



Article

Predicting Fecal Indicator Bacteria Using Spatial Stream Network Models in A Mixed-Land-Use Suburban Watershed in New Jersey, USA

Tsung-Ta David Hsu ¹, Danlin Yu ^{2,*}  and Meiyin Wu ^{1,3}

¹ New Jersey Center for Water Science and Technology, Montclair State University, Montclair, NJ 07043, USA

² Department of Earth and Environmental Studies, Montclair State University, Montclair, NJ 07043, USA

³ Department of Biology, Montclair State University, Montclair, NJ 07043, USA

* Correspondence: yud@montclair.edu; Tel.: +1-973-655-4313

Abstract: Good water quality safeguards public health and provides economic benefits through recreational opportunities for people in urban and suburban environments. However, expanding impervious areas and poorly managed sanitary infrastructures result in elevated concentrations of fecal indicator bacteria and waterborne pathogens in adjacent waterways and increased waterborne illness risk. Watershed characteristics, such as urban land, are often associated with impaired microbial water quality. Within the proximity of the New York–New Jersey–Pennsylvania metropolitan area, the Musconetcong River has been listed in the Clean Water Act’s 303 (d) List of Water Quality-Limited Waters due to high concentrations of fecal indicator bacteria (FIB). In this study, we aimed to apply spatial stream network (SSN) models to associate key land use variables with *E. coli* as an FIB in the suburban mixed-land-use Musconetcong River watershed in the northwestern New Jersey. The SSN models explicitly account for spatial autocorrelation in stream networks and have been widely utilized to identify watershed attributes linked to deteriorated water quality indicators. Surface water samples were collected from the five mainstem and six tributary sites along the middle section of the Musconetcong River from May to October 2018. The log₁₀ geometric means of *E. coli* concentrations for all sampling dates and during storm events were derived as response variables for the SSN modeling, respectively. A nonspatial model based on an ordinary least square regression and two spatial models based on Euclidean and stream distance were constructed to incorporate four upstream watershed attributes as explanatory variables, including urban, pasture, forest, and wetland. The results indicate that upstream urban land was positively and significantly associated with the log₁₀ geometric mean concentrations of *E. coli* for all sampling cases and during storm events, respectively ($p < 0.05$). Prediction of *E. coli* concentrations by SSN models identified potential hot spots prone to water quality deterioration. The results emphasize that anthropogenic sources were the main threats to microbial water quality in the suburban Musconetcong River watershed. The SSN modeling approaches from this study can serve as a novel microbial water quality modeling framework for other watersheds to identify key land use stressors to guide future urban and suburban water quality restoration directions in the USA and beyond.

Keywords: fecal indicator bacteria; land use; microbial water quality modeling; spatial autocorrelation; spatial stream network models; urban and suburban water quality



Citation: Hsu, T.-T.D.; Yu, D.; Wu, M. Predicting Fecal Indicator Bacteria Using Spatial Stream Network Models in A Mixed-Land-Use Suburban Watershed in New Jersey, USA. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4743. <https://doi.org/10.3390/ijerph20064743>

Academic Editor: Paul B. Tchounwou

Received: 3 February 2023

Revised: 3 March 2023

Accepted: 6 March 2023

Published: 8 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Good water quality ensures the well-being of urban and suburban residents, as it safeguards public health and provides economic benefits through recreational opportunities [1,2]. However, microbial water quality is often impaired due to expanded impervious areas and the improper management of sanitary infrastructures, such as combined sewer overflows and failing septic tanks, resulting in elevated concentrations of fecal indicator

bacteria and waterborne pathogens in urban and suburban waterways [1]. Therefore, it is essential to monitor microbial water quality regularly. Microbial water quality is assessed using fecal indicator bacteria (FIB), such as *E. coli*, fecal coliform, and enterococci. These FIBs originate from the gastrointestinal tract of human or warm-blooded animals and are introduced into the urban and suburban waterbodies through fecal matter [3]. Although FIBs are not pathogens per se, their presence indicates the potential cooccurrence of true pathogens [4]. Water quality criteria were hence established to protect public health based on epidemiological studies that investigate the relationships between the rates of gastrointestinal illness among swimmers and the levels of FIBs [5–8]. For example, the most recent USEPA’s Recreational Water Quality Criteria uses a geometric mean of 126 CFU of *E. coli* per 100 mL to indicate an estimated illness rate of 36 per 1000 primary contact recreators in fresh water [4]. Exceedance of the water quality threshold indicates an elevated risk of waterborne illness, and the waterbody is deemed impaired and needs to be listed on the List of Impaired Waters (i.e., Clean Water Act 303 (d) List). Water quality restoration plans, such as total maximum daily load (TMDL) plans, are then to be developed to help guide pollution reduction efforts in order to attain water quality standards [9]. Development of a TMDL plan requires identifying and quantifying sources of pollutants and determining the degree of reduction needed to meet the applicable water quality standards [10].

Mathematic models are heavily involved in the process of TMDL development for pollutant load estimation and source allocation in order to identify and quantify the degree of contamination in a waterbody for better remedial plans [11]. Mechanistic models that are based on underlying hydrological and biogeochemical processes, such as streambed bacterial deposition and decay in the natural system, have been widely utilized for TMDL development, such as Soil and Water Assessment Tool (SWAT) and Hydrological Simulation Program-FORTRAN (HSPF) [10,12,13]. These process-based models often demand more expertise in understanding detailed mechanisms and require more time in developing models [11]. On the other hand, empirical models that are based on statistical approaches to estimate pollutant loads or identify associations between stressors and response variables are also widely used in practice [11]. Without needing to first understand the underlying complex processes regarding the fate and transport of contaminants, empirical approaches serve as a statistical alternative to support TMDL development [11]. A variety of regression models have been applied to identify key environmental variables linked to deteriorating water quality indicators, including linear, logistic, and Poisson regressions [14–17]. Linear regressions are highly accessible, easy to implement and have been widely applied in water quality data analysis. They are broadly accepted among water resource stakeholders [18]. Building a valid regression model, however, often requires a relatively large number of observations and enhanced knowledge to identify the relationships between the response and input variables. While the latter can often be established under a theoretical framework and multivariate and/or nonparametric exploration, acquiring a large number of observations could prove to be difficult, especially when sample collection and processing are costly and take a long time. Oftentimes, compromises have to be made, accepting a larger variance for the sake of a relatively small sample size to proceed with the intended analysis. In addition, linear regression assumes no heteroscedasticity among the residuals and the independence of observations [19]. In reality, since “near things are more related than distant things” [20], observations from adjacent geographic locations tend to have similar characteristics. The term “spatial autocorrelation” or “autocovariance” describes the similarity of measurements as a function of the distance separating them [21,22]. Studies that consider spatial autocorrelation often use Euclidean distance to measure the straight-line distance between measurements and apply it in the terrestrial ecosystems [22]. However, due to the constraints of movement of aquatic organisms and materials within stream channels, for stream water quality modeling, along-channel in-stream distance might be more appropriate [22].

Modeling water quality with stream distance had not previously been well explored until the last decade. Money and Carter [23] developed a predictive model of *E. coli* based

on turbidity data with river-based distance. In addition to turbidity, other water quality and quantity parameters and watershed characteristics, such as land use type and locations of point source pollutions, often play essential roles in affecting microbial water quality, and have been widely considered in TMDL implementation plan development [24]. To take into consideration the potential spatial autocorrelation in the modeling process, spatial stream network (SSN) models were developed and applied in recent studies [21,25–28]. An SSN model employs a pre-established ArcGIS toolset and open source R package to provide an empirical approach to modeling stream water quality in response to watershed attributes with the consideration of spatial autocorrelation to better reflect stream characteristics [29,30]. Autocovariance of the SSN models was constructed based on moving average functions to address spatial autocorrelation and measure the degree of dependence among observations in the stream systems [29]. The conventional autocovariance model assumes isotropy, indicating that the autocorrelation among observations depends only on the distance separating them, but not on the direction [31]. Moving average functions address the direction-based autocorrelation in the stream systems in tail-up or tail-down models [29,31]. Since SSN modeling relies less on understanding underlying hydrological and biogeochemical processes contributing to water quality but utilizes readily available land use/land cover (LULC) data, it provides greater potential for wider applications by stakeholders engaging in water quality issues in urban and suburban environments with limited resources [22,26]. SSN modeling can also predict FIB at a predetermined interval along the stream channel, providing opportunities to identify hot spots for further water quality monitoring and restoration plans [32].

The Musconetcong River, from which we took water samples for the current study, is located in northwestern New Jersey and forms part of the National Wild and Scenic River System in recognition of its remarkable recreational resources [33,34]. Urban and suburban water recreation provides opportunities to promote physical and mental health among residents, foster environmental stewardship and community engagement, and support tourism and the local economy [2,34]. However, the water quality of the Musconetcong River has been historically impaired due to intensive urbanization and roadway development [33]. In fact, it has been listed in the New Jersey Integrated Water Quality Monitoring and Assessment Report (303 (d) List) due to the high abundance of fecal coliforms in northwestern New Jersey. Historically, human wastewater was not connected to sewer systems in the watershed, and various types of domestic and industrial wastewater effluents were discharged into the river through dysfunctional septic tanks. Other non-point source originated from livestock operations, wildlife, etc., contributing considerably to fecal contaminations throughout this mixed-land-use watershed [35]. In 2003, a TMDL plan was established to address microbial water quality deterioration here, requiring a 93% reduction in fecal coliforms [35]. Since then, various water quality restoration projects have been undertaken, including riparian buffers, prescribed grazing, herd reduction, sinkhole closure, and green infrastructure [33]. In 2018, a water quality monitoring study concluded that there had been a significant improvement in microbial water quality over the past decade; however, further reductions in *E. coli* loads were still required to meet the TMDL goal [33]. Local water quality stakeholders have been developing additional monitoring strategies and water quality restoration plans, with delisting of the Musconetcong River from the 303 (d) List as the goal.

In this study, we aimed to develop empirical models with consideration of the spatial autocorrelation and associated key watershed variables with fecal indicator bacteria in the suburban mixed-land-use Musconetcong River watershed. Specifically, our goal was to analyze patterns of *E. coli* concentrations in response to upstream land use attributes through spatial stream network models. We hypothesize that the elevated *E. coli* concentrations in the Musconetcong River were strongly associated with the upstream urban land use. The results from this study will not only provide insights into identifying key land use stressors to guide future water quality restoration directions in this suburban Musconetcong River watershed, but also serve as a potential water quality modeling framework for different

watersheds in the USA and beyond when resources are limited to urban and suburban water quality stakeholders.

2. Materials and Methods

2.1. Water Quality, Land Use and Precipitation Data

E. coli was used as a fecal indicator bacterium as it is more reflective of recent fecal contamination from warm-blooded animals compared to fecal coliforms or total coliforms [4]. Surface water samples were collected twice a month from May to October 2018, with three additional samplings from June to August for a total of 21 events at 5 mainstem and 6 tributary sites located throughout the middle section of the Musconetcong River watershed (Figure 1). No samples were collected when stream channels were dry. The sites were selected based on past water quality monitoring activities showing the significant potential of fecal contaminations. Grab samples were collected aseptically into 1 L sterilized polypropylene bottles and were placed immediately in a cooler during transit. Enumeration of *E. coli* was based on an EPA-approved method using mColiBlue24[®] Broth (Hach Method 10029). In brief, samples were filtered through mixed cellulose ester membrane filters (0.45 mm, 47 mm, Hach, Loveland Colorado) and then soaked with mColiBlue24[®] Broth selective medium. The filters were subsequently incubated at 35 °C for 24 h. After incubation, colonies showing a blue/indigo color were recorded as *E. coli*. Tentative *E. coli* colonies were further verified using brilliant green bile (BGB) and lauryl tryptose broth (LTB). The final results were reported as colony forming units (CFU) per 100 mL of water samples. After the processing of the samples, we checked the weather data on the days when samples were collected and determined if certain samples were collected during a storm event based on the weather data. All *E. coli* data were uploaded to the Water Quality Portal (<https://www.waterqualitydata.us/>, accessed on 5 December 2021).

The 2015 land use/land cover data were acquired from the New Jersey Department of Environmental Protection (NJDEP), Bureau of Geographic Information Systems. A modified Anderson coding system was used to categorize land use data [36]. The total drainage area for the whole Musconetcong River watershed was 120,941.71 acres, with 53.2% forest, 21.2% urban, and 14.0% agricultural land. Four land use categories were extracted as explanatory variables for spatial stream network (SSN) modeling, including urban (1000), pasture (2120), forest (4000) and wetland (6000). Due to the sampling cost and length, the sample size in the current study was not particularly large (21 events). While more detailed land use land cover products, such as the USGS's Multi-Resolution Land Characteristics (MRLC) land cover, are available, to avoid losing too many degrees of freedom in the empirical study, we elected to use the relatively coarse land use land cover categorization. Urban land use classification covered land use characteristics from rural single units to high-density multiple residential dwellings, commercial buildings, and industrial areas. Houses with various dwelling units or a high density in the study area were all connected to septic tanks instead of sewage pipes for waste disposal (Personal Communication with the Musconetcong Watershed Association). The study's purpose was to investigate in general how urbanization within a watershed impacts water quality. The pasture subgroup was selected from the agricultural land use categories (2000), as failing septic systems and poor livestock management practices were previously listed as major threats to microbial water quality [35]. Precipitation data were obtained from NJDEP, the Department of Water Monitoring & Standards (<https://njdep.rutgers.edu/rainfall/>, accessed on 5 December 2021). The rainfall threshold for a storm event was defined as 12.7 mm (0.5 inch) within 36 h, the same definition used in the Watershed Restoration and Protection Plan for the Musconetcong River Watershed created by the Rutgers Cooperative Extension Water Resources Program [37]. A summary of analytical methods and data sources used in this study is shown in Table 1.

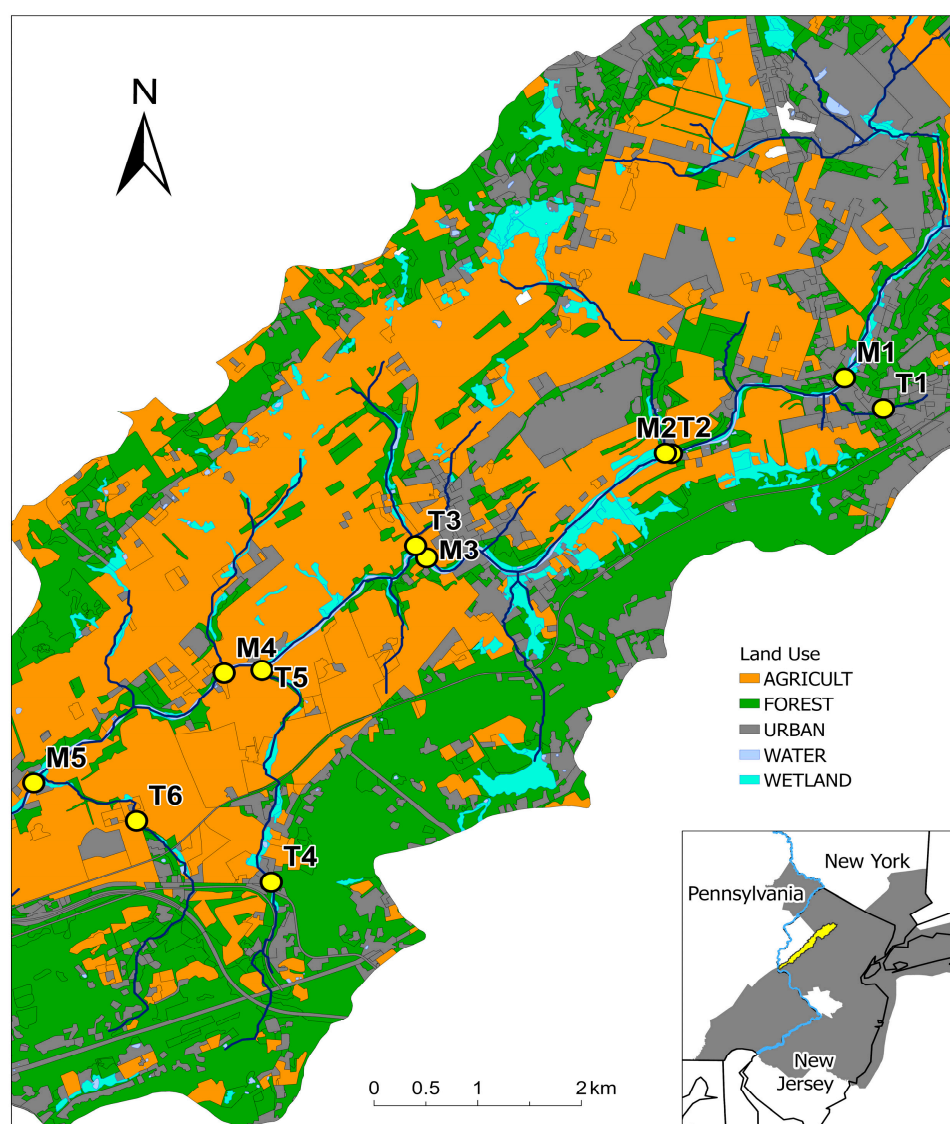


Figure 1. Sampling locations and adjacent land use types in the Musconetcong River watershed in this study. Sampling locations are denoted by yellow dots. The gray area on the map inlet indicates the New York–New Jersey–Pennsylvania metropolitan area (Source: United State Census Bureau).

Table 1. Analytical methods and data sources used in this study.

Data	Sources/Methods
<i>E. coli</i>	EPA approved Hach mColiBlue24 [®]
Urban land use	NJDEP, Bureau of Geographic Information Systems
Pasture land use	NJDEP, Bureau of Geographic Information Systems
Forest land use	NJDEP, Bureau of Geographic Information Systems
Wetland land use	NJDEP, Bureau of Geographic Information Systems
Precipitation	NJDEP, Department of Water Monitoring & Standards

2.2. Spatial Stream Network Modeling

To explore the impact of land use in the watershed on water quality, we extended from traditional regression analysis to adopt the spatial stream network model in this study [22,38,39]. Spatial stream network models are generalized linear mixed models allowing explanation of variance in observations with both fixed and random effects due

to spatial autocorrelation among geographically observed events [40,41], which can be expressed as follows:

$$y = X\beta + z + \varepsilon$$

where y is a vector of observations, X is a matrix for fixed effects that explain the variance of the observations that can be captured by general spatial patterns, β is a vector of coefficients for X , z accounts for random effects resulting from spatial autocorrelation that cannot be explained by fixed effects, and ε is a vector of independent random errors [21,32]. The spatial autocorrelation structures that describe the random effects can be modeled with linear with sill, Matérn, exponential, Epanechnikov, and spherical models [29]. Parameters for spatial autocovariance functions include nugget, partial sill, and range. The nugget of the autocovariance function describes the portion of variance that cannot be explained, the partial sill accounts for the variance that depends on the distance among observations, and the range indicates the minimal distance at which observations are no longer spatially correlated [32]. SSN models can also incorporate autocovariance models based on Euclidean distance, accounting for terrestrial and atmospheric factors that are stream-independent [22]. Both tail-up and tail-down autocovariance structures were established to estimate spatial autocorrelation. In tail-up models, moving average functions run upstream only and split at confluences. Spatial weights are required to estimate the proportion of upstream tributary influence based on flow volume, watershed area, or other relevant attributes [22]. Tail-up models are only applicable to flow-connected sites, where water directly flows from an upstream to a downstream site characterized by passive downstream movement, e.g., temperature, bacteria, or sediment [22]. In tail-down models, moving average functions run in the downstream direction unilaterally until reaching the most downstream location of the stream network, indicating the possibility of autocorrelation among all locations [29]. In addition to flow-connected sites, tail-down models can be applied to flow-unconnected sites, where some upstream movement is expected to facilitate the connectivity for a given attribute, e.g., fish or macroinvertebrates [22].

In order to perform spatial stream network (SSN) modeling, an SSN object was first created using STARS geoprocessing toolset [21] in ArcGIS 10.8 (ESRI, Redlands, CA, USA). STARS is used to build a landscape network (LSN) as a personal geodatabase to represent spatial relationships, such as flow connectivity, direction, and distance [29] and transform the geometry, attribute data, and topological relationships among features of GIS datasets into an SSN object that can be easily accessed and analyzed in R statistical software with the *ssn* package [30] for SSN modeling. The process included building a landscape network (LSN), creating reach contributing areas [42], calculating RCA attributes and area, accumulating watershed attributes, incorporating the sampling locations into the LSN, calculating watershed attributes, and calculating upstream distance, segment proportion influence, and additive function [43]. In the watershed of the current study, we obtained a total of 427 nodes in the stream network. The entire process of generating the SSN object was automated, and once generated in ArcGIS, the SSN object was then exported to R statistical software [44] for SSN modeling.

In this study, the *ssn* package in R [30,39] was used to fit spatial stream network models to the observation data based on 11 sampling locations. Two models were established and calibrated. The first model used the geometric mean concentrations of *E. coli* for all sampling events as the response variable; the second model used the geometric mean concentrations of *E. coli* during storm events only as the response variable. Shapiro–Wilk tests were used to test the normality of the distributions of the original values [45]. If the normality assumption was not satisfied, the values of geometric mean concentrations were log-transformed. A nonspatial model based on an ordinary least squares regression and two spatial models based on Euclidean and stream distance, respectively, were established to incorporate four upstream watershed attributes as explanatory variables, including urban, pasture, forest, and wetland. Percentage upstream land use values were derived from dividing the total accumulated area of upstream land use by the total upstream catchment area processed and generated via the STARS toolset. Variance inflation factors (VIF) were

calculated to check multicollinearity among the four land use variables. Tail-up models were applied in the spatial autocovariance functions for the SSN modeling based on stream distance, as the movement of bacteria along the stream networks is characterized by passive downstream transport [22]. Likelihood approaches were utilized to estimate the regression coefficients b for the matrix of fixed effects [46]. Maximum likelihood (ML) estimators have been shown to be greatly biased when the number of observations is small [47], and thus restricted maximum likelihood (REML) was used. An alpha value of 0.05 was used to evaluate whether the relationship between watershed variables (urban, agricultural, forest, or wetland) and *E. coli* concentrations in the models was statistically significant. The root mean square prediction error (RMSPE) of the leave one out cross validation (LOOCV) was used as the criterion to compare the predictive capability among models. LOOCV was performed by excluding one observation at a time. Submodels were calculated and compared with the excluded samples [48]. A lower RMSPE of the LOOCV indicated a better model performance [32]. The coefficient of determination (R^2) was used to measure the proportion of variance in the observations explained by the fixed or random effects of each model.

2.3. Predicting *E. coli* Concentrations

One of the practical uses for establishing the empirical SSN model for the Musconetcong River watershed is that we can use the model to predict the concentration of *E. coli* along the river. Predictions were performed for the geometric mean concentrations of *E. coli* for all sampling events and during storm events based on Euclidean distance and stream distance models, respectively. A total of 816 predictive sites were assigned evenly across the Musconetcong River stream network at a 300-m interval. Determination of predictive sites was conducted using the STARS geoprocessing toolset [21] in ArcGIS 10.8. Predictions of *E. coli* concentrations were performed in the R statistical software [44] with SSN package [30]. The results were illustrated on stream maps illustrating both the $\log_{10}$ geometric mean concentrations of *E. coli*. The capability of predicting the concentration of *E. coli* provides both a means to assess modeling performance and a supplement to the regular sampling routine for water quality monitoring and management.

3. Results

3.1. Microbial Water Quality and Land Use

All 21 sampling events were recorded for most of the locations, except for T1 (3) and T3 (19). Among them, three were defined as storm events (accumulated rainfall within 36 h was 13.2, 22.9, and 41.9 mm, respectively), except for T1 (one event, accumulated rainfall was 22.9 mm). The geometric mean concentrations of *E. coli* for all sampling events among the mainstem sites were similar (240 to 296 CFU/100 mL), while those among the tributary sites had a greater variation (73.3 to 1355.6 CFU/100 mL). Four of the six tributary sites exceeded the geometric mean threshold of *E. coli* (126 CFU/100 mL) compared to one of the five mainstem sites for all sampling events (Figure 2). During storm events, the geometric mean concentrations of *E. coli* among the mainstem sites ranged from 303 to 811 CFU/100 mL), while those among the tributary sites also had a greater variation (585.7 to 7300 CFU/100 mL). All of the sites exceeded the geometric mean threshold of *E. coli* (Figure 2). The upstream land use characteristics of sampling locations were similar among the mainstream sites, while those of tributary sites exhibited a greater variety as each site demonstrated unique patterns in this suburban mixed-land-use watershed (Figure 3). The percentage of upstream urban land ranged from 21.3% to 22.5% and from 8.2% to 57.4% for the mainstem and tributary sites, respectively. The percentage of upstream pasture varied from 0.0% to 0.2% and 0.1% to 20.3% for the mainstem and tributary sites, respectively. The percentage of upstream forest ranged from 41.0% to 42.9% and from 14.6% to 58.7% for the mainstem and tributary sites, respectively. The percentage of upstream wetland ranged from 6.3% to 6.7% and 0.3% to 6.4% for the mainstem and tributary sites, respectively. Overall, pasture land dominated the study area, with urban lands clustered toward the

upper portion. Forest land is concentrated toward the upstream area of tributaries in the study area (Figure 1).

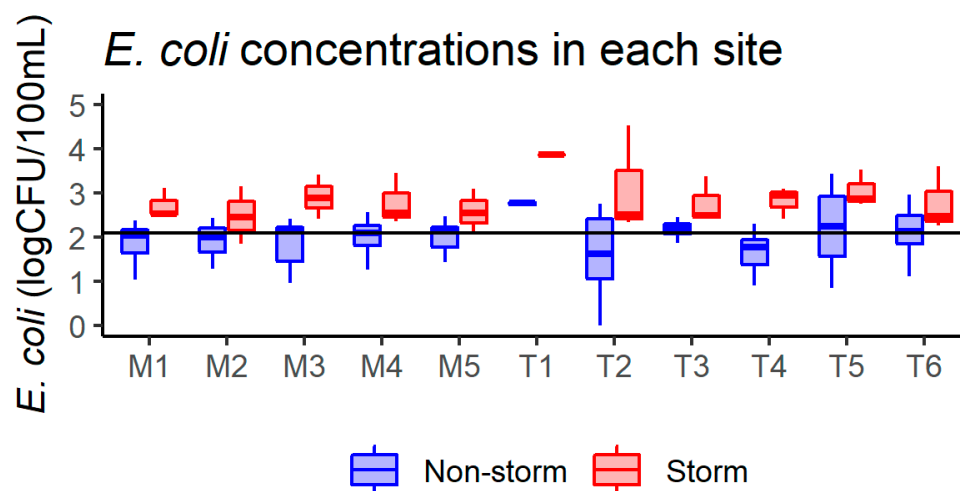


Figure 2. *E. coli* concentrations in a logarithmic scale for each site in this study. Red boxes indicate range of *E. coli* concentrations during storm events. Blue boxes indicate range of *E. coli* concentrations during non-storm events. The threshold of the rainfall amount for a storm event is 12.7 mm (0.5 inch) within 36 h. Black line denotes the geometric mean threshold of *E. coli* concentration of 126 CFU/100 mL (2.1 log₁₀ CFU/100 mL). M indicates mainstem sites, whereas T indicates tributary sites of the sampling locations.

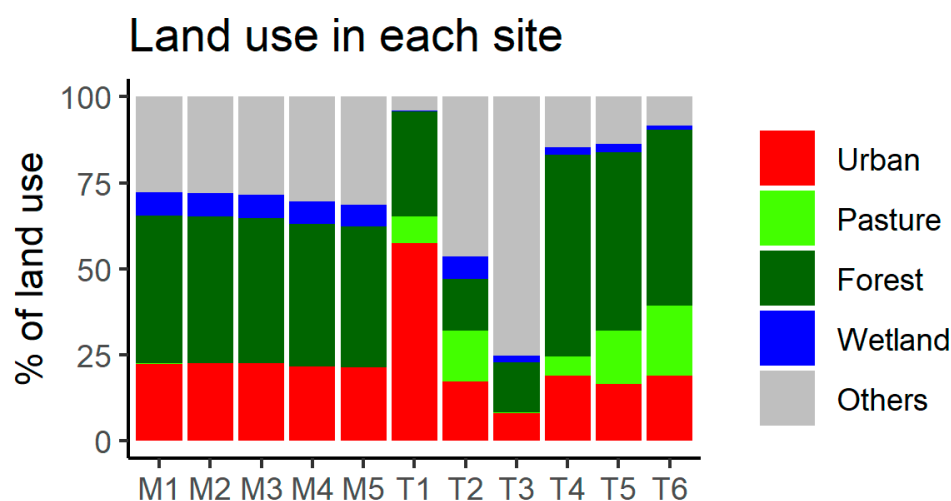


Figure 3. Percentage of each land use for each site in this study. Only the land use classifications used for the spatial stream network (SSN) models in this study are shown, including urban (red), pasture (light green), forest (dark green), and wetlands (blue). M indicates mainstem sites, whereas T indicates tributary sites of the sampling locations.

3.2. Spatial Stream Network Modeling

Three individual spatial stream network (SSN) models with three distinct autocovariance structures were constructed for each response variable in this study, including no spatial correlation (ordinary least squares), straight-line distance (Euclidean distance), and along-channel in-stream distance (stream distance). Variance inflation factors (VIF) for all four land use variables were below 2, indicating that no multicollinearity among them was detected. The original values of the geometric mean concentrations of *E. coli* were not normally distributed. Therefore, the original values were log-transformed to satisfy the normal distribution assumption. Table 2 shows the SSN model autocovari-

ance structures, coefficients for explanatory variables, and performance. Applying SSN models identified that upstream urban land was positively and significantly associated with the $\log_{10}$ geometric mean concentrations of *E. coli* for all events and during storm events, respectively ($p < 0.05$). Upstream pasture land was positively and significantly correlated with the $\log_{10}$ geometric mean concentrations of *E. coli* during storm events only ($p < 0.05$). Although insignificantly, the geometric mean concentrations of *E. coli* for all sampling events demonstrated negative correlations with upstream wetland ($p = 0.06$). SSN models improved the overall coefficient of determination (R^2) as the additional portion of variance can be explained by the random effects attributed to the spatial autocorrelation. In fact, based on stream distance, the overall SSN models explained nearly 100% of variance compared to the ordinary least squares models, with 20% of variance being unaccounted for. The lowest level of unexplained variance was found for SSN models based on stream distance for both response variables (Nugget < 0.0001). Prediction errors also improved from non-spatial to spatial models. SSN models based on Euclidean distance showed the lowest value of LOOCV RMSPE for $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events (1.110), while the lowest value of LOOCV RMSPE was seen for $\log_{10}$ geometric mean concentrations of *E. coli* during storm events modeled based on stream distance (1.290). Overall, the distance with no spatial autocorrelation observed for $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events was shorter than that for the storm events based on either Euclidean (Range 2.31 vs. 3.11 km) or stream distance (Range 2.87 vs. 3.37). The best stream distance autocovariance functions were linear with sill tail-up for both $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events and during storm events.

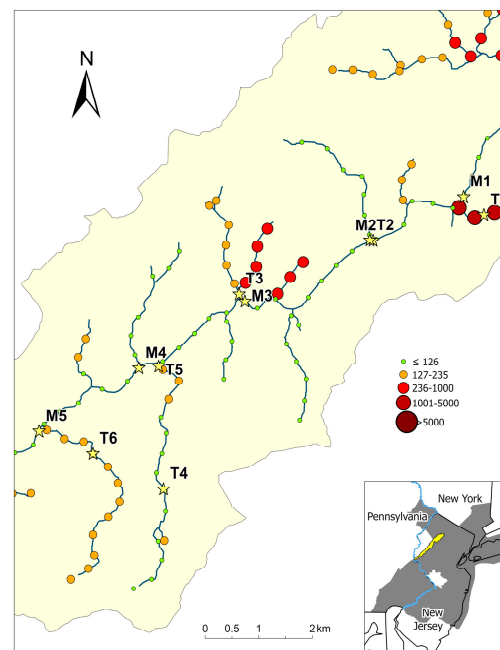
3.3. Predicting *E. coli* Concentrations

Euclidean distance and stream distance models were chosen to predict the $\log_{10}$ geometric means concentrations of *E. coli* for all events and during storm events, respectively, due to lower root mean square percentage error (RMSPE) values (Table 2). Figure 4 shows both the geometric mean concentrations of *E. coli* of the 11 sampling locations (stars) as well as the predicted geometric mean concentrations of *E. coli* (solid circles) in the study area. Overall, the predicted geometric mean concentrations of *E. coli* were higher during storm events (Figure 4b) than for all sampling events (Figure 4a). Predictive values for the geometric mean concentrations of *E. coli* ranged from 1 to 5012 CFU/100 mL and from 43 to 30,903 CFU/100 mL for all events and during storm events, respectively.

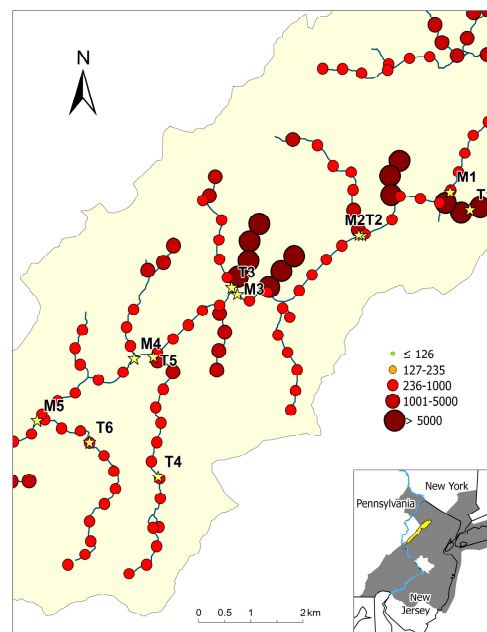
Table 2. SSN model autocovariance structures, coefficients for explanatory variables, and performance for two response variables (overall geometric mean of *E. coli* and geometric mean of *E. coli* during storm events).

Response Variable	Autocovariance					Coefficients for Explanatory Variables				Model Performance			
	Model	Structure	Nugget	Partial Sill	Range (km)	% Urban	% Pasture	% Forest	% Wetland	Fixed Effect R ²	Random Effect R ²	Total R ²	LOOCV RMSPE
log ₁₀ Geomean (overall)	OLS	Nugget only	0.041	NA	NA	* 0.018	−0.003	−0.006	^ −0.066	0.81	NA	0.81	1.234
	Euclidean Distance	Gaussian Euclidean + Nugget	0.028	0.014	2.31	* 0.018	−0.003	−0.005	^ −0.068	0.83	0.06	0.89	1.110
	Stream Distance	LinearSill Tailup + Nugget	0.000	0.040	2.87	* 0.018	−0.003	−0.006	^ −0.068	0.81	0.19	1.00	1.226
log ₁₀ Geomean (storm)	OLS	Nugget only	0.054	NA	NA	* 0.021	0.014	−0.008	−0.026	0.83	NA	0.83	1.396
	Euclidean Distance	Gaussian Euclidean + Nugget	0.033	0.005	3.11	* 0.022	0.016	−0.008	−0.022	0.84	0.02	0.86	1.346
	Stream Distance	LinearSill Tailup + Nugget	0.000	0.038	3.37	* 0.021	* 0.019	−0.009	−0.024	0.83	0.17	1.00	1.290

* $p < 0.05$, ^ $p = 0.06$.



(a)



(b)

Figure 4. Prediction of (a) overall geometric mean concentrations of *E. coli* and (b) geometric mean concentrations of *E. coli* during storm events. Color and size of the circles denote the geometric mean concentrations of *E. coli*, with larger and darker circles indicating greater geometric mean concentrations of *E. coli*. Stars indicate the sampling locations in this study.

4. Discussion

4.1. Spatial Stream Network Modeling Performance

Spatial stream network modeling provides an empirical approach for urban and suburban water quality stakeholders to analyze the spatial distribution of parameters without the need to first understand the underlying hydrological and biogeochemical processes that lead to water quality impairment. The application of SSN models has been introduced

to a wide range of water quality indicators, such as temperature [49], dissolved oxygen [50], total phosphorus [26], and macroinvertebrates [51,52]. However, the modeling of fecal indicator bacteria using SSN models had not been documented until recently [32,53]. Holcomb and Messier [53] modeled fecal coliforms in a mixed-land-use watershed in North Carolina and identified agricultural land use, forest cover, antecedent precipitation, and temperature as being strongly associated with mean fecal coliform concentrations. Neill and Tetzlaff [32] used *E. coli* as a fecal indicator bacterium in an agriculture-dominant watershed with sporadic urban development in Scotland and found that the anthropogenic impact index (lumped indicator for potential contamination from human point sources) was significantly correlated with 5th, 50th, and 95th percentile *E. coli* concentrations. Similar to Neill and others (2018), modeling based on stream distance in this study demonstrated improvement in the total R square and root mean square percentage error (RMSPE) for geometric mean concentrations of *E. coli* during storm events. However, only an improvement in the total R square was observed for the model based on stream distance for geometric mean concentrations of *E. coli* for all sampling events. This may be due to limitations regarding the number of sites as well as the locations of the mainstem sites, as they were relatively highly clustered compared to the overall watershed. An improvement in RMSPE was not observed in Holcomb et al. (2018) either. In this study, the best autocovariance functions were linear with sill with stream distance for both $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events and during storm events, indicating a pattern in which the variability among observations increases linearly with the separation distance until it reaches the maximum difference [54]. Overall, the range for the SSN models for storm events was greater than that for all sampling events, and the stream distance models had a greater range than the Euclidean distance models did, indicating heavier impacts on *E. coli* concentrations from upstream areas during storm events.

4.2. Suburban Land Use and Microbial Water Quality

Various factors could lead to the deterioration of microbial water quality in urban and suburban waterways [1]. In this study, upstream urban land was identified as being positively correlated with either $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events or during storm events only. Using the same SSN modeling approach, Neill and Tetzlaff [32] identified human (leaking sewage pipes and failing septic tanks) and farmyard sources as being significantly associated with 5th, 50th, and 95th percentile concentrations of *E. coli*. A variety of mechanisms, such as failing onsite wastewater treatment systems (i.e., septic tanks), combined sewer overflows (CSOs), sanitary sewer overflows (SSOs) and urban runoffs can introduce human fecal contamination into adjacent waterways in urban and suburban watersheds [1]. Increased risks of childhood emergency department visits and infectious diarrhea associated with combined sewer overflow and septic sites have been documented [55,56]. In addition to urban land use, upstream pasture land was found to be positively and significantly associated with $\log_{10}$ geometric mean concentrations of *E. coli* during storm events in this study. Holcomb et al. (2018) also reported similar significant relationships between agricultural land use and fecal coliform. Ill-considered agricultural practices in suburban watersheds, such as concentrated animal feeding operations, overgrazing and manure applications could lead to water quality deterioration [57,58]. Davies-Colley et al. (2004) also demonstrated that a dairy cow herd produced more than 50,000 CFU/100 mL of *E. coli* when crossing a stream. A significant relationship was found between upstream pastureland use and *E. coli* for storm events only in this study, illustrating the impact of agricultural runoff on microbial water quality during storm events. For instance, higher concentrations of *E. coli* and *Salmonella* were identified in irrigation ponds at two produce farms after rain events [59]. Noteworthy, a negative association (although only marginally significant, $p = 0.06$) was identified between upstream wetlands and geometric mean concentrations of *E. coli* for all sampling events. Wetlands can provide essential ecosystem services, such as water purification and runoff reductions [60]. Reductions in fecal indicator bacteria or pathogens provided by wetlands

have been well documented [61–63]. Hsu et al. (2017) reported an average 22.3% reduction in *E. coli* across two wetlands receiving inflow from urban waterways [63]. The results from this study also reinforce the concept of using wetlands as the best management practice for microbial water quality restoration. For example, a variety of constructed wetlands have been implemented to treat domestic and agricultural wastewater as well as urban stormwater runoff [64].

4.3. Extreme Weather Conditions and Microbial Water Quality

All of the geometric mean concentrations of *E. coli* during storm events exceeded the geometric mean threshold of *E. coli* of 126 CFU/100 mL in this study. Extreme weather events, such as heavy precipitation, have been well documented to degrade water quality by elevating the concentrations of FIB in urban and suburban waterways through stormwater runoff and combined and sanitary overflows [1]. These heavy rainfall events significantly contributed to waterborne disease outbreaks in the United States and Canada [65,66]. In the United States, among the 548 reported outbreaks analyzed from 1948 through 1994, 68% of them were preceded by the highest 20% of precipitation events [65]. In Canada, the heaviest 7% of precipitation events increased the relative odds of an outbreak by 2.3-fold [66]. In fact, more than 400,000 cases of acute gastrointestinal illness (AGI) were attributed to a drinking water treatment plant overwhelmed by high turbidity load after a period of heavy precipitation in Milwaukee, Wisconsin in 1993 [67]. Downstream of the Musconetcong River in Philadelphia, where the Delaware River serves as the drinking water source, a significant increase in waterborne AGI following precipitation above the 95th percentile was documented [68]. Under the current trend of global climate change, the National Climate Assessment projected an almost 50% increase in the total annual precipitation falling in the heaviest cases (1%) by the late 21st century under the higher scenario (RCP 8.5) in the northeastern United States [69]. This will further increase the frequency and intensity of urban stormwater runoffs and combined and sanitary overflows, degrading microbial water quality and increasing potential public health risk in the Musconetcong River and downstream drainage area in the future if no additional water quality protection action is implemented. The land use variables identified from the SSN modeling results can be taken into consideration when developing water quality restoration or climate resiliency plans to address water quality and public health concerns in urban and suburban areas, such as establishing municipal separate storm sewer system (MS4), expanding public sanitary sewer systems, regulating private sewage disposal practice, and promoting the use of green infrastructure.

4.4. Predicting Fecal Contamination Hot Spots and Future Directions

Since field water sampling is time-consuming and costly, water quality modeling can serve as a cost-effective and timely approach to enhance existing monitoring programs in conjunction with actual field samples. It can also provide stakeholders with a mean to quickly identify “hot spots” during the site selection process for further water monitoring and water quality enhancement actions. Spatial stream network predictive modeling provides a feasible approach to predict riverine *E. coli* concentrations with a predetermined stream-distance interval. The predicted results can be used to prioritize sites for further monitoring and subsequent restoration practices. Combined with local knowledge, the modeling results can be tailored to form a localized site-specific management approach. In this study, predictive sites near T1 in a catchment basin with a high percentage of urban land showed elevated *E. coli* concentrations for all sampling events and during storm events. Another example is shown in the next four predictive sites downstream of T4. The predicted value for *E. coli* for all sampling events was below the geometric mean threshold (126 CFU/100 mL). However, the predicted value for storm events was above the single-sample maximum threshold (235 CFU/100 mL), indicating the significant impact of agricultural runoff as a result of heavy precipitation on microbial water quality. Different localized water quality restoration approaches should be developed for both areas, such as

improving the maintenance of septic systems in urbanized areas and establishing riparian buffers in agriculture-dominant sites.

For TMDL process development, SSN models become a cost-effective approach along with field sampling programs for water quality stakeholders to engage in watershed management with the support of readily available land use land cover data, open-source R packages, and geographic information systems. In fact, the current study area was located in one of the TMDL segments of the Musconetcong River, and additional segments are located both upstream and downstream of the current study area [35]. Therefore, the same SSN approach can be extended to additional TMDL segments within the Musconetcong River watershed to provide a comprehensive understanding for integrated water quality management (NJDEP, 2003). Incorporating additional sampling locations is essential to encompass a larger watershed to improve SSN modeling. Money and Carter [23] suggested “a minimum of 10–50 data points should exist to construct a correlation model depending on the watershed size”. In our current study, because of cost and the size of our watershed, we only collected 21 data points to establish the model. The sample size is on the small end and does not have sufficient degrees of freedom for us to set up a model training and testing routine. While water monitoring is an ongoing endeavor for environmental management teams at both the local and state levels, we hope in the future that expanding sampling locations will allow model validation to encompass both training and validation datasets. It was also suggested that the placement of sampling locations in relation to the whole watershed could impact the outcome of the spatial modeling [38]. For instance, a better overall error map may be produced by placing sites close to the origins of tributaries, while a good estimate of tail-up autocovariance functions could be generated by selecting sites between confluences. In addition, a spatial stream predictive model in conjunction with other field and laboratory observations (e.g., sanitary survey or microbial source tracking) can identify sites with greater contamination potential and sources of fecal contamination. This is essential for a thorough watershed management plan to address impairment and TMDL source reductions.

Although beyond the scope of SSN modeling, temporal variations also affect the outcome when modeling microbial water quality. Holcomb and Messier [53] included a time component in their geostatistical models and obtained improved prediction errors than models considering spatial effects only. The effects of precipitation patterns could also be investigated when the temporal effects are incorporated into spatial modeling to provide a comprehensive understanding of microbial water quality dynamics in response to a variety of environmental variables and weather scenarios. Overall, the current study provides critical insights into assessing the amount of fecal contamination to guide further monitoring and management activities and serves as potential water quality modeling framework for urban and suburban watersheds in the USA and beyond.

5. Conclusions

In this study, we successfully applied spatial stream networks (SSN) to model elevated concentrations of *E. coli* in the suburban mixed-land-use Musconetcong River watershed in response to upstream land use attributes, including urban, pasture, forest, and wetland. The SSN model is essentially a spatial statistical model designed for modeling the statistical relationships between variables on stream networks [23]. The modeling results suggest that upstream urban land was positively and significantly associated with $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events and during storm events, respectively. Upstream pasture land was positively and significantly correlated with $\log_{10}$ geometric mean concentrations of *E. coli* during storm events only. Although only marginally significantly, the $\log_{10}$ geometric mean concentrations of *E. coli* for all sampling events demonstrated a negative relationship with upstream wetland land use as per the SSN model calibration. Applying SSN modeling based on spatial distance also demonstrated improved model performance over non-spatial models. The SSN modeling results concur with previous findings that anthropogenic sources represent the main threat to microbial water quality in

the Musconetcong River watershed. The prediction of *E. coli* concentrations by SSN models identified potential hot spots prone to water quality deteriorations. Future directions could benefit from incorporating additional sampling locations beyond the current section of the watershed and introducing temporal effects in the spatial modeling. Nevertheless, with the support of publicly available watershed attribute data, a pre-established ArcGIS toolset and the open-source R package, we believe that SSN models can be employed by other urban and suburban water quality stakeholders to assist in their water quality monitoring and restoration needs when the resources are constrained.

Author Contributions: T.-T.D.H.: Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Data Curation, Writing—Original Draft, Visualization; D.Y.: Conceptualization, Methodology, Resources, Writing—Reviewing and Editing, Supervision; M.W.: Resources, Writing—Reviewing and Editing, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: Delaware Watershed Research Fund (DWRF-17-03) for microbial data collection.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Lee Lee, Alessandra Rossi, Kevin Zerbe, Molly Hillenbrand, and Edward Wong of the New Jersey Center for Water Science and Technology as well as Nancy Lawler and Christa Reeves of the Musconetcong Watershed Association for their assistance in microbial data collection. The authors would also like to thank the Delaware Watershed Research Fund for their financial support for microbial data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arnone, R.D.; Walling, J.P. Waterborne pathogens in urban watersheds. *J. Water Health* **2007**, *5*, 149–162. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Schneider, I.E. Urban Water Recreation: Experiences, Place Meanings, and Future Issues. In *The Water Environment of Cities*; Springer: Boston, MA, USA, 2009; pp. 125–140.
3. USEPA. *Health Effects Criteria for Fresh Recreational Waters*; USEPA: Washington, DC, USA, 1984.
4. USEPA. *Recreational Water Quality Criteria*; USEPA: Washington, DC, USA, 2012.
5. Prüss, A. Review of epidemiological studies on health effects from exposure to recreational water. *Int. J. Epidemiol.* **1998**, *27*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Wymer, L.J.; Dufour, A.P. A model for estimating the incidence of swimming-related gastrointestinal illness as a function of water quality indicators. *Environmetrics* **2002**, *13*, 669–678. [\[CrossRef\]](#)
7. Almeida, C.; González, S.O.; Mallea, M.; González, P. A recreational water quality index using chemical, physical and microbiological parameters. *Environ. Sci. Pollut. Res.* **2012**, *19*, 3400–3411. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Wymer, L.J.; Wade, T.J.; Dufour, A.P. Equivalency of risk for a modified health endpoint: A case from recreational water epidemiology studies. *BMC Public Health* **2013**, *13*, 459. [\[CrossRef\]](#)
9. Keller, A.A.; Cavallaro, L. Assessing the US Clean Water Act 303(d) listing process for determining impairment of a waterbody. *J. Environ. Manag.* **2008**, *86*, 699–711. [\[CrossRef\]](#)
10. Benham, B.L.; Baffaut, C.; Zeckoski, R.W.; Mankin, K.R.; Pachepsky, Y.A.; Sadeghi, A.M.; Brannan, K.M.; Soupir, M.L.; Habersack, M.J. Modeling bacteria fate and transport in watersheds to support TMDLs. *Trans. ASABE* **2006**, *49*, 987–1002. [\[CrossRef\]](#)
11. National Research Council. *Assessing the TMDL Approach to Water Quality Management*; National Research Council: Washington, DC, USA, 2001.
12. Cho, K.H.; Pachepsky, Y.A.; Oliver, D.M.; Muirhead, R.W.; Park, Y.; Quilliam, R.S.; Shelton, D.R. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. *Water Res.* **2016**, *100*, 38–56. [\[CrossRef\]](#)
13. Jha, M.K.; Wolter, C.F.; Schilling, K.E.; Gassman, P.W. Assessment of Total Maximum Daily Load Implementation Strategies for Nitrate Impairment of the Raccoon River, Iowa. *J. Environ. Qual.* **2010**, *39*, 1317–1327. [\[CrossRef\]](#)
14. Mas, D.M.L.; Ahlfeld, D.P. Comparing artificial neural networks and regression models for predicting faecal coliform concentrations. *Hydrol. Sci. J.* **2007**, *52*, 713–731. [\[CrossRef\]](#)
15. Cha, Y.; Park, M.-H.; Lee, S.-H.; Kim, J.H.; Cho, K.H. Modeling spatiotemporal bacterial variability with meteorological and watershed land-use characteristics. *Water Res.* **2016**, *100*, 306–315. [\[CrossRef\]](#)

16. Herrig, I.; Seis, W.; Fischer, H.; Regnery, J.; Manz, W.; Reifferscheid, G.; Böer, S. Prediction of fecal indicator organism concentrations in rivers: The shifting role of environmental factors under varying flow conditions. *Environ. Sci. Eur.* **2019**, *31*, 59. [\[CrossRef\]](#)
17. Rossi, A.; Wolde, B.T.; Lee, L.H.; Wu, M. Prediction of recreational water safety using *Escherichia coli* as an indicator: Case study of the Passaic and Pompton rivers, New Jersey. *Sci. Total. Environ.* **2020**, *714*, 136814. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Herrig, I.M.; Böer, S.I.; Brennholt, N.; Manz, W. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. *Water Res.* **2015**, *85*, 148–157. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Acevedo, M.F. *Data Analysis and Statistics for Geography, Environmental Science, and Engineering*; CRC Press: Boca Raton, FL, USA, 2012.
20. Tobler, W.R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **1970**, *46*, 234–240. [\[CrossRef\]](#)
21. Peterson, E.E.; Hoef, J.V.; Isaak, D.J.; Falke, J.A.; Fortin, M.-J.; Jordan, C.E.; McNyset, K.; Monestiez, P.; Ruesch, A.S.; Sengupta, A.; et al. Modelling dendritic ecological networks in space: An integrated network perspective. *Ecol. Lett.* **2013**, *16*, 707–719. [\[CrossRef\]](#)
22. Isaak, D.J.; Peterson, E.E.; Hoef, J.M.V.; Wenger, S.J.; Falke, J.A.; Torgersen, C.E.; Sowder, C.; Steel, E.A.; Fortin, M.-J.; Jordan, C.E.; et al. Applications of spatial statistical network models to stream data. *WIREs Water* **2014**, *1*, 277–294. [\[Cross-Ref\]](#)
23. Money, E.S.; Carter, G.P.; Serre, M.L. Modern space/time geostatistics using river distances: Data integration of turbidity and *E. coli* measurements to assess fecal contamination along the Raritan River in New Jersey. *Environ. Sci. Technol.* **2009**, *43*, 3736–3742. [\[CrossRef\]](#) [\[PubMed\]](#)
24. USEPA. *Using Microbial Source Tracking to Support TMDL Development and Implementation*; USEPA: Washington, DC, USA, 2011.
25. Isaak, D.J.; Hoef, J.M.V.; Peterson, E.E.; Horan, D.L.; Nagel, D.E. Scalable population estimates using spatial-stream-network (SSN) models, fish density surveys, and national geospatial database frameworks for streams. *Can. J. Fish. Aquat. Sci.* **2017**, *74*, 147–156. [\[CrossRef\]](#)
26. Scown, M.W.; McManus, M.G.; Carson, J.H., Jr.; Nietch, C.T. Improving Predictive Models of In-Stream Phosphorus Concentration Based on Nationally-Available Spatial Data Coverages. *JAWRA J. Am. Water Resour. Assoc.* **2017**, *53*, 944–960. [\[CrossRef\]](#)
27. ElSaadani, M.; Krajewski, W.F.; Zimmerman, D.L. River network based characterization of errors in remotely sensed rainfall products in hydrological applications. *Remote Sens. Lett.* **2018**, *9*, 743–752. [\[CrossRef\]](#)
28. Gendaszek, A.S.; Dunham, J.B.; Torgersen, C.E.; Hockman-Wert, D.P.; Heck, M.P.; Thorson, J.; Mintz, J.; Allai, T. Land-Cover and Climatic Controls on Water Temperature, Flow Permanence, and Fragmentation of Great Basin Stream Networks. *Water* **2020**, *12*, 1962. [\[CrossRef\]](#)
29. Hoef, J.M.V.; Peterson, E.; Theobald, D. Spatial statistical models that use flow and stream distance. *Environ. Ecol. Stat.* **2006**, *13*, 449–464. [\[CrossRef\]](#)
30. Hoef, J.M.V.; Peterson, E.E. A Moving Average Approach for Spatial Statistical Models of Stream Networks. *J. Am. Stat. Assoc.* **2010**, *105*, 6–18. [\[CrossRef\]](#)
31. Barry, R.P.; Jay, M.V.H. Blackbox Kriging: Spatial Prediction without Specifying Variogram Models. *J. Agric. Biol. Environ. Stat.* **1996**, *1*, 297–322. [\[CrossRef\]](#)
32. Neill, A.J.; Tetzlaff, D.; Strachan, N.J.C.; Hough, R.L.; Avery, L.M.; Watson, H.; Soulsby, C. Using spatial-stream-network models and long-term data to understand and predict dynamics of faecal contamination in a mixed land-use catchment. *Sci. Total Environ.* **2018**, *612*, 840–852. [\[CrossRef\]](#)
33. Hunt, A.; Wu, M.; Hsu, T.-T.; Roberts-Lawler, N.; Miller, J.; Rossi, A.; Lee, L. Picking Up Where the TMDL Leaves Off: Using the Partnership Wild and Scenic River Framework for Collaborative River Restoration. *Sustainability* **2021**, *13*, 1878. [\[CrossRef\]](#)
34. Stinnette, I.; Taylor, M.; Kerr, L.; Pirani, R.; Lipuma, S.; Lodge, J. *State of the Estuary 2018*; Hudson River Foundation: New York, NY, USA, 2018.
35. New Jersey Department of Environmental Protection. *Total Maximum Daily Loads for Fecal Coliform to Address 28 Streams in the Northwest Water Region*; New Jersey Department of Environmental Protection: Camden, NJ, USA, 2003.
36. Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, R.E. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; US Government Printing Office: Washington, DC, USA, 1976.
37. Restoration, R.W. Protection Plan for the Musconetcong River Watershed from Hampton to Bloomsbury. In *Water Quality Monitoring Data Report*; Rutgers Cooperative Extension Water Resources Program: New Brunswick, NJ, USA, 2011.
38. Garreta, V.; Monestiez, P.; Hoef, J.M.V. Spatial modelling and prediction on river networks: Up model, down model or hybrid? *Environmetrics* **2010**, *21*, 439–456. [\[CrossRef\]](#)
39. Ver Hoef, J.M.; Peterson, E.; Clifford, D.; Shah, R. SSN: An R Package for Spatial Statistical Modeling on Stream Networks. *J. Stat. Softw.* **2014**, *56*, 1–45.
40. Dubin, R.A. Spatial autocorrelation and neighborhood quality. *Reg. Sci. Urban Econ.* **1992**, *22*, 433–452. [\[CrossRef\]](#)
41. Anselin, L.; Bera, A.K.; Florax, R.; Yoon, M.J. Simple diagnostic tests for spatial dependence. *Reg. Sci. Urban Econ.* **1996**, *26*, 77–104. [\[CrossRef\]](#)
42. Paule-Mercado, M.; Ventura, J.; Memon, S.; Jahng, D.; Kang, J.-H.; Lee, C.-H. Monitoring and predicting the fecal indicator bacteria concentrations from agricultural, mixed land use and urban stormwater runoff. *Sci. Total Environ.* **2016**, *550*, 1171–1181. [\[CrossRef\]](#) [\[PubMed\]](#)

43. Peterson, E.E. *STARS: Spatial Tools for the Analysis of River Systems version 2.0. 6—A tutorial*; Queensland University of Technology: Brisbane, Australia, 2017.
44. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
45. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [\[CrossRef\]](#)
46. Peterson, E.E.; Merton, A.A.; Theobald, D.M.; Urquhart, N.S. Patterns of Spatial Autocorrelation in Stream Water Chemistry. *Environ. Monit. Assess.* **2006**, *121*, 571–596. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Cressie, N.; Lahiri, S. The Asymptotic Distribution of REML Estimators. *J. Multivar. Anal.* **1993**, *45*, 217–233. [\[CrossRef\]](#)
48. Beleites, C.; Baumgartner, R.; Bowman, C.; Somorjai, R.; Steiner, G.; Salzer, R.; Sowa, M.G. Variance reduction in estimating classification error using sparse datasets. *Chemom. Intell. Lab. Syst.* **2005**, *79*, 91–100. [\[CrossRef\]](#)
49. Detenbeck, N.E.; Morrison, A.C.; Abele, R.W.; Kopp, D.A. Spatial statistical network models for stream and river temperature in New England, USA. *Water Resour. Res.* **2016**, *52*, 6018–6040. [\[CrossRef\]](#)
50. Post, C.J.; Cope, M.P.; Gerard, P.D.; Mastro, N.M.; Vine, J.R.; Stiglitz, R.Y.; Hallstrom, J.O.; Newman, J.C.; Mikhailova, E.A. Monitoring spatial and temporal variation of dissolved oxygen and water temperature in the Savannah River using a sensor network. *Environ. Monit. Assess.* **2018**, *190*, 274. [\[CrossRef\]](#)
51. Frieden, J.C.; Peterson, E.E.; Webb, J.A.; Negus, P.M. Improving the predictive power of spatial statistical models of stream macroinvertebrates using weighted autocovariance functions. *Environ. Model. Softw.* **2014**, *60*, 320–330. [\[CrossRef\]](#)
52. Larsen, S.; Bruno, M.C.; Vaughan, I.P.; Zolezzi, G. Testing the River Continuum Concept with geostatistical stream-network models. *Ecol. Complex.* **2019**, *39*, 100773. [\[CrossRef\]](#)
53. Holcomb, D.A.; Messier, K.P.; Serre, M.L.; Rowny, J.G.; Stewart, J.R. Geostatistical Prediction of Microbial Water Quality Throughout a Stream Network Using Meteorology, Land Cover, and Spatiotemporal Autocorrelation. *Environ. Sci. Technol.* **2018**, *52*, 7775–7784. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Franklin, R.B.; Mills, A.L. *The Spatial Distribution of Microbes in the Environment*; Springer: Dordrecht, The Netherlands, 2007.
55. Brokamp, C.; Beck, A.F.; Muglia, L.; Ryan, P. Combined sewer overflow events and childhood emergency department visits: A case-crossover study. *Sci. Total Environ.* **2017**, *607*–*608*, 1180–1187. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Borchardt, M.A.; Chyou, P.-H.; DeVries, E.O.; Belongia, E.A. Septic system density and infectious diarrhea in a defined population of children. *Environ. Health Perspect.* **2003**, *111*, 742–748. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Crowther, J.; Kay, D.; Wyer, M.D. Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: Relationships with land use and farming practices. *Water Res.* **2002**, *36*, 1725–1734. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Burkholder, J.; Libra, B.; Weyer, P.; Heathcote, S.; Kolpin, D.; Thorne, P.S.; Wichman, M. Impacts of Waste from Concentrated Animal Feeding Operations on Water Quality. *Environ. Health Perspect.* **2007**, *115*, 308–312. [\[CrossRef\]](#)
59. Harris, C.S.; Tertuliano, M.; Rajeev, S.; Vellidis, G.; Levy, K. Impact of storm runoff on Salmonella and Escherichia coli prevalence in irrigation ponds of fresh produce farms in southern Georgia. *J. Appl. Microbiol.* **2018**, *124*, 910–921. [\[CrossRef\]](#)
60. Mitsch, W.J.; Gosselink, J.G. *Wetlands*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
61. Molleda, P.; Blanco, I.; Ansola, G.; de Luis, E. Removal of wastewater pathogen indicators in a constructed wetland in Leon, Spain. *Ecol. Eng.* **2008**, *33*, 252–257. [\[CrossRef\]](#)
62. Hathaway, J.M.; Hunt, W.F.; Jadlocki, S. Indicator Bacteria Removal in Storm-Water Best Management Practices in Charlotte, North Carolina. *J. Environ. Eng.* **2009**, *135*, 1275–1285. [\[CrossRef\]](#)
63. Hsu, T.-T.D.; Mitsch, W.J.; Martin, J.F.; Lee, J. Towards sustainable protection of public health: The role of an urban wetland as a frontline safeguard of pathogen and antibiotic resistance spread. *Ecol. Eng.* **2017**, *108*, 547–555. [\[CrossRef\]](#)
64. Biswal, B.K.; Balasubramanian, R. Constructed Wetlands for Reclamation and Reuse of Wastewater and Urban Stormwater: A Review. *Front. Environ. Sci.* **2022**, *10*, 201. [\[CrossRef\]](#)
65. Curriero, F.C.; Patz, J.A.; Rose, J.B.; Lele, S. The Association Between Extreme Precipitation and Waterborne Disease Outbreaks in the United States, 1948–1994. *Am. J. Public Health* **2001**, *91*, 1194–1199. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Thomas, K.M.; Charron, D.F.; Waltner-Toews, D.; Schuster, C.; Maarouf, A.R.; Holt, J.D. A role of high impact weather events in waterborne disease outbreaks in Canada, 1975–2001. *Int. J. Environ. Health Res.* **2006**, *16*, 167–180. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Mac Kenzie, W.R.; Hoxie, N.J.; Proctor, M.E.; Gradus, M.S.; Blair, K.A.; Peterson, D.E.; Kazmierczak, J.J.; Addiss, D.G.; Fox, K.R.; Rose, J.B.; et al. A Massive Outbreak in Milwaukee of Cryptosporidium Infection Transmitted through the Public Water Supply. *N. Engl. J. Med.* **1994**, *331*, 161–167. [\[CrossRef\]](#) [\[PubMed\]](#)
68. De Roos, A.J.; Kondo, M.C.; Robinson, L.F.; Rai, A.; Ryan, M.; Haas, C.N.; Lojo, J.; Fagliano, J.A. Heavy precipitation, drinking water source, and acute gastrointestinal illness in Philadelphia, 2015–2017. *PLoS ONE* **2020**, *15*, e0229258. [\[CrossRef\]](#)
69. Dupigny-Giroux, L.A.; Mecray, E.L.; Lemcke-Stampone, M.D.; Hodgkins, G.A.; Lentz, E.E.; Mills, K.E.; Lane, E.D.; Miller, R.; Hollinger, D.Y.; Solecki, W.D.; et al. Northeast. In *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment*; U.S. Global Change Research Program: Washington, DC, USA, 2018; Volume II.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.