CONCEPTS & SYNTHESIS
EMPHASIZING NEW IDEAS TO STIMULATE RESEARCH IN ECOLOGY

Ecology, 0(0), 2016, pp. 1–15

The basis function approach for modeling autocorrelation in ecological data

Trevor J. Hefley,[1,2,4] Kristin M. Broms,[1] Brian M. Brost,[1] Frances E. Buderman,[1] Shannon L. Kay,[2] Henry R. Scharf,[2] John R. Tipton,[2] Perry J. Williams,[1,2] and Mevin B. Hooten[3,2,1]

[1]*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA*
[3]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, Colorado 80523 USA*

Abstract. Analyzing ecological data often requires modeling the autocorrelation created by spatial and temporal processes. Many seemingly disparate statistical methods used to account for autocorrelation can be expressed as regression models that include basis functions. Basis functions also enable ecologists to modify a wide range of existing ecological models in order to account for autocorrelation, which can improve inference and predictive accuracy. Furthermore, understanding the properties of basis functions is essential for evaluating the fit of spatial or time-series models, detecting a hidden form of collinearity, and analyzing large data sets. We present important concepts and properties related to basis functions and illustrate several tools and techniques ecologists can use when modeling autocorrelation in ecological data.

Key words: autocorrelation; Bayesian model; collinearity; dimension reduction; semiparametric regression; spatial statistics; time series.

Introduction

Ecological processes interact at multiple temporal and spatial scales, generating complex spatio-temporal patterns (Levin 1992). The science of ecology is concerned with understanding, describing, and predicting components of these spatio-temporal patterns using limited and noisy observations. An important consideration when developing an ecological model is how to include the spatial, temporal, or spatio-temporal aspects of the process (Legendre 1993). For example, species distribution models are used to predict and infer how the occurrence and abundance of plants and animals varies across space and time (Elith and Leathwick 2009, Hefley and Hooten 2016). The abundance of a species within a patch of habitat might depend on environmental covariates (e.g., minimum annual temperature), but might also depend on the abundance in surrounding patches. In other words, the abundance in nearby patches may be more similar than could be explained by environmental conditions alone (Tobler 1970, Legendre and Fortin 1989). When the value of an observation depends on its proximity to other observations, the observations are said to be autocorrelated (Table 1). Disentangling autocorrelation from the effect of environmental covariates is critical to inferring endogenous and exogenous factors that influence populations and ecosystems (Borcard et al. 1992). Moreover, properly accounting for autocorrelation is necessary for obtaining reliable statistical inference (Fieberg and Ditmer 2012, Hefley et al. 2016).

Isolating the effect of autocorrelation in an ecological model can be accomplished by including a function that captures the dependence among observations that are close in space or time. The mathematical form of the function that best describes the autocorrelation is always unknown and may be complex, but can be approximated by a combination of simple basis functions. Most ecologists have encountered basis functions (e.g., polynomial regression), but may not be aware of the breadth of situations in which they can be used to model autocorrelation in ecological data. For example, basis functions are used in semiparametric models, such as generalized additive models (GAMs), but are also implicitly used in spatial or time-series models. Understanding how basis functions can be used to model autocorrelation is essential for evaluating the fit of spatial, time-series, or spatio-temporal models, detecting a hidden form of collinearity (Hodges

Manuscript received 16 June 2016; revised 18 October 2016; accepted 24 October 2016. Corresponding Editor: Paul B. Conn.
[4]Present address: Department of Statistics, Kansas State University, Manhattan, Kansas 66506 USA. E-mail: thefley@ksu.edu

and Reich 2010), and facilitating the analysis of large data sets (Wikle 2010). More importantly, employing the basis function approach enables ecologists to tailor many commonly used models to account for autocorrelation, which can improve inference and predictive accuracy (Hooten et al. 2003, Conn et al. 2015, Buderman et al. 2016).

We have three goals in this paper: (1) introduce concepts and terminology related to basis functions and autocorrelation; (2) demonstrate the connections between commonly used methods to model autocorrelation; and (3) develop a working knowledge of the basis function approach so ecologists can devise ways to model autocorrelation in commonly used ecological models. We first introduce the concepts of basis functions and then first-order and second-order model specifications. Then we present three examples to illustrate these concepts: a standard regression model, a time-series model, and a spatial model, each applied to different types of ecological data. We include supplementary material comprised of tutorials that contain additional descriptions, data, and example computer code. By illustrating a diversity of modeling approaches, we encourage researchers to consider multiple perspectives when modelling autocorrelation in ecological data.

Basis Functions

We begin by describing the basis function approach in a linear regression setting. Consider a simple linear regression model of pelagic bioluminescence density as a function of water depth (Fig. 1a, Gillibrand et al. 2007)

$$\mathbf{y} = \alpha_0 \mathbf{z}_0 + \alpha_1 \mathbf{z}_1 + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{y}$ is defined ($\equiv$) as an $n \times 1$ vector of bioluminescence density (i.e., $\mathbf{y} \equiv (y_1, \ldots, y_n)'$ is a set of n observations), $\mathbf{z}_0$ is an $n \times 1$ vector of ones ($\mathbf{z}_0 \equiv (1, \ldots, 1)'$), $\mathbf{z}_1$ is an $n \times 1$ vector that contains the depth in meters of the observed bioluminescence sources ($\mathbf{z}_1 \equiv (\text{depth}_1, \ldots, \text{depth}_n)'$), α_0 is the intercept, α_1 is a regression coefficient, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector that contains independent and normally distributed error terms with variance σ^2_ε (i.e., $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \ldots, \varepsilon_n)'$ where $\varepsilon_i \sim \text{N}(0, \sigma^2_\varepsilon)$). In this simple linear regression model, the basis coefficients are α_0 and α_1, and the basis vectors $\mathbf{z}_0$ and $\mathbf{z}_1$ are the depths raised to the power 0 and 1 (Table 1). It is not common to refer to transformations of a covariate as basis vectors; however, the transformations form a "basis" of possible values in covariate space. The function that transforms a covariate into a basis vector is referred to as a basis function. Although the terms basis function and basis vector tend to be used interchangeably, basis functions are continuous functions, whereas basis vectors are the output from a function at a finite number of points (e.g., where depth was measured; Table 1; see Appendix S1 for further discussion). The collection of basis vectors resulting from transformations of a covariate defined by a basis function are known collectively as a basis expansion. For example, in Eq. 1, the collection of vectors $\mathbf{z}_0$ and $\mathbf{z}_1$ are the basis expansion determined by transformations of the covariate depth using a power function. Finally, as in simple linear regression, the expected density of bioluminescence at the measured depths is the linear combination of basis vectors and their coefficients $\alpha_0\mathbf{z}_0 + \alpha_1\mathbf{z}_1$ (Fig. 1a).

It is clear from Fig. 1a that the simple linear regression model does not adequately capture the relationship between bioluminescence and depth. A more flexible basis expansion that better captures the relationship is the polynomial regression model that includes the quadratic effect of depth

$$\mathbf{y} = \alpha_0 \mathbf{z}_0 + \alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2 + \boldsymbol{\varepsilon}, \quad (2)$$

where α_2 is the basis coefficient for the squared effect of depth ($\mathbf{z}_2 = \mathbf{z}_1^2$; Fig. 1b). Some models that use basis

Table 1. Glossary of terms and definitions.

Term	Definition
Autocorrelation	Correlation between observations based on some measure of distance or time that exists after the influence of all covariates is accounted for
Basis expansion	A collection of basis vectors from a single covariate
Basis vector	Any transformation of a covariate
Basis function	Any mathematical function that transforms a covariate
Compact support	A support that does not include all possible locations or time points
Correlation function	A function that describes the autocorrelation between observations
Correlation matrix	A positive semi-definite matrix whose elements are the correlation between observations
Covariate	Any quantity that can be measured and is associated with an observation (e.g., the time or spatial location of the observation)
Dependence	Correlation between observations defined on a general space (spatial or temporal dependence is equivalent to autocorrelation)
First-order specification	When a function that models the dependence is specified in the mean (expected value) of a probability distribution
Global support	A support that includes all possible locations or time points
Second-order specification	When a function that models dependence is specified in the covariance of a probability distribution
Support	The set of locations or time points where the basis function results in non-zero values

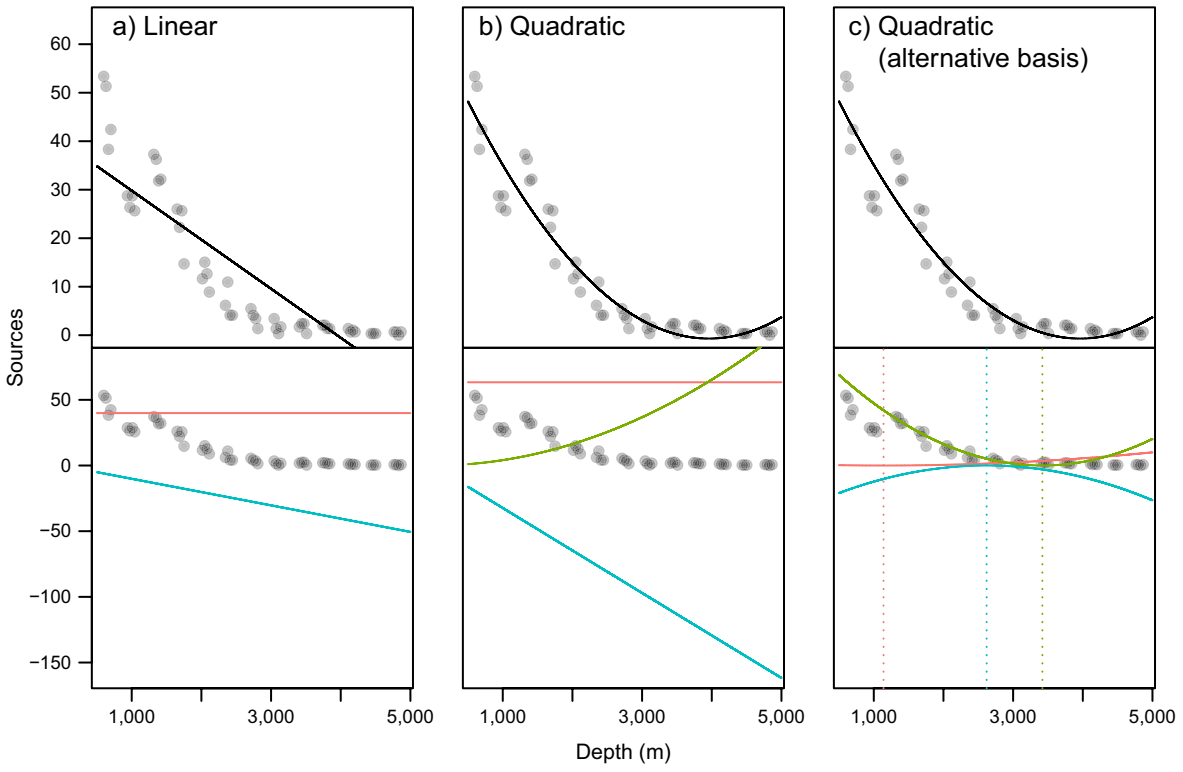

FIG. 1. Scatterplots showing the density of pelagic bioluminescence (sources) vs. water depth. The top panels show fitted regression models (black lines). The corresponding basis vectors multiplied by the estimated coefficients (colored curves) are shown in the bottom panels. (a) Simple linear regression model (Eq. 1) with corresponding constant (red) and linear basis vectors (blue). (b) Polynomial regression model (Eq. 2) with corresponding constant (red), linear (blue), and quadratic basis vectors (green). (c) The same polynomial regression model as that shown in panel (b), except with basis vectors calculated relative to three water depths (k_j in Eq. 3 where k_1 = 1,140 m, k_2 = 2,620 m, and k_3 = 3,420 m; vertical colored lines). Note that the basis vectors are multiplied by the estimated coefficients and summed to produce the fitted curves (black lines). See Appendix S2 for an interactive version of this figure.

functions can be respecified, which can facilitate interpretation, increase computational efficiency, and improve numerical stability of estimation algorithms. For example, we can respecify Eq. 2 using a different basis expansion, but in a way that yields the exact same model

$$\mathbf{y} = \alpha_1(\mathbf{z}_1 - k_1)^2 + \alpha_2(\mathbf{z}_1 - k_2)^2 + \alpha_3(\mathbf{z}_1 - k_3)^2 + \boldsymbol{\varepsilon}, \quad (3)$$

where $\mathbf{z}_1$ is an $n \times 1$ vector of the observed depths in meters, k_j is the jth depth of interest (j = 1, 2, 3), and α_j is the basis coefficient. The two basis expansions in Eqs. 2 and 3 have different basis vectors and will yield different estimates of α_j, but result in the exact same polynomial curve when fit to the data; mathematically, this can be shown by expanding the quadratic terms in Eq. 3. For example, let k_1 = 1,140 m, k_2 = 2,620 m, and k_3 = 3,420 m, and compare the basis vectors and predicted bioluminescence (cf. Fig. 1b, c). An interactive figure for this example can be found in Appendix S2.

Even if the specifications result in identical models, there are many reasons why one basis expansion might be preferred over others. For example, the number of parameters in the model can be reduced if a basis expansion results in some basis coefficients that can be assumed to be zero (see confidence intervals for α_j in Appendix S2). In addition, some basis expansions might have scientific interpretations. For example, the model in Eq. 3 states that the density of bioluminescence is a function of the distance between the observed depth and three locations in the water column that we might believe are biologically important. Finally, some basis expansions may reduce the correlation among terms in the model. For example, the coefficient of determination (R^2) for the basis vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ in Eq. 2 is 0.96, whereas the maximum R^2 among the three basis vectors in Eq. 3 is 0.25 (Fig. 1b, c). The reduced correlation can improve performance of the parameter estimation algorithms, demonstrating the benefit of respecifying a model with a different but equivalent basis expansion.

Model assumptions

A critical assumption of models that use basis functions is that a linear combination of basis vectors adequately approximates the unknown relationship between the observed response and the spatial location. In a regression context, this is analogous to assuming the

covariates, or transformation of covariates, adequately approximate the expected response. For example, it is assumed in Eqs. 2 and 3 that a linear combination of the basis vectors (e.g., $\alpha_0\mathbf{z}_0 + \alpha_1\mathbf{z}_1 + \alpha_2\mathbf{z}_2$) adequately represents the unknown functional relationship between depth and the density of bioluminescence.

More formally, assuming a basis expansion is adequate for approximating an unknown functional relationship implies that the basis vectors "span" the space containing the unknown function explaining the relationship between the covariate and the response. In the bioluminescence example, the basis vectors span the set of all second-degree polynomial functions because any second-degree polynomial function of depth is as a linear combination of the basis vectors in Eqs. 2 or 3. By selecting these basis vectors, we are assuming that the true underlying relationship between depth and the density of bioluminescence is modelled appropriately as a second-degree polynomial. The two collections of basis vectors in Eqs. 2 and 3 both span the same space, which is another way of understanding why the estimated curves in Fig. 1b and c are identical.

Generalizations

Now consider an unknown function $\eta(x)$ that describes a pattern or process in nature that generates autocorrelation over the space of interest x. For example, $\eta(x)$ could describe the similarity in abundance among habitat patches in geographic space, population regulation influenced by endogenous factors in temporal space, or how net primary productivity changes with temperature in covariate space. Even though the true form of the function $\eta(x)$ is unknown, we can approximate it with a combination of simple basis functions. We combine the basis functions in a linear equation such that $\eta(x) \approx \sum_{j=1}^{m} \alpha_j f_j(x)$, using a general notation that consists of m basis functions $f_j(x)$ ($j = 1, \ldots, m$). In the polynomial regression model (Eq. 2), for example, $f_1(x) = x^0$, $f_2(x) = x^1$, and $f_3(x) = x^2$. In what follows, we use matrix notation and write $\boldsymbol{\eta} \equiv \mathbf{Z}\boldsymbol{\alpha}$, where $\boldsymbol{\eta}$ is an $n \times 1$ vector representing an approximation of the unknown function $\eta(\mathbf{x})$ at the n locations $\mathbf{x}$ in the space of interest, $\mathbf{Z}$ is an $n \times m$ matrix containing the basis vectors, and $\boldsymbol{\alpha}$ is an $m \times 1$ vector of basis coefficients. We also use the matrix notation $\mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is an $n \times p$ matrix of traditional covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of traditional regression coefficients. In some applications, there is no practical difference between including basis vectors in $\mathbf{X}$ or $\mathbf{Z}$; however, the choice of notation is used to designate whether the coefficients are treated as fixed ($\mathbf{X}$) or random ($\mathbf{Z}$) effects in what follows.

Model Specifications

An important concept for understanding the equivalence of certain spatial and time series models is first-order and second-order model specifications (Table 1; Cressie and Wikle 2011, Hobbs and Hooten 2015), which are also known in the mixed-model literature as G-side and R-side specifications (Littell et al. 2006, Stroup 2012). First-order and second-order specifications differ in terms of whether a function describing autocorrelation is contained in the mean (expected value) or the covariance of a probability distribution. A general understanding of hierarchical models or mixed models is necessary for what follows (Ruppert et al. 2003, Littell et al. 2006, Wood 2006, Ruppert et al. 2009, Stroup 2012, Hodges 2013, Hobbs and Hooten 2015).

First-order specification

A first-order specification refers to a model that contains a function in the mean structure of a distribution for describing autocorrelation (Table 1). Consider the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Assuming independent and normally distributed errors ($\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\varepsilon \mathbf{I})$), we can write this model as

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2_\varepsilon \mathbf{I}). \quad (4)$$

As before, an assumption of Eq. 4 is that a linear combination of the covariates serves as a good approximation to the unknown relationship with the mean (expected value) of the response. For example, we might assume that bioluminescence density can be modeled as a quadratic effect of depth (e.g., Eq. 1 and Fig. 1b); however, the model in Eq. 4 is inadequate because autocorrelation in the residuals is evident (e.g., bioluminescence densities that occurs within certain ranges of depth occur entirely above or below the fitted curve; Fig. 1a). A linear combination of basis vectors may be added to the mean to improve model fit and satisfy model assumptions, such that

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \sigma^2_\varepsilon \mathbf{I}). \quad (5)$$

The basis expansion ($\mathbf{Z}$) accounts for additional complexity in the mean structure (e.g., lack of fit due to autocorrelation). The covariates in $\mathbf{X}$ may or may not include an effect of time or spatial location (e.g., linear and quadratic effect of depth).

Second-order specification

A second-order specification refers to a model that contains a function in the covariance of a probability distribution for describing the autocorrelation. Consider the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\varepsilon \mathbf{I})$, $\boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\alpha \mathbf{R})$, and the random effect $\boldsymbol{\eta}$ results in correlated errors. Using integration, we can express the model as

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2_\varepsilon \mathbf{I} + \sigma^2_\alpha \mathbf{R}), \quad (6)$$

where $\mathbf{R}$ is a correlation matrix that accounts for autocorrelation among observations. The correlation matrix $\mathbf{R}$ is often specified using a correlation function that depends on a distance measure between two observations in the space of interest (Table 1). In the bioluminescence

example, autocorrelation that remains in the residuals after fitting a regression model could be explicitly modeled using a second-order specification in Eq. 6.

Equivalent specifications

In some situations, the first-order and second-order specifications result in the same model. When a model has an equivalent first- or second-order specification, it is advantageous to convert between the two specifications for efficient implementation of models that account for autocorrelation and to assess collinearity among covariates and basis vectors. To demonstrate equivalent model specifications for a specific case, we make the additional assumption that the basis coefficients in Eq. 5 are normally distributed random effects (i.e., $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \sigma_\alpha^2 \mathbf{I})$). Equivalent probability density functions can be obtained by integrating the first-order specification in Eq. 5

$$\begin{aligned} \mathbf{y} &\sim \int \mathrm{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \sigma_\varepsilon^2 \mathbf{I}) \mathrm{N}(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}) d\boldsymbol{\alpha} \\ &= \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma_\varepsilon^2 \mathbf{I} + \sigma_\alpha^2 \mathbf{Z}\mathbf{Z}'), \end{aligned} \tag{7}$$

where equivalence between the first- and second-order specifications holds if the correlation matrix $\mathbf{R}$ is the outer product of the basis expansion $\mathbf{Z}$ (i.e., $\mathbf{R} \equiv \mathbf{Z}\mathbf{Z}'$). The integration in Eq. 7 effectively "moves" the autocorrelation modeled by the basis vectors in the mean structure to the covariance structure. For example, consider a mixed-effects model where $\mathbf{Z}$ is used to represent autocorrelation due to a site or grouping effect among observations

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \tag{8}$$

where, y_1 and y_2 were observed at the first site, y_3 and y_4 were observed at the second site, etc. If we assume the basis coefficients are normally distributed random effects, then

$$\begin{aligned} \mathbf{R} &= \mathbf{Z}\mathbf{Z}' \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \end{aligned} \tag{9}$$

where $\mathbf{R}$ is called the compound symmetry correlation matrix (Littell et al. 2006, Zuur et al. 2009, Stroup 2012). The model that is obtained by using the first-order specification that treats site as a random effect (Eq. 8) is identical to the model obtained by specifying a second-order model using a compound symmetry correlation matrix (Eq. 9).

Although one may start with a basis expansion $\mathbf{Z}$, many methods developed to model autocorrelation start by choosing a correlation matrix $\mathbf{R}$. When starting with a correlation matrix $\mathbf{R}$, a basis expansion $\mathbf{Z}$ can be obtained by decomposing (factoring) the correlation matrix (e.g., using spectral decomposition; Lorenz 1956, Cressie and Wikle 2011:156). Consider the regression model in Eq. 6 and let $\mathbf{R}(\phi)$ be an order-one autoregressive correlation matrix (AR(1))

$$\mathbf{R}(\phi) = \begin{bmatrix} 1 & \phi^1 & \phi^2 & \cdots & \phi^{n-1} \\ \phi^1 & 1 & \phi^1 & \cdots & \phi^{n-2} \\ \phi^2 & \phi^1 & 1 & \cdots & \phi^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix}, \tag{10}$$

where $-1 < \phi < 1$ and n is the total number of observations (e.g., in a time series). The AR(1) correlation matrix (Eq. 10) is commonly used in time-series analysis to model temporal correlation that diminishes geometrically with a rate of decay that depends on ϕ. When a correlation matrix or basis expansion depends on parameters, we include the parameters in parentheses (e.g., $\mathbf{R}(\phi)$ and $\mathbf{Z}(\phi)$).

A correlation matrix can be decomposed to produce basis vectors that are useful in the first-order specifications. One approach to obtain basis vectors from $\mathbf{R}(\phi)$ is the spectral decomposition: $\mathbf{R}(\phi) = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$, where $\mathbf{Q}$ are the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with elements that contain the eigenvalue associated with each eigenvector (note that we have suppressed the notation for dependence of $\mathbf{Q}$ and $\boldsymbol{\Lambda}$ on ϕ for brevity; Cressie and Wikle 2011:156–157). Using the spectral decomposition, the basis expansion can be written as $\mathbf{Z}(\phi) = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}$. For example, if three observations $\mathbf{y} \equiv (y_1, y_2, y_3)'$ were collected at times $t = 1, 2, 3$, the AR(1) correlation matrix using $\phi = 0.5$ is

$$\mathbf{R}(0.5) = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}. \tag{11}$$

The spectral decomposition of $\mathbf{R}(0.5)$ is

$$\begin{aligned} \mathbf{R}(0.5) &= \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}' \\ &= \begin{bmatrix} -0.54 & -0.71 & 0.45 \\ -0.64 & 0 & -0.77 \\ -0.54 & 0.71 & 0.45 \end{bmatrix} \begin{bmatrix} 1.84 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0.41 \end{bmatrix} \\ &\quad \begin{bmatrix} -0.54 & -0.64 & -0.54 \\ -0.71 & 0 & 0.71 \\ 0.45 & -0.77 & 0.45 \end{bmatrix}. \end{aligned} \tag{12}$$

The matrices of eigenvectors ($\mathbf{Q}$) and eigenvalues ($\boldsymbol{\Lambda}$) in Eq. 12 can be used to construct the basis expansion

$$\mathbf{Z}(0.5) = \mathbf{Q}\boldsymbol{\Lambda}^{1/2} = \begin{bmatrix} -0.74 & -0.61 & 0.29 \\ -0.87 & 0 & -0.49 \\ -0.74 & 0.61 & 0.29 \end{bmatrix}. \tag{13}$$

Alternatively, one might use the eigenvectors $\mathbf{Q}$ as basis vectors (i.e., $\mathbf{Z}(\phi) \equiv \mathbf{Q}$; Griffith and Peres-Neto 2006),

which would require specifying a non-constant variance for the basis coefficients such that $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\alpha \boldsymbol{\Lambda})$.

Converting models between first- and second-order specifications using the techniques we have presented is critical for harnessing the power of basis functions for modeling autocorrelation in ecological data. For example, converting a first-order specification to the equivalent second-order specification is important when implementing efficient numerical techniques because analytical integration (e.g., Eq. 7) is often more stable and efficient when compared to numerical integration (e.g., numerical quadrature; Markov chain Monte Carlo [MCMC]; Finley et al. 2015).

Generalized models

Basis functions can be incorporated into the mean structure of any response distribution (e.g., Poisson, binomial). For example, generalized linear mixed models can include random effects for the coefficients associated with basis vectors that account for autocorrelation (Bolker et al. 2009). Similarly, basis functions can also be embedded within Bayesian hierarchical models to account for autocorrelation (e.g., Conn et al. 2015, see Example 3). In the examples that follow, we assume that the basis coefficients are normally distributed random effects; however, this is not a necessary assumption. As with any generalized linear mixed model or Bayesian hierarchical model, distributions for random effects other than the normal can be used (e.g., gamma, *t*-distribution; Higdon 2002, Lee et al. 2006, Gelman et al. 2013, Johnson et al. 2013, Hobbs and Hooten 2015).

Example 1: Pelagic Bioluminescence vs. Depth Gradient

In *Basis Functions*, we initially modeled the density of bioluminescence as a function of depth using coefficients that were fixed effects; however, depth can also be thought of as the spatial location in the water column. Thus, it is natural to model the density of bioluminescence using a spatial model instead of regressing on depth directly. In the three model specifications used to capture spatial autocorrelation that follow, $\mathbf{X}$ is an $n \times 1$ matrix of ones and $\boldsymbol{\beta}$ is a scalar intercept term. As a result, the influence of depth is modeled in either the basis expansion $\mathbf{Z}$ or the correlation matrix $\mathbf{R}$, in accordance with the first-order or second-order specification, respectively.

Spatial regression model: a second-order specification

Consider the model in Eq. 6 where the correlation matrix $\mathbf{R}(\phi)$ is specified using a parametric correlation function that depends on a range parameter ϕ (Cressie and Wikle 2011, Banerjee et al. 2014). The range parameter controls how the correlation diminishes as the distance between two locations increases. For this example, we use a Gaussian correlation function

$$r_{ij}(\phi) = e^{-d^2_{ij}/\phi}, \tag{14}$$

where d_{ij} is the distance between locations i and j (note that $d_{ij} = 0$ for $i = j$) and $r_{ij}(\phi)$ is the element in the ith row and jth column of $\mathbf{R}(\phi)$. In the bioluminescence example, d_{ij} is the difference in depth between observations i and j. Given the second-order specification, it is not immediately clear how to estimate the influence of depth on bioluminescence (i.e., $\beta_0 + \boldsymbol{\eta}$), which requires the spatial random effect $\boldsymbol{\eta}$. To predict bioluminescence at the observed (and unobserved) depths using the second-order specification, we used best linear unbiased prediction (see Robinson [1991] for derivation). The predicted spatial random effect for the observed depths, given estimates of all other parameters, can be obtained using

$$\hat{\boldsymbol{\eta}} = \hat{\sigma}^2_\alpha \mathbf{R}(\hat{\phi}) \left(\hat{\sigma}^2_\varepsilon \mathbf{I} + \hat{\sigma}^2_\alpha \mathbf{R}(\hat{\phi}) \right)^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right) \tag{15}$$

where the "hat" notation indicates that the parameter is an estimate (e.g., the maximum likelihood estimate). The fitted spatial model (Fig. 2a; $\hat{\beta}_0 + \hat{\boldsymbol{\eta}}$) captures fine scale (local) variability better than the polynomial regression model (Fig. 1b, c; Appendix S3).

Spatial regression model: a first-order specification

The spatial regression model can also be implemented using basis vectors. Consider the first-order specification from Eq. 5, where $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \sigma^2_\alpha \mathbf{I})$ and $\mathbf{Z}(\phi)$ is obtained from a spectral decomposition of $\mathbf{R}(\phi)$. Using Eq. 15, basis coefficients are equivalent to

$$\hat{\boldsymbol{\alpha}} = \hat{\sigma}^2_\alpha \mathbf{Z}(\hat{\phi})' \left(\hat{\sigma}^2_\varepsilon \mathbf{I} + \hat{\sigma}^2_\alpha \mathbf{Z}(\hat{\phi}) \mathbf{Z}(\hat{\phi})' \right)^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right). \tag{16}$$

Because $\mathbf{R}(\hat{\phi}) \equiv \mathbf{Z}(\hat{\phi})\mathbf{Z}(\hat{\phi})'$, by definition, $\mathbf{Z}(\hat{\phi})\hat{\boldsymbol{\alpha}}$ is the same as $\hat{\boldsymbol{\eta}}$ in Eq. 15. The expected bioluminescence (i.e., $\hat{\beta}_0 + \mathbf{Z}(\hat{\phi})\hat{\boldsymbol{\alpha}}$) from the first-order specification is shown in Fig. 2b (see Appendix S3 for details). The fitted values from the first- and second-order specifications are exactly the same (cf. Fig. 2a, b) because both specifications result in an equivalent model.

Even if the initial model formulation is a second-order specification that uses a correlation function, the equivalent first-order specification is often useful. Three important uses for the first-order specification are (1) it allows for assessment of collinearity between covariates in $\mathbf{X}$ and basis vectors in $\mathbf{Z}(\phi)$ (see Example 2); (2) basis vectors can be visualized and certain types of basis expansions have useful ecological interpretation (Griffith and Peres-Neto 2006); and (3) certain types of basis expansions are useful for dimension reduction required to fit models to large data sets (see Example 3; Wikle 2010). We demonstrate the utility of the first-order specification in the following examples.

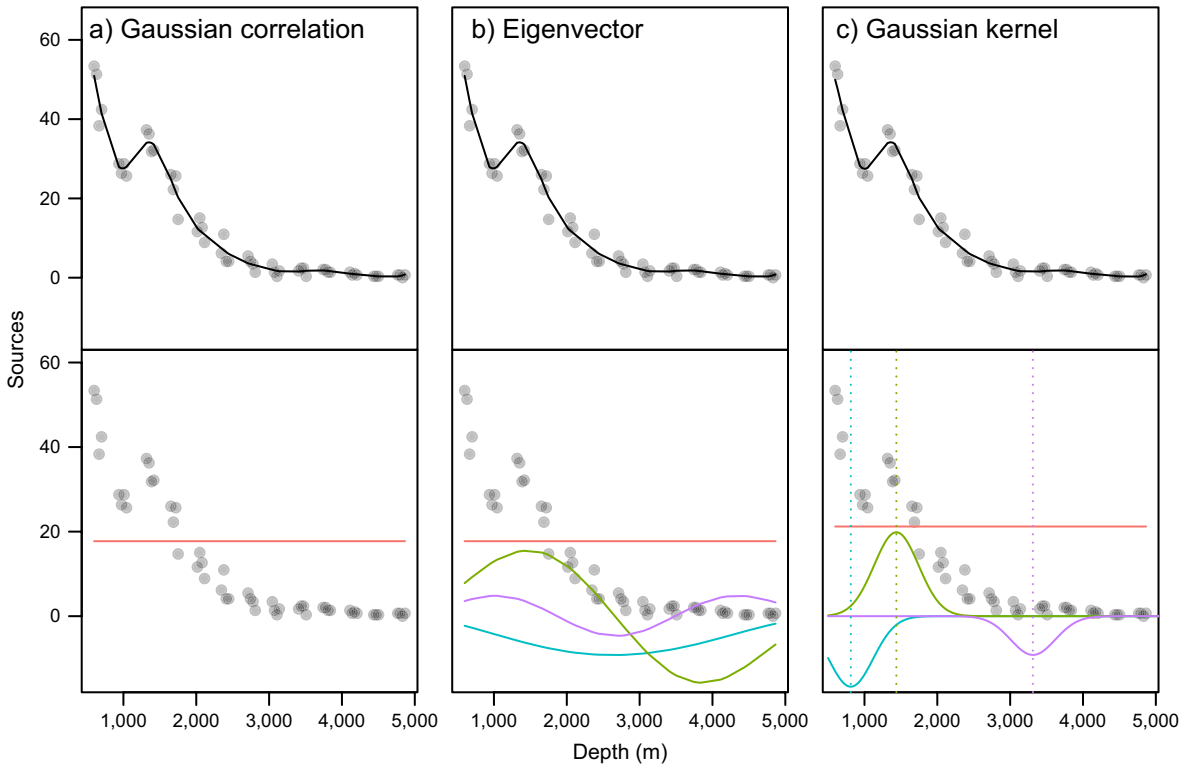

FIG. 2. Scatterplots showing the density of pelagic bioluminescence (sources) vs. water depth. The top panels shows the fitted curve (black lines) obtained from a second-order model specification that uses (a) a Gaussian correlation function, (b) the equivalent first-order specification that uses eigenvectors, and (c) a first-order specification that uses a Gaussian kernel basis function. The bottom panels show the intercept term (red), eigenvectors (b), and Gaussian kernel basis vectors (c). For illustrative purposes, only the product of three basis vectors and coefficients are shown (with knots located at the vertical lines in panel c). There are 51 eigenvectors and coefficients that sum to produce the fitted curve (black line) in panel b, and 17 kernel basis vectors that sum to produce the fitted curve (black line) in panel c. See Appendix S2 for an interactive version of this figure.

Modeling spatial autocorrelation using kernel basis functions

Another method that can be used to model autocorrelation is kernel regression. Kernel regression is a semiparametric regression technique widely used by statisticians and the machine learning community (Bishop 2006, Hastie et al. 2009, James et al. 2013). Regression models that employ kernel basis functions are written using the first-order specification. The commonly used Gaussian kernel basis function is defined as

$$z_{ij}(\phi) \propto e^{-2d_{ij}^2/\phi}, \tag{17}$$

where $z_{ij}(\phi)$ is the element in the ith row and jth column of $\mathbf{Z}(\phi)$, and d_{ij} is the distance between the ith data point and the jth knot ($j = 1, \ldots, m$ where m is the number of basis vectors). Knots are locations in the space of interest where each basis vector is anchored (e.g., k_j in the bioluminescence example; Table 1). In Fig. 2c, we show the expected density of bioluminescence predicted from the kernel regression model (i.e., $\hat{\beta}_0 + \mathbf{Z}(\hat{\phi})\hat{\boldsymbol{\alpha}}$). Comparison of the eigenvectors and kernel basis vectors reveals that the two types of basis vectors look different, but the fitted curves are nearly equivalent (cf. Fig. 2b, c). Importantly, as the number of basis vectors and knots increases to infinity (on a grid), the first-order model specification that uses a Gaussian kernel basis function (Eq. 17) converges to the second-order specification that uses a Gaussian correlation function (Eq. 14; Higdon 2002). An interactive figure that allows users to experiment with basis functions for this example appears in Appendix S2.

Regression models that use kernel basis functions are useful because they allow for more flexible correlation structures compared to models that rely on standard correlation functions (Barry and Ver Hoef 1996, Higdon 2002, Sampson 2010; Table 2). Further, the number of basis vectors and coefficients can be controlled by the user depending on the level of computational efficiency required. Choosing the dimension of the basis expansion (m) to be less than n is known as dimension reduction. For example, converting the $n \times n$ correlation matrix in the second-order spatial model to the equivalent first-order specification requires $m = n$, eigenvectors; however, kernel regression used a pre-selected number of kernel basis vectors ($m = 17$ for this example).

Table 2. Common types of bases, important properties, and references.

Basis type	Orthogonal	Support	Notable use	Reference
Eigenvector	Yes	Global	Dimension reduction and detecting collinearity between covariates and basis vectors in second-order models	Hodges and Reich (2010), Cressie and Wikle (2011: Chapter 5), Hodges (2013: Chapter 10)
Fourier	Yes	Global	Large data sets with a smooth effect of autocorrelation	Paciorek (2007*a*), Cressie and Wikle (2011: Chapter 3)
Kernel	No	Global or compact	Large data sets or a directional effect of autocorrelation	Higdon (2002), Peterson and Ver Hoef (2010), Sampson (2010)
Piecewise linear	No	Compact	Implementing numerical solutions to stochastic partial differential equations	Lindgren et al. (2011), Krainski et al. (2016)
Polynomial	No	Global	Modeling simple nonlinear effects of autocorrelation	Ruppert et al. (2003: Chapter 2), James et al. (2013: Chapter 7)
Splines	No	Global or compact	Large data sets with smooth effects of autocorrelation	Ruppert et al. (2003: Chapter 3), Wood (2006), Hastie et al. (2009: Chapter 5), James et al. (2013: Chapter 7)
Wavelets	Yes	Global and compact	Modeling discontinuous effects of autocorrelation	Nason (2010)

Dimension reduction usually relies on first-order model specification and is helpful for modeling autocorrelation in large data sets, allowing for statistical inference that would otherwise be computationally infeasible. For example, Katzfuss (*in press*) developed a multi-resolution basis function approximation for second-order model specifications that employs correlation functions and demonstrated the approach by modeling 271,014 estimates of total precipitable water obtained from the Microwave Integrated Retrieval System satellites. Modeling autocorrelation in large spatial data sets, such as those collected from automated sensing instruments on satellites and aircraft, would not be feasible using traditional methods (e.g., second-order models that rely on correlation functions).

Example 2: Population Trend

Ecologists often look for temporal trends in population abundance to evaluate whether populations are increasing or decreasing. One way to infer if a population is increasing or decreasing is to fit a trend line to a time series of relative abundance. Northern bobwhite quail (*Colinus virginianus*) are a common species that occurs throughout a large portion of the United States, but are declining in abundance in many regions (Veech 2006). For this example, we estimate the trend in bobwhite quail abundance in an area of Nebraska (Fig. 3a) using the simple linear model

$$\mathbf{y} \sim \mathrm{N}(\beta_0 + \beta_1 \mathbf{t}, \sigma_\varepsilon^2 \mathbf{I}), \tag{18}$$

where $\mathbf{y}$ is a $t \times 1$ vector containing the index of population size from each time period and $\mathbf{t}$ is the corresponding vector of times. Parameter estimation using maximum likelihood results in $\hat{\beta}_1 = -1.16$ and a 95% confidence interval of [−1.88, −0.44]. The fitted regression model suggests a decline in abundance; however, a clear pattern in the residuals is present, possibly due to underlying population dynamics of bobwhite quail (Appendix S4: Fig. S1). If autocorrelation is ignored, the uncertainty associated with regression coefficients will be underestimated and may cause the researcher to overstate the statistical significance of the decline (Cressie 1993, Hoeting 2009).

Time series model: a second-order specification

One approach for modeling the autocorrelation generated by endogenous population dynamics is to assume that the population size (or some transformation thereof) can be modeled as $\mathbf{y} \sim \mathrm{N}(\beta_0 + \beta_1 \mathbf{t}, \sigma_\varepsilon^2 \mathbf{I} + \sigma_\alpha^2 \mathbf{R}(\phi))$, where $\mathbf{R}(\phi)$ can be any appropriate correlation matrix. The AR(1) correlation matrix (Eq. 10) might be appropriate for modelling autocorrelation in populations because it arises from a process where $\eta_t = \phi\eta_{t-1} + \nu_t$ and $\nu_t \sim \mathrm{N}(0, \sigma_\alpha^2/(1-\phi^2))$ (Rue and Held 2005:1–3, Littell et al. 2006:175–176); similar specifications of difference equations are used as stochastic models of population dynamics (Dennis et al. 2006). When we account for the autocorrelation in the bobwhite quail time series using the AR(1) correlation matrix (Eq. 10), we obtain $\hat{\beta}_1 = -1.10$ and a 95% confidence interval of [−2.61, 0.41]. The 95% confidence interval now covers zero and is approximately twice as wide compared to the simple linear model that does not account for autocorrelation (Eq. 18). The fitted trend lines for the two models ($\hat{\beta}_0 + \hat{\beta}_1 \mathbf{t}$) appear nearly identical, but when the temporal process $\boldsymbol{\eta}$ is included ($\hat{\beta}_0 + \hat{\beta}_1 \mathbf{t} + \hat{\boldsymbol{\eta}}$; where $\boldsymbol{\eta}$ is estimated using Eq. 15), the fit is much better because the residuals appear to be uncorrelated (see Appendix S4: Figs. S1–S4).

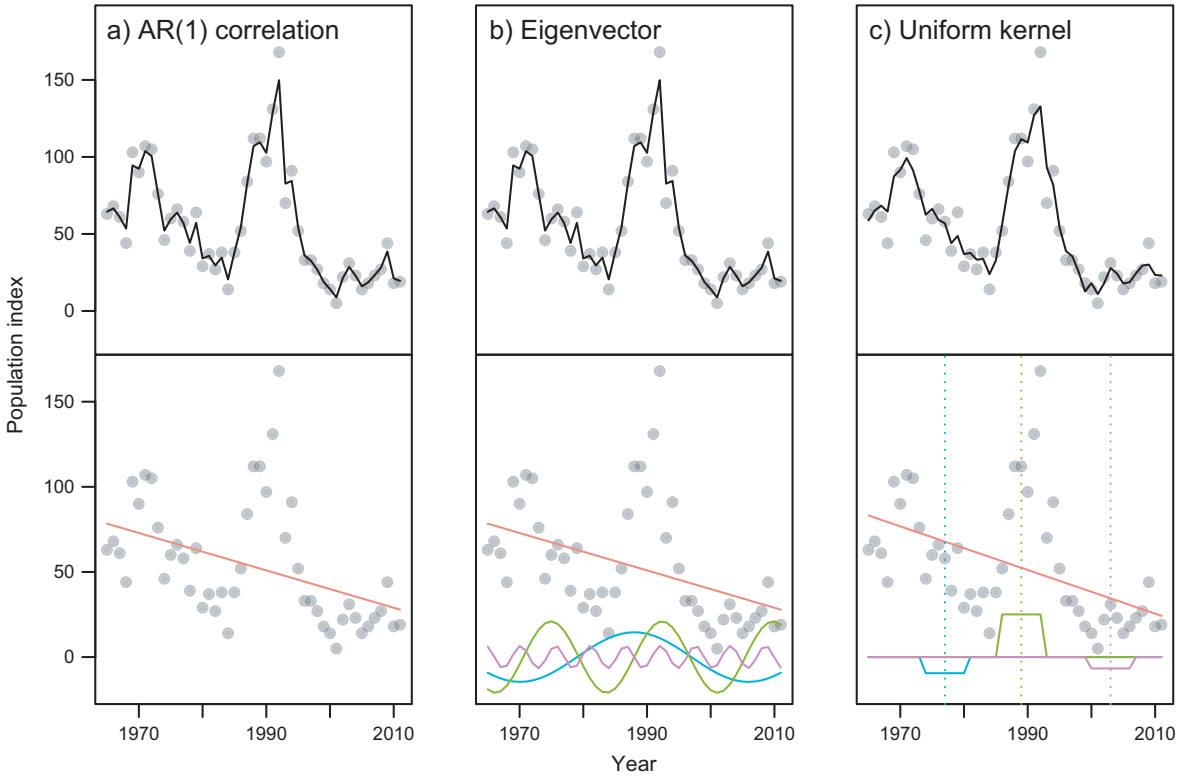

FIG. 3. Scatterplots of bobwhite quail population counts (points) from Nemaha County, Nebraska, USA. The top panels show fitted regression models (black lines) obtained from a second-order specification that uses (a) an AR(1) correlation matrix, (b) the equivalent first-order specification that uses eigenvectors, and (c) a first-order specification that uses a compactly supported uniform kernel basis function. The bottom panels show the fixed effects term (red), three eigenvectors (b), and three compactly supported kernel basis vectors with knots located at the vertical lines (c). All basis vectors are multiplied by basis coefficients. See Hefley et al. (2013) for a detailed description of the data.

Time series models: first-order specifications

To demonstrate two different first-order model specifications that account for temporal autocorrelation, we use eigenvectors obtained from a spectral decomposition of the AR(1) correlation matrix (Fig. 3b), as well as a compactly supported uniform kernel basis function with knots placed at each year (Fig. 3c, Table 1). In contrast to the spatial model used in the bioluminescence example, the AR(1) correlation matrix does not have a corresponding kernel that can be used as an approximation. Consequently, the first-order model that uses a uniform kernel basis function results in a different fit to the data when compared to the model that uses an AR(1) correlation matrix (Fig. 3b, c). Both models, however, appear to capture the temporal autocorrelation and result in similar estimates ($\hat{\beta}_1 = -1.28$, 95% confidence interval [−2.59, 0.03] for the uniform kernel basis function).

Our example with bobwhite quail demonstrates that it is important to check for collinearity between basis vectors and covariates when inference on parameters, rather than prediction, is desired. As with traditional regression models, severe collinearity can negatively influence inference (Dormann et al. 2013). The potential for collinearity is evident in the first-order specification and can be easily checked by calculating the correlation between covariates and basis vectors; in contrast, the collinearity is effectively "hidden" in the correlation matrix of the second-order specification (Hodges and Reich 2010, Hanks et al. 2015). Checking for collinearity in second-order models involves obtaining the equivalent first-order model specification (see Appendix S4). In the quail time series, for example, the coefficient of determination between the covariate year (**t** in Eq. 18) and second eigenvector is $R^2 = 0.80$, which is high enough to degrade inference (see Hefley et al. 2016 for alternative approaches).

EXAMPLE 3: PREDICTING THE DISTRIBUTION OF A SPECIES

In this example, we fit three different models that account for spatial autocorrelation to illustrate concepts presented in previous sections. Many ecological studies aim to predict the presence or abundance of a species at unsampled locations using species distribution models applied to count, presence–absence, or presence-only data (Elith and Leathwick 2009, Hefley and Hooten 2016). Generalized linear mixed models with a spatial random effect are well-suited to model a species

distribution using count or presence–absence data (Bolker et al. 2009). For example, Hooten et al. (2003) used a binary spatial regression model to predict the probability of pointed-leaved tick trefoil (*Desmodium glutinosum*) occurring in 10 × 10 m plots across a 328-ha area from presence–absence data collected at 216 plots (Fig. 4a). A common problem when predicting the distribution of a species is that data are sparse relative to the study area. For this study, only 0.66% of the plots within the study area were sampled (Fig. 4a). Consequently, Hooten et al. (2003) specified a second-order spatial random effect to increase the predictive ability of a binary regression model and to account for spatial autocorrelation generated by a complex

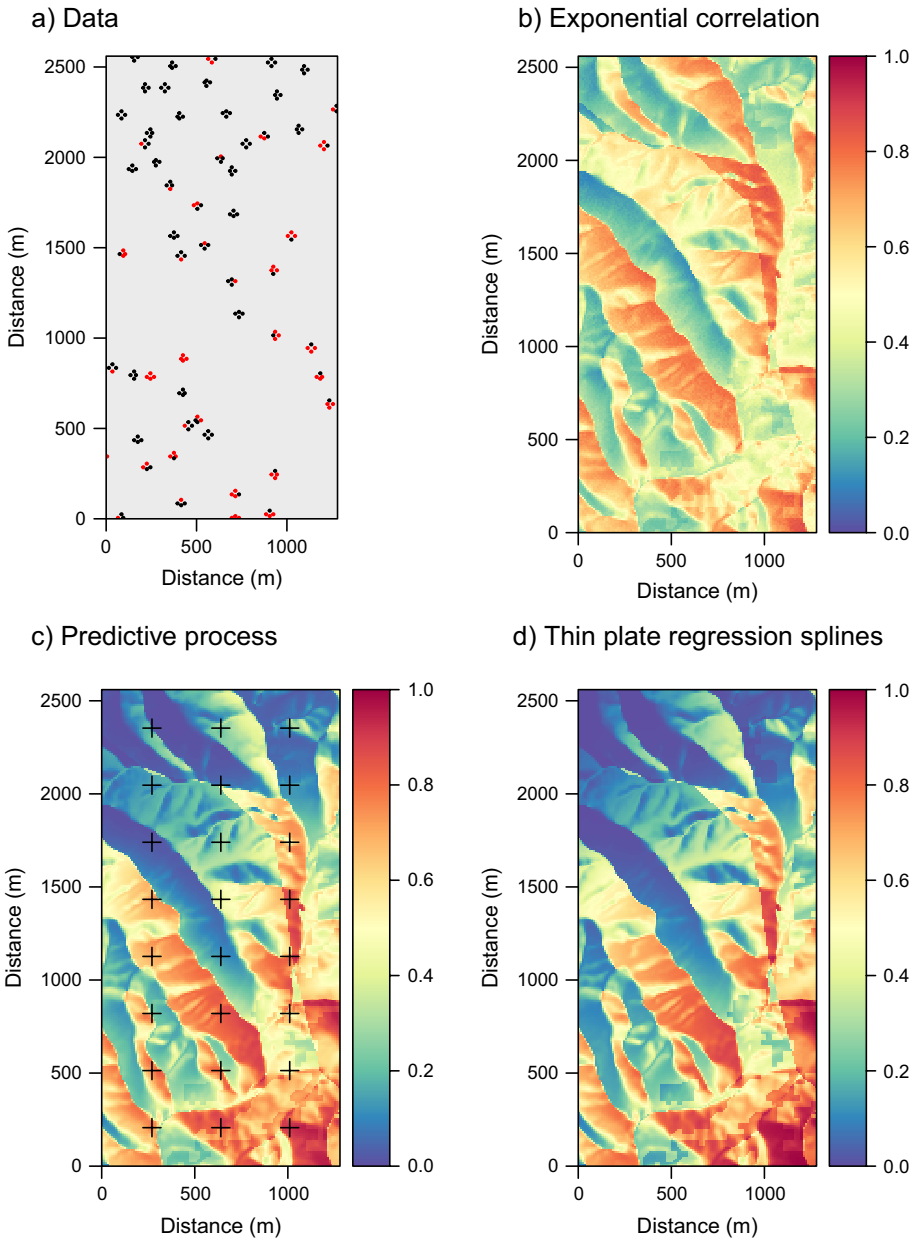

FIG. 4. Prediction domain from the Missouri Ozark Forest Ecosystem Project presented in Hooten et al. (2003). Red and black points (a) represent the 10 × 10 m plot locations that were sampled (n = 216) and whether pointed-leaved tick trefoil was present (red) or absent (black). The heat maps (b–d) show the predicted probability of occurrence in 32,768 plots from a binary spatial regression model (b; Eq. 19), a reduced dimension binary spatial regression model using predictive process basis functions (Eq. 21) with knots located within the prediction domain represented by + (c), and a generalized additive model that uses thin plate regression splines (d).

ecological process. A suitable second-order spatial binary model for presence–absence data is

$$\begin{aligned}\mathbf{y} &\sim \text{Bernoulli}\,(g(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\eta}))\\ \boldsymbol{\eta} &\sim \text{N}\left(\mathbf{0},\sigma_\alpha^2\mathbf{R}(\phi)\right),\end{aligned} \tag{19}$$

where $\mathbf{y}$ is an $n \times 1$ vector with elements equal to 1 if the species is present and 0 if the species is absent at a sampled location, $g(\cdot)$ is an inverse link function, and $\boldsymbol{\eta}$ is a vector of spatially autocorrelated random effects. As in Hooten et al. (2003), we specified the correlation matrix in Eq. 19 using an exponential correlation function

$$r_{ij}(\phi) = e^{-d_{ij}/\phi}, \tag{20}$$

where d_{ij} is the distance between locations i and j. Although there are numerous ways to implement the binary regression model, we adopt a Bayesian approach. The associated prior distributions and covariates are described in Hooten et al. (2003). After fitting the model, we predicted the probability of occurrence at all 32,768 plots within the study area. The predicted probability of occurrence depends on several covariates and has a strong spatial component (Fig. 4b).

Evaluating the likelihood for any second-order model requires inverting the correlation matrix $\mathbf{R}(\phi)$. For the geostatistical (continuous space) spatial model, inverting the correlation matrix has a computational cost that increases according to the cube of the sample size. For this example, when $n = 216$, fitting the Bayesian spatial model requires approximately 45-s per 1,000 MCMC samples obtained on a laptop computer with a 2.8-GHz quad-core processor, 16 GB of RAM, and optimized basic linear algebra subprograms, but would require approximately 1 h to obtain 1,000 MCMC samples from the same model if the sample size was $n = 1{,}000$.

For large spatial data sets, a variety of computationally efficient implementations can be used to model the spatial autocorrelation. The majority of efficient methods involve modeling the spatial autocorrelation using basis functions and a first-order model specification, which can be specified to result in dimension reduction. Unlike the Gaussian kernel basis function that approximates a Gaussian correlation function (see bioluminescence example), there is no kernel basis function that approximates an exponential covariance function (see Higdon 2002: fig. 2 or Banerjee et al. 2014:387). Therefore, we illustrate dimension reduction, using two different types of basis functions: the predictive process and thin plate regression splines. The predictive process is similar to kernel regression methods, except the basis expansion is slightly different and the basis coefficients are correlated in a reduced dimension of geographic space (Banerjee et al. 2008, 2014). The predictive process approach models the spatial process by smoothing over a finite number of representative locations as follows

$$\begin{aligned}\mathbf{Z}(\phi) &\equiv \mathbf{C}(\phi)\mathbf{R}^*(\phi)^{-1}\\ \boldsymbol{\alpha} &\sim \text{N}\left(0,\sigma_\alpha^2\mathbf{R}^*(\phi)\right),\end{aligned} \tag{21}$$

where $\mathbf{R}^*(\phi)$ is the $m \times m$ correlation matrix for the preselected knots (Fig. 4c) and $\mathbf{C}(\phi)$ is the $n \times m$ cross-correlation matrix between the observed data and knots. By examining Eq. 21, it can be seen that the predictive process is similar to kernel basis functions, but relies on a correlation function to specify the matrix of basis vectors and the spatial correlation among basis coefficients. Using the predictive process method with $m = 50$ knots, the Bayesian model requires approximately 3-s per 1,000 MCMC samples obtained and the predicted probability of occurrence appears similar when compared to the second-order spatial model (cf. Fig. 4b, c; Appendix S5). Furthermore, the predictive process method can be implemented using readily available software (Finley et al. 2015).

Generalized additive models are similar to models that use spatial random effects, but rely on techniques and basis functions commonly used in semiparametric regression (Ruppert et al. 2003). Specifying a GAM typically requires choosing a type of basis function, the number of basis vectors, and the location and number of knots. A common difference between the previous methods we have demonstrated and GAMs is the type of basis functions used. Many different basis functions are used to specify GAMs and introductions can be found in Ruppert et al. (2003), Wood (2006), Hastie et al. (2009), and James et al. (2013). Although GAMs can be implemented under a Bayesian paradigm (Crainiceanu et al. 2005, Gelman et al. 2013: Chapter 20), penalized maximum likelihood methods are commonly used (Wood 2006). For illustrative purposes, we implement a GAM using thin plate regression splines to model the spatial autocorrelation. Briefly, thin plate regression splines are a type of basis function that result in a one- or two-dimensional smooth effect of the spatial autocorrelation. Using a GAM framework, thin plate regression splines can be implemented in standard software and may be particularly useful for very large data sets (e.g., $n \approx 10^6$), requiring approximately 2-s to fit the model to our data using 50 basis coefficients (Wood et al. 2015). The predicted probability of occurrence is shown in Fig. 4d and is comparable to both specifications of the Bayesian spatial model (Fig. 4). We expected similarity between the GAM and the spatial model because there is a connection between first-order models that use spline basis functions and second-order spatial models (Nychka 2000).

Discussion

Autocorrelation: the two cultures

"What is one person's (spatial) covariance structure may be another person's mean structure (Cressie 1993:25)." This quote highlights that, within subfields of statistics that focus on dependent data (e.g., spatial statistics), there is no general consensus on whether the influence of autocorrelation should be specified in the mean or covariance structure of a probability distribution. The utility of first-order specified models that use

basis functions and second-order specified models that use covariance functions for dependent data has been a topic of discussion for several decades among statisticians (e.g., Cressie 1989 and comments by Wahba 1990, Laslett 1994 and comments by Handcock et al. 1994 and Mächler 1994). As we have demonstrated, there are many cases where the two approaches result (exactly or approximately) in the same model. With regard to which method to use, there are entire books written about correlation functions from a single perspective (e.g., Kriging in a spatial context; Stein 1999) and about certain classes of basis functions (Nason 2010; Table 2). Given the diversity of approaches, it is difficult to make specific recommendations. Our goal is to encourage researchers to consider both perspectives, rather than one or the other.

First-order or second-order?

Models that use second-order specifications can be converted to the equivalent first-order specification to assess collinearity among basis vectors and covariates of interest (Hodges and Reich 2010, Hefley et al. 2016). Modeling the autocorrelation using a first-order specification can be beneficial when the autocorrelation does not follow a standard correlation function, such as the case with data collected from streams and rivers (Peterson and Ver Hoef 2010, Sampson 2010, Isaak et al. 2014) or moving animals (Buderman et al. 2016; Hooten and Johnson, *in press*). The first-order specification might be more appealing when specifying theory-based ecological models (e.g., using partial differential equations) because the first-order model specification is naturally hierarchical (Wikle and Hooten 2010). With a first-order specification, the conditional distribution of the data (or unobserved latent process) can be selected and knowledge of the process can be incorporated into the mean structure (Wikle 2003, Hooten and Wikle 2008, Wikle and Hooten 2010; Williams et al., *in press*). Second-order models can also be based on ecological theory (Bolker and Pacala 1997, 1999), but may be more challenging than first-order models to understand and specify (Hanks, *in press*); however, second-order specifications can facilitate more computationally efficient algorithms by exchanging numerical integration algorithms for analytical solutions. Thus, many contemporary models for autocorrelation are specified in terms of first-order structure and then converted to second-order structure for implementation (Finley et al. 2015).

Choosing basis functions

Choosing basis functions requires an understanding of both the underlying ecological process and the properties of the basis functions. For example, a property of the polynomial basis function is that it has a global support; thus, an observation at one location influences the fit of the model at another location, regardless of how far apart the two locations are (Table 1). This is why polynomial basis expansions often fail to model fine scale structure (cf. Figs. 1b and 2b). From an ecological perspective, the global support of polynomial basis functions implies that the underlying ecological process is connected across the entire space of interest. If the ecological process is thought to have discontinuities, then basis functions that capture discontinuous structure and have compact support are a better choice (e.g., the uniform kernel used in Example 2; Table 2).

When selecting a basis function to model autocorrelation, standard model checking procedures are critical to ensure that model assumptions are met (e.g., checking for correlated residuals, collinearity, lack of fit, overfitting). Formal model selection may also be useful for selecting the optimal basis functions (Gelfand et al. 2012, Gelman et al. 2013: Chapter 20, Hooten and Hobbs 2015).

Computational considerations may be important when choosing a basis function. For example, orthogonal basis functions often result in more stable computational algorithms because the basis vectors are independent, obviating collinearity among basis vectors. We illustrated only a few of the many basis functions that could be used; thus, we recommend that practitioners become familiar with the variety of options to ensure that the chosen basis function matches the goals of the study. To facilitate this, we provided a brief summary of common basis functions, their properties, and useful references in Table 2.

Implementation

Typically, only a small number of covariates are included in a regression model, but one may want to include as many or more basis vectors as there are observations. For example, there are as many eigenvectors as there are unique locations in the dataset when the correlation matrix is specified using a correlation function. When many basis vectors are used to model autocorrelation, the model can overfit the data. Adding constraints to high-dimensional estimation problems is a common technique to prevent overfitting. Such methods include regularization, penalized maximum likelihood estimation (e.g., ridge regression), treating the basis coefficients as random effects, or using a prior distribution that induces shrinkage (regularization) in a Bayesian model. There are important connections among methods that impose constraints to prevent overfitting that we have not presented here, but are important to understand when implementing models that use basis functions (Hooten and Hobbs 2015).

When fitting models to data sets where dimension reduction is required, there is a trade-off between the reduction in dimension and the fit of the model. The fit of the model is influenced by dimension reduction because choosing the number of basis vectors to include in a model is an implicit form of regularization (Gelman et al. 2013: Chapter 20, Hooten and Hobbs 2015). Determining which basis functions are optimal for approximating correlation

functions, how many basis vectors are needed, and the locations of knots are active areas of research (Gelfand et al. 2012). A general rule of thumb is to choose fewer basis vectors than the number of unique locations in the data set, but as large as possible given the computational restrictions so that the predictions are accurate (e.g., for out-of-sample data). A detailed summary of dimension reduction approaches is beyond the scope of this paper, but technical introductions can be found in Paciorek (2007*b*), Wikle (2010), Cressie and Wikle (2011), and Banerjee et al. (2014).

Basis function model specifications have also become popular in spatio-temporal modeling, both in environmental (Wikle 2002) and ecological applications (Hooten and Wikle 2007). In practice, we find that understanding the properties of basis functions is critical to implementing computationally efficient Bayesian hierarchical models that account for spatial, temporal, or spatio-temporal autocorrelation. In addition, using basis functions as part of a Bayesian hierarchical model makes many spatiotemporal models accessible to users of JAGS, NIMBLE, Stan, and WinBugs (Crainiceanu et al. 2005, Wood 2016). Basis function components can be added to existing hierarchical models to account for autocorrelation using the tools and techniques presented in this paper. Furthermore, for many standard ecological models (e.g., Poisson, negative binomial regression, etc.), basis function models are implemented under a GAM framework in R packages such as mgcv (Wood 2006) or in packages specifically developed for spatial and spatio-temporal data analysis such as spBayes (Finley et al. 2015). If the basis functions are not parameter dependent (e.g., Moran eigenvector maps), basis functions can also be used in standard software for linear or generalized linear mixed models (e.g., R packages lme4 and MCMCglmm; Griffith and Peres-Neto 2006).

Inference and collinearity

For some applications, collinearity among covariates and basis vectors might occur and the development of remedial methods is a current topic of research in spatial statistics (Reich et al. 2006, Hodges and Reich 2010, Paciorek 2010, Hodges 2013, Hughes and Haran 2013, Hanks et al. 2015, Hefley et al. 2017). The effects of collinearity among covariates and basis vectors have been noted in the ecological literature as well, particularly in a spatial context (Kühn 2007, Bini et al. 2009, Hooten et al. 2013, Johnson et al. 2013), in time series generated from animal movement (Fieberg and Ditmer 2012), and in population trajectories (Hefley et al. 2016). In our experience, addressing collinearity among covariates and basis vectors is a difficult challenge in applied problems. In some cases, the conventional wisdom that applies to collinearity among covariates can also be applied to basis vectors, but new intuition is needed when basis coefficients are treated as random effects (Hodges and Reich 2010, Paciorek 2010, Hodges 2013, Hanks et al. 2015, Murakami and Griffith 2015). As with collinearity among covariates in linear regression models, there is no clear remedy for extreme cases.

Conclusion

Ecologists face many choices when specifying models. One important choice is how to model autocorrelation, which is not limited to specific domains and can occur in any space (e.g., covariate space, time, three-dimensional Euclidean space). Many methods used to model autocorrelation are general and can be understood as generalized linear mixed models that employ basis expansions and treat basis coefficients as random effects. Using the basis function approach, we find that many of the commonly used ecological models can be modified to incorporate autocorrelation.

Acknowledgments

We thank Paul Conn, Evan Cooch, Perry de Valpine, Devin Johnson, Maxwell Joseph, Jay Ver Hoef, Hadley Wickham, and four anonymous reviewers for valuable insight and early discussions about this work. The authors acknowledge support for this research from USGS G14AC00366 and NSF DMS 1614392. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2014. Hierarchical modeling and analysis for spatial data. CRC Press, Boca Raton, Florida, USA.

Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang. 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society B 70:825–848.

Barry, R. P., and J. Ver Hoef. 1996. Blackbox kriging: spatial prediction without specifying variogram models. Journal of Agricultural, Biological, and Environmental Statistics 1: 297–322.

Bini, M., et al. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography 32:193–204.

Bishop, C. 2006. Pattern recognition and machine learning. Springer, New York, New York, USA.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24: 127–135.

Bolker, B., and S. W. Pacala. 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theoretical Population Biology 52: 179–197.

Bolker, B. M., and S. W. Pacala. 1999. Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. American Naturalist 153:575–602.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73: 1045–1055.

Buderman, F. E., M. B. Hooten, J. S. Ivan, and T. M. Shenk. 2016. A functional model for characterizing long-distance movement behaviour. Methods in Ecology and Evolution 7: 264–273.

Conn, P. B., D. S. Johnson, J. M. V. Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235–252.

Crainiceanu, C., D. Ruppert, and M. P. Wand. 2005. Bayesian analysis for penalized spline regression using WinBUGS. Journal of Statistical Software 14:1–24.

Cressie, N. 1989. Geostatistics. American Statistician 43: 197–202.

Cressie, N. 1993. Statistics for spatial data. John Wiley & Sons, Hoboken, New Jersey, USA.

Cressie, N., and C. Wikle. 2011. Statistics for spatio-temporal data. Wiley, Hoboken, New Jersey, USA.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating density dependence, process noise, and observation error. Ecological Monographs 76:323–341.

Dormann, C. F., et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677.

Fieberg, J., and M. Ditmer. 2012. Understanding the causes and consequences of animal movement: a cautionary note on fitting and interpreting regression models with time-dependent covariates. Methods in Ecology and Evolution 3:983–991.

Finley, A. O., S. Banerjee, and A. E. Gelfand. 2015. spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software 63: 1–28.

Gelfand, A. E., S. Banerjee, and A. O. Finley. 2012. Spatial design for knot selection in knot-based dimension reduction models. Pages 142–169 *in* J. Mateu and W. G. Müller, editors. Spatio-temporal design: advances in efficient data acquisition. Wiley, Hoboken, New Jersey, USA.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. 2013. Bayesian data analysis. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Gillibrand, E., A. Jamieson, P. Bagley, A. Zuur, and I. Priede. 2007. Seasonal development of a deep pelagic bioluminescent layer in the temperate NE Atlantic Ocean. Marine Ecology Progress Series 341:37–44.

Griffith, D. A., and P. R. Peres-Neto. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87:2603–2613.

Handcock, M. S., K. Meier, and D. Nychka. 1994. Comment on kriging and splines: an empirical comparison of their predictive performance in some applications. Journal of the American Statistical Association 89:401–403.

Hanks, E. M. *In press*. Modeling spatial covariance using the limiting distribution of spatio-temporal random walks. Journal of the American Statistical Association. http://dx.doi.org/10.1080/01621459.2016.1224714

Hanks, E. M., E. M. Schliep, M. B. Hooten, and J. A. Hoeting. 2015. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics 26:243–254.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. Second edition. Springer Series in Statistics. Springer, New York, New York, USA.

Hefley, T. J., and M. B. Hooten. 2016. Hierarchical species distribution models. Current Landscape Ecology Reports 1: 87–97.

Hefley, T. J., M. B. Hooten, J. M. Drake, R. E. Russell, and D. P. Walsh. 2016. When can the cause of a population decline be determined? Ecology Letters 19:1353–1362.

Hefley, T. J., A. J. Tyre, and E. E. Blankenship. 2013. Statistical indicators and state–space population models predict extinction in a population of bobwhite quail. Theoretical Ecology 6:319–331.

Hefley, T. J., M. B. Hooten, E. M. Hanks, R. E. Russell, and D. P. Walsh. 2017. The Bayesian group lasso for confounded spatial data. Journal of Agricultural, Biological, and Environmental Statistics. DOI:10.1007/s13253-016-0274-1

Higdon, D. 2002. Space and space-time modeling using process convolutions. Pages 38–56 *in* C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, editors. Quantitative methods for current environmental issues. Volume 3754. Springer, New York, New York, USA.

Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.

Hodges, J. S. 2013. Richly parameterized linear models: additive, time series, and spatial models using random effects. CRC Press, Boca Raton, Florida, USA.

Hodges, J. S., and B. J. Reich. 2010. Adding spatially-correlated errors can mess up the fixed effect you love. American Statistician 64:325–334.

Hoeting, J. A. 2009. The importance of accounting for spatial and temporal correlation in analyses of ecological data. Ecological Applications 19:574–577.

Hooten, M. B., E. M. Hanks, D. S. Johnson, and M. W. Alldredge. 2013. Reconciling resource utilization and resource selection functions. Journal of Animal Ecology 82: 1146–1154.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85: 3–28.

Hooten, M. B., and D. S. Johnson. *In press*. Basis function models for animal movement. Journal of the American Statistical Association. http://dx.doi.org/10.1080/01621459.2016.1246250

Hooten, M. B., D. R. Larsen, and C. K. Wikle. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. Landscape Ecology 18: 487–502.

Hooten, M. B., and C. K. Wikle. 2007. Shifts in the spatio-temporal growth dynamics of shortleaf pine. Environmental and Ecological Statistics 14:207–227.

Hooten, M. B., and C. K. Wikle. 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. Environmental and Ecological Statistics 15:59–70.

Hughes, J., and M. Haran. 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society B 75: 139–159.

Isaak, D. J., et al. 2014. Applications of spatial statistical network models to stream data. Wiley Interdisciplinary Reviews: Water 1:277–294.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An introduction to statistical learning. Springer, New York, New York, USA.

Johnson, D. S., R. R. Ream, R. G. Towell, M. T. Williams, and J. D. L. Guerrero. 2013. Bayesian clustering of animal abundance trends for inference and dimension reduction. Journal of Agricultural, Biological, and Environmental Statistics 18:299–313.

Katzfuss, M. *In press*. A multi-resolution approximation for massive spatial datasets. Journal of the American Statistical Association. http://dx.doi.org/10.1080/01621459.2015.1123632

Krainski, E., F. Lindgren, D. Simpson, and H. Rue. 2016. The R-INLA tutorial on SPDE models. http://www.math.ntnu.no/inla/r-inla.org/tutorials/spde/spde-tutorial.pdf

Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distributions 13:66–69.

Laslett, G. M. 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications. Journal of the American Statistical Association 89:391–400.

Lee, Y., J. A. Nelder, and Y. Pawitan. 2006. Generalized linear models with random effects: unified analysis via H-likelihood. CRC Press, Boca Raton, Florida, USA.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74:1659–1673.

Legendre, P., and M. J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107–138.

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

Lindgren, F., H. Rue, and J. Lindström. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society B 73:423–498.

Littell, R. C., W. W. Stroup, G. A. Milliken, R. D. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. SAS Institute, Cary, North Carolina, USA.

Lorenz, E. N. 1956. Empirical orthogonal functions and statistical weather prediction. http://www.o3d.org/abracco/Atlantic/Lorenz1956.pdf

Mächler, M. 1994. Comment on kriging and splines: an empirical comparison of their predictive performance in some applications. Journal of the American Statistical Association 89:403–405.

Murakami, D., and D. A. Griffith. 2015. Random effects specifications in eigenvector spatial filtering: a simulation study. Journal of Geographical Systems 17:311–331.

Nason, G. 2010. Wavelet methods in statistics with R. Springer Science & Business Media, New York, New York, USA.

Nychka, D. W. 2000. Spatial-process estimates as smoothers. Pages 393–424 *in* M. G. Schimek, editor. Smoothing and regression: approaches, computation, and applications. Wiley, Hoboken, New Jersey, USA.

Paciorek, C. J. 2007*a*. Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. Journal of Statistical Software 19:1–38.

Paciorek, C. J. 2007*b*. Computational techniques for spatial logistic regression with large data sets. Computational Statistics and Data Analysis 8:3631–3653.

Paciorek, C. 2010. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. Statistical Science 25:107–125.

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

Reich, B. J., J. S. Hodges, and V. Zadnik. 2006. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 62:1197–1206.

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. Statistical Science 6:15–32.

Rue, H., and L. Held. 2005. Gaussian Markov random fields: theory and applications. CRC Press, Boca Raton, Florida, USA.

Ruppert, D., P. Wand, and R. Carroll. 2003. Semiparametric regression. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK.

Sampson, P. D. 2010. Constructions for nonstationary spatial processes. Pages 119–130 *in* A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, editors. Handbook of spatial statistics. CRC Press, Bocan Raton, Florida, USA.

Stein, M. L. 1999. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media, New York, New York, USA.

Stroup, W. W. 2012. Generalized linear mixed models: modern concepts, methods and applications. CRC Press, Boca Raton, Florida, USA.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234–240.

Veech, J. A. 2006. Increasing and declining populations of northern bobwhites inhabit different types of landscapes. Journal of Wildlife Management 70:922–930.

Wahba, G. 1990. Comment on Cressie. American Statistician 44:255–258.

Wikle, C. K. 2002. A kernel-based spectral model for non-Gaussian spatio-temporal processes. Statistical Modelling 2: 299–314.

Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394.

Wikle, C. K. 2010. Low rank representations for spatial processes. Pages 107–118 *in* A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, editors. Handbook of spatial statistics. CRC Press, Boca Raton, Florida, USA.

Wikle, C. K., and M. B. Hooten. 2010. A general science-based framework for dynamical spatio-temporal models. Test 19: 417–451.

Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, M. R. Bower, and T. J. Hefley. *In press*. An integrated data model to estimate spatio-temporal occupancy, abundance, and colonization dynamics. Ecology. DOI: 10.1002/ecy.1643

Wood, S. 2006. Generalized additive models: an introduction with R. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, Boca Raton, Florida, USA.

Wood, S. N. 2016. Just Another Gibbs Additive Modeller: Interfacing JAGS and mgcv. https://arxiv.org/abs/1602.02539

Wood, S. N., Y. Goude, and S. Shaw. 2015. Generalized additive models for large data sets. Journal of the Royal Statistical Society C 64:139–155.

Zuur, A., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer Science & Business Media, New York, New York, USA.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.1674/suppinfo