Journal of Food Protection 86 (2023) 100045

Contents lists available at ScienceDirect

Journal of Food Protection

journal homepage: www.elsevier.com/locate/jfp

Research Paper

Spatial Versus Nonspatial Variance in Fecal Indicator Bacteria Differs Within and Between Ponds

Claire M. Murphy [1], Daniel L. Weller [1,2], Reza Ovissipour [3], Renee Boyer [1], Laura K. Strawn [1,*]

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA 24061, USA
[2] Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY USA
[3] Department of Food Science and Technology, Virginia Tech Seafood Agricultural Research and Extension Center, Hampton, VA 23669, USA

ARTICLE INFO

Keywords:
Escherichia coli
FSMA
Microbial water quality
Produce Safety Rule
Representative sample
Water testing

ABSTRACT

Surface water environments are inherently heterogenous, and little is known about variation in microbial water quality between locations. This study sought to understand how microbial water quality differs within and between Virginia ponds. Grab samples were collected twice per week from 30 sampling sites across nine Virginia ponds (n = 600). Samples (100 mL) were enumerated for total coliform (TC) and *Escherichia coli* (EC) levels, and physicochemical, weather, and environmental data were collected. Bayesian models of coregionalization were used to quantify the variance in TC and EC levels attributable to spatial (e.g., site, pond) versus nonspatial (e.g., date, pH) sources. Mixed-effects Bayesian regressions and conditional inference trees were used to characterize relationships between data and TC or EC levels. Analyses were performed separately for each pond with ≥3 sampling sites (5 intrapond) while one interpond model was developed using data from all sampling sites and all ponds. More variance in TC levels were attributable to spatial opposed to nonspatial sources for the interpond model (variance ratio [VR] = 1.55) while intrapond models were pond dependent (VR: 0.65–18.89). For EC levels, more variance was attributable to spatial sources in the interpond model (VR = 1.62), compared to all intrapond models (VR < 1.0) suggesting that more variance is attributable to nonspatial factors within individual ponds and spatial factors when multiple ponds are considered. Within each pond, TC and EC levels were spatially independent for sites 56–87 m apart, indicating that different sites within the same pond represent different water quality for risk management. Rainfall was positively and pH negatively associated with TC and EC levels in both inter- and intrapond models. For all other factors, the direction and strength of associations varied. Factors driving microbial dynamics in ponds appear to be pond-specific and differ depending on the spatial scale considered.

Fecal contamination of surface waterways poses a public health risk. Surface waters, including agricultural and recreational water, with fecal contamination have been identified as the likely source of multiple outbreaks of gastrointestinal and respiratory illness (Craun et al., 2006; Food and Drug Administration, 2019; Graciaa et al., 2018; Greene et al., 2008; Lee et al., 2002; U.S Environmental Protection Agency, 2009). From 2000–2014, public health officials reported 140 outbreaks associated with untreated recreational water which resulted in at least 4,958 cases of disease and two deaths (Graciaa et al., 2018). Among the 95 outbreaks with a confirmed infectious etiology, enteric pathogens caused 84 % (80/95) (Graciaa et al., 2018). Surface water is also often used in the agricultural environment for practices such as irrigation, fertigation, frost protection, pesticide application, and dust abatement. If surface water used for agricultural purposes is contaminated with pathogenic bacteria, foodborne outbreaks may result. For example, in the 2008 *Salmonella* Saintpaul outbreak associated with raw jalapeño and serrano peppers, the outbreak strain was sourced to the agricultural water used on the farm (Centers for Disease Control Prevention, 2008). Due to the public health risks associated with fecal contamination of surface water sources, there is considerable interest in understanding factors that impact fecal bacteria presence and concentration in surface waters.

Methods for assessing water quality and safety in recreational water, and surface water for produce production rely heavily on the use of fecal indicator bacteria (FIB) testing (Food and Drug Administration, 2015; U.S Environmental Protection Agency, 2012). FIB monitoring is routinely used to assess the acceptability of surface water quality for produce production and recreation use; and if neces-

* Corresponding author at: Department of Food Science and Technology, Virginia Tech, 1230 Washington Street, SW, Blacksburg, VA 24061, USA.
E-mail address: Lstrawn@vt.edu (L.K. Strawn).

https://doi.org/10.1016/j.jfp.2023.100045
Received 5 September 2022; Accepted 18 January 2023
Available online 24 January 2023

sary, determine if corrective actions are needed (Food and Drug Administration, 2015; U.S Environmental Protection Agency., 2012). However, variations in microbial water quality due to spatial (data dependent on geographic location, e.g., latitude and longitude, waterway) and temporal (data relating to time, e.g., day, month, season) factors, or the inherent heterogeneity of freshwater systems, need to be considered when developing monitoring programs. A variety of spatial and temporal factors have been identified in previous research to influence FIB levels in surface waters (Draper et al., 2016; Duris et al., 2013; Francy et al., 2013; Green et al., 2021; Murphy et al., 2022; Schilling et al., 2009; Truitt et al., 2018; Weller et al., 2020; Wilkes et al., 2009; Zhang et al., 2020). For example, data from 68 streams in New York State (US) found that the density of upstream pig farms and turbidity were positively associated with higher *Escherichia coli* concentrations, while the presence of upstream goat/sheep farms, dissolved oxygen, and pH had a negative effect (Green et al., 2021). Additionally, a study in Pennsylvania (US) found that stream sites with greater urban land use and less forest land use were positively associated with increased FIB levels (Duris et al., 2013).

Most FIB-based monitoring programs do not consider sampling frequency; therefore, they may not accurately capture temporal variation in water quality. Recent studies have shown that the sampling frequencies proposed in several regulations/standards may not be sufficient at capturing overall water quality (Draper et al., 2016; Havelaar et al., 2017; Weller et al., 2020). Previous research demonstrated the first proposed version of the Food Safety Modernization Act's Produce Safety Rule Agricultural Water (Subpart E) standards of using 20 samples to calculate a geometric mean (GM) and statistical threshold value (STV) were not sufficient to characterize the bacteriological quality of irrigation ponds from Florida (US), as GM and STV were highly variable from subset to subset (Havelaar et al., 2017). Furthermore, little data have been provided on the spatial scale needed for the effective implementation of FIB monitoring in ponds, and prior studies have shown that FIB levels differ between sampling locations within a single surface water source (Pachepsky et al., 2018; Quilliam et al., 2011; Schilling et al., 2009; Weller et al., 2020). This highlights the importance of considering sampling temporal variation, sampling site variation, and scale when developing and implementing FIB monitoring programs.

Promising applications for understanding water quality are based on models that aim to predict microbial water quality for surface waterways based on a variety of spatial and temporal factors (Belias et al., 2021; Green et al., 2021; Murphy et al., 2022; Strawn et al., 2013; Topalcengiz et al., 2017; Weller et al., 2020; Weller et al., 2020). Advanced statistical approaches may be used to determine combinations of environmental, meteorological, spatial, and temporal factors to evaluate, and determine if changes to FIB monitoring efforts are needed. Therefore, the overall aim of this study was to characterize spatial and temporal variability in FIB levels within and between ponds to determine (i) the distance at which microbial water quality within a given waterway becomes spatially independent, (ii) if more variance in FIB levels is attributable to spatial or nonspatial factors and (iii) identify factors associated with inter- and intrapond microbial water quality.

Methods

Study design and water sampling

Nine ponds in Montgomery County, Virginia (US) were used in this study for understanding inter- and intrapond water quality. From the nine ponds, a total of 30 sampling sites were used for sample collection. Sampling locations within and number of samples collected from each pond were based on varying location (i.e., inflow, outflow, middle), feasibility of safe sample collection, and pond size (pond-A: number of sites (N) = 3, pond-B: N = 2, pond-C: N = 1, pond-D: N = 10, pond-E: N = 1, pond-F: N = 5, pond-G: N = 1, pond-H: N = 3, pond-I: N = 4). From the 30 sites, water samples were collected twice a week for a 10-week period from September to November 2021 (Total number of samples collected = 600). Water samples were collected directly into sterile 1-L wide-mouth bottles using a 3.6 m sampling pole (Nasco). As previously described (Strawn et al., 2013; Truitt et al., 2018), samples were taken a minimum of 2 m from the shoreline and 0.3 m below the surface to ensure sediment was not disturbed. All samples were transported on ice and processed within 6 h of sample collection according to methods deemed appropriate under FSMA's Produce Safety Rule Agricultural Water Subpart E (Food and Drug Administration, 2015).

Enumeration of fecal indicator bacteria (FIB)

Fecal indicator bacteria (FIB; total coliform and *E. coli*) levels were enumerated using Colilert Quanti-Tray 2000 kit (IDEXX, Westbrook) following the manufacturer's instructions (https://www.idexx.com/en/water/water-products-services/quanti-tray-system/). Quanti-Trays were incubated at 35 ± 2°C for 24 ± 2 h. Sample dilutions (1:10 and 1:100) were performed; therefore, upper and lower limits of quantification were 241,960 MPN/100 mL (5.4 $\log_{10}$ MPN/100 mL) and 10 MPN/100 mL (1 $\log_{10}$ MPN/100 mL), respectively.

Metadata collection

For each sample collection date, data related to meteorological factors were obtained from WeatherStem's data mining tool (https://en.weatherstem.com/data) using the closest weather station to each sampling site. The weather stations were within approximately 700–2500 m of each pond (as a bird flies; estimated using Google Earth pro: https://www.google.com/earth/versions/#download-pro). In total, 19 different meteorological factors were obtained, including wind speed (km/h), air temperature (°C), relative humidity (%), rainfall (mm), solar radiation (watts/m^2), and UV radiation (UV index). Direct meteorological measurements were acquired for the day of sampling and the three days prior to sample collection. Summary statistics for continuous weather factors are shown in Table S1.

During each sampling event, physicochemical water quality factors were collected from each sampling site. Measurements were obtained in triplicate for turbidity (NTU), water temperature (°C), pH, conductivity (µs/cm), and dissolved oxygen (mg/mL). Each physicochemical water quality factor was measured in the field with the YSI ProDSS (Yellow Springs, OH). Summary statistics for physicochemical water quality factors are shown in Figure S1.

To characterize the land use and pond features, constant and variable site observations were collected. The perimeter (m) and area (m^2) of each pond were estimated using Google Earth Pro and the polygon tool (https://www.google.com/earth/versions/#download-pro). Pond size is summarized in Table S2. During the first sampling, constant site observations were collected for each of the 30 sampling sites. Constant observations included pond location for sample collection (i.e. inflow, outflow, middle), dam (y/n), buildings within 100 m (y/n), parking lot within 100 m (y/n), road within 100 m (y/n), adjacent ditch to pond (y/n), pipe in or out of pond (y/n), surrounding vegetation (e.g., grassland, woods), pond bottom (e.g., silt, sand, gravel), livestock adjacent to pond (y/n), recreation adjacent to pond (y/n), and slope to pond (e.g., bottom of incline, flat). Constant observations are summarized in Table S3. During each of the 20 sampling events, variable site observations were collected for each of the 30 sampling sites. Variable observations included the presence of people (y/n), dogs (y/n), geese (y/n), ducks (y/n), other animals (y/n), fishing (y/n), surface biofilms (scummy, oily film on the surface of pond y/n), boating (y/n), submerged vegetation (y/n), emergent vegetation (y/n), trash (y/n), and algae (y/n).

Statistical analyses

All analyses were performed in R version 4.0.2 (R Foundation for Statistical Computing). Initial descriptive analyses were performed, and summary statistics were calculated for all continuous factors (e.g., FIB levels, weather conditions). Similar to prior literature (Murphy et al., 2022), for samples where FIB levels were above the upper limit of detection (LOD), one-half greater than the upper LOD was used. Similarly, where FIB levels were below the lower LOD, one-half of the lower LOD was used. When aiming to quantify spatial variance, analyses were performed separately for each pond with ≥3 sampling sites (referred to as intrapond, total intrapond models = 5) and using all available data (referred to as interpond).

Intercept-only Bayesian models of coregionalization (MCOR) were fit to characterize the ratio of spatial variance (variance dependent on geographic location) to nonspatial variance (variance independent of geographic location) for each FIB, and to quantify the effective ranges (continuous factor: the distance at which FIB levels became independent) for each pond and across all ponds (Finley et al., 2007, 2013; Plummer et al., 2006; Wackernagel, 2003; Weller et al., 2022). Since the aim of the intercept-only MCOR was to quantify the variance ratio, individual pond models were not implemented for the four ponds where all samples were collected from less than three sites (pond-B, pond-C, pond-E, pond-G). In order to fit the MCOR, semivariograms, which chart out the autocorrelation of how sample values vary with distance, were fit using GPS coordinates and FIB levels to help generate informative priors for phi (3/effective range), tau (nugget, or nonspatial variance), and rho (partial sill, or spatial variance) using an exponential covariance model using the *geoR* package (Weller et al., 2022). The MCOR were fit, and estimates were quantified using the *spBayes* and *coda* packages (Finley et al., 2007, 2013; Plummer et al., 2006). As previously described (Weller et al., 2022), ten thousand iterations (10 K) were performed for each MCOR, but the first 70 % (7000) were discarded as burn-in. For the intra- and interpond MCOR, the 25th, 50th, and 75th percentiles for the phi, tau, rho, and variance ratio (rho/tau) were calculated.

Bayesian mixed models were implemented to characterize the differences in $\log_{10}$ FIB levels between ponds and to determine inter- and intrapond differences attributable to each individual physicochemical, weather, and environmental factor exclusively. In the interpond models, a fixed effect of "days sincefirst sampling" and a random effect of site nested in pond was used. For the intrapond models, a fixed effect of "days since first sampling" and random effect of site was used. The temporal factors of “days since first sampling” were integers that represented the number of days since the first sampling event. For example, the first sampling event occurred on 9/14/2021 and was assigned a value for “days since first sampling” of 0. Thus, the sampling events occurring on 9/21/2021 and 10/14/2021 were assigned a value of 7 and 30 for “days since first sampling”, respectively. Fixed and random effects for inter- and intrapond models were used to account for pseudoreplication and temporal autocorrelation. Separate models for total coliform and *E. coli* levels were fit for each of the five intrapond and the one interpond data with each individual site observation, physicochemical water quality parameter, and meteorological factor as covariates.

Models were fit with the default package priors utilizing the *brms* package using six, 5,001 iteration chains, and thinning = 10 (Bürkner, 2017a, 2017b). The maximum a posteriori (MAP) and 95 % credibility interval (CrI) for the effect estimates, as well as, the probability of direction (PD) and regional of practical equivalence overlap (ROPE) were extracted using the *bayestestR* package (Makowski, Ben-Shachar & Chen et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). The MAP is interpreted like a beta coefficient from a frequentist model (e.g., change in $\log_{10}$ MPN of FIB/100 mL associated with either a one-unit change for a continuous covariate or the change from reference level to a different level for a categorical covariate). The PD is an indicator of whether the direction of the association is positive or negative; with values near 1.0 indicating greater certainty that the effect of the factor is truly positive and values near zero indicating greater certainty that the effect of the factor is truly negative (Makowski, Ben-Shachar & Chen et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). Thus, PD values of 0.95, 0.975, 0.995, and 0.9995 correspond to two-sided frequentist P values of 0.10, 0.05, 0.01, and 0.001, respectively (Makowski, Ben-Shachar & Chen et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). The practical significance (PS) refers to the magnitude of the difference and is a uni-directional equivalence test that indicates if the effect is both nonnegligible, and in a given direction based on values larger than 0.5 (Makowski, Ben-Shachar & Chen et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). The ROPE indicates if the parameter is outside of a range of practically negligible effect with ROPE percentage closer to zero signifying more confidence that the given factor has a substantial effect on FIB levels. ROPE values > 97.5 % indicate a negligible effect (factor has a not-noticeable, insignificant effect), between 2.5 % and 97.5 % denote an uncertain effect (effect of factors is unknown and unclear), and < 2.5 % indicate a nonnegligible effect (factor has a noticeable, significant effect) (Makowski, Ben-Shachar & Chen et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). Having a 95 % CrI that includes zero is not necessarily indicative of the absence of an association and should be considered alongside the PD, PS, and ROPE for interpretation.

To understand, identify, and visualize ranked combinations of physicochemical, weather, and environmental factors associated with an increased or decreased levels of FIB, conditional inference trees were implemented. Conditional inference trees were used because they can account for collinearity and correlated factors, can account for hierarchical relationships, and are easily interpreted. Trees were developed utilizing the *mlr* and *partykit* packages. Tenfold cross-validation repeated three times was used to tune hyper-parameters. To prevent overfitting, the maxdepth (maximum depth to which a tree can be grown) was set to a maximum of 3 and the mincriterion (the value of the test statistic that must be exceeded in order to implement a split) was set to 0.95. The outcome of the trees was $\log_{10}$ MPN/100 mL FIB concentration.

Results

The surface water used in this study had various uses and were not all irrigation, livestock, and/or recreation ponds. The sizes of the ponds sampled ranged from approximately 123–736 m in perimeter and 490–23,500 m^2 in area (Table S2). Of the 30 sampling sites, 7 were pond inflows and 6 pond outflows (Table S3). The ponds represented four unique pond bottom sediment materials with one pond each of clay, gravel, and sand and the remaining six being silt. Additional data on the ponds and sampling sites can be found in Table S3.

For total coliform (TC) levels, 10 samples were above the LOD (>5.4 $\log_{10}$ MPN/100 mL) and eight samples were below the lower LOD (<1.0 $\log_{10}$ MPN/100 mL). For *E. coli*, 44 samples were below the lower LOD, and no counts were above the upper LOD. For the counts above the upper LOD (>241960 MPN/100 mL), the value of 5.6 $\log_{10}$ MPN/100 mL (362940 MPN/100 mL) was used (1.5 time the upper LOD). For the counts below the lower LOD (10 MPN/mL), the value of 0.7 $\log_{10}$ MPN/100 mL (5 MPN/100 mL) was used (0.5 time the lower LOD). The interpond mean and median $\log_{10}$ MPN/100 mL values for TC levels were 4.2 (Range = 0.7, 5.6) and 3.2 (Interquartile Range [IQR] = 4.0, 4.6), respectively (Table S4). For *E. coli*, the interpond mean and median $\log_{10}$ MPN/100 mL values were 2.4 (Range = 0.7, 4.8) and 2.5 (IQR = 1.7, 3.0), respectively

(Table S4). Mean intrapond TC and *E. coli* levels ranged from 3.4 to 4.6 and 1.2 to 2.9 $\log_{10}$ MPN/100 mL, respectively. Pond-I had the lowest mean $\log_{10}$ MPN/100 mL for both FIB targets (Table S4). Pond-A and Pond-D had the highest TC levels, while Pond-F had the highest *E. coli* levels (Table S4). All summary statistics for inter- and intrapond FIB levels can be found in Table S4.

Variance attributable to spatial versus nonspatial sources

The interpond variance ratio, which indicates the variance attributable to spatial versus nonspatial sources, was 1.55 and 1.62 for TC and *E. coli*, respectively (Table 1). Since the variance ratio was >1.0, more variance in FIB levels was attributable to spatial than nonspatial sources. TC variance ratios ranged from 0.65 to 18.89 for each of the five individual ponds. However, variance ratios explaining *E. coli* levels were <1.0 for all five intrapond models indicating that for individual ponds, more variance in *E. coli* levels was attributable to nonspatial sources (compared to spatial sources).

The distance at which microbial water quality became spatially independent

In the intercept-only MCOR, the effective range (the range at which FIB levels became spatially independent) varied substantially based on the scale of analysis (the scale used to analyze the data, e.g., intrapond vs interpond). According to the intercept-only MCOR, the effective range for TC and *E. coli* levels in interpond models was 756.4 m and 1,512.1 m, respectively (Table 1). The interpond effective range for *E. coli* levels was twice that of TC levels, suggesting that the pond accounted for more variability in *E. coli* levels than TC levels. For individual ponds, the effective range for TC and *E. coli* levels ranged from 68.70 to 86.28 and 56.99 to 69.00 m, respectively (Table 1).

Associations between physicochemical, weather, and environmental data and FIB levels

Table S5 shows the differences in FIB levels between ponds, using the pond with the lowest mean FIB levels as reference (pond-I). Nonnegligible (noticeable) differences exist between both the TC and *E. coli* levels between the ponds; therefore, pond was used as a random factor in Bayesian mixed models. According to the results of the Bayesian mixed models, rain in mm 0–24 h before sample collection (BSC) had a substantial nonnegligible positive effect on both TC and *E. coli* levels, regardless of the scale of analysis (i.e., intrapond vs interpond (Table 2). Interestingly, the positive association between rain and FIB levels occurred for both rain volume 0–24 h BSC (mm), and if a rain event occurred 0–24 h BSC (Y/N), indicating that rain 0–24 h BSC, regardless of volume, increased FIB levels (Table 2). Of the intraponds, Pond-F was the only individual pond model where rain 0–24 h BSC (volume or presence/absence) was not associated with FIB levels (Table 2). For the interpond models, a rain event 0–24 h BSC and a positive one mm unit change in rain volume resulted in 0.45 and 0.79 $\log_{10}$ MPN/100 mL increases in *E. coli* levels, respectively. As for physicochemical water quality parameters, pH had a negative association with FIB levels. For TC levels, the interpond and two intrapond models showed a decrease between -0.22 and -0.44 $\log_{10}$ MPN/100 mL for a one positive unit change in pH (Table 2). Similarly, the interpond and four out of five intrapond models demonstrated a negative association (-0.27 to -0.66 $\log_{10}$ MPN/100 mL) between *E. coli* levels and pH. Results of Bayesian mixed models for *E. coli* levels also demonstrated an association with dissolved oxygen (mg/mL) and $\log_{10}$ turbidity ($\log_{10}$ NTU) for pond-D and pond-I, respectively (Table 2). However, since the ROPE values of these two factors were between 2.5 % and 97.5 % (0.13 and 0.14, respectively), the effects on *E. coli* levels are uncertain based on (ROPE $>$ 97.5 % = negligible effect and $<$2.5 % = nonnegligible effect). As for associations with physicochemical water quality parameters and meteorological factors, it is important to note that the nonnegligible factors and the directions of the effect estimate were consistent between TC and *E. coli* levels for both the interpond and intrapond models. No additional weather or physicochemical water quality factor demonstrated any notable association with FIB levels (PS $\geq$ 0.50, PD $\geq$ 0.75, ROPE $\leq$ 0.25).

Results of Bayesian mixed models for environmental site observations are listed in Table S6. Overall, the effect of environmental site observation on FIB levels was dependent on the pond under consider-

Table 1
Results from the intercept-only Bayesian models of coregionalization (MOCR) for Ponds[a]

	Effective Range (m)[b]			Partial Sill[c]			Nugget[d]			Variance Ratio[e]		
	Median	Quartiles		Median	Quartiles		Median	Quartiles		Median	Quartiles	
		1st	3rd		1st	3rd		1st	3rd		1st	3rd
Total Coliforms (TC)												
Interpond	756.43	2194.59	861.08	0.82	0.32	1.40	0.53	0.52	0.56	1.55	0.62	2.50
Pond-A	74.20	164.20	49.33	0.18	0.15	0.21	0.04	0.01	0.05	4.33	14.01	4.62
Pond-D	73.17	147.56	49.91	0.20	0.18	0.23	0.03	0.01	0.04	5.81	27.35	5.50
Pond-F	68.70	166.02	44.50	0.09	0.04	0.15	0.15	0.08	0.19	0.65	0.48	0.79
Pond-H	75.52	187.03	48.70	0.53	0.45	0.61	0.03	0.01	0.03	18.89	32.73	13.94
Pond-I	86.28	244.90	52.80	0.79	0.36	1.22	0.92	0.80	1.02	0.85	0.44	1.19
E. coli												
Interpond	1512.10	2377.18	1289.77	0.60	0.50	0.68	0.37	0.36	0.39	1.62	1.39	1.89
Pond-A	68.56	178.89	43.64	0.10	0.05	0.12	0.24	0.19	0.28	0.42	0.28	0.43
Pond-D	60.59	106.84	8.60	0.03	0.02	0.0	0.32	0.29	0.34	0.10	0.07	0.12
Pond-F	56.99	96.65	39.59	0.06	0.03	0.07	0.27	0.23	0.30	0.22	0.15	0.25
Pond-H	68.06	137.42	45.77	0.08	0.02	0.09	0.42	0.36	0.49	0.19	0.06	0.20
Pond-I	69.00	142.51	45.67	0.05	0.03	0.06	0.41	0.37	0.45	0.13	0.09	0.12

[a] Since the aim of intercept-only Bayesian models of coregionalization (MOCR) was to quantify the variance ratio (spatial/nonspatial variance), individual pond models were not implemented for the four ponds where samples were collected from three or less sites: Pond-B, Pond-E, Pond-G, and Pond-C. Data for these ponds were used in the Interpond model.

[b] Effective range was the distance in meters at which *E. coli* levels became spatially independent.

[c] Partial sill = spatial variance.

[d] Nugget = nonspatial variance.

[e] Variance Ratio = variance attributable to spatial versus nonspatial sources (Partial Sill/Nugget).

Table 2
Results of Bayesian mixed models[a] that demonstrated an association for interpond and intrapond[b] between total coliform (TC) and *E. coli* levels ($\log_{10}$ MPN/100 mL) and water quality and weather factors [i.e., practical significance (PS) ≥ 0.50, probability of direction (PD) ≥ 0.75, and overlap with the region of practical equivalence (ROPE) ≤ 0.25]

Factor	Pond	MAP[c]	95 % CrI[d]	PD[e]	PS[f]	ROPE[g]
Total Coliform ($\log_{10}$ MPN/100 mL)						
Rain Event 0–24 h BSC[h] (Y/N)						
	Interpond	0.47	0.27, 0.64	1.00**	1.00	0.00**
	Pond-A	0.27	−0.07, 0.63	0.93*	0.82	0.17
	Pond-D	0.35	0.11, 0.58	1.00**	0.98	0.00**
	Pond-H	0.98	0.26, 1.75	1.00**	0.99	0.00**
	Pond-I	0.82	0.01, 1.52	0.97**	0.94	0.02**
Rain Volume 0–24 h BSC (mm)						
	Interpond	0.69	0.46, 0.97	1.00**	1.00	0.00**
	Pond-A	0.53	0.09, 1.06	0.99**	0.96	0.00**
	Pond-D	0.79	0.42, 1.09	1.00**	1.00	0.00**
	Pond-H	1.76	0.71, 2.83	1.00**	1.00	0.00**
	Pond-I	1.05	−0.04, 2.07	0.97**	0.96	0.02**
pH						
	Interpond	−0.22	−0.40, −0.03	1.00**	0.90	0.08
	Pond-D	−0.35	−0.68, −0.04	0.98**	0.93	0.04*
	Pond-F	−0.44	−0.89, 0.06	0.96**	0.91	0.07
E. coli ($\log_{10}$ MPN/100 mL)						
Dissolved Oxygen (mg/mL)		−0.12	−0.15, −0.09	1.00**	0.85	0.13
	Pond-D					
Log Turbidity (NTU)						
	Pond-I	0.22	−0.03, 0.48	0.96**	0.85	0.14
Rain Event 0–24 h BSC (Y/N)						
	Interpond	0.45	0.30, 0.62	1.00**	1.00	0.00**
	Pond-A	0.35	−0.12, 0.89	0.92*	0.84	0.13
	Pond-D	0.22	0.01, 0.45	0.98**	0.89	0.10
	Pond-H	0.86	0.17, 1.39	0.99**	0.99	0.00**
	Pond-I	0.74	0.30, 1.19	1.00**	1.00	0.00**
Rain Volume 0–24 h BSC (mm)						
	Interpond	0.79	0.57, 1.00	1.00**	1.00	0.00**
	Pond-A	0.92	0.22, 1.59	1.00**	0.99	0.00**
	Pond-D	0.39	0.07, 0.66	0.99**	0.96	0.02**
	Pond-H	0.96	0.14, 1.87	0.98**	0.97	0.00**
	Pond-I	1.23	0.60, 1.77	1.00**	1.00	0.00**
pH						
	Interpond	−0.27	−0.41, −0.10	1.00**	0.98	0.00**
	Pond-A	−0.34	−0.71, 0.05	0.96**	0.89	0.09
	Pond-D	−0.60	−0.89, −0.27	1.00**	1.00	0.00**
	Pond-F	−0.56	−1.08, −0.02	0.99**	0.96	0.02**
	Pond-I	−0.66	−1.36, −0.02	0.98**	0.95	0.02**

[a] For the Interpond model, a fixed effect of date and random effect of site ID nested in pond was used. For individual pond models with more than one sampling site, a fixed effect of date and random effect of site was used. Fixed and random effects were used to account for pseudo-replication and temporal autocorrelation.
[b] Ponds with more than three sampling sites were used to investigate intrapond relationships.
[c] Maximum a posteriori estimate (MAP).
[d] 95% credibility interval (CrI).
[e] Probability of direction (PD) values of 0.95, 0.975, 0.995, and 0.9995 correspond to two-sided frequentist P values of 0.10, 0.05, 0.01, and 0.001, respectively (McEgan et al., 2013; Murphy et al., 2022). PD ≥ 0.95 is marked by **, while PD ≥ 0.90 but < 0.95 is marked by *.
[f] Practical significance (PS). Values should be larger than 0.5 to indicate practical significance.
[g] Regional of practical equivalence (ROPE). The following cutoffs were used for ROPE interpretation: >99 % negligible effect, >97.5 % probably negligible effect, between 2.5 % and 97.5 % uncertain effect, <2.5 % nonnegligible effect, <1% significant effect (Makowski et al., 2019; Makowski et al., 2019). ROPE ≤ 0.025 is marked by **, while ROPE ≤ 0.05 but >0.025 is marked by *.
[h] BSC – before sample collection.

ation. For TC levels, 17 unique environmental site observations resulted in at least one model meeting the criteria of PS ≥ 0.5, PD ≥ 0.75, and ROPE ≤ 0.25. However, of those 17 factors, no factor was associated with more than two models indicating that environmental site observations driving TC levels were pond or scale dependent (Table S6). Additionally, of these 17 factors, there were only two associations where the effects on TC levels were certain to be nonnegligible (PD ≥ 0.95 and ROPE ≤ 0.025): buildings (within 100 m, Y/N) and recreation adjacent to pond (Y/N) increased TC levels by 2.33 and 2.36 $\log_{10}$ MPN/100 mL, respectively, for pond-I (Table S6).

For *E. coli* levels, 16 unique environmental site observations resulted in at least one model meeting the criteria of PS ≥ 0.5, PD ≥ 0.75, and ROPE ≤ 0.25 (Table S6). The presence of surface biofilm was associated with *E. coli* levels in the interpond model and four out of five intrapond models (pond-A, pond-D, pond-H, pond-I), while the remaining 15 environmental site observations were associated with less than two (≤2) models (Table S6). However, the direction of the effect differed between models with the presence of a surface biofilm having a positive association with pond-A and a negative association with the interpond, pond-D, pond-H, and pond-I. Overall, the presence of surface biofilms had a nonnegligible effect (PD ≥ 0.95 and ROPE ≤ 0.025) on only two models: interpond (MAP = −0.34) and pond-D (MAP = −80.33) models (Table S6). This inconsistency in the effect direction was also observed for the association of *E. coli* levels and waterfowl (duck and geese), location within a pond (e.g., middle vs inflow), and the presence of submerged vegetation

(Table S6). Furthermore, of the 16 environmental site observations, only six were certain to be nonnegligible (PD ≥ 0.95 and ROPE ≤ 0.025) associations, including a positive association between the presence of waterfowl and *E. coli* levels in pond-A (MAP = 0.50) and a negative association between emergent vegetation and *E. coli* levels in pond-F (MAP = -0.36) (Table S6).

Conditional inference trees

To characterize and visualize hierarchical associations between FIB levels and factors, separate conditional inference trees were implemented for TC and *E. coli*. Conditional inference trees divide data into smaller subsets, splitting from the root node (the variable that best splits the data) by attribute/parameter. Splits are made based on which attributes generate the highest information gain and result in the most homogenous subsets. In both the TC and *E. coli* conditional inference tree, pond was the root node, thus, the trees confirmed that the individual pond had a substantial impact on both TC and *E. coli* levels (Figs. 1 and 2). Based on the conditional inference tree, samples collected from pond-A, pond-B, pond-D, pond-E, pond-F, and pond-G during warmer air temperatures (days since first sampling ≤23 indicating warmer air temperature), when it rained, had the highest levels of TC bacteria (Fig. 1). Samples collected from pond-C, pond-H, and pond-I from sampling sites not within 100 m of buildings had the lowest levels of TC bacteria (Fig. 1). Referring to the *E. coli* conditional inference tree, samples collected during a rain event from ponds, where the bottom was clay, gravel, or silt or collected from sampling sites without the presence of a pipe, with low dissolved oxygen (≤9.9 mg/mL) had the highest *E. coli* levels (Fig. 2). The lowest *E. coli* levels came from ponds with sandy bottoms when the relative humidity was less than 85.4 % (Fig. 2).

Discussion

While freshwater systems are intrinsically complex, the overall findings of this study highlight the importance of scale when designing water quality monitoring programs or future water quality studies. Moreover, this study found that the factors driving microbial dynamics in ponds appeared to be pond-specific and that nearby sites (56 to 87 m) within the same pond represent separate water sources (i.e., separate microbial water quality) for risk assessment and management purposes. These results have implications for assessing surface water used for produce production, as currently proposed in the latest draft of Subpart E for FSMA's Produce Safety Rule (Food and Drug Administration., 2022) and suggest that water quality and potential surface water interventions are site specific and depend on local spatial–temporal factors.

Scale of analysis influenced observed sources of variance

The results of this study demonstrated that generally as scale increased (from intrapond to interpond), the variance attributable to spatial sources increased and the variance attributable to nonspatial sources decreased. This finding is consistent with previous literature demonstrating that factors driving FIB levels (TC and *E. coli*) in surface waters appeared dependent on the scale of analysis (Badgley et al., 2019; Rafi et al., 2018; Weller et al., 2022). A recent study in New York State (US) found substantial deviations in the percent variance attributable to spatial and nonspatial sources between waterways and watersheds (Weller et al., 2022) and concluded that factors associated with water quality differed by scale. Additionally, a prior study, conducted in both Texas (US) and Oklahoma (US), found a negative relationship between *E. coli* concentration and watershed size and stream order (a classification of streams based on their numbers of tributaries: the greater the stream order, the more tributaries) with *E. coli* concentrations generally decreasing with increasing stream order and watershed size (Rafi et al., 2018).

Meteorological drivers were consistently associated with FIB levels across spatial Scales, while associations with spatial factors appeared to be Pond-Specific

In the study reported here, spatial factors were extremely pond dependent; no spatial factor was associated with $\log_{10}$ FIB levels in

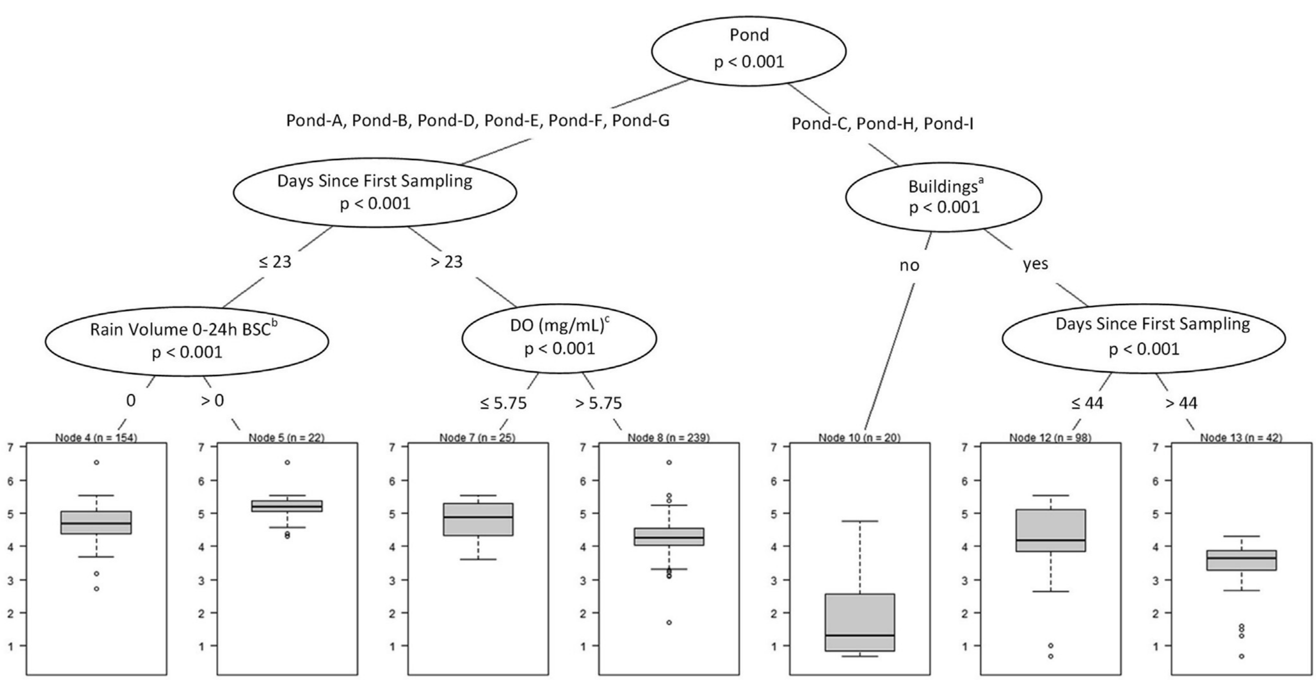

Figure 1. Conditional inference tree used to visualize hierarchical relationships between factors and total coliform (TC) levels (log10 MPN/100 mL). [a]Building within 100 m (y/n); [b]Rain volume (mm) 0–24 h before sample collection (BSC); and [c]dissolved oxygen (mg/mL).

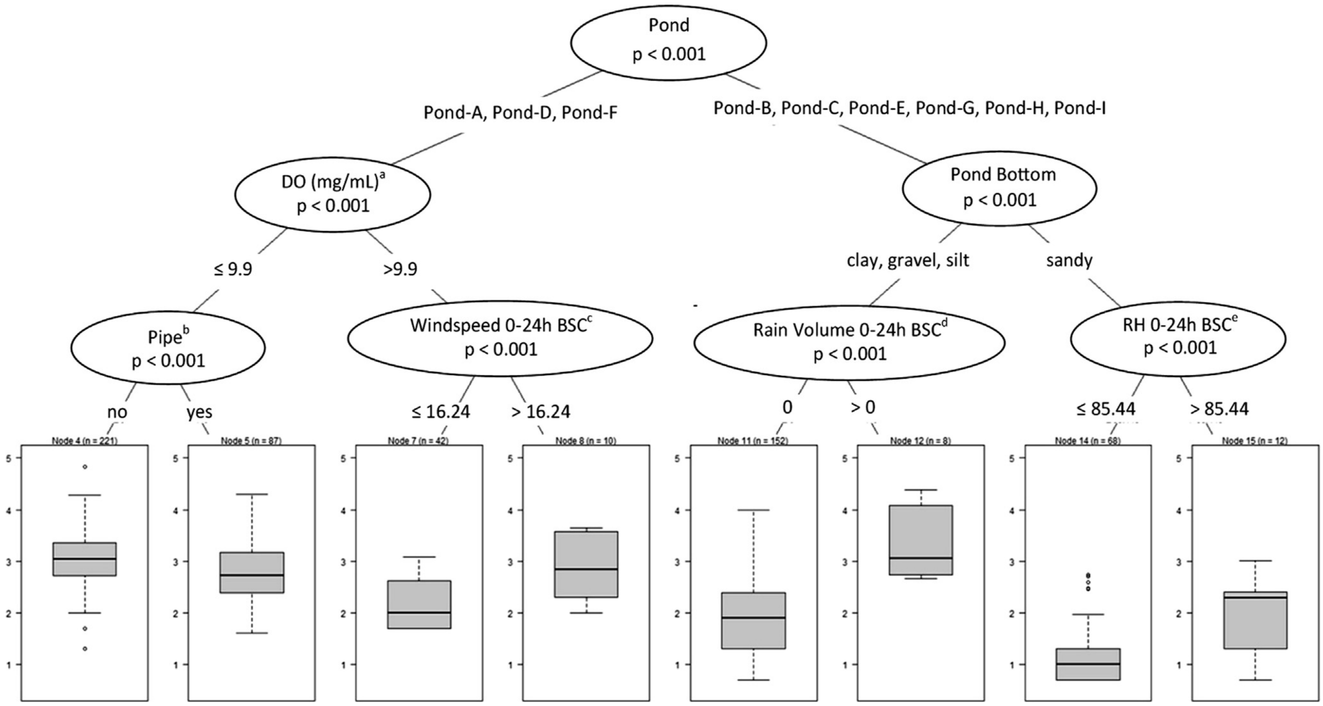

Figure 2. Conditional inference tree used to visualize hierarchical relationships between factors and *E. coli* levels ($\log_{10}$ MPN/100 mL). [a]Dissolved oxygen (mg/mL); [b]Pipe in or out of pond (y/n); [c]Windspeed (km/h) 0–24 h before sample collection (BSC); [d]Rain volume (mm) 0–24 h before sample collection (BSC); and [e]Relative humidity (%) 0–24 h before sample collection (BSC).

all ponds. Moreover, even when there was evidence of an association between the presence of surface biofilms and *E. coli* level for four of the five ponds, the direction, uncertainty, and strength varied considerably. These biofilms form on the surface of the ponds when algae or animals die at a rate faster than the naturally occurring bacteria can degrade it. Previous research has similarly shown that the spatial factors driving FIB levels in surface water varied between water sources and are often stream order/watershed size-dependent (Badgley et al., 2019; Murphy et al., 2022; Pandey et al., 2012; Partyka et al., 2018; Weller et al., 2020, 2022; Won et al., 2013). For example, random forest analysis identified that sampling site, with sampling sites being dependent on and exclusive to a waterway, as the top-ranked factor associated with *E. coli* levels in Arizona (US) canals (Weller et al., 2020), and as the second most important predictor for the human-associated fecal source gene HF183 and *E. coli* for water samples collected from nine watersheds in southwest Virginia (US) (Badgley et al., 2019). The findings of these and other studies support spatial factors driving fecal contamination are likely influenced by differences in land use type and watershed/waterway. Therefore, spatial factors should not be regarded as inherently risky or not risky during risk assessments of surface water, instead such categorization needs to be context-dependent for when, how, and what the water will be used for.

Meteorological factors were consistently associated with FIB levels across inter- and intrapond models. The research presented here demonstrated that rainfall, regardless of amount, resulted in an increase in $\log_{10}$ MPN/100 mL of both TC and *E. coli* levels, regardless of inter- or intrapond scale. It is well established that rainfall events can lead to runoff, resuspension of sediments, and sewage overflows (Fluke et al., 2019; Kistemann et al., 2002; Nagels et al., 2002; Pachepsky et al., 2011), which may result in increased levels of microorganisms in surface water. Yet previous studies evaluating correlations/associations between precipitation (rainfall) and microbial water quality yield inconsistent results (Francy et al., 2013; McEgan et al., 2013; Saturday et al., 2021; Schilling et al., 2009; Topalcengiz et al., 2017; Weller et al., 2022; Wilkes et al., 2009, 2011; Won et al., 2013). For example, studies in produce-growing regions of Florida (US) and Ontario (CAN) found varying results; one study observed no strong correlation between FIB levels and sampled water from Florida ponds, and another study found an association between FIB levels and sampled water from Ontario streams (McEgan et al., 2013; Wilkes et al., 2009). These differences suggest that the correlation or associations between elevated microbial levels and rainfall in surface waters may be dependent on characteristics of the individual water source and impacted by different regions, seasons, and interactions. The relationship between FIB levels and pH appears to be well established in prior literature, and in agreement with the findings reported here (Draper et al., 2016; Green et al., 2021; Gu et al., 2013; Topalcengiz et al., 2017). Model outputs in the study reported here observed that pH was consistently associated with a decrease in TC and *E. coli* $\log_{10}$ MPN/100 mL levels in individual ponds, and across all ponds. Overall, the findings reported here suggest that meteorological (e.g., rainfall) and physicochemical water quality (e.g., pH) factors could be incorporated as supplementary indicators into traditional microbial indicator-based water quality monitoring programs for ponds in the Southwestern VA produce-growing region (US). Integration of these data types into monitoring efforts could help guide decision-making in real time. However, additional follow-up work looking at mechanistic analysis would be fruitful in understanding how exact changes in variables impact changes in other variables.

Previous literature has established physicochemical water quality (e.g., turbidity, dissolved oxygen) and weather (e.g., air temperature, relative humidity) factors to be associated with microbial water quality (Belias et al., 2021; Green et al., 2021; Murphy et al., 2022; Strawn et al., 2013; Topalcengiz et al., 2017; Weller et al., 2020, 2020). While meteorological and physicochemical water quality factors other than rainfall and pH were associated with FIB levels in this study, the strength of association and the depth in the surrogate trees suggested that factors were pond dependent and rely on complex interactions. For instance, dissolved oxygen was associated with levels of TC only for six ponds when the air temperature was colder (days since first sampling > 23). This further emphasizes the relationship between microbial water quality, and environmental and physicochemical factors is complex, and varies over space (e.g., between waterway, water type, region) and time (e.g., season, day of sampling). However, there are some limitations to this study and caution should be used when generalizing the data. The study reported here only investigated one region of VA (US) and did not investigate across multiple years or seasons; thus, associations may have been missed. The conclusions surrounding attributable variance and drivers of pond water quality are specific to the geographic region, and sampling timeframe outlined in the study.

Data suggest sites within individual ponds should be regarded as separate sources

One of the aims of this study was to determine the distance at which FIB levels (TC and *E. coli*) within and between ponds were spatially independent. To the authors' knowledge, this is the first study to quantify intrapond distance to understand how spatial variability impacts FIB levels. Based on the MCOR, FIB levels were spatially independent for sites ~100 m apart within individual ponds suggesting that these sites should be considered as separate sources for monitoring purposes. The sizes of the ponds sampled in this study ranged from approximately 490 to 23,500 m^2 in area (Table S2). Considering the distances at which sampling sites are independent and the sizes of the ponds, it may be concluded that a single water sample inaccurately reflected the overall quality of water. This is especially important as larger operations with larger irrigation watercourses may have more than one intake pump within the same waterway. Thus, growers and produce industry stakeholders should use cautions when making conclusions on the acceptability of their surface water based on only taking samples from one or a limited number of sampling locations.

Previous literature has also demonstrated differences in FIB levels at different points within a waterway (Jenkins et al., 2012; Quilliam et al., 2011). For example, a study in Wales (UK) reported there were significant differences in *E. coli* levels on opposite riverbanks (Quilliam et al., 2011), while a study in Georgia (US) demonstrated that outflow concentrations of FIB levels were significantly lower than inflow concentrations in a pond (Jenkins et al., 2012). These results suggest that produce growers using surface water should consider testing more locations within individual water sources to gain a more complete understanding of their water source. However, increased locations and frequency may lead to an economic burden for produce growers who have previously self-reported that testing is economically demanding (Astill et al., 2018, 2019; Perry et al., 2019; Strohbehn et al., 2018). While the microbial water quality within a pond does vary, variance models showed more variability in *E. coli* levels was attributable to pond, than the sampling site location. Therefore, if water quality monitoring is only feasible for a small number of samples, it is more important to prioritize sampling across ponds on a farm at least once, rather than sampling one pond at many sites and another pond not at all.

Overall, the findings of this research demonstrate that while there are a few factors (rain and pH) that appear to be consistently associated with FIB levels in and between ponds, the majority of spatial factors driving FIB levels are pond dependent. Additionally, since FIB levels were found to be spatially independent for sites 56–87 m apart within individual ponds, a one-size-fits-all approach to water quality fails to account for the complexity of freshwater ecosystems. Since relatively few studies have performed similar analyses (Badgley et al., 2019; Rafi et al., 2018; Weller et al., 2022), future studies are needed in other produce-growing regions with other water types (beyond ponds), and over multiple growing seasons to determine if these findings are reproducible. Models built using larger, more representative datasets are needed, including multiregional and multiyear datasets, for the development of tools (i.e., models) that can be integrated into water quality monitoring plans, and subsequent risk management strategies. Moreover, given that there is uncertainty about the correlation between FIB levels (specifically *E. coli*) in surface water, and the overall prevalence of pathogens, future work should be performed on inter- and intrapond variations on the presence/absence of pathogens of interest (e.g., *Salmonella*) in surface waters. Additionally, future work taking samples at different depths might be useful to understand further intrapond variations. Therefore, the study reported here provides a framework for future studies and contributes data on the spatial versus temporal variance in FIB levels within and between ponds.

Declaration of Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Specialty Crops Research Initiative project 2019-51181-30016 from the US Department of Agriculture National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the US Department of Agriculture.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jfp.2023.100045.

References

Astill, G., Minor, T., Calvin, L., & Thornsbury, S. (2018). *Before Implementation of the food safety modernization act's produce rule: a survey of US produce growers.* https://www.ers.usda.gov/webdocs/publications/89721/eib-194.pdf?v=0. Accessed 31 March 2022.

Astill, G., Minor, T., Thornsbury, S., & Calvin, L. (2019). *US produce growers' decision making under evolving food safety standards.* https://www.ers.usda.gov/webdocs/publications/93242/eib-210.pdf?v=904.8. Accessed 31 March 2022.

Badgley, B. D., Steele, M. K., Cappellin, C., Burger, J., Jian, J., Neher, T. P., ... Wagner, R. (2019). Fecal indicator dynamics at the watershed scale: variable relationships with land use, season, and water chemistry. *The Science of the Total Environment, 697* 134113.

Belias, A., Brassill, N., Roof, S., Rock, C., Wiedmann, M., & Weller, D. (2021). Cross-validation indicates predictive models may provide an alternative to indicator organism monitoring for evaluating pathogen presence in Southwestern US agricultural water. *Frontiers in Water, 87*.

Bürkner, P. -C. 2017. *Advanced Bayesian multilevel modeling with the R Package brms*. arXiv preprint arXiv:1705.11123.

Bürkner, P.-C. (2017b). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software, 80*, 1–28.

Centers for Disease Control Prevention (2008*). Multistate outbreak of Salmonella saintpaul infections linked to raw produce (FINAL UPDATE)*. https://www.cdc.gov/salmonella/2008/raw-produce-8-28-2008.html. Accessed 15 August 2022.

Craun, M. F., Craun, G. F., Calderon, R. L., & Beach, M. J. (2006). Waterborne outbreaks reported in the United States. *Journal of Water and Health, 4*, 19–30.

Draper, A. D., Doores, S., Gourama, H., & Laborde, L. F. (2016). Microbial survey of Pennsylvania surface water used for irrigating produce crops. *Journal of Food Protection, 79*, 902–912.

Duris, J. W., Reif, A. G., Krouse, D. A., & Isaacs, N. M. (2013). Factors related to occurrence and distribution of selected bacterial and protozoan pathogens in Pennsylvania streams. *Water Research, 47*, 300–314.

Finley, A. O., Banerjee, S., & Carlin, B. P. (2007). spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software, 19*, 1–24.

Finley, A. O., S. Banerjee, & Gelfand, A. E. (2013). *spBayes for large univariate and multivariate point-referenced spatio-temporal data models.* arXiv preprint arXiv:1310.8192.

Fluke, J., González-Pinzón, R., & Thomson, B. (2019). Riverbed sediments control the spatiotemporal variability of *E. Coli* in a highly managed, Arid River. *Frontiers in Water, 4*.

Food and Drug Administration (2015). *Standards for the growing, harvesting, packing, and holding of produce for human consumption.* https://www.regulations.gov/document/FDA-2011-N-0921-18558. Accessed 14 August 2022.

Food and Drug Administration (2019). *Environmental assessment of factors potentially contributing to the contamination of romaine lettuce implicated in a multi-state outbreak of E. coli O157:H7*. https://www.fda.gov/food/outbreaks-foodborne-illness/environmental-assessment-factors-potentially-contributing-contamination-romaine-lettuce-implicated. Accessed 15 September 2022.

Food and Drug Administration (2022). *FSMA proposed rule on agricultural water*. https://www.fda.gov/food/food-safety-modernization-act-fsma/fsma-proposed-rule-agricultural-water#:~:text=Under%20the%20proposed%20requirements%2C%20covered,produce%20or%20food%20contact%20surfaces. Accessed 14 August 2022.

Francy, D. S., Stelzer, E. A., Duris, J. W., Brady, A. M., Harrison, J. H., Johnson, H. E., & Ware, M. W. (2013). Predictive models for *Escherichia coli* concentrations at inland lake beaches and relationship of model variables to pathogen detection. *Applied and Environmental Microbiology, 79*, 1676–1688.

Graciaa, D. S., Cope, J. R., Roberts, V. A., Cikesh, B. L., Kahler, A. M., Vigar, M., ... Montgomery, S. P. (2018). Outbreaks associated with untreated recreational water—United States, 2000–2014. *American Journal of Transplantation, 18*, 2083–2087.

Green, H., Wilder, M., Wiedmann, M., & Weller, D. L. (2021). Integrative survey of 68 non-overlapping upstate new york watersheds reveals stream features associated With aquatic fecal contamination. *Frontiers in Microbiology*. 2125.

Greene, S., Daly, E., Talbot, E., Demma, L., Holzbauer, S., Patel, N., ... Lynch, M. (2008). Recurrent multistate outbreak of *Salmonella* newport associated with tomatoes from contaminated fields, 2005. *Epidemiology and Infection, 136*, 157–165.

Gu, G., Luo, Z., Cevallos-Cevallos, J. M., Adams, P., Vellidis, G., Wright, A., & van Bruggen, A. H. (2013). Factors affecting the occurrence of *Escherichia coli O157* contamination in irrigation ponds on produce farms in the Suwannee River watershed. *Canadian Journal of Microbiology, 59*, 175–182.

Havelaar, A. H., Vazquez, K. M., Topalcengiz, Z., Muñoz-Carpena, R., & Danyluk, M. D. (2017). Evaluating the US food safety modernization act produce safety rule standard for microbial quality of agricultural water for growing produce. *Journal of Food Protection, 80*, 1832–1841.

Jenkins, M. B., Endale, D. M., Fisher, D. S., Adams, M. P., Lowrance, R., Newton, G. L., & Vellidis, G. (2012). Survival dynamics of fecal bacteria in ponds in agricultural watersheds of the piedmont and coastal plain of Georgia. *Water Research, 46*, 176–186.

Kistemann, T., Claben, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J., ... Exner, M. (2002). Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Applied and Environmental Microbiology, 68*, 2188–2197.

Lee, S. H., Levy, D. A., Craun, G. F., Beach, M. J., & Calderon, R. L. (2002). Surveillance for waterborne-disease outbreaks – United States, 1999–2000. *Morbidity and Mortality Weekly Reports, 51*, 1–47.

Makowski, D., Ben-Shachar, M. S., Chen, S., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology, 10*, 2767.

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw., 4*, 1541.

McEgan, R., Mootian, G., Goodridge, L. D., Schaffner, D. W., & Danyluk, M. D. (2013). Predicting *Salmonella* Populations from biological, chemical, and physical indicators in Florida surface waters. *Applied and Environmental Microbiology, 79*, 4094–4105.

Murphy, C. M., Strawn, L. K., Chapin, T. K., McEgan, R., Gopidi, S., Friedrich, L., … Danyluk, M. D. (2022). Factors Associated With *E. coli* Levels in and *Salmonella* Contamination of Agricultural Water Differed Between North and South Florida Waterways. *Frontiers in Water* 3.

Nagels, J., Davies-Colley, R., Donnison, A., & Muirhead, R. (2002). Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. *Water Science and Technology, 45*, 45–52.

Pachepsky, Y., Kierzewski, R., Stocker, M., Sellner, K., Mulbry, W., Lee, H., & Kim, M. (2018). Temporal stability of *Escherichia Coli* concentrations in Waters of Two Irrigation Ponds in Maryland. *Applied and Environmental Microbiology, 84*, e01876–e01917.

Pachepsky, Y., Shelton, D. R., McLain, J. E., Patel, J., & Mandrell, R. E. (2011). Irrigation waters as a source of pathogenic microorganisms in produce: A review. *Advances in Agronomy, 113*, 75–141.

Pandey, P. K., Soupir, M. L., Haddad, M., & Rothwell, J. J. (2012). Assessing the impacts of watershed indexes and precipitation on spatial in-stream *E.* Coli Concentrations. *Ecological Indicators, 23*, 641–652.

Partyka, M. L., Bond, R. F., Chase, J. A., & Atwill, E. R. (2018). Spatiotemporal variability in microbial quality of western US agricultural water supplies: A multistate study. *Journal of Environmental Quality, 47*, 939–948.

Perry, B., Shaw, A., Johnsen, E., Enderton, A., Strohbehn, C. H., & Naeve, L. (2019). Assessment of midwest growers' needs for compliance with the food safety modernization act produce safety rule. *Food Protection Trends, 39*, 212–217.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News, 6*, 7–11.

Quilliam, R. S., Clements, K., Duce, C., Cottrill, S. B., Malham, S. K., & Jones, D. L. (2011). Spatial variation of waterborne *Escherichia coli* – Implications for routine water quality monitoring. *Journal of Water and Health, 9*, 734–737.

Rafi, K., Wagner, K. L., Gentry, T., Karthikeyan, R., & Dube, A. (2018). *Escherichia coli* concentration as a function of stream order and watershed size. *Journal of Environmental Quality, 47*, 949–957.

Saturday, A., Lyimo, T. J., Machiwa, J., & Pamba, S. (2021). Spatial and temporal variations of faecal indicator bacteria in Lake Bunyonyi, South-Western Uganda. *SN Applied Science, 3*, 697.

Schilling, K. E., Zhang, Y.-K., Hill, D. R., Jones, C. S., & Wolter, C. F. (2009). Temporal variations of *Escherichia coli* concentrations in a large midwestern river. *Journal of Hydrology, 365*, 79–85.

Strawn, L. K., Fortes, E. D., Bihn, E. A., Nightingale, K. K., Gröhn, Y. T., Worobo, R. W., ... Bergholz, P. W. (2013). Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. *Applied and Environmental Microbiology, 79*, 588–600.

Strohbehn, C. H., Enderton, A., Shaw, A. M., Perry, B., Overdiep, J., & Naeve, L. (2018). Determining what growers need to comply with the food safety modernization act produce safety rule. *Journal of Extension, 56* n7.

Topalcengiz, Z., Strawn, L. K., & Danyluk, M. D. (2017). Microbial quality of agricultural water in central Florida. *PLoS One1, 12*, e0174889.

Truitt, L. N., Vazquez, K. M., Pfuntner, R. C., Rideout, S. L., Havelaar, A. H., & Strawn, L. K. (2018). Microbial quality of agricultural water used in produce preharvest production on the eastern shore of Virginia. *Journal of Food Protection, 81*, 1661–1672.

U.S Environmental Protection Agency (2012). *Recreational water quality criteria*. https://www.epa.gov/wqc/recreational-water-quality-criteria-and-methods. Accessed 7 Febuary 2022.

U.S Environmental Protection Agency (2009). *Review of published studies to characterize relative risks from different sources of fecal contamination in recreational water.* https://sarasota.wateratlas.usf.edu/upload/documents/Fecal%20Contamination%20in%20Recreational%20Waters.pdf. Accessed 7 Febuary 2022.

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications*. Springer Science & Business Media.

Weller, D., Brassill, N., Rock, C., Ivanek, R., Mudrak, E., Roof, S., ... Wiedmann, M. (2020). Complex interactions between weather, and microbial and physicochemical water quality impact the likelihood of detecting foodborne pathogens in agricultural water. *Frontiers in Microbiology, 11*, 134.

Weller, D. L., Love, T. M., Belias, A., & Wiedmann, M. (2020). Predictive models may complement or provide an alternative to existing strategies for assessing the enteric pathogen contamination status of Northeastern streams used to provide water for produce production. *Frontiers in Sustainable Food Systems, 4*, 151.

Weller, D. L., Murphy, C. M., Johnson, S., Green, H., Michalenko, E. M., Love, T. M. T., & Strawn, L. K. (2022). *Land use, weather, and water quality factors associated with fecal contamination of Northeastern streams that span an urban-rural gradient* (p. 3). Water: Front.

Weller, D. L., Weller, D. E., Strawn, L. K., & Love, T. M. (2022). *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*. bioRxiv

Wilkes, G., Edge, T., Gannon, V., Jokinen, C., Lyautey, E., Medeiros, D., ... Lapen, D. R. (2009). Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Research, 43*, 2209–2223.

Wilkes, G., Edge, T., Gannon, V. P., Jokinen, C., Lyautey, E., Neumann, N., ... Topp, E. (2011). Associations among pathogenic bacteria, parasites, and environmental and land use factors in multiple mixed-use watersheds. *Water Research, 45*, 5807–5825.

Won, G., Kline, T. R., & LeJeune, J. T. (2013). Spatial-temporal variations of microbial water quality in surface reservoirs and canals used for irrigation. *Agricultural Water Management, 116*, 73–78.

Zhang, X., Zhi, X., Chen, L., & Shen, Z. (2020). Spatiotemporal variability and key influencing factors of river fecal coliform within a typical complex watershed. *Water Research, 178* 115835.