

A geostatistical state-space model of animal densities for stream networks

DANIEL J. HOCKING,^{1,4} JAMES T. THORSON,² KYLE O'NEIL,³ AND BENJAMIN H. LETCHER³

¹*Department of Biology, Frostburg State University, Frostburg, Maryland 21532 USA*

²*Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington 98112 USA*

³*Leetown Science Center, S.O. Conte Anadromous Fish Research Laboratory, U.S. Geological Survey, One Migratory Way, Turners Falls Massachusetts 01376 USA*

Abstract. Population dynamics are often correlated in space and time due to correlations in environmental drivers as well as synchrony induced by individual dispersal. Many statistical analyses of populations ignore potential autocorrelations and assume that survey methods (distance and time between samples) eliminate these correlations, allowing samples to be treated independently. If these assumptions are incorrect, results and therefore inference may be biased and uncertainty underestimated. We developed a novel statistical method to account for spatiotemporal correlations within dendritic stream networks, while accounting for imperfect detection in the surveys. Through simulations, we found this model decreased predictive error relative to standard statistical methods when data were spatially correlated based on stream distance and performed similarly when data were not correlated. We found that increasing the number of years surveyed substantially improved the model accuracy when estimating spatial and temporal correlation coefficients, especially from 10 to 15 yr. Increasing the number of survey sites within the network improved the performance of the nonspatial model but only marginally improved the density estimates in the spatiotemporal model. We applied this model to brook trout data from the West Susquehanna Watershed in Pennsylvania collected over 34 yr from 1981 to 2014. We found the model including temporal and spatiotemporal autocorrelation best described young of the year (YOY) and adult density patterns. YOY densities were positively related to forest cover and negatively related to spring temperatures with low temporal autocorrelation and moderately high spatiotemporal correlation. Adult densities were less strongly affected by climatic conditions and less temporally variable than YOY but with similar spatiotemporal correlation and higher temporal autocorrelation.

Key words: brook trout; dendritic network; detection probability; Gaussian random fields; spatially explicit; spatiotemporal.

INTRODUCTION

Ecologists are concerned with understanding the abundance and distribution of organisms in space and time, as well as the biological processes and interactions that cause these patterns. Surveys are frequently employed to estimate spatiotemporal variation in abundance, with the goal of inferring biological process. However, most statistical methods used in ecology have not explicitly accounted for spatial correlation in the data beyond including covariates that are themselves spatially autocorrelated and random effects related to study design (e.g., ANOVA, GLM, linear and generalized linear mixed models; Hocking et al. 2013, Peterman and Semlitsch 2014, DeWeber and Wagner 2014). Therefore to use these regression methods, researchers must design their studies to ensure that sample points are spaced such that statistical residuals are not correlated. It is difficult to know a priori how close is too close. Any residual autocorrelation violates regression model assumptions and leads to biased results and potentially incorrect inference regarding population distributions and environmental relationships (Dormann et al. 2007). Additionally, information about the

spatial and temporal patterns provides potentially interesting ecological insights that would not be gained if the data were collected in a way to avoid autocorrelation. For these reasons, a large field of spatial statistics has been developed and applied to ecological problems (e.g., Ross et al. 2012, Thorson et al. 2014, Conn et al. 2015).

Streams in a network are likely to have significant correlation in time and space because of regional weather and the hydrologic connections allowing movements and gradients of chemical and physical properties. For example, stream flow and temperature are predictably correlated along the network and it is important to account for this correlation when modeling these systems (Caissie 2006, Ver Hoef et al. 2006, Peterson et al. 2013). Similarly, organisms living in streams are likely to respond to these underlying conditions and their movements are often restricted to the dendritic network creating spatial correlation in the abundance and distribution of stream organisms (Grant et al. 2007, Peterson et al. 2013, Isaak et al. 2014). Spatial models that use Euclidean distance are likely to perform poorly in stream networks because streams in close overland proximity can be completely unconnected or have large hydrologic distances (Ver Hoef et al. 2006). A variety of statistical models have been developed to account for spatial correlations in dendritic networks. These include, but are not limited to, deriving valid covariance relationships for linear models

Manuscript received 2 September 2017; revised 8 May 2018; accepted 14 May 2018. Corresponding Editor: Kristin Mai Broms.

⁴E-mail: djhocking@frostburg.edu

(Peterson et al. 2007) and linear mixed models with moving averages that account for hydrologic distance and flow (Ver Hoef et al. 2006). Some models also include “tail-up,” “tail-down,” or “two-tail” correlations to account for directional autocorrelation (Peterson and Ver Hoef 2010, Ver Hoef and Peterson 2010). Additionally, block Kriging has been used for spatial averaging (Isaak et al. 2017) and splines accounting for network topology and confluence points have been used effectively to model nonlinear trends in stream networks (Donnell et al. 2014).

While these models provide improved inference for many types of data, there are limitations with the current approaches. Current models account for spatial correlations but do not allow for changing spatial correlations over time as with spatiotemporal models (Peterson et al. 2013). A second limitation is the inability to distinguish between process and observation error to account for imperfect detection (Peterson et al. 2013). When performing count surveys of organisms, the probability of detecting each individual in the population is often <1 (imperfect). This results in a problem of inference regarding the populations and environmental effects on the population, particularly when the probability of detection is variable in time and space. To address this issue, a variety of hierarchical models have been developed separating information regarding abundance and detection (e.g., Royle 2004, Royle and Dorazio 2008, Dail and Madsen 2012, Zipkin et al. 2014). However, these models frequently do not account for spatial correlation among sites explicitly (although exceptions exist; Royle and Wikle 2005). Those that do account for spatial autocorrelation often use random group effects assuming clustered sites to be more similar to each other than to other clusters (Hocking et al. 2013, DeWeber and Wagner 2014). This coarse grouping does not allow for autocorrelation as a function of distance. For example, if sampling is done in a series of transects, all sites within a transect are treated the same (Hocking et al. 2013, Peterman and Semlitsch 2013, Milanovich et al. 2015) even though it is likely that adjacent sites are more correlated than distant sites at the opposite ends of the transect. A final limitation of current spatial stream models is the computational challenges with analyzing large networks due to estimating large covariance structures (Peterson et al. 2013).

We describe a novel and generalizable hierarchical model that includes spatiotemporal autocorrelation while accounting for imperfect detection. It also addresses unexplained random variation in abundance not explained by deterministic covariates of abundance (log-normal overdispersion; Harrison 2014). We assessed the spatial component of this model with simulations varying the two parameters of the Ornstein-Uhlenbeck (OU) process used to define the spatial relationships in the network. The OU process is a stochastic process that is similar to a continuous version of a discrete autoregressive (AR1) model with particular properties described in the *Ornstein-Uhlenbeck process for spatial variation* section of the Materials and Methods. This makes it especially well-suited for modeling spatial relationships with distance along a stream network. We also performed a simulation study to evaluate the effects of spatial and temporal replication on model performance. We then applied this model to brook trout (*Salvelinus fontinalis*) data from the

West Susquehanna watershed within Pennsylvania, USA. These data were collected by the Pennsylvania Boat and Fish Commission and are similar to stream fish surveys conducted by state and federal agencies and other researchers throughout the United States. Brook trout were of particular interest as the only native trout in the eastern United States and are threatened by climate and land-use change, overfishing, and exotic species (Hudy et al. 2008).

MATERIALS AND METHODS

Overview

In the following, we assume that data arise from a sampling design where N sites are visited in each of T years (we use vector-matrix notation throughout) and that the same site is never sampled twice in a given year. These N sites are embedded within a stream network where there is only one unique path from each site to every other site (i.e., the stream network is acyclic), and each sample is conducted by eliminating the possibility of movement out of the sampled area (i.e., by placing nets above and below a selected stream segment) and then repeatedly counting and removing all individuals that are observed. We use the term “triple-pass depletion sampling” for this design, given that there are three removal samples conducted in each sampling occasion. Each removal sample has a lower expected count than the previous (because previous sampling has removed individuals), so this triple-pass design allows the detection probability to be estimated from the slope of this decline among passes.

We then modeled intensity $\lambda(s, t)$ at time t and site s (numbers per 100-m stream reach, that is, where distances are measured along a one-dimensional stream reach) from count data following a Poisson distribution as a log-linked generalized linear mixed model with components representing the effect of measured habitat variables, as well as otherwise unexplained spatial, temporal, spatiotemporal, and independent variation. Although some measured independent variables $[\mathbf{x}(s)]$ will correlate with the spatial patterns of animal densities, additional unmeasured factors likely affect the spatial patterns in densities. We include a spatial variation component, $\varepsilon(s)$, in our hierarchical regression model to account for the fact that locations closer together within a network may potentially have more similar densities than more distant locations in ways not predicted by the independent variables. This spatial correlation could result from any number of factors such as density-dependent movement of individuals, underlying geology, physiochemical correlation of the flowing water, or other insufficiently measured spatially correlated network characteristics. Similarly, insufficiently measured factors or complex interactions can result in temporal autocorrelation across space and population dynamics dependent on densities the previous year. We include $\delta(t)$ as a measure of temporal variation. There may also be interactions between space and time that influence the pattern of densities within a network and we include $v(s, t)$ to account for potential spatiotemporal variation (defined as spatial residuals that vary among years). In many ecological systems, there are also many microhabitat variables and other local unexplained variation that result in

variance in excess of predictions from a Poisson distribution when modeling count data (Harrison 2014). We include $\alpha(s, t)$ as overdispersion that is independent among sites and years. The overdispersion parameter is assumed to be normal, independent and identically distributed among sites with variance σ_{iid}^2 [$\alpha(s, t) \sim N(0, \sigma_{iid}^2)$] and is therefore distinct from $v(s, t)$, which is correlated among all sites s in a given year t . We therefore specify

$$\ln(\lambda(s, t)) = \mathbf{x}(s)\boldsymbol{\gamma} + \varepsilon(s) + \delta(t) + v(s, t) + \alpha(s, t), \quad (1)$$

where $\mathbf{x}(s)$ is a row-vector of measured variables affecting density (which includes an intercept term) and $\boldsymbol{\gamma}$ is the estimated impact of these variables on log-density (fixed-effect regression coefficients). Additional descriptions of all model parameters are found in Table 1.

To account for imperfect detection while sampling, we modeled counts $c_d(s, t)$ for depletion pass d ($d \in \{1, 2, 3\}$), site (s), and year (t) assuming that each individual is equally likely to be captured in a given depletion pass. This assumption results in a Poisson distribution for the first pass

$$c_1(s_i, t_i) \sim \text{Poisson}(p_i \times \lambda(s_i, t_i) \times \varphi_i) \quad (2a)$$

where φ_i is the offset for length of stream sampled by observation i (length of survey/100 m) so all densities are the abundance of fish per 100 m of stream length, and p_i is the probability that each individual in the vicinity of observation i will be captured (where this capture probability potentially varies among observations). Counts in the second and third passes are then dependent upon not being captured in the earlier passes

$$c_2(s_i, t_i) \sim \text{Poisson}((1 - p_i) \times p_i \times \lambda(s_i, t_i) \times \varphi_i) \quad (2b)$$

and

$$c_3(s_i, t_i) \sim \text{Poisson}\left(\left(1 - p_i\right)^2 \times p_i \times \lambda(s_i, t_i) \times \varphi_i\right). \quad (2c)$$

In the following, we include variation in detectability among sites and years:

$$p_i = 1 - \exp(-\exp(\mu_p + \eta_i)) \quad (2d)$$

where Eq. 2d represents an inverse complementary-log-log (“cloglog”) link function for detection probability, given parameter μ_p representing average log-detection probability, and independent unexplained variation across sites and years, $\eta_i \sim N(0, \sigma_\eta^2)$, where σ_η^2 is an estimated parameter governing the magnitude of variation in log-detectability among sites and years. We use the cloglog link function so that $\exp(\eta_i)$ is interpreted as the fishing efficiency for each pass of sample i relative to the average sample. Detectability parameters (μ_p , σ_η^2 , and η_i) are estimated simultaneously with parameters representing spatial and spatiotemporal variation in intensity $\lambda(s, t)$. Refer to Table 1 for summary descriptions of all model parameters. The detection formulation could easily be adjusted for repeated site visits rather than depletion sampling where that is the preferred sampling method.

TABLE 1. (A) Description of parameters used in the model and (B) description of data used in the model.

Variable	Definition	Description
A) Parameters		
Overall		
$\lambda(s, t)$	mean abundance	mean abundance at time t and site s
$\boldsymbol{\gamma}^T$	coefficients	vector of fixed-effect coefficients on abundance
Detection		
μ_p	detection rate	mean log-rate of capturing an individual given that it is present at site s and time t (“log-detection rate”)
η_i	variation in detection	variation in log-detection rate samples
σ_η^2	detection variance	variance of $\eta(i)$
Spatial		
$\varepsilon(s)$	spatial contribution	spatial contribution to abundance following an Ornstein-Uhlenbeck (OU) process
$\sigma_s^2(s)$	spatial variance	variance between site s and its parent-site s_{parent} following an OU process
θ_ε	spatial decorrelation per kilometer	exponential spatial decay rate in the correlation between parent and child nodes
σ_ε^2	asymptotic spatial variation	parameter governing asymptotic variation in the spatial OU process for two sites that are far apart
$\rho_s(s)$	spatial correlation	spatial correlation for site s and its parent-site s_{parent} , resulting from an OU process
Temporal		
$\delta(t)$	temporal variation	temporal variation in abundance resulting from ARI autoregressive process
ρ_δ	temporal correlation per year	temporal correlation in the annual ARI process
σ_δ^2	temporal variance	variance describing the temporal ARI process
Spatiotemporal		
$v(s, t)$	spatiotemporal variation	spatiotemporal variation in abundance resulting from OU process
$\sigma_{st}^2(s)$	spatiotemporal variance	spatiotemporal variance between site s and its parent-site s_{parent}
ρ_{st}	spatiotemporal correlation per year	temporal decay rate of spatiotemporal variation, representing correlation for a given site in year t and year $t + 1$
$\rho_v(s)$	spatiotemporal correlation	spatiotemporal correlation between site s and its parent-site s_{parent} , resulting from an OU process
θ_v	spatiotemporal decorrelation per kilometer	spatial decay rate for spatiotemporal variation
σ_v^2	asymptotic spatiotemporal variance	parameter governing asymptotic variance describing the spatiotemporal OU process for two sites that are far apart
Independent		
$\alpha(s, t)$	overdispersion	random log-normal variation beyond Poisson expectation (also termed overdispersion or nugget)

TABLE 1. (Continued)

Variable	Definition	Description
σ_{iid}^2	overdispersion variance	variance term for the Poisson log-normal overdispersion term [$\alpha_i(s)$]
B) Data		
$\mathbf{x}(s)$	covariate data	row vector of measured variables affecting abundance (which includes an intercept term)
$c_d(s, t)$	count data	counts of fish for depletion pass d ($d \in \{1, 2, 3\}$), site and year t assuming that each individual is equally likely to be captured in a depletion pass given that it was not removed during a previous pass
ϕ_i	offset	relative length of stream surveyed to standardize abundance to fish per 100 m (offset = $\text{length}_{\text{survey}}/100$)

Note: See *Materials and Methods* for relevant equations and detailed descriptions.

Spatiotemporal correlations on a stream network

Working within a dendritic stream network, Euclidean distances are unlikely to represent the spatial similarity of population dynamics. Therefore, we instead approximate the similarity between two sites (i.e., correlations in spatial variation $\epsilon(s)$ and spatiotemporal variation $v(s, t)$ by the minimum distance between sites along the stream network (hydrologic distance). To do so, we augment the set of sampled sites (termed “sampling nodes”) with a set of “branching sites” (termed “branching nodes”) where two streams join. We then identify the “root” of the network as the node that is downstream of all other nodes in the network. We then move upstream from this root node and identify the nearest node along the network (or nearest nodes if the root is a branching node). In this case, we label the root node as the “parent” and the nearest node (or nodes) as “children.” Then starting from these children, we again move upstream to the nearest node or nodes and again record the parent–child relation between these nodes. This process is continued until we have reached the headwaters (or the highest sampling nodes) in each stream in the network. This description of the network has the important characteristics that each node is the child in one, and only one, parent–child relationship.

We also assume that changes in variables along the network are “memory-less,” that is, the value of a variable ϵ defined at a set of points along a stream segment follows a first-order Markov process where sites that are not connected by a child–parent relationship are statistically independent given a fixed value of ϵ at all other sites. This property arises from the assumption that the value of ϵ varies while moving along a stream network following a first-order stochastic differential equation.

Given these two properties (that the stream network is acyclic and that spatial variation is a first-order Markov process), we can calculate the conditional probability distribution for $\epsilon(s)$ each site as a function of its value $\epsilon(s_{\text{parent}})$ at the parent node s_{parent} for that site s , and the distance $|s - s_{\text{parent}}|$ between s and its parent s_{parent} . This allows us to

factor the joint probability of a spatial variable $\text{Pr}(\epsilon)$ into a series of easy-to-calculate conditional probabilities

$$\text{Pr}(\epsilon) = \prod_{s=1}^S \text{Pr}(\epsilon(s) | \epsilon(s_{\text{parent}})). \tag{3}$$

We further assume that variation in ϵ arises from a mean-reverting Weiner process with movement along the network. A Weiner process is a continuous stochastic process with independent increments often used to describe Brownian motion. Adding a mean-reverting component results in an Ornstein-Uhlenbeck process with the properties of being stationary, Gaussian, and Markovian for the right-hand side of Eq. 3, as we now describe in detail. This model is identical to the tail-down exponential model from Ver Hoef and Peterson (2010), although defining it as we do allows for easy computation within standard computational software.

Ornstein-Uhlenbeck process for spatial variation

We used the Ornstein-Uhlenbeck process to represent the spatial relationships along the network. The OU process implies that a child node will be correlated with its parent node as a function of distance following

$$\epsilon(s) | \epsilon(s_{\text{parent}}) \sim N(\rho_s(s) \times \epsilon(s_{\text{parent}}), \sigma_s^2(s)). \tag{4}$$

The variance, $\sigma_s^2(s)$, for site s conditional on the value of its parent s_{parent} given an OU process is

$$\sigma_s^2(s) = \frac{\sigma_\epsilon^2}{2\theta_\epsilon} \left(1 - e^{-2\theta_\epsilon |s - s_{\text{parent}}|} \right), \tag{5}$$

where θ_ϵ is the exponential rate of decay in correlation between child and parent nodes with larger values represent less correlation, $|s - s_{\text{parent}}|$ is the stream distance between parent and child nodes, and σ_ϵ^2 governs the asymptotic variance from an OU process for two sites that are far apart. The expected correlation between points in the network is then represented by $\rho_s(s)$ where

$$\rho_s(s) = e^{-\theta_\epsilon |s - s_{\text{parent}}|}. \tag{6}$$

Eqs. 5 and 6 are specified such that the pointwise variance of $\epsilon(s)$ (i.e., the variance $\epsilon(s)$ was drawn again from the same stochastic process) is $\sigma_\epsilon^2 \times (2\theta_\epsilon)^{-1}$.

First-order autocorrelation for temporal variation

We include a temporal term $\delta(t)$ for each year t to represent years that are higher or lower than expected across all sites. We model vector δ (representing $\delta(t)$ in all years) using first-order autocorrelation

$$\delta = \text{MVN}(\mathbf{0}, \sigma_t^2 \mathbf{R}_\delta) \tag{7}$$

where σ_t^2 is the variance in this temporal term, and \mathbf{R}_δ is the correlation matrix for a first-order autocorrelation process

$$\mathbf{R}_\delta(t, t^*) = \rho_\delta^{|t - t^*|} \tag{8}$$

where $\mathbf{R}_\delta(t, t^*)$ is the correlation between years t and t^* , separated by $|t - t^*|$ years, and ρ_δ is an estimated parameter representing the correlation in δ for two adjacent years.

Ornstein-Uhlenbeck process for spatiotemporal variation

We similarly used the OU process to represent the spatiotemporal relationships along the network. We use the vector $\mathbf{v}(s)$ to represent the spatiotemporal term $v(s, t)$ for all years t , and it varies along the network as an OU process

$$\mathbf{v}(s) \sim \text{MVN}(\rho_v(s) \times \mathbf{v}(s_{\text{parent}}), \sigma_{\text{st}}^2(s) \mathbf{R}_v) \quad (9)$$

where $\sigma_{\text{st}}^2(s)$ is again the spatiotemporal variance for site s given the vector $\mathbf{v}(s_{\text{parent}})$ for its parent, s_{parent}

$$\sigma_{\text{st}}^2(s) = \frac{\sigma_v^2}{2\theta_v} \left(1 - e^{-2\theta_v|s-s_{\text{parent}}|}\right). \quad (10)$$

The parameter $\rho_v(s)$ is the correlation due to spatial similarity

$$\rho_v(s) = e^{-\theta_v|s-s_{\text{parent}}|} \quad (11)$$

and \mathbf{R}_v is the correlation due to temporal similarity, which we assume follows first-order autocorrelation (Eq. 8 but replacing ρ_δ with ρ_{st} , where ρ_{st} is an estimated parameter representing the temporal correlation between two adjacent years in spatiotemporal variation $v(s, t)$. We assumed that the decorrelation distance was identical for spatial and spatiotemporal variation (i.e., $\theta_v = \theta_\epsilon$), but the variance in the OU process was independent for the spatial and spatiotemporal components. This assumption could be relaxed in the future but may require large networks with a large amount of spatially and temporally replicated data to fit, potentially beyond what is available for most studies.

Parameter estimation

We estimate parameters within a mixed-effects model, while treating variation in detectability ($\eta(s, t)$) as well as overdispersion ($\alpha(s, t)$), temporal ($\delta(t)$), spatial ($\epsilon(s)$), and spatiotemporal ($v(s, t)$) variation in density as random effects. We estimate parameters by maximizing the marginal likelihood function with respect to fixed effects, where the marginal likelihood function is calculated using the Laplace approximation to approximate the integral across random effects. We confirm that the model is converged by ensuring that the gradient of the marginal likelihood with respect to each fixed effect is within ± 0.001 and that the Hessian matrix is positive definite. Parameter estimation was conducted using Template Model Builder (Kristensen et al. 2016) within the R statistical platform (R Development Core Team 2016). We note that the Ornstein-Uhlenbeck process results in an exponential correlation function, and Ver Hoef et al. (2006) show that this correlation function is valid (i.e., will result in a positive definite covariance). Log-density is the sum of a series of multivariate normal random effects (Eq. 1, with the exception of $\mathbf{x}(s)\boldsymbol{\gamma}$ where $\boldsymbol{\gamma}$ is a fixed effect) and each multivariate normal distribution is valid (has a positive definite covariance). Therefore, log-density is

itself valid, as follows from the properties of an additive covariance function. Further research could formally decompose the proportion of variance in log-density that is attributed to each additive component in Eq. 1, although we do not do so here. Future studies could also expand upon or modify the framework used here, although changes may not be identifiable (i.e., the Jacobian matrix of sufficient statistics for the data with respect to parameters might be rank-deficient) or estimable (the Hessian matrix of the marginal log-likelihood at the maximum likelihood estimator may not be positive definite). We recommend future analyses check estimability using automatic differentiation (as we have done here), and future theoretical work should examine identifiability in spatiotemporal models (e.g., following methods in Hunter and Caswell 2009)

Spatial simulations

We conducted simulations to evaluate model performance. The first set of simulations was designed to test the ability to estimate spatial correlations and how well the model estimated abundance with varying levels of spatial autocorrelation compared with a nonspatial model. We included a single covariate on density that differed by location but was not spatially autocorrelated ($\boldsymbol{\gamma}^T = [2.3, 0.5]$; intercept, covariate coefficient). We used a mean density of 10 fish per 100 m ($x_1(s) = \ln(10)$). Both models were identical except for the inclusion of the spatial variation component ($\epsilon(s)$). These models were performed for a single year and assuming no extra-Poisson overdispersion so temporal, spatiotemporal, and independent components, $\delta(t)$, $v(s, t)$, and $\alpha(s, t)$ were not included in the models. We assumed a constant detection rate of 0.5 ($\mu_p = \ln(0.5)$) and three-pass depletion sampling at each location. Therefore, the probability of detecting an individual that remained in the stream on any given pass was 50%.

For the spatial model, we simulated data with all combinations of θ_ϵ in $\{0.5, 1, 2, 3\}$ and σ_ϵ in $\{0.1, 0.25, 0.5, 0.75\}$. These values of θ_ϵ represent a large range in correlations such that when $\theta_\epsilon = 0.1$ then $\rho_s(s) = 0.607$ whereas when $\theta_\epsilon = 3$ then $\rho_s(s) = 0.050$ at points 1 km apart.

We ran 200 simulations for each combination of θ_ϵ and σ_ϵ and fit each simulated data set with the spatial model described (single year with no temporal or spatiotemporal variation) and with a nonspatial model. We varied the spatial decorrelation per kilometer, θ_ϵ , and the asymptotic spatial variance, σ_ϵ , independently, but only examine the effects of the combined spatial component ($\theta_\epsilon\sigma_\epsilon$) because θ_ϵ and σ_ϵ are not independently identifiable. We ran the simulation using the White River watershed in Vermont with 359 nodes because it was a reasonably sized network with sufficient distances and numbers of nodes to be diverse but not so large as to make simulation of the network correlations excessively long. Distances between child and parent nodes ranged from 0.17 to 5.13 km with a mean of 1.13 km. The R code for simulating the data is *available online*.⁵

⁵ https://github.com/djhocking/Trout_GRF/blob/master/Code/Spatial_Simulations.R

Spatiotemporal simulation experiment

We also wanted to understand the effect of spatial and temporal replication on model performance. We simulated 300 independent data sets for the White River in Vermont over 20 yr for each of the 359 nodes. For each simulation, we randomly sampled the data to represent surveying various numbers of sites and years (all combinations of 4, 8, 10, 15, and 20 yr with 25, 50, 100, and 359 sites). For each survey combination and simulation, we fit the spatiotemporal model including spatial, temporal, and spatiotemporal dynamics (matching the data generating model) and a temporal model with no spatial or spatiotemporal dynamics. For each simulation, we used $\theta_v = \theta_\varepsilon = 0.3$, $\sigma_\varepsilon = 0.5$, $\rho_\delta = 0.6$, $\sigma_t = 0.2$, $\sigma_v = 0.4$, $\rho_{st} = 0.7$, detection probability $P = 0.5$, and density coefficients $\gamma^T = [2.3, 0.2]$, where the first value in γ^T is the log-mean intercept and the second value is the coefficient (slope) of a site-level covariate. All R and TMB functions for these simulations can be found in the Data S1 supplemental materials with additional information in Metadata S1.

Brook trout case study

As the only trout native to eastern U.S. streams and rivers, brook trout are a species of social and economic importance in the region. State and federal agencies as well as organizations such as Trout Unlimited and the Eastern Brook Trout Joint Venture (EBTJV) have particular interest in supporting viable populations of brook trout. As such, there have been numerous recent modeling efforts to estimate occupancy, abundance, and population dynamics in response to landscape conditions, climate change, and management actions (DeWeber and Wagner 2014, Kanno et al. 2015, Letcher et al. 2015, Bassar et al. 2016). However, these models generally do not account for spatial correlations beyond using random regional, watershed, or sub-basin effects.

We identified the West Susquehanna, Pennsylvania watershed for our case study because it was a moderately large network with a high density of good quality stream fish data over a long-time period. The electrofishing data were collected by the state of Pennsylvania Boat and Fish Commission using standard methods common across agencies and researchers throughout the eastern United States. We did not use the West Susquehanna watershed in our simulations because it is much larger than the White River network, with many more confluences, which would greatly slow the data simulation.

The West Susquehanna watershed contained 11,220 nodes, comprised of 349 survey sites and 10,871 stream reaches. Sites were surveyed a total of 34 yr from 1981 and 2014. There was a total of 683 site visits with a mean of 2.0 and a range of 1–21 visits per site. The total drainage area of the watershed was 18,068 km² and the smallest stream had a cumulative drainage area of 0.4 km². The median drainage area was 4.4 km². The mean distance between nodes in the network was 1.37 km and ranged from 0.001 to 11.61 km with a median of 1.11 km.

The watershed was primarily forested (mean percent forest cover = 79%) but with a range from 0% to 100% within individual stream catchments. We used percent forest cover as a fixed-effect covariate in our model along with surficial

coarseness, mean air temperatures from the summer (previous year), fall (previous year), winter, and spring prior to summer fish surveys, and mean daily precipitation for the same seasons. Daily temperature and precipitation data were obtained from daymet (Thornton et al. 1997, 2016) and spatially aggregated to the catchment scale. The surficial coarseness was the percentage of the catchment area covered by a parent soil material with texture described as sandy, gravelly, or a combination of the two. These classifications were obtained from the USDA National Resources Conservation Sciences Soil Survey Geographic Database (SSURGO; Soil Survey Staff 2015). Forest cover data were obtained from the 2011 National Land Cover Database (NLCD; Homer et al. 2015). All basin characteristics were calculated as spatial sums (precipitation) or means within each zonal catchment layer as delineated based on the truncated NHDHRDV2 flowlines. All details and ArcPython scripts are *available online*.⁶ The covariate summary statistics for the West Susquehanna watershed are presented in Table 2.

We used the National Hydrography Dataset high-resolution flowlines truncated to >0.75 km² drainage area for spatial consistency and exclusion of highly ephemeral streams (flowlines *available online*).⁷ Any survey locations or other points of interest were then snapped to the flowlines. All survey points and confluences, including the base of the network and the terminal headwaters, were considered network nodes. Except for the base node, the distance from each child node was calculated to its downstream parent node to define the network relationships and distances. All hydrography processing was done using ArcPython in ArcGIS v10.2 (Environmental Systems Research Institute, Redlands, California, USA). The full description of the process, scripts, and links to the hydrography data is archived *online*.⁶ The hydrography for the region from Maine to Virginia, USA, can be downloaded by hydrologic unit code 2 (*available online*).⁸ All

TABLE 2. Summary table of covariate values for the West Susquehanna watershed.

Variable	Mean	Minimum	Maximum
Forest cover (%)	79.15	0	100
Surficial coarseness (% sandy/gravelly)	6.62	0	100
Previous summer mean temperature (°C)	17.74	15.21	21.66
Previous fall mean temperature (°C)	3.49	−0.09	7.3
Winter mean temperature (°C)	−1.77	−7.99	2.87
Spring mean temperature (°C)	14.63	10.31	17.31
Previous summer mean precipitation (mm)	3.78	1.59	8.92
Previous fall mean precipitation (mm)	2.99	1.29	5.01
Winter mean precipitation (mm)	2.58	1.1	4.73
Spring mean precipitation (mm)	2.91	1.42	6.9

Note: Susquehanna watershed was defined by catchment draining into each stream reach.

⁶ <http://conte-ecology.github.io/shedsGisData/>

⁷ <https://nhd.usgs.gov/index.html>

⁸ <http://ecosheds.org/assets/nhdhrd/v2/>

continuous covariates were standardized by subtracting the mean and dividing by the standard deviation for computational efficiency. None of these variables had Pearson correlations >0.60.

Model selection

For young of the year (YOY) and adult brook trout independently, we compared eight models with different combinations of spatial, temporal, and spatiotemporal correlations (2 × 2 × 2 factorial design; Table 3). All other components of the model including fixed-effect covariates were identical in all models. Meteorological conditions during the previous summer were used in the adult models but were excluded in the YOY models because spawning does not occur until the fall. We used Akaike’s information criterion (AIC) to select the best model balancing model fit and model complexity (Burnham 2004, Burnham et al. 2010).

RESULTS

Spatial simulations

Overall, we found that the spatial model had greater precision for estimates of covariate effects or population density than the nonspatial model. We found that the spatial model estimated the spatial component ($\theta_\epsilon \sigma_\epsilon$) well when there was strong spatial correlation but tended to slightly underestimate the component when the spatial decorrelation was low (θ_ϵ large; Fig. 1). The spatial model estimated the mean intensity (expected fish per 100 m) across the watershed much better than the nonspatial model when there was moderate to high spatial decorrelation rates (Fig. 1) and the mean uncertainty (SE) of the estimated intensity was much larger for the nonspatial model compared with the spatial model when the spatial decorrelation was large (Fig. 1). We used the difference between model predictions and true values to calculate the root-mean-squared error (RMSE) as an assessment of model predictive accuracy. The RMSE was far larger for the nonspatial model compared with the spatial model across all values of θ_ϵ (Fig. 1), indicating that

abundance estimates at individual locations were much more accurate for the spatial model. This difference in uncertainty was largest with high levels of spatial correlation. The fixed-effect coefficient for the single covariate (γ^T) was estimated well across all values of θ_ϵ , but the variation in this estimate was slightly smaller for the spatial model, especially at higher levels of spatial correlation (Fig. 1).

The range of σ_ϵ also significantly influenced the parameter estimates and the differences between spatial and nonspatial models. The spatial model recovered $\theta_\epsilon \sigma_\epsilon$ well with very slight underestimation on average, except when σ_ϵ was small (0.1), which resulted in good recovery on average but extremely high variation among simulations (Fig. 2). The spatial and nonspatial models performed similarly in the estimation of mean intensity across the watershed when the true value of σ_ϵ was small but the spatial model was more accurate and more precise compared with the nonspatial model as the level of σ_ϵ increased (Fig. 2). The uncertainty in mean network intensity went up for the nonspatial model as σ_ϵ increased but was constant for the spatial model across levels of σ_ϵ (Fig. 2). The RMSE was again much smaller for the spatial model compared with the nonspatial model as σ_ϵ increased. The variability in the RMSE also increased greatly for the nonspatial model as σ_ϵ increased (Fig. 2). The fixed-effect coefficient was estimated well for both models but the uncertainty increased in the nonspatial model as σ_ϵ increased (Fig. 2).

Spatiotemporal simulation study

We found the mean network intensity was estimated fairly well for both the spatial and nonspatial models, but both models tended to slightly underestimate abundance slightly when few years were surveyed (Fig. 3). The value of $\theta_\epsilon \sigma_\epsilon$ was underestimated with <15 yr of data while the estimates of $\theta_0 \sigma_0$ were proportionally overestimated with <15 yr of data (Fig. 3). This same pattern was observed with fewer than 100 sampled sites (Fig. 4). Similarly, for both the spatial and nonspatial model, it took 15–20 yr to accurately recover the temporal autocorrelation, although it was still slightly underestimated by the spatial model. The variability in the temporal process was recovered well with the spatial model regardless of the number of years surveyed but the nonspatial model had more variation among simulations with increasing years surveyed (Fig. 4). The value of the fixed-effect covariate, γ , was estimated well for both models regardless of the number of years sites were sampled but the variation in the estimation was consistently lower for the spatial model (Fig. 3).

The number of sites sampled similarly influenced the estimation of the spatial and spatiotemporal components, with an increasing number of sites improving the estimates, especially from 25 to 100 sites (Fig. 4). The RMSE improved with an increasing number of sites sampled for the spatial model and was lower (more accurate) for the spatial model with 100 or more sites compared with the nonspatial model (Fig. 4). The fixed-effect coefficient was recovered well for both models although the estimate was biased low for the nonspatial model with only 25 sites. The precision in the fixed-effect estimate improved with the number of sites sampled and was consistently better for the spatial model.

TABLE 3. Description of models compared with the Akaike information criterion (AIC) for adult and young of the year (YOY) brook trout populations in the West Susquehanna watershed.

Number	Model	Model components
1	basic	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \alpha(s, t)$
2	spatial (S)	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \epsilon(s) + \alpha(s, t)$
3	temporal (T)	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \delta(t) + \alpha(s, t)$
4	S + T	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \epsilon(s) + \delta(t) + \alpha(s, t)$
5	spatiotemporal (ST)	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + v(s, t) + \alpha(s, t)$
6	S + ST	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \epsilon(s) + v(s, t) + \alpha(s, t)$
7	T + ST	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \delta(t) + v_t(s) + \alpha(s, t)$
8	S + T + ST	$\ln(\lambda(s, t)) = \gamma^T \mathbf{x}(s)_i + \epsilon(s) + \delta(t) + v(s, t) + \alpha(s, t)$

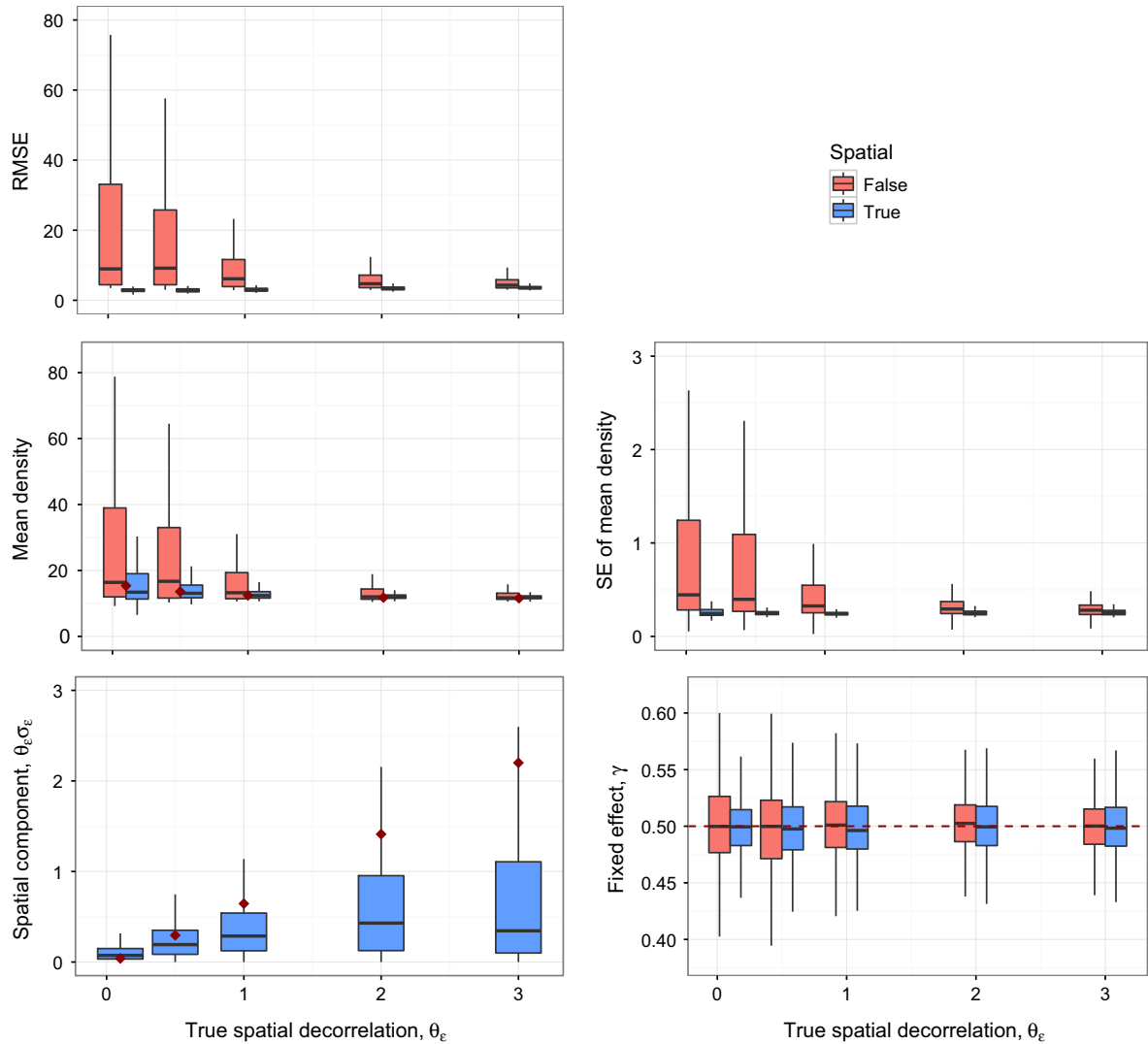


FIG. 1. Parameter estimates across different values of θ_ϵ (spatial decorrelation per kilometer) from the spatial simulation study varying θ_ϵ and σ_ϵ (the asymptotic spatial variance; Eq. 5). Lower values of θ_ϵ represent higher correlation with distance. Parameter estimates' accuracy (root mean square error, RMSE) were compared for the spatial model (blue; Model 2 in Table 3) and a nonspatial model (red; Model 1 in Table 3). Values in the box plots represent the combined uncertainty from 200 simulations and variation in simulated levels of σ_ϵ . Red points represent the true simulation values. The midline represents the median and the boxes extend from the 25th to the 75th percentile (Interquartile range). Whiskers extend from the smallest to the largest observation but not beyond 1.5 times the interquartile range.

Despite reasonable estimates of mean abundance and fixed effects in many simulations, the nonspatial model (Model 3 in Table 3) generally did not sufficiently recover the heterogeneity and spatial pattern in density as seen in Fig. 5.

Brook trout case study

The top YOY model included temporal and spatiotemporal components. The null model was the worst and any model with a spatial or spatiotemporal component was ranked higher than the temporal-only model (Table 3). For adult brook trout, the spatiotemporal model and the temporal plus spatiotemporal model were the top two models with a ΔAIC of 0.3 (Table 4). We chose to draw inference from the temporal plus spatiotemporal model for the easiest direct comparison with the YOY. The most complex model

(containing temporal, spatial, and spatiotemporal components from Eq. 1) failed to converge with the adult data and was excluded from model comparison.

From the top models, we estimated the temporal and spatiotemporal model parameters along with the fixed effects, detection probabilities, and overdispersion terms. Adults exhibited strong temporal correlation per year ($\rho_\delta = 0.59$) with low variability ($\sigma_\tau = 0.16$), whereas YOY exhibited no temporal correlation per year ($\rho_\delta = -0.05$) but high stochastic temporal variability ($\sigma_\tau = 0.76$). The estimated values of the spatiotemporal decay θ_0 were at the lower end of what we tested with simulations for both YOY (0.13) and adults (0.16), indicating high spatiotemporal correlation (~50% at 5 km; Fig. 6). The estimates of the spatiotemporal standard deviation σ_0 were high for YOY (0.65) and adults (0.59). The combination of the two parameters indicates extremely high

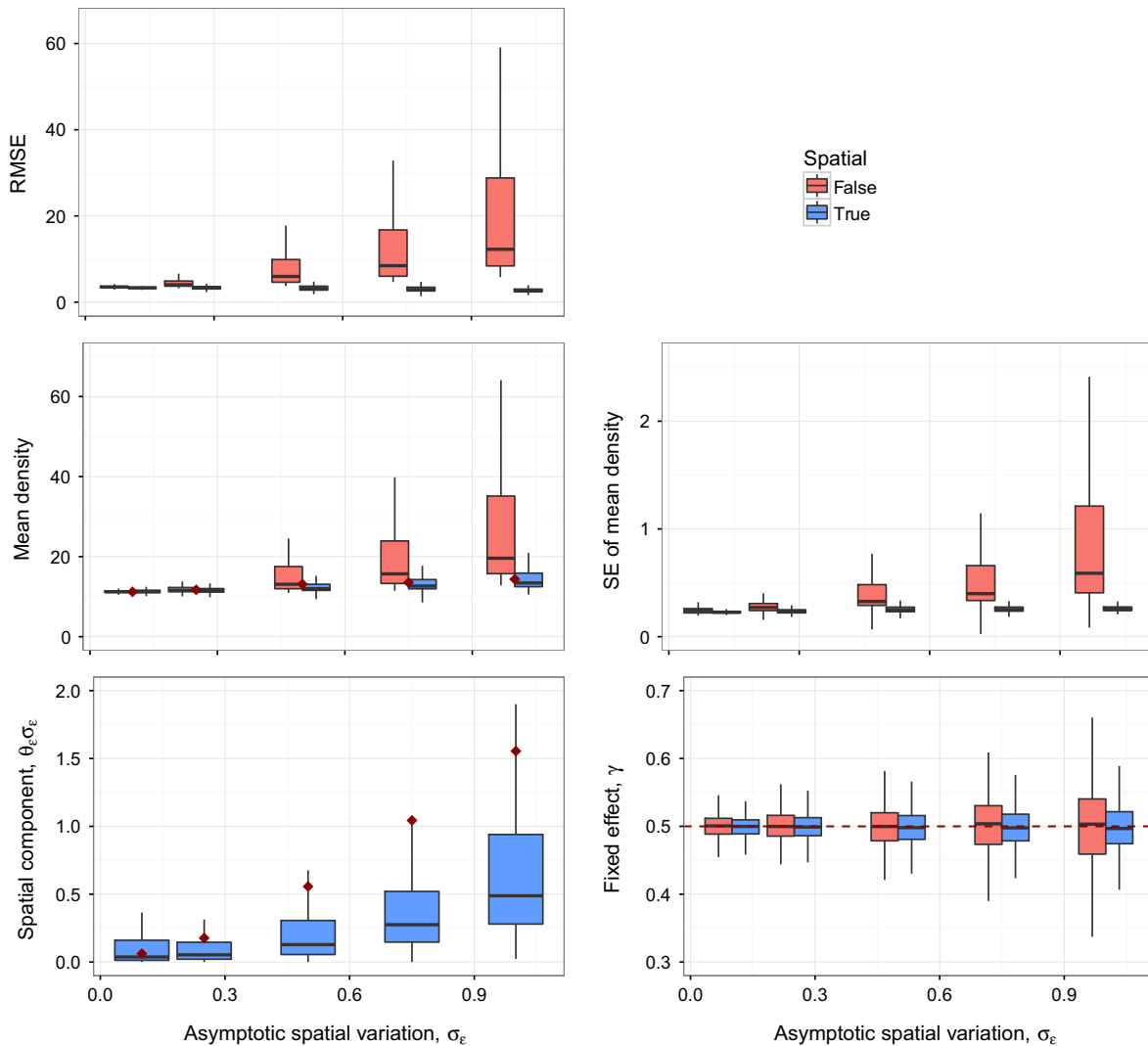


FIG. 2. Parameter estimates across different values of σ_ϵ from the spatial simulation study varying θ_ϵ and σ_ϵ (Eq. 5). Parameter estimates' accuracy (RMSE) were compared for the spatial model (blue; Model 2 in Table 3) and a nonspatial model (red; Model 1 in Table 3). Ranges in the box plots represent the combined uncertainty from 200 simulations and variation in simulated levels of θ_ϵ .

spatiotemporal autocorrelation, which is revealed by the very high estimate of temporal decay ρ_{st} of 0.98 and 0.97 for YOY and adults, respectively (Table 5). Forest cover, the previous year's mean summer temperature, spring temperature, and, to a lesser extent, the previous fall mean temperature were all important predictors of adult abundance. For YOY, only forest cover and mean spring temperature had substantial effects on abundance. Seasonal precipitation did not influence abundance for YOY or adults (Table 5).

DISCUSSION

We developed a geostatistical model for estimating animal densities within dendritic networks while accounting for imperfect detection. Spatial simulations demonstrated improved estimates of animal densities even at relatively low levels of spatial correlations compared with traditional non-spatial models (Fig. 1). Even when the spatial decay rate (θ_ϵ) was one (36% correlation at 1 km and virtually zero

correlation at 10 km), the spatial model had significantly higher predictive accuracy of reach-level density. There were no scenarios where the spatial model performed worse than the nonspatial model for estimating mean density.

Similarly, we demonstrated the benefits of our model over a large range of years and surveyed sites through simulation. The accuracy improved with increasing number of years that sites were surveyed (Fig. 3; RMSE). However, there was a large improvement in recovery of the spatial and spatiotemporal components of the model given 15–20 yr of data. Although there is moderately high uncertainty in the estimation of the spatial and spatiotemporal components (in Fig. 3), this is likely due in part to combining simulation replication uncertainty with variation among sites while holding the number of years constant. Similarly, the variation in recovery of the spatial and spatiotemporal components was likely inflated (in Fig. 4) because of combining simulation uncertainty with variation in the number of years surveyed while only holding the number of sites constant.

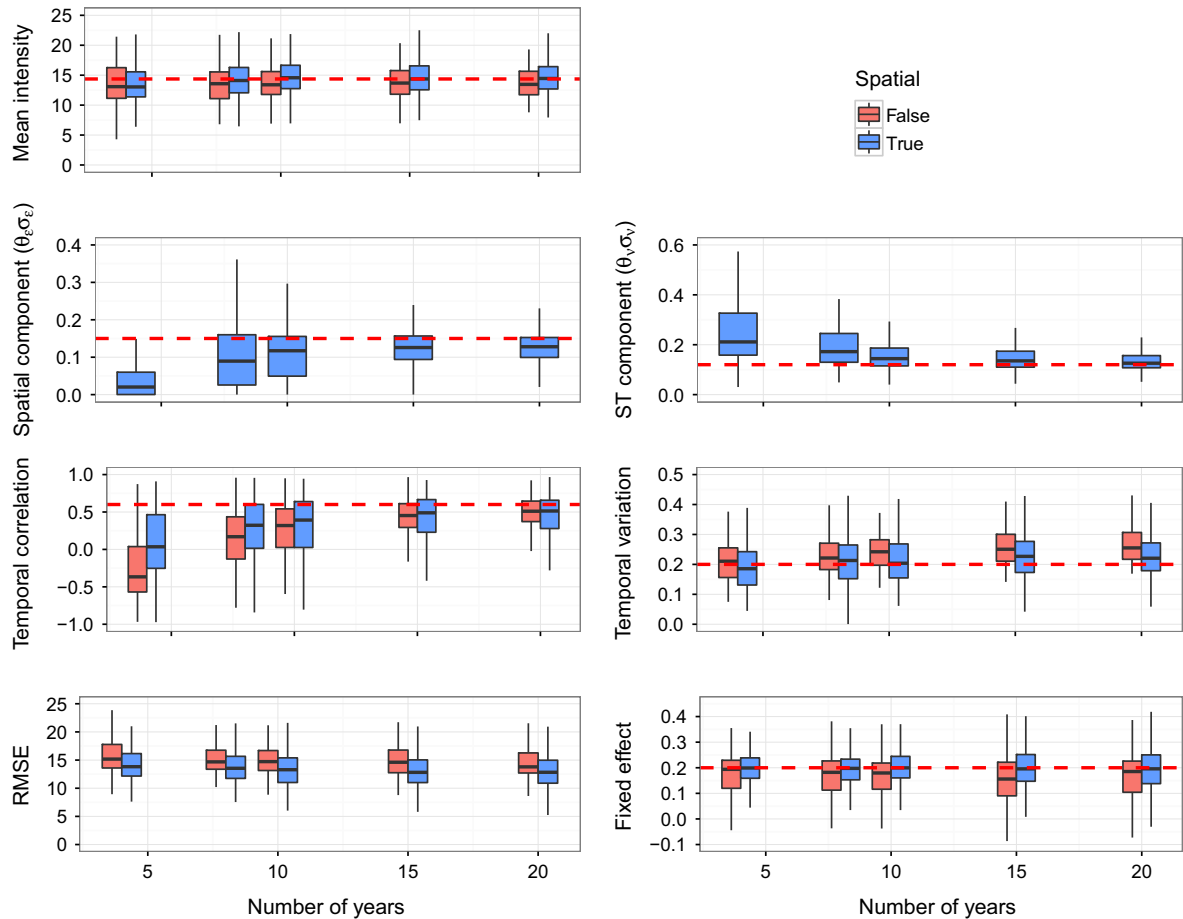


FIG. 3. Results from the power simulation study showing the effect of varying the number years each site is surveyed. Variation described by each box plot resulted from the 300 simulations and varying the number of sites (while holding the number of years constant). For each simulation, we used $\theta_0 = \theta_e = 0.3$, $\sigma_e = 0.5$, $\rho_s = 0.6$, $\sigma_r = 0.2$, $\sigma_v = 0.4$, $\rho_{st} = 0.7$, detection probability $P = 0.5$, site-level covariate on abundance $\gamma^T = 0.2$, and mean abundance = 10. Dashed red lines represent true values. Refer to Table 1 for definitions of all parameters.

However, the spatial model showed clear improvement in recovery of the spatial correlation and accuracy of local density estimation (RMSE) with an increased number of surveyed sites. Based on these limited simulations, we recommend aiming for at least 15 yr of data for 100 sites (given spacing of sites similar to those considered here). However, further investigation is warranted to explore the effects of having a collection of sites that are visited at different intervals as is the case with many freshwater fisheries data sets. It is possible that only a subset of sites would have to be visited each year to adequately characterize the spatiotemporal dynamics. Although this may appear as a large number of sites and years, many state agencies already have these data from long interest in freshwater fisheries stock status. For some watersheds, multiple agencies and NGOs might have to pool data to have sufficient replication furthering the argument for regional cross-boundary databases. When fewer years of data are available, the reduced spatial model without the temporal or spatiotemporal components can be run on each year of data with improved estimation in comparison with a traditional nonspatial model (Figs. 2, 3).

The non-spatiotemporal model was also good at recovering densities but did poorly at estimating temporal

correlations and variability (Figs. 4, 5). Given the large improvements of the spatial model compared with the nonspatial model in the single year simulation (Figs. 2, 3), this may not be a general pattern and may depend on network topology and sampling density within the network and over time. Additionally, the nonspatial model will not be as useful for estimating local densities in unsampled stream reaches. Many management actions occur at small scales and therefore understanding local population dynamics is important for prioritizing local actions and understanding the effects of those actions, particularly in an adaptive management framework. Such a situation could occur for decisions that are repeated and adjusted based on population responses such as stocking programs or setting stream-level fishing regulations (e.g., barbless hooks, catch and release, take limits). Even for one-time decisions such as in-stream habitat modification, and dam or culvert removal at a local site, it is important to have good estimates of local, rather than just watershed, abundance because the local change in abundance can help prioritize the location of the next project.

This spatiotemporal model can readily be applied to existing standard electrofishing data from state and federal agencies. Using this model with brook trout data collected the

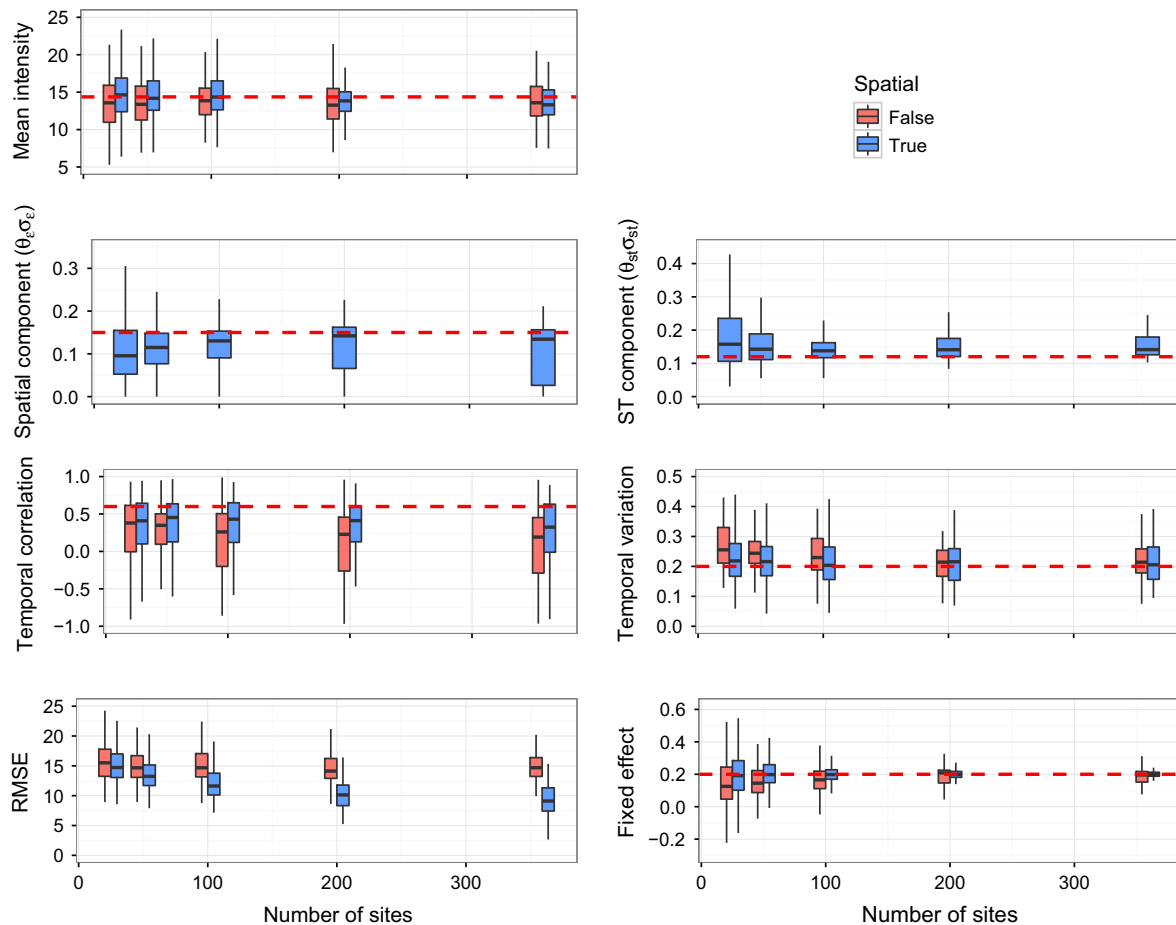


FIG. 4. Results from the power simulation study showing the effect of varying the number of sites surveyed. Variation described by each box plot resulted from the 300 simulations and varying the number of years each site was surveyed (while holding the number of sites constant). For each simulation, we used $\theta_v = \theta_e = 0.3$, $\sigma_e = 0.5$, $\rho_\delta = 0.6$, $\sigma_t = 0.2$, $\sigma_v = 0.4$, $\rho_{st} = 0.7$, detection probability $P = 0.5$, site-level covariate on abundance $\gamma^T = 0.2$, and mean abundance = 10. Refer to Table 1 for definitions of all parameters.

Pennsylvania Boat and Fish Commission, we demonstrated improved model fit compared with basic nonspatial models even accounting for increased model complexity (i.e., using AIC). For adult brook trout, the spatiotemporal model and the model with temporal and spatiotemporal components outperformed all other models (Table 4). Similarly, the temporal plus spatiotemporal model performed best with the YOY data (Table 4). In addition to evidence from model comparisons, the estimated coefficient values for brook trout data fell within our range of simulations indicating that the estimates are reliable. Removing spatial or temporal covariates could change which model was selected based on AIC. It would also change the estimate of the spatial and spatiotemporal correlations because they can be interpreted as the latent correlations not explained by the fixed effects resulting from unmeasured, complex, biotic and abiotic interactions exhibiting spatial autocorrelation.

Both YOY and adult densities were positively associated with forest cover and negatively associated with spring temperatures (Table 5). This finding is similar to brook trout model results from a rangewide occupancy model (Wagner et al. 2014). A recent review of salmonid fish response to environmental drivers (Kovach et al. 2016) also found

negative effects of increased seasonal temperature on trout populations, supporting our estimate of a strong negative summer temperature effect. Similar results were also found for an Adirondack lake (Robinson et al. 2010), streams in West Virginia (Huntsman and Petty 2014) and Michigan (Grossman et al. 2012), and from demographic models for brook trout in Shenandoah National Park (SNP; Kanno et al. 2015, 2016). We also found that temperature had a larger effect on YOY than on adults. Similarly, Bassar et al. (2016) found that population dynamics in a small stream system were largely driven by the effects of yearly temperature variation on YOY.

We estimated only weak relationships between seasonal precipitation and trout density. This is in stark contrast to the strongly negative effects of winter precipitation found in SNP (Kanno et al. 2015, 2016). Topographical and geological differences may help explain the divergent effects of precipitation estimated for the two study areas. Trout habitat in SNP is high elevation and high gradient while the sites we studied are more variable in elevation, aspect, and gradient potentially obscuring precipitation effects. It is likely that precipitation will have a much greater effect on trout populations in high gradient, nonporous sites. It is also likely that

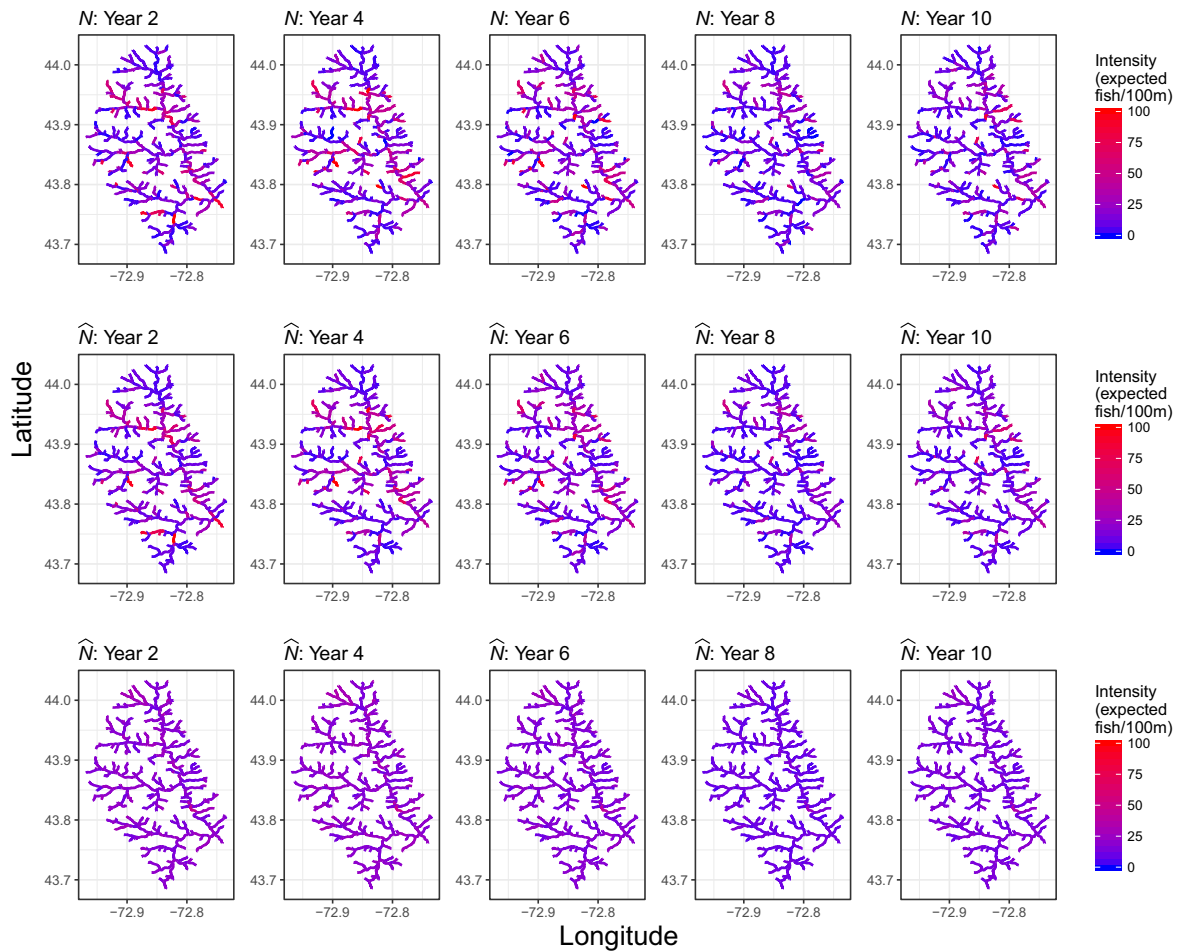


FIG. 5. Example of a spatiotemporal simulation of the abundance along a stream network over time (intensity). The top row shows the true (simulated) abundances and the middle row depicts the mean expected abundance based on the matching spatiotemporal model (Eq. 1). The bottom row shows the mean expected density for a model with temporal autocorrelation but no spatial or spatiotemporal contributions (Model 3 in Table 3). For each simulation, we used $\theta_v = \theta_e = 0.3$, $\sigma_e = 0.5$, $\rho_\delta = 0.6$, $\sigma_t = 0.2$, $\sigma_v = 0.4$, $\rho_{st} = 0.7$, detection probability $P = 0.5$, site-level covariate on abundance $\gamma^T = [2.3, 0.2]$.

we underestimated the importance of precipitation in general because we estimated effects of seasonal precipitation means over an area without large spatial variability in precipitation patterns. Floods can have dramatic effects on salmonids, including year class loss (Letcher and Terrick 1998, Carline and McCullough 2003) and, in extreme cases, local extirpation (Vincenzi et al. 2014). Recolonization ability, habitat complexity, and high fecundity, however, all contribute to high resilience of brook trout populations to floods (Nislow et al. 2002, Roghair et al. 2002, George et al. 2015). It is possible that this model could be used to assess the spatial and spatiotemporal decay rates related to the effects of major flood events such as hurricanes in the future. This could be important as flood frequency and severity is expected to increase with climate change in some parts of the world (Hirabayashi et al. 2013).

Adult brook trout exhibited higher temporal correlation and less unexplained random variation (overdispersion SD; Table 6) in density compared with YOY. This supports previous findings of high YOY variability due to temperature and flow conditions along with other stochastic events (Carlson

and Letcher 2003, Xu et al. 2010, Kanno et al. 2015, 2016). Both adults and YOY densities exhibited similar levels of spatiotemporal correlation with relatively slow decorrelation with distance as evidenced by the low spatiotemporal decay rates ($\theta_v = 0.16$, and 0.13 , respectively) and high asymptotic spatiotemporal variances (Table 6). The effect of these parameters can be seen in Fig. 6, which shows correlation with distance. For example, correlation is approximately 50% at 5 km and 25% at 10 km for YOY. Adult correlations are only slightly lower than for YOY with hydrologic distance. There is virtually no autocorrelation in densities of YOY or adults at distances beyond 20 km. This suggests that these populations are generally operating independently (minimal density-dependent reshuffling through movement) and that landscape, land-use, and meteorological variables are sufficient to describe any long-distance correlations in brook trout population dynamics. Our model can also be used to predict densities at unsampled sites to aid natural resource managers in making decisions at locations when there is insufficient time to collect local data prior to when an impact or management action will occur.

This is unsurprising given the general short movements and high genetic differentiation of brook trout over relatively short distances (Whiteley et al. 2013). This exceedingly high spatiotemporal correlation per year indicates a slow rate of change in the spatial patterning (i.e., high densities sites tended to maintain relatively high densities,

TABLE 4. Comparison of brook trout models for the West Susquehanna watershed using AIC.

Number	Model	AIC	Δ AIC
Adult			
5	ST	9,408	0.0
7	T + ST	9,408	0.3
4	S + T	9,583	175
2	S	9,588	180
3	T	9,783	375
1	basic	9,794	387
6	S + ST	NA	NA
8	S + T + ST	NA	NA
YOY			
7	T + ST	9,592	0
8	S + T + ST	9,596	4
5	ST	9,663	71
6	S + ST	9,666	74
4	S + T	9,739	147
2	S	9,801	209
3	T	9,925	333
1	basic	10,048	456

Note: The models (6 and 8) that included spatial and spatiotemporal components failed to converge with adult data and were not used in the comparison.

TABLE 5. Summary of parameter estimates from the model including temporal and spatiotemporal components for adult brook trout in the West Susquehanna watershed.

Parameter	Adult		YOY	
	Estimate	SE	Estimate	SE
Intercept	-2.45	0.13	-3.41	0.30
Forest cover	0.82	0.12	1.12	0.16
Surficial coarseness	0.01	0.06	0.04	0.08
Summer temperature	-0.26	0.05		
Fall temperature	0.09	0.03	0.02	0.11
Winter temperature	-0.01	0.03	0.05	0.11
Spring temperature	-0.16	0.05	-0.68	0.16
Summer precipitation	-0.02	0.01		
Fall precipitation	0.05	0.02	0.01	0.04
Winter precipitation	0.04	0.02	-0.01	0.05
Spring precipitation	0.05	0.02	-0.06	0.06
Detection rate (μ_p)	1.35	0.02	1.08	0.02
Detection SD (σ_η)	0.21	0.02	0.30	0.02
Temporal correlation per year (ρ_δ)	0.59	0.26	-0.05	0.21
Temporal SD (σ_t)	0.16	0.06	0.76	0.13
Spatiotemporal decorrelation per year (ρ_{st})	0.97	0.01	0.98	0.01
Spatiotemporal decorrelation per kilometer (θ_b)	0.16	0.03	0.13	0.02
Asymptotic spatiotemporal SD (σ_b)	0.59	0.06	0.65	0.07
Overdispersion SD (σ_{iid})	0.36	0.04	0.53	0.04

Notes: The YOY model did not include the previous summer temperature or precipitation since they were not yet laid as eggs. Parameters are defined in Table 1. The first 11 parameters were fixed effects on abundance contained in the vector of coefficients γ^T .

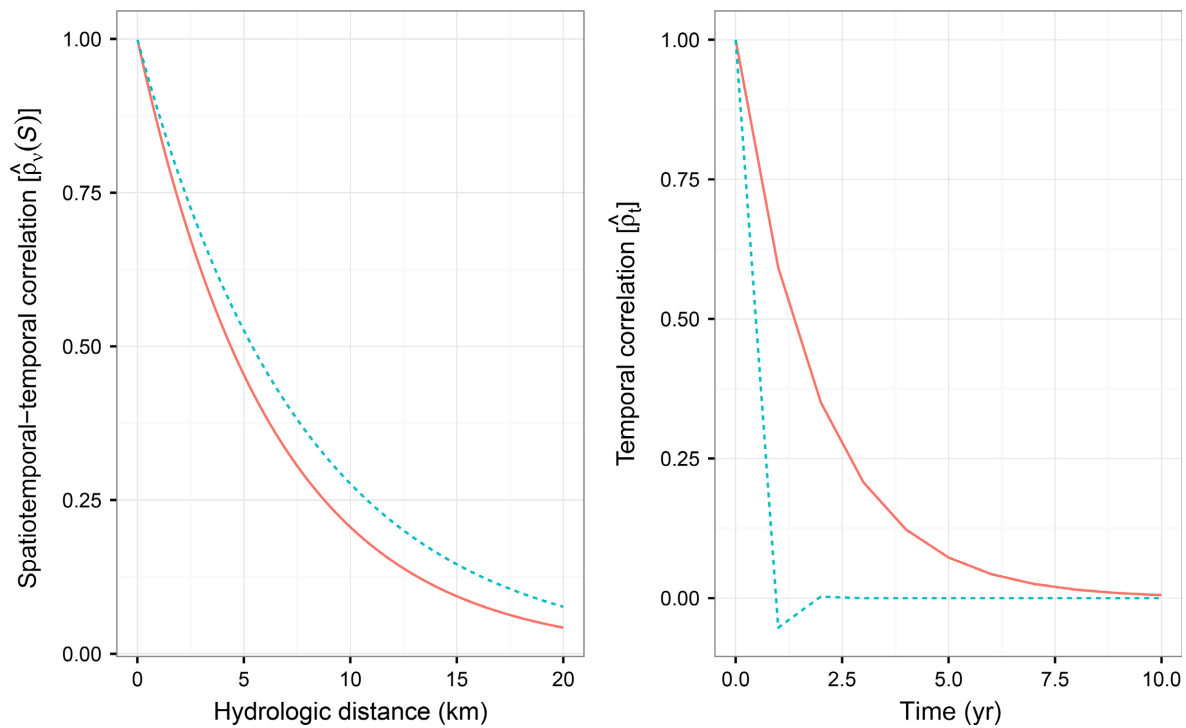


FIG. 6. Spatiotemporal and temporal decay curves with distance for adult and young of the year (YOY) brook trout in the West Susquehanna watershed for the model including temporal and spatiotemporal components. The spatiotemporal correlation, $\rho_v(s)$, is the expected correlation between parent and child nodes for a given distance (Eq. 11) Adults are represented by the solid red line and YOY by the dashed blue line.

indicating some temporal stability in local habitat quality or preference). The limited brook trout dispersal in headwater streams and the fine-scale variation in habitat quality combine to create pockets of high abundances. For example, stream temperatures can vary dramatically at very small spatial scales due to local ground water input (Snyder et al. 2015) and topography and wood in streams can create small pools (Bisson et al. 1987). Both cool water reaches in the summer and pools generally create high-quality habitat for brook trout.

In summary, we demonstrated good recovery of spatial and temporal components and good accuracy in estimating local fish densities across a stream network. Our model can be used to improve precision when estimating local densities in a network compared with traditional nonspatial models while providing additional information about the spatiotemporal population dynamics of these organisms. Given that the spatial model always performed as well or better than the nonspatial model, we recommend our approach for analysis of data even when there is previous indication of slight spatial correlations

ACKNOWLEDGMENTS

We thank E. Childress for discussions related to the use of data and inference from the models, and J. Ver Hoef, J. Hastie, M. McClure, and two anonymous reviewers for comments that improved the manuscript. D. Hocking was partially supported by a USGS Mendenhall Fellowship and funding through the DOI Northeast Climate Science Center aided this project. B. Letcher, D. Hocking, and J. Thorson conceived of the concept and J. Thorson developed the statistical framework. K. O'Neil conducted the stream network and GIS analysis. D. Hocking and J. Thorson conducted the simulations and statistical analyses. All authors contributed to writing and editing of the manuscript and approved of the final version.

LITERATURE CITED

- Bassar, R. D., B. H. Letcher, K. H. Nislow, and A. R. Whiteley. 2016. Changes in seasonal climate outpace compensatory density-dependence in eastern brook trout. *Global Change Biology* 22:577–593.
- Bisson, P. A., R. E. Bilby, M. D. Bryant, C. A. Dolloff, G. B. Grette, R. A. House, M. L. Murphy, K. V. Koski, and J. R. Sedell. 1987. Large woody debris in forested streams in the Pacific Northwest: past, present, and future. Pages 143–190 in E. O. Salo and T. W. Cundy, editors. *Streamside management: forestry and fishery interactions*. University of Washington, Seattle, Washington, USA.
- Burnham, K. P. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* 33:261–304.
- Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2010. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65:23–35.
- Caissie, D. 2006. The thermal regime of rivers: a review. *Freshwater Biology* 51:1389–1406.
- Carline, R. F., and B. J. McCullough. 2003. Effects of floods on brook trout populations in the Monongahela National Forest, West Virginia. *Transactions of the American Fisheries Society* 132:1014–1020.
- Carlson, S. M., and B. H. Letcher. 2003. Variation in brook and brown trout survival within and among seasons, species, and age classes. *Journal of Fish Biology* 63:780–794.
- Conn, P. B., D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, and J. M. London. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. *Ecological Monographs* 85:235–252.
- Dail, D., and L. Madsen. 2012. Estimating open population site occupancy from presence-absence data lacking the robust design. *Biometrics* 69:146–156.
- DeWeber, J. T., and T. Wagner. 2014. Predicting brook trout occurrence in stream reaches throughout their native range in the Eastern United States. *Transactions of the American Fisheries Society* 144:11–24.
- Donnell, D. O., A. Rushworth, A. W. Bowman, and E. S. Marian. 2014. Flexible regression models over river networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 63: 47–63.
- Dormann, C. F., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628.
- George, S. D., B. P. Baldigo, A. J. Smith, and G. R. Robinson. 2015. Effects of extreme floods on trout populations and fish communities in a Catskill Mountain river. *Freshwater Biology* 60:2511–2522.
- Grant, E. H. C., W. H. Lowe, and W. F. Fagan. 2007. Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters* 10:165–175.
- Grossman, G. D., A. Nuhfer, T. Zorn, G. Sundin, and G. Alexander. 2012. Population regulation of Brook Trout (*Salvelinus fontinalis*) in Hunt Creek, Michigan: a 50-year study. *Freshwater Biology* 57:1434–1448.
- Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2:e616.
- Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae. 2013. Global flood risk under climate change. *Nature Climate Change* 3: 816–821.
- Hocking, D. J., K. J. Babbitt, and M. Yamasaki. 2013. Comparison of silvicultural and natural disturbance effects on terrestrial salamanders in northern hardwood forests. *Biological Conservation* 167:194–202.
- Homer, C., J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown. 2015. Completion of the 2011 national land cover database for the conterminous United States—representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81:345–354.
- Hudy, M., T. M. Thieling, N. Gillespie, and E. P. Smith. 2008. Distribution, status, and land use characteristics of subwatersheds within the native range of Brook Trout in the eastern United States. *North American Journal of Fisheries Management* 28:1069–1085.
- Hunter, C. M., H. Caswell. 2009. Rank and redundancy of multi-state mark-recapture models for seabird populations with unobservable states. Pages 797–825 in D. L. Thomson, E. G. Cooch and M. J. Conroy, editors. *Modeling demographic processes in marked populations*. Springer US, Boston, Massachusetts, USA.
- Huntsman, B. M., and J. T. Petty. 2014. Density-dependent regulation of brook trout population dynamics along a core-periphery distribution gradient in a central Appalachian watershed. *PLoS ONE* 9:e91673.
- Isaak, D. J., J. M. Ver Hoef, E. E. Peterson, D. L. Horan, and D. E. Nagel. 2017. Scalable population estimates using spatial-stream-network (SSN) models, fish density surveys, and national geospatial database frameworks for streams. *Canadian Journal of Fisheries and Aquatic Sciences* 74:147–156.
- Isaak, D. J., et al. 2014. Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water* 1:277–294.
- Kanno, Y., B. H. Letcher, N. P. Hitt, D. A. Boughton, J. E. B. Wofford, and E. F. Zipkin. 2015. Seasonal weather patterns drive population vital rates and persistence in a stream fish. *Global Change Biology* 21:1856–1870.

- Kanno, Y., K. C. Pregler, N. P. Hitt, B. H. Letcher, D. J. Hocking, and J. E. B. Wofford. 2016. Seasonal temperature and precipitation regulate brook trout young-of-the-year abundance and population dynamics. *Freshwater Biology* 61:88–99.
- Kovach, R. P., C. C. Muhlfeld, R. Al-chokhachy, J. B. Dunham, B. H. Letcher, and J. L. Kershner. 2016. Impacts of climatic variation on trout: a global synthesis and path forward. *Reviews in Fish Biology and Fisheries* 26:135–151.
- Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. 2016. TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software* 70:1–21.
- Letcher, B. H., and T. D. Terrick. 1998. Maturation of male age-0 Atlantic salmon following a massive, localized flood. *Journal of Fish Biology* 53:1243–1252.
- Letcher, B. H., et al. 2015. Robust estimates of environmental effects on population vital rates: an integrated capture–recapture model of seasonal brook trout growth, survival and movement in a stream network. *Journal of Animal Ecology* 84:337–352.
- Milanovich, J. R., D. J. Hocking, W. E. Peterman, and J. A. Crawford. 2015. Effective use of trails for assessing terrestrial salamander abundance and detection: a case study at Great Smoky Mountains national park effective use of trails for assessing terrestrial salamander abundance and detection: a case study at great smoky M. *Natural Areas Journal* 35:590–598.
- Nislow, K. H., F. J. Magilligan, C. L. Folt, and B. P. Kennedy. 2002. Within-basin variation in the short-term effects of a major flood on stream fishes and invertebrates. *Journal of Freshwater Ecology* 17:305–318.
- Peterman, W. E., and R. D. Semlitsch. 2013. Fine-scale habitat associations of a terrestrial salamander: the role of environmental gradients and implications for population dynamics. *PLoS ONE* 8: e62184.
- Peterman, W. E., and R. D. Semlitsch. 2014. Spatial variation in water loss predicts terrestrial salamander distribution and population dynamics. *Oecologia* 176:357–369.
- Peterson, E. E., D. M. Theobald, and J. M. Ver Hoef. 2007. Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology* 52:267–279.
- Peterson, E. E., and J. M. Ver Hoef. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology* 91:644–651.
- Peterson, E. E., et al. 2013. Modelling dendritic ecological networks in space: an integrated network perspective. *Ecology Letters* 16:707–719.
- R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org
- Robinson, J. M., D. C. Josephson, B. C. Weidel, and C. E. Kraft. 2010. Influence of variable interannual summer water temperatures on Brook Trout growth, consumption, reproduction, and mortality in an unstratified Adirondack lake. *Transactions of the American Fisheries Society* 139:685–699.
- Roghair, C. N., C. A. Dolloff, and M. K. Underwood. 2002. Response of a Brook Trout population and instream habitat to a catastrophic flood and debris flow. *Transactions of the American Fisheries Society* 131:718–730.
- Ross, B. E., M. B. Hooten, and D. N. Koons. 2012. An accessible method for implementing hierarchical models with spatio-temporal abundance data. *PLoS ONE* 7:e49395.
- Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, Boston, Massachusetts, USA.
- Royle, J. A., and C. K. Wikle. 2005. Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics* 12:225–243.
- Snyder, C. D., N. P. Hitt, and J. A. Young. 2015. Accounting for groundwater in stream fish thermal habitat responses to climate change. *Ecological Applications* 25:1397–1419.
- Soil Survey Staff. 2015. Soil Survey Geographic Database (SSURGO). <https://websoilsurvey.nrcs.usda.gov/>
- Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* 190:214–251.
- Thornton, P. E., M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook. 2016. Daymet: daily surface weather data on a 1-km grid for North America, version 3. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.
- Thorson, J. T., H. J. Skaug, K. Kristensen, A. O. Shelton, E. J. Ward, J. H. Harms, and J. A. Benante. 2014. The importance of spatial models for estimating the strength of density dependence. *Ecology* 96:1202–1212.
- Ver Hoef, J. M., and E. E. Peterson. 2010. A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association* 105:6–18.
- Ver Hoef, J. M., E. Peterson, and D. Theobald. 2006. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* 13:449–464.
- Vincenzi, S., A. J. Crivelli, W. H. Satterthwaite, and M. Mangel. 2014. Eco-evolutionary dynamics induced by massive mortality events. *Journal of Fish Biology* 85:8–30.
- Wagner, T., J. T. Deweber, J. Detar, D. Kristine, and J. A. Sweka. 2014. Spatial and temporal dynamics in brook trout density: implications for population monitoring. *North American Journal of Fisheries Management* 34:258–269.
- Whiteley, A. R., J. A. Coombs, M. Hudy, Z. Robinson, A. R. Colton, K. H. Nislow, and B. H. Letcher. 2013. Fragmentation and patch size shape genetic structure of brook trout populations. *Canadian Journal of Fisheries and Aquatic Sciences* 70:678–688.
- Xu, C. L., B. H. Letcher, and K. H. Nislow. 2010. Size-dependent survival of brook trout *Salvelinus fontinalis* in summer: effects of water temperature and stream flow. *Journal of Fish Biology* 76:2342–2369.
- Zipkin, E. F., J. T. Thorson, K. See, H. J. Lynch, E. H. C. Grant, Y. Kanno, R. B. Chandler, B. H. Letcher, and J. A. Royle. 2014. Modeling structured population dynamics using data from unmarked individuals. *Ecology* 95:22–29.

SUPPORTING INFORMATION

Additional supporting information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/eap.1767/full>

DATA AVAILABILITY

GIS and stream network data were archived and publicly available with supporting information with the Spatial Hydro-Ecological Decision System (SHEDS) at <http://ecosheds.org>. Trout data were kindly provided by the Pennsylvania Fish and Boat Commission and are available upon request. Please contact Jason Detar (PFBC; jdetar@state.pa.us) for details. Simulation data used in this manuscript were generated using custom R functions. All functions and scripts to allow for replication are available online at https://github.com/djhocking/Trout_GRF