*Journal:* Methods in Eco Evo [6000-7000 words including captions and references!]

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

Daniel J. Hocking[[1]](#footnote-2), *Frostburg State University, Department of Biology, Frostburg, MD 21532.*

James T. Thorson*, Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, 98112.*

Kyle O’Neil, *US Geological Survey, Leetown Science Center, S.O. Conte Anadromous Fish Research Laboratory, One Migratory Way, Turners Falls, MA 01376.*

Benjamin H. Letcher, *US Geological Survey, Leetown Science Center, S.O. Conte Anadromous Fish Research Laboratory, One Migratory Way, Turners Falls, MA 01376.*

**Running Title:** Spatio-temporal model for abundance in streams

**Word Count:**

Disclaimer: This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the US Geological Survey (USGS), it does not represent any official finding or policy.

**Abstract**

Fish and wildlife populations will have dynamics that are 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 under-estimated. We developed a novel statistical method to account for spatio-temporal correlations within dendritic stream networks, while accounting for imperfect detection in the surveys. Through simulations, we found this model outperformed 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 years. Increasing the number of survey sites within the network improved the performance of the non-spatial model but only marginally improved the density estimates in the spatio-temporal model. We applied this model to Brook Trout data from the West Susquehanna Watershed in Pennsylvania collected over 34 years from 1981 - 2014. We found the model including temporal and spatio-temporal 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 spatio-temporal correlation. Adult densities were less strongly affected by climatic conditions and less temporally variable than YOY but with similar spatio-temporal correlation and higher temporal autocorrelation.

**Keywords:** Spatially-explicit, spatio-temporal, dendritic network, Gaussian random fields, Brook Trout, detection probability

**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 spatio-temporal 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 (e.g. ANOVA, GLM, linear and generalized linear mixed models). 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, Onn 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 is predictably correlated along the network and it important to account for 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 (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 et al. 2010, Ver Hoef and Peterson 2010). Additionally, block Kriging has been used for spatial interpolation (Ver Hoef et al. 2006).

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 less than one (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 (Zipkin et al. n.d., e.g. Royle 2004, Royle and Dorazio 2008, Dail and Madsen 2012). 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 (Peterman and Semlitsch 2013, Hocking et al. 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 below. 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 U.S. 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). 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 that 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 density at time and site (numbers per 100 meter stream reach, i.e., where distances are measured along a one-dimensional stream reach) as a log-linked linear regression model with components representing the effect of measured habitat variables, as well as otherwise unexplained spatial, temporal, spatio-temporal, and independent variation. Although some measured independent variables [ 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, 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 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 to account for potential spatio-temporal 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 as overdispersion that is independent among sites and years (i.e., the nugget in a geostatistical model). The overdispersion parameter is assumed to be normally distributed among sites with an independent and identically distributed variance parameter []. We therefore specify:

(1)

where is a row-vector of measured variables affecting density (which includes an intercept term) and is the estimated impact of these variables on log-density (fixed-effed regression coefficients). Additional descriptions of all model parameters are found in Table 1.

To account for imperfect detection while sampling, we modeled counts for depletion pass (), site and year 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:

(2a)

where is the offset for length of stream sampled by observation (length of survey / 100 m) so all abundances are relative to fish per 100 m of stream length, and is the probability that each individual present at site and time for observation will be captured (this probability potentially varies among sites and years). Counts in the second and third passes are then dependent upon not being captured in the earlier passes:

(2b)

and

(2c)

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

(2d)

where Eq. 2d represents an inverse complementary-log-log (“cloglog”) link function for detection probability, given parameter representing average log-detection probability, and unexplained variation, where is an estimated parameter governing the magnitude of variation in log-detectability among sites and years. Detectability parameters (, , and ) are estimated simultaneously with parameters representing spatial and spatio-temporal variation in density . Refer to Table 1 for summary descriptions of all model parameters.

*Spatio-temporal 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 and spatio-temporal variation ) 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”, i.e., the value of a variable 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 (conditional on the value of at all other sites). This property arises from the assumption that the value of 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 each site as a function of its value at the parent node for that site , and the distance between and its parent . This allows us to factor the joint probability of a spatial variable into a series of easy-to-calculate conditional probabilities:

(3)

We further assume that variation in 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.

*Ornstein-Uhlenbeck process for spatial variation*

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

(4)

The variance, for site conditional on the value of its parent given an OU process is

(5)

where is the exponential rate of decay in correlation between child and parent nodes with distance and larger values represent less correlation, is the distance between parent and child nodes, and 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 where

(6)

Eq. 5-6 are specified such that the pointwise variance of (i.e., the variance was drawn again from the same stochastic process) is .

*First-order autocorrelation for temporal variation*

We include a temporal term for each year to represent years that are better or worse on average across all sites. We model vector (representing in all years) using first-order autocorrelation

(7)

where is the variance in this temporal term, and is the correlation matrix for a first-order autocorrelation process:

(8)

where is the correlation between years and , separated by years, and is an estimated parameter representing the correlation in for two adjacent years.

*Ornstein-Uhlenbeck process for spatio-temporal variation*

We similarly used the OU process to represent the spatio-temporal relationships along the network. We use the vector to represent the spatio-temporal term for all years , and it varies along the network as an OU process:

(9)

where is again the spatio-temporal variance for site given the vector for its parent :

(10)

is the correlation due to spatial similarity:

(11)

and is the correlation due to temporal similarity, which we assume follows first-order autocorrelation (eq. Eq. 8, but replacing with , where is an estimated parameter representing the temporal correlation between two adjacent years in spatio-temporal variation ). We assumed that the decorrelation distance was identical for spatial and spatio-temporal variation (i.e., ), 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 () as well as overdispersion (), temporal (), spatial (), and spatio-temporal () 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. Parameter estimation is conducted using Template Model Builder (Kristensen et al. 2016) within the R statistical platform (citation).

**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 non-spatial model. To accomplish this, we simulated data with all combinations of in {0.1, 0.5, 1, 2, 3} and in {0.1, 0.25, 0.5, 0.75, 1.0}. We used the standard deviations (e.g. ) rather than the variance (e.g. when specifying the models for convenience. These values of represent a large range in correlations such that when = 0.1 then (90.5% correlation) whereas when then (5.0% correlation) at points 1 km apart. We also included a single covariate on abundance that differed by location but was not spatially autocorrelated We ran 200 simulations for each combination of and and fit each simulated dataset with the spatial model described (single year with no temporal or spatiotemporal variation) and with a non-spatial model. Both models accounted for imperfect detection, observed following a 3-pass depletion survey and were identical except for the inclusion of the spatial component of the model. The probability of detecting an individual that remained in the stream on any given pass was 50%. 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 km to 5.13 km with a mean of 1.13 km. The R code for simulating the data can be found at <https://github.com/djhocking/Trout_GRF/blob/master/Code/Spatial_Simulations.R>

**Spatiotemporal Power Analysis**

We also wanted to understand the effect of spatial and temporal replication on model performance. We simulated 200 independent data sets for the White River in Vermont over 20 years 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 years 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 , , , , , , detection probability , and abundance coefficients, where the first value in is the log-mean intercept and the second value is the coefficient (slope) of a site-level covariate.

**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 2015, 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, PA 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 U.S. 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 years from 1981 and 2014. There was a total of 683 site visits with a mean of 2.0 and a range of 1 to 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 forest (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; ref). Forest cover data was 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 can be found at <http://conte-ecology.github.io/shedsGisData/>. The covariate summary statistics for the West Susquehanna watershed are presented in Table 2.

We used the National Hydrography Dataset high-resolution flowlines (<https://nhd.usgs.gov/index.html>) truncated to >0.75 km2 drainage area for spatial consistency and exclusion of highly ephemeral streams. 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. The full description of the process, scripts, and links to the hydrography data is archived at (<http://conte-ecology.github.io/shedsGisData/>). The hydrography for the region from Maine to Virginia, USA can be downloaded by hydrologic unit code 2 at <http://ecosheds.org/assets/nhdhrd/v2/>.

*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 x 2 x 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**

We found that the spatial model estimated the spatial correlation decay rate () well when there was strong spatial correlation but tended to slightly underestimate the decay rate when the correlation was low ( large; Figure 2a). Similarly, the variability in the estimation of was larger as the value of increased (Figure 2a). The spatial model estimated the mean abundance across the watershed much better than the non-spatial model (Figure 2b) and the mean uncertainty of the overall abundance was relatively similar among models, but the variation in this uncertainty was much larger for the non-spatial model compared with the spatial model when the spatial correlation was large (Figure 2c). The root mean squared error (RMSE) was far larger for the non-spatial model compared with the spatial model across all values of (Figure 2d), 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 () was estimated well across all values of , but the variation in this estimate was slightly smaller for the spatial model, especially at higher levels of spatial correlation (Figure 2e).

The range of also significantly influenced the parameter estimates and the differences between spatial and non-spatial models. At all levels, was recovered well using the spatial model with very slight underestimation on average (Figure 3a). The spatial and non-spatial models performed similarly in the estimation of mean abundance across the watershed when the true value of was small but the spatial model was more accurate and more precise compared with the non-spatial model as the level of increased (Figure 3b). The uncertainty in mean network abundance went up for the non-spatial model as increased but was constant for the spatial model across levels of (Figure 3c). The RMSE was again much smaller for the spatial model compared with the non-spatial model as increased. The variability in the RMSE also increased greatly for the non-spatial model as increased (Figure 3d). The fixed effect coefficient was estimated well for both models but the uncertainty increased in the non-spatial model as increased (Figure 3e).

**Spatiotemporal Power Analysis**

We found the mean network abundance was estimated fairly well for both the spatial and non-spatial models, but the spatial model tended to underestimate abundance slightly when few years were surveyed (Figure 4). However, the RMSE was lower for the spatial model compared with the non-spatial model and the difference increased with the number of years surveyed. The value of was recovered well regardless of the number of years sampled, but the imprecision when estimating decreased (improved) with the number of years sampled (Figure 4) and even more dramatically with the number of sites sampled (Figure 5). The variance in the spatial process ( was underestimated and the spatiotemporal variance () overestimated when sites were sampled for 10 or fewer years (Figure 4). Similarly, for both the spatial and non-spatial model it took 10-15 years to accurately recover the temporal autocorrelation, although the variability in the temporal process was recovered with approximately eight years of sampling in the non-spatial model and possibly as little as four years with the spatial model (Figure 5). The value of the fixed effect covariate, , 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 spatiotemporal model (Figure 4).

The number of sites sampled similarly influenced the estimation of the spatial and spatiotemporal variance terms, with an increasing number of sites improving the spatial variance estimation and to a lesser extent, the spatiotemporal variance (Figure 5). The RMSE of the spatial model improved with an increasing number of sites sampled but it did not significantly improve for the non-spatial model (Figure 5). The fixed effect coefficient was recovered well for both models and the precision improved with the number of sites sampled. Despite reasonable estimates of mean abundance and fixed effects in many simulations, the non-spatial model (Model 3 in Table 3), generally did not sufficiently recover the heterogeneity and spatial pattern in density as seen in Figure 6.

**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 4). For adult Brook Trout, the spatiotemporal model and the temporal plus spatiotemporal model were the top two models with a delta AIC 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 spatio-temporal 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 spatio-temporal model parameters along with the fixed effects, detection probabilities, and overdispersion terms. Adults exhibited strong temporal autocorrelation (= 0.59) with low variability (= 0.16), whereas YOY exhibited no temporal autocorrelation (= -0.05) but high stochastic temporal variability (= 0.76). The estimated values of the spatio-temporal decay were at the lower end of what we tested with simulations for both YOY (0.13) and adults (0.16), indicating high spatio-temporal correlation (~50% at 5 km; Figure 7). The estimates of the spatio-temporal standard deviation were high for YOY (0.65) and adults (0.59). The combination of the two parameters indicate extremely high spatiotemporal autocorrelation, which is revealed by the very high estimate of temporal decay of 0.98 and 0.97 for YOY and adults, respectively (Table 6). 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).

[I recommend adding 2-3 sentence paragraph summarizing that figure of total abundance for YOY and adult, along with your point about no stock-recruit relationship but a YOY-adult relationship being visible]

**Discussion**

We have 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 (Figure 2). Even when the spatial decay rate ( 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 non-spatial model for estimating total abundance.

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 (Figure 4; RMSE). However, there was a large improvement in recovery of the spatial and spatio-temporal components of the model with between 10 and 15 years of data. Although there is moderately-high uncertainty in the estimation of the spatial and spatio-temporal variances (in Figure 4), 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 spatio-temporal components was likely inflated (in Figure 5) because of combining simulation uncertainty with variation in the number of years surveyed while only holding the number of sites constant. 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 years of data for 100 sites. 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 spatio-temporal 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.

Although the non-spatial model was good at recovering the mean network density, it still had higher (worse) RMSE than the spatial model (Figures 5 & 6). This indicates that the non-spatial model does well at estimating average density but is not as good at estimating density at individual locations (also seen in Figure 6). 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 spatio-temporal model can readily be applied to existing standard electrofishing data from state and federal agencies. Using this model with brook trout data collected the Pennsylvania Boat and Fish Commission, we demonstrated improved model fit compared with basic non-spatial models even accounting for increased model complexity (i.e. using AIC). For adult brook trout, the spatio-temporal model and the model with temporal and spatio-temporal components outperformed all other models (Table 4). Similarly, the temporal plus spatio-temporal 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.

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 range-wide 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, reflecting 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. (2015) 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, non-porous sites. It is also likely that 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 (Roghair et al. 2002, Nislow et al. 2002, George et al. 2015). It is possible that this model could be used to assess the spatial and spatio-temporal 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 … Both adults and YOY densities exhibited similar levels of spatio-temporal correlation with relatively slow decorrelation with distance as evidenced by the low spatio-temporal decay rates (0.16, and 0.13, respectively) and high asymptotic spatio-temporal variances (Table 6). The effect of these parameters can be seen in Figure 7, 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. 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 spatio-temporal correlation per year indicates a slow rate of change in the spatial patterning (i.e. high densities sites tended to maintain relatively high densities, indicating some temporal stability in local habitat quality or preference). [relate to other research].

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 non-spatial models while providing additional information about the spatio-temporal population dynamics of these organisms. Given that the spatial model always performed as well or better than the non-spatial model, we recommend our approach for analysis of data even when there is previous indication of slight spatial correlations

**Acknowledgements**

We thank the Pennsylvania Boat and Fish Commission for providing access to fish survey data and E. Childress for discussions related to the use of data and inference from the models. DJH was supported by a USGS Mendenhall Fellowship and funding through the DOI Northeast Climate Science Center during this project.

**Data Accessibility**

To enable readers to locate archived data, authors should list the database and the respective accession numbers or DOIs for all data from the manuscript that has been made publicly available in this section. E.g:  
- Species descriptions: uploaded as online supporting information- Phylogenetic data: TreeBASE Study accession no. Sxxxx- R scripts: uploaded as online supporting information- Sample locations, IMa2 input files and microsatellite data: DRYAD entry doi: xx.xxxx/dryad.xxxx

**References**

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.

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.

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. 2015. 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.

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.

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.

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, D. Theobald, J. M. Ver Hoef, E. Peterson, and D. Theobald. 2006. Spatial statistical models that use flow and stream distance. Environmental and Ecological Statistics 13:449–464.

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.

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., E. E. Peterson, J. M. Ver Hoef, S. J. Wenger, J. A. Falke, C. E. Torgersen, C. Sowder, S. E. Ashley, M. Fortin, C. E. Jordan, A. S. Ruesch, N. Som, and P. Monestiez. 2014. Applications of spatial statistical network models to stream data. WIREs Water:1–18.

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.

Letcher, B. H., P. Schueller, R. D. Bassar, K. H. Nislow, A. Coombs, K. Sakrejda, M. Morrissey, D. B. Sigourney, R. Whiteley, M. J. O. Donnell, and T. L. Dubreuil. 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.

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.

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.

Onn, P. A. U. L. B. C., D. E. S. J. Ohnson, J. A. Y. M. V. E. R. H. Oef, M. E. B. H. Ooten, and J. O. M. L. Ondon. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235–252.

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.

Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. a Falke, M.-J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, A. Sengupta, N. Som, E. A. Steel, D. M. Theobald, C. E. Torgersen, and S. J. Wenger. 2013. Modelling dendritic ecological networks in space: an integrated network perspective. Ecology letters 16:707–19.

Peterson, E. E., J. M. Ver Hoef, and M. Jay. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. Ecology 91:644–651.

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.

Royle, J. A., and C. K. Wikle. 2005. Efficient statistical mapping of avian count data. Environmental and Ec 12:225–243.

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. ORNL DAAC, 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.

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.

Zipkin, E.F., Thorson, J.T., See, K., Lynch, H.J., Grant, E.H.C., Kanno, Y., Chandler, R.B., Letcher, B.H. & Royle, J.A. (2014). Modeling structured population dynamics using data from unmarked individuals. Ecology, 95, 22–29.

**Tables**

**tables and figures**

Supporting Information and Appendices

1. Corresponding Author:Frostburg State University, Department of Biology, 201 Compton Science Center, 101 Braddock Road, Frostburg, MD 21532. Email - [*djhocking@frostburg.edu*](mailto:djhocking@frostburg.edu)*.. Phone: 301-687-4343.* [↑](#footnote-ref-2)