Spatial Autoregressive Models for Statistical Inference from Ecological Data

Jay M. Ver Hoef[1], Erin E. Peterson[2], Mevin B. Hooten[3], Ephraim M. Hanks[4], and Marie-Josée Fortin[5]

[1]Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115
tel: (907) 456-1995 E-mail: jay.verhoef@noaa.gov
[2] ARC Centre for Excellence in Mathematical and Statistical Frontiers (ACEMS) and the Institute for Future Environments, Queensland University of Technology
[3] U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology
Department of Statistics, Colorado State University
[4] Department of Statistics, The Pennsylvania State University
[5] Department of Ecology and Evolutionary Biology, University of Toronto

September 8, 2017

Abstract

Ecological data often exhibit spatial pattern, which can be modeled as autocorrelation. Conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models are network-based models (also known as graphical models) specifically designed to model spatially autocorrelated data based on neighborhood relationships. We identify and discuss six different types of practical ecological inference using CAR and SAR models, including: 1) model selection, 2) spatial regression, 3) estimation of autocorrelation, 4) estimation of other connectivity parameters, 5) spatial prediction, and 6) spatial smoothing. We compare CAR and SAR models, showing their development and connection to partial correlations. Special cases, such as the intrinsic autoregressive model (IAR), are described. CAR and SAR models depend on weight matrices, whose practical development uses neighborhood definition and row-standardization. Weight matrices can also include ecological covariates and connectivity structures, which we emphasize, but have been rarely used. Trends in harbor seals (*Phoca vitulina*) in southeastern Alaska from 463 polygons, some with missing data, are used to illustrate the six inference types. We develop a variety of weight matrices and CAR and SAR spatial regression models are fit using maximum likelihood and Bayesian methods. Profile likelihood graphs illustrate inference for covariance parameters. The same data set is used for both prediction and smoothing, and the relative merits of each are discussed. We show the nonstationary variances and correlations of a CAR model and demonstrate the effect of row-standardization. We include several take-home messages for CAR and SAR models, including 1) choosing between CAR and IAR models, 2) modeling ecological effects in the covariance matrix, 3) the appeal of spatial smoothing, and 4) how to handle isolated neighbors. We highlight several reasons why ecologists will want to make use of autoregressive models, both directly and in hierarchical models, and not only in explicit spatial settings, but also for more general connectivity models.

Key Words: Conditional autoregressive, simultaneous autoregressive, CAR, SAR, IAR, geostatistics, prediction, smoothing

INTRODUCTION

Ecologists have long recognized that data exhibit spatial patterns (Watt, 1947). These patterns were often expressed as spatial autocorrelation (Sokal and Oden, 1978), which is the tendency for sites that are close together to have more similar values than sites that are farther from each other. When spatial autocorrelation exists in data, ecologists often use spatial statistical models because the assumption of independent errors is violated, making many conventional statistical methods inappropriate (Cliff and Ord, 1981; Legendre, 1993). Areal data are a type of spatial ecological data that involve polygons or area-referenced data with measured values from the polygons (e.g., animal counts from game management areas). Often, ecological data collected in nearby polygons are more similar than those farther apart due to similar habitat conditions, biological processes such as migration or dispersal, and human impacts or management interventions. For example, higher animal counts or occupancy often form spatial clusters on the landscape (Thogmartin et al., 2004; Poley et al., 2014; Broms et al., 2014), plant measurements from a set of plots may be spatially patterned (Agarwal et al., 2005; Bullock and Burkhart, 2005; Huang et al., 2013), or global species diversity can exhibit geographic patterns when represented as a coarse-scale grid (Tognelli and Kelt, 2004; Pedersen et al., 2014). For these types of spatial data, spatial information can be encoded using neighborhoods, which leads to spatial autoregressive models (Lichstein et al., 2002). The two most common spatial autoregressive models are the conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models (Haining, 1990; Cressie, 1993). CAR and SAR models form a large class of spatial statistical models. Ecological data often exhibit spatial pattern, and while CAR and SAR models have been used in ecology, they should be used more often. Our objective is to review CAR and SAR models in a practical way, so that their potential may be more fully realized and used by ecologists, and we begin with an overview of their many uses.

Statistical Inference from CAR and SAR Models

We motivate the uses of spatial autoregressive models by considering typical (and not so typical, but useful) objectives where CAR and SAR models have been used for statistical inference in ecological studies: 1) model selection, 2) spatial regression, 3) estimation of autocorrelation, 4) estimation of other connectivity parameters, 5) spatial prediction, and 6) spatial smoothing (Table 1). There are many other interesting objectives in ecology, but these six are especially relevant for spatial modeling with CAR and SAR. When residual spatial autocorrelation is found based on, for example, Moran's I (Moran, 1948; Sokal and Oden, 1978), none of the objectives in Table 1 could be accomplished rigorously (in a probabalistic framework, using likelihoods for model selection and parameter estimates with confidence intervals) without modeling spatial autocorrelation. When data are collected on spatial areal (also called lattice, Cressie, 1993) units, SAR and CAR models provide the most straighforward and well-studied approach for accomplishing any of these objectives. We motivate each objective in turn and provide examples of studies in which autoregressive models were used.

Model selection (objective 1) can reveal important relationships between the response (i.e., dependent variable) and predictor variables. There are a plethora of model comparison methods, or multimodel inferences, based on Akaike Information Criteria (AIC, Akaike, 1973), Deviance Information Criteria (DIC, Spiegelhalter et al., 2002), etc., that are generally available (e.g., Burnham and Anderson, 2002; Hooten and Hobbs, 2015). CAR and SAR covariance matrices may be part of some or all models, and choosing a model, or comparing various CAR and SAR models, may be an important goal of the investigation. For example, Cassemiro et al. (2007) compared classical regression models assuming independence with SAR models while simultaneously selecting covariates using AIC when studying metabolism in amphibians. Qiu and Turner (2015) used SAR models for random errors along with model averaging in a study of landscape heterogeneity. Tognelli and Kelt (2004) compared CAR and SAR based on autocorrelation in residuals, choosing SAR for an analysis of factors affecting mammalian species richness in South America. In recent theoretical developments, Song and De Oliveira (2012) provided details on comparing various CAR and SAR models using Bayes factors. Zhu et al. (2010) extended the least absolute shrinkage and selection operator (LASSO, Tibshirani, 1996) using the least angle regression algorithm (LARS, Efron et al., 2004) to CAR and SAR models.

Regression analysis (objective 2) focuses on understanding relationships between predictor and response variables. Gardner et al. (2010) used a spatial CAR regression model to show that the probability of wolverine occupancy depended on predictors related to elevation and human influence in the plots. Returning to an example above, Cassemiro et al. (2007) found that several environmental predictors, including temperature, net primary productivity, annual actual evapotransiration, etc., helped explain species richness for amphibians. Agarwal et al. (2005) used a CAR model to study the effect of landscape variables, including road and population density, on deforestation. Using a SAR model for the spread of invasive alien plant species, Dark (2004) found relationships with elevation, road density, and native plant species richness. Beale et al. (2010) provided a review of spatial regression methods, including CAR and SAR. In many of these models, the autoregressive component was a latent random effect in a generalized linear mixed model, (also viewed as a hierarchical model (Cressie et al., 2009) or a state-space model (de Valpine and Hastings, 2002)), where the response variable was count (Clayton and Kaldor, 1987), binary (Gardner et al., 2010), or ordinal (Agarwal et al., 2005). Later, we provide more discussion of CAR and SAR in hierarchical models.

Understanding the strength of autocorrelation in spatial data (objective 3) can reveal connectivity and interrelatedness of ecological systems. Gardner et al. (2010) used a Bayesian CAR model to estimate the autocorrelation parameter ρ, with credible intervals to show uncertainty. Lichstein et al. (2002) also provided estimates of the CAR autocorrelation parameter for three different bird species, along with likelihood ratio tests against the null hypothesis that they were zero. Similarly, but for SAR models, Bullock and Burkhart (2005) used likelihood ratio tests to show significant estimates of several thousand tree species/location combinations with both positive and negative autocorrelation parameters.

Objective 4, understanding direct covariate effects on autocorrelation, is almost never used in ecological models, or in other disciplines. Typically, for regression, we model covariates affecting the mean of the response variable. For example, for the ith response variable Y_i, $E[Y_i] = \mu_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots$, where β_p is the pth regression coefficient, and $x_{p,i}$ is the pth covariate for the ith variable. Here, covariates are only part of the fixed effects and hence affect autocorrelation indirectly through the residual error. Typically, autocorrelation is controlled by the single parameter ρ, which scales the strength of autocorrelation. However, as for the mean μ_i (and through the likelihood), we can model the effect of multiple measurements (covariates) *between* pairs of response variables (locations for spatial data). For example, if $\rho_{i,j}$ is the correlation between site i and j, Swe can let $\rho_{i,j} = \theta_0 + \theta_1 x_{1,i,j} + \ldots$, where $x_{1,i,j}$ is a covariate defined between the ith and jth locations (e.g., a variable thought to impede or promote animal dispersal or gene flow). This direct influence of covariates on autocorrelation may be of interest in ecological studies concerned with connectivity (for a landscape-genetic example, see Hanks and Hooten, 2013) and we provide an example of how graphical models (mathematical constructs of points, or "nodes," connected by lines, or "edges") can be used to address this objective later.

Prediction at unsampled locations (objective 5) is a common goal in spatial analyses. An example of prediction using CAR models is given in both Magoun et al. (2007) and Gardner et al.

(2010), who modeled occupancy of wolverines from aerial surveys (also see Johnson et al., 2013a). There were three types of observations: 1) plots that were surveyed with observed animals, 2) plots that were surveyed with no animals, and 3) unsurveyed plots. Predictions for unsurveyed plots provided probabilities of wolverine occurrence. Huang et al. (2013) predicted N_2O in pastures with missing samples using CAR models, and Thogmartin et al. (2004) used CAR models to predict Cerulean Warblers abundance in the midwest US. Despite these examples, and the fact that geostatistics and time series are largely focused on prediction (at unsampled locations) and forecasting (at unsampled times in the future), respectively, there are few examples of prediction using CAR and SAR models in ecology, or other disciplines.

To conceptualize smoothing (objective 6), imagine that disease rates in conservation districts are generally low, say less than 10% based on thousands of samples, but spatially patterned with areas of lower and higher rates. However, one conservation district has but a single sample that is positive for the disease. It would be unrealistic to estimate the whole conservation district to have a 100% disease rate based on that single sample. CAR and SAR models can be used to create rates that smooth over observed data by using values from nearby districts to provide better estimates. For examples, see Beguin et al. (2012) and Evans et al. (2016). Entire books have been written on the subject (e.g., Elliot et al., 2000; Pfeiffer et al., 2008; Lawson, 2013b), and spatial smoothing of diseases form the introductions to CAR and SAR models in many textbooks on spatial statistics (Cressie, 1993; Waller and Gotway, 2004; Schabenberger and Gotway, 2005; Banerjee et al., 2014). Smoothing generally occurs when there is a complete census of areal units (e.g., agriculatural production in plots, or disease counts from counties). In the past, ecologists often sampled from plots, and rarely had a complete census, so they used this objective infrequently. However, increasingly advanced instruments (e.g., LIDAR, Campbell and Wynne, 2011) are yielding remotely sensed data with complete spatial coverage, allowing more opportunties for smoothing. In addition, smoothing over measurement error is attractive for hierarchical (Cressie et al., 2009) and state-space (de Valpine and Hastings, 2002) models.

Our review shows that CAR and SAR models are used for many types of statistical inference from ecological data, yet some highly cited ecological papers have incorrectly compared CAR/SAR to geostatistical models, incorrectly formulated the CAR model, and have given incorrect relationships between CAR and SAR models (details are given in Appendix A). We emphasize that good statistical practice with CAR and SAR models depends on more and better information. When ecological data are collected in spatial areal units, CAR and SAR models are often the most appropriate approach for accounting for spatial autocorrelation, and are thus essential tools for making valid inference on spatial data. To understand them better, we first compare CAR and SAR to geostatistical models.

Autoregressive Models and Geostatistics

A common framework for statistical inference in ecology is regression or, more generally, a generalized linear model (GLM), in which variation in the response variable is modeled as a function of predictor variables (or covariates). A key assumption in these models is that each response variable is independent from all others, after accounting for the covariate effects. When the response variables are collected in space, it is very common for the residuals resulting from a regression or GLM analysis to show spatial autocorrelation. Such autocorrelation violates the independence assumption, and can make standard results, such as confidence or credible intervals, invalid (Cliff and Ord, 1981; Legendre, 1993).

Instead of assuming independence, spatial statistical models directly account for spatial autocorrelation through modeling the covariance matrix $\mathbf{\Sigma}$ of the residuals as a function of the locations where the response variable, contained in the vector $\mathbf{y}$, were collected. For example, when the observations are point-referenced (i.e., each y was collected at a location with known GPS coordinates), geostatistical methods are often used (e.g., Turner et al., 1991). In a geostatistical model, the covariance of two observations is modeled directly as a function of the distance between the spatial locations where the observations were collected. For example, under the exponential covariance model (Chiles and Delfiner, 1999, p. 84), covariance decays exponentially with distance d_{ij} between observations

$$\mathrm{Cov}(y_i, y_j) = \Sigma_{ij} = \sigma^2 e^{-d_{ij}/\phi}, \tag{1}$$

which makes observations that occur close to each other in space highly correlated, while observations very far from each other nearly independent. Extending a regression model to allow for spatial autocorrelation (e.g., Ver Hoef et al., 2001) keeps inference on regression parameters from being invalidated by residual autocorrelation.

Geostatistical models directly model the covariance between spatial locations, and have been developed specifically for point-referenced data. However, a wide range of ecological studies collect aggregate observations from areal regions such as quadrats or pre-specified spatial polygons. In this setting, one could use a geostatistical model, such as the exponential covariance model (1), but this requires specifying a point to represent each areal unit, for example the centroid of each areal unit (e.g., Ver Hoef and Cressie, 1993). While this is possible, another class of spatial covariance models have been developed specifically to take advantage of the characteristics of areal data, the *autoregressive* spatial models. In these models, a network of connections between neighboring areal units is specified, and spatial dependence is specified through a model that conditions on observations at neighboring locations. This conditional spatial dependence can be shown to define the *inverse* of a covariance matrix (also known as the precision matrix, the term that we will use henceforth). Inverting this precision matrix then results in a spatial covariance matrix $\mathbf{\Sigma}$ defined by the network structure of the neighbor relationships. We illustrate with a simple example next.

An Example of a Spatial Autoregressive Model

To introduce autoregressive models, and illustrate how the network structure of an autoregressive model results in spatial autocorrelation, we consider a simple setup in which observations are collected at nine locations arranged in a 3×3 grid (Fig. 1).

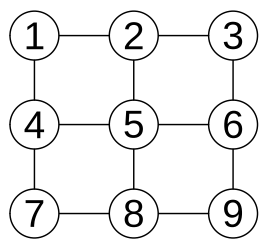

Figure 1: Spatial arrangement of sites in a simple 3×3 grid, where the numbers label each site.

In a geostatistical model where the observations were obtained at a point-referenced location, we could define spatial autocorrelation based on the distance between sites (Eq. 1). In an autoregressive model, spatial autocorrelation is defined by neighborhood (network) structure. In Fig. 1, we have defined neighborhood structure based on nearest neighbors in each cardinal direction. Neighbors are shown by the vertical and horizontal lines, so site 1 has two neighbors, labeled 2 and 4, etc. We can capture these neighborhood relationships in a matrix. For Fig. 1, let

$$
\mathbf{W} = \begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{pmatrix} \tag{2}
$$

be the matrix that indicates neighbor relationships, where a one in the jth column for the ith row indicates that site j is a neighbor of site i, otherwise the entry is zero. The rows and columns correspond to the numbered sites in Fig. 1. Under a CAR model for spatial autocorrelation, which we explore in more detail in the next Section, the spatial precision matrix $\mathbf{\Sigma}^{-1}$ is defined as $(\mathbf{I} - \rho\mathbf{W})$, where ρ is an autocorrelation parameter and $\mathbf{I}$ is a diagonal matrix of all ones. The resulting spatial covariance matrix $\mathbf{\Sigma}$, which describes spatial correlation based on the neighborhood structure in $\mathbf{W}$, is obtained by inverting the precision matrix

$$
\mathbf{\Sigma} = (\mathbf{I} - \rho\mathbf{W})^{-1} = \begin{pmatrix}
1.10 & 0.26 & 0.06 & 0.26 & 0.12 & 0.04 & 0.06 & 0.04 & 0.02 \\
0.26 & 1.16 & 0.26 & 0.12 & 0.29 & 0.12 & 0.04 & 0.07 & 0.04 \\
0.06 & 0.26 & 1.10 & 0.04 & 0.12 & 0.26 & 0.02 & 0.04 & 0.06 \\
0.26 & 0.12 & 0.04 & 1.16 & 0.29 & 0.07 & 0.26 & 0.12 & 0.04 \\
0.12 & 0.29 & 0.12 & 0.29 & 1.24 & 0.29 & 0.12 & 0.29 & 0.12 \\
0.04 & 0.12 & 0.26 & 0.07 & 0.29 & 1.16 & 0.04 & 0.12 & 0.26 \\
0.06 & 0.04 & 0.02 & 0.26 & 0.12 & 0.04 & 1.10 & 0.26 & 0.06 \\
0.04 & 0.07 & 0.04 & 0.12 & 0.29 & 0.12 & 0.26 & 1.16 & 0.26 \\
0.02 & 0.04 & 0.06 & 0.04 & 0.12 & 0.26 & 0.06 & 0.26 & 1.10
\end{pmatrix}, \tag{3}
$$

where in this example, $\rho = 0.2$.

We use this simple example to illustrate that 1) geostatistical models are defined by actual spatial distance, while CAR and SAR models are defined by neighborhoods, and 2) geostatistical models specify the covariance matrix $\mathbf{\Sigma}$ directly, whereas CAR and SAR models specify the precision matrix. We also note that it is not immediately obvious how the covariance matrix will behave based on our neighborhood definitions (because of the nonlinear nature of a matrix inverse). For example, the variances on the diagonal of Eq. 3 are not all equal. Notice the covariances for site 1 (the off-diagonal elements in the first row of $\mathbf{\Sigma}$), showing that site 1 is most highly correlated with sites 2 and 4, but also nonzero correlation with non-neighbors. Wall (2004) found some surprising and unusual behavior for CAR and SAR models. Our goal is to demystify CAR and SAR models, and provide practical suggestions for use of these models in ecological analyses.

CAR and SAR models are prevalent in the literature, and the six objectives listed above (Table 1) show that these models are essential tools for the analysis of ecological data. Our goals are as follows: 1) to explain how these models are obtained, 2) provide insight and intuition on how they work, 3) to compare CAR and SAR models, and 4) provide practical guidelines for their use. Using harbor seal (*Phoca vitulina*) trends, we provide an example for further illustration of the objectives given in Table 1. We then discuss important topics that have received little attention so far. For example, there is little guidance in the literature on handling isolated (unconnected) sites, or how to choose between a CAR model and a special case of the CAR model, the intrinsic autoregressive model (IAR). We provide such guidance, and finish with five take-home messages that deserve more attention.

SPATIAL AUTOREGRESSIVE MODELS

Spatial relationships for CAR and SAR models are based on a graphical model, or a network, where, using terminology from graphical models (e.g., Lauritzen, 1996; Whittaker, 2009), sites are called nodes (circles in Fig. 1) and connections are called edges (lines in Fig. 1). Edges can be defined in many ways, but a common approach is to create an edge between adjoining units in geographic space or any network space. Statistical models based on graphical spatial structure are sometimes known as Gaussian Markov random fields (e.g., Rue and Held, 2005). For notation, let

Y_i be a random variable used to model observations at the ith node, where $i = 1, 2, \ldots, N$, and all Y_i are contained in the vector $\mathbf{y}$. Then consider the spatial regression framework,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}, \tag{4}$$

where the goal is to model a first-order mean structure that includes covariates (i.e., predictor variables, $\mathbf{X}$, measured at the nodes) with regression coefficients $\boldsymbol{\beta}$, as well as a latent spatial random error $\mathbf{z}$, where $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and independent error $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I})$. Note that $\mathbf{z}$ is not directly measured, and instead must be inferred using a statistical model. The spatial regression framework becomes a spatial autoregressive model when the covariance matrix, $\boldsymbol{\Sigma}$, for $\mathbf{z}$, takes one of two main forms: 1) the SAR model,

$$\boldsymbol{\Sigma} \equiv \sigma_Z^2((\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}'))^{-1}, \tag{5}$$

or, 2) the CAR model,

$$\boldsymbol{\Sigma} \equiv \sigma_Z^2(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}. \tag{6}$$

Here, spatial dependence between Z_i and Z_j is modeled by $\mathbf{B} = \{b_{ij}\}$ and $\mathbf{C} = \{c_{ij}\}$ for the SAR and CAR models, respectively, where $b_{ii} = 0$ and $c_{ii} = 0$ and $\mathbf{M} = \{m_{ij}\}$ is a diagonal matrix (all off-diagonal elements are 0), where m_{ii} is proportional to the conditional variance of Z_i given all of its neighbors. The spatial dependence matrices are often developed as $\mathbf{B} = \rho\mathbf{W}$ and $\mathbf{C} = \rho\mathbf{W}$, where $\mathbf{W}$ is a weights matrix and ρ controls the strength of dependence. For the example in Eq. 3, we used a CAR model (Eq. 6) with $\mathbf{C} = \rho\mathbf{W}$, where $\mathbf{W}$ was given in Eq. 2, and $\sigma_Z^2 = 1$, $\mathbf{M} = \mathbf{I}$, and $\rho = 0.2$.

To help understand autoregressive models, consider partial correlation (e.g., Snedecor and

Cochran, 1980, pg. 361), which is the idea of correlation between two variables after "controlling," or holding fixed, the values for all other variables. If $\mathbf{\Sigma}^{-1} = \mathbf{\Omega} = \{\omega_{i,j}\}$, then the partial correlation between random variables Z_i and Z_j is $-\omega_{ij}/\sqrt{\omega_{ii}\omega_{jj}}$ (Lauritzen, 1996, pg. 120), which, for normally distributed data, is equivalent to conditional dependence. For the example in Fig. 1 and Eq. 2, $\mathbf{\Sigma}^{-1} = (\mathbf{I} - 0.2\mathbf{W})$ and so the partial correlation between sites 1 and 2 is 0.2. Thus, we can see that the CAR model, in particular, allows the modeler to directly specify partial correlations (or covariances), rather than (auto)correlation directly. That is, we are in control of specifying the off-diagonal matrix values of $\mathbf{W}$ in $\mathbf{\Sigma}^{-1} = \sigma_Z^2\mathbf{M}^{-1}(\mathbf{I} - \rho\mathbf{W})$, and therefore we are specifying the partial correlations. The SAR model case is similar, though instead of directly specifying partial correlations, as is done with $(\mathbf{I} - \mathbf{C})$ in the CAR model, the SAR specification involves modeling a square root, $(\mathbf{I} - \mathbf{B})$, of the precision matrix. Contrast this with geostatistics, where we are in control of specifying $\mathbf{\Sigma}$, and therefore we directly specify the (auto)correlations. In both cases, we generally use a functional parameterization, rather than specify every matrix entry individually. For CAR and SAR models, the specification is often based on neighbors (e.g., partial correlation exists between neighbors that share a boundary, conditional on all other sites), and for geostatistics, the specification is based on distance (e.g., correlation depends on an exponential decay with distance). For CAR models, if $c_{ij} = 0$, then sites i and j are partially uncorrelated; otherwise there is partial dependence. Note that diagonal elements b_{ii} and c_{ii} are always zero. For $\mathbf{z}$ (a SAR or CAR random variable) to have a proper statistical distribution, ρ must lie in a range of values that allows $(\mathbf{I} - \mathbf{B})$ to have an inverse and $(\mathbf{I} - \mathbf{C})$ to have positive eigenvalues; that is, ρ cannot be chosen arbitrarily, and its range depends on the weights in $\mathbf{W}$ (later, we discuss elements of $\mathbf{W}$ other than 0 and 1).

The statistical similarities among the SAR and CAR models are obvious; they both rely on a latent Gaussian specification, a weights matrix, and a correlation parameter. In that sense, both the SAR and CAR models can be implemented similarly. However, there are key differences between SAR and CAR models that are fundamentally important because they impact inference gained from these models. As such, we describe each model in more detail and provide practical advice.

SAR Models

One approach for building the SAR model begins with the usual regression formulation described in Eq. 4. Instead of modeling the correlation of $\mathbf{z}$ directly, an explicit autocorrelation structure is imposed,

$$\mathbf{z} = \mathbf{B}\mathbf{z} + \boldsymbol{\nu}, \tag{7}$$

where the spatial dependence matrix, $\mathbf{B}$, is relating $\mathbf{z}$ to itself, and $\boldsymbol{\nu} \sim \mathrm{N}(\mathbf{0}, \sigma_Z^2\mathbf{I})$. These models are generally attributed to Whittle (1954). Solving for $\mathbf{z}$, note that $(\mathbf{I} - \mathbf{B})^{-1}$ must exist (Cressie, 1993; Waller and Gotway, 2004), and then $\mathbf{z}$ has zero mean and covariance matrix $\boldsymbol{\Sigma} = \sigma_Z^2((\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}'))^{-1}$. The spatial dependence in the SAR model comes from the matrix $\mathbf{B}$ that causes the simultaneous autoregression of each random variable on its neighbors. When constructing $\mathbf{B} = \rho\mathbf{W}$, the weights matrix $\mathbf{W}$ does not have to be symmetric because it does not appear directly in the inverse of the covariance matrix (i.e., precision matrix). For the example in Eq. 2, the covariance matrix is

$$\boldsymbol{\Sigma} = ((\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W}'))^{-1} = \begin{pmatrix} 1.37 & 0.67 & 0.23 & 0.67 & 0.46 & 0.20 & 0.23 & 0.20 & 0.09 \\ 0.67 & 1.60 & 0.67 & 0.46 & 0.87 & 0.46 & 0.20 & 0.33 & 0.20 \\ 0.23 & 0.67 & 1.37 & 0.20 & 0.46 & 0.67 & 0.09 & 0.20 & 0.23 \\ 0.67 & 0.46 & 0.20 & 1.60 & 0.87 & 0.33 & 0.67 & 0.46 & 0.20 \\ 0.46 & 0.87 & 0.46 & 0.87 & 1.93 & 0.87 & 0.46 & 0.87 & 0.46 \\ 0.20 & 0.46 & 0.67 & 0.33 & 0.87 & 1.60 & 0.20 & 0.46 & 0.67 \\ 0.23 & 0.20 & 0.09 & 0.67 & 0.46 & 0.20 & 1.37 & 0.67 & 0.23 \\ 0.20 & 0.33 & 0.20 & 0.46 & 0.87 & 0.46 & 0.67 & 1.60 & 0.67 \\ 0.09 & 0.20 & 0.23 & 0.20 & 0.46 & 0.67 & 0.23 & 0.67 & 1.37 \end{pmatrix}, \tag{8}$$

using $\rho = 2$. Eq. 8 can be compared to Eq. 3. The constraints to allow $(\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}')$, when $\mathbf{B} = \rho\mathbf{W}$, to be a proper precision matrix are best explored through the eigenvectors and eigenvalues of $\mathbf{W}$. If $\lambda_{[1]} < 0$ is the smallest eigenvalue, and $\lambda_{[N]} > 0$ is the largest eigenvalue of $\mathbf{W}$, then $1/\lambda_{[1]} < \rho < 1/\lambda_{[N]}$ is sufficient for an inverse of $(\mathbf{I} - \mathbf{B})$ to exist. This is a sufficient, but not a necessary, condition. It is possible to specify a SAR model that does not satisfy this condition, but this is almost never done in practice, and we do not explore it further here. For Eq. 2, the minimum eigenvalue is -2.828 and the maximum is 2.828, with no eigenvalues equal to zero, so Eq. 2 can be made into a proper covariance matrix and ρ must be between $\pm$ 0.354.

The model created by Eq. 4 and Eq. 7 has been termed the "spatial error" model version of SAR models. An alternative is to simultaneously autoregress the response variable and the errors, $\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (Anselin, 1988), yielding the "SAR lag model" (Kissling and Carl, 2008),

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}, \tag{9}$$

which allows the matrix $\mathbf{W}$ to smooth covariates in $\mathbf{X}$ as well as creating autocorrelation in the error for $\mathbf{y}$ (e.g., Hooten et al., 2013). A final version is to simultaneously autoregress both response and a separate random effect $\boldsymbol{\nu}$ (e.g., "SAR mixed model" Kissling and Carl, 2008),

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\nu} + \boldsymbol{\varepsilon}. \tag{10}$$

CAR Models

The term "conditional" in the CAR model is used because each element of the random process is specified conditionally on the values of the neighboring nodes. The CAR model is typically specified as

$$Z_i | \mathbf{z}_{-i} \sim \mathrm{N}\left(\sum_{\forall c_{ij} \neq 0} c_{ij} z_j, m_{ii}\right), \tag{11}$$

where $\mathbf{z}_{-i}$ is the vector of all Z_j where $j \neq i$, $\mathbf{C}$ is the spatial dependence matrix with c_{ij} as its i, jth element, $c_{ii} = 0$, and $\mathbf{M}$ is zero except for diagonal elements m_{ii}. Note that m_{ii} may depend on the values in the ith row of $\mathbf{C}$. In this parameterization, the conditional mean of each Z_i is weighted by values at neighboring nodes. The variance component, m_{ii}, is also conditional on the neighboring nodes and is thus nonstationary, varying with node i. In contrast to SAR models, it is not obvious that Eq. 11 can lead to a full joint distribution for all random variables; however, this was demonstrated by Besag (1974) using Brook's lemma (Brook, 1964) and the Hammersley-Clifford theorem (Hammersley and Clifford, 1971; Clifford, 1990). For $\mathbf{z}$ to have a proper statistical distribution, $(\mathbf{I} - \mathbf{C})$ must have positive eigenvalues and $\mathbf{\Sigma} = \sigma_n^2(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$ must be symmetric, which requires that

$$\frac{c_{ij}}{m_{ii}} = \frac{c_{ji}}{m_{jj}}, \;\; \forall\, i, j. \tag{12}$$

For CAR models, when $\mathbf{C} = \rho\mathbf{W}$, $\mathbf{W}$ and ρ can be constrained in exactly the same way as for SAR models; if $1/\lambda_{[1]} < \rho < 1/\lambda_{[N]}$ for $\lambda_{[1]}$ the smallest, and $\lambda_{[N]}$ the largest eigenvalues of $\mathbf{W}$, then $\mathbf{I} - \rho\mathbf{W}$ will have positive eigenvalues.

A special case of the CAR model, called the intrinsic autoregressive model (IAR) (Besag and Kooperberg, 1995), occurs when Eq. 11 is parameterized as

$$Z_i \sim \mathrm{N}\left(\sum_{j \in \mathcal{N}_i} z_j / |\mathcal{N}_i|, \tau^2 / |\mathcal{N}_i|\right), \tag{13}$$

where $\mathcal{N}_i$ are all of the locations defined as neighbors of the ith location, $|\mathcal{N}_i|$ is the number of neighbors of the ith location, and τ^2 is a constant variance parameter. In Eq. 13, the conditional mean of each random variable is the average of its neighbors, and the variance is proportional to the inverse of the number of neighbors. Next, we discuss the creation of weights based on averages of neighboring values.

Row-standardization

We begin a discussion of the weights matrix, $\mathbf{W}$, which applies to both SAR and CAR models. Consider the simplest case, where a one in $\mathbf{W}$ indicates a connection (an edge) between sites i and j and a zero indicates no such connection, as in Eq. 2. For site i, let us suppose that there are $|\mathcal{N}_i|$ neighbors, so there are $|\mathcal{N}_i|$ ones in the ith row of $\mathbf{W}$. In terms of constructing random variables, this implies that Z_i is the *sum* of its neighbors, and summing increases variance. Generally, if left uncorrected, it will not be possible to obtain a covariance matrix in this case. As an analog, consider the first-order autoregressive (AR1) model from time series, where $Z_{i+1} = \phi Z_i + \nu_i$, and ν_i is an independent random variable. It is well-known that $\phi = 1$ is a random walk, and anything with $|\phi| \geqslant 1$ will not have a variance because the series "explodes" (e.g., Hamilton, 1994, pg. 53). There is a similar phenomenon for SAR and CAR models. In our simple example, for the construction $\rho\mathbf{W}$, the value $\rho|\mathcal{N}_i|$ effectively acts like ϕ, and both should be less than 1 to yield a proper statistical model. For example, consider the case where all locations are on an evenly-spaced rectangular grid of infinite size where each node is connected to 4 neighbors, called a rook's neighborhood; one each up, down, left, and right (as in Fig. 1). It is well-known that spatial autoregressive models for this example must have $|\rho| < 1/4$ (Haining, 1990, pg. 82) (compare this to the finite grid in Fig. 1, which had $|\rho| < 0.354$). More generally, $|\rho| < 1/n$ if all sites have exactly n neighbors, $|\mathcal{N}_i| = n$ for all sites, to keep variance under control. This leads to the idea of row-standardization.

If we divide each row in $\mathbf{W}$ by $w_{i,+} \equiv \sum_j w_{ij}$, then, again thinking in terms of constructing random variables, each Z_i is the *average* of its neighbors, which decreases variance. This is similar to what is expressed in Eq. 13. Row-standardization of Eq. 2 yields

$$\mathbf{W}_+ = \begin{pmatrix} 0.00 & 0.50 & 0.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.33 & 0.00 & 0.33 & 0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.00 & 0.00 \\ 0.33 & 0.00 & 0.00 & 0.00 & 0.33 & 0.00 & 0.33 & 0.00 & 0.00 \\ 0.00 & 0.25 & 0.00 & 0.25 & 0.00 & 0.25 & 0.00 & 0.25 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.00 & 0.33 & 0.00 & 0.00 & 0.00 & 0.33 \\ 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.33 & 0.00 & 0.33 & 0.00 & 0.33 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.50 & 0.00 \end{pmatrix}, \tag{14}$$

which is an asymmetric matrix. For the CAR models, if $\mathbf{W}_+$ is an asymmetric matrix with each row in $\mathbf{W}$ divided by $w_{i,+}$, then $m_{ii} = \tau^2/w_{i,+}$ (the ith diagonal element of $\mathbf{M}$) satisfies Eq. 12. Note that an additional variance parameter for $m_{i,i}$ will not be identifiable from σ_Z^2 in Eq. 6, so the row-standardized CAR model can be written equivalently as,

$$\mathbf{\Sigma} = \sigma_Z^2(\mathbf{I} - \rho\mathbf{W}_+)^{-1}\mathbf{M}_+ = \sigma_Z^2(\text{diag}(\mathbf{W1}) - \rho\mathbf{W})^{-1}, \tag{15}$$

where $\mathbf{1}$ is a vector of all ones and diag($\mathbf{v}$) creates a matrix of all zeros except the vector $\mathbf{v}$ is on the diagonal. For both CAR and SAR models, regardless of the number of neighbors, when using row standardization, it is sufficient for $|\rho| < 1$, which is very convenient. Row standardization simplifies the bounds of ρ and makes optimization easier to implement.

Moreover, consider again the case of an evenly-spaced rectangular grid of points, but this time of finite size, again using a rook's neighborhood. Using row standardization, points in the interior of the rectangle are averaged over 4 neighbors, and they will have smaller variance than those at the perimeter, averaged over 3 neighbors, and the highest variance will be locations in the corners, averaged over 2 neighbors. Hence, in general, variance increases toward the *perimeter*. Without row standardization, even when ρ controls overall variance, locations in the *middle*, summed over more neighbors, have higher variance than those at the perimeter. Using the example in Eq. 2,

$$\mathbf{\Sigma} = (\mathbf{I} - \rho\mathbf{W}_+)^{-1}\mathbf{M}_+ = \begin{pmatrix} 0.72 & 0.27 & 0.15 & 0.27 & 0.15 & 0.10 & 0.15 & 0.10 & 0.08 \\ 0.27 & 0.53 & 0.27 & 0.15 & 0.19 & 0.15 & 0.10 & 0.10 & 0.10 \\ 0.15 & 0.27 & 0.72 & 0.10 & 0.15 & 0.27 & 0.08 & 0.10 & 0.15 \\ 0.27 & 0.15 & 0.10 & 0.53 & 0.19 & 0.10 & 0.27 & 0.15 & 0.10 \\ 0.15 & 0.19 & 0.15 & 0.19 & 0.40 & 0.19 & 0.15 & 0.19 & 0.15 \\ 0.10 & 0.15 & 0.27 & 0.10 & 0.19 & 0.53 & 0.10 & 0.15 & 0.27 \\ 0.15 & 0.10 & 0.08 & 0.27 & 0.15 & 0.10 & 0.72 & 0.27 & 0.15 \\ 0.10 & 0.10 & 0.10 & 0.15 & 0.19 & 0.15 & 0.27 & 0.53 & 0.27 \\ 0.08 & 0.10 & 0.15 & 0.10 & 0.15 & 0.27 & 0.15 & 0.27 & 0.72 \end{pmatrix},$$

with $\rho = 0.8$. The variances are on the diagonal, and these should be compared to Eq. 3. For an error process in Eq. 4, higher variance near the perimeter makes more sense (as in many kriging error maps), and, with a more natural and consistent range of values for ρ, row-standardization is beneficial.

Using row-standardization, and setting $\rho = 1$ in Eq. 11 leads to the IAR model in Eq. 13. In our AR1 analogy, this is equivalent to $\phi = 1$. In this case, $\mathbf{\Sigma}^{-1}$ is singular (i.e., does not have an inverse), and $\mathbf{\Sigma}$ does not exist. It can be verified that Eq. 14 has a zero eigenvalue. While this may seem undesirable, random walks and Brownian motion are stochastic processes without covariance matrices (Codling et al., 2008). Considering how they are constructed, it helps to think of the variances and covariances being defined on the increments; the differences between adjacent variables. For these increments, the variances and covariances are well-defined. The IAR distribution is improper, however it is similarly well-defined on spatial increments or contrasts. To make the IAR proper, an additional constraint can be included, $\sum_i Z_i = 0$. In essence, this constraint allows all of the random effects to vary except one, which is subsequently used to ensure that the values sum to zero as a whole. Geometrically, the sum-to-zero constraint can be thought of as anchoring the process near zero for the purposes of random errors in a model. With such a constraint, the IAR model is appealing as an error process in Eq. 4, forming a flexible surface where there is no autocorrelation parameter ρ to estimate. The IAR model is called a first-order intrinsic Gaussian Markov random field (Rue and Held, 2005, p. 93); higher orders are possible but we do not discuss them here.

The Choice of Spatial Neighborhood Structure

There is little guidance in the literature on how to choose the neighborhood structure in autoregressive models. One reason for this is that there is rarely a clear scientific understanding of the mechanism behind spatial autocorrelation; rather, in most ecological modeling, our scientific understanding of the system is used to model the mean structure, and modeling spatial autocorrelation is a secondary consideration. The formulation in Eq. 4 suggests that the spatial random effect $\mathbf{z}$ can be thought of as a missing covariate that is spatially smooth, but there are other possibilities as well. Hanks (2017) shows that the long-time limiting distribution of a spatio-temporal random walk can result in a spatial random effect with SAR covariance, indicating that SAR models can be seen as the covariance that results when the spatio-temporal process being studied could be approximated by a random walk. This is an example of a SAR model arising from a mechanism that may match a scientific question.

In the absence of a scientific motivation for spatial autocorrelation, one way to view autoregressive models is as a modeling choice required to relax the assumption of independence of $\mathbf{y}$ in Eq. 4, conditional on $\mathbf{X}$ and $\mathbf{z}$, when it is not true. In a regression analysis, one might consider multiple transformations of a response variable to satisfy the assumption of normality of residuals. These transformations are not, in general, motivated by scientific understanding, but rather by modeling expediency. Similarly, in a spatial analysis, one might consider multiple autoregressive models, such as SAR and CAR models with different neighborhood structures. A final model could be chosen based on AIC, DIC, or other similar criteria. In this situation, there is little to be gained by trying to interpret the CAR or SAR model that best fits the data. Rather, the researcher should focus interpretation on mean effects (objective 2) or prediction (objective 5), and recognize that choosing a good neighborhood structure can improve both of these objectives.

The neighborhood structure of a CAR or SAR model depends on the connected nodes in the network; these are almost always defined as the areal units on which one has observations. This choice can have unintended consequences, as it implies that the process being studied only exists on the specified areal units. This would be appropriate, for example, when one is modeling recruitment of a species with a known geographic extent, and when the data collection has encompassed the entire range of the species. As noted above, autoregressive models that use row-standardization tend to have higher marginal variance at the perimeter of the network – this corresponds with the assumption that we are often less certain about the state of a system at its boundaries than we are in more central spatial locations.

This assumption makes little sense when the system being studied is known to extend beyond the spatial range of the study. In this case, there is no obvious reason to assume that higher variance would occur at the perimeter of the study region. Instead, it would be more appropriate to extend the range of the spatial random effect by creating a buffer region of areal units on the boundary of the study region (e.g., Lindgren et al., 2011). While these buffer areal units would not have observations associated with them, they would stabilize the marginal variance of the spatial random effect, and would be appropriate whenever the process under study is known to extend beyond the spatial domain of the data.

More Weighting – Accounting for Functional and Structural Connectivity

So far, we have reviewed standard spatial autoregressive models. Now, we want to consider their more general formulation as graphical, or network models. In general, the autoregressive component is an "error" process, and not often of primary interest (compared to prediction or estimating fixed effects parameters, $\boldsymbol{\beta}$). However, for ecological networks, there is a great deal of interest in studying spatial connectivity, or equivalently spatial autocorrelation. We discuss other weighting schemes for autoregressive models that have been very rarely, or never, used, but would provide valid autocorrelation models for studying connectivity in ecology. In particular, although the decomposition is not unique, we introduce weighting schemes for the $\mathbf{W}$ matrix that can separate and clarify structural and functional components in network connectivity. By structural, we mean correlation that is determined by physical proximity, such as geographic neighborhoods, a distance measure, etc. By functional, we mean correlation that is affected by dispersal, landscape characteristics, and other covariates of interest, which we illustrate next.

Consider a spatial network of nodes and edges, with the response variable measured at nodes, putting us in the setting of SAR and CAR models. Let $\mathbf{e}_{ij}$ be a characteristic of an edge between the ith and jth nodes. The structural aspects can be accommodated in the neighborhood structure – the binary representation of connectivity contains the idea of neighborhood structure. Then edge weights, w_{ij}, between the ith and jth nodes could combine functional and structural connectivity if they are modeled as,

$$w_{ij} = \begin{cases} f(\mathbf{e}_{ij}, \boldsymbol{\theta}), & j \in \mathcal{N}_i, \\ 0, & j \notin \mathcal{N}_i, \end{cases} \tag{16}$$

where $\boldsymbol{\theta}$ is a p-vector of parameters. To clarify, consider the case where $\mathbf{x}_i$ is a vector of p habitat characteristics of the ith node, $\mathbf{e}_{i,j} = (\mathbf{x}_i + \mathbf{x}_j)/2$, and $f(\mathbf{e}_{ij}, \boldsymbol{\theta}) = \exp(\mathbf{e}'_{ij}\boldsymbol{\theta})$ (Hanks and Hooten, 2013). This allows a model of the effect that habitat characteristics at the nodes has on connectivity. If $\theta_h < 0$, then an increase in the hth habitat characteristic results in a smaller edge weight and greater resistance to network connectivity. However, if $\theta_h > 0$, then an increase in the hth habitat characteristic results in a larger edge weight and less resistance to network connectivity. In this example, the mean of the habitat characteristics found at the two nodes, $(\mathbf{x}_i + \mathbf{x}_j)/2$, was used, but any other function of the two values could also be used (e.g., difference) if it makes ecological sense. Alternatively, $f(\mathbf{e}_{ij}, \boldsymbol{\theta})$ could be something that is directly measured on edges, such as a sum of pixel weights in a shortest path between two nodes from a habitat map.

For a matrix representation of Eq. 16, let $\mathbf{F}(\boldsymbol{\theta})$ be a matrix of functional relationships for *all* edges, let $\mathbf{B}$ be a binary matrix indicating neighborhood structure, and $\mathbf{W} = \mathbf{F}(\boldsymbol{\theta}) \odot \mathbf{B}$, where $\odot$ is the Hadamard (direct, or element by element) product. Then $\mathbf{F}(\boldsymbol{\theta}) \odot \mathbf{B}$ allows a decomposition for exploring structural and functional changes in connectivity by manipulating each separately. Of course, this must respect the restrictions described above for SAR and CAR models, and the parameters need to be estimated, which we discuss in the section on fitting methods.

Comparing CAR to SAR with Practical Guidelines

With a better understanding of SAR and CAR models, we now compare them more closely and make practical recommendations for their use; see also Wall (2004). First, we generally do not recommend versions of the SAR model given by Eq. 9 and Eq. 10. It is difficult to understand how smoothing/lagging covariates and extra random effects contribute to model performance, nor to our understanding, and these models performed poorly in ecological tests (Dormann et al., 2007; Kissling and Carl, 2008). Henceforth, we only discuss the error model defined by Eq. 7.

A SAR model can be written as a CAR model and vice versa, although almost all published accounts on their relationships are incomplete (Ver Hoef et al., 2017). Cressie (1993, pg. 408) demonstrated how a SAR model with four neighbors (rook's neighbor) results in a CAR model that involves all eight neighbors (queen's neighbor) plus rook's move to the second neighbors. It is evident from Eq. 5 that specifying first-order neighbors in $\mathbf{B}$ will result in non-zero partial correlations between second-order neighbors because of the product $(\mathbf{I}-\mathbf{B})(\mathbf{I}-\mathbf{B})'$ in the precision matrix. Hence, SAR models have a reputation as being less "local" (averaging over more neighbors, so causing more smoothing) than the CAR models. In fact, using the same construction $\rho\mathbf{W}$ for both SAR and CAR models, Wall (2004) showed that correlation (in $\boldsymbol{\Sigma}$, not partial correlation) increases more rapidly with ρ in SAR models than CAR models, which is also apparent when comparing Eq. 3 to Eq. 8.

Regarding restrictions on ρ, Wall (2004) also showed strange behavior for negative values of ρ. In geostatistics, there are very few models that allow negative spatial autocorrelation, and, when they do, it cannot be strong. Thus, in most situations, ρ may be constrained to be positive. The fact that $\mathbf{W}$ in SAR models is not required to be symmetric may seem to be an advantage over CAR models. However, we point out that this is illusory from a modeling standpoint, although it may help conceptually in formulating the models. For an analogy, again consider the AR1 model from time series. The model is specified as $Z_{i+1} = \phi Z_i + \nu_i$, so it seems like there is dependence only on previous times. However, the correlation matrix is symmetric, and $\mathrm{corr}(Z_i, Z_{i+t}) = \mathrm{corr}(Z_i, Z_{i-t}) = \phi^t$. Note also that this shows that specifying partial correlations as zero (or conditional independence), does not mean that marginal correlation is zero (i.e., $\mathrm{corr}(Z_i, Z_{i+t}) \neq 0$ for all t lags). The same is true for CAR and SAR models. In fact, the situation is less clear than for the AR1 models, where $\mathrm{corr}(Z_i, Z_{i+t}) = \phi^t$ regardless of i. For CAR and SAR models, two sites that have the same "distance" from each other will have different correlation, depending on whether they are near the center of the spatial network, or near the perimeter; that is, correlation is nonstationary, just like the variance as described in the Section on row-standardization.

CAR and SAR in Hierarchical Models

We now focus on the use of CAR and SAR spatial models within a hierarchical model. To discuss these models more specifically and concretely, in the example and following discussion, consider the following hierarchical structure that forms a general framework for all that follows,

$$\begin{aligned}
&\mathbf{y} \sim [\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\xi}],\\
&\boldsymbol{\mu} \equiv \mathbf{X}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon},\\
&\mathbf{z} \sim [\mathbf{z}|\boldsymbol{\Sigma}] \equiv \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}),\\
&\boldsymbol{\Sigma}^{-1} \equiv \mathbf{F}(\mathbf{N}, \mathbf{D}, \rho, \boldsymbol{\theta}, \ldots),\\
&\boldsymbol{\varepsilon} \sim [\boldsymbol{\varepsilon}|\sigma^2] \equiv \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),
\end{aligned} \tag{17}$$

where $[\cdot]$ denotes a generic statistical distribution (Gelfand and Smith, 1990), with the variable on the left of the bar and conditional variables or parameters on the right of the bar. Here, let $\mathbf{y}$ contain random variables for the potentially observable data, which could be further partitioned into $\mathbf{y} = (\mathbf{y}_o', \mathbf{y}_u')'$, where $\mathbf{y}_o$ are observed and $\mathbf{y}_u$ are unobserved. Then $[\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\xi}]$ is typically the data model, with a distribution such as Normal (continuous ecological data, such as plant biomass), Poisson (ecological count data, such as animal abundance), or Bernoulli (ecological binary data, such as occupancy), which depends on a mean $\boldsymbol{\mu}$ with link function g, and other parameters $\boldsymbol{\xi}$. The mean $\boldsymbol{\mu}$ has the typical spatial-linear mixed-model form, with design matrix $\mathbf{X}$ (containing covariates, or explanatory variables), regression parameters $\boldsymbol{\beta}$, spatially autocorrelated errors $\mathbf{z}$, and independent errors $\boldsymbol{\varepsilon}$. We let the random effects, $\mathbf{z}$, be a zero-mean multivariate-normal distribution with covariance matrix $\boldsymbol{\Sigma}$. In a geostatistical spatial-linear model, we would model $\mathbf{\Sigma}$ directly with covariance functions based on distance like the exponential, spherical, and Matern (Chiles and Delfiner, 1999). The variance σ^2, of the independent component $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, is called the nugget effect. However, in CAR and SAR models, and as described above, we model the precision matrix, $\mathbf{\Sigma}^{-1}$. We denote this as a matrix function, $\mathbf{F}$, that depends on other information (e.g., a neighborhood matrix $\mathbf{N} = \mathbf{B}$ or $\mathbf{C}$, a distance matrix $\mathbf{D}$, and perhaps others). We isolate the parameter ρ that controls the strength of autocorrelation. Note, however, there could be other parameters, $\boldsymbol{\theta}$, that form the functional relationships among $\mathbf{N}, \mathbf{D}, \ldots$, and $\mathbf{\Sigma}^{-1}$. In a Bayesian analysis, we could add further priors, but here we give just the essential model components that provide most inferences for ecological data. The model component to be estimated or predicted from Eq. 17 is identified in Table 1. Note that a joint distribution for all random quantities can be written as $[\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\nu}][\mathbf{z}|\mathbf{\Sigma}][\boldsymbol{\varepsilon}|\sigma^2]$, but the only observable data come from $\mathbf{y}$. The term likelihood is used when the joint distribution is considered a function of all unknowns, given the observed data, which we denote $L(\cdot|\mathbf{y})$, and this often forms the basis for fitting models (discussed next) and model comparison (Table 1).

Fitting Methods for Autoregressive Models

Maximum likelihood estimation is one of the most popular estimation methods for spatial models (Cressie, 1993), but it can be computationally expensive. Earlier, when computers were less powerful, methods were devised to trade efficiency (on bias and consistency) for speed, such as pseudolikelihood (Besag, 1975) and coding (Besag, 1974) for CAR models, among others (Cressie, 1993). Both CAR and SAR models are well-suited for maximum likelihood estimation (Banerjee et al., 2014). For spatial models, the main computational burden in geostatistical models is inversion of the covariance matrix; for CAR and SAR models, the inverse of the covariance matrix is what we actually model, simplifying computations (Paciorek, 2013). Thus, only the determinant of the covariance matrix needs computing, and fast methods are available (Pace and Barry, 1997a,b), while if matrices do need inverting, sparse matrix methods can be used (Rue and Held, 2005). In addition, for Bayesian Markov chain Monte Carlo methods (MCMC; Gelfand and Smith, 1990), CAR models are ready-made for conditional sampling because of their conditional specification.

Spatial autoregressive models are often used in generalized linear models, which can be viewed as hierarchical models, where the spatial CAR model is generally latent in the mean function in a hierarchical modeling framework. Indeed, one of their most popular uses is for "disease-mapping," whose name goes back to Clayton and Kaldor (1987); see Lawson (2013a) for book-length treatment. These models can be treated as hierarchical models (Cressie et al., 2009), where the data are assumed to arise from a count distribution, such as Poisson, but then the log of the mean parameter has a CAR/SAR model to allow for extra-Poisson variation that is spatially patterned (e.g., Ver Hoef and Jansen, 2007). Note that this provides a full likelihood, unlike the quasi-likelihood often used for overdispersion for count data (Ver Hoef and Boveng, 2007). A similar hierarchical framework has been developed as a generalized linear model for occupancy, which is a binary model, but then the logit (or probit) of the mean parameter has a CAR/SAR model to allow for extra-binomial variation that is spatially patterned (Magoun et al., 2007; Gardner et al., 2010; Johnson et al., 2013a; Broms et al., 2014; Poley et al., 2014). CAR and SAR models can be embedded in more complicated hierarchical models as well (e.g., Ver Hoef et al., 2014). Sometimes that may be too slow, and a fast general-purpose approach to fitting these types of hierarchical models, which depends in part on the sparsity of the CAR covariance matrix, is integrated nested Laplace approximation (INLA, Rue et al., 2009). INLA has been used in generalized linear models for ecological data (e.g., Haas et al., 2011; Aarts et al., 2013), spatial point patterns (Illian et al., 2013), and animal movement models (Johnson et al., 2013b), among others. The growing popularity of INLA is due in part to its fast computing for approximate Bayesian inference on the marginal distributions of latent variables.

EXAMPLE: HARBOR SEAL TRENDS

We used trends in harbor seals (*Phoca vitulina*) to illustrate the models and approaches for inference described in previous sections. Harbor seals are abundant along the northwest coast of the United States and Canada to Alaska (Pitcher and Calkins, 1979). Management of harbor seals is important due to subsistence and reliance on these animals by Native Americans (Wolfe et al., 2009). Consequently, interest in harbor seals led to many studies that have documented abundance and trend in Oregon and Washington (Harvey et al., 1990; Huber et al., 2001; Jeffries et al., 2003; Brown et al., 2005), British Columbia (Bigg, 1969; Olesiuk et al., 1990; Olesiuk, 1999), and Alaska (Pitcher, 1990; Frost et al., 1999; Small et al., 2003; Boveng et al., 2003; Ver Hoef and Frost, 2003; Mathews and Pendleton, 2006).

The study area is shown in Fig. 2 and contains 463 polygons used as survey sample units along the mainland, and around islands, in Southeast Alaska. Based on genetic sampling, this area has been divided into 5 different "stocks" (or genetic populations). Over a 14-year period, at various intervals per polygon, seals were counted from aircraft. Using those counts, a trend for each polygon was estimated using Poisson regression. Any polygons with less than two surveys were eliminated, along with trends (linear on the log scale) that had estimated variances greater than 0.1. This eliminated sites with small sample sizes. We treated the estimated trends, on the log scale, as raw data, and ignored the estimated variances. These data are illustrative because we expected the trends to show geographic patterns (more so than abundance which varied widely in polygons) and stock structure connectivity, along with stock structure differences in mean values. The data were also continuous in value, thus we modeled the trends with normal distributions to keep the modeling simpler and the results more evident. A map of the estimated trend values (that we henceforth treat as raw data) is given in Fig. 3, showing 463 polygons, of which 306 had observed values and 157 were missing.

For neighborhood structures, we considered three levels of neighbors. The first-order neighbors were based on any two polygons sharing one or more boundary point, and were computed using the *poly2nb* function in the *spdep* package (Bivand and Piras, 2015) in R (R Core Team, 2016). Some polygons were isolated, so they were manually connected to the nearest polygon in space using straight-line (Euclidean) distance between polygon centroids. The first-order neighbors are shown graphically in Fig. 4a with a close-up of part of the study area given in Fig. 4b. Let $\mathbf{N}_1$ be a matrix of binary values, where a 1 indicates two sites are first-order neighbors, and a 0 otherwise. Then second-order neighbors, which include neighbors of first-order neighbors, were easily obtained in the matrix $\mathbf{N}_2 = \mathcal{I}(\mathbf{N}^2)$. Here, $\mathcal{I}(\cdot)$ is an indicator function on each element of the matrix, being 0 only when that element is 0, and 1 otherwise. A close-up of some of the second-order neighbors is shown in Fig. 4c. The fourth-order neighbor matrix was obtained as $\mathbf{N}_4 = \mathcal{I}(\mathbf{N}_2^2)$, and a close-up is shown in Fig. 4d.

We considered covariance constructions that elaborated the three different neighborhood definitions. Let $\mathbf{N}_i; i = 1, 2, 4$ be a neighborhood matrix as described in the previous paragraph. Let $\mathbf{S}$ be a matrix of binary values that indicate whether two sites are in different stocks; that is, if site i and j are in the same stock, then $\mathbf{S}[i,j] = 0$, otherwise $\mathbf{S}[i,j] = 1$. Finally, let the i,jth entries in $\mathbf{D}$ be the Euclidean distance between the centroids of the ith and jth polygons. Then the most elaborate CAR/SAR model we considered was

$$\mathbf{W} = \mathbf{N}_i \odot \mathbf{F}(\boldsymbol{\theta}) = \mathbf{N}_i \odot \exp(-\mathbf{S}/\theta_1) \odot \exp(-\mathbf{D}/\theta_2). \tag{18}$$

We use Eq. 18 in Eq. 5 and Eq. 6, where for SAR models $\mathbf{B} = \rho\mathbf{W}$ or $\mathbf{B} = \rho\mathbf{W}_+$, and for CAR models $\mathbf{C} = \rho\mathbf{W}; \mathbf{M} = \mathbf{I}$ or $\mathbf{C} = \rho\mathbf{W}_+; \mathbf{M} = \mathbf{M}_+$. Note that, when considering the spatial regression model in Eq. 4, $\text{var}(\mathbf{y}) = \mathbf{\Sigma} + \sigma^2_\varepsilon \mathbf{I}$ would also be possible; for example, for a first-order CAR model, $\text{var}(\mathbf{y}) = \sigma^2_Z(\mathbf{I} - \rho\mathbf{W})^{-1} + \sigma^2_\varepsilon\mathbf{I}$. However, when $\rho = 0$, then σ^2_Z and σ^2_ε are not identifiable. In fact, as ρ goes from 1 to 0, it allows for diagonal elements to dominate in $(\mathbf{I} - \rho\mathbf{W})^{-1}$, and there seems little reason to add $\sigma^2_\varepsilon\mathbf{I}$. We evaluated some models with the additional component $\sigma^2_\varepsilon\mathbf{I}$, but σ^2_ε was always estimated to be near 0, so few of those models are presented. The exception is the IAR model, where conceptually ρ is fixed at one.

Our construction is unusual due to the $\exp(-\mathbf{S}/\theta_1)$ component. We interpret θ_1 as an among-stock connectivity parameter. Connectivity is of great interest to ecologists, and by its very definition it is about relationships *between* two nodes. Therefore, it is naturally modeled through the covariance matrix, which is also concerned with this *second-order* model property. Recall that, within stock, all entries in $\mathbf{S}$ will be zero, and hence those same entries in $\exp(-\mathbf{S}/\theta_1)$ will be one. Now, if *among* stocks there is little correlation, then θ_1 should be very small, causing those entries in $\exp(-\mathbf{S}/\theta_1)$ to be near zero. On the other hand, if θ_1 is very large, then there will be high correlation among stocks, and thus the stocks are highly connected with respect to the behavior of the response variable, justifying our interpretation of the parameter. When used in conjunction with the neighborhood matrix, the $\exp(-\mathbf{S}/\theta_1)$ component helps determine if there is additional correlation due to stock structure (low values of θ_1, meaning low connectivity) or whether the neighborhood definitions are enough (θ_1 very large, meaning high connectivity). Similarly, the $\exp(-\mathbf{D}/\theta_2)$ component models the edge weights of neighboring areal units in the autoregressive graph as an exponentially-decreasing function of distance between centroids. While this component is similar in form to the exponential covariance function (Eq. 1) in geostatistical models, the geostatistical model makes the covariance decay exponentially with distance, while in this autoregressive model, the edge weights in $\mathbf{W}$, which help define the precision matrix, decay exponentially with distance. Similar models for edge weights have been employed in other studies to allow for flexible autoregressive models (e.g., Cressie and Chan, 1989; Hanks et al., 2016).

We fit model Eq. 4 with a variety of fixed effects and covariance structures, and a list of those models is given in Table 2. We fit models using maximum likelihood (except for the IAR model, which does not have a likelihood, as discussed earlier), and details are given in Appendix B. The resulting maximized values of 2*log-likelihood are given in Fig. 5. Of course, some models are generalizations of other models, with more parameters, and will necessarily have a better fit. Methods such as Akaike Information Criteria (AIC, Akaike, 1973), Bayesian Information Criteria (BIC, Schwarz, 1978), or others (see, e.g., Burnham and Anderson, 2002; Hooten and Hobbs, 2015), can be used to select among these models. This is an example of objective 1 listed in Table 1. For AIC, each additional parameter adds a "penalty" of 2 that is subtracted from the maximized 2*log-likelihood. Fig. 5 shows the number of model parameters along the x-axis, and dashed lines at increments of two help evaluate models. For example, XC4RD has 8 parameters, so, using AIC for model selection, it should be at least 2 better than a model with 7 parameters. If one prefers a likelihood-ratio approach, then a model with one more parameter should be better by a χ-squared value on 1 degree of freedom, or 3.841. We note that there appears to be high variability among model fits, depending on the neighborhood structure (Fig. 5). Several authors have decried the general lack of exploration of the effects of neighborhood definition and choice in weights (Best et al., 2001; Earnest et al., 2007), and our results support their contention that this deserves more attention. In particular, it is interesting that row-standardized CAR models give substantially better fits than unstandardized, and CAR is much better than SAR. Note, however, that these comparisons may not hold for other data sets. Also, for row-standardized CAR models, fit worsens going from first-order to second-order neighborhoods, but then improves when going to fourth-order. Using distance between centroids had little effect until fourth-order neighborhoods were used. By an AIC criteria, model XC4RD, with 8 parameters, would be the best model because it achieved an equal model fit as XC4RDS and XC4RDU with 9 parameters, but was also more than 2 better than any of the models with 7 parameters. For model XC4RDS, the parameter θ_1 was very large, making $\exp(-\mathbf{S}/\theta_1)$ nearly constant at 1, so this model component could be dropped without changing the likelihood. Also, the addition of the uncorrelated random errors (model XC4RDU) had an estimated variance σ_ε^2 near zero, and left the likelihood essentially unchanged.

As an example of objective 2 from Table 1, the estimation of fixed effects parameters, for 3 different models, are given in Table 3. The model is overparameterized, so the parameter μ is essentially the estimate for stock 1. For example, for the XU model, exp(-0.079) = 0.92, giving an estimated trend of about 8% average decrease per year for sites from stock 1. It is significantly different from 0, which is equivalent to no trend, at $\alpha = 0.05$. This inference is obtained by taking the estimate and dividing by the standard error, and then assuming that ratio is a standard normal distribution under the null hypothesis that $\mu = 0$. The other estimates are *deviations* from μ, so stock 2 is estimated to have exp(-0.079 + 0.048) = 0.97, or a decrease of about 3% per year. A P-value for stock 2 is obtained by assuming that the estimate divided by the standard error has a standard normal distribution under the model of no difference in means, which is 0.111, and is interpreted as the probability of obtaining the stock 2 value, or larger, if it had the same mean as stock 1. It appears that stocks 3–5 have increasing trends, and that they are significantly different from stock 1 at $\alpha = 0.05$ when tested individually. In comparison, model XC4R, using maximum likelihood estimates (MLE), and Bayesian estimates (MCMC), are given in the middle two sets of columns of Table 3. The MLE estimates and standard errors for the best-fitting model, according to AIC (model XC4RD), are shown in the last set of columns in Table 3, which are very similar to the XC4R model. Further contrasts between trends in stocks are possible by using the variance-covariance matrix for the estimated fixed effects for MLE estimates, or finding the posterior distribution of the contrasts using MCMC sampling in a Bayesian approach.

Several aspects of Table 3 deserve comment. First, consistent with much literarure, notice that the standard errors for the spatial error models are larger than for the independence model XU, leading to greater uncertainty about the fixed effects estimates (Cliff and Ord, 1981; Anselin and Griffith, 1988; Legendre, 1993; Lennon, 2000; Ver Hoef et al., 2001; Fortin and Payette, 2002). Also, the Bayesian posterior standard deviations are somewhat larger than those of maximum likelihood. This is often observed in spatial models when using Bayesian methods, where the uncertainty in estimating the covariance parameters is expressed in the standard errors of the fixed effects, whereas for MLE the covariance parameters are fixed at their most likely values (Handcock and Stein, 1993).

More recently, researchers have been examining the effect of autocorrelation on the shifting values of the correlation coefficients themselves. For example, in Table 3, when going from classical multiple regression (Model XU), assuming independent residuals, to any of the spatial models, the regression coefficients change. When errors are spatially autocorrelated, the classical regression model is unbiased for estimating the coefficients (but not the standard errors of the coefficients) (Cressie, 1993; Schabenberger and Gotway, 2005; Hawkins et al., 2007; Dormann, 2007), so interest centers on whether spatial models are more efficient (that is, unbiased like classical regression, but generally closer to the true value). It has been argued that spatial models generally move coefficients closer to their true values (e.g., Ver Hoef and Cressie, 1993; Kühn, 2007; Dormann, 2007), while more extensive analysis showed ambiguous results that depended on the shift metric and the model (Bini et al., 2009).

Moreover, some have argued that classical regression coefficients may be preferred if the covariates have strong spatial correlation of their own (a topic called spatial confounding, Clayton et al., 1993; Reich et al., 2006). To explain, imagine that there are two highly autocorrelated covariates, and they are collinear (cross-correlated) as well, but we only observe one of them. The effect of the unobserved covariate will end up in the error term, causing autocorrelation there that is strongly correlated to the observed covariate. This can cause unreliable estimation of the regression coefficient for the observed covariate. The extent of this effect and how to correct for it (if at all) is the subject of current interest and debate (e.g., Hodges and Reich, 2010; Paciorek, 2010; Hughes and Haran, 2013; Hanks et al., 2015). These issues, from proper confidence interval coverage, to shifting regression coefficients, to spatial confounding, occur for all spatial (and temporal) regression models, including CAR, SAR, and geostatistical models. These are actively evolving research areas, and we add little except to make ecologists aware of them.

For objective 3 from Table 1, consider the curves in Fig. 6. We fit all combinations of CAR and SAR models, with and without row-standardization, for the first-, second-, and fourth-order neighbors (12 possible models). All such models had 7 parameters, and a few of the models are listed in Table 2. The likelihood profiles for ρ of the three best-fitting models are shown in Fig. 6. The peak value for XC4R shows that this is the best model, and the MLE for ρ for this model is 0.604. This curve also provides a likelihood-based confidence interval, known as a profile likelihood confidence interval (Box and Cox, 1964), which essentially inverts a likelihood-ratio test. A $100(1-\alpha)\%$ confidence interval for a given parameter is the set of all values such that a two-sided test of the null hypothesis that the parameter is the maximum likelihood value would not be rejected at the α level of significance (i.e., the MLE value minus a χ-squared value with one degree of freedom, which is 3.841 if $\alpha = 0.05$). These are all values above the dashed line in Fig. 6 for model XC4R, or, in other words, the endpoints of the confidence interval are provided by the intersection of the dashed line with the curve, which has a lower bound of 0.113 and an upper bound of 0.868. We also show the posterior distribution of ρ for the same model, XC4R, using a Bayesian analysis. The posterior mean was 0.687, with a 95% credible interval ranging from 0.315 to 0.933. The Bayesian estimate used improper uniform priors, so the joint posterior distribution of all parameters will be proportional to the likelihood. The difference between the MLE and the Bayesian estimates for the XC4R model is due to the fact that the MLE is the peak of the likelihood jointly (with all other parameters at their peak), whereas the Bayesian posterior is a marginal distribution (all other parameters have essentially been integrated over by the MCMC algorithm). Nonetheless, the MLE and Bayesian inferences are quite similar.

Fig. 7 shows likelihood profiles for the other parameters in the covariance matrix. For the best model, XC4RD, the solid line in Fig. 7a shows a peak for $\log(\theta_2)$ at 3.717, forming the maximum likelihood estimate and relating to objective 4 from Table 1. Once again, we show a dashed line at the maximized 2*log-likelihood (413.447) minus a χ-squared value at $\alpha = 0.05$ on one degree of freedom (3.841) to help visualize a confidence interval for θ_2 (the profile likelihood confidence interval given by all values of the solid line that are above the dashed line). The log-likelihood drops rapidly from the MLE ($\hat{\theta}_2 = 3.717$) on the left, intersecting the dashed line and forming a lower bound at 2.894, whereas the upper limit is unbounded. We return to the notion of stock connectivity in Fig. 7b. The profile likelihood for θ_1 for Model XC4RDS is given by the solid line. The likelihood is very flat for larger values of θ_1, and in fact it is continuously increasing at an imperceptible rate. Thus, the MLE is the largest value in the parameter range, which we clipped at $\log(\theta_1) = 10$. A lower bound is at $\log(\theta_1) = 0.525$, whereas the upper limit is unbounded again.

Continuing with further inferences from the model, we consider prediction (objective 5) from Table 1. Algorithms for both prediction and smoothing are given in Appendix C. Kriging is a spatial prediction method associated with geostatistics (Cressie, 1990). However, for any covariance matrix, the prediction equations can be applied regardless of how that covariance matrix was developed. We used universal kriging, that is, we included stock effects as covariates, (Huijbregts and Matheron, 1971; Cressie, 1993, pg. 151) to predict all unsampled polygons (black polygons in Fig. 3) using the XC4RD model. Note that kriging, as originally formulated, is an exact interpolator (Cressie, 1993, pg. 129) that "honors the data" (Schabenberger and Gotway, 2005, p. 252) by having predictions take on observed values at observed sites. In Fig. 8a we show the raw observations along with the predictions, making a complete map for all sites. Of course, what distinguishes predictions using statistical models, as opposed to deterministic algorithms (e.g., inverse distance weighting, Shepard, 1968) is that statistical predictions provide uncertainty estimates for each prediction (Fig. 8B). When kriging is used as an exact interpolator, the values are known at observed sites, so the prediction variances are zero at observed sites. Hence, we only show the prediction standard errors for polygons with missing data.

We also use the more traditional smoother for CAR and SAR models, such as those used in (Clayton and Kaldor, 1987), forming objective 6 from Table 1. For model XC4RD, without any independent component, this is essentially equivalent to leave-one-out-cross-validation. That is, the conditional expectation, which is obtained directly from Eq. 11 (after adjusting for estimated covariate effects) is used rather than the observed value at each location. When the covariance matrix is known, for normally distributed data, ordinary kriging is also the conditional expectation (Cressie, 1993, p. 108, 174). Hence, the predicted and smoothed values, using the conditional expectation, are given in Fig. 8c; note then, that the predictions are equivalent to Fig. 8a at the unsampled locations. Two extremes in smoothing approaches are 1) kriging as an exact predictor, that is, it leaves the data unchanged (Figs. 8a), and 2) removing observed data to replace them with conditional expectations based on neighbors (Fig. 8c). In fact, both are quite unusual for a smoothing objective. Generally, a model is adopted with a spatial component, and a noisy measurement error or independent component. Smoothing then involves finding a compromise between the spatial component and the raw, observed data. As an example for these data, consider the XI4RU model, which has an IAR component plus an uncorrelated error component. The IAR model has very high autocorrelation ($\rho = 1$), but here we allowed it to be a mixture with uncorrelated error, and the relative values of σ_Z^2 and σ_ε^2 will determine how much autocorrelation is estimated for the data. Under this model, predictions for observed data can fall between the very smooth IAR predictions and the very rough observed data. When such a model is formulated hierarchically (Eq. 17), often in a Bayesian context, predictions exhibit a property called shrinkage (Fay and Herriot, 1979), where predictions of observed values are some compromise between an ultra-smooth fit from a pure IAR model, and the roughness of the raw data (Fig. 8d). The amount of shrinkage depends on the relative values of σ_Z^2 and σ_ε^2. In fact, this is usually the case when CAR and SAR models are used in a generalized linear model setting because the conditional independence assumption (e.g., of a Poisson distribution) is analogous to the $\sigma_\varepsilon^2\mathbf{I}$ component. Note that a Bayesian perspective is not a requirement, a similar objective is obtained using filtered kriging (Waller and Gotway, 2004, pg. 306) when there are both spatial and uncorrelated variance components.

Finally, to complete the example, we return to the idea of nonstationarity in variances and covariances. Stationarity is the notion that statistical properties remain constant (stationary) as we move through space. Stationarity means that correlation between neighbors on the edge of the study area are the same as those in the middle, and that variances are constant throughout. Notice that, as claimed earlier, row-standardization causes variance to decrease with the numbers of neighbors (which are generally greater in the interior of a study area in contrast to the perimeter) for model XC4R (Fig. 9a), but it is not a simple function of neighbors alone, as it depends in complicated ways on the whole graphical (or network) structure. In contrast, variance generally increases with the number of neighbors without row-standardization (model XC4) of the neighborhood matrix (Fig. 9a). Correlation also decreases with neighbor order, although not as dramatically as one might expect (Fig. 9b), and not at all (on average) between first-order to second-order when the neighborhood matrix is not row-standardized. Box plots summarize all possible correlations as a function of distance between centroids (binned into classes, Fig. 9c,d), which show that while correlation generally decreases with distance between centroids, there is a great deal of variation. Also recall that the MLE for ρ, which is a parameter in the precision matrix, for model XC4R was 0.604 (Fig. 6), but for the covariance matrix, correlations are much lower (Fig. 9b-d). Because weights are developed for partial correlations, or for the inverse of the covariance matrix, when we examine the covariance matrix itself, the diagonal elements are non-constant, in contrast to typical geostatistical models. It is important to realize that there is no direct calculation between the estimated ρ value in the CAR or SAR model and the correlations in the covariance matrix; only that higher ρ generally means higher correlations throughout the covariance matrix. One can always invert the fitted CAR or SAR model to obtain the full covariance matrix, and this can then be inspected and summarized if needed (e.g., Fig. 9), thus improving model diagnostics and our understanding of the fitted model.

DISCUSSION AND CONCLUSIONS

Autoregressive models are an important class of spatial models that have rarely been explained in practical terms. We provide the following summary of CAR and SAR models.

1. Intuition on CAR and SAR models can be obtained by considering the relationship between autoregressive weights and partial correlations.

2. Row standardization is generally a good idea after choosing initial neighborhoods and

weights. This will result in CAR models that are generally more local for a given set of neighbors because, for that same set of neighbors, the SAR model squares the weights matrix, creating neighbors of neighbors in the precision matrix.

3. The IAR model is a special case of the CAR model that uses row standardization and fixes the autocorrelation parameter at one, which leads to an improper covariance matrix; however, much like a similar AR1 model, or Brownian motion, these are still useful models.

In addition, we presented six objectives, some of which are common, and others less so, in which spatial autoregressive models could be used. We fit a variety of CAR/SAR models using MLE and MCMC methods to an example data set to illustrate all six objectives outlined in the Introduction. In what follows, we provide further discussion on 5 take-home messages: 1) thoughts on choosing between CAR and IAR models, 2) modeling ecological effects in the covariance matrix, 3) the appeal of spatial smoothing, 4) how to handle isolated neighbors, and 5) software considerations.

The choice of IAR versus CAR is confusing, and while both are often described in the literature together, there is little guidance on choosing between them. One advantage of the IAR is that it has one less parameter to estimate. It was proposed by Besag and Kooperberg (1995) in part based on the following: they noticed that for a certain CAR model, ρ in Eq. 6 needed to be 0.999972 to have a marginal correlation near 0.75 (indeed, compare the estimate of $\hat{\rho} = 0.604$ in our example yielding the correlations seen in Fig. 9). In many practical applications, ρ was often estimated to be very near 1, so Besag and Kooperberg (1995) suggested the IAR model as a flexible spatial surface that has one less parameter to estimate. On the other hand, critics noticed that it may force spatial smoothness where none exists (e.g., Leroux et al., 2000). Our point of view is best explored through the hierarchical model Eq. 17. Consider an example of count data, where the data model, $\mathbf{y} \sim [\mathbf{y}|g(\boldsymbol{\mu})]$, conditional on the mean $g(\boldsymbol{\mu})$, is composed of independent Poisson distributions. Hence, there are no extra variance parameters $\boldsymbol{\nu}$, but rather the independent, nonstationary variance component is already determined because it is equal to the mean. In this case, we recommend the CAR model to allow flexibility in modeling the diagonal of the covariance matrix (the CAR model can allow for smaller ρ values, which essentially allows for further uncorrelated error). On the other hand, if $[\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\nu}]$ has a free variance parameter in $\boldsymbol{\nu}$ (e.g., the product of independent normal or negative binomial distributions), then we recommend the IAR model to decrease confounding between the diagonal of $\boldsymbol{\Sigma}$, essentially controlled by ρ, and the free variance parameter in $\boldsymbol{\nu}$.

The results for Figs. 6 and 7 have confidence intervals that are quite wide. In general, uncertainty is much higher when trying to estimate covariance parameters than regression (fixed effect) parameters. Nevertheless, the covariance models that we constructed demonstrate that it is possible to examine the effect of covariates in the covariance structure (see also Hanks and Hooten, 2013). In other words, it is possible to make inference on connectivity parameters in the covariance matrix, but, they may be difficult to estimate with much precision if the data are measured only on the nodes. In our harbor seal example, when stock effects were put into the mean structure, there was abundant evidence of different effects, but when that effect was put into the covariance matrix, the precision was quite low. It is important to put connectivity effects into the covariance matrix (in many cases, that will be the only place that makes sense), but realize that they may be difficult to estimate well without large data sets.

From an ecological viewpoint, why do spatial smoothing? Geostatistics had a tradition where modelers were often adamant that no smoothing occur ("honoring the data," Schabenberger and Gotway, 2005, p. 224). That tradition is often unknowingly continued with uncritical use of kriging formulas for prediction. For example, if we assume that the observed values, without error, are part of the process of interest, then notice from Fig. 3 that the largest value is 0.835 from the legend on the right. Recalling that these are trends, on the log scale, the observed value from the data was $\exp(0.835) = 2.3$, or more than doubling each year. That is clearly not a sustainable growth rate and is likely due to small sample sizes and random variation. That same value from Fig. 8c is $\exp(0.039) = 1.04$, or about 4% growth per year, which is a much more reasonable estimate of growth. The largest smoothed value in Fig. 8c, back on the exponential scale, was 1.083, or about 8% growth per year, and the largest value in Fig. 8d, back on the exponential scale, was 1.146, or about 15% growth per year. These values are similar to published estimates of harbor seal growth rates in natural populations (e.g., Hastings et al., 2012). Fig. 8c,d also clarifies the regional trends, which are difficult to see among the noise in Fig. 3 or by simply filling in the missing sites with predictions (Fig. 8a). For these reasons, smoothing is very popular in disease-mapping applications, and it should be equally attractive for a wide variety of ecological applications. In particular, the XI4RU model (Fig. 8d) is appealing because it uses the data to determine the amount of smoothing. However, we also note that when used in hierarchical models where, for the data model, the variance is fixed in relation to the mean (e.g., binomial, Bernoulli, and Poisson), the amount of smoothing will be dictated by the assumed variance of the data model. In such cases, we reiterate the discussion on choosing between CAR and IAR.

A rarely discussed consideration is the case of isolated nodes (sites with no neighbors) when constructing the neighborhood matrix. Having a row of zeros in $\mathbf{B}$ in Eq. 5, or in $\mathbf{C}$ in Eq. 6, will cause problems. It is even easier to see that we cannot divide by zero in Eq. 13, or during row-standardization. Instead, we suggest that the covariance matrix be constructed as

$$\begin{pmatrix} \sigma_I^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{pmatrix},$$

where we show the data ordered such that all isolated sites are first, and their corresponding covariance matrix is $\sigma_I^2\mathbf{I}$. The matrix $\mathbf{\Sigma}$ is the CAR or SAR covariance matrix for the sites connected by neighbors. Note that one of the main issues here is the separation of the variance parameters, σ_I^2 and σ_Z^2 in Eq. 5 or Eq. 6. As seen in Eq. 13, the autoregressive variance is often scaled by the number of neighbors, and because the isolated sites have no neighbors, it is prudent to give them their own variance parameter.

Our final take-home message concerns questions that a user should ask when fitting autoregressive models with existing software packages. Does the software check the weights to ensure the covariance matrix will be proper? It may be computationally expensive to check it internally, which lessens the appeal of the autoregressive models, and the software may trust the user to give it a valid weights matrix. Does the software use row-standardization internally? How does the software handle isolated sites? These are special issues that only pertain to CAR and SAR models, so we suggest investigation of these issues so that the software output can be better understood.

In closing, we note that "networks," and network models, are seeing increasing use throughout science, including ecology (Borrett et al., 2014). Looking again at Fig. 4, if we remove the polygon boundaries, these are network models. Spatial information, in the way of neighborhoods, was used to create the networks. Thus, more general concepts for CAR and SAR models are the graphical models (Lauritzen, 1996; Whittaker, 2009). A better understanding of these models will lead to their application as network models when data are collected on the nodes of the network, and they can be extended beyond spatial data. This provides a rich area for further model development and research that can include, modify, and enhance the autoregressive models.

Acknowledgments

This research began from a working group on network models at the Statistics and Applied Mathematical Sciences (SAMSI) 2014-15 Program on Mathematical and Statistical Ecology. The project received financial support from the National Marine Fisheries Service, NOAA. Aerial surveys were authorized under a Marine Mammal Protection Act General Authorization (LOC No. 14590) issued to the Marine Mammal Laboratory. The findings and conclusions in the paper of the NOAA author(s) do not necessarily represent the views of the reviewers nor the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm names does not imply an endorsement by the U.S. Government.

Data and Code Accessibility

An R package called spAReco was created that contains all data and code. This document was created using knitr, and the manuscript combining latex and R code is also included in the package. The package can be downloaded at https://github.com/jayverhoef/spAReco.git, with instructions for installing the package.

References

Aarts, G., Fieberg, J., Brasseur, S., and Matthiopoulos, J. (2013), "Quantifying the effect of habitat availability on species distributions," *Journal of Animal Ecology*, 82, 1135–1145.

Agarwal, D. K., Silander, J. A., Gelfand, A. E., Dewar, R. E., and Mickelson, J. G. (2005), "Tropical deforestation in Madagascar: analysis using hierarchical, spatially explicit, Bayesian regression models," *Ecological Modelling*, 185, 105–131.

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, eds. Petrov, B. and Csaki, F., Budapest: Akademiai Kiado, pp. 267–281.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht, the Netherlands: Kluwer Academic Publishers.

Anselin, L. and Griffith, D. A. (1988), "Do spatial effecfs really matter in regression analysis?" *Papers in Regional Science*, 65, 11–34.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL, USA: Chapman and Hall/CRC Press.

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., and Elston, D. A. (2010), "Regression analysis of spatial data," *Ecology Letters*, 13, 246–264.

Beguería, S. and Pueyo, Y. (2009), "A comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation," *Global Ecology and Biogeography*, 18, 273–279.

Beguin, J., Martino, S., Rue, H., and Cumming, S. G. (2012), "Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation," *Methods in Ecology and Evolution*, 3, 921–929.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

— (1975), "Statistical analysis of non-lattice data," *The Statistician*, 179–195.

Besag, J. and Kooperberg, C. (1995), "On conditional and intrinsic autoregressions," *Biometrika*, 82, 733–746.

Best, N., Cockings, S., Bennett, J., Wakefield, J., and Elliott, P. (2001), "Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographical scale and ecological bias," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 164, 155–174.

Bigg, M. A. (1969), *The harbour seal in British Columbia*, no. 172, Fisheries Research Board of Canada.

Bini, L., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Akre, T. S., Albaladejo, R. G., Albuquerque, F. S., Aparicio, A., Araújo, M. B., Baselga, A., Beck, J., et al. (2009), "Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression," *Ecography*, 32, 193–204.

Bivand, R. and Piras, G. (2015), "Comparing Implementations of Estimation Methods for Spatial Econometrics," *Journal of Statistical Software*, 63, 1–36.

Borrett, S. R., Moody, J., and Edelmann, A. (2014), "The Rise of Network Ecology: Maps of the topic diversity and scientific collaboration," *Ecological Modelling*, 293, 111–127.

Boveng, P. L., Bengtson, J. L., Withrow, D. E., Cesarone, J. C., Simpkins, M. A., Frost, K. J., and Burns, J. J. (2003), "The abundance of harbor seals in the Gulf of Alaska," *Marine Mammal Science*, 19, 111–127.

Box, G. E. P. and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 26, 211–252.

Broms, K. M., Johnson, D. S., Altwegg, R., and Conquest, L. L. (2014), "Spatial occupancy models applied to atlas data show Southern Ground Hornbills strongly depend on protected areas," *Ecological Applications*, 24, 363–374.

Brook, D. (1964), "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems," *Biometrika*, 51, 481–483.

Brown, R. F., Wright, B. E., Riemer, S. D., and Laake, J. (2005), "Trends in abundance and current status of harbor seals in Oregon: 1977–2003," *Marine Mammal Science*, 21, 657–670.

Bullock, B. P. and Burkhart, H. E. (2005), "An evaluation of spatial dependency in juvenile loblolly pine stands using stem diameter," *Forest Science*, 51, 102–108.

Burnham, K. P. and Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, New York: Springer-Verlag Inc.

Campbell, J. B. and Wynne, R. H. (2011), *Introduction to Remote Sensing*, Guilford Press.

Cassemiro, F. A., Diniz-Filho, J. A. F., Rangel, T. F. L., and Bini, L. M. (2007), "Spatial autocorrelation, model selection and hypothesis testing in geographical ecology: Implications for testing metabolic theory in New World amphibians," *Neotropical Biology and Conservation*, 2, 119–126.

Chiles, J.-P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: John Wiley & Sons.

Clayton, D. and Kaldor, J. (1987), "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping," *Biometrics*, 43, 671–681.

Clayton, D. G., Bernardinelli, L., and Montomoli, C. (1993), "Spatial correlation in ecological analysis," *International Journal of Epidemiology*, 22, 1193–1202.

Cliff, A. D. and Ord, J. K. (1981), *Spatial Processes: Models and Applications*, vol. 44, London, U.K.: Pion.

Clifford, P. (1990), "Markov random fields in statistics," in *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, eds. Grimmett, R. G. and Welsh, D. J. A., New York, NY, USA: Oxford University Press, pp. 19–32.

Codling, E. A., Plank, M. J., and Benhamou, S. (2008), "Random walk models in biology," *Journal of the Royal Society Interface*, 5, 813–834.

Cressie, N. (1990), "The origins of kriging," *Mathematical Geology*, 22, 239–252.

Cressie, N., Calder, K. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009), "Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling," *Ecological Applications*, 19, 553–570.

Cressie, N. and Chan, N. H. (1989), "Spatial modeling of regional variables," *Journal of the American Statistical Association*, 84, 393–401.

Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-temporal Data*, Hoboken, New Jersey: John Wiley & Sons.

Cressie, N. A. C. (1993), *Statistics for Spatial Data, Revised Edition*, New York: John Wiley & Sons.

Dark, S. J. (2004), "The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis," *Diversity and Distributions*, 10, 1–9.

de Valpine, P. and Hastings, A. (2002), "Fitting population models incorporating process noise and observation error," *Ecological Monographs*, 72, 57–76.

Dormann, C. F. (2007), "Effects of incorporating spatial autocorrelation into the analysis of species distribution data," *Global ecology and biogeography*, 16, 129–138.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., and Wilson, R. (2007), "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review," *Ecography*, 30, 609–628.

Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., and Beard, J. (2007), "Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models," *International Journal of Health Geographics*, 6, 54, doi = 10.1016/j.csda.2011.10.022.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least angle regression," *The Annals of Statistics*, 32, 407–499.

Elliot, P., Wakefield, J. C., Best, N. G., Briggs, D., et al. (2000), *Spatial epidemiology: methods and applications.*, Oxford, UK: Oxford University Press.

Evans, T. S., Kirchgessner, M. S., Eyler, B., Ryan, C. W., and Walter, W. D. (2016), "Habitat influences distribution of chronic wasting disease in white-tailed deer," *The Journal of Wildlife Management*, 80, 284–291.

Fay, R. E. and Herriot, R. A. (1979), "Estimates of income for small places: an application of

James-Stein procedures to census data," *Journal of the American Statistical Association*, 74, 269–277.

Fortin, M.-J. and Payette, S. (2002), "How to test the significance of the relation between spatially autocorrelated data at the landscape scale: a case study using fire and forest maps," *Ecoscience*, 9, 213–218.

Frost, K. J., Lowry, L. F., and Ver Hoef, J. M. (1999), "Monitoring the trend of harbor seals in Prince William Sound, Alaska, after the Exxon Valdez oil spill," *Marine Mammal Science*, 15, 494–506.

Gardner, C. L., Lawler, J. P., Ver Hoef, J. M., Magoun, A. J., and Kellie, K. A. (2010), "Coarse-Scale Distribution Surveys and Occurrence Probability Modeling for Wolverine in Interior Alaska," *The Journal of Wildlife Management*, 74, 1894–1903.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398–409.

Haas, S. E., Hooten, M. B., Rizzo, D. M., and Meentemeyer, R. K. (2011), "Forest species diversity reduces disease risk in a generalist plant pathogen invasion," *Ecology Letters*, 14, 1108–1116.

Haining, R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, UK: Cambridge University Press.

Hamilton, J. D. (1994), *Time Series Analysis*, vol. 2, Princeton, NJ, USA: Princeton University Press.

Hammersley, J. M. and Clifford, P. (1971), "Markov fields on finite graphs and lattices," Unpublished Manuscript.

Handcock, M. S. and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Hanks, E. M. (2017), "Modeling spatial covariance using the limiting distribution of spatio-temporal random walks," *Journal of the American Statistical Association*, 0, In Press, bibtex: Hanks2017.

Hanks, E. M. and Hooten, M. B. (2013), "Circuit theory and model-based inference for landscape connectivity," *Journal of the American Statistical Association*, 108, 22–33.

Hanks, E. M., Hooten, M. B., Knick, S. T., Oyler-McCance, S. J., Fike, J. A., Cross, T. B., Schwartz, M. K., et al. (2016), "Latent spatial models and sampling design for landscape genetics," *The Annals of Applied Statistics*, 10, 1041–1062.

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

Harvey, J. T., Brown, R. F., and Mate, B. R. (1990), "Abundance and distribution of harbor seals (Phoca vitulina) in Oregon, 1975-1983," *Northwestern Naturalist*, 65–71.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York, NY: Springer.

Hastings, K. K., Small, R. J., and Pendleton, G. W. (2012), "Sex-and age-specific survival of harbor seals (*Phoca vitulina*) from Tugidak Island, Alaska," *Journal of Mammalogy*, 93, 1368–1379.

Hawkins, B. A., Diniz-Filho, J. A. F., Bini, L. M., De Marco, P., and Blackburn, T. M. (2007), "Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology," *Ecography*, 30, 375–384.

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

Hooten, M. and Hobbs, N. (2015), "A guide to Bayesian model selection for ecologists," *Ecological Monographs*, 85, 3–28.

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

Huang, X., Grace, P., Hu, W., Rowlings, D., and Mengersen, K. (2013), "Spatial prediction of N_2O emissions in pasture: a Bayesian model averaging analysis," *PLoS ONE*, 8, e65039.

Huber, H. R., Jeffries, S. J., Brown, R. F., Delong, R. L., and Vanblaricom, G. (2001), "Correcting aerial survey counts of harbor seals (Phoca vitulina richardsi) in Washington and Oregon," *Marine Mammal Science*, 17, 276–293.

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

Huijbregts, C. and Matheron, G. (1971), "Universal kriging (an optimal method for estimating and contouring in trend surface analysis)," in *Proceedings of Ninth International Symposium on Techniques for Decision-making in the Mineral Industry*, ed. McGerrigle, J. I., The Canadian Institute of Mining and Metallurgy, Special Volume 12, pp. 159–169.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P., and Travis, J. M. (2013), "Fitting complex ecological point process models with integrated nested Laplace approximation," *Methods in Ecology and Evolution*, 4, 305–315.

Jeffries, S., Huber, H., Calambokidis, J., and Laake, J. (2003), "Trends and status of harbor seals in Washington State: 1978-1999," *The Journal of wildlife management*, 207–218.

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., and Pond, B. A. (2013a), "Spatial occupancy models for large data sets," *Ecology*, 94, 801–808.

Johnson, D. S., Hooten, M. B., and Kuhn, C. E. (2013b), "Estimating animal resource selection from telemetry data using point process models," *Journal of Animal Ecology*, 82, 1155–1164.

Keitt, T. H., Bjørnstad, O. N., Dixon, P. M., and Citron-Pousty, S. (2002), "Accounting for spatial pattern when modeling organism-environment interactions," *Ecography*, 25, 616–625.

Kissling, W. D. and Carl, G. (2008), "Spatial autocorrelation and the selection of simultaneous autoregressive models," *Global Ecology and Biogeography*, 17, 59–71.

Kühn, I. (2007), "Incorporating spatial autocorrelation may invert observed patterns," *Diversity and Distributions*, 13, 66–69.

Lauritzen, S. L. (1996), *Graphical Models*, Oxford, UK: Oxford University Press.

Lawson, A. B. (2013a), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Boca Raton, FL: CRC Press.

— (2013b), *Statistical Methods in Spatial Epidemiology*, Chichester, UK: John Wiley & Sons.

Legendre, P. (1993), "Spatial autocorrelation: trouble or new paradigm?" *Ecology*, 74, 1659–1673.

Lennon, J. J. (2000), "Red-shifts and red herrings in geographical ecology," *Ecography*, 23, 101–113.

Leroux, B. G., Lei, X., and Breslow, N. (2000), "Estimation of disease rates in small areas: A new mixed model for spatial dependence," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, eds. Halloran, M. E. and Berry, D. A., Springer, pp. 179–191.

Lichstein, J. W., Simons, T. R., Shriner, S. A., and Franzreb, K. E. (2002), "Spatial autocorrelation and autoregressive models in ecology," *Ecological Monographs*, 72, 445–463.

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

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, 10, 325–337.

Magoun, A. J., Ray, J. C., Johnson, D. S., Valkenburg, P., Dawson, F. N., and Bowman, J. (2007), "Modeling wolverine occurrence using aerial surveys of tracks in snow," *Journal of Wildlife Management*, 71, 2221–2229.

Mathews, E. A. and Pendleton, G. W. (2006), "Declines in harbor seal (*Phoca vitulina*) numbers in Glacier Bay National Park, Alaska, 1992–2002," *Marine Mammal Science*, 22, 167–189.

Moran, P. A. (1948), "The interpretation of statistical maps," *Journal of the Royal Statistical Society. Series B (Methodological)*, 10, 243–251.

Olesiuk, P. F. (1999), "An assessment of the status of harbour seals (*Phocu vitulinu*) in British Columbia," Canadian stock assessment secretariat research document 99/33, Fisheries and Oceans Canada, Ottawa, Ontario, Canada.

Olesiuk, P. F., Bigg, M. A., and Ellis, G. M. (1990), "Recent trends in the abundance of harbour seals, Phoca vitulina, in British Columbia," *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 992–1003.

Pace, R. K. and Barry, R. (1997a), "Fast spatial estimation," *Applied Economics Letters*, 4, 337–341.

— (1997b), "Sparse spatial autoregressions," *Statistics & Probability Letters*, 33, 291–297.

Paciorek, C. J. (2010), "The importance of scale for spatial-confounding bias and precision of spatial regression estimators," *Statistical Science*, 25, 107.

— (2013), "Spatial models for point and areal data using Markov random fields on a fine grid," *Electronic Journal of Statistics*, 7, 946–972.

Pedersen, R. Ø., Sandel, B., and Svenning, J.-C. (2014), "Macroecological evidence for competitive regional-scale interactions between the two major clades of mammal carnivores (Feliformia and Caniformia)," *PLOS ONE*, 9, e100553.

Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K. B., Rogers, D. J., and Clements, A. C. (2008), *Spatial Analysis in Epidemiology*, Oxford, UK: Oxford University Press Oxford.

Pitcher, K. W. (1990), "Major decline in number of harbor seals, *Phoca vitulina richardsi*, on Tugidak Island, Gulf of Alaska," *Marine Mammal Science*, 6, 121–134.

Pitcher, K. W. and Calkins, D. G. (1979), "Biology of the harbor seal, *Phoca vitulina richardsi*, in the Gulf of Alaska," Tech. Rep. Research Unit 229, Contract 03-5-002-69, US. Department of Commerce, National Oceanic and Atmospheric Administration, Outer Continental Shelf Environmental Assessment Program Final Report 19(1983), Anchorage, AK.

Poley, L. G., Pond, B. A., Schaefer, J. A., Brown, G. S., Ray, J. C., and Johnson, D. S. (2014), "Occupancy patterns of large mammals in the Far North of Ontario under imperfect detection and spatial autocorrelation," *Journal of Biogeography*, 41, 122–132.

Qiu, J. and Turner, M. G. (2015), "Importance of landscape heterogeneity in sustaining hydrologic ecosystem services in an agricultural watershed," *Ecosphere*, 6, 1–19.

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

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

Rue, H. and Held, L. (2005), *Gauss Markov Random Fields: Theory and Applications*, Boca Raton, FL, USA: Chapman and Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 319–392.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton, Florida: Chapman Hall/CRC.

Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, 6, 461–464.

Shepard, D. (1968), "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM National Conference*, ACM, pp. 517–524.

Small, R. J., Pendleton, G. W., and Pitcher, K. W. (2003), "Trends in abundance of Alaska harbor seals, 1983–2001," *Marine Mammal Science*, 19, 344–362.

Snedecor, G. W. and Cochran, W. G. (1980), *Statistical Methods, Seventh Edition*, Iowa State University.

Sokal, R. R. and Oden, N. L. (1978), "Spatial autocorrelation in biology: 1. Methodology," *Biological Journal of the Linnean Society*, 10, 199–228.

Song, J. J. and De Oliveira, V. (2012), "Bayesian model selection in spatial lattice models," *Statistical Methodology*, 9, 228–238.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

Thogmartin, W. E., Sauer, J. R., and Knutson, M. G. (2004), "A Hierarchical Spatial Model of Avian Abundance With Application to Cerulean Warblers," *Ecological Applications*, 14, 1766–1779.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tognelli, M. F. and Kelt, D. A. (2004), "Analysis of determinants of mammalian species richness in South America using spatial autoregressive models," *Ecography*, 27, 427–436.

Turner, S. J., O'Neill, R. V., Conley, W., Conley, M. R., and Humphries, H. C. (1991), "Pattern and scale: statistics for landscape ecology," in *Quantitative Methods in Landscape Ecology*, eds. Turner, S. J. and Gardner, R. H., Springer-Verlag, pp. 17–49.

Ver Hoef, J. M. and Boveng, P. L. (2007), "Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?" *Ecology*, 88, 2766–2772.

Ver Hoef, J. M., Cameron, M. F., Boveng, P. L., London, J. M., and Moreland, E. E. (2014), "A spatial hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea," *Statistical Methodology*, 17, 46–66.

Ver Hoef, J. M. and Cressie, N. (1993), "Spatial statistics: Analysis of field experiments," in *Design and Analysis of Ecological Experiments*, eds. Scheiner, S. M. and Gurevitch, J., Chapman & Hall Ltd, pp. 319–341.

Ver Hoef, J. M., Cressie, N., Fisher, R. N., and Case, T. J. (2001), "Uncertainty and spatial linear models for ecological data," in *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications*, eds. Hunsaker, C. T., Goodchild, M. F., Friedl, M. A., and Case, T. J., New York, NY: Springer-Verlag, pp. 265–282.

Ver Hoef, J. M. and Frost, K. J. (2003), "A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska," *Environmental and Ecological Statistics*, 10, 201–219.

Ver Hoef, J. M., Hanks, E. M., and Hooten, M. B. (2017), "On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models," *Journal of the American Statistical Association*, submitted.

Ver Hoef, J. M. and Jansen, J. K. (2007), "Space-time zero-inflated count models of harbor seals," *Environmetrics*, 18, 697–712.

Wall, M. M. (2004), "A close look at the spatial structure implied by the CAR and SAR models," *Journal of Statistical Planning and Inference*, 121, 311–324.

Waller, L. A. and Gotway, C. A. (2004), *Applied Spatial Statistics for Public Health Data*, John Wiley and Sons, New Jersey.

Watt, A. S. (1947), "Pattern and process in the plant community," *Journal of ecology*, 35, 1–22.

Whittaker, J. (2009), *Graphical Models in Applied Multivariate Statistics*, Chichester, UK: Wiley Publishing.

Whittle, P. (1954), "On stationary processes in the plane," *Biometrika*, 41, 434–449.

Wolfe, R. J., Fall, J. A., and Stanek, R. T. (2009), "The subsistence harvest of harbor seals and sea lions by Alaska Natives in 2004," Tech. Rep. No. 303, Alaska Department of Fish and Game, Division of Subsistence.

Zhu, J., Huang, H.-C., and Reyes, P. E. (2010), "On selection of spatial linear models for lattice data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 389–402.

TABLES

Table 1: Common objectives when using spatial autoregressive models. Notation for the model components comes from Eq. 17.

Objective	Description	Model Component
1. Model Comparison & Selection	CAR and SAR models are often part of a spatial (generalized) linear model. One goal, prior to further inference, might be to compare models, and then choose one. The choice of the form of a CAR or SAR model may be important in this comparison and selection.	$L(\cdot\|\mathbf{y})$
2. Regression	The goal is to estimate the spatial regression coefficients, which quantify how an explanatory variable "affects" the response variable.	$\boldsymbol{\beta}$
3. Autocorrelation	The goal is to estimate the "strength" of autocorrelation, especially if it represents an ecological idea such as spatial connectivity, which quantifies how similarly sites change in the residual errors, after accounting for regression effects.	ρ
4. Connectivity Structure	The goal is to estimate covariate effects on connectivity (neighborhood) structure. Although rarely used, covariates can be included in the precision matrix to see how they affect connectivity structure (causing more or less correlation).	$\boldsymbol{\theta}$
5. Prediction	This is the classical goal of geostatistics, and is rarely used in CAR and SAR models. However, if sites have missing data, prediction is possible.	$\mathbf{y}_u$ and/or $\boldsymbol{\mu}_u$
6. Smoothing	The goal is to create values at spatial sites that smooth over observed data by using values from nearby locations to provide better estimates.	$g(\boldsymbol{\mu})$

Table 2: A variety of candidate models used to explore spatial autoregressive models for the example data set. For fixed effects, the $\mathbf{1}$ indicates an overall mean in the model, and $\mathbf{X}_{\text{stock}}$ includes an additional categorical effect for each stock. A $[\cdot]_+$ around a matrix indicates row-standardization, and for CAR models, $[\mathbf{M}]_+$ is the appropriate diagonal matrix for such row standardization. The matrices themselves are described in the text. For model codes, m indicates an overall mean only, whereas X indicates the additional stock effect in the fixed effects. C indicates a CAR model, S a SAR model, and I an IAR model. A 1 indicates a first-order neighborhood, 2 a second-order neighborhood, and 4 a fourth-order neighborhood. R indicates row-standardization. D indicates inclusion of Euclidean distance within neighborhoods, S a cross stock connectivity matrix. U at the end indicates inclusion of an additive random effect of uncorrelated variables.

Model Code	Fixed Effects	Covariance Model	No. Parms
mU	$\mathbf{1}$	$\sigma_\varepsilon^2\mathbf{I}$	2
mC1R	$\mathbf{1}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_1]_+)^{-1}[\mathbf{M}]_+$	3
XU	$\mathbf{X}_{\text{stock}}$	$\sigma_\varepsilon^2\mathbf{I}$	6
XC1R	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_1]_+)^{-1}[\mathbf{M}]_+$	7
XC1	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho\mathbf{W}_1)^{-1}$	7
XS1R	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2[(\mathbf{I}-\rho[\mathbf{W}_1]_+)(\mathbf{I}-\rho[\mathbf{W}_1]_+)]^{-1}$	7
XS1	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2[(\mathbf{I}-\rho\mathbf{W}_1)(\mathbf{I}-\rho\mathbf{W}_1)]^{-1}$	7
XC2R	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_2]_+)^{-1}[\mathbf{M}]_+$	7
XC4R	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_4]_+)^{-1}[\mathbf{M}]_+$	7
XC4	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho\mathbf{W}_4)^{-1}$	7
XI4RU	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-[\mathbf{W}_4]_+)^{-1}[\mathbf{M}]_+ + \sigma_\varepsilon^2\mathbf{I}$ (improper)	7
XC4RD	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_4\odot\exp(-\mathbf{D}/\theta_2)]_+)^{-1}[\mathbf{M}]_+$	8
XC4RDS	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_4\odot\exp(-\mathbf{D}/\theta_2)\odot\exp(-\mathbf{S}/\theta_1)]_+)^{-1}[\mathbf{M}]_+$	9
XC4RDU	$\mathbf{X}_{\text{stock}}$	$\sigma_Z^2(\mathbf{I}-\rho[\mathbf{W}_4\odot\exp(-\mathbf{D}/\theta_2)]_+)^{-1}[\mathbf{M}]_+ + \sigma_\varepsilon^2\mathbf{I}$	9

Table 3: Estimated fixed effects for several models listed in Table 2. Both the estimate (Est.) and estimated standard error (Std.Err.) are given for each model. All models use maximum likelihood estimates (MLE), except for XC4R model, we distinguish the MLE estimate with -MLE, and a Bayesian estimate using Markov chain Monte Carlo with -MCMC.

	XU		XC4R-MLE		XC4R-MCMC		XC4RD	
Parameter	Est.	Std.Err	Est.	Std.Err	Est.	Std.Err.	Est.	Std.Err
μ	-0.079	0.0225	-0.080	0.0288	-0.082	0.0330	-0.077	0.0290
$\beta_{\text{stock 2}}$	0.048	0.0298	0.063	0.0379	0.063	0.0429	0.058	0.0386
$\beta_{\text{stock 3}}$	0.093	0.0281	0.095	0.0355	0.097	0.0386	0.092	0.0356
$\beta_{\text{stock 4}}$	0.132	0.0279	0.135	0.0346	0.138	0.0406	0.132	0.0346
$\beta_{\text{stock 5}}$	0.084	0.0259	0.093	0.0327	0.096	0.0378	0.089	0.0330

FIGURES

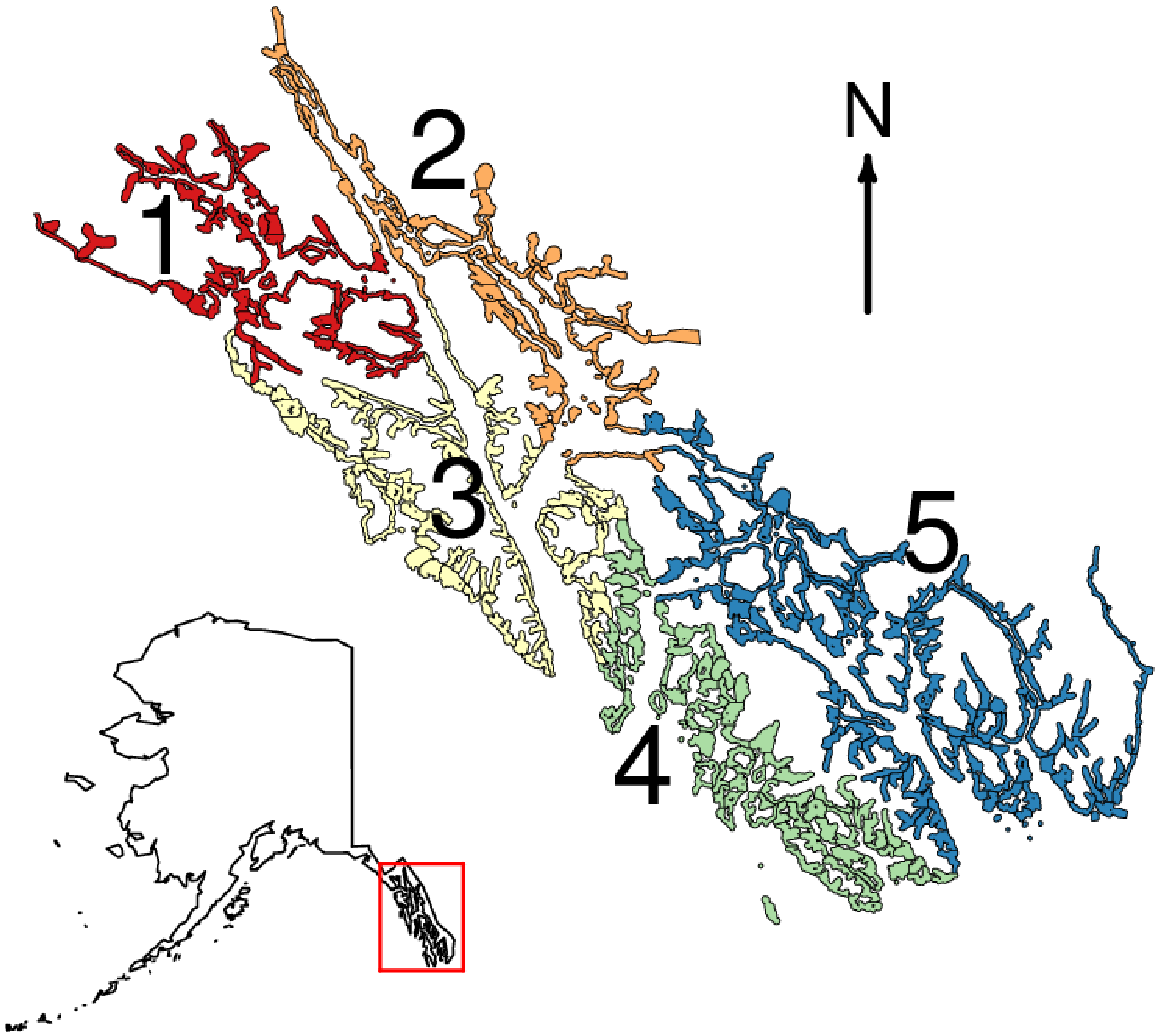

Figure 2: Study area in Southeast Alaska, outlined in red in the lower left figure. Survey polygons were established around the coast of the mainland and all islands, which were surveyed for harbor seals. The study area comprises 5 stocks, each with their own color, and are numbered for further reference.

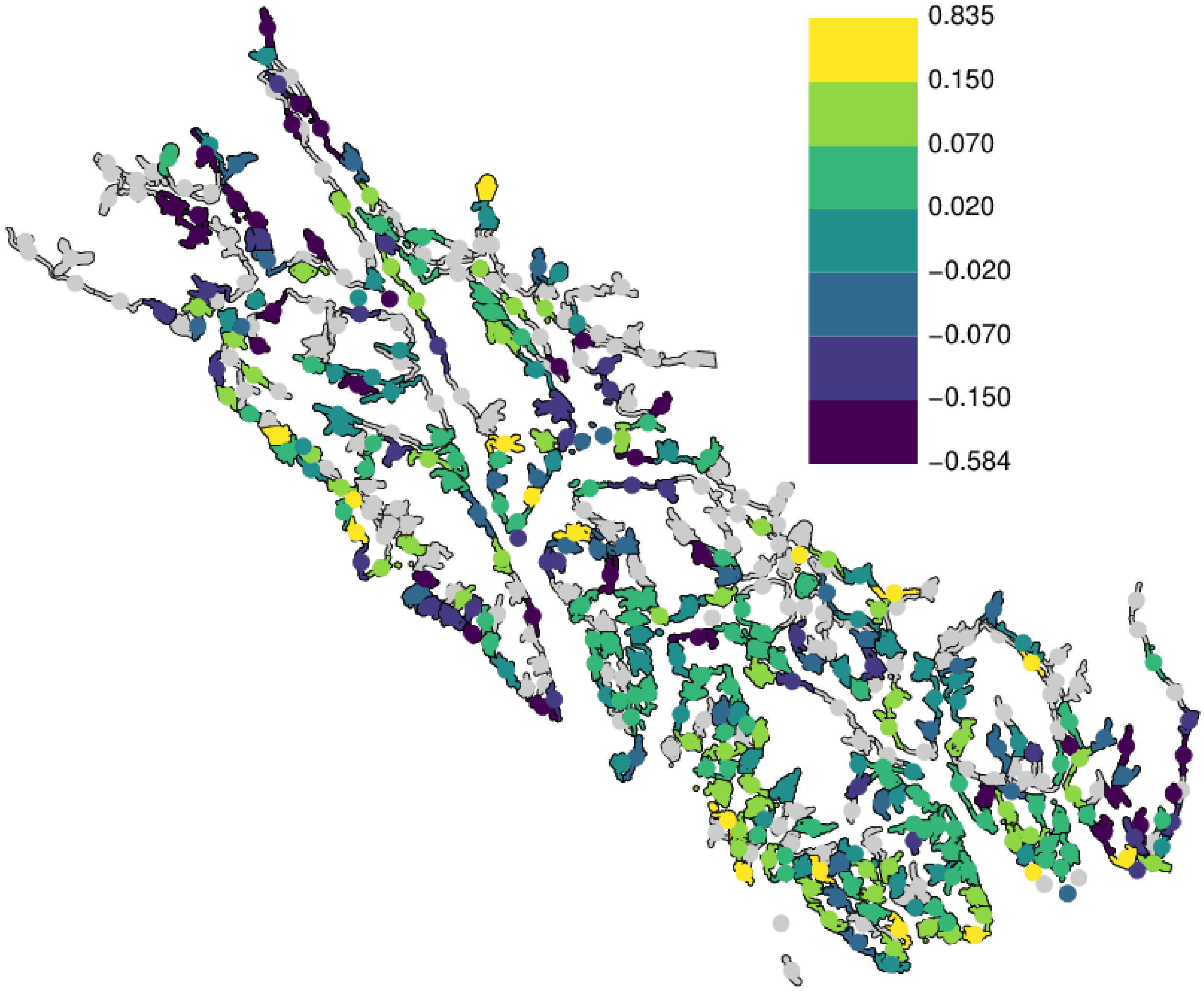

Figure 3: Map of the estimated trends (used as our raw data), where polygons are colored by their trend values. The light grey polygons have missing data. Because some polygons were small and it was difficult to see colors in them, all polygons were also overwritten by a circle of the same color. The trend values were categorized by colors, with increasing trends in yellows and greens, and decreasing trends in blues and violets, with the cutoff values given by the color ramp.

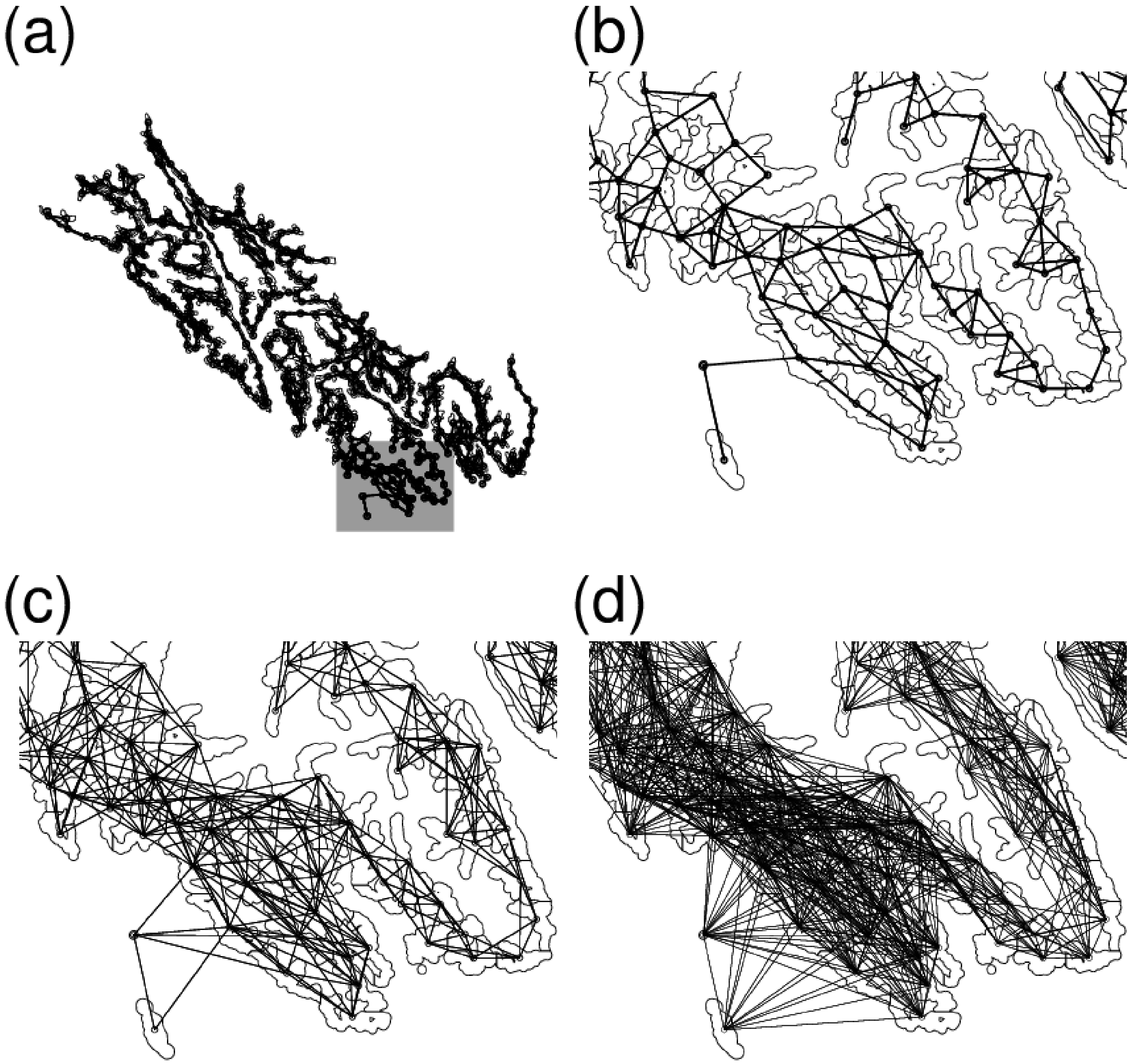

Figure 4: First, second, and fourth-order neighbor definitions for the survey polygons. (a) First-order neighbors for all polygons. The grey rectangle is the area for a closer view in the following subfigures: (b) first-order neighbors; (c) second-order neighbors; and (d) fourth-order neighbors.

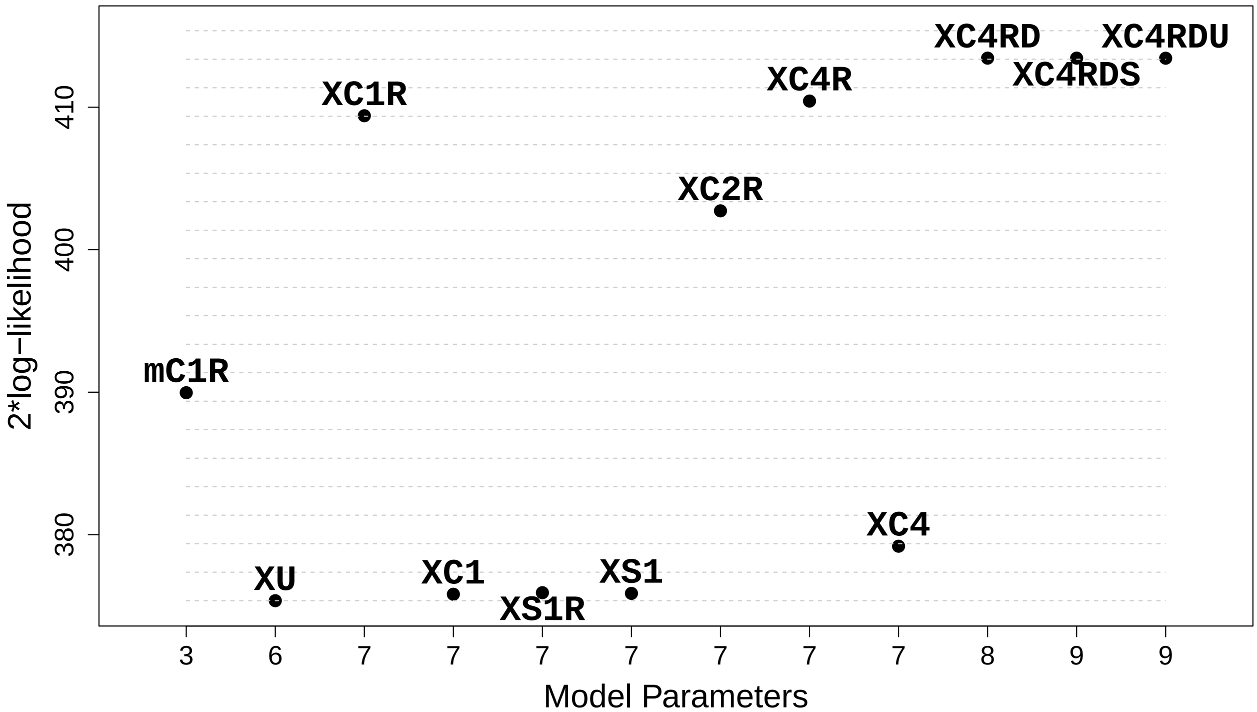

Figure 5: Two times the log-likehood for the optimized (maximized) fit for the models given in Table 2. Model mU had a much lower value (350.2) and is not shown. Starting with model XU, the dashed grey lines show increments of 2, which helps evaluate the relative importance of models by either an AIC or a likelihood-ratio test criteria.

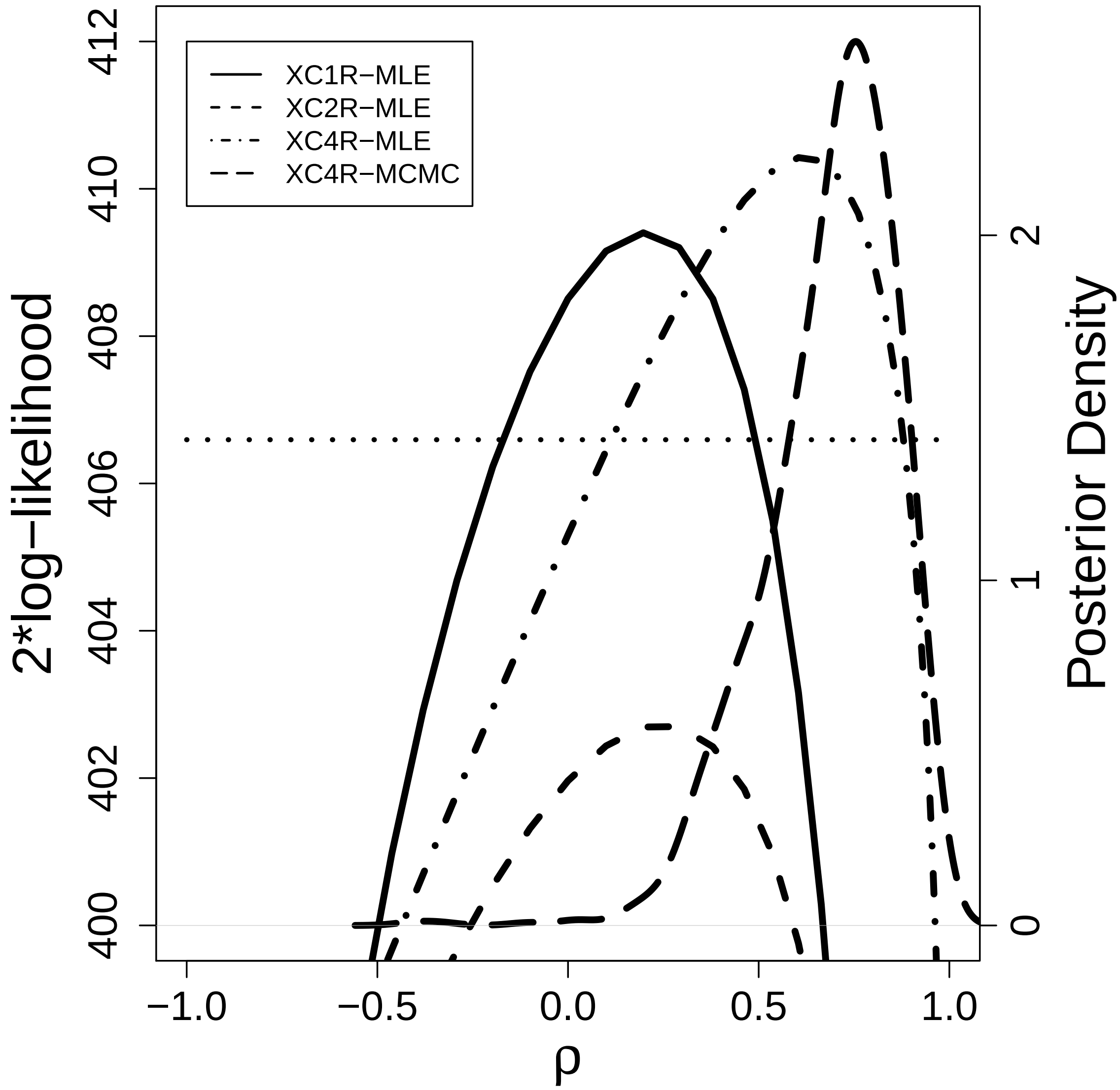

Figure 6: The different lines show 2∗log-likelihood profiles of ρ for three different models, listed in the legend. If the model is followed by -MLE, then the maximum of the profile provides the maximum likelihood estimate, and the 2∗log-likelihood is given by the left y-axis, while if it is followed by MCMC, then it is the posterior distribution from a Bayesian model with a uniform prior on ρ, and the density is given by the right y-axis. The horizontal dotted line is the maximum value for XC4R minus 3.841, the 0.05 α-level value of a χ-squared distribution on one degree of freedom.

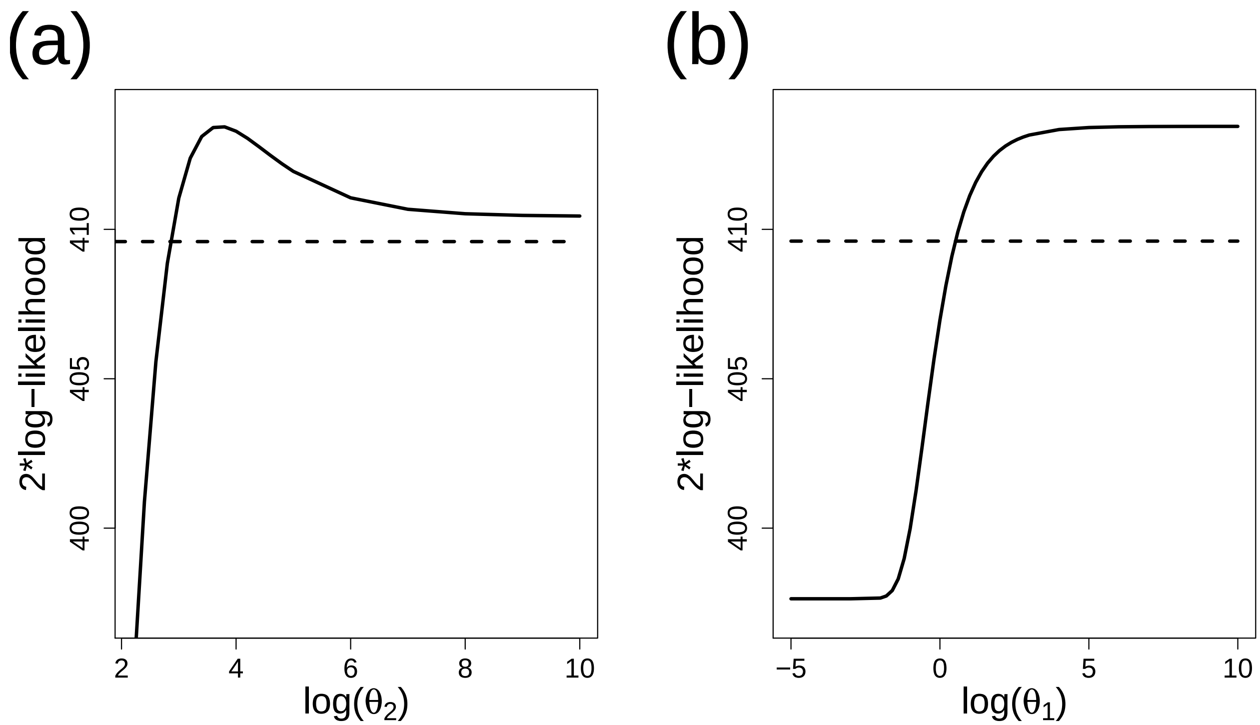

Figure 7: A) The solid line is the 2*Log-likelihood profile of θ_2 for model XC4RD. B) The solid line is the 2*Log-likelihood profile of θ_1 for model XC4RDS. For each figure, the horizontal dashed line is the maximum value for the model minus 3.841, the 0.05 α-level value of a χ-squared distribution on one degree of freedom.

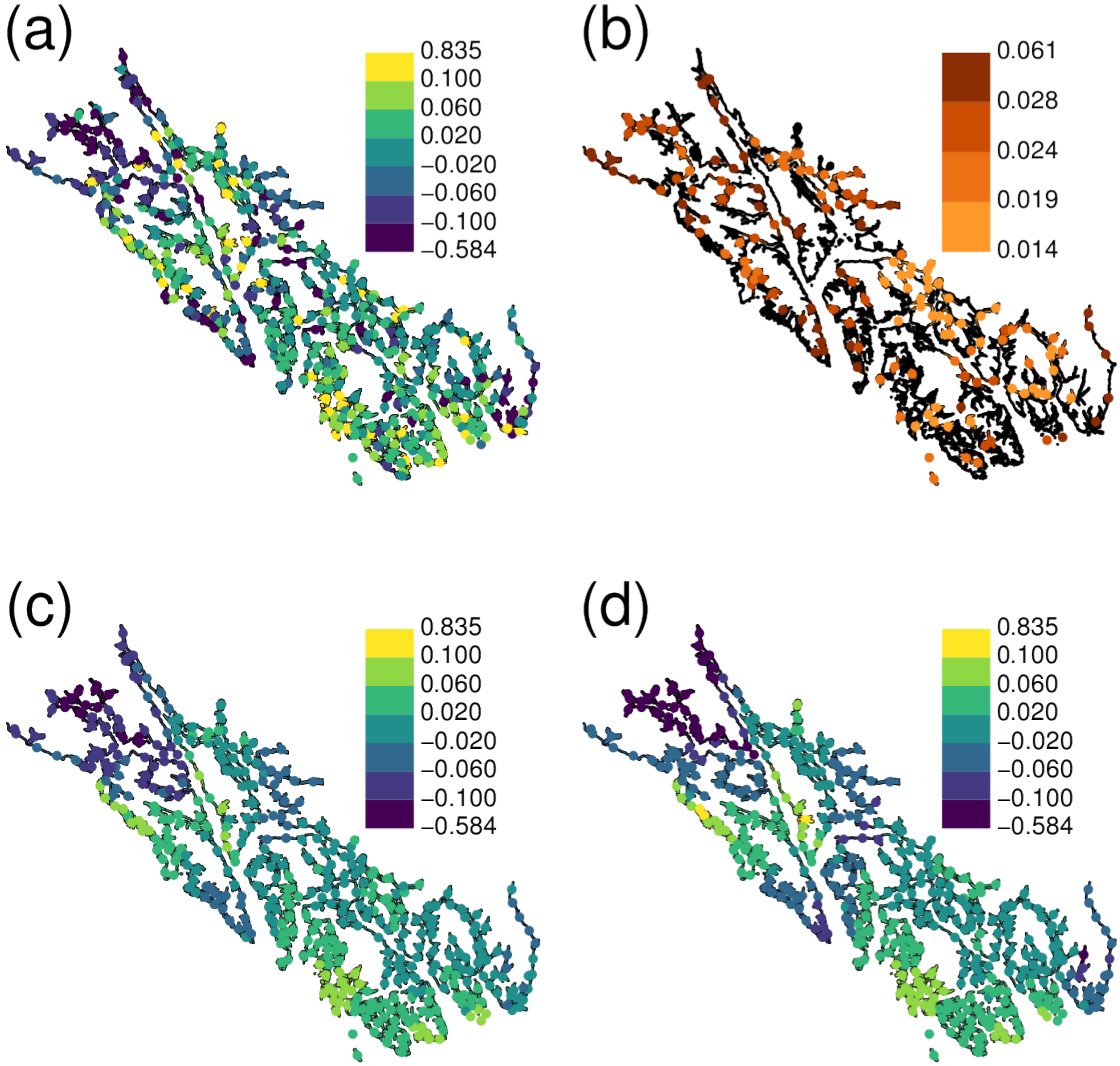

Figure 8: Predictions and smoothing for the harbor-seal stock-trend data. (a) Predictions, using universal kriging from the XC4R model, at unsampled locations have been added to the raw observed data from sampled locations. (b) Prediction standard errors for unsampled locations using universal kriging from XC4R. (c) Smoothing over all locations using conditional expectation based on the XC4R model. (d) Smoothing over all locations by using posterior predictions (mean of posterior distributions) using the XI4RU model in a Bayesian hierarchical model.

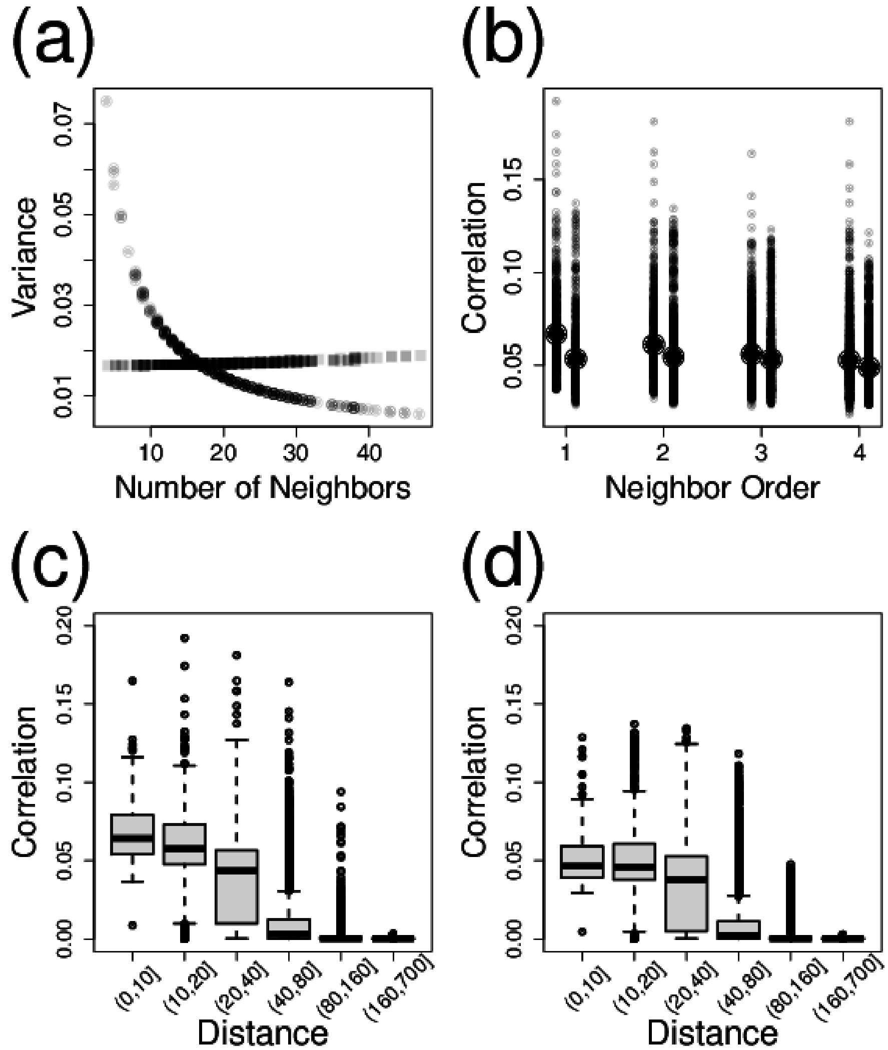

Figure 9: Nonstationarity illustrated for XC4 model, and the same model using row-standardization, XC4R. A) Marginal variances of the multivariate covariance matrix (diagonal elements of $\mathbf{\Sigma}$) as a function of the numbers of neighbors, where circles indicate XC4R and squares indicate XC4. Each symbol is partially transparent. B) All pairwise correlations as a function of the neighborhood order between sites. On the left of each neighbor order is XC4R, and on the right is XC4. The larger circle is the average value. C) and D) Boxplots of pairwise correlation as a function of distance between polygon centroids, binned into classes, for models XC4R and XC4, respectively.

1 APPENDIX A: Misconceptions and Errors in the Literature

The fact that CAR and SAR models are developed from the precision matrix, in contrast to geostatistical models being developed for the covariance matrix, has caused some confusion in the ecological literature. For example, in comparing geostatistical models to SAR models, Beguería and Pueyo (2009) stated "Semivariogram models account for spatial autocorrelation *at all possible distance lags*, and thus they do not require *a priori* specification of the window size and the covariance structure," (emphasis by the original authors). CAR and SAR models also account for spatial autocorrelation at all possible lags, as seen in Fig. 9c,d. In a temporal analogy, the autoregressive AR1 time series models also account for autocorrelation at all possible lags, where the conditional specification $Z_{i+1} = \phi Z_i + \nu_i$, with ν_i an independent random shock and $|\phi| < 1$, implies that $\mathrm{corr}(Z_i, Z_{i+t}) = \phi^t$ for all t. In fact, if we restrict $0 < \phi < 1$, then this can be reparameterized as $\mathrm{corr}(Z_i, Z_{i+t}) = \exp(-t(-\log(\phi)))$, which is an exponential geostatistical model with range parameter $-\log(\phi)$. While there are interesting results in Beguería and Pueyo (2009), a restriction on the range of autocorrelation is not a reason that CAR/SAR models might perform poorly against a geostatistical model. The important concept is that the autoregressive specification is local in the precision matrix and not in the covariance matrix.

CAR models are often incorrectly characterized. For example, Keitt et al. (2002) characterized CAR models as: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{C}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$, with a stated covariance matrix of $\sigma^2(\mathbf{I} - \rho\mathbf{C})^{-1}$, where $\mathbf{C}$ is symmetric. Their actual implementation may have been correct, and there are excellent and important results in Keitt et al. (2002); however, the construction they used leads to a SAR covariance matrix of $\sigma^2(\mathbf{I} - \rho\mathbf{C})^{-1}(\mathbf{I} - \rho\mathbf{C})^{-1}$ if $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ and $\mathbf{C}$ is symmetric. Even to characterize a CAR model as $\sigma^2(\mathbf{I} - \rho\mathbf{C})^{-1}$ with symmetric $\mathbf{C}$ is overly restrictive, as we have demonstrated that an asymmetric $\mathbf{C}$ with the proper $\mathbf{M}$ will still satisfy

Eq. 12, or alternatively that $\mathbf{\Sigma}^{-1} = (\mathbf{M}^{-1} - \mathbf{C})/\sigma^2$, where $\mathbf{C}$ is symmetric but $\mathbf{M}^{-1}$ is not necessarily constant on the diagonals. In fact, constraining a CAR model to $\sigma^2(\mathbf{I} - \rho\mathbf{C})^{-1}$ does not allow for row-standardized models. These mistakes are perpetuated in Dormann et al. (2007), and we have seen similar errors in describing CAR models as SAR models in other literature, presentations, and help sites on the internet.

Dormann et al. (2007) also claimed that any SAR model is a CAR model, which agrees with the literature (e.g., Cressie, 1993, p. 409), but then they show an incorrect proof (it is also incorrect in Haining (1990, p. 89), and likely beginning there), because they do not consider that $\mathbf{C}$ for a CAR model must have zeros along the diagonal. In fact, we demonstrate in Ver Hoef et al. (2017) that, despite statistical and ecological literature to the contrary, CAR models and SAR models can be written equivalently, and we provide details.

2 APPENDIX B: Maximum Likelihood Estimation for CAR/SAR Models with Missing Data

We begin by finding analytical solutions when we can, and then substituting them into the likelihood to reduce the number of parameters as much as possible for the full covariance matrix. Assume a linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is a vector of response variables, $\mathbf{X}$ is a design matrix of full rank, $\boldsymbol{\beta}$ is a vector of parameters, and the zero-mean random errors have a multivariate normal distribution, $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a patterned covariance matrix; i.e., it has non-zero off-diagonal elements. Suppose that $\boldsymbol{\Sigma}$ has parameters $\{\theta, \boldsymbol{\rho}\}$ and can be written as $\boldsymbol{\Sigma} = \theta\mathbf{V}_{\boldsymbol{\rho}}$, where θ is an overall variance parameter and $\boldsymbol{\rho}$ are parameters that structure $\mathbf{V}_{\boldsymbol{\rho}}$ as a non-diagonal matrix, and we show the dependency as a subscript. Note that $\boldsymbol{\Sigma}^{-1} = \mathbf{V}_{\boldsymbol{\rho}}^{-1}/\theta$. Recall that the maximum likelihood estimate of $\boldsymbol{\beta}$ for any $\{\theta, \boldsymbol{\rho}\}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{y}$. By substituting $\hat{\boldsymbol{\beta}}$ into the normal likelihood equations, -2 times the log-likelihood for a normal distribution is

$$\mathcal{L}(\theta, \boldsymbol{\rho}|\mathbf{y}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \log(|\boldsymbol{\Sigma}|) + n\log(2\pi),$$

where n is the length of $\mathbf{y}$, but this can be written as,

$$\mathcal{L}(\theta, \boldsymbol{\rho}|\mathbf{y}) = \mathbf{r}_{\boldsymbol{\rho}}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{r}_{\boldsymbol{\rho}}/\theta + n\log(\theta) + \log(|\mathbf{V}|) + n\log(2\pi) \tag{B.1}$$

where $\mathbf{r}_{\boldsymbol{\rho}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ (notice that $\hat{\boldsymbol{\beta}}$ is a function of $\boldsymbol{\rho}$, so we show that dependency for $\mathbf{r}$ as well). Conditioning on $\boldsymbol{\rho}$ yields

$$\mathcal{L}(\theta|\boldsymbol{\rho}, \mathbf{y}) = \mathbf{r}_{\boldsymbol{\rho}}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{r}_{\boldsymbol{\rho}}/\theta + n\log(\theta) + \text{terms not containing } \theta$$

and minimizing (analytically) for θ involves setting

$$\frac{\partial\mathcal{L}(\theta|\boldsymbol{\rho}, \mathbf{y})}{\partial\theta} = -\mathbf{r}_{\boldsymbol{\rho}}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{r}_{\boldsymbol{\rho}}/\theta^2 + n/\theta$$

equal to zero, yielding the maximum likelihood estimate

$$\hat{\theta} = \mathbf{r}_{\boldsymbol{\rho}}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{r}_{\boldsymbol{\rho}}/n. \tag{B.2}$$

Substituting Eq. B.2 back into Eq. B.1 yields the -2*log-likelihood as a function of $\boldsymbol{\rho}$ only,

$$\mathcal{L}(\boldsymbol{\rho}|\mathbf{y}) = n\log(\mathbf{r}_{\boldsymbol{\rho}}'\mathbf{V}_{\boldsymbol{\rho}}^{-1}\mathbf{r}_{\boldsymbol{\rho}}) + \log(|\mathbf{V}_{\boldsymbol{\rho}}|) + n(\log(2\pi) + 1 - \log(n)). \tag{B.3}$$

Equation Eq. B.3 can be minimized numerically to yield the MLE $\hat{\boldsymbol{\rho}}$, and then $\hat{\theta} = \mathbf{r}_{\hat{\boldsymbol{\rho}}}'\mathbf{V}_{\hat{\boldsymbol{\rho}}}^{-1}\mathbf{r}_{\hat{\boldsymbol{\rho}}}/n$, and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}_{\hat{\boldsymbol{\rho}}}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}_{\hat{\boldsymbol{\rho}}}^{-1}\mathbf{y}$ yield analytical solutions for MLEs after obtaining (numerically) the MLE for $\boldsymbol{\rho}$.

We developed the inverse covariance matrix $\boldsymbol{\Sigma}_A^{-1} = \text{diag}(\mathbf{W}\mathbf{1}) - \rho\mathbf{W}$, and here we use $\boldsymbol{\Sigma}_A$ to denote it is for *all* locations, those with observed data as well as those without. Without missing data, Eq. B.3 can be evaluated quickly by factoring out an overall variance parameter from $\boldsymbol{\Sigma}_A^{-1}$ and using sparse matrix methods to quickly and efficiently evaluate $|\mathbf{V}_{\boldsymbol{\rho}}|$ by recalling that $|\mathbf{V}_{\boldsymbol{\rho}}| = 1/|\mathbf{V}_{\boldsymbol{\rho}}^{-1}|$. However, when there are missing data, there is no guarantee that $\mathbf{V}_{\boldsymbol{\rho}}$ will be sparse. The obvious and direct approach is to first obtain $\mathbf{\Sigma}_A = (\mathbf{\Sigma}_A^{-1})^{-1}$, and then obtain $\mathbf{V}_{\boldsymbol{\rho}} = \mathbf{\Sigma}[\mathbf{i},\mathbf{i}]$, where $\mathbf{i}$ is a vector of indicators that subsets the rows and columns of $\mathbf{\Sigma}$ to only those for sampled locations. Then, a third step is a second inverse to find $\mathbf{V}_{\boldsymbol{\rho}}^{-1}$. However, this is computationally expensive. A faster way uses results from partitioned matrices and Schur complements. In general, let the square matrix $\mathbf{\Sigma}$ with dimensions $(m+n)\times(n+m)$ be partitioned into block submatrices,

$$\underset{(m+n)\times(m+n)}{\mathbf{\Sigma}} = \begin{bmatrix} \underset{m\times m}{\mathbf{A}} & \underset{m\times n}{\mathbf{B}} \\ \underset{n\times m}{\mathbf{C}} & \underset{n\times n}{\mathbf{D}} \end{bmatrix}$$

with dimensions given below each matrix. Assume $\mathbf{A}$ and $\mathbf{D}$ are nonsingular. Then define the matrix function $\mathbf{S}(\mathbf{\Sigma},\mathbf{A}) = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ as the Schur complement of $\mathbf{\Sigma}$ with respect to $\mathbf{A}$. Likewise, there is a Schur complement with respect to $\mathbf{D}$ by reversing the roles of $\mathbf{A}$ and $\mathbf{D}$. Using Schur complements, it is well-known (e.g., Harville, 1997, p. 97) that an inverse for a partitioned matrix $\mathbf{\Sigma}$ is,

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1} \\ -\mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1} \end{bmatrix}$$

Then, note that $\mathbf{A}^{-1} = \mathbf{S}(\mathbf{\Sigma}^{-1},\mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1})$; that is, if we already have $\mathbf{\Sigma}^{-1}$, then $\mathbf{A}^{-1}$ is the Schur complement of $\mathbf{\Sigma}^{-1}$ with respect to the rows and columns that correspond to $\mathbf{D}$. Additionally, the largest matrix that we have to invert is $[\mathbf{S}(\mathbf{\Sigma},\mathbf{A})^{-1}]^{-1}$, which is $n\times$n and has dimension less than $\mathbf{\Sigma}$, and only one inverse is required. Thus, if we let $\mathbf{A}$ correspond to the m rows and columns of the observed locations, and $\mathbf{D}$ correspond to the n rows and columns of the missing data, then this provides a quick and efficient way to obtain $\mathbf{V}_{\boldsymbol{\rho}}^{-1}$ from $\mathbf{\Sigma}_A^{-1}$, and the largest inverse required is of dimension $n\times n$, where n is the number of sites with missing data.

3 APPENDIX C: Prediction and Smoothing

In what follows, we provide the formulas used in creating Fig. 8. For universal kriging, the formulas can be found in Cressie and Wikle (2011, p. 148),

$$\hat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \mathbf{c}_i\boldsymbol{\Sigma}_{-i}^{-1}(\mathbf{y}_{-i} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\hat{y}_i$ is the prediction for the ith node, $\mathbf{x}_i$ is a vector containing the covariate values for the ith node, $\mathbf{X}$ is the design matrix for the covariates (fixed effects), $\mathbf{c}_i$ is a vector containing the fitted covariance between the ith site and all *other* sites with observed data, $\boldsymbol{\Sigma}$ is the fitted covariance matrix among all observed data, $\mathbf{y}$ is a vector of observed values for the response variable, and $\hat{\boldsymbol{\beta}} = \mathbf{X}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$ is the generalized least squares estimate of $\boldsymbol{\beta}$. The covariance values contained in $\mathbf{c}_i$ and $\boldsymbol{\Sigma}$ were obtained using maximum likelihood estimate for the parameters as detailed in Appendix B. We include the $-i$ subscript on $\boldsymbol{\Sigma}_{-i}^{-1}$ and $\mathbf{y}_{-i}$ to indicate that, when smoothing, we predict at the ith node by removing that datum from $\mathbf{y}$, and by removing its corresponding rows and columns in $\boldsymbol{\Sigma}$. If the value is missing, then prediction proceeds using all observed values. Hence, Fig. 8a contains the observed values plus the predicted values at nodes with missing values, while Fig. 8c contains predicted values at all nodes, where any observed value at a node was removed and predicted with the rest of the observed values. The prediction standard errors are given by,

$$\hat{\mathrm{se}}(\hat{y}_i) = \sqrt{\mathbf{c}_i'\boldsymbol{\Sigma}_{-i}^{-1}\mathbf{c}_i + \mathbf{d}_i'(\mathbf{X}_{-i}'\boldsymbol{\Sigma}_{-i}^{-1}\mathbf{X}_{-i})^{-1}\mathbf{d}_i}$$

where $\mathbf{d}_i = \mathbf{x}_i' - \mathbf{X}_{-i}'\boldsymbol{\Sigma}_{-i}^{-1}\mathbf{c}_i$.

For the IAR model smoothing in Fig. 8d, we used the WinBUGS (Lunn et al., 2000) software, final version 1.4.3. The model code is very compact and given below:

```
model
  {
    for(i in 1:N) {
      trend[i] ~ dnorm(mu[i],prec)
      mu[i] <- beta0 + beta[stockid[i]] + b[i]
          }
    b[1:N] ~ car.normal(adj[], weights[], num[], tau)
    beta0 ~ dnorm(0,.001)
    beta[1] ~ dnorm(0,.001)
    beta[2] ~ dnorm(0,.001)
    beta[3] ~ dnorm(0,.001)
    beta[4] <- 0
    beta[5] ~ dnorm(0,.001)
    prec ~ dgamma(0.001,0.001)
    sigmaEps <- sqrt(1/prec)
    tau  ~ dgamma(0.5, 0.0005)
    sigmaZ <- sqrt(1 / tau)
  }
```

The means of the MCMC samples from the posterior distributions of `mu[i]` were used for the IAR smoothing for the ith location in Fig. 8d.