. Bathing Water Quality Directive. Directive 2006/7/EC Official
599 Journal of the European Union.
- 600 21. SA DWAF. 1996. Water Quality Guidelines. South Africa.
- 601 22. FDA. 2013. United States Food and Drug Administration. Standards for the growing,
602 harvesting, packing, and holding of produce for human consumption.
- 603 23. California Leafy Greens Marketing Agreement. 2017. Commodity Specific Food Safety
604 Guidelines for the Production and Harvest of Lettuce and Leafy Greens. California Leafy
605 Green Handler Marketing Board, Sacramento, CA, California, United States.
- 606 24. Arizona Leafy Greens Marketing Agreement, California Leafy Greens Marketing
607 Agreement. 2012. Commodity specific food safety guidelines for the production and
608 harvest of lettuce and leafy greens. California Leafy Green Handler Marketing Board,
609 Phoenix, AZ, Arizona, United States.
- 610 25. Havelaar AH, Vazquez KM, Topalcengiz Z, Muñoz-Carpena R, Danyluk MD. 2017.
611 Evaluating the U.S. Food Safety Modernization Act Produce Safety Rule Standard for
612 Microbial Quality of Agricultural Water for Growing Produce. *J Food Prot* 80:1832–1841.

- 613 26. McEgan R, Mootian G, Goodridge LD, Schaffner DW, Danyluk MD. 2013. Predicting
614 *Salmonella* populations from biological, chemical, and physical indicators in Florida
615 surface waters. *Appl Environ Microbiol* 79:4094–4105.
- 616 27. Edge TA, El-Shaarawi A, Gannon V, Jokinen C, Kent R, Khan IUH, Koning W, Lapen D,
617 Miller J, Neumann N, Phillips R, Robertson W, Schreier H, Scott A, Shtepani I, Topp E,
618 Wilkes G, van Bochove E. 2012. Investigation of an Environmental Benchmark for
619 Waterborne Pathogens in Agricultural Watersheds in Canada. *J Environ Qual* 41:21.
- 620 28. Truitt LN, Vazquez KM, Pfunter RC, Rideout SL, Havelaar AH, Strawn LK. 2018.
621 Microbial Quality of Agricultural Water Used in Produce Preharvest Production on the
622 Eastern Shore of Virginia. *J Food Prot* 81:1661–1672.
- 623 29. Wall G, Clements D, Fisk C, Stoeckel D, Woods K, Bihn E. 2019. Meeting Report: Key
624 Outcomes from a Collaborative Summit on Agricultural Water Standards for Fresh
625 Produce. *Compr Rev Food Sci Food Saf* 18:723–737.
- 626 30. Francy DS, Stelzer EA, Duris JW, Brady AMGG, Harrison JH, Johnson HE, Ware MW.
627 2013. Predictive models for *Escherichia coli* concentrations at inland lake beaches and
628 relationship of model variables to pathogen detection. *Appl Environ Microbiol* 79:1676–
629 88.
- 630 31. Green H, Wilder M, Wiedmann M, Weller D. 2021. Integrative Survey of 68 Non-
631 overlapping Upstate New York Watersheds Reveals Stream Features Associated With
632 Aquatic Fecal Contamination. *Front Microbiol* 0:2125.
- 633 32. Partyka ML, Bond RF, Chase JA, Atwill ER. 2018. Spatiotemporal variability in
634 microbial quality of western US agricultural water supplies: A multistate study. *J Environ*
635 *Qual* 47:939–948.
- 636 33. Rao G, Eisenberg J, Kleinbaum D, Cevallos W, Trueba G, Levy K, Rao G, Eisenberg
637 JNS, Kleinbaum DG, Cevallos W, Trueba G, Levy K. 2015. Spatial Variability of
638 *Escherichia coli* in Rivers of Northern Coastal Ecuador. *Water* 7:818–832.
- 639 34. Rafi K, Wagner KL, Gentry T, Karthikeyan R, Dube A. 2018. *Escherichia coli*
640 Concentration as a Function of Stream Order and Watershed Size. *J Environ Qual* 47:949–
641 957.
- 642 35. Wyer MD, Kay D, Morgan H, Naylor S, Clark S, Watkins J, Davies CM, Francis C,
643 Osborn H, Bennett S. 2018. Within-day variability in microbial concentrations at a UK
644 designated bathing water: Implications for regulatory monitoring and the application of
645 predictive modelling based on historical compliance data. *Water Res X* 1:100006.
- 646 36. Crosby SC, Spiller NC, Tietz KE, Cooper JR, Fraboni PJ. 2019. Temporal and spatial
647 variability of instream indicator bacteria (*Escherichia coli*) and implications for water
648 quality monitoring. *Environ Monit Assess* 191.
- 649 37. Buck O, Niyogi DK, Townsend CR. 2004. Scale-dependence of land use effects on water
650 quality of streams in agricultural catchments. *Environ Pollut* 130:287–299.
- 651 38. Hurley T, Mazumder A. 2013. Spatial scale of land-use impacts on riverine drinking
652 source water quality. *Water Resour Res* 49:1591–1601.

- 653 39. Environmental Protection Agency U, of Water O. 2010. Sampling and Consideration of
654 Variability (Temporal and Spatial) For Monitoring of Recreational Waters.
- 655 40. Quilliam RS, Clements K, Duce C, Cottrill SB, Malham SK, Jones DL. 2011. Spatial
656 variation of waterborne *Escherichia coli* – implications for routine water quality
657 monitoring. *J Water Health* 9:734–737.
- 658 41. Weller DL, Love TMT, Belias A, Wiedmann M. 2020. Predictive models may
659 complement or provide an alternative to existing strategies for managing enteric pathogen
660 contamination of Northeastern streams used for produce production. *Front Sustain Food*
661 *Syst* 4:561517.
- 662 42. ESRI. 2014. ArcGIS Desktop: Release 10.2.2. 10.2.2. Environmental Systems Research
663 Institute, Redlands, CA.
- 664 43. R Core Development Team, Team. 2017. R: A Language and Environment for Statistical
665 Computing. 10.1. R Foundation for Statistical Computing, Vienna, Austria.
- 666 44. Wei T. 2013. corrplot: Visualization of a correlation matrix. R package version 0.73. 0.73.
- 667 45. Bates D, Maechler M, Bolker B, Walker S. 2014. lme4: Linear mixed-effects models
668 using Eigen and S4. R package version 1.1-7. *J Stat Softw*.
- 669 46. Barton K. 2016. MuMIn: Multi-Model Inference. R package version 1.15.6.
- 670 47. Harrand S, Weller D, Mercedes Illas-Ortiz P, Wiedmann M, Strawn L. Tracking and
671 modeling of *Listeria monocytogenes* contamination in spinach fields from planting to
672 harvest.
- 673 48. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer New York,
674 Dordrecht; New York.
- 675 49. Ishii S, Sadowsky MJ. 2008. *Escherichia coli* in the Environment: Implications for Water
676 Quality and Human Health. *Microbes Environ* 23:101–108.
- 677 50. Finley AO, Banerjee S, Carlin BP. 2007. spBayes: An R Package for Univariate and
678 Multivariate Hierarchical Point-referenced Spatial Models. *J Stat Softw* 19:1–24.
- 679 51. Finley AO, Banerjee S, Gelfand AE. 2015. spBayes for Large Univariate and Multivariate
680 Point-Referenced Spatio-Temporal Data Models. *J Stat Softw* 63:1–28.
- 681 52. Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output
682 analysis for MCMC - Open Research Online. *R News* 6.
- 683 53. Partyka ML, Bond RF, Chase JA, Atwill ER. 2018. Spatial and temporal variability of
684 bacterial indicators and pathogens in six California reservoirs during extreme drought.
685 *Water Res* 129:436–446.
- 686 54. Dorner SM, Anderson WB, Gaulin T, Candon HL, Slawson RM, Payment P, Huck PM.
687 2007. Pathogen and indicator variability in a heavily impacted watershed. *J Water Health*
688 5:241–257.
- 689 55. Pachepsky YA, Allende A, Boithias L, Cho K, Jamieson R, Hofstra N, Molina M. 2018.

- 690 Microbial Water Quality: Monitoring and Modeling. *J Environ Qual* 47:931–938.
- 691 56. R.D. Harmel, R. Karthikeyan, T. Gentry, R. Srinivasan. 2010. Effects of Agricultural
692 Management, Land Use, and Watershed Scale on *E. coli* Concentrations in Runoff and
693 Streamflow. *Trans ASABE* 53:1833–1841.
- 694 57. Wiens JA. 1989. Spatial scaling in ecology. *Funct Ecol* 3:385–197.
- 695 58. Reicherts JD, Emerson CW. 2009. Monitoring bathing beach water quality using
696 composite sampling. *Environ Monit Assess* 2009 1681 168:33–43.
- 697 59. Kinzelman JL, Dufour AP, Wymer LJ, Rees G, Pond KR, Bagley RC. 2009. Comparison
698 of Multiple Point and Composite Sampling for Monitoring Bathing Water Quality.
699 <http://dx.doi.org/101080/07438140609353887> 22:95–102.
- 700 60. Lothrop N, Bright KR, Sexton J, Pearce-Walker J, Reynolds KA, Verhoughstraete MP.
701 2018. Optimal strategies for monitoring irrigation water quality. *Agric Water Manag*
702 199:86–92.
- 703 61. Reicherts J. 2008. Using Composite Sampling Techniques to Monitor Bathing Beach
704 Water Quality in Kalamazoo County, Michigan. Masters Theses.
- 705 62. Bertke EE. 2007. Composite Analysis for *Escherichia coli* at Coastal Beaches. *J Great*
706 *Lakes Res* 33:335–341.
- 707 63. Nagels JW, Davies-Colley RJ, Donnison AM, Muirhead RW. 2002. Faecal contamination
708 over flood events in a pastoral agricultural stream in New Zealand. *Water Sci Technol*
709 45:45–52.
- 710 64. Muirhead RW, Davies-Colley RJ, Donnison AM, Nagels JW. 2004. Faecal bacteria yields
711 in artificial flood events: quantifying in-stream stores. *Water Res* 38:1215–1224.
- 712 65. Zhou K, Sassi HP, Morrison CM, Duan JG, Gerba CP. 2017. Resuspension of
713 *Escherichia coli* and MS2 Bacteriophage from Bed Sediment in Irrigation Canals . *J Irrig*
714 *Drain Eng* 143.
- 715 66. Fluke J, González-Pinzón R, Thomson B. 2019. Riverbed Sediments Control the
716 Spatiotemporal Variability of *E. coli* in a Highly Managed, Arid River. *Front Water* 1:4.
- 717 67. Crump BC, Peterson BJ, Raymond PA, Amon RMW, Rinehart A, McClelland JW,
718 Holmes RM. 2009. Circumpolar synchrony in big river bacterioplankton. *Proc Natl Acad*
719 *Sci U S A* 106:21208–21212.
- 720 68. Whitlock JE, Jones DT, Harwood VJ. 2002. Identification of the sources of fecal
721 coliforms in an urban watershed using antibiotic resistance analysis. *Water Res* 36:4273–
722 4282.
- 723 69. Varness KJ, Pacha RE, Lapen RF. 1978. Effects of dispersed recreational activities on the
724 microbiological quality of forest surface water. *Appl Environ Microbiol* 36:95–104.
- 725 70. Christensen H, Pacha R, Varness K, Lapen R. 1978. Human use in a dispersed recreation
726 area and its effect on water quality, p. 107–119. *In* Proceedings, recreational impact on
727 wildlands. U.S. Department of Agriculture, Seattle, WA.

- 728 71. Dasher DH, Urban L V, Dvoracek MJ, Fish EB, Abstract A. 1981. Effects of Recreation
729 on Water Quality in Guadalupe Mountains National Park.
- 730 72. Hendry GS, Leggatt EA. 1982. Some effects of shoreline cottage development on lake
731 bacteriological water quality. *Water Res* 16:1217–1222.
- 732 73. Flack JE, Medine AJ, Hansen-Bristow KJ. 1988. Stream Water Quality in a Mountain
733 Recreation Area. *Mt Res Dev* 8:11.
- 734 74. Sliva L, Williams DD. 2001. Buffer zone versus whole catchment approaches to studying
735 land use impact on river water quality. *Water Res* 35:3462–3472.
- 736 75. Staley ZR, Grabuski J, Sverko E, Edge TA. 2016. Comparison of microbial and chemical
737 source tracking markers to identify fecal contamination sources in the Humber River
738 (Toronto, Ontario, Canada) and associated storm water outfalls. *Appl Environ Microbiol*
739 82:6357–6366.
- 740 76. Sauer EP, VandeWalle JL, Bootsma MJ, McLellan SL. 2011. Detection of the human
741 specific *Bacteroides* genetic marker provides evidence of widespread sewage
742 contamination of stormwater in the urban environment. *Water Res* 45:4081–4091.
- 743 77. Garcia-Armisen T, Servais P. 2007. Respective contributions of point and non-point
744 sources of *E. coli* and enterococci in a large urbanized watershed (the Seine river, France).
745 *J Environ Manage* 82:512–518.

746

Table 1: Number of samples tested for the four focal water quality parameters of the present study collected from each water type and overall.

	Lakes ^a		Streams		Total		Enumeration Method
	Samples	Sites	Samples	Sites	Samples	Sites	
<i>E. coli</i>	1,738	92	4,025	158	5,763	250	EPA Method 1604 ^b
Total Coliforms	1,166	78	3,375	148	4,541	226	EPA Method 1604 ^b
Total Suspended Solids	935	50	3,864	153	4,799	203	SM 2450 D-2011 ^c
Turbidity	1,128	57	3,742	156	4,870	213	EPA Method 180.1 ^d

^a In addition to streams and lakes, the CSI collected samples from canals/flood control channels and ponds. All sampled canals linked two lakes, drew water from a lake, served as a flood control channel for a lake, or connected a lake to another waterway. Therefore, canal data were grouped with the lake data. Due to the small number of ponds (N=2), pond data were pooled with lake data as well. In the present table, 114, 99, 113, and 115 of the “lake” samples correspond to samples collected from canals/flood control samples and used for *E. coli*, total coliform, total suspended solids, and turbidity enumeration, respectively. Eight of the “lake” samples correspond to samples collected from a pond and used for *E. coli*, total coliform, total suspended solids, and turbidity enumeration.

^b https://www.epa.gov/sites/default/files/2015-08/documents/method_1604_2002.pdf

^c <https://www.standardmethods.org/doi/10.2105/SMWW.2882.030>

^d https://www.epa.gov/sites/default/files/2015-08/documents/method_180-1_1993.pdf

Figure 1: The top map depicts the watersheds for the sampled streams (plum) as well as lakes (blue) and cities (black dots) in the study region. The four maps along the bottom show, from left to right, the average relative concentration of *E. coli*, total coliforms, total suspended solids, and turbidity in samples collected from each canal, lake, and stream site. Larger bubbles indicate higher concentration, smaller bubbles indicate lower concentration. Some samples were not tested for all analytes, so the bottom maps cover a smaller area than the top map. Watersheds omitted from the smaller maps had no data for the four focal parameters of this study.

Figure 2: Distribution of log₁₀ *E. coli*, total coliform, total suspended solids, and turbidity for baseflow versus stormflow events, each type of water source, and each HUC-4/Finger Lake watershed providing data for one or more of the four focal water quality parameters (C = Cayuga Lake, K = Keuka Lake, Se = Seneca Lake, Su = Susquehanna River).

Figure 3: Variance in key water quality parameters in individual lakes (L), streams (S), and HUC-4/Finger Lake watersheds (W) accounted for when select spatial or temporal factors were included in random effects models (see Fig S4 for the full results). Results are only shown for lakes, streams, and watersheds with at least 20 samples. The variance in water quality for all lakes sampled, all streams, and all samples is represented by a green star in the corresponding column. Week of year = No. of weeks since Jan 1st.

Figure 4: Results of a multivariable Bayesian models of coregionalization that assessed the association between environmental parameters and log₁₀ *E. coli* levels. The 1st and 3rd quartiles, and median values for each estimate are shown. The boxplots show the distribution of the estimates for individual streams, lakes, and HUC-4/Finger Lake watersheds, while the estimate for models built using the data from all lakes samples, all streams samples, and all samples is represented by a green star in the corresponding column. % spatial = percent of variance attributable to spatial effects. Urban, cropland, and pasture refer to the inverse-distance weighted proportion of the upstream watershed under these covers. TSS= Total suspended solids. The reference-level for mouth and outfall are these features being absent (as opposed to present), for flow event is baseflow (as opposed to stormflow), and for month is April.

Supplemental Materials:

Table S1: Summary statistics for biological, chemical, and physical water quality parameters in Upstate New York surface water sources sampled between 2000 and 2020; five land-use parameters are also summarized. The laboratory method used is cited in the final column.

Water Quality Parameter	No. of Samples From				Years	Summary Statistics					Method
	Canal	Lake	Stream	Total		Min	Max	Mean	Stand. Dev.	Med	
Biological Parameters ^a											
Chlorophyll (ug/L)	0	97	0	97	2009-2017	0.25	48.50	5.64	6.77	4.30	EPA Method 446.0
<i>E. coli</i> (CFU/100-mL)	114	1,624	4,025	5,763	2003-2020	0.00	5.33*10 ⁵	2.35*10 ⁴	1.53*10 ⁴	1.20*10 ²	EPA Method 1604
Fecal Coliform (CFU/100-mL)	0	0	168	168	2006-2019	10.00	1.05*10 ⁶	3.55*10 ⁴	1.14*10 ⁵	6.34*10 ²	SM 9222D-2006
Total Coliforms (CFU/100-mL)	99	1067	3,375	4,541	2002-2018	5.00	1.00*10 ⁶	2.08*10 ⁴	3.91*10 ⁴	7.50*10 ³	EPA Method 1604
Chemical Parameters											
Hardness											
Calcium (mg CaCO3/L)	0	15	35	50	2006-2009	0.00	230.00	126.25	61.30	119.50	Lamotte Kit 5858
Total (mg CaCO3/L)	95	446	7,954	8,495	2002-2020	3.60	1,259.00	133.41	86.57	120.00	SM 2340C-2011
Nitrogen ^b											
Ammonia (mg/L)	0	26	345	371	2002-2020	0.00	30.70	0.24	1.72	0.05	SM 4500-NH3 D-2011/E-2011
Nitrate (mg/L)	100	549	4,285	4,934	2002-2020	0.00	31.90	1.29	2.43	0.53	SM 4500-NH3/E-2011
Organic (mg/L)	48	335	2,964	3,347	2002-2020	0.02	65.60	0.68	1.80	0.38	SM 4500-NH3 D-2011/E-2011
Total (mg/L)	11	34	365	410	2018-2019	0.06	33.10	2.81	4.84	0.96	SM 4500-NH3 D-2011/E-2011
Oxygen											
Chemical Demand (mg/L)	3	9	264	276	2002-2018	4.60	200.00	20.43	21.24	13.50	SM 5220D-2018
Dissolved (mg/L)	0	363	5,065	5,428	2002-2020	0.40	18.70	8.92	2.22	9.00	Lamotte Kit 5860
Phosphorous ^c											
Soluble Reactive (mg/L)	115	878	4,439	5,432	2004-2020	0.20	3735.00	56.59	196.74	14.60	EPA Method 365.3
Total (mg/L)	116	882	4671	5669	2002-2020	0.60	9710.00	111.82	321.03	28.10	EPA Method 365.3
Radioactivity											

Alpha (pCi/L)	3	6	107	116	2012-2014	-0.93	5.52	0.98	1.04	0.67	EPA 900.00
Beta (pCi/L)	3	6	107	116	2012-2014	0.44	6.91	2.12	1.06	1.97	EPA 900.00
Alkalinity (mg CaCO ₃ /L)	95	431	3,315	3,841	2002-2020	2.43	554.00	129.09	63.98	120.00	SM 2320B-2011
Aluminum (mg/L)	0	3	3	4	2016-2018	0.02	0.18	0.09	0.08	0.08	EPA Method 200.7
Arsenic (mg/L)	2	2	2	6	2016-2018	0.00	0.02	0.02	0.01	0.02	EPA Method 200.8
Barium (mg/L)	3	11	94	108	2012-2018	0.01	1.20	0.06	0.13	0.04	EPA Method 200.8
Boron (mg/L)	0	2	5	7	2015-2019	0.04	0.21	0.07	0.06	0.04	EPA Method 200.7
Bromide (mg/L)	0	0	1	1	2014	0.05	0.05	0.05	NA	0.05	EPA Method 300.0
Cadmium (mg/L)	0	0	2	2	2018	0.00	0.00	0.00	0.00	0.00	EPA Method 200.8
Calcium (mg/L)	0	2	6	8	2016-2019	39.10	101.00	76.35	22.93	81.80	EPA Method 200.7
Chloride (mg/L)	111	627	4,214	4,952	2002-2020	0.82	3666.00	41.92	83.33	28.00	SM 4500-CL-C-2011
Chromium (mg/L)	2	2	2	6	2016-2018	0.00	0.01	0.00	0.01	0.00	EPA Method 200.8
Conductivity (uS/cm)	109	599	9,157	9,865	2002-2020	3.20	10350.00	364.64	279.55	322.00	SM 2510B-2011
Copper (mg/L)	2	2	2	6	2016-2018	0.00	0.00	0.00	0.00	0.00	EPA Method 200.8
Fluoride (mg/L)	0	0	3	3	2019	0.13	0.20	0.15	0.04	0.13	EPA Method 300.0
Iron (mg/L)	0	2	2	4	2016-2018	0.01	0.23	0.12	0.09	0.11	EPA Method 200.8
Lead (mg/L)	0	0	2	2	2018	0.00	0.00	0.00	0.00	0.00	EPA Method 200.8
Magnesium (mg/L)	0	2	0	2	2016	11.60	12.80	12.20	0.85	12.20	EPA Method 200.7
Manganese (mg/L)	0	2	2	4	2016-2018	0.01	0.03	0.02	0.01	0.02	EPA Method 200.7
Mercury (mg/L)	2	2	2	6	2016-2018	0.00	0.00	0.00	0.00	0.00	EPA Method 254.1
pH	97	542	8,306	8,945	2002-2020	6.00	9.50	7.66	0.52	7.75	pH Meter
Selenium (mg/L)	0	2	2	4	2016-2018	0.02	0.02	0.02	0.00	0.02	EPA Method 200.8
Sodium (mg/L)	0	2	2	4	2016-2018	34.90	125.00	64.97	41.79	50.00	EPA Method 200.7
Strontium (mg/L)	3	9	92	104	2012-2018	0.04	1.00	0.12	0.11	0.10	EPA Method 200.7

Sulfate (mg/L)	9	32	463	504	2002-2019	2.50	174.00	13.68	10.05	11.50	SM 4500-SO4 E-2011
Surfactants (mg/L)	3	6	92	101	2012-2014	0.02	0.12	0.03	0.02	0.02	SM 5540C-2018
Land Cover (Inverse-Distance Weighted % in Upstream Watershed)											
Cropland	0	0	9,614	9,614	2002-2020	0.00	80.18	10.73	16.30	4.05	-
Developed	0	0	9,614	9,614	2002-2020	1.02	71.66	9.75	8.75	6.65	-
Forest-Wetland	0	0	9,614	9,614	2002-2020	0.00	98.75	54.18	18.36	55.87	-
Pasture	0	0	9,614	9,614	2002-2020	0.00	60.60	24.06	11.13	22.52	-
Physical											
Sediment											
Secchi Disk Depth (m)	0	153	0	153	2007-2019	0.75	7.30	3.22	1.10	3.00	
Total Dissolved Solids (mg/L)	3	10	129	142	2003-2019	0.00	1505.00	248.80	139.74	225.00	SM 4540 C- 2011
Total Solids (mg/L)	0	2	221	223	2002-2020	0.63	772.00	200.64	146.67	203.00	SM 2540 B- 2011
Total Suspended Solids (mg/L)	113	821	3,864	4,799	2002-2020	0.00	3890.00	45.40	188.02	5.00	SM 2450 D- 2011
Turbidity (NTU)	115	1,013	3,742	4,870	2002-2020	0.08	4960.00	43.39	184.17	5.97	EPA Method 180.1
Water Temp. (C)	108	880	8,425	9,413	2002-2020	-8.40	31.00	11.29	7.13	11.20	Thermometer

^a Standard plate count was enumerated for one stream sample in 2012 and was 7.50 log₁₀ CFU/100-mL.

^b Nitrite was measured in 10 lake and 1 stream samples between 2006 and 2014. The mean, minimum, maximum, standard deviation and median were 0.90, 4.76, 1.40, 1.13, and 1.05 mg/L, respectively.

^c Total dissolved phosphorous was measured in 8 lake and 1 stream samples between 2006 and 2014. The mean, minimum, maximum, standard deviation and median were 5.3, 37.3, 9.5, 10.4, and 6.1 mg/L, respectively.

Table S2: Results from the intercept-only Bayesian models of coregionalization.

Scale of Analysis	Waterway ^a	Number of		Effective Range (meters) ^b			Partial Sill ^c			Nugget ^d			Variance Ratio ^e		
		Samples	Sites	Median	Quartiles		Median	Quartiles		Median	Quartiles		Median	Quartiles	
					1st	3rd		1st	3rd		1st	3rd		1st	3rd
<i>E. coli</i>															
Lake															
	Cayuga Lake	630	26	669	1027	498	0.03	0.03	0.06	0.66	0.63	0.68	19.81	24.95	12.40
	Dryden Lake	133	3	4	7	4	0.06	0.04	0.08	0.40	0.36	0.44	7.06	9.92	5.64
	Keuka Lake	65	6	117	279	69	0.26	0.23	0.34	0.69	0.63	0.78	2.63	2.78	2.29
	Seneca Lake	84	4	6	10	4	0.07	0.04	0.11	0.72	0.64	0.81	10.48	14.45	7.28
Stream															
	Big Stream	135	7	10	20	6	0.33	0.09	0.46	0.87	0.67	0.99	2.60	7.71	2.14
	Burroughs Creek	25	5	32	64	22	0.05	0.03	0.08	0.64	0.51	0.84	13.58	15.57	10.08
	Canoga Creek	26	3	214	408	136	0.28	0.18	0.45	0.50	0.42	0.63	1.77	2.27	1.40
	Catharine Creek	121	8	319	596	200	0.12	0.09	0.14	0.43	0.39	0.48	3.61	4.26	3.39
	Cayuga Inlet	1060	38	96	231	57	0.06	0.05	0.07	0.63	0.61	0.65	11.36	13.20	9.75
	Direct Streams	266	15	544	1306	355	0.04	0.03	0.06	0.95	0.89	1.01	21.81	27.01	16.35
	Deans Creek	66	4	5	7	4	0.06	0.04	0.09	1.01	0.90	1.13	16.66	22.92	12.97
	Mill Creek	74	4	181	316	131	0.04	0.03	0.05	0.92	0.83	1.02	23.96	32.98	18.74
	Paines Creek	71	4	285	570	183	0.04	0.03	0.06	0.87	0.78	0.98	20.34	26.74	17.46
	Fall Creek	994	26	721	1034	464	0.13	0.12	0.16	0.51	0.49	0.52	3.79	4.26	3.37
	Great Gully	17	3	5	9	4	0.27	0.19	0.37	0.48	0.36	0.67	1.75	1.90	1.83
	Kashong Creek	37	3	6	11	4	0.07	0.05	0.13	0.82	0.69	0.97	11.65	15.31	7.17
	Keuka Outlet	115	8	124	217	71	0.39	0.33	0.51	0.67	0.60	0.73	1.70	1.82	1.41
	Milliken Creek	22	5	5	7	4	0.22	0.18	0.26	0.01	0.01	0.02	0.06	0.04	0.09
	Pleasant Grove Brook	60	2	47	104	26	0.18	0.07	0.41	0.42	0.35	0.49	2.30	4.90	1.19
	Pony Hollow Creek	66	6	5	8	3	0.09	0.07	0.16	0.38	0.32	0.44	4.32	4.66	2.79
	Reeder Creek	82	5	6	12	4	0.04	0.03	0.07	0.56	0.49	0.62	14.89	18.57	8.37
	Salmon Creek	298	14	5	9	4	0.05	0.03	0.10	0.64	0.57	0.69	12.06	17.89	6.72
	Six Mile Creek	938	17	68	120	39	0.04	0.03	0.04	0.56	0.54	0.58	14.53	19.77	12.86
	Taughannock Creek	258	9	567	843	429	0.02	0.02	0.03	0.67	0.63	0.72	29.73	34.48	24.37

Watershed	Trumansburg Creek	286	9	98	159	59	0.09	0.07	0.11	1.09	1.03	1.17	12.44	15.36	10.42
	Yawger Creek	36	5	5	11	4	0.11	0.08	0.17	0.64	0.53	0.77	5.64	6.96	4.46
	Cayuga Lake	4294	157	358	1006	259	0.15	0.13	0.16	0.63	0.62	0.65	4.20	4.77	4.06
	Keuka Lake	115	8	164	308	90	0.28	0.21	0.39	0.70	0.63	0.76	2.51	3.02	1.94
	Seneca Lake	385	24	778	1511	349	0.13	0.10	0.17	0.75	0.71	0.79	5.74	6.76	4.67
	Susquehanna River	66	6	4	6	3	0.15	0.11	0.20	0.34	0.30	0.40	2.28	2.73	1.97
Total Coliforms															
Lake	Cayuga Lake	541	19	1151	2429	712	0.03	0.02	0.05	0.43	0.41	0.44	13.26	17.02	8.25
	Dryden Lake	120	3	4	7	4	0.06	0.04	0.08	0.40	0.36	0.44	7.06	9.92	5.64
	Keuka Lake	37	4	195	332	121	0.10	0.06	0.18	0.23	0.20	0.26	2.24	3.07	1.47
	Seneca Lake	54	4	124	262	87	0.10	0.07	0.13	0.32	0.28	0.35	3.15	4.02	2.59
Stream	Big Stream	90	7	5	9	4	0.03	0.02	0.05	0.41	0.36	0.46	15.86	16.84	9.13
	Burroughs Creek	16	4	6	9	4	0.41	0.22	0.58	0.07	0.02	0.25	0.17	0.10	0.44
	Canoga Creek	19	3	453	594	386	16.59	12.40	23.23	0.07	0.03	0.13	0.00	0.00	0.01
	Catharine Creek	87	6	408	615	280	0.07	0.05	0.10	0.23	0.21	0.25	3.07	3.97	2.49
	Cayuga Inlet	924	38	162	356	110	0.03	0.02	0.03	0.45	0.44	0.47	17.43	19.77	14.85
	Direct Streams	193	15	1635	2571	1111	0.08	0.04	0.27	0.55	0.52	0.59	7.05	11.85	2.18
	Deans Creek	47	4	5	9	4	0.08	0.04	0.17	0.55	0.43	0.65	7.02	10.31	3.77
	Mill Creek	54	4	332	665	219	0.03	0.02	0.06	0.39	0.35	0.45	11.43	15.22	7.47
	Paines Creek	52	4	25	55	16	0.06	0.04	0.07	0.51	0.46	0.58	9.29	11.59	7.87
	Fall Creek	929	26	862	1076	741	0.14	0.12	0.17	0.44	0.42	0.45	3.16	3.53	2.66
	Kashong Creek	24	3	4	6	4	0.50	0.40	0.60	0.02	0.01	0.06	0.05	0.02	0.10
	Keuka Outlet	68	7	7	19	4	0.17	0.08	0.24	0.21	0.11	0.32	1.23	1.28	1.32
	Pony Hollow Creek	48	6	5	9	4	0.03	0.02	0.04	0.16	0.14	0.19	5.14	5.58	4.74
	Reeder Creek	62	5	5	7	4	0.06	0.03	0.08	0.13	0.11	0.16	2.43	3.46	1.96
	Salmon Creek	240	14	5	9	4	0.05	0.02	0.12	0.46	0.39	0.51	9.38	17.45	4.28
	Six Mile Creek	823	16	50	193	19	0.02	0.01	0.03	0.38	0.37	0.40	15.85	25.33	13.11
	Taughannock Creek	208	8	799	1374	562	0.03	0.01	0.05	0.43	0.40	0.46	16.66	28.15	9.30
	Trumansburg Creek	250	9	5	10	4	0.04	0.03	0.07	0.65	0.61	0.69	16.40	23.90	10.20

Watershed	Yawger Creek	14	5	129	213	95	0.07	0.04	0.10	0.17	0.12	0.22	2.51	3.19	2.22
	Cayuga Lake	3675	142	28116	29383	26132	0.36	0.28	0.72	0.47	0.46	0.47	1.29	1.66	0.66
	Keuka Lake	68	7	36	93	8	0.10	0.07	0.14	0.25	0.21	0.28	2.40	3.06	2.03
	Seneca Lake	265	22	8	20	5	0.05	0.03	0.07	0.32	0.29	0.34	6.51	9.25	4.89
	Susquehanna River	48	6	4	5	3	0.16	0.07	0.20	0.02	0.01	0.13	0.11	0.17	0.65
Total Suspended Solids															
Lake	Cayuga Lake	610	24	718	1490	463	0.06	0.04	0.07	0.46	0.44	0.48	7.91	10.04	6.57
	Dryden Lake	141	3	215	372	148	0.04	0.03	0.07	0.22	0.20	0.24	5.09	7.97	3.52
	Keuka Lake	72	7	1253	2564	698	0.05	0.04	0.07	0.50	0.45	0.57	9.65	11.12	7.59
	Seneca Lake	87	5	9	22	5	0.03	0.02	0.09	0.78	0.70	0.87	25.06	33.22	9.30
Stream	Big Stream	131	7	652	1111	441	0.04	0.02	0.10	0.88	0.80	0.95	22.78	32.78	9.55
	Burroughs Creek	24	4	7	15	4	0.02	0.01	0.03	0.50	0.42	0.59	31.96	30.43	20.08
	Canoga Creek	22	3	343	574	241	0.04	0.03	0.07	0.65	0.55	0.82	16.63	21.33	11.71
	Catharine Creek	117	8	42	91	20	0.07	0.05	0.08	0.49	0.44	0.54	7.42	8.60	6.37
	Cayuga Inlet	989	31	7754	7868	6704	0.12	0.06	0.17	0.00	0.00	0.04	0.02	0.00	0.23
	Direct Streams	229	15	4	6	4	0.10	0.08	0.15	0.45	0.39	0.49	4.57	4.84	3.30
	Deans Creek	57	4	5	8	4	0.04	0.03	0.06	0.42	0.36	0.50	9.73	12.07	7.81
	Mill Creek	63	4	5	10	4	0.04	0.03	0.09	0.62	0.54	0.70	13.89	17.00	7.77
	Paines Creek	62	4	6	12	4	0.03	0.03	0.09	0.40	0.34	0.46	11.61	12.88	5.37
	Fall Creek	1047	24	795	1238	622	0.04	0.02	0.06	0.35	0.34	0.36	9.56	13.46	5.60
	Great Gully	17	3	9	21	4	0.06	0.05	0.08	0.55	0.39	0.77	9.15	8.29	10.15
	Kashong Creek	37	3	5	9	4	0.04	0.03	0.08	0.39	0.33	0.47	8.65	10.58	6.13
	Keuka Lake Tributaries	24	8	416	787	277	0.02	0.02	0.04	0.31	0.27	0.35	15.95	16.25	9.23
	Keuka Outlet	117	8	8	23	4	0.03	0.02	0.06	0.41	0.38	0.45	12.15	18.93	8.14
	Milliken Creek	22	5	23	32	18	0.26	0.07	0.45	0.07	0.03	0.26	0.26	0.46	0.57
	Pleasant Grove Brook	52	2	14	33	6	0.19	0.06	0.30	0.13	0.10	0.18	0.70	1.75	0.59
	Pony Hollow Creek	44	5	5	7	4	0.03	0.02	0.06	0.30	0.24	0.37	10.74	13.73	6.50
	Reeder Creek	81	4	149	286	99	0.06	0.04	0.08	0.52	0.47	0.60	9.00	11.12	7.44
	Salmon Creek	274	12	225	419	142	0.03	0.02	0.06	0.68	0.64	0.72	20.94	38.92	13.11

Watershed	Six Mile Creek	924	17	175	814	76	0.06	0.05	0.07	0.66	0.64	0.69	11.89	14.28	10.17
	Taughannock Creek	262	9	358	624	253	0.04	0.03	0.05	0.61	0.58	0.65	15.50	20.05	12.05
	Trumansburg Creek	222	7	20	76	6	0.07	0.05	0.10	0.53	0.49	0.58	8.04	10.16	5.79
	Yawger Creek	36	5	363	595	257	0.04	0.03	0.06	0.39	0.34	0.46	9.42	10.69	8.14
Turbidity Lake	Cayuga Lake	4162	143	5352	6668	4284	0.06	0.05	0.07	0.51	0.51	0.52	8.49	9.47	7.41
	Keuka Lake	141	16	841	1506	579	0.06	0.05	0.08	0.43	0.39	0.47	6.58	8.15	5.57
	Seneca Lake	377	24	171	383	106	0.04	0.03	0.06	0.65	0.62	0.69	15.04	18.62	11.91
	Susquehanna River	47	8	59	141	41	0.02	0.02	0.03	0.31	0.27	0.36	13.05	15.32	11.95
Stream	Cayuga Lake	632	24	394	912	257	0.09	0.07	0.11	0.36	0.34	0.37	4.04	4.85	3.37
	Dryden Lake	144	3	194	444	116	0.04	0.03	0.06	0.14	0.13	0.15	3.33	4.09	2.50
	Keuka Lake	8	3	6	10	4	0.05	0.04	0.06	0.91	0.65	1.24	18.65	18.64	19.08
	Seneca Lake	26	3	5	11	4	0.51	0.35	0.67	0.10	0.02	0.24	0.20	0.05	0.36
Stream	Big Stream	43	6	248	448	182	0.07	0.04	0.09	0.61	0.54	0.68	9.16	13.56	7.99
	Burroughs Creek	25	5	62	165	29	0.07	0.04	0.12	0.24	0.08	0.36	3.55	1.88	3.01
	Canoga Creek	22	3	5	8	4	0.23	0.09	0.56	0.15	0.02	0.42	0.65	0.24	0.75
	Catharine Creek	32	4	485	726	356	0.06	0.04	0.13	0.06	0.04	0.13	1.00	1.00	1.00
	Cayuga Inlet	1053	34	3931	3989	3658	0.08	0.02	0.13	0.00	0.00	0.02	0.01	0.00	0.14
	Direct Streams	235	16	4	5	3	0.07	0.05	0.12	0.39	0.34	0.43	5.74	6.59	3.49
	Deans Creek	55	4	61	104	43	0.06	0.04	0.08	0.37	0.32	0.43	6.37	7.95	5.23
	Mill Creek	66	4	5	9	4	0.04	0.02	0.05	0.51	0.44	0.57	14.25	20.26	11.23
	Paines Creek	64	4	6	12	4	0.03	0.02	0.05	0.37	0.33	0.42	11.63	15.90	7.86
	Fall Creek	1075	25	1025	1597	767	0.03	0.02	0.04	0.23	0.23	0.24	6.95	9.13	5.94
	Great Gully	13	3	5	10	4	0.04	0.03	0.07	0.44	0.25	1.47	10.42	8.50	21.62
	Keuka Outlet	15	5	6	10	4	0.02	0.01	0.03	0.75	0.57	0.87	32.72	38.85	26.27
	Milliken Creek	22	5	6	11	4	0.06	0.04	0.08	0.04	0.03	0.08	0.75	0.67	1.01
	Pleasant Grove Brook	70	11	18	28	14	0.60	0.46	0.80	0.01	0.00	0.01	0.01	0.01	0.01
	Pony Hollow Creek	55	5	5	8	4	0.03	0.02	0.07	0.19	0.16	0.22	6.10	8.02	3.17
	Reeder Creek	34	5	248	448	182	0.07	0.04	0.09	0.61	0.54	0.68	9.16	13.56	7.99

Watershed	Salmon Creek	255	12	6	10	5	0.08	0.03	0.10	0.51	0.48	0.56	6.28	14.71	5.48
	Six Mile Creek	985	16	1760	5143	392	0.10	0.08	0.12	0.50	0.49	0.52	5.10	5.86	4.20
	Taughannock Creek	264	9	429	831	244	0.03	0.02	0.04	0.48	0.45	0.51	16.49	19.92	12.57
	Trumansburg Creek	227	7	9	19	5	0.05	0.04	0.07	0.41	0.38	0.44	7.58	10.77	6.09
	Yawger Creek	29	5	60	105	42	0.10	0.07	0.15	0.23	0.19	0.27	2.29	2.73	1.75
	Cayuga Lake	4283	158	4791	5814	3847	0.07	0.06	0.08	0.39	0.38	0.39	5.86	6.78	5.20
	Keuka Lake	15	5	5	7	4	0.25	0.07	0.47	0.41	0.09	0.70	1.61	1.31	1.50
	Seneca Lake	109	15	421	858	262	0.07	0.05	0.10	0.43	0.39	0.47	6.55	8.65	4.85
	Susquehanna River	60	10	797	2799	351	0.07	0.06	0.10	0.23	0.20	0.26	3.10	3.36	2.60

^a Because intercept-only Bayesian models of coregionalization aimed to quantify the variance ratio (non-spatial/spatial variance), individual waterway models were not implemented for the four waterways where all samples were collected from a single site, Chapel Pond, Lake Ridge Creek, Glen Eldridge Creek, or Townline Creek; data for these creeks were used in the watershed and statewide models.

^b Effective range is the distance in meters at which *E. coli* levels became spatially independent.

^c Partial sill = spatial variance.

^d Nugget = non-spatial variance.

^e Variance Ratio = variance attributable to non-spatial versus spatial sources (Nugget/Partial Sill)

Table S3: Results from the random effects models; blank cells indicate that for the given combination of parameter, scale of analysis, and waterway that the model could not be implemented for the random effect in the column. For instance, waterway could not be included as a random effect for the waterway models. Similarly, all of the sites along Big Stream were in a single county so county could not be included as a random effect for the Big Stream models.

Outcome	Scale of Analysis	Waterway	Random Effect									
			County	Site	Waterway	Watershed	Water Type	Day of the Week	Month	Season	No. of Weeks Since Jan. 1 st	Year
<i>E. coli</i>												
	Statewide											
		Statewide	0.07	0.22	0.12	0.17	0.17	0.10	0.14	0.13	0.22	0.06
		All-Lake	0.15	0.50	0.21	0.20	0.28	0.16	0.03	0.01	0.17	0.05
		All-Stream	0.09	0.23	0.14	0.22	0.40	0.10	0.14	0.13	0.22	0.06
	Lake											
		Cayuga Lake	0.34	0.05			0.00	0.09	0.16	0.15	0.29	0.05
		Chapel Pond									0.02	
		Keuka Lake		0.30			0.00	0.32	0.09	0.13	0.23	0.00
		Seneca Lake	0.03	0.20				0.43	0.11	0.05	0.16	0.00
	Stream											
		Big Stream		0.07				0.42	0.27	0.24	0.37	0.09
		Burroughs Creek		0.00				0.18	0.46	0.33	0.66	0.05
		Canoga Creek		0.66				0.21	0.07	0.00	0.08	0.07
		Catharine Creek	0.00	0.19				0.23	0.13	0.15	0.19	0.06
		Cayuga Inlet		0.07			0.09	0.22	0.4	0.21	0.52	0.16
		Direct Streams		0.00				0.31	0.63	0.52	0.68	0.15
		Deans Creek		0.00				0.18	0.71	0.60	0.72	0.13
		Mill Creek		0.00			0.13	0.37	0.7	0.47	0.75	
		Paines Creek		0.00				0.25	0.57	0.51	0.67	0.15
		Fall Creek	0.01	0.27			0.37	0.14	0.12	0.10	0.23	0.05
		Great Gully		0.00				0.16	0.22	0.17	0.51	0.00
		Kashong Creek		0.00				0.80	0.31	0.13	0.77	0.00
		Keuka Outlet		0.40			0.02	0.30	0.10	0.14	0.25	0.00
		Milliken Creek		0.03				0.49	0.49	0.56	0.51	0.05

Total Coliforms	Watershed	Pleasant Grove Brook	0.76				0.05	0.00	0.00	0.46	0.03	
		Pony Hollow Creek	0.04				0.28	0.41	0.23	0.44	0.32	
		Reeder Creek	0.01				0.33	0.4	0.31	0.54	0.02	
		Salmon Creek	0.00	0.01			0.08	0.29	0.09	0.60	0.42	
		Six Mile Creek		0.05			0.33	0.26	0.19	0.52	0.16	
		Taughannock Creek	0.00	0.00		0.00	0.11	0.25	0.19	0.58	0.09	
		Trumansburg Creek	0.00	0.20			0.03	0.03	0.05	0.19	0.09	
		Yawger Creek		0.00			0.36	0.00	0.05	0.84	0.4	
	Lake	Cayuga Lake	0.05	0.41	0.22		0.28	0.12	0.04	0.03	0.14	0.06
		Keuka Lake		0.40			0.02	0.297	0.10	0.14	0.25	0.00
		Seneca Lake	0.11	0.2	0.13			0.49	0.25	0.19	0.38	0.04
		Susquehanna River		0.04				0.28	0.41	0.23	0.44	0.32
	Statewide	Statewide	0.04	0.13	0.08	0.15	0.31	0.05	0.14	0.13	0.29	0.21
		All-Lake	0.08	0.54	0.18	0.09	0.26	0.14	0.03	0.01	0.16	0.07
		All-Stream	0.05	0.13	0.09	0.18	0.56	0.04	0.14	0.15	0.30	0.22
	Lake	Cayuga Lake	0.00	0.03			0.00	0.08	0.14	0.09	0.28	0.19
		Dryden Lake		0.00				0.37	0.15	0.08	0.14	0.16
		Keuka Lake		0.26			0.04	0.368	0.20	0.30	0.40	0.11
		Seneca Lake	0.15	0.14				0.301	0.24	0.1	0.25	0.00
	Stream	Big Stream		0.00				0.44	0.50	0.43	0.56	0.17
		Burroughs Creek		0.53				0.10	0.17	0.00	0.55	0.51
		Canoga Creek		0.55				0.00	0.00	0.00	0.03	0.00
		Catharine Creek	0.00	0.10				0.09	0.3	0.08	0.41	0.05
		Cayuga Inlet		0.03			0.17	0.18	0.38	0.23	0.5	0.33
		Direct Streams		0.00				0.21	0.48	0.43	0.67	0.61
		Deans Creek		0.00				0.00	0.55	0.53	0.70	0.63
		Mill Creek		0.00			0.47	0.36	0.46	0.35	0.73	

Total Suspended Solids	Watershed	Paines Creek	0.03				0.10	0.50	0.52	0.65	0.59
		Fall Creek	0.00	0.22		0.13	0.18	0.22	0.08	0.32	0.19
		Kashong Creek		0.00			0.73	0.36	0.39	0.90	0.00
		Keuka Outlet		0.33		0.05	0.28	0.21	0.35	0.48	0.24
		Pony Hollow Creek		0.01			0.46	0.57	0.14	0.75	0.26
		Reeder Creek		0.00			0.57	0.19	0.1	0.46	0.11
		Salmon Creek	0.02	0.03			0.11	0.38	0.06	0.64	0.38
		Six Mile Creek		0.03			0.30	0.28	0.3	0.62	0.13
		Taughannock Creek	0.00	0.00			0.11	0.53	0.32	0.67	0.16
		Trumansburg Creek	0.00	0.00			0.07	0.18	0.20	0.35	0.12
		Yawger Creek		0.09			0.00	0.83	0.43	0.82	0.38
Total Suspended Solids	Statewide	Cayuga Lake	0.04	0.43	0.20	0.33	0.13	0.04	0.04	0.20	0.15
		Keuka Lake		0.33		0.05	0.28	0.21	0.35	0.48	0.24
		Seneca Lake	0.06	0.07	0.06		0.5	0.37	0.26	0.51	0.13
		Susquehanna River		0.01			0.46	0.57	0.14	0.75	0.26
		Statewide	0.01	0.11	0.04	0.01	0.18	0.22	0.06	0.01	0.2
		All-Lake	0.08	0.21	0.10	0.06	0.18	0.25	0.12	0.01	0.17
		All-Stream	0.00	0.08	0.03	0.00	0.00	0.23	0.06	0.01	0.20
	Lake	Cayuga Lake	0.12	0.12		0.05	0.27	0.18	0.01	0.26	0.06
		Dryden Lake		0.06			0.00	0.08	0.02	0.13	0.02
		Keuka Lake		0.29		0.16	0.46	0.00	0.01	0.20	0.00
		Seneca Lake	0.06	0.06			0.59	0.02	0.01	0.22	0.01
	Stream	Big Stream		0.00			0.54	0.10	0.05	0.41	0.06
		Burroughs Creek		0.00			0.57	0.00	0.10	0.77	0.20
		Canoga Creek		0.00			0.79	0.12	0.00	0.89	0.45
		Catharine Creek	0.03	0.15			0.17	0.24	0.19	0.41	0.12
		Cayuga Inlet		0.24		0.31	0.48	0.25	0.00	0.43	0.16

Turbidity	Watershed	Direct Streams		0.01			0.30			0.56		0.53		0.50		0.23													
		Deans Creek		0.00			0.27			0.6		0.67		0.55		0.09													
		Mill Creek		0.00			0.20			0.25			0.57		0.47		0.62												
		Paines Creek		0.00						0.25			0.55		0.49		0.56		0.20										
		Fall Creek		0.10			0.08			0.06			0.17			0.25		0.06		0.32		0.04							
		Great Gully					0.00						0.08			0.13		0.09		0.71		0.00							
		Kashong Creek					0.00						0.76			0.11		0.00		0.85		0.00							
		Keuka Lake Tributaries					0.10						0.41			0.29				0.41									
		Keuka Outlet					0.23			0.10			0.53			0.02		0.02		0.31		0.00							
		Milliken Creek					0.10						0.00			0.29		0.13		0.28		0.00							
		Pleasant Grove Brook					0.89						0.27			0.00		0.00		0.00		0.14							
		Pony Hollow Creek					0.00						0.39			0.43		0.41		0.37		0.07							
		Reeder Creek					0.01						0.55			0.15		0.11		0.33		0.06							
		Salmon Creek		0.00			0.01						0.13			0.53		0.07		0.76		0.51							
		Six Mile Creek					0.10						0.32			0.15		0.04		0.48		0.11							
		Taughannock Creek		0.03			0.02			0.00			0.30			0.29		0.10		0.71		0.11							
		Trumansburg Creek		0.00			0.08						0.19			0.13		0.06		0.39		0.04							
		Yawger Creek					0.00						0.39			0.04		0.03		0.71		0.00							
		Turbidity	Statewide	Cayuga Lake		0.01			0.11			0.02			0.18			0.20			0.06		0.01		0.20		0.05		
				Keuka Lake		0.00			0.19			0.01			0.11			0.50			0.03		0.03		0.24		0.00		
				Susquehanna River		0.64			0.81									0.00			0.28		0.17		0.24		0.20		
Statewide				0.01			0.17			0.10			0.00			0.28			0.21			0.08		0.00		0.20		0.05	
All-Lake				0.10			0.48			0.22			0.00			0.32			0.16			0.23		0.09		0.25		0.05	
All-Stream				0.01			0.14			0.15			0.00			0.00			0.23			0.08		0.00		0.21		0.05	
Lake	Cayuga Lake			0.15			0.21			0.11			0.22			0.19			0.01		0.28		0.06						
	Dryden Lake					0.14						0.00			0.10			0.02		0.12		0.02							
	Seneca Lake		0.03			0.03						0.627			0.51			0.00		0.55		0.56							
	Stream																												

	Big Stream	0.00			0.53	0.41	0.12	0.64	0.62
	Burroughs Creek	0.02			0.52	0.11	0.00	0.82	0.22
	Canoga Creek	0.00			0.56	0.14	0.00	0.9	0.49
	Catharine Creek	0.00	0.27		0.70	0.38	0.04	0.32	0.05
	Cayuga Inlet	0.31		0.41	0.47	0.24	0.00	0.42	0.15
	Direct Streams	0.00			0.55	0.65	0.4	0.62	0.25
	Deans Creek	0.01			0.41	0.61	0.52	0.6	0.21
	Mill Creek	0.00		0.24	0.60	0.69	0.34	0.67	
	Paines Creek	0.00			0.48	0.71	0.55	0.69	0.21
	Fall Creek	0.11	0.11	0.14	0.18	0.26	0.04	0.32	0.03
	Great Gully	0.00			0.00	0.00	0.00	0.98	0.00
	Keuka Outlet	0.00		0.00	0.85	0.17	0.17	0.83	0.67
	Milliken Creek	0.28			0.15	0.07	0.13	0.00	0.03
	Pleasant Grove Brook	0.87			0.00	0.27	0.04	0.33	0.08
	Pony Hollow Creek	0.00			0.30	0.46	0.3	0.48	0.08
	Reeder Creek	0.00			0.41	0.50	0.09	0.48	0.11
	Salmon Creek	0.00	0.03		0.06	0.53	0.03	0.77	0.55
	Six Mile Creek	0.19			0.32	0.15	0.05	0.44	0.1
	Taughannock Creek	0.03	0.01	0.00	0.31	0.33	0.09	0.75	0.11
	Trumansburg Creek	0.00	0.03		0.21	0.11	0.03	0.48	0.01
	Yawger Creek	0.14			0.52	0.00	0.00	0.77	0.01
Watershed									
	Cayuga Lake	0.00	0.23	0.16	0.25	0.21	0.08	0.01	0.19
	Keuka Lake	0.00			0.00	0.85	0.17	0.17	0.83
	Susquehanna River	0.46	0.80		0.00	0.37	0.23	0.24	0.33

Table S4: Association between *E. coli* levels and environmental factors according to the multivariable Bayesian models of coregionalization.

Models	Median Coefficient Estimate (50% Credibility Interval)								
	Log ₁₀ Total Susp. Solids (mg/L)	Flow Event ^a (Baseflow= Reference)	IDW ^b Prop. of Upstream Watershed (%)			Waterway Type (Lake=Reference) ^c		Stream Mouth ^d (Absent= Reference)	Outfall ^e (Absent= Reference)
			Pasture	Crop	Dev.	Canal	Stream		
Statewide ^f	0.52 (0.51, 0.53)	0.43 (0.41, 0.44)	0.01 (0.01, 0.01)	0.01 (0.01, 0.02)	0.01(0.01 , 0.02)	0.27 (0.10, 0.44)	0.43 (0.29, 0.57)	-0.10 (-0.15,-0.05)	0.39 (0.30, 0.50)
Watersheds									
Cayuga Lake ^g	0.51 (0.50, 0.52)	0.42 (0.41, 0.44)	0.01 (0.01, 0.01)	0.02 (0.01, 0.02)	0.06 (0.01, 0.18)	0.74 (0.54, 1.07)	0.89 (0.73, 1.07)	-0.07 (-0.13, -0.02)	-0.17 (-0.30, - 0.06)
Keuka Lake	0.37 (0.28, 0.45)	1.26 (1.11, 1.040)	-0.61 (- 0.85, - 0.40)	-0.04 (- 0.07, - 0.01)	-0.07 (- 0.09, - 0.05)	-	-	-	-
Seneca Lake	0.50 (0.47, 0.54)	0.65 (0.58, 0.72)	0.00 (- 0.00, 0.01)	0.01 (0.01, 0.02)	0.01 (0.01, 0.02)	-	-	-0.18 (-0.29, -0.07)	0.85 (0.66, 1.04)
Susquehann a River	-0.15 (-0.29, - 0.02)	0.64 (0.38, 0.91)	-0.04 (- 0.08, - 0.00)	-	0.07 (- 0.04, 0.18)	-	-	-	-
All-Lakes ^h	0.46 (0.44, 0.48)	0.61 (0.57, 0.65)	-	-	-	0.19 (0.03, 0.36)	-	0.37 (0.24, 0.49)	-
Cayuga Lake	0.45 (0.43, 0.48)	0.32 (0.18, 0.48)	-	-	-	-	-	-	-
Chapel Pond	2.36 (1.93, 2.85)	-0.03 (-0.27, 0.23)	-	-	-	-	-	-	-
Dryden Lake	0.19 (0.10, 0.28)	0.28 (0.18, 0.40)	-	-	-	-	-	-	-
Keuka Lake	0.49 (0.41, 0.57)	0.75 (0.53, 0.94)	-	-	-	-	-	-	-
Seneca Lake	0.48 (0.40, 0.56)	0.76 (0.56, 0.95)	-	-	-	-	-	-	-
All-Streams	0.52 (0.51, 0.53)	0.42 (0.40, 0.43)	0.01 (0.01, 0.01)	0.01 (0.01, 0.01)	0.03 (0.02, 0.03)	-	-	-0.02 (-0.14, 0.09)	0.48 (0.38, 0.59)
Big Stream	0.35 (0.27, 0.42)	1.14 (0.99, 1.29)	-0.00 (-	-	0.04	-	-	-0.15 (-0.30, -0.01)	0.45 (0.26,

			0.01, 0.01)		(0.02, 0.05)				0.67)
Burroughs Creek	0.25 (0.04, 0.46)	0.27 (-0.07, 0.67)	0.04 (- 0.14, 0.19)	0.03 (0.00, 0.07)	-	-	-	-	-
Canoga Creek	0.50 (0.28, 0.71)	0.36 (-0.07, 0.83)	0.45 (0.02, 0.86)	-	-	-	-	-	-
Catherine Creek	0.58 (0.51, 0.64)	0.19 (0.06, 0.32)	0.01 (- 0.06, 0.08)	-0.02 (- 0.09, 0.04)	0.01 (- 0.04, 0.06)	-	-	-0.25 (-0.76, 0.27)	-
Cayuga Inlet	0.43 (0.38, 0.49)	-0.27 (-0.36, - 0.19)	-0.01 (- 0.01, - 0.00)	0.01 (0.00, 0.01)	0.00 (- 0.01, 0.02)	-	-	-	-
Direct Streams	0.43 (0.38, 0.49)	-0.28 (-0.36, - 0.20)	-0.01 (- 0.05, - 0.04)	0.01 (0.00, 0.01)	0.00 (- 0.01, 0.01)	-	-	-	-
Deans Creek	0.54 (0.42, 0.66)	-0.40 (-0.56, - 0.25)	-0.01 (- 0.01, - 0.00)	-	0.06 (0.03, 0.08)	-	-	-	-
Lake Ridge Creek	0.90 (0.70, 1.12)	-0.08 (-0.32, 0.16)	-	-8.15 (- 12.98, - 2.57)	-	-	-	-	-
Mill Creek	0.39 (0.30, 0.48)	-0.11 (-0.25, 0.06)	-0.01 (- 0.02, - 0.002)	-	0.01 (- 0.01, 0.04)	-	-	-	-
Paines Creek	0.53 (0.41, 0.65)	-0.18 (-0.39, 0.05)	-0.04 (- 0.06, - 0.03)	-	0.01 (- 0.03, 0.03)	-	-	-	-
Townline Creek	0.32 (0.20, 0.47)	-0.70 (-0.86, - 0.54)	0.04 (0.02, 0.05)	-	-	-	-	-	-
Fall Creek	0.51 (0.48, 0.53)	0.31 (0.28, 0.34)	0.02 (0.02, 0.03)	0.05 (0.04, 0.05)	0.01 (- 0.00, 0.01)	-	-	0.21 (0.10, 0.29)	-0.77 (-0.91, - 0.62)
Glen Eldridge Creek	0.49 (0.34, 0.65)	-0.20 (-0.61, 0.19)	-	-	-	-	-	-	-
Great Gully	-0.03 (-0.35, 0.30)	1.63 (1.03, 2.24)	0.02 (0.01,	-	-	-	-	-	-

			0.02)						
Kashong Creek	1.23 (1.09, 1.36)	-0.24 (-0.56, 0.09)	-0.10 (-0.13, -0.06)	-	0.05 (0.01, 0.08)	-	-	-	-
Keuka Outlet	0.37 (0.29, 0.45)	1.23 (1.10, 1.36)	-0.64 (-0.71, -0.57)	-0.04 (-0.05, -0.03)	-0.07 (-0.08, -0.07)	-	-	0.91 (0.79, 1.02)	1.70 (1.57, 1.84)
Milliken Creek	0.09 (-0.04, 0.23)	-	-0.07 (-0.89, -0.05)	-	-	-	-	-0.76 (-0.99, -0.48)	-
Pleasant Grove Brook	0.17 (0.00, 0.34)	0.17 (-0.19, 0.54)	-	-	-	-	-	-	-
Pony Hollow Creek	-0.16 (-0.27, -0.58)	0.57 (0.37, 0.77)	-0.01 (-0.03, 0.002)	-0.04 (-0.05, -0.20)	-	-	-	-	-
Reeder Creek	0.06 (0.01, 0.11)	1.53 (1.44, 1.63)	-0.01 (-0.02, 0.01)	-	-	-	-	-0.14 (-0.56, 0.25)	-
Salmon Creek	0.36 (0.32, 0.40)	1.14 (1.06, 1.23)	-0.03 (-0.03, 0.02)	-	-0.00 (-0.03, 0.02)	-	-	0.17 (0.07, 0.27)	-
Six Mile Creek	0.48 (0.46, 0.50)	0.35 (0.32, 0.39)	0.01 (0.00, 0.01)	0.00 (-0.02, 0.02)	0.04 (0.02, 0.05)	-	-	-	-
Taughannock Creek	0.57 (0.54, 0.61)	0.46 (0.41, 0.53)	-0.01 (-0.02, -0.00)	-0.01 (-0.01, -0.00)	0.02 (0.00, 0.03)	-	-	0.04 (-0.06, 0.13)	-
Trumansburg Creek	0.66 (0.61, 0.71)	0.65 (0.57, 0.74)	-0.02 (-0.03, 0.00)	-0.01 (-0.02, 0.01)	-0.03 (-0.04, 0.01)	-	-	-	0.39 (0.18, 0.60)
Yawger Creek	-0.13 (-0.1, 0.02)	1.54 (1.29, 1.79)	0.17 (0.12, 0.21)	0.09 (0.05, 0.13)	-0.11 (-0.20, -0.01)	-	-	0.54 (0.30, 0.80)	-

^a If the sample was collected under baseflow or stormflow conditions.

^b IDW=inverse distance weighted proportion of land in upstream watershed under given land cover

^c Waterway type refers to the type of waterway from which the samples were collected.

^d If the sample was collected at the stream mouth or upstream of the mouth.

^e If there was an outfall at the sampling site or not.

^f In the statewide model, water type was included as a covariate with lake as the reference-level. The statewide model estimated that *E. coli* levels were 0.27 log₁₀ CFU/100-mL (50% Credibility Interval [CI]=0.10, 0.44) higher for canals/flood control channels compared to lakes, 0.28 log₁₀ CFU/100-mL (CI =-0.72, 0.17) lower for springs compared to lakes, and 0.43 log₁₀ CFU/100-mL (CI =0.29, 0.57) higher in streams compared to lakes. Watershed was also included as a covariate in the statewide model with the Cayuga Lake watershed as the reference-level. Compared to samples collected from the Cayuga Lake watershed, *E. coli* levels were 0.69 log₁₀ CFU/100-mL (CI =0.54, 0.84) higher in samples collected from the Keuka Lake Watershed, 0.25 log₁₀ CFU/100-mL (CI =0.17, 0.33) higher in samples collected from the Seneca Lake Watershed, and 0.55 log₁₀ CFU/100-mL (CI =0.73, 0.38) lower in samples collected from the Susquehanna River watershed.

^g In the watershed model for Cayuga Lake, water type was included as a covariate with lake as the reference-level. The Cayuga Lake watershed model estimated that *E. coli* levels were 0.74 log₁₀ CFU/100-mL (CI =0.54, 0.95) higher for canals/flood control channels compared to lakes, 0.07 log₁₀ CFU/100-mL (CI =-0.03, 0.53) lower for springs compared to lakes, and 0.37 log₁₀ CFU/100-mL (CI =0.24, 0.49) higher in streams compared to lakes.

^h In the all-lake model, water type was included as a covariate with lake as the reference-level. The all-lake model estimated that *E. coli* levels were 0.19 log₁₀ CFU/100-mL (CI =0.03, 0.36) higher for canals/flood control channels compared to lakes, 0.35 log₁₀ CFU/100-mL (CI =-0.74, 0.04) lower for ponds compared to lakes, and 0.89 log₁₀ CFU/100-mL (CI =0.73, 1.07) higher in streams compared to lakes.

Table S5: Association between *E. coli* levels and temporal factors according to the multivariable Bayesian models of coregionalization.

Model	Median Coefficient Estimate (50% Credibility Interval)												E R a	Nonspati al:Spatia l Variance b	
	Year	Month (April=Reference)													
		Jan.	Feb.	Mar.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.			
Statewide	-0.01 (- 0.01, - 0.00)	0.06 (0.02, 0.11)	-0.26 (- 0.30, - 0.22)	-0.20 (- 0.23, - 0.17)	0.34 (0.31, 0.36)	0.70 (0.68, 0.73)	0.59 (0.57, 0.62)	0.64 (0.62, 0.67)	0.66 (0.63, 0.68)	0.59 (0.57, 0.62)	0.30 (0.27, 0.32)	0.19 (0.16, 0.22)			
Watersheds															
Cayuga Lake	-0.00 (- 0.01, - 0.00)	0.06 (0.01, 0.09)	-0.27 (- 0.31, - 0.23)	-0.20 (- 0.22, - 0.17)	0.33 (0.30, 0.36)	0.72 (0.69, 0.74)	0.60 (0.57, 0.63)	0.63 (0.61, 0.66)	0.65 (0.62, 0.68)	0.55 (0.52, - /58)	0.27 (0.25, 0.30)	0.18 (0.15, 0.21)			
Keuka Lake	-0.03 (- 0.06, - 0.01)	-	-	-	0.49 (0.34, 0.69)	-0.84 (- 1.03, - 0.67)	0.57 (0.46, 0.68)	0.54 (0.41, 0.68)	0.77 (0.68, 0.90)	0.93 (0.80, 1.06)	0.57 (0.44, 0.74)	-			
Seneca Lake	-0.00 (- 0.06, - 0.04)	-	-	-0.03 (- 0.14, 0.09)	0.41 (0.31, 0.51)	0.47 (0.38, 0.56)	0.64 (0.57, 0.71)	0.56 (0.47, 0.64)	0.61 (0.53, 0.69)	0.81 (0.73, 0.90)	0.27 (0.18, 0.37)	-			
Susquehann a River	0.25 (0.16, 0.34)	-	0.65 (0.24, 1.10)	-0.22 (- 0.62, 0.15)	0.55 (0.30, 0.79)	1.13 (0.88, 1.39)	0.49 (0.14, 0.88)	0.31 (0.08, 0.54)	1.56 (1.23, 1.94)	0.54 (0.30, 0.79)	-	-			
All-Lake	0.00 (- 0.00, 0.01)	-0.07 (- 0.20, 0.06)	-0.27 (- 0.42, - 0.15)	-0.30 (- 0.36, - 0.24)	0.31 (0.24, 0.37)	0.60 (0.53, 0.65)	0.52 (0.47, 0.57)	0.52 (0.46, 0.58)	0.63 (0.57, 0.68)	0.60 (0.54, 0.66)	0.42 (0.35, 0.49)	0.19 (0.10, 0.27)			
Cayuga Lake	-0.00 (- 0.00, - 0.01)	-0.20 (- 0.36, - 0.05)	-0.23 (- 0.38, - 0.09)	-0.36 (- 0.44, - 0.29)	0.28 (0.21, 0.35)	0.75 (0.68, 0.82)	0.54 (0.48, 0.61)	0.51 (0.45, 0.59)	0.66 (0.59, 0.73)	0.60 (0.53, 0.66)	0.44 (0.36, 0.51)	0.18 (0.09, 0.26)			
Chapel Pond	0.36 (0.11, 0.60)	-	-	0.15 (- 0.19, 0.48)	-	2.33 (1.88, 2.85)	-	-	1.56 (1.10, 2.07)	-	-	-			
Dryden Lake	-0.00 (- 0.02, 0.01)	-	-	-0.01 (- 0.16, 0.15)	-	-0.20 (- 0.32, - 0.05)	-	-	0.13 (- 0.01, 0.26)	-	-	-			
Keuka Lake	-0.03 (- 0.06, 0.003)	-	-	0.30 (0.07, 0.52)	0.48 (0.26, 0.71)	0.40 (0.21, 0.60)	0.33 (0.20, 0.50)	0.39 (0.21, 0.57)	0.19 (0.002, 0.35)	0.52 (0.34, 0.72)	0.18 (- 0.07, 0.41)	-			
Seneca Lake	-0.02 (- 0.05, 0.01)	-	-	0.30 (0.09, 0.56)	0.46 (0.24, 0.68)	0.43 (0.19, 0.64)	0.34 (0.18, 0.51)	0.42 (0.23, 0.61)	0.24 (0.05, 0.41)	0.53 (0.34, 0.73)	0.19 (- 0.03, 0.42)	-			
All-Streams	-0.01 (- 0.01, -	0.08 (0.04,	-0.26 (- 0.29, -	-0.18 (- 0.21, -	0.34 (0.31,	0.72 (0.69,	0.63 (0.60,	0.67 (0.64,	0.67 (0.64,	0.58 (0.55,	0.28 (0.25,	0.19 (0.16,			

	0.00)	0.13)	0.22)	0.15)	0.37)	0.75)	0.65)	0.70)	0.70)	0.61)	0.31)	0.22)
Big Stream	-0.09 (- 0.12, - 0.07)	-	-	-0.13 (- 0.32, 0.05)	0.20 (0.01, 0.37)	-0.15 (- 0.35, 0.05)	0.56b (0.44, 0.69)	0.63 (0.47, 0.78)	0.50 (0.36, 0.64)	0.61 (0.47, 0.77)	0.20 (- 0.03, 0.44)	-
Burroughs Creek	-0.10 (- 0.17, - 0.03)	-	-	-0.48 (- 0.71, - 0.22)	-	0.97 (0.66, 1.31)	-	-	0.44 (0.16, 0.69)	-	-	-
Canoga Creek	-0.06 (- 0.14, 0.01)	-	-	-0.28 (- 0.52, - 0.03)	-	0.51 (0.07, 0.92)	-	-	0.06 (- 0.16, 0.27)	-	-	-
Catherine Creek	-0.03 (- 0.05, - 0.00)	-	-	-0.14 (- 0.32, 0.04)	-0.12 (- 0.31, 0.09)	0.45 (0.31, 0.60)	0.60 (0.49, 0.71)	0.38 (0.23, 0.52)	0.44 (0.31, 0.57)	0.75 (0.61, 0.89)	0.23 (0.08, 0.38)	-
Cayuga Inlet	-0.04 (- 0.05, - 0.04)	-	-	-0.44 (- 0.55, - 0.32)	0.45 (0.33, 0.54)	1.00 (0.92, 1.09)	0.91 (0.82, 1.01)	1.88 (1.75, 2.01)	-	1.43 (1.33, 1.53)	-	0.75 (0.64, 0.85)
Direct Streams	-0.04 (- 0.05, - 0.04)	-	-	-0.41 (- 0.52, - 0.31)	0.45 (0.34, 0.55)	1.02 (0.95, 1.11)	0.94 (0.85, 1.03)	1.92 (1.78, 1.06)	-	1.45 (1.36, 1.56)	-	0.78 (0.67, 0.88)
Deans	-0.05 (- 0.07, - 0.04)	-	-	-0.20 (- 0.37, 0.02)	-0.18 (- 0.36, 0.01)	1.17 (1.01, 1.32)	0.89 (0.70, 1.08)	2.29 (1.02, 2.55)	-	1.75 (1.53, 1.98)	-	0.75 (0.56, 0.96)
Lake Ridge	-0.02 (- 0.03, 0.01)	-	-	-1.26 (- 1.51, - 1.01)	0.41 (0.15, 0.64)	0.05 (- 0.12, 0.25)	0.45 (0.27, 0.63)	-0.19 (- 0.62, 0.29)	-	0.11 (- 0.21, 0.38)	-	0.78 (0.51, 1.02)
Mill	-0.03 (- 0.05, - 0.02)	-	-	-0.18 (- 0.38, 0.02)	1.20 (1.02, 1.38)	1.04 (0.86, 1.19)	1.02 (0.81, 1.21)	2.13 (1.84, 2.40)	-	1.20 (1.02, 1.38)	-	0.77 (0.56, 0.98)
Paines	-0.06 (- 0.08, - 0.04)	-	-	-0.90 (- 1.13, - 0.68)	0.43 (0.18, 0.68)	0.89 (0.70, 1.04)	0.63 (0.42, 0.83)	1.31 (0.99, 1.64)	-	1.06 (0.82, 1.28)	-	0.74 (0.51, 0.98)
Townlin e	-0.06 (- 0.08, - 0.05)	-	-	-0.54 (- 0.74, - 0.34)	0.90 (0.69, 1.10)	1.31 (1.16, 1.44)	1.56 (1.36, 1.76)	2.23 (1.92, 2.55)	-	2.32 (2.11, 2.53)	-	2.32 (2.12, 2.53)
Fall Creek	-0.02 (- 0.02, - 0.01)	0.10 (0.003, 0.20)	-0.39 (- 0.48, - 0.31)	-0.06 (- 0.12, 0.008)	0.49 (0.40, 0.57)	0.79 (0.73, 0.85)	0.62 (0.54, 0.70)	0.81 (0.72, 0.91)	0.69 (0.63, 0.76)	0.78 (0.62, 0.93)	0.51 (0.44, 0.58)	0.19 (0.13, 0.26)
Glen Eldridge Creek	-0.06 (- 0.21, 0.13)	-	-	-0.51 (- 0.70, - 0.32)	-	-	-	-	-0.04 (- 0.25, 0.14)	-	-	-
Great Gully	-0.24 (- 0.45, - 0.05)	-	-	0.03 (- 0.48, 0.44)	-	-	-	-	-0.24 (- 0.65, 0.16)	-	-	-
Kashong	-0.004 (-	-	-	-	0.30	0.80	0.63	-	-	1.01	0.12	-

Creek	0.06, (0.05)				(0.06, 0.53)	(0.63, 0.97)	(0.50, 0.74)			(0.86, 1.15)	9_0.02, (0.27)		
Keuka Outlet	-0.04 (-0.06, -0.01)	-	-	-	0.51 (0.34, 0.65)	-0.89 (-1.04, -0.69)	0.54 (0.43, 0.64)	0.50 (0.38, 0.61)	0.72 (0.62, 0.82)	0.88 (0.78, 1.00)	0.55 (0.40, 0.70)	-	
Milliken Creek	0.36 (0.13, 0.56)	-	-	-	-1.47 (-1.77, -1.15)				-0.62 (-0.92, -0.36)	-0.57 (-0.75, -0.41)		-	
Pleasant Grove Brook	0.00 (-0.02, 0.02)	-	-	-	-0.03 (-0.29, 0.21)	0.32 (0.00, 0.64)	0.06 (-0.20, 0.35)	-0.31 (-0.57, 0.01)	-0.23 (-0.47, 0.05)	-0.38 (-0.73, 0.05)		-	
Pony Hollow Creek	0.27 (0.20, 0.35)		0.70 (0.34, 1.06)	-0.10 (-0.46, 0.23)	0.55 (0.33, 0.75)	1.03 (0.80, 1.25)	0.60 (0.22, 0.92)	0.25 (0.04, 0.45)	1.58 (1.22, 1.94)	0.43 (0.21, 0.65)		-	
Reeder Creek	0.03 (0.01, 0.05)	-	-	0.12 (-0.07, 0.30)	0.85 (0.69, 1.02)	0.91 (0.73, 1.10)	1.12 (0.98, 1.27)	1.10 (0.94, 1.26)	0.84 (0.69, 1.00)	0.71 (0.54, 0.87)	0.84 (0.66, 1.01)	-	
Salmon Creek	-0.01 (-0.02, 0.01)	-0.45 (-0.84, -0.10)	-0.34 (-0.67, 0.04)	-0.51 (-0.61, 0.40)	0.00 (-0.07, 0.08)	0.04 (-0.07, 0.14)		0.36 (0.28, 0.44)	0.68 (0.40, 0.95)	0.67 (0.56, 0.79)	-0.04 (-0.12, 0.05)	0.10 (-0.00, 0.20)	
Six Mile Creek	0.01 (0.01, 0.02)	0.24 (0.18, 0.30)	-0.22 (-0.29, 0.15)	0.01 (-0.04, 0.06)	0.46 (0.41, 0.53)	0.94 (0.89, 0.99)	0.86 (0.80, 0.91)	0.99 (0.92, 1.06)	0.78 (0.73, 0.83)	0.84 (0.79, 0.90)	0.49 (0.43, 0.55)	0.30 (0.20, 0.40)	
Taughannock Creek	-0.01 (-0.02, 0.00)	-0.29 (-0.60, 0.04)	-0.43 (-0.52, 0.34)	-0.55 (-0.68, 0.44)	0.47 (0.37, 0.58)	0.39 (0.31, 0.49)	0.43 (0.35, 0.53)	0.25 (0.16, 0.36)	0.08 (-0.04, 0.20)	0.31 (0.23, 0.39)	-0.40 (-0.51, 0.31)	0.07 (-0.20, 0.07)	
Trumansburg Creek	-0.03 (-0.04, 0.02)	-0.56 (-0.94, 0.15)	-0.26 (-0.39, 0.12)	-0.63 (-0.84, 0.45)	0.03 (-0.13, 0.20)	0.31 (0.15, 0.47)	0.10 (-0.05, 0.24)	-0.008 (-0.14, 0.14)	-0.20 (-0.41, 0.01)	0.05 (-0.06, 0.17)	-0.06 (-0.20, 0.09)	-0.01 (-0.23, 0.19)	
Yawger Creek	-0.52 (-0.63, -0.42)	-	-	-1.00 (-1.41, -0.61)	0.50 (0.22, 0.78)	0.50 (0.18, 0.88)			1.04 (0.68, 1.36)	0.12 (-0.14, 0.38)		-	

^a ER = median effective range, or the distance in meters at which *E. coli* levels became spatially independent.

^b Variance Ratio = variance attributable to non-spatial versus spatial sources (Median Estimate for Nugget/Median Estimate for Partial Sill)

Figure S1: Number of samples collected from **lakes** and **streams** each year since collection started. Fewer samples were collected in 2020 than 2019 since data were downloaded in April 2020; the most recent sample was collected on April 6th, 2020.

Figure S2: Seasonal (W=winter, Sp=spring, Su=summer, F=fall) and monthly differences in the distribution of log₁₀ *E. coli*, total coliform, total suspended solids, and turbidity in canal, lake, and streams samples.

Figure S3: Loess-smoothed lines showing trends in the log10 concentration of *E. coli*, total coliforms, total suspended solids, and turbidity against days since Jan 1st (i.e., over one year), and over the 18 years samples were collected.

Figure S4: (S2A) Loess-smoothed lines showing East-West and North-South trends in (Top Row) the IDW percent of the watersheds under **crop**, **developed**, **forest-wetland**, and **pasture** cover, and (Bottom Row) the log10 concentration of ***E. coli***, **total coliforms**, **total suspended solids**, and **turbidity**. Note that in FigS2A the trends in land cover approx. mirror the trends in water quality. (S2B) Loess-smoothed lines show trends in the log10 concentration of ***E. coli***, **total coliforms**, **total suspended solids**, and **turbidity** against IDW percent of upstream watershed under crop, developed, forest-wetland, or pasture cover.

S2A

S2B

Figure S5: Percent of variance in four water quality parameters that can be attributed to spatial effects in intercept-only Bayesian models of coregionalization for individual lakes (L), streams (S), and HUC-4/Finger Lake watersheds. The percent spatial variance calculated using the 1st and 3rd quartiles, and median values for the partial sill and nugget are shown [% spatial variance = (partial sill)/(partial sill+nugget)].