Check for updates

Ecological Applications, 32(1), 2022, e02485
© 2021 by the Ecological Society of America

Partitioning macroscale and microscale ecological processes using covariate-driven non-stationary spatial models

Charlotte F. Narr [1,2,4] Pavel Chernyavskiy,[3] and Sarah M. Collins [2]

[1]*Southern Illinois University in Carbondale, Carbondale, Illinois 62901 USA*
[2]*Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming 82071 USA*
[3]*Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia 22903 USA*

Citation: Narr, C. F., P. Chernyavskiy, and S. M. Collins. 2022. Partitioning macroscale and microscale ecological processes using covariate-driven non-stationary spatial models. Ecological Applications 32(1): e02485. 10.1002/eap.2485

Abstract. Ecological inference requires integrating information across scales. This integration creates a complex spatial dependence structure that is most accurately represented by fully non-stationary models. However, ecologists rarely use these models because they are difficult to estimate and interpret. Here, we facilitate the use of fully non-stationary models in ecology by improving the interpretability of a recently developed non-stationary model and applying it to improve our understanding of the spatial processes driving lake eutrophication. We reformulated a model that incorporates non-stationary correlation by adding environmental predictors to the covariance function, thereby building on the intuition of mean regression. We created ellipses to visualize how data at a given site correlate with their surroundings (i.e., the range and directionality of underlying spatial processes). We applied this model to describe the spatial dependence structure of variables related to lake eutrophication across two different regions: a Midwestern United States region with highly agricultural landscapes, and a Northeastern United States region with heterogeneous land use. For the Midwest, increases in forest cover increased the homogeneity of the residual spatial structure of total phosphorus, indicating that macroscale processes dominated this nutrient's spatial structure. Conversely, high forest cover and baseflow reduced the spatial homogeneity of chlorophyll *a* residuals, indicating that microscale processes dominated for chlorophyll *a* in the Midwest. In the Northeast, increases in urban land use and baseflow decreased the homogeneity of phosphorus concentrations indicating the dominance of microscale processes, but none of our covariates were strongly associated with the residual spatial structure of chlorophyll *a*. Our model showed that the spatial dependence structure of environmental response variables shifts across space. It also helped to explain this structure using ecologically relevant covariates from different scales whose effects can be interpreted intuitively. This provided novel insight into the processes that lead to eutrophication, a complex and pervasive environmental issue.

Key words: *chlorophyll* a; *eutrophication; LAGOS database; non-stationary model; phosphorus.*

Introduction

Integrating information across scales is a fundamental problem in ecology (Levin 1992). Because ecological processes are driven by factors at multiple, interacting spatial scales, they often exhibit complex spatial dependence structure. Dealing with this structure in a replicable way is becoming increasingly important as the amalgamation of ever larger datasets (e.g. NEON, GLEON, LAGOS, LUCAS, etc.) allows researchers to ask questions at broad spatial scales and fine tune the scale of an analysis. Choosing a single scale for any ecological analysis is risky because it limits inference and may produce biased estimates (Denny et al. 2004, McCullough et al. 2019). Inference from analyses at a single scale is particularly risky if cross-scale interactions exist (Peters et al. 2004, Soranno 2014). For example, the relationship between wetland cover in the riparian lake buffer and lake total phosphorus concentrations is scale dependent (Fergus et al. 2011). As a result, large scale analyses (spanning six States in the United States) indicate that local % wetland has no effect on lake total phosphorus concentrations, but regional analyses show that this effect is negative in regions with low levels of agricultural activity and positive in regions with high levels of agricultural activity (Soranno et al. 2014). Presenting patterns across all scales that may affect a given process might mitigate these risks (Denny et al. 2004), but creating a framework with enough flexibility to estimate and describe these patterns in an intuitive way remains a challenge. Methods for integrating scales have improved our

Manuscript received 25 August 2020; revised 31 March 2021; accepted 18 May 2021; final version received 23 September 2021. Corresponding Editor: Mark Risser.
[4] E-mail: charlotte.narr@siu.edu

understanding of the ecosystem consequences of many phenomena, for example, evolution, biogeochemical fluxes, and biogeography, but the continued improvement and adoption of these methods is a critical component of ecosystem management (Chave 2013).

One common way for ecologists to handle issues of scale is to integrate information across scales and then account for spatial or temporal autocorrelation. However, when spatial or temporal autocorrelation is accounted for without investigating its source, large amounts of unexplained spatial variation often remain, and important ecological insights are difficult to extract (Fergus et al. 2016). Analyzing the spatial structure that many methods work to "look beyond" can increase the explanatory power of ecological models, improving our understanding of ecological relationships (Pickett and Cadenasso 1995). For example, Lapierre et al. (2018) explicitly analyzed the spatial structure of variables describing water quality and environmental predictors of water quality. They found that the strength of the relationship between water quality and environmental predictors increased as the similarity between their spatial structures increased, suggesting that patterns in spatial structure both induce and reflect ecologically important information.

How can ecologists integrate data across scales in a way that retains the important information hidden within a dataset's spatial dependence structure? One useful option is to analyze the factors that shape spatial structure at the same time variation across sites is analyzed. In other words, integrating across scales can be facilitated by analyzing the covariance (which describes the spatial dependence structure) along with the variance (which describes the mean variance across sites). Analyzing spatial structure is challenging because the range, or extent of spatial correlation among sites, can vary across a given study region. This probably violates a key assumption of many spatial statistics methods used by ecologists, second-order stationarity, which assumes that the specified spatial correlation structure does not vary across the study region (Fortin and Dale 2016). Many ecological phenomena are also heavily influenced by directional processes, such as wind and elevation, that may violate another assumption of many spatial models, isotropy. Isotropy assumes that the strength of spatial dependence is independent of direction. Conversely, anisotropy occurs when the strength of spatial dependence varies with direction. For example, if the prevailing wind direction is east, we would expect that sites located on the west-to-east axis are more correlated than those on the north-to-south axis. However, even under stationary anisotropic models, an identical "prevailing" direction is used throughout the study region, which is often unrealistic for many ecological processes. Therefore, accurate modeling of many ecological phenomena requires non-stationary spatial models that relax these assumptions and allow the covariance structure – in addition to the mean – to vary across the study region.

Following seminal work by Higdon et al. (1999) and later by Paciorek and Schervish (2006), non-stationary models have been applied to a variety of complex environmental processes. For example, Schmidt and Guttorp (2011) used non-stationary spatial models to study solar radiation in British Columbia, Canada and mean temperatures in Colorado, USA. More recently, non-stationary models have been used to study annual precipitation: Ingebrigtsen and Lindgren (2014) focus on Norway, whereas Xu and Gardoni (2018) work with data collected in Colorado, USA. Unfortunately, many of these methods are notoriously difficult to estimate and can be difficult to interpret, which limits their utility to non-expert users.

Recent progress in statistical methods has improved the ease of estimation and interpretability of non-stationary models, making these new models well suited to integrate ecological information across scales. The spatial dependence structure of an environmental process is typically modeled by a covariance function into which elements of non-stationary correlation can be incorporated. Models that build on the intuition of mean regression by adding environmental predictors to the covariance function are particularly useful in investigating the relationship between ecological covariates and properties of spatial covariance (Calder 2008, Reich et al. 2011, Vianna Neto and Schmidt 2014, Risser and Calder 2015). When data from different scales are included as covariates, these models effectively integrate information across scales. For example, Risser and Calder (2015) use a Bayesian covariate-driven non-stationary model to describe the effects of elevation and east–west slope on the spatial dependence structure of precipitation in Colorado. Similarly, Vianna Neto et al. (2014) develop a non-stationary model specifically for directional covariates (e.g., prevailing wind direction) and applied it to ozone pollution in the eastern United States. Risser and Turek (2020) develop an R package (*BayesNSGP*) that focused on making non-stationary model computation easier for moderately sized data sets, for example, annual precipitation for the entire continental United States. These projects illustrate the ongoing interest in non-stationary spatial processes and the potential of non-stationary models for integrating information across multiple scales, but these have very rarely been adopted by ecologists (but see Schmidt and Rodríguez 2011 and Schmidt and Rodríguez 2015). The paucity of non-stationary models in ecology may be because these models can be difficult to implement by non-statisticians, because model parameters can be difficult to interpret (but see Risser and Turek 2020 for improvements in interpretability), or simply because these models are not well known to ecologists.

Application

In this paper, our objective is to facilitate the use of fully non-stationary models in ecology. We do this by integrating and building on the models of Vianna Neto et al. (2014) and Risser and Calder (2015). We reformulate covariate-driven non-stationary models to emphasize the interpretability of model parameters that measure the influence of ecological variables within the spatial covariance portion of the model. In contrast with some previous work, we implement linear relationships between environmental covariates and the anisotropy angle and ratio using typical log and logit link functions. The linearity of these relationships is important because anisotropy angles and ratios can be increased and reduced by changing covariate values. Our covariates inform the shapes of local spatial correlation that are easy to visualize. Specifically, following visualization in Higdon et al. (1999) and Paciorek and Schervish (2006), we characterize variation in spatial dependence using locally stationary anisotropy ellipses. The orientation and size of the ellipses are determined by the effect of ecologically relevant covariates on the structure of the spatial dependence. These ellipses represent how data at each site correlate with their surroundings, thereby reflecting the directionality of the underlying spatial processes. Our models are estimated in a fully Bayesian setting using the adaptive No-U-Turn Hamiltonian Monte Carlo (NUTS HMC) method implemented in Stan 2.17 (Carpenter et al. 2017) in R v.3.5 language (R Core Team 2019). In the context of complex correlated data, HMC has been shown to consistently outperform traditional MCMC samplers (Metropolis-Hastings and Gibbs) in terms of effective samples per minute and model diagnostics (Monnahan and Thorson 2017, Monnahan and Kristensen 2018, Nishio and Arakawa 2019).

To further facilitate the use of fully non-stationary models among ecologists, we demonstrate the application of our reformulated model and evaluate its robustness and utility by using it to describe the spatial dependence structure of a complex and pervasive environmental issue, eutrophication. By evaluating the performance of our model under different datasets and parameterizations, we illustrate the model's strengths and limitations for different types of ecological datasets. Eutrophication, caused by nutrient inputs to freshwater systems from human activities such as agriculture and urban development, is responsible for the deterioration of coastal water quality and 20% of impaired river or stream miles (Howarth and Sharpley 2002, Gilinsky et al. 2009). Human activities contribute nutrients to freshwater at multiple scales from point sources, such as wastewater treatment plants, to N deposition that can affect entire regions. Lakes in some watersheds are also more sensitive to nutrient inputs than others (Stoddard et al. 2016). As a result, freshwater nutrient concentrations have dependence structures that are likely to vary over space. Understanding how these spatial dependencies structure the interaction between nutrient sources and watershed characteristics to predict lake nutrient concentrations and trophic status is a persistent challenge in ecosystem management.

The spatial dependence structure of nutrient concentrations in lakes probably reflects the main drivers of nutrient inputs into lakes, i.e., watershed transport capacity and land use. Generally, when transport capacity is low, fine-scale or within-lake properties exert a stronger influence on nutrient concentrations, but when transport capacities are high, larger scale or watershed level activities can become more important (Fraterrigo and Downing 2008). Therefore, urban areas often serve as nutrient sources to nearby lakes but, in areas where baseflow is high, the nutrients from urban areas may travel further, contaminating lakes that are more distant from cities. This high baseflow could create zones of increased homogeneity in the spatial dependence structure of nutrients in lakes down gradient from other lakes. The effect of baseflow on the spatial structure might then be well represented by a thin ellipse (an ellipse with a high ratio of length to width) angled in the direction of flow.

To examine this possibility, we applied our non-stationary model to describe the spatial dependence structure of lake nitrogen (N), phosphorus (P), and chlorophyll *a* (from this point forwards "Chl*a*") concentrations, as well as N:P ratios across two different regions in the United States: first, a Midwestern region including the States of Iowa, Wisconsin, and Illinois, and second, a Northeastern region including the State of New York. Investigating the spatial dependence structures of these four aspects of water quality in two distinct regions allowed us to examine the performance of our model across eight datasets and four response variables.

We also explored the role of covariate choice in model performance. We suspected that our non-stationary model would perform better when the spatial covariance function was parameterized with covariates that exhibit strong spatial structure. To examine this possibility, we parameterized the spatial dependence structure of each response variable in each region with the same three sets of covariates. This design allowed us to assess model performance across covariance parameters (holding the variance portion of the model constant) and within covariance parameters (allowing the covariance portion of the model to vary). Because we expected the spatial dependence structure of lake nutrients to reflect the combined effects of watershed transport capacity and land use, we parameterized the spatial covariance portion of our models with covariates representing combinations of both the movement (baseflow) and potential sources of nutrients within a watershed. In each location, and for each response variable, we estimated the spatial covariance structure using three different sets of variables: (1) baseflow and total N deposition in the watershed, (2) baseflow and the percent urban land use in each watershed, and (3) baseflow and the percent of forest cover in each watershed (i.e. not developed for urban or agricultural activities). Finally, we conducted analyses to determine if the spatial structure of these covariates affected our ability to estimate their effects within the context of covariate-driven non-stationary models.

Methods

Stationary spatial models

The least complex stationary spatial models are isotropic, where the extent of spatial correlation – measured by the range parameter ρ inside the covariance function – remains constant in all directions and at all locations (Fig. 1a). This produces circular, concentric correlation contours, with stronger spatial correlation near the locations where data have been observed. Real-world spatial processes are rarely truly isotropic, however isotropic models may serve as useful approximations of more complex processes, especially if the study region is relatively small and geographically homogeneous.

Stationary anisotropic models (Fig. 1b) are a natural extension of isotropic models, where we elongate the correlation contours at all locations by the so-called anisotropy ratio (λ) along some direction (θ), both of which are estimated from the data. Isotropy is a special case of anisotropy, in which the anisotropy ratio, which is the ratio of major-to-minor axis of each ellipse, is set to 1. Although anisotropic models are theoretically more general, the assumption that there exists one identical predominant direction across the study region is often tenuous. This assumption probably holds true only if the study region is relatively small and the geophysical mechanism that induces directionality is well defined, for example, air pollution and prevailing wind patterns.

More formally, in the context of stationary anisotropic models, the correlation structure at each site ($\boldsymbol{s}$) is described by an ellipse $\boldsymbol{\Sigma}(\boldsymbol{s})$, where all intersite distances are rotated counterclockwise through an angle θ and then compressed in that direction by the anisotropy ratio λ. By compressing distances in direction θ, we thereby induce greater spatial correlation in that direction, as spatial correlation functions intuitively assume that correlation is greater at small intersite distances. By definition, stationary models assume that θ and λ do not change with site, so we can write $\boldsymbol{\Sigma}(\boldsymbol{s}) = \boldsymbol{\Sigma}$ for all sites. We parameterize anisotropy ellipses following Schabenberger and Gotway (2005) and Bass and Sahu (2017), such that:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

As λ approaches 1, the ellipses become less elongated and start to resemble circles, and so spatial dependence approaches isotropy. Consistent with this, when $\lambda = 1$, anisotropy angle θ is not estimable. Distances are computed on the rotated and scaled coordinates and these transformed distances are input into a spatial correlation function, for example the exponential, such that $\rho(h^*, \delta) = \exp(-\delta h^*)$, where $h^*_{ij} = \sqrt{(\boldsymbol{s_i} - \boldsymbol{s_j})^T H (\boldsymbol{s_i} - \boldsymbol{s_j})}$ with $H = \begin{pmatrix} \lambda^2 \cos^2\theta + \sin^2\theta & (\lambda^2 - 1)\cos\theta\sin\theta \\ (\lambda^2 - 1)\cos\theta\sin\theta & \cos^2\theta + \lambda^2\sin^2\theta \end{pmatrix}$.

The likelihood of the outcome variable $Y(\boldsymbol{s})$ is assumed to be multivariate Normal: $Y(\boldsymbol{s}) \sim \mathrm{MVN}\left(X(\boldsymbol{s})\boldsymbol{\beta},\ (\boldsymbol{\Omega})_{ij} = \alpha^2\rho\left(h^*_{ij},\ \delta\right) + \tau^2\mathbf{I}\right)$ with spatially

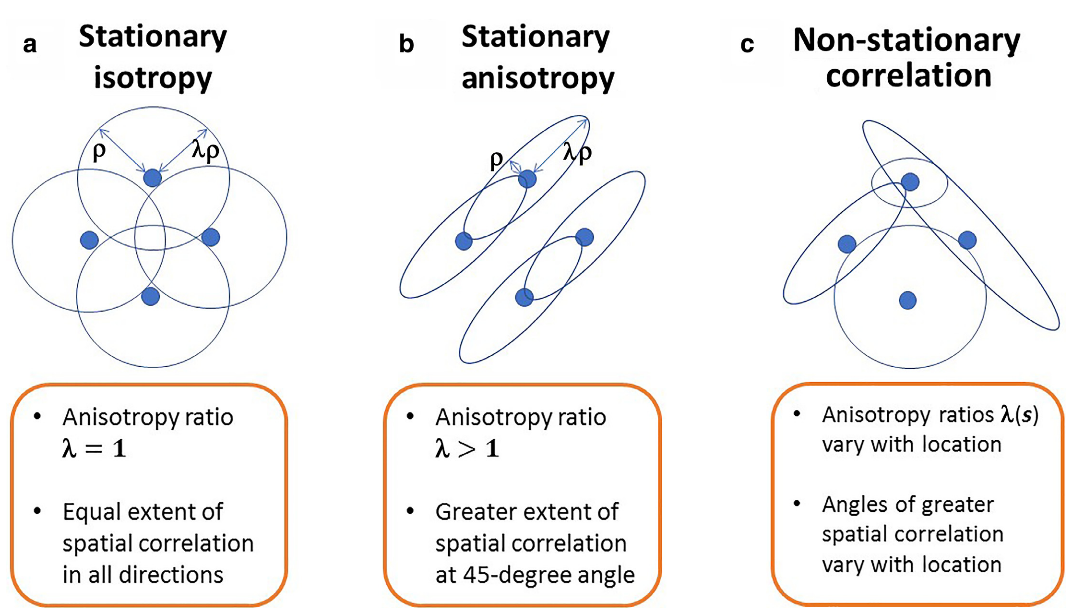

Fig. 1. Stylized examples of spatial correlation contours with four locations under stationary isotropy (a), stationary anisotropy (b), and non-stationarity (c). Anisotropy ratio (lambda) and baseline range (rho) are labeled on the figure.

varying mean $X(s)\beta$ and an $N \times N$ variance–covariance matrix $\mathbf{\Omega}$.

Non-stationary spatial models

There are many potential reasons why non-stationary spatial correlation arises. Among these, the most common reason is that the study region is sufficiently large such that a homogeneous (i.e., stationary) spatial correlation structure is no longer feasible (e.g., Sampson 2010). Non-stationary models are estimated specifically to relax the assumption that $\mathbf{\Sigma}(\boldsymbol{s}) = \mathbf{\Sigma}$, allowing the anisotropy ellipse – and therefore the direction and strength of spatial dependence – to vary at each site ($\boldsymbol{s}$) (Fig. 1c). This is akin to additionally allowing the variance–covariance matrix to depend on site inside the likelihood function: $Y(\boldsymbol{s}) \sim \text{MVN}(X(\boldsymbol{s})\boldsymbol{\beta},\ \mathbf{\Omega}(\boldsymbol{s}))$. The estimation task becomes considerably more complex, in which we must now estimate one anisotropy ellipse per location. Furthermore, methods that introduce this level of flexibility into the spatial correlation structure while retaining model interpretability tend to be relatively uncommon (Kleiber and Nychka 2012).

Here, we extend ideas recently introduced by Risser and Calder (2015) and Vianna Neto et al. (2014), who suggest using spatially indexed covariates to inform how anisotropy varies in space. Risser and Calder (2015) adapt a covariance regression (e.g., see Hoff and Niu 2012) to construct the spatial correlation kernel at each spatial location as a quadratic function of observed covariates. Vianna Neto et al. (2014) allow spatially varying directional effects by setting the angle of greater spatial correlation to be a deterministic function of a vector-valued directional covariate. Both papers operate under a Bayesian estimation paradigm by sampling from the posterior distribution using a Gibbs sampler. Furthermore, both papers feature relatively small spatial datasets with 195 and 48 locations, respectively, ostensibly to avoid the well known "big N problem" in spatial statistics (e.g., Gelfand and Banerjee 2017). Therefore, we introduce spatially varying angles as $\theta(\boldsymbol{s})$ and spatially varying ratios as $\lambda(\boldsymbol{s})$ and specify hierarchical linear processes for each one, such that:

$$\pi \times \text{logit}^{-1}(\theta(\boldsymbol{s})) \sim N(\alpha_T + X_T(\boldsymbol{s})\boldsymbol{\beta}_T, \psi_T)$$
$$\text{and} \log(\lambda(\boldsymbol{s})) \sim N(\alpha_L + X_L(\boldsymbol{s})\boldsymbol{\beta}_L, \psi_L)$$

where the symbol $\sim$ is read "distributed as," parameters α_T and α_L serve as "baseline" angle and ratios, and matrices $X_T(\boldsymbol{s})$ and $X_L(\boldsymbol{s})$ are design matrices for angles and ratios, respectively. In introducing these design matrices, we gain the ability to investigate the association between any general covariates and the spatially varying angles and ratios separately. Finally, we note that with our specification, we aim to disentangle systematic variation (via the mean structure of the two Normal distributions) from random variation, or variation due to noise (via the two "error" standard deviations ψ_T and ψ_L).

By specifying linear processes for each spatially varying anisotropy parameter, we can understand the impact of covariates using familiar GLM-like interpretations. For example, if the coefficient $\beta_{L_1} > 0$, this implies that the anisotropy ratio – and consequently the size of the ellipse and the geographic extent of spatial dependence – increase by $\exp\left(\beta_{L_1}\right)$ for each unit change in the covariate x_1. The intercepts for each process reflect the "baseline" anisotropy ratio and angle across the region. With centered covariates, this "baseline" would be the estimated anisotropy ratio and angle over locations with average values for all covariates. Note that we are free to specify quadratic terms and interactions inside the two design matrices, in addition to using both quantitative and categorical covariates. Furthermore, we can suggest evidence of statistically significant non-stationarity by recording whether 0 is inside the credible interval for either β_T and β_L. More formally, one could also designate a Region of Practical Equivalence (ROPE), for example a range between −0.1 and 0.1 (Kruschke 2014), and perform a formal Bayesian hypothesis test. Of course, we could encounter the case where there exists non-stationarity in angles, but not in ratios, and vice versa. Finally, stationary models are a special case of our model: we recover anisotropy when $\beta_L = 0$, $\beta_T = 0$, and the "baseline" $\lambda > 1$; we recover isotropy when $\beta_L = 0$, $\beta_T = 0$ and the "baseline" $\lambda = 1$.

The anisotropy ellipse at each location is a function of $\theta(\boldsymbol{s})$ and $\lambda(\boldsymbol{s})$, such that:

$$\mathbf{\Sigma}(\boldsymbol{s}) = \begin{pmatrix} \cos[\theta(\boldsymbol{s})] & -\sin[\theta(\boldsymbol{s})] \\ \sin[\theta(\boldsymbol{s})] & \cos[\theta(\boldsymbol{s})] \end{pmatrix} \begin{pmatrix} \lambda(\boldsymbol{s}) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos[\theta(\boldsymbol{s})] & \sin[\theta(\boldsymbol{s})] \\ -\sin[\theta(\boldsymbol{s})] & \cos[\theta(\boldsymbol{s})] \end{pmatrix}$$

Following Paciorek and Schervish (2006), once each $\mathbf{\Sigma}(\boldsymbol{s})$ is defined and assuming a stationary variance $\alpha^2 + \tau^2$, we combine (more formally: convolve) N such ellipses into a positive-definite non-stationary variance–covariance matrix through:

$$(\mathbf{\Omega})_{ij}^{NS} = \alpha^2 |\Sigma(s_i)|^{\frac{1}{4}} |\Sigma(s_j)|^{\frac{1}{4}} |(\Sigma(s_i) + \Sigma(s_j))/2|^{-\frac{1}{2}} R^S\left(\sqrt{Q(s_i,\ s_j)}\right) + \tau^2 I$$

$$Q(s_i,\ s_j) = (s_i - s_j)^T \left((\Sigma(s_i) + \Sigma(s_j))/2\right)^{-1} (s_i - s_j)$$

where R^S is any stationary isotropic spatial correlation function, α^2 is the spatial variance, and τ^2 is the nugget variance. We use the usual exponential spatial correlation function throughout this manuscript. The likelihood of the outcome variable is again assumed to be multivariate Normal with mean $X'(s)\boldsymbol{\beta}$ and a non-stationary variance–covariance matrix $\boldsymbol{\Omega}^{NS}$.

To further clarify our estimation strategy, we combine all parameters to be estimated into the vector $\boldsymbol{\Psi} = \{\boldsymbol{\beta}, \alpha_T, \boldsymbol{\beta}_T, \alpha_L, \boldsymbol{\beta}_L, \psi_T, \psi_L, \alpha^2, \tau^2\}$, which parameterizes the mean function, the covariance function, the spatial variance, and the nugget variance, respectively. Our models focus on the marginal likelihood $f(Y|\boldsymbol{\Psi})$, which marginalizes over the N spatial effects and has a closed form for the multivariate Normal: $Y(s) \sim \text{MVN}\left(X'(s)\boldsymbol{\beta},\ \boldsymbol{\Omega}^{NS}\right)$. The posterior distribution is then specified as: $p(\boldsymbol{\Psi}|Y) \propto f(Y|\boldsymbol{\Psi}) \times p(\boldsymbol{\Psi})$, where the priors are assumed to be *a priori* independent. Relative to the conditional likelihood, working with the marginal likelihood is known to reduce the computational burden and produce a computationally better behaved variance–covariance matrix (Banerjee and Carlin 2015).

To contrast our model specification against others, in the Risser and Calder (2015) model, covariates interact via a quadratic function with the entire anisotropy ellipse via a covariance regression: not anisotropy angles and ratios separately. This means that the coefficients in Risser and Calder's model are only estimable up to a constant. Therefore, we would be unable to discern positive associations from negative associations, for example. Risser and Calder allow the variance (α^2) to be non-stationary and itself depend on a set of parameters, an extension we do not pursue in the current paper. The model in Vianna Neto et al. (2014) allows only the angles to vary in space as a deterministic function of directional covariates and assumes all other covariance parameters are stationary. Finally, the seminal Paciorek and Schervish (2006) model uses two additional N-dimensional Gaussian processes, each with its own variance and range parameter, to allow anisotropy angles and ratios to vary smoothly in space. Although this approach is intuitive, it is computationally prohibitive, as noted by the authors in their paper.

Data and model parameterization

We applied our statistical model to limnological and geospatial data from the LAGOS-NE$_{\text{LIMNO}}$ v.1.087.3 (Soranno and Bacon 2017, Soranno and Cheruvelil 2019) accessed using the *lagosne* R package (Stachelek and Oliver 2017). This database includes limnological information synthesized from 87 State agency, federal agency, university, tribal and citizen science water quality monitoring programs, and has undergone extensive quality assurance and control to ensure data are comparable (Soranno 2015). We used this dataset to explore the utility of our non-stationary model across a range of response variables, datasets, and covariates by parameterizing it in different ways using the following steps: (1) choosing four response variables of major importance for water quality, (2) identifying a suite of potential predictor variables for these response variables from three spatial scales, (3) identifying two different regions within the United States, with an appropriate number of lakes for model exploration, (4) creating a reduced set of uncorrelated predictors for each response variable within each region to use to parameterize the mean (non-spatial) portion of the model, (5) choosing three sets of covariates to parameterize the covariance (spatial) portion of the model. Each of these steps is outlined in detail below.

In step one, we chose four limnological response variables from LAGOS-NE$_{\text{LIMNO}}$ that are related to lake eutrophication: total nitrogen (TN), total phosphorus (TP), chlorophyll *a* (Chl*a*), and the stoichiometric ratio of total nitrogen to total phosphorus (N:P). N:P is related to nutrient limitation of primary production, but it can be more difficult to explain cross-lake spatial patterns in N:P compared with nutrient concentrations (Collins et al. 2017). We calculated the decadal median value for each response variable. In step two, we developed candidate models using predictor variables that have been shown to drive water quality at continental scales (Taranu and Gregory-Eaves 2008, Read 2015, Collins et al. 2017, Lapierre et al. 2018). Predictor variables were generated from LAGOS-NE$_{\text{GEO}}$ v1.05 (Soranno and Cheruvelil 2017), and either related to climate and hydrology at the 8-digit hydrological unit code (HUC 8) watershed scale, land use at the lake watershed scale, or individual characteristics of lakes (e.g., lake depth, lake surface area).

In step three, we identified two distinct regions within the United States that would allow us to examine the success and utility of our non-stationary model in different environmental contexts: the Midwest (Iowa, Illinois, and Wisconsin), and the Northeast (New York). We selected these regions because we wanted to test the model in contiguous areas with data from 100 to 300 lakes, a broad range of water quality conditions, and a broad range of environmental conditions for multiple types of covariates (Fig. 2). To maintain consistency between the two regions while limiting potential multicollinearity, we added a fourth step examining correlations between a single, large set of potential predictors for each region and removing predictors until there were no correlations >0.5. This process produced a set of predictors that included: average HUC 8 baseflow, average HUC 8 TN deposition, % urban land use in the individual lake watershed, % forest land cover in the individual lake watershed, lake area, maximum lake depth, lake connection (whether a lake is isolated, a headwater lake, a drainage lake, or a drainage lake with upstream lakes), and the ratio of lake area to watershed area, which is a common proxy for water residence time. Using this common set of predictors, we conducted ordinary least squares regression (OLS) to determine which predictors explained the most variation in TP, TN, N:P, and Chl*a* in each location and, as a result, belong in the mean portion of the non-stationary model. The parameters used in the mean

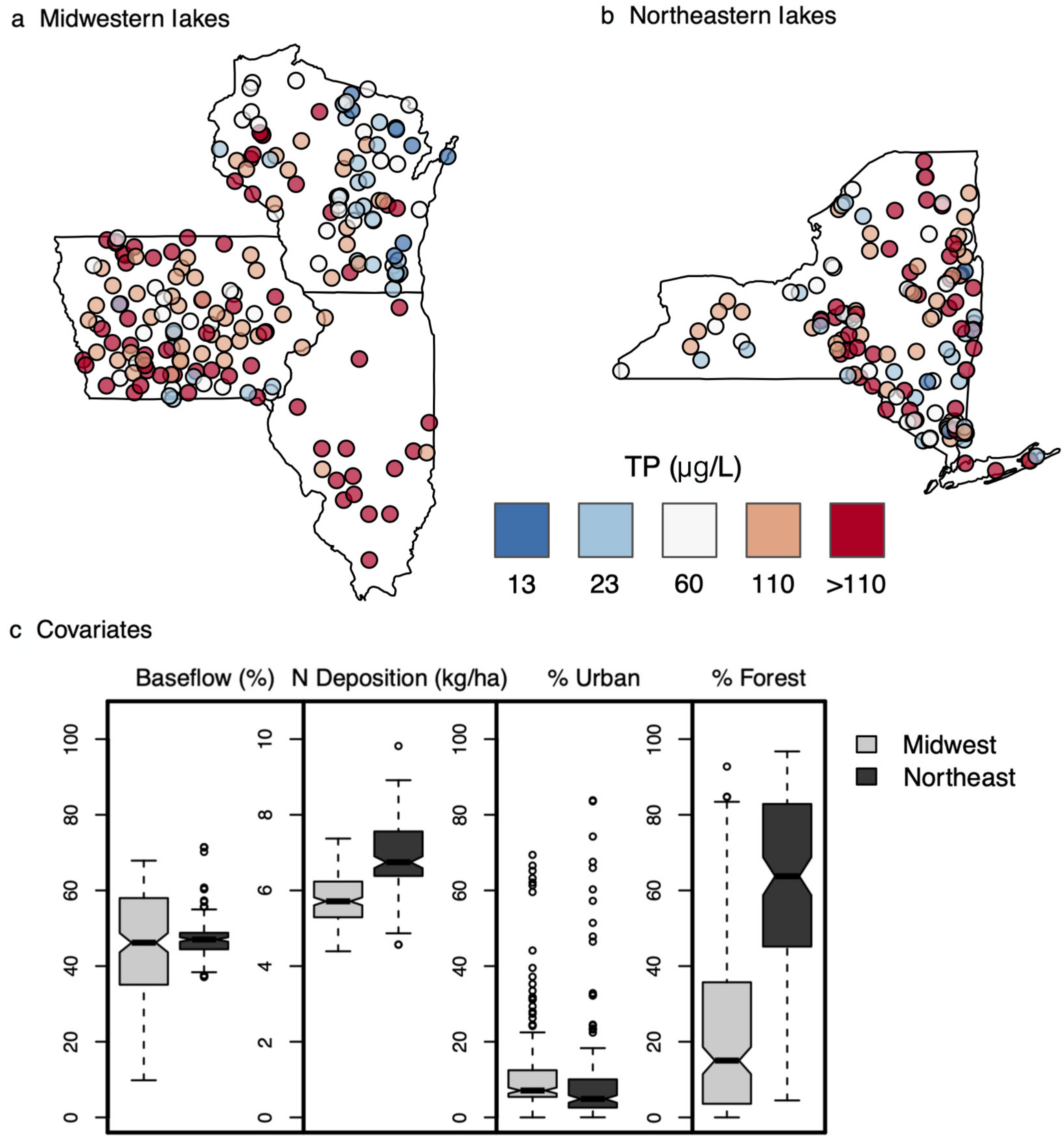

FIG. 2. Map of the two study regions showing total phosphorus (TP) data in each lake (a, b) and boxplot (c) showing the distributions of covariates included in the spatial covariance portion of the model covariate distributions for each study region.

and covariance portions of each model are shown in Table 1.

In the fifth and final step of model parameterization, we created three sets of covariates to include in the non-stationary portion of the model. To facilitate the computation of these models, we limited the number of parameters in the non-stationary portion of the model to two variables that represented the transport and source of nutrients within a watershed. In each set of covariates, baseflow was used to represent the transport of nutrients, but we varied the second parameter to reflect either broad (N deposition) or finer scale (percent land use) watershed processes. Therefore, for each response variable in each region, we held the non-spatial portion of the model constant and parameterized the spatial portion of the model (both the ratio and angle of the spatial dependence structure) with (1) baseflow and total N deposition in the watershed, (2) baseflow and the percent urban land use in each watershed, and (3) baseflow and the percent of forest cover in each watershed (i.e. not developed for urban or agricultural activities). All covariates were centered and standardized.

Interpreting model output

To determine the ability of our model to accurately estimate the effects of covariates on the spatial structure of the residuals, we ran our Bayesian models for 1,000

TABLE 1. Predictor variables from LAGOS-NE$_{GEO}$ v1.05 included in the stationary portion of the model for each response variable and region.

Response	Region	Fixed effects
TN	MW	N deposition, Watershed % forest, Lake2 Depth, Lake Width to area ratio
	NE	N deposition, Watershed % forest, Lake Depth, Lake Width to area ratio
TP	MW	Baseflow, Watershed % forest, Lake Depth, Lake Connectivity
	NE	N deposition, Watershed % forest, Lake Depth, Lake Width to area ratio
Chl*a*	MW	Baseflow, Watershed % forest, Lake Depth, Lake Connectivity, Total Lake N and P
	NE	Baseflow, Watershed % forest, Lake Width to area ratio, Lake Total N
N:P	MW	Baseflow, Lake Depth, Lake Connectivity
	NE	N deposition, Baseflow, Watershed % forest, Lake Width to area ratio

Notes: We selected specific predictors for each parameterization to avoid multicollinearity and to explain the most variation for each response variable within a region according to ordinary least squares regression. Each of these stationary portions of the model was run with three different non-stationary portions of the model. The non-stationary portion of the model (both the angle and ratio) was specified by (1) baseflow and total N deposition in the watershed, (2) baseflow and the percent urban land use in each watershed, and (3) baseflow and the percent of forest cover in each watershed (i.e. not developed for urban or agricultural activities).

iterations following a 300 iteration warm-up period for each model, using an "adapt delta" parameter of 0.9. For models that were used for ecological inference, these specifications produced effective sample sizes that ranged between 137 for percent urban's effect on the anisotropy angle and 2952 for maximum lake depth's effect on the mean.

Information criteria (e.g., Akaike information criterion [AIC], Bayesian information criteria [BIC], deviance information criterion [DIC], Watanabe-Akaike information criterion [WAIC]) are routinely used to compare statistical models and select one that is expected to perform best for the purpose of out-of-sample prediction. However, lakes included in the LAGOS-NE database are known to be a biased sample of all North American lakes (Stanley et al. 2019), making it generally inappropriate to make predictions on out-of-sample lakes based on these data. In addition, interpolation to unobserved lakes is not our main research focus for the current paper. In Bayesian data analysis, even when prediction is not the goal, information criteria can offer clues about the robustness of model specification and the informativeness of priors. To that end, we computed the WAIC (Watanabe 2013) for the models that converged, in addition to other indicators of MCMC performance (R-hat, Effective Sample Size, divergent transitions). For ecological inference, we validated the results of models with low effective sample sizes by doubling the total number of iterations. Code for accessing data, running models, and making figures is available (see *Open Research*). Priors for each parameterization of the model are shown in Appendix S1: Table S1.

To compare the performance of the model across each parameterization, we examined three metrics of model performance: the number of divergent transitions, the smallest effective sample size (across all parameters), and the largest R-hat statistic (across all parameters). Divergent transitions can indicate potential pathological areas in the posterior due to a problematic combination of likelihood and priors and/or non-identifiability of model parameters. Common problematic areas comprise funnel-like shapes or, more generally, any region of extreme curvature in the posterior (Betancourt 2017). Model specifications that produce many divergent transitions indicate certain areas of the posterior have not been sampled and therefore invalidates further inference. A few divergent transitions can still occur for trivial numerical errors in computing the posterior and therefore can still lead to valid inference. Small effective sample sizes reflect autocorrelation in successive MCMC samples and are detrimental to posterior inference, often indicating that chains are not efficiently exploring the posterior. Finally, large R-hat statistics can indicate a lack of convergence in MCMC chains, further invalidating posterior inference. Therefore, it is desirable to have the smallest number of divergent transitions, the largest effective sample size, and R-hat statistics that are closest to 1.

We suspected that our non-stationary model would perform better when the spatial covariance function was parameterized with covariates that exhibit strong spatial structure. We tested this prediction by first estimating the Effective Degrees of Freedom (EDF) and % deviance explained of a Generalized Additive Model smoothing term on the spatial coordinates for each covariate, then we graphically compared the EDF and % deviance of each covariate to the three metrics of model performance described above.

RESULTS

MCMC performance

Our metrics (number of divergent transitions, smallest effective sample size, and the largest R-hat) suggest that there is substantial variation in the ability of our approach to reliably estimate the effects of covariates on the residual spatial structure of our selected water quality variables. This variation in model performance was not associated with any specific region, water quality response variable (i.e., TP, TN, Chl*a*, or N:P), or covariate (N deposition, baseflow, or watershed land use/-cover). However, the number of divergent transitions was lower for covariates with stronger spatial structure, as indicated by the EDF and % deviance explained of a

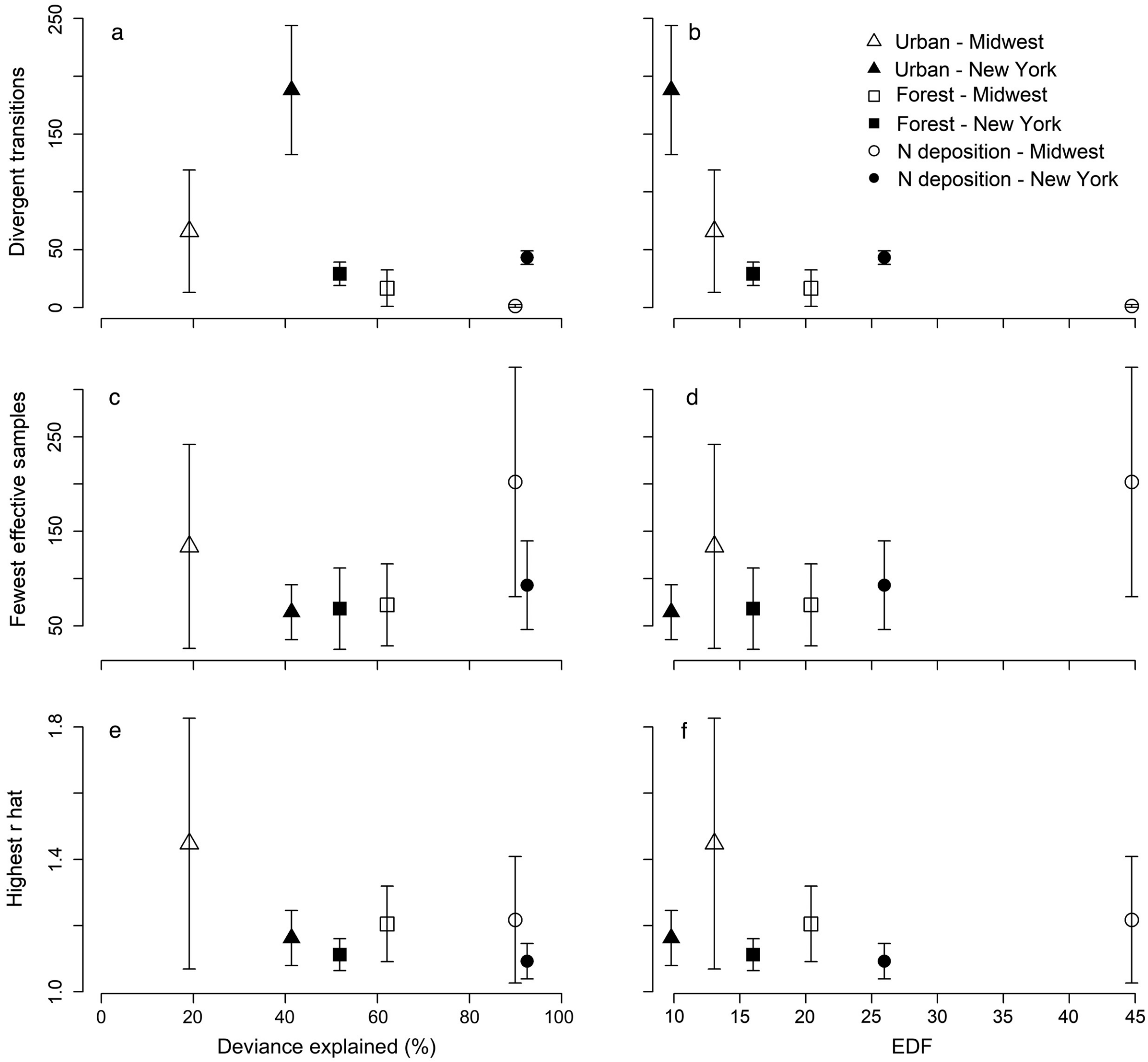

FIG. 3. (a–f) Three metrics of model success as a function of the Effective Degrees of Freedom (EDF) and percent deviance explained by a spatial smoothing term for each covariate estimated via a General Additive Models (GAMs). Covariates with higher EDFs and % deviance explained reflect the presence of stronger spatial structure in the covariate values. Generally, confidence in model parameter estimates increases with fewer divergent transitions, higher effective sample sizes, and lower R-hats. Error bars represent standard errors across all indicators of water quality under consideration.

Generalized Additive Model smoothing term estimated for each covariate (Fig. 3a, b). Furthermore, there was no clear relationship between the lowest effective sample size, the highest R-hat and the strength of the covariates' spatial structure (Fig. 3c–f).

Model comparison using WAIC

We used WAICs computed based on marginal log-likelihood for those models that converged to produce reliable results (Table 2). Compared with their stationary counterparts, the non-stationary models attain a lower marginal WAIC in all cases except for Chl*a* in the Midwest region. In that case, misspecification of the non-stationary model is evident by the effective number of parameters being unrealistically large. In other cases, we show that non-stationarity among TP covariances is best explained by baseflow and % urban in New York and by baseflow and % forest in the Midwest. This result is intuitive for New York because the credible intervals for the two coefficients comfortably exclude 0 (Fig. 4a). For the Midwest, only the credible interval for % forest excludes 0, but this model still attains the lowest WAIC among the three specifications tested. In all cases, the effective number of parameters for the non-stationary models is somewhat overestimated. Whereas the spatial effects do not count toward effective parameters inside the marginal WAIC, more research is needed to

TABLE 2. Marginal Watanabe-Akaike Information Criteria (WAIC) for models that have converged.

		Non-stationary model			Stationary model‡		
Response†	Covariance structure	ELPD	p_{WAIC}	WAIC	ELPD	p_{WAIC}	WAIC
New York							
Chl*a*	baseflow + % forest	−170.1	12.1	**340.2**	−178.6	4.4	357.3
TP	baseflow + % forest	−734.4	722.1	1468.8			
TP	baseflow + N dep.	−321.1	280.7	642.2	−141.8	4.0	283.6
TP	baseflow + % urban	−136.5	32.5	**273.1**			
Midwest							
Chl*a*	baseflow + % forest	−1908.1	1767	3816.3	−413.6	5.8	**827.1**
TP	baseflow + % forest	−325.8	22.5	**651.5**			
TP	baseflow + N dep.	−335.8	12.5	671.6	−346.8	5.1	693.7
TP	baseflow + % urban	−334.6	19.2	669.1			

Notes: Our reformulated non-stationary models are compared vs. stationary model for two regions (Midwest and New York), two response variables (Chl*a* and TP), and three covariance structure specifications (baseflow + forest, baseflow + N deposition, baseflow + urban). ELPD is the Expected Log Predictive Density; p_{WAIC} is the effective number of parameters; WAIC is the marginal WAIC. The model that attains the lowest WAIC for each response variable is bolded.

† All outcome variables were log-transformed prior to analysis.

‡ Stationary model has no covariates in its covariance function, therefore only one set of values per response variable is provided.

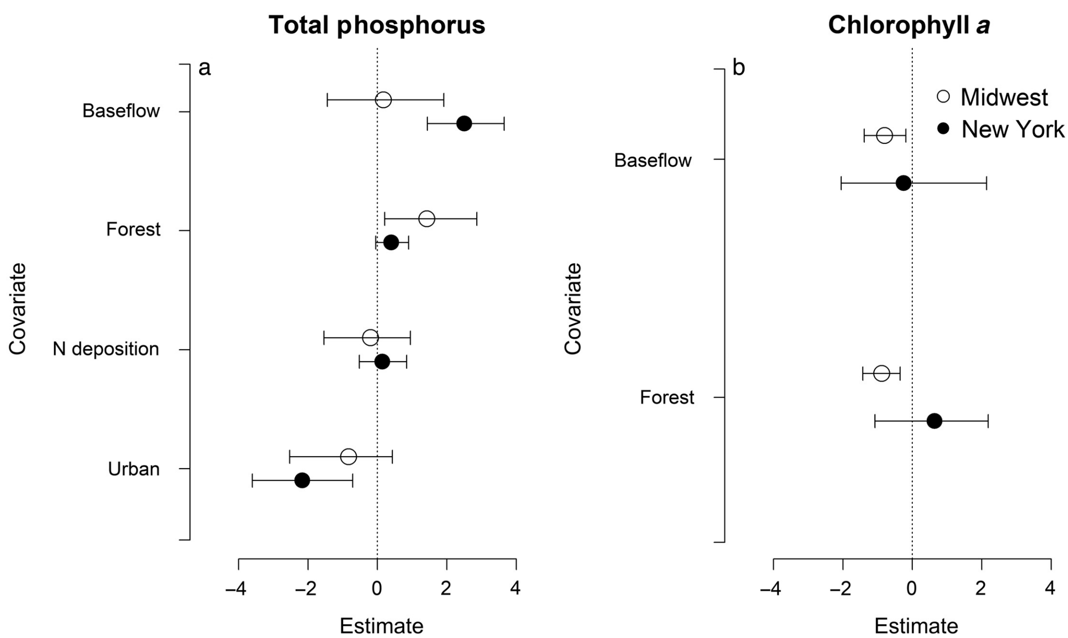

FIG. 4. Association of environmental covariates on the residual spatial structure of lake total phosphorus (TP) (a) and chlorophyll *a* (Chl*a*) (b) in the Midwestern and Northeastern regions. Higher estimates reflect the presence of larger anisotropy ellipses, indicating greater spatial homogeneity. Error bars indicate upper and lower 90% Highest Probability Density Intervals. Baseflow was included as a covariate in all models, but the effect shown on TP is estimated from a model that includes watershed urban land use as a covariate. Baseflow estimates from the other models for TP are not shown because they did not produce reliable estimates (effective sample size < 100, and R-hat values >1.03). Similarly, N deposition and urban land use are not shown for Chl*a* because our models did not produce reliable estimates for these covariates.

understand how parameters inside the covariance function count toward the total number of effective parameters.

Microscale and macroscale patterns in lake water quality

Our reformulated non-stationary model described the effects of ecologically relevant covariates on the residual spatial structure of TP and Chl*a*. For TN and N:P, the parameter estimates in the covariance function were either unreliable (effective sample size < 100, and R-hat values >1.03) or uninformative. The estimated effects of baseflow, N deposition, and watershed percent forest on N:P ratios in the Northeastern region overlapped with zero. We focus on TP and Chl*a* for the remainder of the results. Residual spatial structure of TP and Chl*a* varies

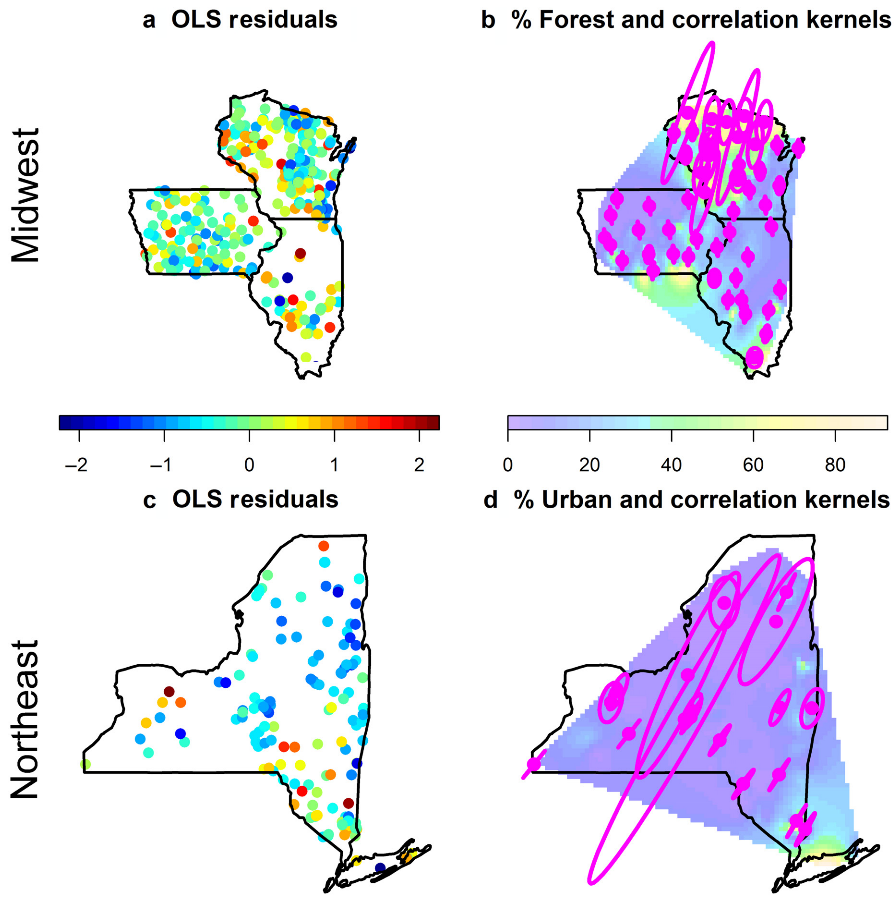

FIG. 5. Residual spatial process of lake total phosphorus (TP) in the Midwestern region (a) and Northeastern region (c). Ellipses generated by our covariate-driven non-stationary model reflect the estimated effect of watershed percent forest in the Midwest (b) and baseflow and watershed percent urban in the Northeast (d). Percent forest (b) and percent urban (d) are shown with ellipses for context.

markedly across the Midwest and New York. Our model describes this variation using ellipses of locally stationary kernels that vary in size and angle of orientation as a linear function of different environmental covariates (Figs. 5, 6).

Our model indicates that the covariates that explain non-stationary correlation in the spatial structure of lake TP and Chl*a* in the Midwestern region are different from those in the Northeast (Fig. 4). For TP in the Midwest, the shapes of the ellipses were best described by forested land cover in the lake watershed, such that more forested areas produced larger ellipses, reflecting stronger spatial dependence and therefore greater spatial homogeneity (Fig. 4a). In the Northeast, we found that the shapes of the ellipses were best described by baseflow and urban land use in the lake watershed: higher baseflows and watersheds with smaller percentages of urban areas were associated with larger ellipses and therefore more spatial homogeneity. Conversely, increases in both % forest and baseflow produced smaller ellipses for Chl*a* in the Midwest, indicating weak residual spatial dependence. In other words, the mean function in the Midwest essentially explains almost all geographic variation of Chl*a*. None of our covariates were strongly associated with the residual spatial structure of Chl*a* in the Northeast (Fig. 4b).

DISCUSSION

Consistent with previous work and our expectations, our reformulated non-stationary spatial model showed that the spatial dependence structure of environmental data varies across space. Our model helps to explain this structure using ecologically relevant covariates whose effects can be interpreted intuitively. Because these covariates were taken from multiple scales (e.g. at the watershed and basin levels), our model effectively integrates information across scales. Importantly, this integration of scales provided novel insight into the processes that lead to eutrophication, a complex and pervasive environmental issue. We demonstrated the presence of non-stationary spatial correlation of two

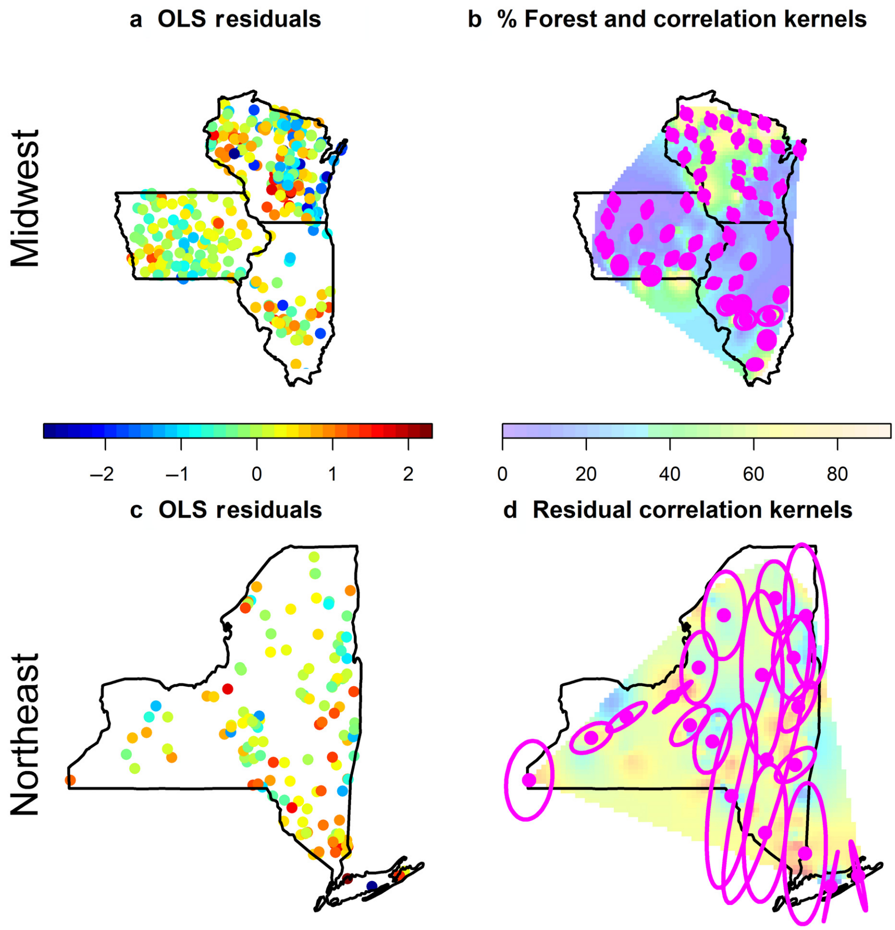

FIG. 6. Residual spatial structure of lake chlorophyll *a* (Chl*a*) in the Midwestern region (a) and Northeastern region (c). Ellipses generated by our covariate-driven non-stationary model reflect the effect of baseflow and watershed percent forest in the Midwest (b) and Northeast (d) on the spatial covariance of the residuals. Percent forest (b) and smoothed residuals (d) are shown with ellipses for context.

indicators of lake water quality (TP and Chl*a*), identified several environmental drivers of this non-stationary correlation, and showed that these drivers are different in the Midwestern vs. Northeastern United States. Our results suggest that the use of stationary models (either isotropic or anisotropic) may be inappropriate for water quality variables and demonstrate how micro and macroscale phenomena work together to shape the water quality of lakes. Our model not only satisfies these assumptions, but it improves our understanding of the relationships between watershed characteristics and indicators of water quality by using these characteristics to explain additional variation in spatial structure.

Enhancing the interpretability of non-stationary models

Our reformulated model detects areas where spatial correlation varies systematically at some regional or continental scale. However, even in non-stationary models, valid variance–covariance matrices require homogenous spatial correlation at very fine scales. In other words, non-stationary models require local stationarity. To this end, Paciorek and Schervish (2006) used Gaussian processes to ensure that their spatial kernels vary smoothly in space. In contrast, our non-stationary model improves interpretability by linking the direction and strength of spatial dependence to environmental covariates via linear models. This essentially mandates that the spatial correlation changes with each covariate and ignores the effect of covariates on local stationarity. Intuitively, covariates that themselves are spatially smooth and therefore exhibit strong spatial structure are better candidates to explain non-stationary correlation in a way that produces a locally stationary process. Conversely, covariates with geographic patterns that resemble checker-board patterns would probably require spatial correlation to change too abruptly to preserve local stationarity. We find evidence to support this assertion with poor model performance when percent urban land use is used inside the covariance function (Fig. 3).

With any Bayesian analysis, priors are important for/ exert a strong influence over parameters that are not well-identified by the data. In our model, and also in our experience with stationary anisotropic models, we find that anisotropy angles are relatively poorly defined compared with anisotropy ratios. Therefore, we recommend experimenting with informative priors for the spatial process for angles, whereas one may leave priors as weakly informative or uninformative for ratios. A major aim of our analysis was to evaluate the robustness of the model, so we kept our priors as weakly informative or uninformative as possible.

Model comparison via information criteria

Model comparison and selection using information criteria – despite being routinely applied to select the "best" model – remains an open question in statistics. Gelman and Hwang (2014) contrast several commonly used criteria (AIC, DIC, WAIC) and conclude that each can be flawed in certain circumstances. For example, the WAIC can be poorly defined with structured data, and the DIC fails (Spiegelhalter et al., 1999, 2014), among other cases, when posterior distributions are poorly summarized by their means. Some researchers go further to question the entire premise of selecting a statistical model based on a single criterion (e.g., see Christian P. Robert's comment published alongside Spiegelhalter et al., 2014).

The task is more complex in the context of spatial models because the analyst must decide whether it is more sensible to use the marginal or the conditional log-likelihood as the basis of their chosen criterion. This choice results in vastly different values and is referred to as the selection of "model focus," which has been comprehensively discussed by Millar (2009) and Celeux, (2006), among others. Computational considerations often influence the analyst's decisions as well: Banerjee et al. (2015) note that marginal likelihoods are often better behaved in spatial models with a Gaussian outcome. However, marginal likelihoods may be intractable when the distribution of the outcome is not Gaussian. To this end, Millar (2018) and Li et al. (2016) recommend using marginal information criteria whenever possible, especially when the number of observations per "cluster" is low, as is common with spatial data.

Among our models that converged, the non-stationary models attained a lower marginal WAIC for all but one response variable. The number of effective parameters – often considered an indicator of misspecification – was expectedly larger in non-stationary models and remained generally close to the number of parameters being estimated. Whereas the spatial effects do not count toward effective parameters inside the marginal WAIC, more research is needed to understand how parameters inside the covariance function count toward the total number of effective parameters. Model evaluation via the conditional WAIC, or the newly developed Leave-One-Out Information Criterion (Vehtari and Gelman 2017) was outside the scope of our work but could be explored in future analyses.

Model inferences for water quality

Our results indicate that the spatial dependence structure of TP concentrations among lakes is best described by watershed percent forest in the Midwestern States of Iowa, Illinois and Wisconsin, but, in the State of New York, it is best described by urbanization and baseflow. Multiple studies link freshwater P inputs to watershed percent forest, urbanization, and baseflow, but we are unaware of any work examining the effects of these watershed characteristics on the spatial structure of P concentrations among lakes. It is generally expected that forested areas have lower levels of anthropogenic P inputs than urbanized ones (Noe and Hupp 2005, Ellison and Brett 2006, Wakida et al. 2014). Our results suggest that lakes in forested areas are also more similar to one another in terms of P concentration than those in urban areas. The increased heterogeneity of lake P concentrations in urban areas suggests that P concentrations in the lakes of urbanized watersheds are governed by fine-scale processes, while those in forested watersheds are controlled by broader scale processes. This is consistent with the well documented diversity of point and non-point P sources in urban landscapes and the complex structural matrix through which these inputs are transported into aquatic ecosystems (Kaushal and Belt 2012). High baseflows appear to counteract the effects of urbanization on the spatial dependence structure of lake P. Baseflow probably increases homogeneity in lakes P by increasing P transport among lakes.

While TP concentrations were more spatially homogenous among lakes in watersheds with high % forest and baseflows, Chl*a* concentrations were more spatially heterogenous among lakes in Midwestern watersheds with these attributes. These results indicate that Chl*a* is governed by finer scale processes than TP in forested, high baseflow catchments, and they highlight the complex relationship between lake nutrient concentrations and trophic status. TP is often directly related to Chl*a* in lakes, but the strength of this relationship is spatially structured (Fergus et al. 2016). This structure may be particularly pronounced in forested water sheds where light limitation reduces the effect of nutrients on chlorophyll (Lowe and Golladay 1986) and wooded wetlands magnify it (Wagner et al. 2011).

We successfully applied this spatial modeling approach for TP and Chl*a*. However, our model did not produce reliable estimates of the effects of covariates on the spatial dependence structure of lake TN in either region or lake N:P ratios in the Midwestern region. Where our model produced reliable estimates for predictors of the spatial dependence of lake N:P ratios (in the Northeast), they overlapped with zero. We were surprised that we were unable to explain the spatial structure of lake TN. Previous studies using both TN and TP as response variables have produced similar results for both elements (e.g., Read et al. 2015, Collins et al. 2017). One potential explanation for the lack of success detecting the spatial dependence structure of lake TN is that the average time series for N data in LAGOS-NE is much shorter than P or Chl*a* (Stanley et al. 2019). As a result, the decadal medians we used in the analysis were based on fewer data points and, potentially, noisier data. We were not surprised that our model failed to detect effects of our covariates on the spatial dependence structure of lake N:P ratios. Spatial patterns in stoichiometric ratios are relatively difficult to explain with the geographic and lake characteristics in LAGOS (Collins et al. 2017).

Application to other ecological datasets

The past decade has produced new insights about macrosystems biology and new data products that can be used to address ecological questions at regional to continental scales. Some databases (including data from LAGOS-NE used in this manuscript) are compilations of numerous existing data collection efforts that were harmonized into a single database (Soranno et al. 2015). The ample amount of water quality data collected by state, federal, university, tribal and citizen monitoring programs across the United States led to spatial data coverage in LAGOS-NE that had a sufficient density of sampling sites over space to successfully employ this modeling approach. Monitoring programs that are designed specifically to collect data for a particular project or agency (e.g., NEON, US EPA National Lakes Assessment) have some sampling design advantages, such as the capacity to select sites in a stratified random design, or a more even availability of data across sites, but these programs typically include fewer collection sites at a lower density. Fitting spatial models to data products with more limited spatial coverage (i.e., fewer sites clustered in close proximity, and fewer sites overall) may be challenging, but our results suggest that we should rise to the challenge because non-stationary spatial modeling approaches can lend new insights into macroscale ecological questions. The high spatial coverage of remotely sensed water quality data (e.g. Ross et al. 2019) is particularly suitable for the application of our non-stationary models. In general, spatial locations that are evenly spaced throughout the study region are known to be optimal for prediction, whereas locations that are clustered together are known to be optimal for parameter estimation (e.g., Zhu and Zhang 2006). To this end, we expect that lake locations in LAGOS-NE should lend themselves well to successful estimation of model parameters. Intuitively, databases that carry water quality data sampled from locations that are clustered together, would make better candidates for future application of our non-stationary models.

Acknowledgments

This work was funded by a seed grant from the College of Arts and Sciences at the University of Wyoming to Sarah M. Collins and Pavel Chernyavskiy. Charlotte F. Narr and Pavel Chernyavskiy contributed equally to this work.

Literature Cited

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2015. Hierarchical modeling and analysis for spatial data. CRC Press, Boca Raton, FL, USA.

Bass, M. R., and S. K. Sahu. 2017. A comparison of centring parameterisations of Gaussian process-based models for Bayesian computation using MCMC. Statistics and Computing 27:1491–1512.

Betancourt, M. 2017. A Conceptual Introduction to Hamiltonian Monte Carlo. https://arxiv.org/abs/1701.02434

Calder, C. A. 2008. A dynamic process convolution approach to modeling ambient particulate matter concentrations. Environmetrics 19:39–48.

Carpenter, B. A., M. D. Gelman, D. Hoffman, B. Lee, M. Goodrich, M. Betancourt, J. Brubaker, P. L. Guo, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76. https://doi.org/10.18637/jss.v076.i01

Celeux, G., et al. 2006. Deviance information criteria for missing data models. Bayesian Analysis 1:651–673.

Chave, J. 2013. The problem of pattern and scale in ecology: What have we learned in 20 years? Ecology Letters 16: 4–16.

Collins, S. M., S. K. Oliver, J. F. Lapierre, E. H. Stanley, J. R. Jones, T. Wagner, and P. A. Soranno. 2017. Lake nutrient stoichiometry is less predictable than nutrient concentrations at regional and sub-continental scales. Ecological Applications 27:1529–1540.

Denny, M. W., B. Helmuth, G. H. Leonard, C. D. G. Harley, L. J. H. Hunt, and E. K. Nelson. 2004. Quantifying scale in ecology: Lessons from a wave-swept shore. Ecological Monographs 74:513–532.

Ellison, M. E., and M. T. Brett. 2006. Particulate phosphorus bioavailability as a function of stream flow and land cover. Water Research 40:1258–1268.

Fergus, C. E., A. O. Finley, P. A. Soranno, and T. Wagner. 2016. Spatial variation in nutrient and water color effects on lake chlorophyll at macroscales. PLoS One 11:e0164592.

Fergus, C. E., P. A. Soranno, K. S. Cheruvelil, and M. T. Bremigan. 2011. Multiscale landscape and wetland drivers of lake total phosphorus and water color. Limnology and Oceanography 56:2127–2146.

Fortin, M.-J., M. R. T. Dale, and J. M. Ver Hoef. 2016. Spatial analysis in ecology. Pages 1–13 *in* Wiley StatsRef: Statistics Reference Online. John Wiley & Sons Ltd, Chichester, UK.

Fraterrigo, J. M., and J. A. Downing. 2008. The influence of land use on lake nutrients varies with watershed transport capacity. Ecosystems 11:1021–1034.

Gelfand, A. E., and S. Banerjee. 2017. Bayesian modeling and analysis of geostatistical data. Annual Review of Statistics and its Application 4:245–266.

Gelman, A., J. Hwang, and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016.

Gilinsky, E., J. M. Capacasa, M. G. Baker & E. S. King. 2009. An urgent call to action: report of the state-EPA nutrient innovations task group. US Environmental Protection Agency, Washington DC.

Higdon, D. M., J. Swall, and J. C. Kern 1999. Non-stationary spatial modeling. Pages 761–768 *in* Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting. Oxford University Press.

Hoff, P. D., and X. Niu. 2012. A covariance regression model. Statistica Sinica 22. https://doi.org/10.5705/ss.2010.051

Howarth, R. W., A. Sharpley, and D. Walker. 2002. Sources of nutrient pollution to coastal waters in the United States: implications for achieving coastal water quality goals. Estuarine Research Federation Estuaries 656:656–676.

Ingebrigtsen, R., F. Lindgren, and I. Steinsland. 2014. Spatial models with explanatory variables in the dependence structure. Spatial Statistics 8:20–38.

Kaushal, S. S., and K. T. Belt. 2012. The urban watershed continuum: Evolving spatial and temporal dimensions. Urban Ecosystems 15:409–435.

Kleiber, W., and D. Nychka. 2012. Nonstationary modeling for multivariate spatial processes. Journal of Multivariate Analysis 112:76–91.

Kruschke, J. K. 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Second edition. Elsevier Inc., London, UK.

Lapierre, J. F., S. M. Collins, D. A. Seekell, K. Spence Cheruvelil, P. N. Tan, N. K. Skaff, Z. E. Taranu, C. E. Fergus, and P. A. Soranno. 2018. Similarity in spatial structure constrains ecosystem relationships: Building a macroscale understanding of lakes. Global Ecology and Biogeography 27:1251–1263.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Li, L., S. Qiu, B. Zhang, and C. X. Feng. 2016. Approximating cross-validatory predictive evaluation in Bayesian latent variable models with integrated IS and WAIC. Statistics and Computing 26:881–897.

Lowe, R. L., S. W. Golladay, and J. R. Webster. 1986. Periphyton response to nutrient manipulation in streams draining clearcut and forested watersheds. Journal of the North American Benthological Society 5:221–229.

McCullough, I. M., K. S. Cheruvelil, S. M. Collins, and P. A. Soranno. 2019. Geographic patterns of the climate sensitivity of lakes. Ecological Applications 29:1–14.

Millar, R. B. 2009. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. Biometrics 65:962–969.

Millar, R. B. 2018. Conditional vs marginal estimation of the predictive loss of hierarchical models using WAIC and cross-validation. Statistics and Computing 28:375–385.

Monnahan, C. C., and K. Kristensen. 2018. No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. PLoS One 13: e0197954.

Monnahan, C. C., J. T. Thorson, and T. A. Branch. 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution 8:339–348.

Nishio, M., and A. Arakawa. 2019. Performance of hamiltonian monte carlo and no-u-turn sampler for estimating genetic parameters and breeding values. Genetics Selection Evolution 51:73.

Noe, G. B., and C. R. Hupp. 2005. Carbon, nitrogen, and phosphorus accumulation in floodplains of Atlantic Coastal Plain rivers, USA. Ecological Applications 15:1178–1190.

Paciorek, C. J., and M. J. Schervish. 2006. Spatial Modelling using a new class of nonstationary covariance functions. Environmetrics 17:483–506.

Peters, D. P. C., R. A. Pielke, B. T. Bestelmeyer, C. D. Allen, S. Munson-McGee, and K. M. Havstad. 2004. Cross-scale interactions, nonlinearities, and forecasting catastrophic events. Proceedings of the National Academy of Sciences of the United States of America 101:15130–15135.

Pickett, S., and M. Cadenasso. 1995. Landscape ecology: Spatial heterogeneity in ecological systems. Science 269:331–334.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Read, E. K., et al. 2015. The importance of lake-specific characteristics for water quality across the continental United States. Ecological Applications 25:943–955.

Reich, B. J., J. Eidsvik, M. Guindani, A. J. Nail, and A. M. Schmidt. 2011. A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration. Annals of Applied Statistics 5. https://doi.org/10.1214/11-AOAS482

Risser, M. D., and C. A. Calder. 2015. Regression-based covariance functions for nonstationary spatial modeling. Environmetrics 26:284–297.

Risser, M. D., and D. Turek. 2020. Bayesian inference for high-dimensional nonstationary Gaussian processes. Journal of Statistical Computation and Simulation 90(16):2902–2928.

Ross, M. R. V., S. N. Topp, A. P. Appling, X. Yang, C. Kuhn, D. Butman, M. Simard, and T. M. Pavelsky. 2019. AquaSat: A Data Set to Enable Remote Sensing of Water Quality for Inland Waters. Water Resources Research 55:10012–10025.

Sampson, P. D. 2010. Constructions for nonstationary spatial processes. Pages 119–130 *in* Handbook of spatial statistics. CRC Press, Boca Raton, Florida, USA.

Schabenberger, O., and C. A. Gotway. 2005. Statistical methods for spatial data analysis. Taylor and Francis Group, LLC, Boca Raton, Florida, USA.

Schmidt, A. M., P. Guttorp, and A. O'Hagan. 2011. Considering covariates in the covariance structure of spatial processes. Environmetrics 22:487–500.

Schmidt, A. M., and M. A. Rodríguez. 2011. Modelling multivariate counts varying continuously in space. Pages 611–628 *in* J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors. Bayesian Statistics 9. Oxford University Press, Oxford, UK.

Schmidt, A. M., M. A. Rodríguez, and E. S. Capistrano. 2015. Population counts along elliptical habitat contours: hierarchical modelling using Poisson-lognormal mixtures with nonstationary spatial structure. Annals of Applied Statistics 9:1372–1393.

Soranno, P. A., et al. 2014. Cross-scale interactions: Quantifying multi-scaled cause-effect relationships in macrosystems. Frontiers in Ecology and the Environment 12:65–73.

Soranno, P. A., et al. 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science and data reuse. Gigascience 4:28.

Soranno, P. A., L. C. Bacon, and M. Beauchene. 2017. LAGOS-NE: A multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. Gigascience 6:12.

Soranno, P. A., and K. S. Cheruvelil. 2017. LAGOS-NE-GEO v1.05: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013.

Environmental Data Initiative. https://doi.org/10.6073/pasta/b88943d10c6c5c480d5230c8890b74a8

Soranno, P. A., and K. S. Cheruvelil. 2019. LAGOS-NE-LIMNO v1.087.3: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013. Environmental Data Initiative. https://doi.org/10.6073/pasta/08c6f9311929f4874b01bcc64eb3b2d7

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2014. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B 76:485–493.

Spiegelhalter, D. J., G. B. Nicola, P. C. Bradley, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64:1–34.

Stachelek, J., and S. Oliver. 2017. LAGOSNE: Interface to the Lake Multi-scaled Geospatial and Temporal Database. R package version 1.1.0. https://cran.r-project.org/package=LAGOSNE

Stanley, E. H., S. M. Collins, N. R. Lottig, S. K. Oliver, K. Webster, K. S. Cheruvelil, and P. A. Soranno. 2019. Biases in lake water quality sampling and implications for macroscale research. Limnology and Oceanography 64:1572–1585.

Stoddard, J. L., J. Van Sickle, A. T. Herlihy, J. Brahney, S. Paulsen, D. V. Peck, R. Mitchell, and A. I. Pollard. 2016. Continental-scale increase in lake and stream phosphorus: are oligotrophic systems disappearing in the United States? Environmental Science and Technology 50:3409–3415.

Taranu, Z. E., and I. Gregory-Eaves. 2008. Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. Ecosystems 11:715–725.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432.

Vianna Neto, J. H. V., A. M. Schmidt, and P. Guttorp. 2014. Accounting for spatially varying directional effects in spatial covariance structures. Journal of the Royal Statistical Society: Series C (Applied Statistics) 63:103–122.

Wagner, T., P. A. Soranno, K. E. Webster, and K. S. Cheruvelil. 2011. Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. Freshwater Biology 56:1811–1824.

Wakida, F. T., S. Martinez-Huato, E. Garcia-Flores, T. D. J. Piñon-Colin, H. Espinoza-Gomez, and A. Ames-López. 2014. Pollutant association with suspended solids in stormwater in Tijuana, Mexico. International Journal of Environmental Science and Technology 11:319–326.

Watanabe, S. 2013. A widely applicable Bayesian information criterion. Journal of Machine Learning Research 14:867–897.

Xu, H., and P. Gardoni. 2018. Improved latent space approach for modelling non-stationary spatial-temporal random fields. Spatial Statistics 23:160–181.

Zhu, Z., and H. Zhang. 2006. Spatial sampling design under the infill asymptotic framework. Environmetrics 17:323–337.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2485/full

Open Research

Data and code are available on Zenodo: https://doi.org/10.5281/zenodo.4813009