Environmental Science & Technology LETTERS

pubs.acs.org/journal/estlcu

Letter

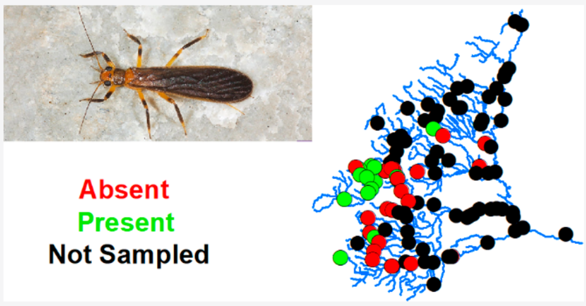
Citizen Science Data Show Temperature-Driven Declines in Riverine Sentinel Invertebrates

Timothy J. Maguire* and Scott O. C. Mundle

Cite This: https://dx.doi.org/10.1021/acs.estlett.0c00206

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: We used the presence and absence data of sentinel invertebrates (stonefly, order Plecoptera) collected by citizen scientists over 17 years to approximate trends in stream health in urban Detroit, Michigan, USA. Citizen science data are commonly collected based on availability of limited funds. Thus, survey locations lack consistent data collection, and missing values are common. While citizen science data sets can be large, on a regional and local scale, they are often undervalued but present an opportunity for managers to inform their decisions if the missing data can be addressed. To overcome this hurdle, here, missing values were modeled with a combination of spatial (inverse distance weighting and spatial stream network), temporal (Bayesian state space), and machine learning (ensemble random forest) models combining atmospheric, hydrologic, and biologic data. Using the estimated missing values, we determined negative population trends in stoneflies driven by stream temperature via a dynamic occupancy model. Urban streams present a challenge to resource managers because data are collected at disparate locations and frequencies and inconsistently recorded. However, we show how a combination of methods with publicly available and citizen science data from across disciplines can inform managers and support land-use decisions.

■ INTRODUCTION

Anthropogenic impact on water quality is a global issue wherever rivers and cities coexist. The effect of development on rivers is ubiquitous to the extent that the phrase "urban stream syndrome" has come to broadly be applied to this resource management issue.[1] Restoration of urban streams is influenced by integrated social, political, and environmental factors.[2] Regulators must identify reference conditions and applicable techniques for restoration while maintaining human requirements in a data-driven process.[3] Biomonitoring of invertebrates is a common data collection response to this challenge but is complicated by a lack of training, unavailable funds, and inconsistent data collection.[4]

Biomonitoring of benthic macroinvertebrates is an established assessment tool as macroinvertebrates are indicators of ecological health.[5] Stoneflies (order Plecoptera) are among those invertebrates which act as indicators.[6] These organisms exist along the stream bed in their larval stage where they are sensitive to low oxygen conditions and high temperatures; stoneflies emerge from the water in winter and become mobile active adults in order to mate.[7,8] Urban stonefly species are also potentially impacted by pesticides, toxic metals, and other deleterious contaminates in urban runoff.[9] Urban land use is generally associated with reduced stonefly populations;[10] thus, resource managers often use stoneflies as a restoration gauge. However, habitat restoration is not always successful in returning lost stonefly populations.[11] One advantage in using stoneflies is that citizen science is a reliable source of presence or absence data.[12,13]

While some citizen science data collection projects are functional collaborations that develop large-scale spatial—temporal models shared between scientists, managers, and the public,[14] many community-based data collection efforts are used as "snapshot" assessments of river segments. Annually accrued citizen science data sets may be publicly available and aggregated into large regulator data sets,[15] but they also exist as orphaned data sets maintained by volunteers and inefficiently used by watershed regulators. In the state of Michigan, USA, there are 27 citizen-based watershed associations promoting riverine ecosystem health and collecting data, but how much of that data is answering the questions they were collected to answer?

Here, we use citizen science data to address the resource management questions underpinning the community action which prompted their collection: is the population of sentinel invertebrates in an urban stream expanding or contracting and why? We hypothesize that populations are changing because of land use impacting water quality and increases in summer stream

Received: March 11, 2020
Revised: April 3, 2020
Accepted: April 3, 2020
Published: April 3, 2020

temperature. With urbanization expanding rapidly throughout the Anthropocene, watersheds will be under increased stress from runoff laden contaminates, changing temperature regimes as a result of a warming climate, and daylighting of streams as overhanging trees are removed. Urban waterways are an important resource to be managed, and here, we show how to make the best use of incomplete data.

MATERIALS AND METHODS

Data. The citizen science-generated data we used to determine sentinel invertebrate population changes were collected on the River Rouge in urban Detroit, Michigan, USA (Figure 1). The river name sporadically appears as Rouge River

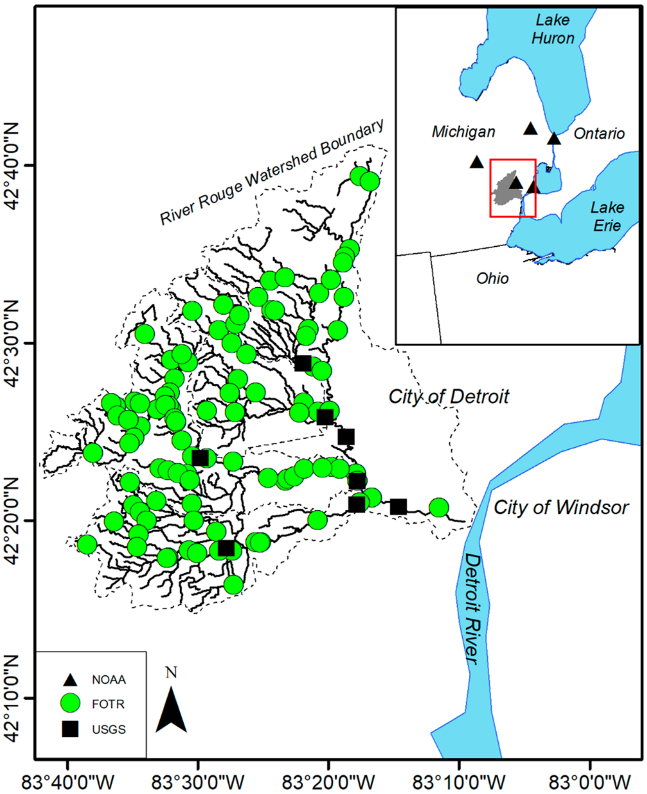


Figure 1. River Rouge watershed is heavily urbanized within the city of Detroit. National Oceanographic and Atmospheric Administration (NOAA) weather stations, United States Geologic Survey (USGS) stream temperature sites, and Friends of the Rouge (FOTR) winter stonefly (order Plecoptera) citizen science survey data were combined to estimate stonefly occupancy, site extinction rate, site colonization rate, and observer efficacy within the River Rouge.

in available literature; here, we use the United States Geologic Survey (USGS) nomenclature "River Rouge". The River Rouge watershed is 92% urban, subject to intensive industrial use, and impacted by soil and groundwater contamination.[16,17] Friends of the Rouge (FOTR) was established in 1986 to promote and clean the River Rouge through outreach and citizen science participation. FOTR winter stonefly surveys have been conducted annually in January from 2002 to 2018 using procedures detailed in the Michigan Clean Water Corps guidance documents (micorps.net/stream-monitoring/stream-documents/). Each site was surveyed for a minimum of 30 min by sampling all available habitats along a 100-m of riverbank with D-frame nets. FOTR recorded the presence or absence of stoneflies at 103 invertebrate survey sites; however, the data collection was sporadic. Every site was not sampled every year. Thus, of the 1751 potential data points, the FOTR data set had 447 recorded values. The missing values were modeled in order to assess the stonefly population changes that have occurred over the 17 years of the data set (Table 1, Supporting Information, and Movie, Supporting Information).

Missing FOTR values were modeled using summer stream temperature and land-use data. Annual summer stream temperatures were not available at the stonefly survey sites; models were used to estimate these values based on National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (www.ncei.noaa.gov) weather data and USGS National Water Information System (waterdata.usgs.gov/nwis) stream data. Data sets across the 2002−2018 range were available at five NOAA weather stations (daily minimum air temperature) and 10 USGS stream stations (mean daily temperature) (Figure 1). Land use was defined by the U.S. Environmental Protection Agency (EPA) National Aquatic Resource Survey StreamCat database (www.epa.gov/national-aquatic-resource-surveys/streamcat).[18] Available within StreamCat, the benthic invertebrate multimetric index (BMMI) defined as a 0 to 1 scale of biological conditions for benthic invertebrate assemblages was used in the missing FOTR stonefly model to define land use. However, 25 of the 103 stream segments of the River Rouge with FOTR sites were not defined (Figure 1); thus, BMMI had to be predicted for those sections. Groundwater baseflow estimates per stream segment within the watershed were retrieved from the Michigan GIS Open Data website (https://gis-michigan.opendata.arcgis.com/).

Models. We used a dynamic occupancy model (DOM) to determine the occurrence, extinction, colonization, and detection probabilities for winter stoneflies in the River Rouge watershed. In order to use the DOM, we first used a combination of initial models in order to define land use, determine summer temperature at the stonefly FOTR sites, and estimate the missing FOTR values (Figure 2).

While stonefly species are sensitive to poor water quality, water quality data for the stonefly sites were not available. BMMI was used as a proxy for water quality as it is defined by land use at the stonefly FOTR sites. To estimate the 25 stonefly FOTR sites which were excluded from the StreamCat database, we used a machine learning approach. BMMI values available through StreamCat were estimated using a random forest model which included 198 land use variables which are also available in the StreamCat database.[19] We recreated the BMMI analysis and filled in the missing BMMI River Rouge values by creating an ensemble of 50 models predicting BMMI from the 154 land use variables available for the River Rouge watershed. The 50 models were randomly constructed as either a random forest or gradient boosting model using the h2o package in R.[20−22] Each of the 50 models was randomly given 50 land use variables of the available 154 to avoid overfitting. Here, 80% of the 78 stonefly FOTR sites with BMMI were used to train the models and 10% to validate the models, while 10% were withheld to test the model predictions. Model outputs were incorporated in an ensemble random forest model, and BMMI was predicted for the 25 omitted sites.

Summer stream temperature was modeled by a combination of spatial and temporal analysis. USGS stream temperature sites were used to predict summer stream temperature at the stonefly FOTR sites; however, the USGS sites were not active all year nor were they active during all years of the FOTR survey range. To model stream temperature on the unobserved days at each

https://dx.doi.org/10.1021/acs.estlett.0c00206
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

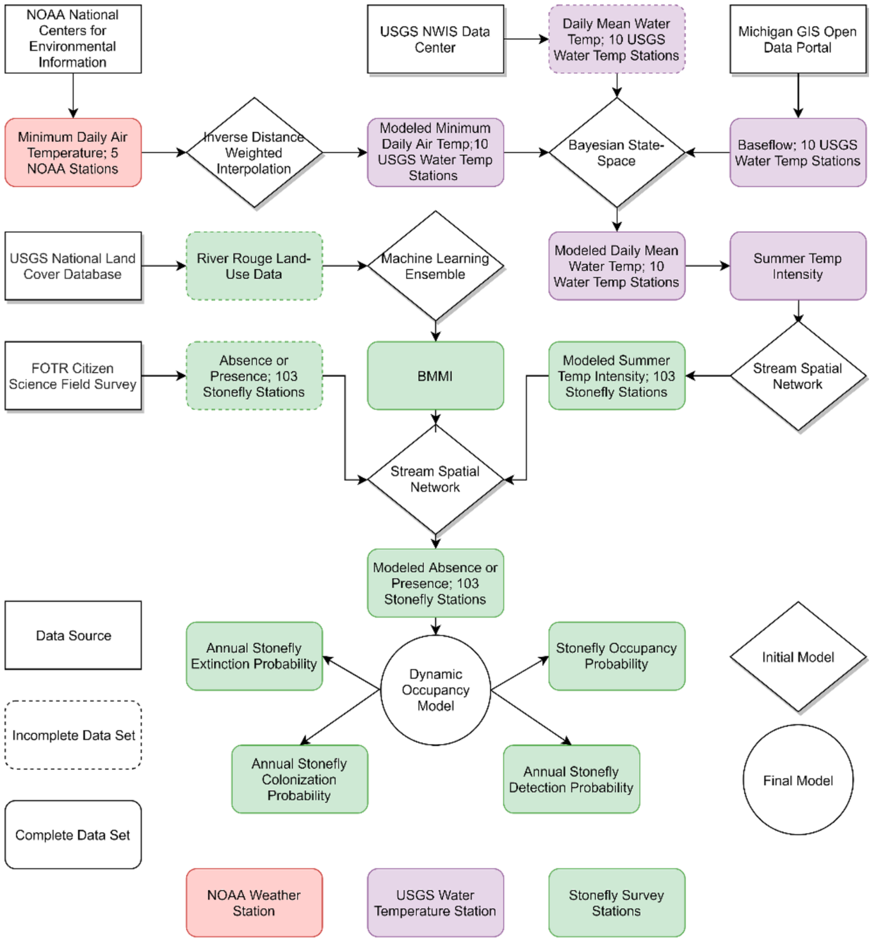


Figure 2. River Rouge stonefly population occupancy was modeled with a dynamic occupancy model (DOM); however, prior to the DOM several data sets had missing values which were estimated. Data on stonefly presence and absence were collected by the Friends of the Rouge (FOTR) citizen science organization. Weather data were downloaded from the National Oceanographic and Atmospheric Administration (NOAA). Stream data were downloaded from the United States Geologic Survey (USGS). Five initial models were used to establish a complete data set of stoneflies for the DOM.

USGS site, a model combining atmospheric temperature and groundwater baseflow was developed. Atmospheric temperature was available throughout the temporal range of the FOTR surveys; however, the NOAA were not co-located with USGS sites. To predict the atmospheric temperature at the USGS sites using the NOAA data, inverse distance weighting (IDW) spatial analysis was used.[23] Baseflow estimates associated with the stream segments containing USGS sites and daily minimum air temperature modeled via IDW were used predictor variables in a Bayesian state-space model of available USGS daily mean stream temperatures in a time series context. Sites had an independent coefficient relating stream temperature to air temperature. A single coefficient for the effect of baseflow was used for all sites. Uninformative priors were used. Gelman Rubin convergence was used to assess if an adequate number of burn-in iterations had been reached, and autocorrelation plots were used to establish an adequate thin.[24] The area between daily stream temperature estimates and 18 °C was calculated from the combined modeled mean daily stream temperature and observations to establish an annual estimate of heat in summer months at the USGS sites.[25] The over 18 °C area allowed us to establish a representative value for the summer stream temperature at the USGS sites. The over 18 °C area was then used to estimate the summer stream temperature at the FOTR sites with a spatial stream network (SNN).[26] The SSN was fit using a Gaussian distribution, elevation as predictor variable, and year as a random effect, and from this SSN, the annual summer stream temperature was estimated at the FOTR sites (Figure 2).

Predicted annual summer stream temperature and BMMI were then used as predictor variables in a final SSN fit to the stonefly survey data. The final SSN was fit with a binomial distribution and treated each site as a random effect. The predicted presence and absence of stoneflies at unobserved sites were extracted from the final SSN and used to complete the stonefly data set in the DOM (Figure 2).

■ RESULTS AND DISCUSSION

The stonefly DOM used to predict occupancy changes in the River Rouge was built on observations and predictions from a SSN; the SSN was generated with land use as a proxy for water quality (BMMI) and summer stream temperature (Figure 2). The BMMI retrieved from StreamCat ranged from 0.164 and 0.5787; the machine learning ensemble of predicting BMMI at the 25 unobserved stream segments had a root-mean-square error (RMSE) of 0.019 and a R^2 of 0.922. Daily mean stream temperature was estimated at USGS stations from IDW-predicted minimum air temperature and baseflow in a state-space model. In the Bayesian state-space model of daily mean stream temperature, the site-independent coefficients relating

https://dx.doi.org/10.1021/acs.estlett.0c00206
Environ. Sci. Technol. Lett. XXXX, XXX, XXX−XXX

each site stream temperature to air temperature ranged from 0.151 to 0.211, and the baseflow coefficient was −0.0015. All were significantly different from zero. Daily stream temperature represented summer temperature by calculating the area represented at each site above 18 °C that had a mean value of 313 and standard error (SE) of 10. Summer stream temperature area above 18 °C was predicted at the FOTR sites based on a SSN of USGS stations temperature predicted by year as a random effect and elevation as a fixed effect. Leave-one-out cross-validation (LOOCV) estimates of R^2 and root-mean-square prediction error were 0.93 and 35.3, respectively.[27] Elevation was a significant predictor of summer temperature ($p < 0.001$).

The SSN to predict missing stonefly absence or presence was built with BMMI and summer temperature as predictor variables and sites as random effects. BMMI was not a significant predictor ($p = 0.75$), while temperature was significant ($p = 0.02$), and LOOCV correctly predicted 82% and 76% of unoccupied and occupied sites, respectively. The SSN used to model the missing data is a flexible tool for resource managers investigating citizen science data within streams. In lieu of absence and presence data, if species abundance or diversity data were of interest, the SSN model could be fit using Poisson or Gaussian distributions. The predicted values for those missing data could then be used in further ecological models, as here the predicted missing FOTR values were used in a DOM.

The DOM predicted that 21.2% (4.3%, SE) of sites were occupied by stoneflies in the FOTR survey area. The probability an unoccupied site is colonized was 1.6% (0.4% SE), and the probability of a site occupied by stoneflies going extinct was 6.4% (1.7% SE). The probability that stoneflies are observed at an occupied site is 81% (3.1% SE).

Missing data are long-standing hurdles in watershed-scale modeling efforts and citizen science data collection. In the FOTR-collected data, 103 sites were represented by 447 recorded values and 1304 missing values. The data used to model the missing FOTR values were rife with missing data. BMMI was missing from 25 of the 103 stream segments with FOTR sites. Daily summer temperature was not available at the FOTR sites, and the USGS temperature data that were available contained missing data. The air temperature data used to model the missing stream temperature data were missing data. Moreover, the FOTR, USGS, and NOAA sites were all in different locations. However, research questions can be addressed using novel combinations of spatial (IDW and spatial stream networks) and statistical (machine learning and Bayesian state space) models to estimate missing values (Figure 2). With limited funds available to address urban stream syndrome, adapting and generating results from incomplete citizen science data can establish which restoration effort and land-use policy should be adopted.

Observer error is a fundamental concern with citizen science data. However, our estimates indicate that an occupied site was recorded as present 81% of the time. The high efficacy of sampling is due to the citizen science training conducted by FOTR and the ecology of the winter stonefly. FOTR uses a Quality Assurance Project Plan where a tiered system of highly trained team leaders oversees each group of citizen scientists. The stonefly dark color against light-colored sampling trays makes stoneflies ideal target species to find. Additionally, accounting for the missing data enabled an estimate of occupancy across the study system (21%). While early emergence of adult stoneflies due to specific annual site conditions could result in individuals being missed by the FOTR survey, the impact of this over the 17 years of the data set would be minimal, and the occupancy metric should be used by urban resource managers as a benchmark of sampling efficacy and future restoration population increases. Managers establishing projects to assess sentinel invertebrates via citizen science projects should consider site selection, the number of samples, and the number of years required as used by the methodology presented (Figure 2). Sites should be added until representative of holistic conditions of the BMMI and temperature throughout the watershed. The number of samples over time can then be increased and our method used to informatively gap fill both spatially and temporally missing data. The number of sampling events required to adequately estimate colonization and extinction rates should be determined by the size of the parameter confidence intervals which will represent the uncertainty in observations and incorporate variability from early emergence.

A holistic picture of site colonization and extinction is important to derive management strategies for urban streams. Observations of organisms in river segments where they had not appeared previously could give a false impression of stream health. Our approach defines stoneflies occurring within the River Rouge watershed as a metapopulation, treating each survey site as a local population.[28,29] Migration between local populations happens along stream segments as adults.[30] The metapopulation of macroinvertebrates is dynamic and will expand into sites even as overall more sites are going extinct. Our results show the stonefly metapopulation decreasing (6.4% extinction probability compared to 1.6% colonization) in the River Rouge watershed as a result of changes in temperature rather than adjacent land use. The practical application of this result is that as development continues within the River Rouge watershed it will be advantageous to maintain overhanging trees along the river course to shade and reduce water temperatures. Additionally, these results suggest that the urban stream population of stoneflies can be enhanced by reducing urban heat island effects with reduced imperviousness and use of reflective building materials along the river. Ameliorating urban stream syndrome is a global resource issue, and strategies compatible with continued mixed use are needed to improved social, political, and environmental factors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.estlett.0c00206.

Table 1: annual observed absence and presence data, along with total unsampled sites (PDF)

Movie: sampling results and spatial distribution of sites from 2002 to 2018 (MP4)

■ AUTHOR INFORMATION

Corresponding Author

Timothy J. Maguire − *Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada N9B 3P4;* orcid.org/0000-0002-7742-769X; Email: maguiret@uwindsor.ca

Author

Scott O. C. Mundle − *Great Lakes Institute for Environmental Research and Department of Chemistry and Biochemistry,*

University of Windsor, Windsor, Ontario, Canada N9B 3P4;
orcid.org/0000-0001-5976-9656

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.estlett.0c00206

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Friends of the Rouge for sharing their data and support. Funding was provided by Canadian Natural Sciences and Engineering Research Council's (NSERC) Discovery Grant program (SOCM).

REFERENCES

(1) Booth, D. B.; Roy, A. H.; Smith, B.; Capps, K. A. Global Perspectives on the Urban Stream Syndrome. *Freshw. Sci.* **2016**, *35*, 412−420.
(2) Findlay, S. J.; Taylor, M. P. Why Rehabilitate Urban River Systems? *Area* **2006**, *38*, 312−325.
(3) Nilsson, C.; Jansson, R.; Malmqvist, B.; Naiman, R. J. Restoring Riverine Landscapes: The Challenge of Identifying Priorities, Reference States, and Techniques. *Ecol. Soc.* **2007**, *12*, 16.
(4) Morse, J. C.; Bae, Y. J.; Munkhjargal, G.; Sangpradub, N.; Tanida, K.; Vshivkova, T. S.; Wang, B.; Yang, L.; Yule, C. M. Freshwater Biomonitoring with Macroinvertebrates in East Asia. *Front. Ecol. Environ.* **2007**, *5*, 33−42.
(5) Gaufin, A. R. Use of Aquatic Invertebrates in the Assessment of Water Quality. In *Biological Methods for the Assessment of Water Quality*; ASTM International, 1973.
(6) Baumann, R. W. Nearctic Stonefly Genera as Indicators of Ecological Parameters (Plecoptera: Insecta). *Great Basin Nat.* **1979**, 241−244.
(7) Nebeker, A. V. Effect of Water Temperature on Nymphal Feeding Rate, Emergence, and Adult Longevity of the Stonefly Pteronarcys Dorsata. *J. Kansas Entomol. Soc.* **1971**, 21−26.
(8) Nebeker, A. V. Effect of Low Oxygen Concentration on Survival and Emergence of Aquatic Insects. *Trans. Am. Fish. Soc.* **1972**, *101*, 675−679.
(9) Macneale, K. H.; Kiffney, P. M.; Scholz, N. L. Pesticides, Aquatic Food Webs, and the Conservation of Pacific Salmon. *Front. Ecol. Environ.* **2010**, *8*, 475−482.
(10) Moore, A. A.; Palmer, M. A. Invertebrate Biodiversity in Agricultural and Urban Headwater Streams: Implications for Conservation and Management. *Ecol. Appl.* **2005**, *15*, 1169−1177.
(11) Violin, C. R.; Cada, P.; Sudduth, E. B.; Hassett, B. A.; Penrose, D. L.; Bernhardt, E. S. Effects of Urbanization and Urban Stream Restoration on the Physical and Biological Structure of Stream Ecosystems. *Ecol. Appl.* **2011**, *21*, 1932−1949.
(12) Fore, L. S.; Paulsen, K.; O'Laughlin, K. Assessing the Performance of Volunteers in Monitoring Streams. *Freshwater Biol.* **2001**, *46*, 109−123.
(13) Edwards, P. M. The Value of Long-Term Stream Invertebrate Data Collected by Citizen Scientists. *PLoS One* **2016**, *11*, No. e0153713.
(14) Sullivan, B. L.; Wood, C. L.; Iliff, M. J.; Bonney, R. E.; Fink, D.; Kelling, S. EBird: A Citizen-Based Bird Observation Network in the Biological Sciences. *Biol. Conserv.* **2009**, *142*, 2282−2292.
(15) Poisson, A. C.; McCullough, I. M.; Cheruvelil, K. S.; Elliott, K. C.; Latimore, J. A.; Soranno, P. A. Quantifying the Contribution of Citizen Science to Broad-scale Ecological Databases. *Front. Ecol. Environ.* **2020**, *18*, 19−26.
(16) Murray, K. S.; Rogers, D. T.; Kaufman, M. M. Heavy Metals in an Urban Watershed in Southeastern Michigan. *J. Environ. Qual.* **2004**, *33*, 163−172.
(17) Lenaker, P. L.; Corsi, S. R.; Borchardt, M. A.; Spencer, S. K.; Baldwin, A. K.; Lutz, M. A. Hydrologic, Land Cover, and Seasonal Patterns of Waterborne Pathogens in Great Lakes Tributaries. *Water Res.* **2017**, *113*, 11−21.
(18) Hill, R. A.; Weber, M. H.; Leibowitz, S. G.; Olsen, A. R.; Thornbrugh, D. J. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *J. Am. Water Resour. Assoc.* **2016**, *52*, 120−128.
(19) Hill, R. A.; Fox, E. W.; Leibowitz, S. G.; Olsen, A. R.; Thornbrugh, D. J.; Weber, M. H. Predictive Mapping of the Biotic Condition of Conterminous U.S. Rivers and Streams. *Ecol. Appl.* **2017**, *27*, 2397−2415.
(20) LeDell, E.; Navdeep, G.; Aiello, S.; Fu, A.; Candel, A.; Click, C.; Kraljevic, T.; Nykodym, T.; Aboyoun, P.; Kurka, M.; Malohlava, M., et al. *h2o: R Interface for 'H2O'*; R package, version 6.0-84, 2019. https://cran.r-project.org/web/packages/h2o/index.html (accessed April 2020).
(21) R Core Team. *R: A Language and Environment for Statistical Computing*, version 3.4.2; R Foundation for Statistical Computing: Vienna, Austria, 2017. https://www.R-project.org/ (accessed April 2020).
(22) RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, Inc.: Boston, MA, 2016. https://rstudio.com/ (accessed April 2020).
(23) Pebesma, E. J. Multivariable Geostatistics in S: The Gstat Package. *Comput. Geosci.* **2004**, *30*, 683−691.
(24) Brooks, S. P.; Gelman, A. General Methods for Monitoring Convergence of Iterative Simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434−455.
(25) Baskerville, G. L.; Emin, P. Rapid Estimation of Heat Accumulation from Maximum and Minimum Temperatures. *Ecology* **1969**, *50*, 514−517.
(26) Isaak, D. J.; Peterson, E. E.; Ver Hoef, J. M.; Wenger, S. J.; Falke, J. A.; Torgersen, C. E.; Sowder, C.; Steel, E. A.; Fortin, M.-J.; Jordan, C. E.; Ruesch, A. S.; Som, N.; Monestiez, P. Applications of Spatial Statistical Network Models to Stream Data. *Wiley Interdiscip. Rev.: Water* **2014**, *1*, 277−294.
(27) Isaak, D. J.; Ver Hoef, J. M.; 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.
(28) Levins, R. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bull. Entomol. Soc. Am.* **1969**, *15*, 237−240.
(29) Hanski, I. Single-Species Metapopulation Dynamics: Concepts, Models and Observations. *Biol. J. Linn. Soc.* **1991**, *42*, 17−38.
(30) Macneale, K. H.; Peckarsky, B. L.; Likens, G. E. Stable Isotopes Identify Dispersal Patterns of Stonefly Populations Living along Stream Corridors. *Freshwater Biol.* **2005**, *50*, 1117−1130.