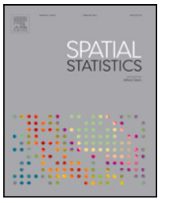
Spatial Statistics 67 (2025) 100902

Contents lists available at ScienceDirect

Spatial Statistics

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



Stochastic spatial stream networks for scalable inferences of riverscape processes

Xinyi Lu [a,*], Andee Kaplan [b], Yoichiro Kanno [c], George Valentine [c,d], Jacob M. Rash [e], Mevin Hooten [f]

[a] Mathematics and Statistics Department, Utah State University, Logan, UT, 84404, USA
[b] Department of Statistics, Colorado State University, Fort Collins, CO, 80523, USA
[c] Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, 80523, USA
[d] US Forest Service Rocky Mountain Research Station, Fort Collins, CO, 80526, USA
[e] North Carolina Wildlife Resources Commission, Marion, NC, 28752, USA
[f] Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX, 78712, USA

ARTICLE INFO

Dataset link: https://www.sciencebase.gov/catalog/item/5f62407d82ce38aaa236148b

Keywords:
Bayesian hierarchical models
Brook trout
Markov random field
Population models
Space–time dynamics

ABSTRACT

Spatial stream networks (SSN) models characterize correlated ecological processes in dendritic ecosystems. Conventional SSN models rely on pre-processed stream networks and point-to-point hydrologic distances. However, this data processing may be labor-intensive and time-consuming over large spatial domains. Therefore, we propose to infer the functional connectivity of stream networks stochastically. Our physically-guided model utilizes the knowledge that water flows from high elevation to low elevation, and flow rate typically increases when two tributaries merge. We also leverage the hierarchical branching architecture of dendritic networks to alleviate computing and reduce uncertainty. Spatial autoregressive models composed of inferred SSNs propagate stochasticity between network connectivity and dynamic ecological processes in a Bayesian framework. We show in simulated examples that our mechanistic model facilitated learning about the functional network and enhanced predictive performance. We also demonstrate our approach in a large-scale case study using native brook trout (*Salvelinus fontinalis*) count data. A population model based on our stochastic SSN outperformed that with a conventional SSN in predicting abundance and expedited the analysis by circumventing data processing.

1. Introduction

Ecological statistics often involves characterizing spatial heterogeneity and dependence of ecological processes (Gelfand et al., 2010; Cressie, 2015). The modeling of processes occurring over complex spatial domains benefits from representing the domain features explicitly (Som et al., 2014; Zimmerman and Ver Hoef, 2024). For example, physical, chemical, and biological dispersal processes along stream networks are governed by watershed configuration and flow connectivity (Cressie et al., 2006; Ver Hoef et al., 2006). Ver Hoef and Peterson (2010) proposed a class of spatial stream network (SSN) models that induce autocorrelation by integrating moving-average functions along stream segments. The SSN models produce statistically valid covariance functions, and have been illustrated in various applications including water temperature (Santos-Fernandez et al., 2022), water quality (Ver Hoef and Peterson, 2010; Scown et al., 2017; McManus et al., 2020), ecosystem metabolism (Rodríguez-Castillo et al., 2019), species

* Corresponding author.
E-mail address: lucy.lu@usu.edu (X. Lu).

https://doi.org/10.1016/j.spasta.2025.100902
Received 29 October 2024; Received in revised form 14 February 2025; Accepted 8 April 2025
Available online 25 April 2025

richness (Peterson and Ver Hoef, 2010; Rodríguez-González et al., 2019), and population abundance (Wenger et al., 2011; Isaak et al., 2017). The popularity of SSN models is facilitated by **R** packages `SSN` (Ver Hoef et al., 2014) and `SSN2` (Dumelle et al., 2024) that streamline model fitting, parameter inference, and spatial prediction for an array of customization options. Fitting SSN models using these packages relies on the Spatial Tools for the Analysis of River Systems (STARS) toolset (Peterson and Ver Hoef, 2014) for ArcGIS (ESRI, 2011) or the `openSTARS` (Kattwinkel et al., 2020) and `SSNbler` (Peterson et al., 2024) **R** packages to process stream networks from geospatial hydrologic databases (e.g., the National Hydrologic Dataset (U.S. Geological Survey, 2016), National Stream Internet Project (Nagel et al., 2015), StreamCat (Hill et al., 2016)). Although rarely discussed in literature, pre-processing SSNs from these databases is labor-intensive and time-consuming over large spatial domains. This is because site coordinates are not always co-located with stream segments in the databases and manual correction is recommended (Peterson et al., 2024). Furthermore, the software programs calculate point-to-point hydrologic distances by traversing the entire network, which often contains more segments than those involved in the study. For example in Lu et al. (2024), count data were collected from 109 streams in western North Carolina to model native brook trout populations, whereas over two thousand streams in the study domain required manual processing to derive the network information for their SSN model. To address the challenges in data pre-processing, we propose to infer the functional connectivity of stream segments probabilistically.

A segment is a section of a stream between two confluences, and a network is a set of hydrologically connected segments. In a physical network, two segments are connected if they have a common confluence. However, our stochastic SSN utilizes only observed segments, and therefore we consider two segments functionally connected if data demonstrate spatial correlation. Two segments that are physically separated may be functionally connected in a stochastic network due to indirect merging (i.e., two observed tributaries flowing into an unobserved main stem at different confluences). Alternatively, two segments that are physically connected may be functionally separated due to disrupted dispersal (i.e., fish movement from main stem to observed tributary impeded by unknown obstacles.). We infer functional connectivity using the unique network architecture of SSNs and laws of gravity. Dendritic networks (e.g., rivers, caves, hedgerows) are represented by a hierarchical structure where blind-ended branches are connected to main stems (Campbell Grant et al., 2007; Brown and Swan, 2010). They are related to tree networks in computer science, whose topology is characterized by the branching and the number of generations (Meador, 2008). Because of the branching hierarchy, we can reconstruct the network by knowing the stem to which each branch is directly connected (Estrada et al., 2014), and because gravity dictates that water flows from high elevation to low elevation and typically increases in flow rate downstream, we can eliminate branching that is physically infeasible.

Conventional SSN models resemble geostatistical models in which covariance is defined continuously over the study domain (Ver Hoef and Peterson, 2010). Our stochastic SSN model, in contrast, represents covariance in areal data where the observation units are segments (Banerjee et al., 2003). Spatial covariance in areal data has been represented in various ways. In a Markov random field (MRF), a direct representation of the covariance structure is also known as a second-order specification, and is commonly achieved using conditional (CAR) or simultaneous autoregressive (SAR) models (Ver Hoef et al., 2018). Alternatively, the second-order structure may be induced by a first-order specification using basis functions (Hefley et al., 2017). Outside of MRFs, spatial dependence may arise mechanistically, for example, following stochastic differential and partial differential equations (Hanks, 2017; Lu et al., 2020; Wright et al., 2022). Nonetheless, most of these methods rely on observed and fixed spatial configurations. On the other hand where spatial configurations are considered random, the inference is often focused on configurations instead of first-order dynamics. For example, statistical network analysis aims at learning local features that influence network connectivity, and the outcomes of these analyses are typically pair-wise connections or similarity measurements (Krackhardt, 1988; Hanks and Hooten, 2013). Our stochastic SSN combines the strengths of MRF modeling and statistical network analysis under a Bayesian framework, where we specify covariance matrices using spatial autoregressive models conditioned on inferred functional connectivity. The cohesive framework allows us to infer both first-order covariate effects on abundance and physically-meaningful network characteristics, and enables uncertainty propagation between the two objectives.

We present our stochastic SSN and an associated population model in Section 2. In Section 3, we demonstrate the method using simulated examples and a case study of native brook trout (*Salvenilus fontinalis*) populations in western North Carolina, USA. Finally, in Section 4, we discuss possible extensions and broader applications of our method.

2. Method

We begin by demonstrating the general stochastic SSN modeling framework which we use to infer connectivity matrices representing spatial dependence among observed segments in a stream network. We then introduce a specific application where the stochastic-SSN-inferred connectivity matrices are used to construct covariance matrices for spatial random effects in a multi-scale spatio-temporal population model under a Bayesian hierarchical framework.

2.1. Stochastic SSN

We use terminology from graph theory (Biggs et al., 1986) and refer to stream segments as nodes. Our method aims to infer functional connectivity, or edges, between pairs of nodes, where we denote the upstream segment a child and the downstream segment its parent. We focus on the commonly observed branching mechanism of downstream merging, so that a child can have at most one direct parent, and a parent can have multiple direct children. We also denote the parent of a parent an ancestor, although knowledge of the full ancestry is not necessary for reconstructing the network. In the following model statements, we represent

scalars and indices with lowercase letters, vectors with bold lowercase letters, matrices and three-dimensional arrays with bold uppercase letters, and sets with script uppercase letters.

Every node in the network, $i = 1, \ldots, n$, is associated with elevation x_i^e and flow (stream discharge) x_i^f. Flow is measured at the segment level; however, elevation is measured at the site level and may vary along the segment. For the purpose of illustration, we use the mean elevation when there exist multiple sites on a segment. Because water flows by gravity and typically increases in volume downstream, we construct the $n \times n$ feasibility matrix, $\boldsymbol{F}$,

$$F_{ij} = \begin{cases} 1, \text{ if } i \neq j \text{ and } x_i^e \geq x_j^e \text{ and } x_i^f \leq x_j^f \\ 0, \text{ o.w.} \end{cases}. \tag{1}$$

Node j is a feasible parent to node i with non-zero probability if $F_{ij} = 1$, and node j is infeasible with probability one if $F_{ij} = 0$. The feasibility matrix is fully observed given elevation and flow, and informs the support for subsequent modeling of parenthood as multinomial random variables.

We denote $\mathcal{P}_i$ the set of feasible parents to node i with $F_{ij} = 1$. When $\mathcal{P}_i = \emptyset$, we define $\mathbb{P}(a_i = 0) = 1$, indicating that a_i is parentless. When $\mathcal{P}_i \neq \emptyset$, we let $\mathbb{P}(a_i = 0) = 1 - \xi$, representing the case where none of the feasible parents is a functional parent to i because the network is partially observed. With probability ξ, the functional parent a_i arises from $\mathcal{P}_i$. The model for a_i is summarized as follows,

$$a_i \sim \begin{cases} \delta(0), & \text{if } |\mathcal{P}_i| = 0 \\ \text{MN}(1; \mathcal{P}_i, \tilde{\boldsymbol{p}}_i), & \text{if } |\mathcal{P}_i| \geq 1 \end{cases}, \tag{2}$$

where $\tilde{\boldsymbol{p}}_i$ is the $(n+1)$-element probability vector, $\boldsymbol{p}_i$, scaled to sum to 1, and

$$p_{ij} = \begin{cases} 1 - \xi, & \text{if } j = 0 \text{ (i.e., parentless)} \\ \xi u(d_{ij}), & \text{if } j \in \mathcal{P}_i \\ 0, & \text{if } j \notin \mathcal{P}_i \end{cases}. \tag{3}$$

We let $u(d_{ij}) = \exp(-d_{ij}\tau)$, where d_{ij} is the geographic distance between the averaged site locations on segments i and j. In general, d_{ij} may be any measure of similarity between nodes i and j as in statistical network analyses, and τ represents the associated linear effect. In practice, we used geographic distance because it required no additional covariate information or pre-processing of hydrologic databases.

Conditioned on the parent vector, $\boldsymbol{a}$, we can construct the connectivity matrix, $\boldsymbol{C}$, in a variety of ways. The matrix $\boldsymbol{C}$ may be asymmetric for processes governed by flow direction, such as diffusion of chemicals. Alternatively, $\boldsymbol{C}$ may be symmetric for processes occurring in both directions, such as movement of stream fish. Furthermore, elements of $\boldsymbol{C}$ may be strictly binary to indicate direct parenthood; they may also be weighted to indicate generation distance. Even when an element of $\boldsymbol{C}$ is zero between a child and its ancestor, the induced correlation between them may still be non-zero, and the strength of which depends on the specific MRF representation (Ver Hoef et al., 2018). For illustration, we specify a symmetric, weighted connectivity matrix as follows

$$C_{ij} = C_{ji} = \begin{cases} 1/g_{ij}, \text{ if } j \text{ is an ancestor of } i \\ 0, \text{ o.w.} \end{cases}, \tag{4}$$

where g_{ij} is the generation distance between nodes i and j. For example, $g_{ij} = 1$ if j is the direct parent of i and $g_{ij} = 2$ if j is the grandparent of i.

2.2. Hierarchical modeling

We seek inference for connectivity matrices using our Bayesian hierarchical framework (Berliner, 1996) to characterize spatio-temporal population dynamics of a stream fish. In our hierarchical model statement, we use index i for site (instead of segment/node as in Section 2.1). We also index connectivity matrix by m (i.e., $\boldsymbol{C}_m$) to distinguish multiple stream networks. Count data were collected from surveys using two sampling protocols in our study. Depletion surveys were conducted at two-thirds of the study sites, where three consecutive samplings efforts (passes) were made at every survey, each pass removing samples from the enclosed segment. Single-pass surveys were conducted at the remaining one-third of the study sites, where one sampling effort (pass) was made at every survey. Following Lu et al. (2024), our data model accounts for imperfect detection in a modified N-mixture framework (Royle, 2004) as follows,

$$y_{i,t,j} \sim \begin{cases} \text{Binomial}\left(N_{i,t},\ q_{i,t}\right), & j = 1 \\ \text{Binomial}\left(N_{i,t} - \sum_{l=1}^{j-1} y_{i,t,l},\ q_{i,t}\right), & j > 1 \end{cases}. \tag{5}$$

Variable $y_{i,t,j}$ is the observed count at site i, $i = 1, \ldots, n$, in year t, $t = 1, \ldots, T$, and pass j, $j = 1, \ldots, J_i$; $N_{i,t}$ is true abundance; and capture probability $q_{i,t} = q_{\text{single}}$ if a single-pass survey occurred, and $q_{i,t} = q_{\text{depletion}}$ if a depletion survey occurred.

Our process model allows dispersion in true abundance conditioned on the associated mean density $\lambda_{i,t}$ and the sampling area A_i in 1000 m^2,

$$N_{i,t} \sim \text{Poisson}(A_i \lambda_{i,t}). \tag{6}$$

Dynamics in mean density are characterized using a log-linear function of fixed and random effects,

$$\log\left(\lambda_{i,t}\right) = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{h}_{i,t}'\boldsymbol{\theta}_i + \epsilon_{w_i} + \nu_{s_i} + \psi_{m_i} + \gamma_i \log t + \eta_t, \tag{7}$$

where $\boldsymbol{x}_i$ are landscape covariates, $\boldsymbol{h}_{i,t}$ are weather-related covariates, and $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{\text{suitable}}$ if site i was in a suitable habitat and $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{\text{unsuitable}}$ otherwise.

The multi-scale spatial random effects highlight the nested architecture of stream habitats, where ϵ_{w_i} denotes the random effect defined on the watershed that contains site i, ν_{s_i} denotes the random effect defined on site i along segment s, and ψ_{m_i} denotes the random effect defined on the segment that contains site i within network m. At the watershed level, we characterize spatial dependence using a CAR model as follows,

$$\boldsymbol{\epsilon} \sim \mathrm{N}\left(\mathbf{0}, \sigma_\epsilon^2(\mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \rho_\epsilon \boldsymbol{W})^{-1}\right), \tag{8}$$

where $\boldsymbol{W}$ is inversely weighted by geographic distances between watershed centroids and $\mathbf{1}$ is a column vector of 1. We let $\rho_\epsilon = 0.99$ to induce moderate correlation (Banerjee et al., 2003). At the segment level, we assume conditional independence among the segments and characterize dependence among sites on the same segment, $s = 1, \ldots, S$, using a geostatistical model as follows,

$$\boldsymbol{\nu}_s \sim \mathrm{N}\left(\mathbf{0}, \sigma_\nu^2 \boldsymbol{\Sigma}_s\right), \tag{9}$$

where $\Sigma_{s,i_1,i_2} = \exp(-\mathrm{dist}_{i_1,i_2}/\phi_\nu)$, dist_{i_1,i_2} is geographic distance between sites i_1 and i_2, and $\phi_\nu > 0$ represents the range of spatial dependence. When a segment has only one site, we let $\nu_{s_i} \sim \mathrm{N}\left(0, \sigma_\nu^2\right)$. Past studies have shown that a combination of geostatistical and stream network models improved spatial predictive ability (Peterson and Ver Hoef, 2010). At the network level, we apply our stochastic SSN model where (Lu et al., 2024) previously adopted a conventional SSN model (Ver Hoef and Peterson, 2010) to characterize dispersal-induced dependence along stream networks. Our study system consisted of M hydrologically separated networks and we assume conditional independence among the networks. We characterize dependence among the segments within the same network, $m = 1, \ldots, M$, based on the inferred connectivity matrix, $\boldsymbol{C}_m$ (Eq. (4)), as follows,

$$\boldsymbol{\psi}_m \sim \mathrm{N}\left(\mathbf{0}, \sigma_\psi^2(\mathrm{diag}(\boldsymbol{C}_m\mathbf{1}) - \rho_\psi \boldsymbol{C}_m)^{-1}\right), \tag{10}$$

where $\rho_\psi = 0.99$. When a network has only one segment, we let $\psi_{m_i} \sim \mathrm{N}\left(0, \sigma_\psi^2\right)$.

Lastly, we model habitat-specific linear trends in log year, where $\gamma_i = \gamma_{\text{suitable}}$ if site i was in a suitable habitat, and $\gamma_i = \gamma_{\text{unsuitable}}$ otherwise. We account for temporal correlation using a first-order autoregressive (AR(1)) model,

$$\eta_t \sim \mathrm{N}\left(\rho_\eta \eta_{t-1},\ \sigma_\eta^2\right),\ t = 2, \ldots, T, \tag{11}$$

where $\rho_\eta \in (-1, 1)$ controls temporal dependence of a stationary process, and $\eta_1 \sim \mathrm{N}\left(0, \sigma_\eta^2\right)$.

2.3. Joint distribution and sampling

We use the bracket notation $[\cdot]$ to represent probability distributions (Gelfand and Smith, 1991). The joint posterior distribution of all model parameters is

$$\begin{aligned}
&\left[\boldsymbol{N}, \boldsymbol{q}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\epsilon}, \sigma_\epsilon^2, \{\boldsymbol{\nu}_s\}, \sigma_\nu^2, \phi_\nu, \{\boldsymbol{\psi}_m\}, \sigma_\psi^2, \{\boldsymbol{C}_m\}, \xi, \tau, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\eta^2, \rho_\eta | \boldsymbol{Y}\right] \\
&\propto [\boldsymbol{Y}|\boldsymbol{N}, \boldsymbol{q}] \times \left[\boldsymbol{N}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\epsilon}, \{\boldsymbol{\nu}_s\}, \{\boldsymbol{\psi}_m\}, \boldsymbol{\gamma}, \boldsymbol{\eta}\right] \times [\boldsymbol{q}] \times [\boldsymbol{\beta}] \times [\boldsymbol{\theta}] \times [\boldsymbol{\gamma}] \times \\
&\quad \left[\boldsymbol{\epsilon}|\sigma_\epsilon^2\right] \times \left[\sigma_\epsilon^2\right] \times \prod_s \left[\boldsymbol{\nu}_s|\sigma_\nu^2, \phi_\nu\right] \times \left[\sigma_\nu^2\right] \times \left[\phi_\nu\right] \times \prod_m \left[\boldsymbol{\psi}_m|\sigma_\psi^2, \boldsymbol{C}_m\right] \times \\
&\quad \left[\sigma_\psi^2\right] \times \prod_m \left[\boldsymbol{C}_m|\xi, \tau\right] \times [\xi] \times [\tau] \times \left[\boldsymbol{\eta}|\sigma_\eta^2, \rho_\eta\right] \times \left[\sigma_\eta^2\right] \times \left[\rho_\eta\right],
\end{aligned}$$

where $\boldsymbol{Y}$ is the array of observed counts, $\boldsymbol{N}$ is the array of true abundance, $\boldsymbol{q} = (q_{\text{single}}, q_{\text{depletion}})'$ is the vector of capture probabilities, $\{\boldsymbol{\nu}_s\}$ is the set of segment-specific random vectors, $\{\boldsymbol{\psi}_m\}$ is the set of network-specific random vectors, and $\{\boldsymbol{C}_m\}$ is the set of network-specific connectivity matrices. We provide a list of prior distributions for the case study in Appendix A. The prior distributions for the capture probabilities are informed by previous studies (Kanno et al., 2015). The remaining parameters are given diffuse prior distributions. Most of the model parameters are sampled using Metropolis–Hastings updates based on their full-conditional distributions (denoted [parameter|·], representing the conditional distribution of the parameter given all other parameters in the model and the data). Specifically, the full-conditional distributions for parameters related to the stochastic SSNs are,

$$\begin{aligned}
[\xi|\cdot] &\propto \prod_m [\boldsymbol{C}_m|\xi, \tau] \times [\xi], \\
[\tau|\cdot] &\propto \prod_m [\boldsymbol{C}_m|\xi, \tau] \times [\tau], \\
[\boldsymbol{C}_m|\cdot] &\propto \left[\boldsymbol{\psi}_m|\sigma_\psi^2, \boldsymbol{C}_m\right] \times [\boldsymbol{C}_m|\xi, \tau].
\end{aligned}$$

Table 1
Parenthood for 10 nodes in a simulated network.

Node	1	2	3	4	5	6	7	8	9	10
Parent	0	10	9	0	10	8	8	4	0	9

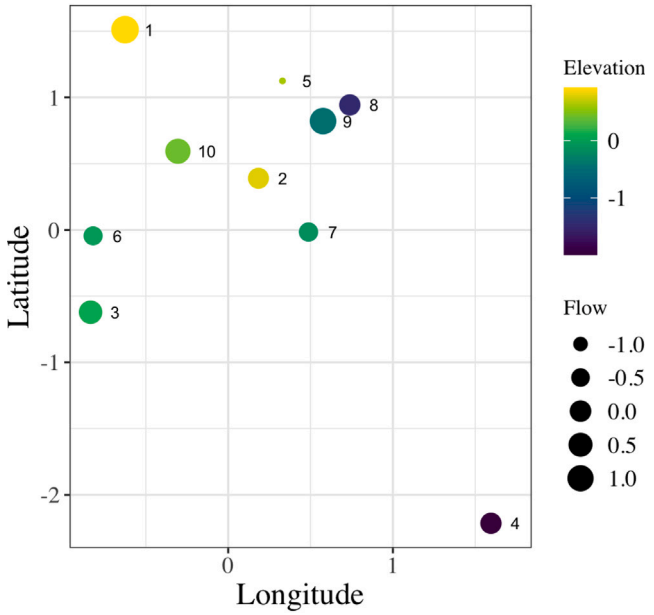


Fig. 1. The simulated network illustrated by nodal attributes in longitude, latitude, elevation, and flow. These attributes defined the feasibility matrix, $\boldsymbol{F}$, deterministically, and $\boldsymbol{F}$ was involved in the probabilistic modeling of parenthood in Table 1.

The computational complexity of our sampling algorithm depends on four conditions: 1. the number of networks; 2. the maximum number of nodes within a network; 3. the maximum number of feasible parents to a node; and 4. the maximum generation distance within a network (i.e., depth). Because our method approximates the physical networks, conditions 1 and 2 may be alleviated via parallelization by reasonably partitioning the study domain into conditionally independent functional networks. For each network m, we sequentially sample the parent of node i, a_i^m, conditioned on the parents of the remaining nodes, $\boldsymbol{a}_{-i}^m$, and reconstruct $\boldsymbol{C}_m$ for every node at every iteration, whose complexity is on the order of the network depth (Eq. (4)). Nonetheless, conditions 3 and 4 did not present a computational bottleneck in our case study because native brook trout populations in North Carolina congregated in headwaters, meaning the functional networks should have short generation distances and each node has a limited number of feasible parents (Fig. 5(a)).

3. Application

3.1. Simulation

The first purpose of this two-part simulation study is to examine the various network attributes corresponding to different parameterizations. We specified $n = 10$ nodes. For each node, $i = 1, \ldots, n$, we sampled elevation, x_i^e, flow, x_i^f, and coordinates in two-dimensions from standard normal distributions (Fig. 1). The simulated elevation and flow represented normalized covariates in the case study. We calculated the feasibility matrix, $\boldsymbol{F}$, using Eq. (1) based on $\boldsymbol{x}^e$ and $\boldsymbol{x}^f$, and we sampled the parents, $\boldsymbol{a}$, using Eq. (2) with $\xi = 0.9$ and $\tau = 1$ (Table 1). The resulting functional network had three components that were disconnected from each other (Fig. 2). The depths as defined by the longest shortest path from one node to another within each component were one, three, and three, from left to right, and the maximum component depth of the network was three. Lastly, we constructed the connectivity matrix, $\boldsymbol{C}$, using Eq. (4) based on $\boldsymbol{a}$.

The parent vector uniquely identifies the functional network. Conditioned on the feasibility matrix, the complexity of the network depends on parameters ξ and τ. We simulated functional networks for various combinations of ξ and τ given $\boldsymbol{F}$, and summarized their complexity using the mode of the number of components and the maximum component depth (Fig. 3). As ξ increased from 0 to 1, the number of components decreased and the maximum component depth increased for all levels of τ, indicating higher overall connectivity. As τ increased from negative to positive, the number of components increased for moderate values of ξ (i.e., $\xi \in (0.25, 0.72)$); the maximum component depth increased then decreased for moderate values of ξ, with peak depths occurring

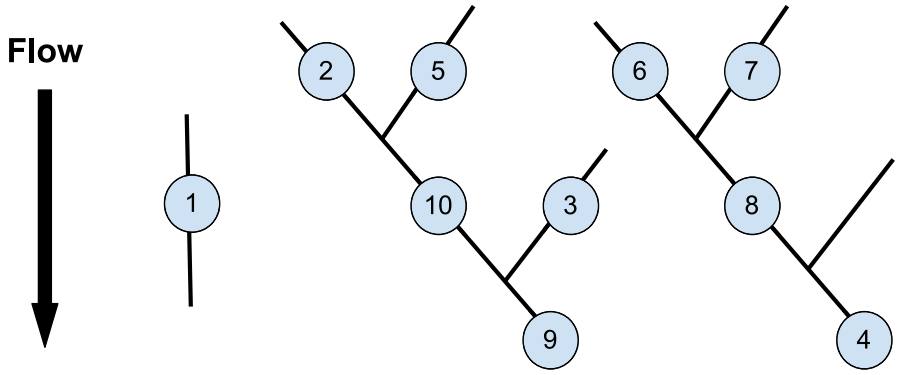


Fig. 2. The simulated network represented by its functional connectivity. Flow direction is from top to bottom. Each node corresponds to a line segment, blind-ended or connected to another segment on one or both ends. The unnumbered segment in the right-most component is an example of an unobserved segment.

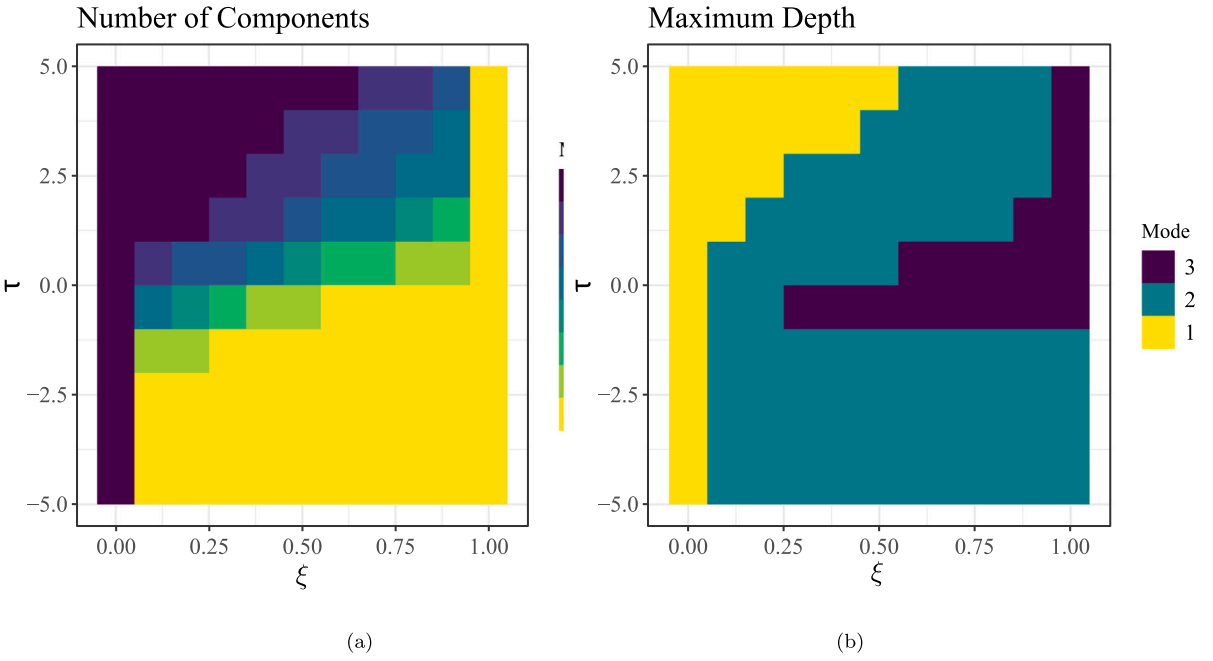


Fig. 3. Heatmaps of simulated network complexity with varying values of ξ and τ.

around $\tau = 0$. This indicated that strong effects of geographic distance (i.e., large magnitude of τ) were associated with less connected networks, especially when the effect was positive.

The second purpose of this two-part simulation study is to demonstrate parameter estimability and predictive ability of our proposed method. We specified $M = 10$ networks, each with $n = 10$ nodes, across $T = 20$ years. We generated connectivity matrices, $\boldsymbol{C}_m$, following Section 2.1, and simulated count data using the spatio-temporal population model,

$$
\begin{aligned}
y_{i,t} &\sim \text{Binom}\left(N_{i,t}, q\right), \\
N_{i,t} &\sim \text{Pois}\left(\lambda_{i,t}\right), \\
\log\left(\lambda_{i,t}\right) &= \boldsymbol{x}_i'\boldsymbol{\beta} + \psi_{m_i} + \eta_t, \\
\boldsymbol{\psi}_m \sim \text{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_m\right),\ \boldsymbol{\Sigma}_m &= \sigma_\psi^2\left(\text{diag}(\boldsymbol{C}_m\boldsymbol{1}) - 0.99\boldsymbol{C}_m\right)^{-1}, \\
\text{for } t = 2, \ldots, T,\ \eta_t \sim \text{N}\left(\rho\eta_{t-1}, \sigma_\eta^2\right),\ \eta_1 &\sim \text{N}\left(0, \sigma_\eta^2\right),
\end{aligned}
$$

where $i = 1, \ldots, n \times M$, $m = 1, \ldots, M$, and $t = 1, \ldots, T$. Vector $\boldsymbol{x}_i$ consisted of an intercept and two covariates from standard normal distributions. We fit the model to simulated data using a Markov Chain Monte Carlo (MCMC) algorithm. We ran the algorithm in

Table 2
Summary of the true parameters, the estimated marginal posterior means (95% credible intervals), and the convergence statistics from simulation.

Parameter	True	Posterior	$\hat{R}$
β_0	−0.5	−0.51 (−0.59, −0.43)	1.04
β_1	0.3	0.32 (0.22, 0.41)	1.00
β_2	1.5	1.50 (1.41, 1.60)	1.00
σ^2_ψ	0.1	0.09 (0.06, 0.14)	1.01
ξ	0.95	0.99 (0.95, 1.00)	1.00
τ	−2.5	−3.27 (−5.99, −1.31)	1.00
σ^2_η	0.1	0.13 (0.07, 0.25)	1.00
ρ	0.7	0.73 (0.40, 0.98)	1.01
q	0.7	0.67 (0.49, 0.88)	1.11

R (Team, 2021) with 40,000 iterations and discarded the first 24,000 burn-in samples. The remaining samples were thinned at a rate of $\frac{1}{10}$. Our model was able to recover the true parameters in simulation with their respective marginal 95% credible intervals (Table 2). The Gelman–Rubin statistics ($\hat{R}$) close to one indicated convergence (Gelman and Rubin, 1992).

Furthermore, our method demonstrated improved spatial predictive ability via a three-fold cross-validation. At each non-overlapping fold, $r = 1, 2, 3$, we randomly designated one-third of the data in both spatial and temporal dimensions as the test set, and the remaining two-thirds of the data as the training set. We fit our proposed model and a null model assuming spatial independence to the training set and predicted for the test set. To predict at a node, i, that was observed, we drew the kth posterior predictive sample of density from the stochastic SSN model based on $\boldsymbol{\beta}^{(k)}$, $\psi_i^{(k)}$, and $\sigma_\eta^{2(k)}$ as follows,

$$\lambda^{(k)}_{\text{Pred},i,t} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}^{(k)} + \psi^{(k)}_{m_i} + \eta^{(k)}_t\right).$$

To predict at a new node, i, with known covariates, we sampled $\psi^{(k)}_{m_i}$ based on $\boldsymbol{\psi}^{(k)}_{m_{-i}}$, $\sigma^{2(k)}_\psi$, and $\boldsymbol{\Sigma}^{(k)}_m$ by kriging as follows,

$$\psi^{(k)}_{m_i} \sim \text{N}\left(-\boldsymbol{\Sigma}^{(k)}_{m_{i,-i}}\left(\boldsymbol{\Sigma}^{(k)}_{m_{-i,-i}}\right)^{-1}\boldsymbol{\psi}^{(k)}_{m_{-i}},\ \sigma^{2(k)}_\psi - \boldsymbol{\Sigma}^{(k)}_{m_{i,-i}}\left(\boldsymbol{\Sigma}^{(k)}_{m_{-i,-i}}\right)^{-1}\boldsymbol{\Sigma}^{(k)}_{m_{-i,i}}\right).$$

Posterior predictive samples of density for the null model were obtained using $\lambda^{(k)}_{\text{Pred},i,t} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}^{(k)} + \eta^{(k)}_t\right)$. We evaluated both models using posterior predictive log likelihood (PPLL) as follows,

$$\text{PPLL}(r) = \int_{\boldsymbol{q}}\int_{\boldsymbol{\Lambda}_{\text{Pred}}} \left[\boldsymbol{Y}^r_{\text{Test}}|\boldsymbol{\Lambda}_{\text{Pred}},\ \boldsymbol{q}\right]\left[\boldsymbol{\Lambda}_{\text{Pred}},\boldsymbol{q}|\boldsymbol{Y}^r_{\text{Train}}\right] d\boldsymbol{\Lambda}_{\text{Pred}} d\boldsymbol{q},$$

where $\boldsymbol{\Lambda}_{\text{Pred}}$ is the matrix of predicted densities. We also used the integrated likelihood to approximate PPLL numerically with a large upper bound for $\boldsymbol{N}_{\text{Pred}}$ as follows,

$$\left[\boldsymbol{Y}^r_{\text{Test}}|\boldsymbol{\Lambda}_{\text{Pred}},\ \boldsymbol{q}\right] = \sum_{\boldsymbol{N}_{\text{Pred}}=\boldsymbol{0}}^{\infty}\left[\boldsymbol{Y}_{\text{Test}}|\boldsymbol{N}_{\text{Pred}},\boldsymbol{q}\right]\left[\boldsymbol{N}_{\text{Pred}}|\boldsymbol{\Lambda}_{\text{Pred}}\right].$$

Our stochastic SSN model outperformed the null model with a better mean score (stochastic SSN: −0.21; null: −0.58).

3.2. Case study

We illustrate our method using population surveys on native brook trout from 173 sites in western North Carolina between 1989 and 2015. Count data were collected from electro-fishing surveys under single-pass or depletion sampling protocols. The sites were distributed across 11 major watersheds (HUC8) and 14 stream networks inhabited by brook trout on both sides of the Eastern Continental Divide in the southern Appalachian Mountains. (Fig. 4).

Like most stream fish, brook trout populations respond to thermal and flow conditions. Previous studies showed that exposures to upper thermal limits and flood events during early life stages can affect abundance negatively (Roghair et al., 2002; Xu et al., 2010; Warren et al., 2012; Kanno et al., 2015). Therefore, we used high summer temperature and high winter and spring flows as weather-related fixed effects, $\boldsymbol{h}_{i,t}$ (Eq. (7)). We obtained daily maximum temperatures from Daymet (Thornton et al., 2022) over a 1 km × 1 km grid, and used the average of daily maxima between June and September to represent high summer temperature at each site. We obtained monthly flow percentiles from NHDPlus (U.S. Geological Survey, 2016), and used the maximum available percentile between December and February to represent high winter flow and that between March and May to represent high spring flow at each segment. These covariates were standardized by site to highlight temporal dynamics of local weather patterns. In addition, we used latitude, elevation, and indicator for slope as landscape fixed effects, $\boldsymbol{x}_i$, where latitude and elevation were standardized over the study region.

We implemented our hierarchical model (Section 2.2) using an MCMC algorithm, and fit to the juvenile (total length < 90 mm) and adult (total length ≥ 90 mm) data separately. We ran the algorithm in R (Team, 2021) with 40,000 iterations and discarded the

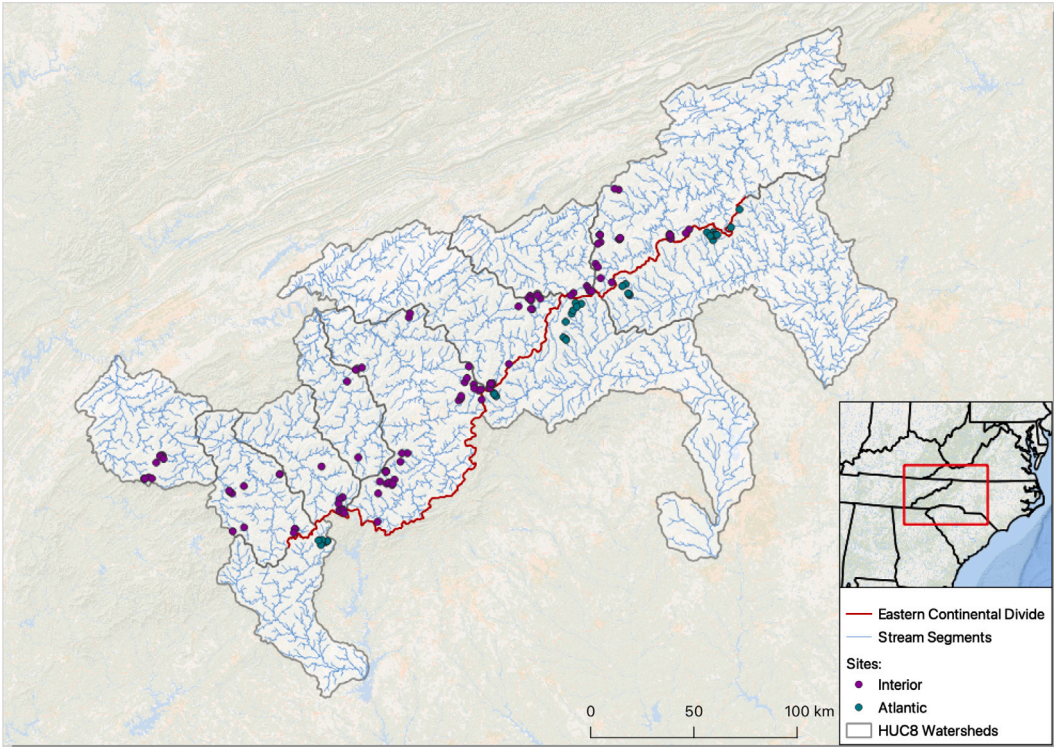


Fig. 4. Map of the study sites in western North Carolina, USA.

first 24,000 burn-in samples. The remaining samples were thinned at a rate of $\frac{1}{10}$. The MCMC algorithm of Lu et al. (2024) required 0.9 h to complete 10,000 iterations on a 2.5 GHz Intel Core i5 processor. In comparison, our algorithm required 1.3 h to complete 10,000 iterations on the same processor. Our analysis did not involve pre-processing, whereas processing the SSNs in Lu et al. (2024) using the STARS tool set required many hours of manual editing by a qualified personnel, in addition to days of program runtime on a 4.1 GHz Intel Xeon processor. Our inference agreed with previous studies Appendix B: Capture efficiencies (per pass) were comparable between single-pass and depletion surveys for both juveniles and adults. For both life stages, mean densities were higher among suitable habitats on the western slope. Western juveniles were most adversely affected by high spring flows, whereas eastern juveniles were most adversely affected by high winter flows. The adult populations were less responsive to seasonal weather patterns than juveniles in general. Most spatial variation was revealed at the segment level among the multi-scale random effects. Temporal variation was smaller and autocorrelation was lower in adult populations than in juveniles.

Specific to our stochastic SSN model, we inferred close-to-one ξ (juvenile: $0.99(0.96, 1.00)$, adult: $0.99(0.96, 1.00)$) and significantly negative τ (juvenile: $-2.78(-3.87, -1.78)$, adult: $-2.71(-4.01, -1.56)$) for both life stages. Our inference on ξ indicated high functional connectivity among the surveyed segments. Our inference on τ was reasonable for the case study where most surveyed segments were head water streams and geographically neighboring segments were hydrologically distant. We demonstrated an example of our inferred functional network in Fig. 5(a). For the purpose of illustration, we visualized at the finest spatial resolution that contained all surveyed segments. The actual network covered a larger spatial domain. Two among the six sites in the SSN were located on the same segment, and their elevations and flows were averaged to obtain the nodal attributes that produced the feasibility matrix

$$\boldsymbol{F}_m = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We derived the point estimate of the functional network using the posterior mode of the parent vector. In this case, vector $\boldsymbol{a}_m = (0, 4, 1, 1, 1)'$ accounted for 76% of all posterior samples, indicating the inferred network had only one component with a depth of three (Fig. 5(b)). Upon comparison between Figs. 5(a) and 5(b), the physical network supported our inference on the direct parenthood of segment 1 to segments 4 and 5 via visual inspection. Segment 1 may be the parent to segment 3 due to downstream merging, yet their dependence appeared much weaker than dependence between segments 1 and 4. The most arguable inference was the direct parenthood to segment 2. The direct parent to segment 2 appeared to be segment 1 but was inferred to be segment 4, even though segments 2 and 1 were still correlated through indirect connection.

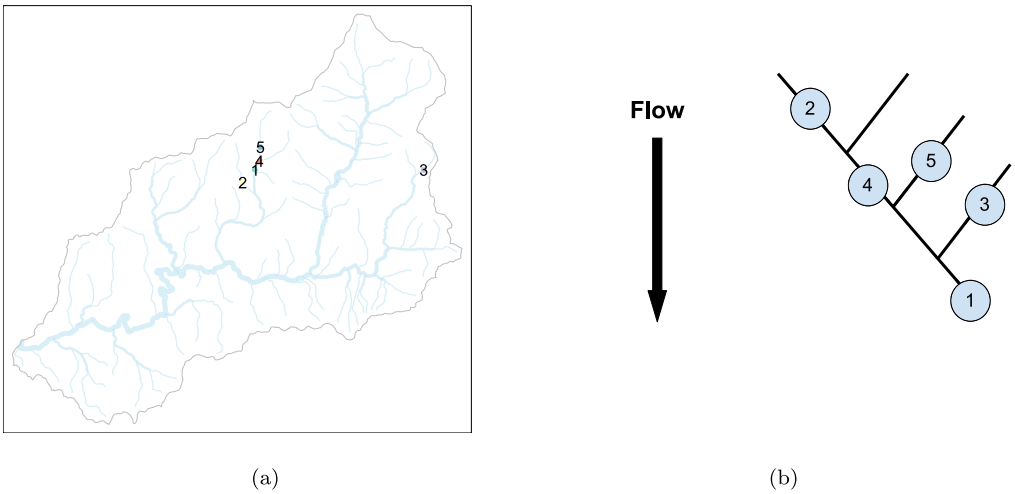


Fig. 5. An example of a physical SSN in (a) and the inferred functional network in (b).

We also compared the predictive ability of our proposed stochastic SSN framework to the conventional SSN of Lu et al. (2024) via a three-fold cross validation scored by PPLL. Similar to the procedure described in Section 3.1, we used a posterior sample of the network level random effect, $\psi_{m_i}^{(k)}$, to predict density at a surveyed segment, and we obtained a posterior predictive sample of $\psi_{m_i}^{(k)}$ by kriging at a new segment conditioned on the inferred functional network. Our stochastic SSN model outperformed the conventional SSN model with a better mean score (stochastic SSN: −0.78; conventional SSN: −0.90).

4. Discussion

We presented a Bayesian framework to characterize spatially correlated processes in dendritic networks. Compared to conventional SSN models, our framework shortened the overall analysis timeline without the need to manually process large-scale stream network data. Our framework retained the representation of branching hierarchies distinctive of conventional SSNs. In addition, we accounted for uncertainty in functional connectivity due to partially observed networks. We showed in simulated examples and a case study that our method enhanced spatial predictive ability compared to models not characterizing functional connectivity.

By incorporating the stochastic SSN within a population model, we were able to infer functional networks using physical principles as well as observed data. Although τ and ξ enhanced our understanding of functional connectivity across the network, we are cautious when interpreting inferred connections by segment. There are several ways by which functional-network-related inference may be refined. First, in a study system characterized by multiple networks, letting ξ and τ vary by network may improve network-specific inference and the model's overall predictive ability. Second, future applications could consider different spatial resolutions upon which the networks are defined. The choice of resolution may rely on empirical knowledge or model comparison, and a multi-resolution approach may be explicitly specified at the network level (Katzfuss, 2017). Third, we may relax the assumptions on feasibility matrices, for example, by allowing flow volume to decrease downstream, to account for various stream and network characteristics such as losing streams and delta structures. Lastly, data on anthropogenic/natural barriers (e.g., dams, culverts, waterfalls) could be used to define feasibility matrices (Warren, Jr. and Pardew, 1998; Porto et al., 1999), for example, by letting $F_{ij} = 0$ if a known barrier exists between nodes i and j. These data should ideally not involve processing large databases so as to maintain our computational advantage.

Other than integrating auxiliary data deterministically into feasibility matrices, we can investigate covariate effects on functional connectivity. In Eq. (3), we can define $u(d_{ij}) \equiv \exp(\|\boldsymbol{d}_i - \boldsymbol{d}_j\|'\boldsymbol{\tau})$, where d_{ij} represents the combined effect of dyadic similarities between segments including climate gradients such as air temperature and precipitation, as well as landscape characteristics such as slope and vegetation cover (Hanks and Hooten, 2013; White et al., 2020). Finally, although our stochastic SSN is defined on segments, we demonstrated in the case study that predictions can be made continuously in space by combining our model with a geostatistical model. Future studies can explore the optimal combination of spatial configurations most suitable for specific modeling objectives.

Funding and acknowledgment

This research was funded by Cooperative Agreement No. G21AC1005 from the U.S. Geological Survey Southeast Climate Adaptation Science Center, and Cooperative Agreement No. F20AC11372 from the U.S. Fish and Wildlife Service. We thank numerous technicians for their assistance in the field.

Table B.3
Estimated posterior means (95% credible intervals) of covariate effects on log density.

Parameter	Juvenile		Adult	
	Unsuitable	Suitable	Unsuitable	Suitable
β_1 (Lat.)	0.33 (0.19, 0.46)		0.27 (0.07, 0.45)	
β_2 (Elev.)	1.77 (1.41, 2.14)		1.56 (1.35, 1.80)	
β_0 (Mean)	0.30 (−0.41, 0.94)	1.35 (0.97, 1.75)	−0.42 (−1.18, 0.03)	0.66 (0.35, 1.04)
θ_1 (Temp)	−0.06 (−0.18, 0.06)	−0.23 (−0.32, −0.13)	0.01 (−0.12, 0.13)	−0.04 (−0.18, 0.05)
θ_2 (W. Flow)	−0.90 (−1.08, −0.71)	−0.24 (−0.33, −0.14)	−0.10 (−0.30, 0.06)	0.03 (−0.07, 0.12)
θ_3 (S. Flow)	−0.50 (−0.69, −0.33)	−0.50 (−0.63, −0.40)	0.12 (−0.01, 0.23)	0.06 (−0.03, 0.16)
γ (Trend)	0.37 (0.19, 0.60)	0.19 (−0.76, 1.33)	0.78 (0.62, 0.96)	0.04 (−0.05, 0.16)

Table B.4
Estimated posterior means (95% credible intervals) of capture probability and random effect parameters.

Parameter	Juvenile	Adult
$q_{\text{depletion}}$ (Depletion Capture)	0.58 (0.56, 0.59)	0.70 (0.68, 0.71)
q_{single} (Single-Pass Capture)	0.59 (0.49, 0.69)	0.70 (0.62, 0.78)
σ_ϵ^2 (Watershed Variance)	0.07 (0.03, 0.18)	0.03 (0.01, 0.07)
σ_ν^2 (Segment Variance)	7.43 (5.47, 10.55)	2.90 (2.19, 3.80)
ϕ_ν (Segment Range)	5.76 (3.22, 9.49)	2.80 (1.25, 4.51)
σ_ψ^2 (Network Variance)	0.01 (0.01, 0.02)	0.01 (0.00, 0.01)
ξ (Network Complexity)	0.99 (0.96, 1.00)	0.99 (0.96, 1.00)
τ (Network Complexity)	−2.78 (−3.87, −1.78)	−2.71 (−4.01, −1.56)
σ_η^2 (Temporal Variance)	0.77 (0.43, 1.29)	0.07 (0.02, 0.16)
ρ_η (Autocorrelation)	0.17 (0.01, 0.47)	0.55 (0.07, 0.91)

Appendix A. Prior distributions

$$
\begin{aligned}
&q_{\text{single}},\ q_{\text{depletion}} \sim \text{Beta}(60, 40) \text{ for YOY},\\
&q_{\text{single}},\ q_{\text{depletion}} \sim \text{Beta}(70, 30) \text{ for adults},\\
&\boldsymbol{\beta},\ \boldsymbol{\theta}_{\text{suitable}},\ \boldsymbol{\theta}_{\text{unsuitable}},\ \gamma_{\text{suitable}},\ \gamma_{\text{unsuitable}} \sim \text{N}(\mathbf{0}, \mathbf{I}),\\
&\log(\sigma_\epsilon),\ \log(\sigma_\nu),\ \log(\sigma_\psi),\ \log(\sigma_\eta) \sim \text{N}(0, 1),\\
&\phi_\nu \sim \text{Unif}(0, 10),\\
&\xi \sim \text{Beta}(1, 1),\\
&\tau \sim \text{N}(0, 1),\\
&\rho_\eta \sim \text{Unif}(-1, 1).
\end{aligned}
$$

Appendix B. Posterior inference for case study

See Tables B.3 and B.4.

Data availability

The datasets and programming codes used in the study will be available on ScienceBase: https://www.sciencebase.gov/catalog/item/5f62407d82ce38aaa236148b.

References

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2003. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC.
Berliner, L.M., 1996. Hierarchical Bayesian time series models. In: Maximum Entropy and Bayesian Methods: Santa Fe, New Mexico, USA, 1995 Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods. Springer, pp. 15–22.
Biggs, N., Lloyd, E.K., Wilson, R.J., 1986. Graph Theory, 1736-1936. Oxford University Press.
Brown, B., Swan, C., 2010. Dendritic network structure constrains metacommunity properties in riverine ecosystems. J. Anim. Ecol. 79 (3), 571–580.

Campbell Grant, E.H., Lowe, W.H., Fagan, W.F., 2007. Living in the branches: Population dynamics and ecological processes in dendritic networks. Ecol. Lett. 10 (2), 165–175.
Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons.
Cressie, N., Frey, J., Harch, B., Smith, M., 2006. Spatial prediction on a river network. J. Agric. Biol. Environ. Stat. 11, 127–150.
Dumelle, M., Peterson, E.E., Ver Hoef, J.M., Pearse, A., Isaak, D.J., 2024. SSN2: The next generation of spatial stream network modeling in R. J. Open Source Softw. 9 (99), 6389.
ESRI, 2011. ArcGIS desktop: Release 10. Redlands, CA: Environmental systems research institute.
Estrada, R., Tomasi, C., Schmidler, S.C., Farsiu, S., 2014. Tree topology estimation. IEEE Trans. Pattern Anal. Mach. Intell. 37 (8), 1688–1701.
Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M., 2010. Handbook of Spatial Statistics. CRC Press.
Gelfand, A.E., Smith, A.F., 1991. Gibbs sampling for marginal posterior expectations. Comm. Statist. Theory Methods 20 (5–6), 1747–1766.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7 (4), 457–472.
Hanks, E.M., 2017. Modeling spatial covariance using the limiting distribution of spatio-temporal random walks. J. Amer. Statist. Assoc. 112 (518), 497–507.
Hanks, E.M., Hooten, M.B., 2013. Circuit theory and model-based inference for landscape connectivity. J. Amer. Statist. Assoc. 108 (501), 22–33.
Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S.L., Scharf, H.R., Tipton, J.R., Williams, P.J., Hooten, M.B., 2017. The basis function approach for modeling autocorrelation in ecological data. Ecology 98 (3), 632–646.
Hill, R.A., Weber, M.H., Leibowitz, S.G., Olsen, A.R., Thornbrugh, D.J., 2016. The stream-catchment (StreamCat) dataset: A database of watershed metrics for the conterminous United States. J. Am. Water Resour. Assoc. 52 (1), 120–128.
Isaak, D.J., Ver Hoef, J.M., Peterson, E.E., Horan, D.L., Nagel, D.E., 2017. Scalable population estimates using spatial-stream-network (SSN) models, fish density surveys, and national geospatial database frameworks for streams. Can. J. Fish. Aquat. Sci. 74 (2), 147–156.
Kanno, Y., Letcher, B.H., Hitt, N.P., Boughton, D.A., Wofford, J.E., Zipkin, E.F., 2015. Seasonal weather patterns drive population vital rates and persistence in a stream fish. Global Change Biol. 21 (5), 1856–1870.
Kattwinkel, M., Szöcs, E., Peterson, E., Schäfer, R.B., 2020. Preparing GIS data for analysis of stream monitoring data: The R package openSTARS. PLoS One 15 (9), e0239237.
Katzfuss, M., 2017. A multi-resolution approximation for massive spatial datasets. J. Amer. Statist. Assoc. 112 (517), 201–214.
Krackhardt, D., 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. Soc. Netw. 10 (4), 359–381.
Lu, X., Kanno, Y., Valentine, G., Rash, J., Hooten, M., 2024. Using multi-scale spatial models of dendritic ecosystems to infer abundance of a stream salmonid. J. Appl. Ecol. 61 (7), 1703–1715.
Lu, X., Williams, P.J., Hooten, M.B., Powell, J.A., Womble, J.N., Bower, M.R., 2020. Nonlinear reaction–diffusion process models improve inference for population dynamics. Environmetrics 31 (3), e2604.
McManus, M.G., D'Amico, E., Smith, E.M., Polinsky, R., Ackerman, J., Tyler, K., 2020. Variation in stream network relationships and geospatial predictions of watershed conductivity. Freshw. Sci. 39 (4), 704–721.
Meador, B., 2008. A survey of computer network topology and analysis examples. Wash. Univ. 2–3.
Nagel, D., Peterson, E., Isaak, D., Ver Hoef, J., Horan, D., 2015. National stream internet protocol and user guide. US For. Serv. Rocky Mt. Res. Stn. Air, Water, Aquat. Environ. Program 2023-2003.
Peterson, E.E., Dumelle, M., Pearse, A.R., Teleki, D., Ver Hoof, J., 2024. An introduction to 'SSNbler': Assembling spatial stream network (SSN) objects in R.
Peterson, E.E., Ver Hoef, J.M., 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. Ecology 91 (3), 644–651.
Peterson, E., Ver Hoef, J., 2014. STARS: An arcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. J. Stat. Softw. 56, 1–17.
Porto, L., McLaughlin, R., Noakes, D., 1999. Low-head barrier dams restrict the movements of fishes in two Lake Ontario streams. North Am. J. Fish. Manag. 19 (4), 1028–1036.
Rodríguez-Castillo, T., Estévez, E., González-Ferreras, A.M., Barquín, J., 2019. Estimating ecosystem metabolism to entire river networks. Ecosystems 22, 892–911.
Rodríguez-González, P.M., García, C., Albuquerque, A., Monteiro-Henriques, T., Faria, C., Guimarães, J.B., Mendonça, D., Simões, F., Ferreira, M.T., Mendes, A., et al., 2019. A spatial stream-network approach assists in managing the remnant genetic diversity of riparian forests. Sci. Rep. 9 (1), 6741.
Roghair, C.N., Dolloff, C.A., Underwood, M.K., 2002. Response of a brook trout population and instream habitat to a catastrophic flood and debris flow. Trans. Am. Fish. Soc. 131 (4), 718–730.
Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60 (1), 108–115.
Santos-Fernandez, E., Ver Hoef, J.M., Peterson, E.E., McGree, J., Isaak, D.J., Mengersen, K., 2022. Bayesian spatio-temporal models for stream networks. Comput. Statist. Data Anal. 170, 107446.
Scown, M.W., McManus, M.G., Carson, Jr., J.H., Nietch, C.T., 2017. Improving predictive models of in-stream phosphorus concentration based on nationally-available spatial data coverages. J. Am. Water Resour. Assoc. 53 (4), 944–960.
Som, N.A., Monestiez, P., Ver Hoef, J.M., Zimmerman, D.L., Peterson, E.E., 2014. Spatial sampling on streams: Principles for inference on aquatic networks. Environmetrics 25 (5), 306–323.
Team, R.C., 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.
Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S., Wilson, B., 2022. Daymet: Daily surface weather data on a 1-km grid for North America, version 4 R1. ORNL DAAC, Oak Ridge, Tennessee, USA.
U.S. Geological Survey, 2016. NHDPlus version 2.
Ver Hoef, J.M., Hanks, E.M., Hooten, M.B., 2018. On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spat. Stat. 25, 68–85.
Ver Hoef, J.M., Peterson, E.E., 2010. A moving average approach for spatial statistical models of stream networks. J. Amer. Statist. Assoc. 105 (489), 6–18.
Ver Hoef, J., Peterson, E., Clifford, D., Shah, R., 2014. SSN: An R package for spatial statistical modeling on stream networks. J. Stat. Softw. 56, 1–45.
Ver Hoef, J.M., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13, 449–464.
Warren, D.R., Robinson, J.M., Josephson, D.C., Sheldon, D.R., Kraft, C.E., 2012. Elevated summer temperatures delay spawning and reduce redd construction for resident brook trout (salvelinus fontinalis). Global Change Biol. 18 (6), 1804–1811.
Warren, Jr., M.L., Pardew, M.G., 1998. Road crossings as barriers to small-stream fish movement. Trans. Am. Fish. Soc. 127 (4), 637–644.
Wenger, S.J., Isaak, D.J., Luce, C.H., Neville, H.M., Fausch, K.D., Dunham, J.B., Dauwalter, D.C., Young, M.K., Elsner, M.M., Rieman, B.E., et al., 2011. Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. Proc. Natl. Acad. Sci. 108 (34), 14175–14180.
White, S.L., Hanks, E.M., Wagner, T., 2020. A novel quantitative framework for riverscape genetics. Ecol. Appl. 30 (7), e02147.
Wright, W.J., Neitlich, P.N., Shiel, A.E., Hooten, M.B., 2022. Mechanistic spatial models for heavy metal pollution. Environmetrics 33 (8), e2760.
Xu, C., Letcher, B.H., Nislow, K.H., 2010. Context-specific influence of water temperature on brook trout growth rates in the field. Freshwater Biol. 55 (11), 2253–2264.
Zimmerman, D.L., Ver Hoef, J.M., 2024. Spatial Linear Models for Environmental Data. CRC Press.