Spatial Autoregressive Models for Ecological Inferences

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

[1]National 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] Department of Statistics, 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, Penn State University
[5] Department of Ecology and Evolutionary Biology, University of Toronto

February 26, 2016

Summary

1. We review conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models for spatial ecological data because they are under-used, or used incorrectly, and this is likely because they are more difficult to understand than geostatistical models. To highlight their usefulness, we identify and discuss six ecological inferences using CAR and SAR models, including model selection, spatial regression, estimation of autocorrelation, estimation of other connectivity parameters, spatial prediction, and spatial smoothing.

2. We compare and constrast CAR and SAR models, showing their development and connection to partial correlations. Special cases, such as the intrinsic autoregressive model (IAR), are also shown. Practical use of CAR and SAR models depends on weight matrices, and neighborhood definition and row-standardization are important concepts when developing such matrices. Weight matrices can also include ecological covariates and connectivity structures, which we encourage but have been rarely used.

3. Trends in harbor seals (*Phoca vitulina*) in southeastern Alaska from 463 polygons, some with missing data, are used to illustrate the six ecological inferences and highlight discussion points from the earlier review. We develop a variety of weight matrices, and CAR and SAR models are fit to data using maximum likelihood and Bayesian methods. We compare models and present spatial regression results for several models. 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.

4. We recommend that ecologists make greater use of autoregressive models, both directly and in hierarchical models, and not only in explicit spatial settings, but also for more general connectivity and graphical models. We conclude with some take-home lessons for CAR and SAR models, including 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.

KEY WORDS: Conditional autoregressive, simultaneous autoregressive, CAR, SAR, IAR, geostatistics, prediction, smoothing

INTRODUCTION

When data are collected at spatial locations, Cressie (1993, p. 8) divides spatial statistical models into two broad classes: 1) point-referenced geostatistical models, and 2) lattice models, also called areal models (Banerjee et al., 2004). The two most common lattice models are the conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models. The defining characteristic for CAR and SAR models is that the data occur on a (possibly irregular) grid, or lattice, with a countable set of nodes. Geostatistical models, on the other hand, are used for data that are spatially continuous (a continuous surface). Another distinction that we find more useful for conceptual understanding is that geostatical models use spatial information to directly model a covariance matrix, whereas CAR and SAR models use spatial information to model the inverse of a covariance matrix (also known as the precision matrix). Modeling the precision matrix is less intuitive than modeling the covariance matrix, which may make it more difficult to understand the implications of modeling decisions. Indeed, in certain circumstances Wall (2004) found some surprising and unusual behavior for CAR and SAR models. The less intuitive precision matrix, and a reputation for unusual behavior, may unfortunately keep some scientists from using these models, or using them incorrectly. For example, papers have incorrectly compared CAR/SAR to geostatistical models (e.g., Beguería and Pueyo, 2009), incorrectly formulated the CAR model (e.g., Keitt et al., 2002), and have given incorrect relationships between CAR and SAR models (e.g., Dormann et al., 2007). We do not dwell on these (details are given in Appendix A), but rather they illustrate that good statistical practice with CAR and SAR models depends on more and better information. CAR and SAR models have many useful applications (e.g., Gelfand et al., 2005; Latimer et al., 2006; Magoun et al., 2007; Hanks and Hooten, 2013). 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.

First we motivate the uses of spatial autoregressive models with typical (and not so typical, but useful) objectives for ecological studies (Table 1). For objective 1, model selection, there are a plethora of model comparison methods, or multimodel inferences, based on AIC, BIC, etc., that are available (see, e.g., Burnham and Anderson, 2002). 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 import goal of the investigation. For objective 2, regression, an early and influential paper on using CAR models for Scottish lip cancer data (Clayton and Kaldor, 1987) found that the covariate "% agriculture" helped explain lip cancer rates among counties in Scotland, in conjunction with a CAR model for spatial autocorrelation. For another example, in a spatial CAR regression model using wolverine data (Gardner et al., 2010), the probability of occupancy depended on covariates related to elevation and human influence in the plots. As an example for objective 3, autocorrelation, using a Bayesian model, Gardner et al. (2010) provide estimates of an autocorrelation parameter, along with credibility intervals. Objective 4, covariate effects on autocorrelation, is almost never used in ecological models, or in other disciplines, but we give an example later. Generally, the neighborhood structure that forms the covariance matrix is considered a nuisance; that is, it is a requirement to acknowledge spatial autocorrelation to make valid inferences on the other objectives in this list. An example of objective 5, prediction, for CAR models is given in Gardner et al. (2010), who modeled occupancy of wolverines from aerial surveys, and 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. Objective 6, smoothing, is the classic idea behind the Scottish lip cancer data (Clayton and Kaldor, 1987). In this example, all counties have observed cancer rates. To conceptualize the smoothing objective, imagine that disease rates in administrative 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 administrative district has a single sample that is positive for the disease. It would be unrealistic to estimate the whole administrative district to have 100% disease rates based on that single sample. The models of Clayton and Kaldor (1987) create rates that smooth over observed data by using values from nearby locations to provide better estimates, in general.

Based on the introduction above, it should be clear that CAR and SAR models can be used in many applications for ecological data. Our objectives are: 1) to explain, in ecological terms, how these models are obtained, 2) give insight and intuition on how they work, 3) to compare CAR and SAR models, and 4) give practical guidelines for their use. We provide an example for further

illustration of the objectives given in Table 1. For discussion, identify areas 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 autocorrelation, or autocovariance, quantitatively represents the degree of statistical dependency between random variables using spatial relationships (Cressie, 1993). It is a common characteristic of ecological data, which are often collected at a specific location in space and in time. Analyzing spatially correlated data requires the use of spatial statistical models because the assumption of independence is violated, making many conventional statistical methods inappropriate.

In general terms, a spatial statistical model accounts for the spatial location of the data (i.e., spatial indexing) in the probabilistic part of the model. A wide variety of methods have been developed and so it is useful to start with a general representation of a spatial statistical model, $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ (Cressie, 1993, pg. 8). Here, Z is an observed or unobserved random variable at location $\mathbf{s}$, which belongs to the spatial domain of interest D. The random variable could represent the presence or absence of a species, percent canopy cover, allele counts, etc., while the domain of interest could be a study area, a management unit, or a biogeographic region, etc. In this paper we focus on spatial models where D is a finite set and where distance is not necessarily uniquely defined. For example, political adminstrative districts (such as counties, boroughs, etc.) have area, any set of them is a finite number; they are spatially well-defined, generally as a polygon in a computer system such as a geographic information system (GIS), but (Euclidean) distance between areal units is not well-defined because that requires two points in a Euclidean coordinate system. Rather than model spatial autocovariance based on Euclidean distance, as is common in geostatistics, spatial relationships for areal data are based on a graphical model, or a network,

where sites are depicted as nodes, and edges denote connections. Edges can be defined in many ways, but a common approach is to create an edge between adjoining administrative units.

The covariance matrices of geostatistical models, where $\mathbf{s}$ varies continuously within D, could be used directly to model data from a finite set of locations within D if Euclidean distance was an appropriate way to express their correlation. The main problem with geostastistical models is that they are guaranteed to be valid only for Euclidean metrics. Consider the case where binary values are used to express connections; the resulting topology is no longer guaranteed to be a Euclidean space. For example, Ver Hoef et al. (2006) show how nodes, connected in a branching network (representing stream nodes), create a topology for which covariance matrices from commonly-used geostatistical models are inappropriate. In this case, one possibility is to create models that are appropriate for the topology (e.g., Ver Hoef and Peterson, 2010), but stream networks are a special topology, and such models are not general. Alternatively, the topology could be deformed to match Euclidean space, such as is done in multidimensional scaling (Curriero, 2006), but this suffers from sensitivity to the exact configuration of points. That is, adding or dropping a node changes the multidimensional scaling. There is a class of statistical models, however, that are ready-made for spatial network models; the spatial autoregressive models. As we demonstrate, these models are specified through the inverse of the covariance matrix, so they are somewhat less intuitive. For that reason, we now discuss them and attempt to give a fuller understanding of their properties.

Spatial autoregressive models are, fundamentally, not spatial at all. They are also known as graphical models (e.g., Lauritzen, 1996; Whittaker, 2009) and Gaussian Markov random fields (e.g., Rue and Held, 2005). However, when spatial information is used to define nodes and edges, they are also known as lattice models (e.g., Cressie, 1993, pg. 8) or areal models (e.g., Banerjee et al., 2004, pg. 69). These models differ from the geostatistical models because D is fixed and finite, rather than continuous (Cressie, 1993, pg. 8). As mentioned previously, an areal unit may not have a single set of spatial coordinates and so the data are indexed by i in the graphical model, rather than $\mathbf{s}$. For example, Z_i is the random variable for the ith node, where $i = 1, 2, \ldots, N$. Now

consider the spatial regression framework,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}, \qquad \text{eqn 1}$$

where the goal is to model a first-order mean structure that includes covariates (i.e., predictor variables, $\mathbf{X}$, measured on nodes), as well as a latent spatial random process $\mathbf{z}$, where $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and independent error $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{\varepsilon}\mathbf{I})$. 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, $\boldsymbol{\Sigma}$, takes one of two main forms: 1) the SAR model,

$$\boldsymbol{\Sigma} \equiv \sigma^2_Z((\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}'))^{-1}, \qquad \text{eqn 2}$$

or, 2) the CAR model,

$$\boldsymbol{\Sigma} \equiv \sigma^2_Z(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}. \qquad \text{eqn 3}$$

Here, spatial dependency 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 ($m_{ij} = 0 \; \forall \; i \neq j$), where m_{ii} is proportional to the conditional variance of Z_i given all of its neighbors. The spatial dependence matrices can be further decomposed into $\mathbf{B} = \rho\mathbf{W}$ and $\mathbf{C} = \rho\mathbf{W}$, where $\mathbf{W}$ is a weights matrix and ρ controls the strength of dependency.

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 others variables. If $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega} = \{\omega_{i,j}\}$, then the partial correlation between random variable 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. Thus, we can see that the CAR model, in particular, allows the modeler to directly specify partial correlations (or covariances), rather than correlation directly. That is, we are in control of specifying the off-diagonal matrix values of $\mathbf{W}$ in $\boldsymbol{\Sigma}^{-1} = \sigma^2_Z\mathbf{M}^{-1}(\mathbf{I} - \boldsymbol{\rho}\mathbf{W})$, 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, which encodes partial correlations. Contrast this to geostatistics, where we are in control of specifying $\mathbf{\Sigma}$, and therefore we are specifying the correlations. In both cases, we generally use some algorithm, often with parameters, rather than specifying every matrix entry individually. For CAR and SAR models, the algorithm is often based on neighbors (e.g., partial correlation exists between neighbors that share a boundary), and for geostatistics the algorithm is based on distance (e.g., correlation depends on an exponential decay with distance). If $b_{ij} = 0$ or $c_{ij} = 0$, they are partially uncorrelated; otherwise there is partial dependence. Note that b_{ii} and c_{ii} are always zero. In order 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 $(\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}')$ or $(\mathbf{I} - \mathbf{C})$ to have an inverse; that is, ρ cannot be chosen arbitrarily, and its range depends on the weights in $\mathbf{W}$.

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 the inferences that are gained from these models. As such, we describe each model in more detail, and later we give more practical advice.

SAR Models

One approach for building the SAR model begins with the usual regression formulation described in (eqn 1). Now, instead of modeling the correlation of $\mathbf{z}$ directly, an explicit autocorrelation structure is imposed,

$$\mathbf{z} = \mathbf{B}\mathbf{z} + \boldsymbol{\nu}, \qquad \text{eqn 4}$$

where the spatial dependence matrix, $\mathbf{B} = \rho\mathbf{W}$, 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 $\sigma_Z^2((\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}'))^{-1}$. The spatial dependency 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). However, there are obvious constraints to allow $(\mathbf{I} - \mathbf{B})(\mathbf{I} - \mathbf{B}')$ to be a precision matrix.. These constraints are best explored through the eigenvectors and eigenvalues of $\mathbf{W}$. The matrix $\mathbf{W}$ must be nonsingular; that is, it cannot have any zero eigenvalues. Also, if $\lambda_{[1]} < 0$ is the smallest eigenvalue, and $\lambda_{[N]} > 0$ is the largest eigenvalue of $\mathbf{W}$, then $1/\lambda_{[1]} < \rho < 1/\lambda_{[N]}$.

The model created by (eqn 1) and (eqn 4) has been termed the "error" model version of SAR models. An alternative is to simultaneously autoregress the whole response variable, not just the errors, $\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (Anselin, 1988), yielding

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}, \qquad \text{eqn 5}$$

which allows the matrix $\mathbf{W}$ to smooth covariates in $\mathbf{X}$ as well as creating autocorrelation in the error for $\mathbf{y}$. A final version is to simultaneously autoregress both response and a separate random effect $\boldsymbol{\nu}$,

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\nu} + \boldsymbol{\varepsilon}, \qquad \text{eqn 6}$$

see, e.g., Kissling and Carl (2008).

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

$$Z_i|\mathbf{z}_{-i} \sim \mathrm{N}\left(\sum_{\forall c_{ij} \neq 0} c_{ij} z_j, m_{ii}\right), \qquad \text{eqn 7}$$

where $\mathbf{z}_{-i}$ is the vector of all Z_j where $j \neq i$, $\mathbf{C} = \rho\mathbf{W}$ is the spatial dependency 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 (eqn 7) 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, $(\mathbf{I}-\mathbf{C})^{-1}$ must exixt and $\mathbf{\Sigma} = \sigma_n^2(\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}$ must be symmetric, which requires that

$$\frac{c_{ij}}{\tau_i^2} = \frac{c_{ji}}{\tau_j^2} \; \forall \; i, j. \qquad \text{eqn 8}$$

For CAR models, $\mathbf{W}$ and ρ are constrained in exactly the same way as for SAR models; $\mathbf{W}$ must be non-singular and $1/\lambda_{[1]} < \rho < 1/\lambda_{[N]}$ for $\lambda_{[1]}$ the smallest, and $\lambda_{[N]}$ the largest eigenvalues of $\mathbf{W}$.

A special case of the CAR model, called the intrinsic autoregressive model (IAR) (Besag and Kooperberg, 1995), occurs when the weights in (eqn 7) occur as,

$$Z_i \sim \mathrm{N}\left(\sum_{j \in \mathcal{N}_i} z_j/|\mathcal{N}_i|, \tau^2/|\mathcal{N}_i|\right), \qquad \text{eqn 9}$$

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 (eqn 9), each random variable is the average of its neighbors, and the variance is proportional to the inverse of the number of neighbors. This idea, of creating weights based on averages of neighboring values, is discussed next.

Row-standardardization

Here, 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 between sites i and j and a zero indicates no such connection. 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 $\mathbf{W}$. Thinking in terms of constructing random variables, this means

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 shock. 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. It is well-known that spatial autoregressive models for this example must have $|\rho| < 1/4$ (Haining, 1990, pg. 82). More generally, $|\rho| < 1/n$ if all sites have exactly n neighbors, $|\mathcal{N}_i| = n \,\forall\, i$, 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. In general then, 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 edges, 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 *edges*. Without row standardization, even when ρ controls overall variance, locations in the *middle*, summed over more neighbors, have higher variance than those at the edges. For an error process in (eqn 1), higher variance near the edges makes more sense, and, with a more natural and consistent range of values for ρ, we highly recommend row-standardization. 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 (eqn 8), and note that τ^2 will not be identifiable from σ^2_Z in (eqn

3), so the row-standardized CAR model can be written equivalently as,

$$\boldsymbol{\Sigma} = \sigma_Z^2(\mathbf{I} - \rho\mathbf{W}_+)^{-1}\mathbf{M}_+ = \sigma_Z^2(\text{diag}(\mathbf{W1}) - \rho\mathbf{W})^{-1}. \qquad \text{eqn 10}$$

Using row-standardization, and setting $\rho = 1$ in (eqn 7) leads to (eqn 9). In our AR1 analogy, this is equivalent to $\phi = 1$. In this case, $\boldsymbol{\Sigma}^{-1}$ is singular (i.e., does not have an inverse), and $\boldsymbol{\Sigma}$ does not exist. While this may seem undesirable, random walks and Brownian motion are stochastic processes without covariance matrices. 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 (eqn 1) that is a flexible surface and there is no autocorrelation parameter to estimate.

More Weighting – Separating Functional from 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 (other scientific and statistical literature), the autoregressive component is an "error" process, and not really of intrinsic 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. Here, 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 neighborhoods, distance, etc., and by functional, we mean correlation that is affected by measured 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. The structural idea can be contained 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,

$$w_{ij} = \begin{cases} f(\mathbf{e}_{ij}, \boldsymbol{\theta}), & j \in \mathcal{N}_i \\ 0, & j \notin \mathcal{N}_i. \end{cases} \qquad \text{eqn 11}$$

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 (eqn 11), 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 Hadmard (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 inference.

Comparison of CAR to SAR and 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 do not recommend versions of the SAR model given by (eqn 5) and (eqn 6). It is difficult to understand how smoothing/lagging covariates and extra random effects contributes 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 (eqn 4).

It is well-known that any SAR model can be written as a CAR model, and Cressie (1993, pg. 408) demonstrates how a SAR model with 4 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. Although it does not appear to be as simple to go from a CAR model to a SAR, contrary to published accounts (Cressie, 1993, pg. 408) and (Banerjee et al., 2004, pg. 86)), it is possible (but not uniquely, without further conditions, which we demonstrate in Appendix B). In fact, it is evident from (eqn 2) that specifying first-order neighbors in $\mathbf{B}$ will result in non-zero partial correlations between second-order neighbors because of the product $(\mathbf{I}-\mathbf{B})(\mathbf{I}-\mathbf{B})$. Hence, SAR models have a reputation as being less "local" than the CAR models. In fact, using the same construction $\rho\mathbf{W}$ for both SAR and CAR models, Wall (2004) shows that correlation (in $\mathbf{\Sigma}$, not partial correlation) increases more rapidly with ρ in SAR models than CAR models.

Regarding restrictions on ρ, Wall (2004) also shows there is very 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. 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 $\text{corr}(Z_i, Z_{i+t}) = \text{corr}(Z_i, Z_{i-t}) = \phi^t$. Note also that this shows that specifying partial correlations as zero (or conditional independence), does not mean that marginal correlation is zero (i.e., $\text{corr}(Z_i, Z_{i+t}) \neq 0$ 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 $\text{corr}(Z_i, Z_{i+t}) = \phi^t$ 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 edge; that is, correlation is nonstationary, just like the variance when we described row-standardization.

CAR and SAR in Hierarchical Models

We now turn our focus to 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,

$$\begin{aligned}
\mathbf{y} &\sim [\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\nu}], \\
\boldsymbol{\mu} &\equiv \mathbf{X}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}, \\
\mathbf{z} &\sim [\mathbf{z}|\boldsymbol{\Sigma}] \equiv \text{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\
\boldsymbol{\Sigma}^{-1} &\equiv \mathbf{F}(\mathbf{N}, \mathbf{D}, \rho, \boldsymbol{\theta}, \ldots), \\
\boldsymbol{\varepsilon} &\sim [\boldsymbol{\varepsilon}|\sigma^2] \equiv \text{N}(\mathbf{0}, \sigma^2\mathbf{I}),
\end{aligned} \qquad \text{eqn 12}$$

where $[\cdot]$ denotes a statistical distribution. 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 $[\mathbf{y}|g(\boldsymbol{\mu}), \boldsymbol{\nu}]$ 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 $\boldsymbol{\nu}$. 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 $\boldsymbol{\Sigma}$ 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 inverse of the covariance matrix, $\mathbf{\Sigma}^{-1}$, also called the precision matrix. 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, and there could be other parameters contained in $\boldsymbol{\theta}$, 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 provide just the essential model components that provide most inferences for ecological data. The model component to be estimated or predicted from (eqn 12) is identified in Table 1. Note that a full 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 are $\mathbf{y}$. The term likelihood is used when the joint distribution is considered a function of all unknowns, given the observed data, which we denote $L(\mathbf{y}|.)$.

Fitting Methods for Autoregressive Models

Maximum likelihood estimation is one of the most popular estimation methods (Cressie, 1993). Earlier, when computers were less powerful, methods were devised to trade efficiency 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., 2004). For spatial models, the main computational burden in geostatistical models is inversion of the covariance matrix; for CAR and SAR models, the inverse is what we actually model. Thus, only the determinant of the covariance matrix needs computing, and fast sparse matrix methods can be used (Pace and Barry, 1997a,b). If matrices 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.

The 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 (2013) for

book-length treatment. These models can be treated as hierarchical models (Cressie et al., 2009), where the data are assumed to be a count model, 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). A similar hierarchical framework has been developed as a generalized linear model for occupancy, which is a binary model, but then the logit 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., 2013; 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 depend in part on the sparsity of the CAR covariance matrix, is integrated nested Laplace approximation (INLA, Rue et al., 2009), which has been used in generalized linear models for ecological data (e.g., Haas et al., 2011; Aarts et al., 2013) and spatial point patterns (Illian et al., 2013), among others. INLA provides fast computing for approximate Bayesian inference on the marginal distributions of latent variables.

EXAMPLE

We used trends in harbor seals (*Phoca vitulina*) for an example to illustrate the models and approaches for inference described earlier. The study area is shown in Fig. 1 and contains 463 polygons 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. Going forward, we will treat the estimated trends, on the log scale, as raw data, and ignore the estimated variances. These data were chosen to be illustrative because we expect the trends to show geographic patterns (more so than abundance which varied widely in polygons), and perhaps differences in mean rates and connectivity

due to stock structure. The data are also continuous in value, so we model the trends with normal distributions to keep the modeling simpler and the results more evident. A map of the raw trend values is given in Fig. 2, 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, 2014). 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. 3a with a close-up of part of the study area given in Fig. 3b. If $\mathbf{N}_1$ is a matrix of binary values, where a 1 indicates two sites are first-order neighbors, and a 0 otherwise, then second-order neighbors, to include neighbors of first-order neigbhors, were easily obtained in the matrix $\mathbf{N}_2 = \mathcal{I}(\mathbf{N}^2)$, where $\mathcal{I}(\cdot)$ 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. 3c. 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. 3d.

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

$$\mathbf{W} = \mathbf{N}_i \odot \mathbf{F}(\boldsymbol{\theta}) = \mathbf{N}_i \odot \exp(-\mathbf{S}/\theta_1) \odot \exp(-\mathbf{D}/\theta_2). \qquad \text{eqn 13}$$

We use (eqn 13) in (eqn 2) and (eqn 3), 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 (eqn 1), $\text{var}(\mathbf{y}) = \boldsymbol{\Sigma} + \sigma_\varepsilon^2\mathbf{I}$ would also be possible; for example, for a first-

order CAR model, $\text{var}(\mathbf{y}) = \sigma_Z^2(\mathbf{I} - \rho\mathbf{W})^{-1} + \sigma_\varepsilon^2\mathbf{I}$. However, when $\rho = 0$, then σ_Z^2 and σ_ε^2 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 $\sigma_\varepsilon^2\mathbf{I}$. We tried some models with the additional component $\sigma_\varepsilon^2\mathbf{I}$, but σ_ε^2 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 interests to ecologists, and by its very definition it is about relationships *between* two items. 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 so 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) or whether the neighborhood definitions are enough (θ_1 very large).

We fit model (eqn 1) 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, and details are given in Appendix C. The resulting maximized value of 2*log-likelihood is given in Fig. 4. 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 for 2*log-likelihood, so number of model parameters are shown along the x-axis, and dashed lines at increments of two help evaluate models in Fig. 4. 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. 4). 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. Also, for row-standardized CAR models, fit worsens going from first-order to second-order neighborhoods, but then gets better when going to fourth-order. Also, perhaps not surprising, using distance between centroids had little effect until fourth-order neighborhoods were used. By an AIC criteria, model XC4RD would be the best model. For model XC4RDS, the parameter θ_1 was very large, making $\exp(-\mathbf{S}/\theta_1)$ 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 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 is a standard normal distribution under the null hypthesis 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 probabililty 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 $\alpha = 0.05$. 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. Notice that for both, the standard errors are larger than for the independence model XU, leading to greater uncertainty about the fixed effects estimates. Also, the Bayesian standard errors 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. 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.

For objective 3 from Table 1, consider the curves in Fig. 5. We tried 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. 5. 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 likelhood value would not be rejected at the α level of significance; i.e., the MLE value minus a chi-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. 5 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% credibility interval ranging from 0.315 to 0.933. The Bayesian estimate used noninformative 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. 6 shows likelihood profiles for the other parameters in the covariance matrix. For the best model, XC4RD, the solid line in Fig. 6a 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 2log-likelihood (413.447) minus a chi-squared value at $\alpha = 0.05$ on one degree of freedom (3.841) to help visualize a confidence interval for θ_2; i.e., the profile likelihood confidence interval given by all values of the solid line that are above the dashed line. The loglikelihood 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. 6b. 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 D. 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. 2) 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. 7a 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 weighted, Shepard, 1968) is that statistical predictions provide standard errors for each prediction (Fig. 7B). 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 (eqn 7) (after adjusting for estimated covariate effects) is used rather than the observed value at each location. Once 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. 7c; note then that the predictions are equivalent to Fig. 7a at the unsampled locations. Two extremes in smoothing approaches are 1) kriging as an exact predictor, that is, it leaves the data unchanged (Figs. 7a), and 2) removing observed data to replace them with conditional expectations based on neighbors (Fig. 7c). 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 as a hierarchical model (eqn 12), often in a Bayesian context, predictions will 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. 7d). 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. 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 edges) for model XC4R (Fig. 8a), 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. 8a). Correlation also decreases with neighbor order, although not as dramatically as one might expect (Fig. 8b), 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 8c,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 for model XC4R was 0.604 (Fig. 5) for the inverse covariance matrix, but for the covariance matrix, correlations are much lower (Fig. 8b-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. 8), and we highly recommend it for model diagnostics and a better understanding of the fitted model.

DISCUSSION AND CONCLUSIONS

We have given six objectives used for spatial autoregressive models, some of which are common, and others less so. We argued that the development of the weights in modeling the covariance matrix for CAR and SAR models is less intuitive than choosing a covariance model in geostatistics, but some intuition can be gained by considering the relationship to partial correlations. We also

argued that row standardization is generally a good idea after choosing initial neighborhoods and weights. CAR models will generally be 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 inverse covariance matrix. The IAR model is a special case of the CAR model that uses row standardization and where the autocorrelation parameter is fixed at one, which leads to an improper covariance matrix; however, much like a similar AR1 model, or Brownian motion, these are still useful models. We used an example data set by fitting a variety of CAR/SAR models using MLE and MCMC methods to illustrate all six objectives outlined in the Introduction. Here, we give further discussion on 5 additional 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. The IAR 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 (eqn 3) needed to be 0.999972 in order to have a marginal correlation of near 0.75 (indeed, compare the estimate of $\hat{\rho} = 0.604$ in our example yielding the correlations seen in Fig. 8). 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 complained 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 (eqn 12). 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 is no extra variance parameter $\boldsymbol{\nu}$, 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 $\mathbf{\Sigma}$, essentially controlled by ρ, and the free variance parameter in $\boldsymbol{\nu}$.

The results for Figs. 5 and 6 have confidence intervals that are quite wide. We believe that, 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. In other words, it is possible to make inference on connectivity types of parameters in the covariance matrix, but, secondly, they may be difficult to estimate with much precision if the data are measured only on the nodes. In our 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 hard to estimate well without a lot of data.

From an ecological viewpoint, why do spatial smoothing? In geology, and mining in particular, modelers were often adamant that no smoothing occur (honoring the data). Notice from Fig. 2 that the largest value is 0.835 from the legend on the right. Recalling that these are trends, on the log scale, that means that 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. 7c 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. 7c, back on the exponential scale, was 1.083, or about 8% growth per year, and the largest value in Fig. 7d, 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. 7c,d also clarifies the regional trends, which are difficult to see among the noise in Fig. 2 or by simply filling in the missing sites with predictions (Fig. 7a) . For these reasons, smoothing is very popular in disease-mapping applications, but they should be equally attractive for a wide variety of ecological applications. In particular, the XI4RU model (Fig. 7d) 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 chosing between CAR and IAR.

One practical consideration that is rarely discussed is the case of isolated sites (those with no neighbors) when constructing the neighborhood matrix. Having a row of zeros in $\mathbf{B}$ in (eqn 2), or in $\mathbf{C}$ in (eqn 3), will cause problems. It is even easier to see that we cannot divide by zero in (eqn 9), 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, e.g., σ_Z^2 in (eqn 2) and (eqn 3). As seen in (eqn 9), 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 software. 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 some 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 increasing throughout science, including ecology (Borrett et al., 2014). Looking again at Fig. 3, 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 at 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

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 NMML. The findings and conclusions in the paper are those of the author(s) and 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 US 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.

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.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL, USA: Chapman and Hall/CRC.

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.

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.

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.

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.

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

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

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

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

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

740 Cressie, N. (1990), "The Origins of Kriging," *Mathematical Geology*, 22, 239–252.

741 Cressie, N., Calder, K. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009), "Accounting
742 for Uncertainty in Ecological Analysis: The Strengths and Limitations of Hierarchical Statistical
743 Modeling," *Ecological Applications*, 19, 553–570.

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

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

748 Curriero, F. C. (2006), "On the use of non-Euclidean distance measures in geostatistics," *Mathe-*
749 *matical Geology*, 38, 907–926.

750 Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies,
751 R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reinek-
752 ing, B., Schröder, B., Schurr, F. M., and Wilson, R. (2007), "Methods to account for spatial
753 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.

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.

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., Schmidt, A. M., Wu, S., Silander, J. A., Latimer, A., and Rebelo, A. G. (2005), "Modelling species diversity through species level hierarchical modelling," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 1–20.

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.

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.

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

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

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

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

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

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

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

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

797 Latimer, A. M., Wu, S., Gelfand, A. E., and Silander Jr, J. A. (2006), "Building statistical models
798 to analyze species distributions," *Ecological Applications*, 16, 33–50.

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

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

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.

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.

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.

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.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

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.

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

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

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

833 Ver Hoef, J. M. and Jansen, J. K. (2007), "Space-time Zero-inflated Count Models of Harbor Seals,"
834 *EnvironMetrics*, 18, 697–712.

835 Ver Hoef, J. M. and Peterson, E. (2010), "A Moving Average Approach for Spatial Statistical
836 Models of Stream Networks (with discussion)," *Journal of the American Statistical Association*,
837 105, 6–18.

838 Ver Hoef, J. M., Peterson, E. E., and Theobald, D. (2006), "Spatial Statistical Models That Use
839 Flow and Stream Distance," *Environmental and Ecological Statistics*, 13, 449–464.

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

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

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

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

TABLES

Table 1: Common objectives when using spatial autoregressive models. Notation for the model components comes from (eqn 12).

Objective	Description	Example Literature	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.	Hooten and Hobbs (2015)	$L(\mathbf{y}\|.)$
2. Regression	The goal is to estimate the spatial regression coefficients, which quantify how an explanatory variable "affects" the response variable.	Clayton and Kaldor (1987)	$\boldsymbol{\beta}$
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.	Gardner et al. (2010)	ρ
4. Neighborhood Structure	The goal is to estimate covariate effects on neighborhood structure. Although rarely used, covariates can be included in the precision matrix to see how they affect neighborhood structure; e.g., causing more or less correlation.	Hanks and Hooten (2013)	$\boldsymbol{\theta}$
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.	Gardner et al. (2010)	$\mathbf{y}_u$ 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.	Clayton and Kaldor (1987)	$g(\boldsymbol{\mu})$

Table 2: A variety of candidate models used to explore spatial autoregressive models for the example data set. For fixed effects, the **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. A C indicates a CAR model, an S a SAR model, and an I an IAR model. A 1 indicates a first-order neighborhood, a 2 a second-order neighborhood, and a 4 a fourth-order neighborhood. An R indicates row-standardization. A D indicates inclusion of Euclidean distance within neighborhoods, and S a cross stock connectivity matrix. A U at the end indicates inclusion of an additive random effect of uncorrelated variables.

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

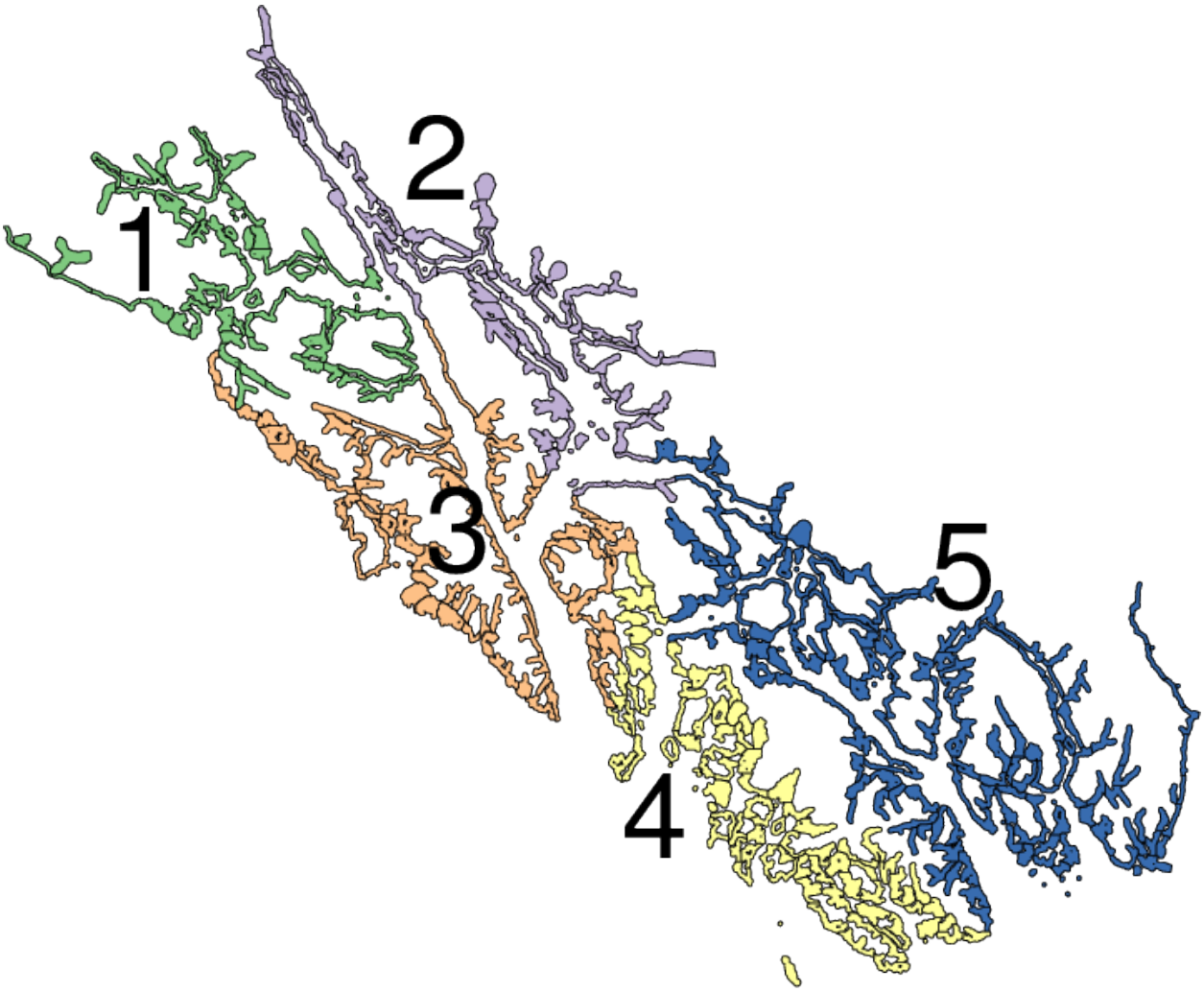

Figure 1: Study area in Southeast Alaska. Survey polygons were established around the coast of the mainland and all islands, which were surveyed for harbor seals. The study area comprises 5 stocks, each with their own color, and are numberedfor further reference.

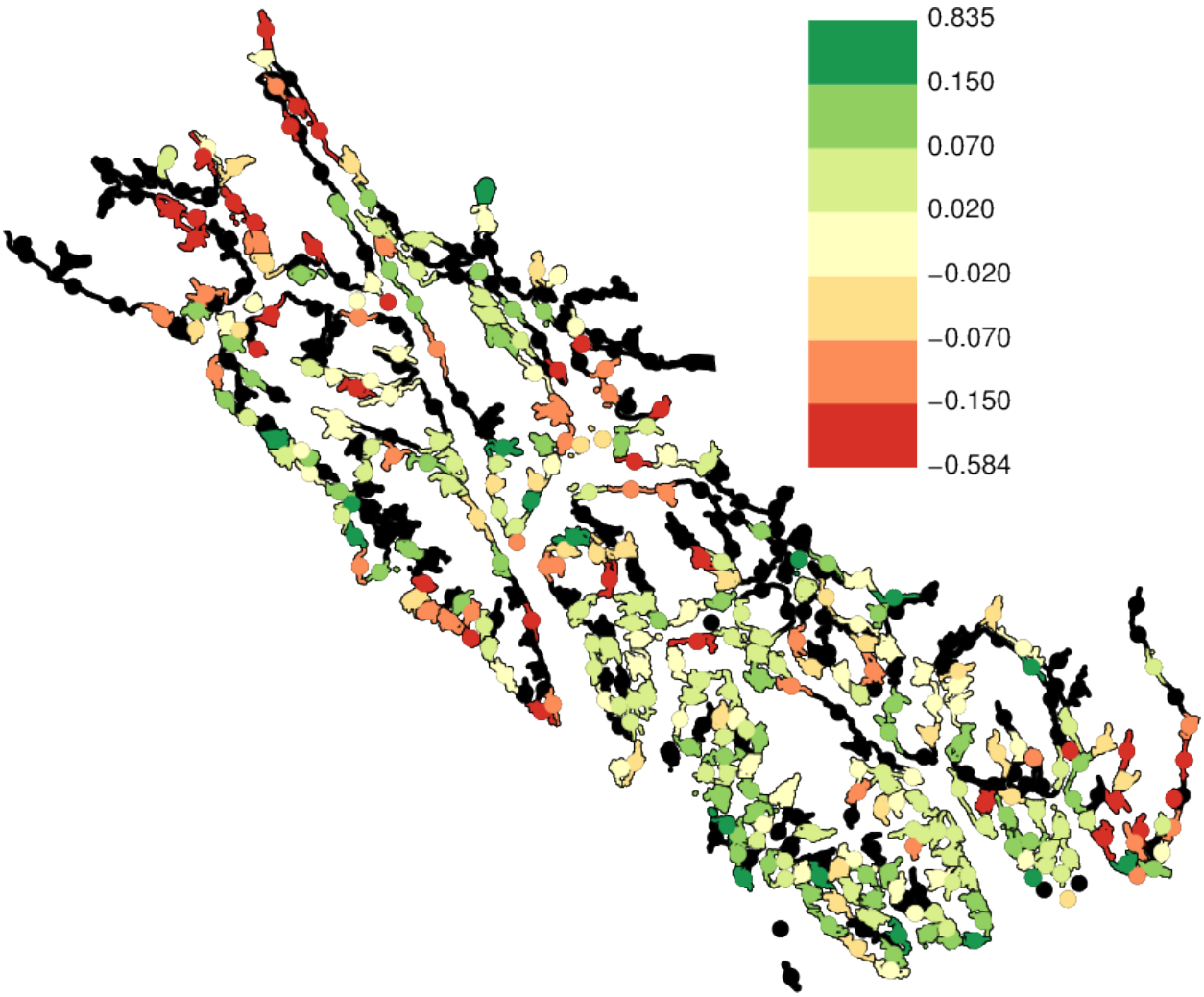

Figure 2: Map of the raw data, where polygons are colored by their trend values. The black polygons have missing data. Because some polygons were small and it was difficult to see colors in them, all polygons were also overwritten by a circle of the same color. The trend values were categorized by colors, with increasing trends in green, and decreasing trends in red, with the cutoff values given by the color ramp.

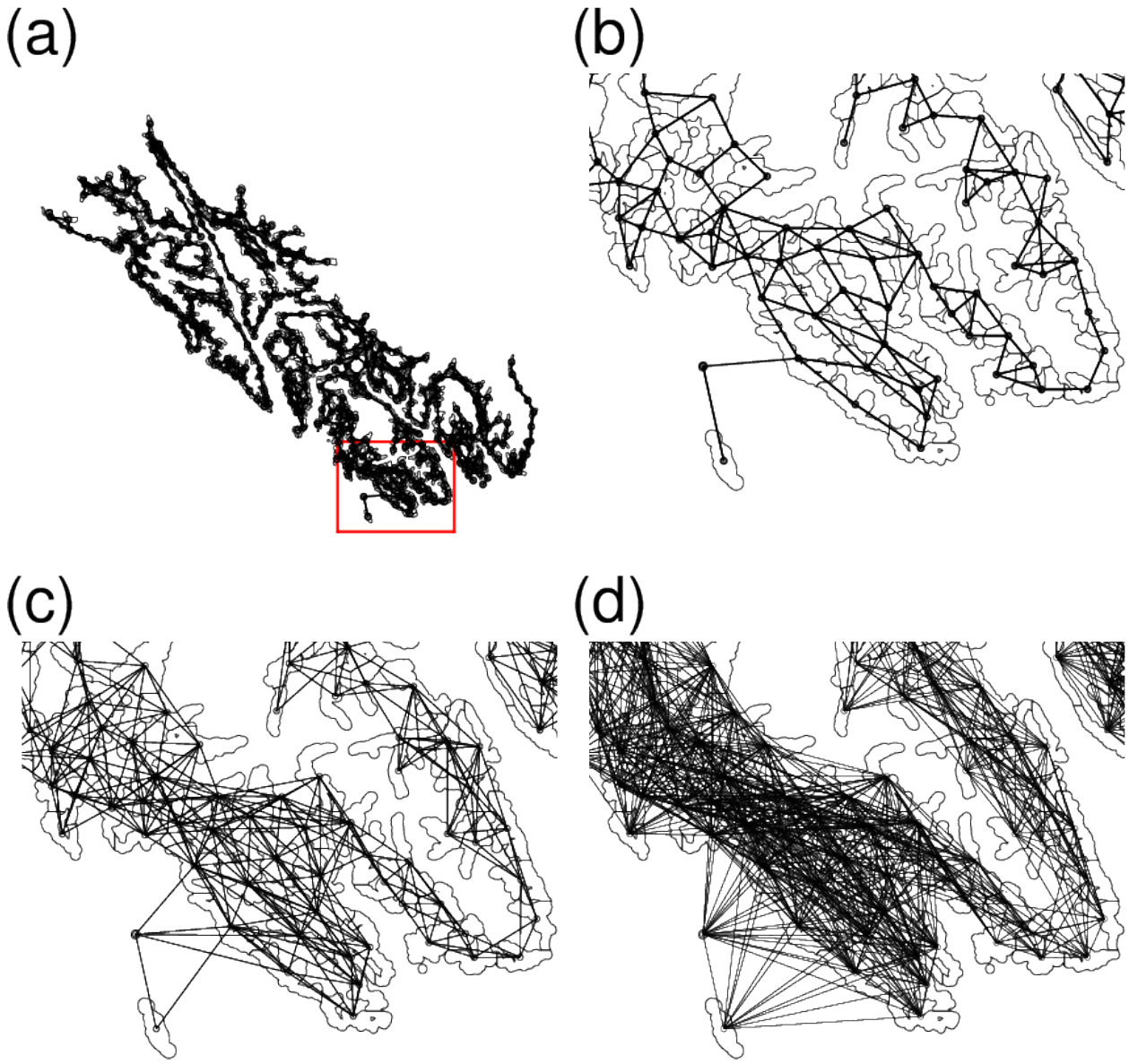

Figure 3: First, second, and fourth-order neighbor definitions for the survey polygons. (a) First-order neighbors for all polygons. The red rectangle is the area for a closer view in the following subfigures: (b) first-order neighbors; (c) second-order neighbors; and (d) fourth-order neighbors.

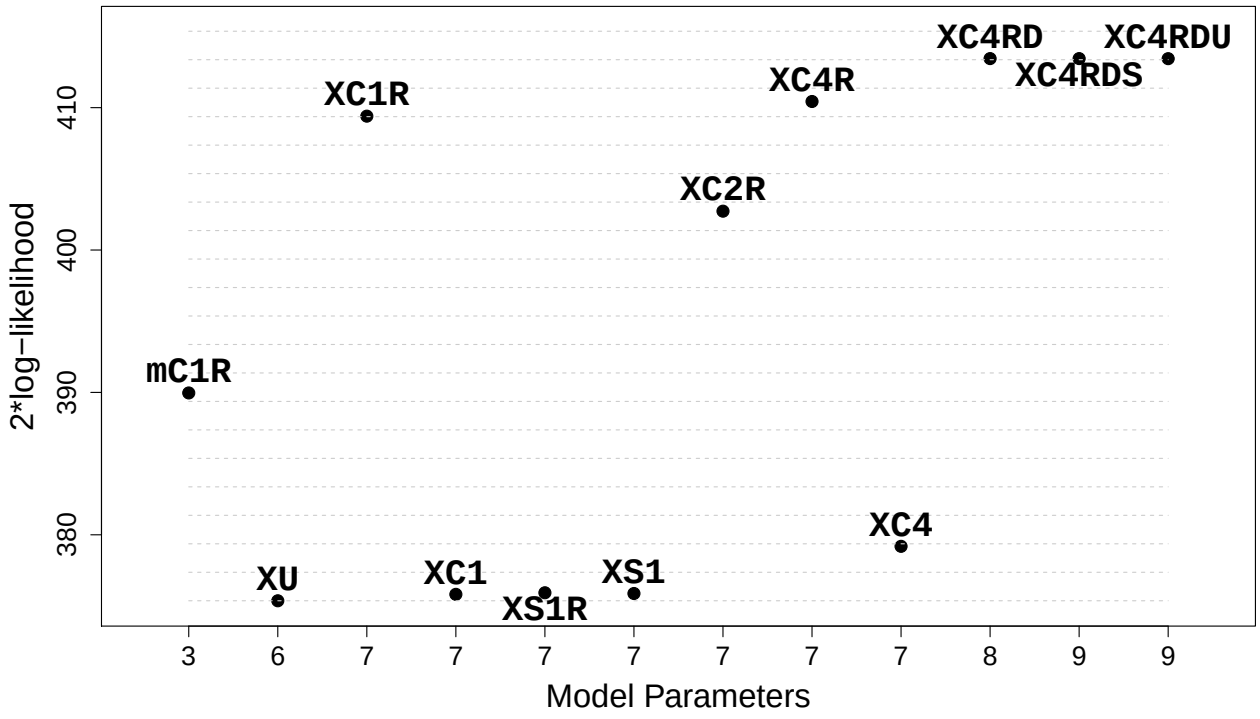

Figure 4: Two times the log-likehood for the optimized (maximized) fit for the models given in Table 2. Model mU had a much lower value (350.2) and is not shown. Starting with model XU, the dashed grey lines show increments of 2, which helps evaluate the relative importance of models by either an AIC or a likelihood-ratio test criteria.

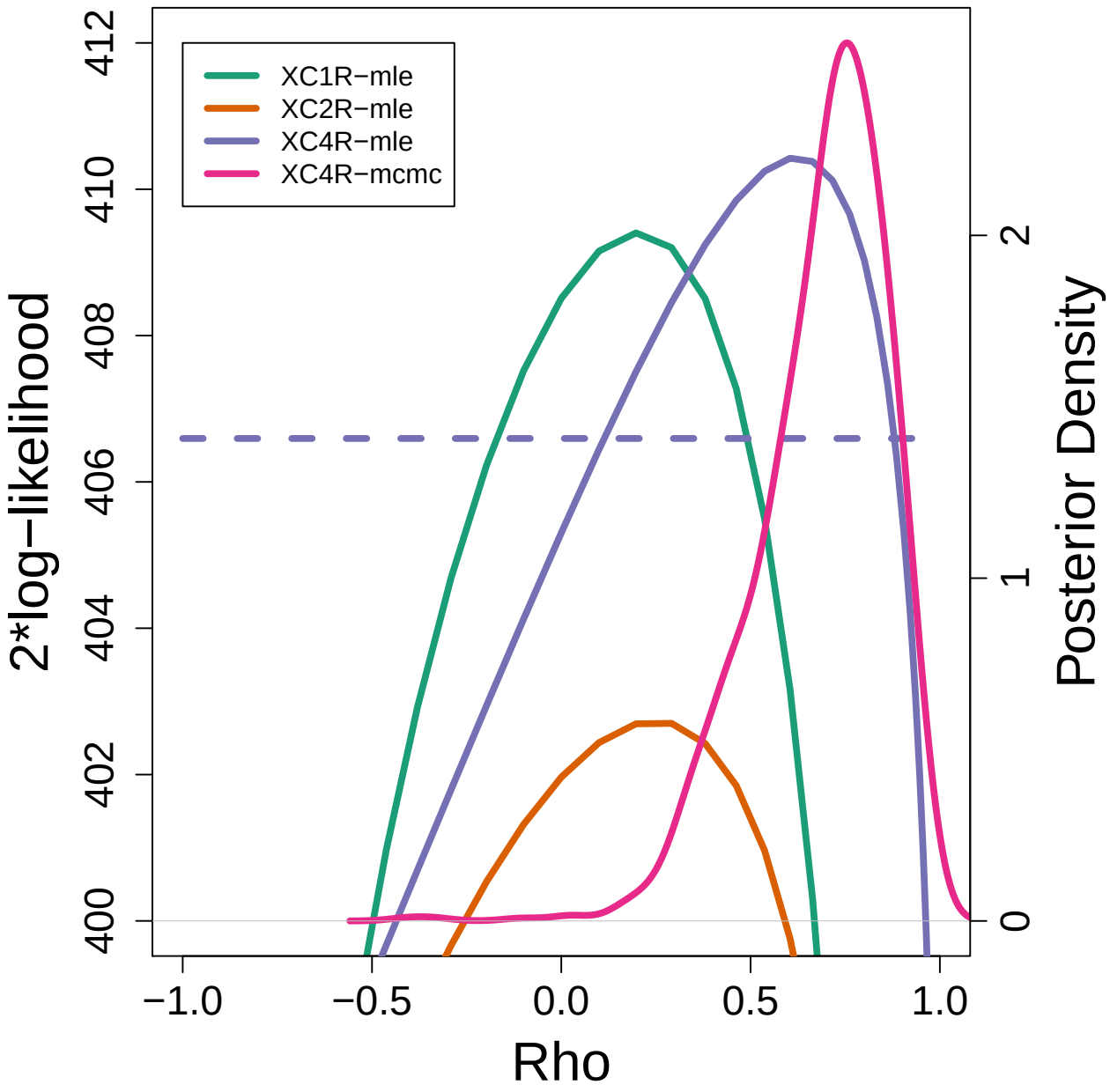

Figure 5: The different colored solid lines show 2Log-likelihood profiles of ρ for three different models, listed in the legend. If the model is followed by -mle, then the maximum of the profile provides the maximum likelihood estimate, and the 2log-likelihood is given by the left y-axis, while if it 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 dashed line is the maximum value for XC4R minus 3.841, the 0.05 α-level value of a chi-squared distribution on one degree of freedom.

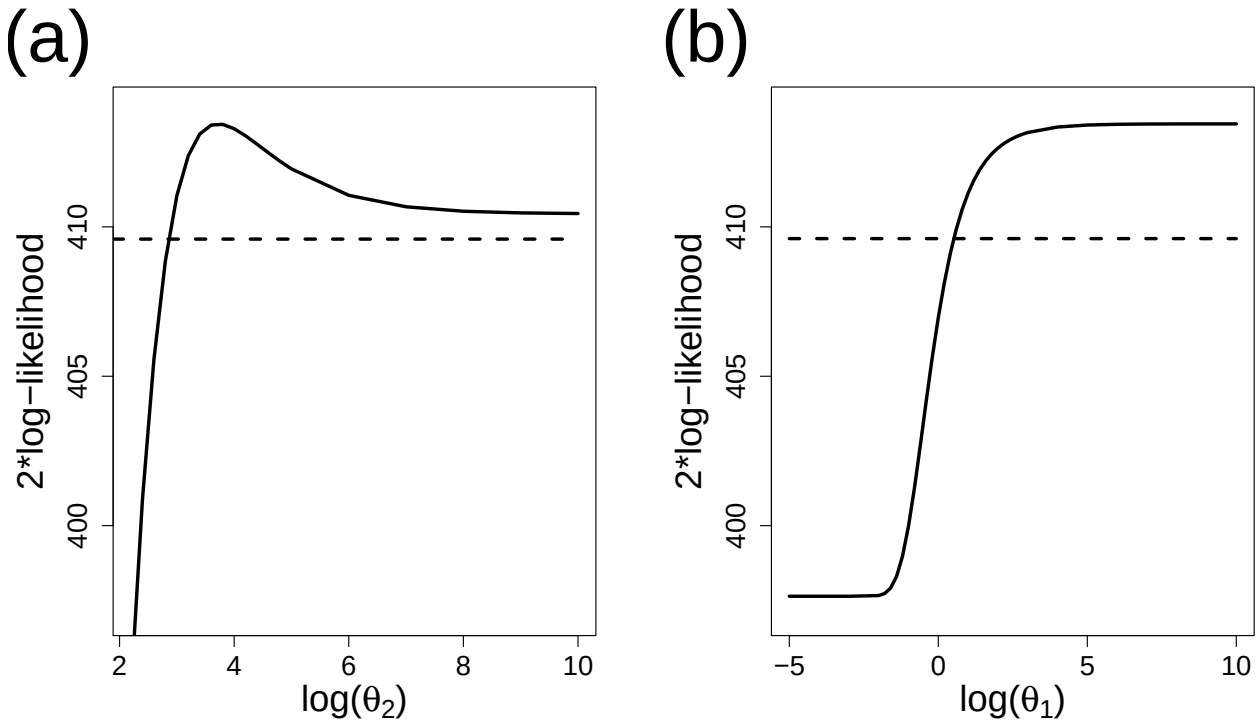

Figure 6: A) The solid line is the 2Log-likelihood profile of θ_2 for model XC4RD. B) The solid line is the 2Log-likelihood profile of θ_1 for model XC4RDS. For each figure, the horizontal dashed line is the maximum value for the model minus 3.841, the 0.05 α-level value of a chi-squared distribution on one degree of freedom.

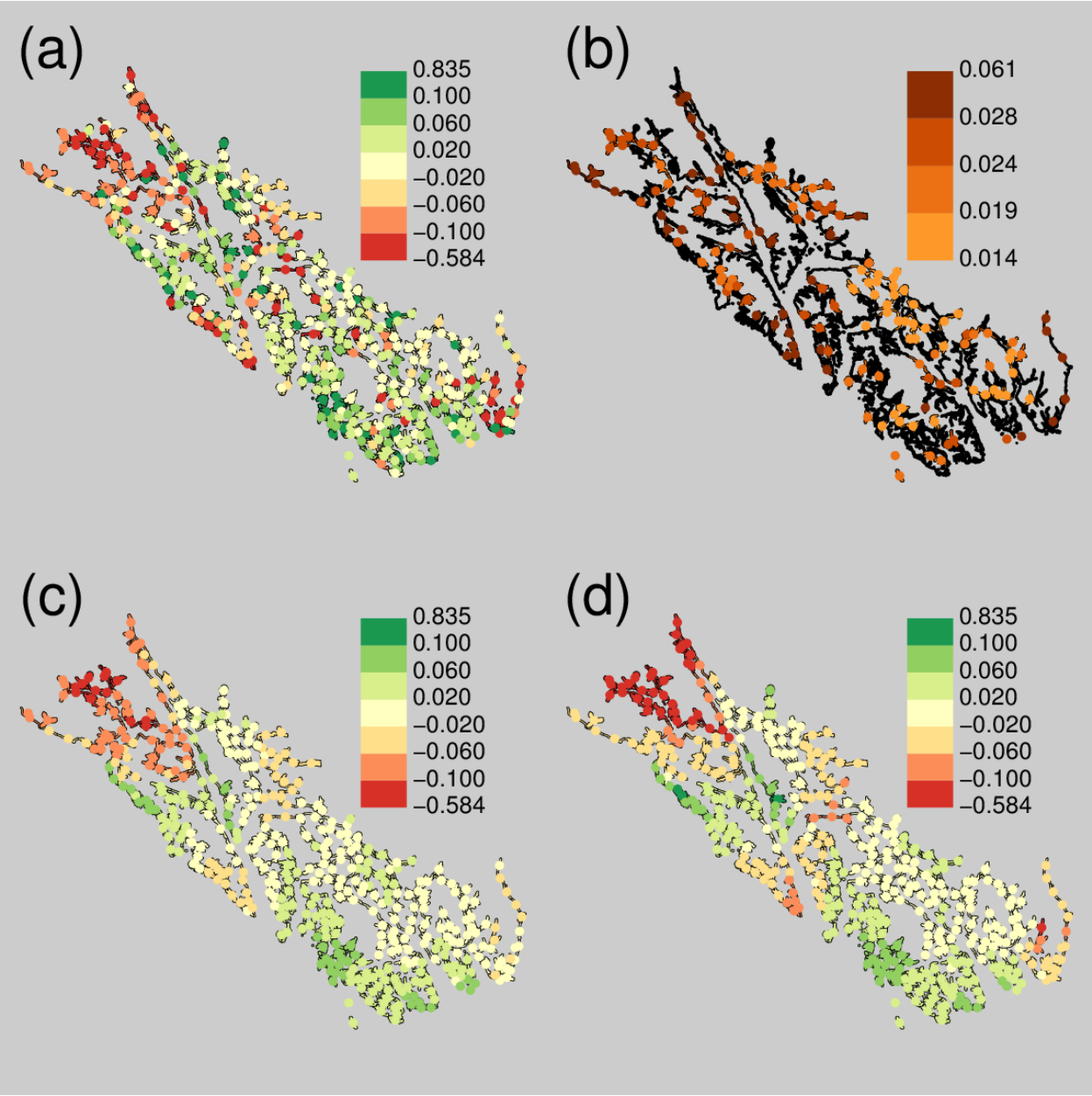

Figure 7: Predictions and smoothing for the stock trend data. (a) Predictions, using universal kriging from the XC4R model, at unsampled locations have been added to the raw observed data from sampled locations. (b) Prediction standard errors for unsampled locations using universal kriging from the XC4R. (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.

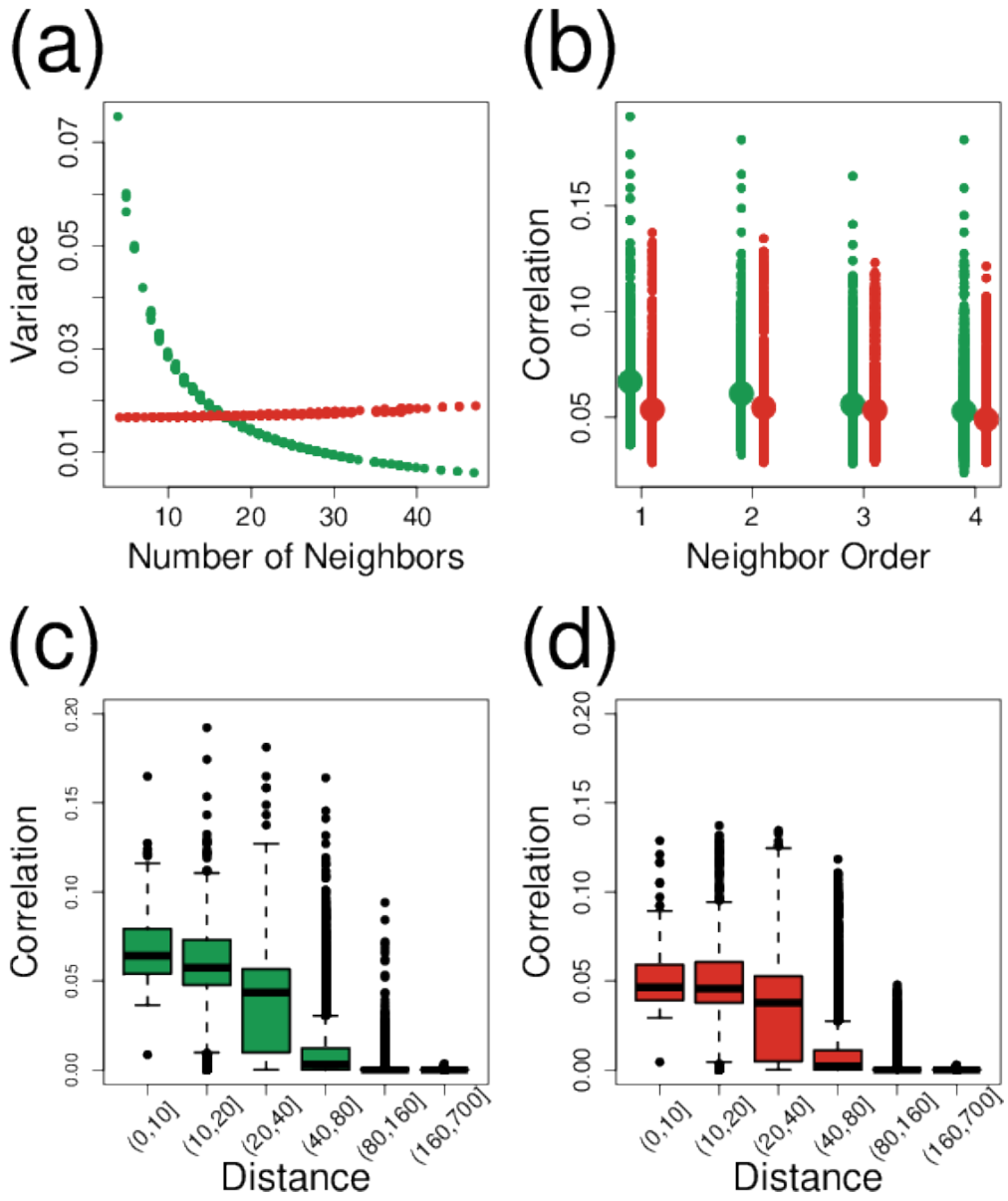

Figure 8: For all subfigures, red is used for model XC4 and green is used for the same model, XC4R, but with row-standardization. A) Marginal variances of the multivariate covariance matrix (diagonal elements of $\mathbf{\Sigma}$) as a function of the numbers of neighbors. B) All pairwise correlations as a function of the neighborhood order between sites. 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: Errors in the Literature

The fact that CAR and SAR models are developed for 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) state "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 8c,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 $\text{corr}(Z_i, Z_{i+t}) = \phi^t$ for all t. In fact, if we restrict $0 < \phi < 1$, then this can be reparameterized as $\text{corr}(Z_i, Z_{i+t}) = \exp(-t(-\log(\phi)))$, 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 autocorrelatin 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) characterize 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. While we believe their implementation may have been correct, and there are excellent and important results in Keitt et al. (2002), their construction leads to a SAR covariance matrix of $\sigma^2(\mathbf{I} - \rho\mathbf{C})^{-1}(\mathbf{I} - \rho\mathbf{C})^{-1}$ if $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ 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 (eqn 8), or alternatively that $\boldsymbol{\Sigma}^{-1} = (\mathbf{M}^{-1} - \mathbf{C})/\sigma^2$, 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 claim 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 Appendix B that, despite literature to the contrary, CAR models and SAR models can be written equivalently, and we give further details.

2 APPENDIX B: Equivalence of CAR and SAR Models

To establish when CAR models can be written as SAR, and vice versa, we need the following equality,

$$(\mathbf{I}-\mathbf{C})^{-1}\mathbf{M} = (\mathbf{I}-\mathbf{B})^{-1}\mathbf{\Lambda}(\mathbf{I}-\mathbf{B}')^{-1}, \tag{B.1}$$

satisfying, for the CAR covariance matrix on the left-hand side, 1) $(\mathbf{I}-\mathbf{C})^{-1}$ exists, 2) $c_{ii} = 0 \,\forall\, i$, and 3) $c_{ij}/m_{ii} = c_{ji}/m_{jj} \;\forall\; i, j$; and for the SAR covariance matrix on the right-hand side, 4) $(\mathbf{I}-\mathbf{B})^{-1}$ exists, and 5) $b_{ii} = 0 \;\forall\; i$. Notice that we write the SAR covariance matrix as $(\mathbf{I}-\mathbf{B})^{-1}\mathbf{\Lambda}(\mathbf{I}-\mathbf{B}')^{-1}$, where $\mathbf{\Lambda}$ is a diagonal matrix, in (B.1), following Cressie (1993, p. 409), which is a little more general than the SAR covariance matrix given in (eqn 2). Ultimately, this demonstration relies on establishing that any zero-mean Gaussian distribution on a finite set of points, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$ can be written as either the left-hand side, or the right-hand side of (B.1). Note, we make use of the following result: if $\mathbf{D}$ is diagonal, and $\mathbf{Q}$ has zero on the diagonals, then both $\mathbf{DQ}$ and $\mathbf{QD}$ have zeros on the diagonal.

First consider obtaining the left-hand side of (B.1); this result is also given by Cressie (1993, p. 434). Write $\mathbf{\Sigma}^{-1} = \mathbf{D} - \mathbf{Q}$, where $\mathbf{D}$ is diagonal and $\mathbf{Q}$ has zeros on the diagonal. Then factor out $\mathbf{D}$ so that $\mathbf{\Sigma}^{-1} = \mathbf{D}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{Q})$, and now let $\mathbf{M} = \mathbf{D}^{-1}$ and $\mathbf{C} = \mathbf{D}^{-1}\mathbf{Q}$, then $(\mathbf{I}-\mathbf{C})^{-1}$ exists because $\mathbf{\Sigma}^{-1}$ and $\mathbf{D}^{-1}$ exist, $\mathbf{C}$ will have zeros on the diagonals, and condition 3) is satisfied by construction (because it is the requirement for symmetry). Thus, $\mathbf{\Sigma}^{-1}$ can be expressed as $\mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$ and $\mathbf{\Sigma} = (\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}$ satisfying conditions 1) to 3).

Next, consider the right-hand side of (B.1). Write $\mathbf{\Sigma}^{-1} = \mathbf{L}\mathbf{L}'$. An important note is that this is not unique. A Cholesky decomposition satisfies this, where $\mathbf{L}$ is lower triangular, or a singular value decomposition could be used, where square roots of the eigenvalues are multiplied in the matrix of eigenvectos. In any case, let $\mathbf{L}\mathbf{L}' = (\mathbf{G}-\mathbf{P})(\mathbf{G}'-\mathbf{P}')$, where $\mathbf{G}$ is diagonal and $\mathbf{P}$ has zero diagonals. Then factor out $\mathbf{G}$ to obtain $\mathbf{L}\mathbf{L}' = (\mathbf{I}-\mathbf{P}\mathbf{G}^{-1})\mathbf{G}\mathbf{G}(\mathbf{I}-\mathbf{G}^{-1}\mathbf{P}')$, and now let $\mathbf{\Lambda}^{-1} = \mathbf{G}\mathbf{G}$ and $\mathbf{B}' = \mathbf{P}\mathbf{G}^{-1}$. Notice that $(\mathbf{P}\mathbf{G}^{-1})' = \mathbf{G}^{-1}\mathbf{P}'$ because $\mathbf{G}^{-1}$ is diagonal. Moreover, $(\mathbf{I}-\mathbf{B})^{-1}$ exists because $\mathbf{L}^{-1}$ and $\mathbf{G}^{-1}$ exist and $\mathbf{B}$ will have zeros on the diagonals. Thus, $\mathbf{\Sigma}^{-1}$ can be expressed as $(\mathbf{I}-\mathbf{B}')^{-1}\mathbf{\Lambda}^{-1}(\mathbf{I}-\mathbf{B})$ and $\mathbf{\Sigma} = (\mathbf{I}-\mathbf{B})^{-1}\mathbf{\Lambda}^{-1}(\mathbf{I}-\mathbf{B}')^{-1}$ satisfying conditions

4) and 5). Note that in order to write it as $(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B}')^{-1}$ we need to find $\mathbf{L}$ with ones on the diagonal.

Hence, we have demonstrated that any zero-mean Gaussian distribution on a finite set of points, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$ can be written as either the left-hand side, or the right-hand side of (B.1), with the important difference that a CAR model is uniquely determined from $\mathbf{\Sigma}$ but a SAR model is not so uniquely determined. That a CAR is unique is easy to see from the algebra used to derive it. To see more fully why a SAR model is not uniquely determined, notice that $\mathbf{\Sigma}^{-1} = \mathbf{L}\mathbf{L}' = \mathbf{L}(\mathbf{A}'\mathbf{A})\mathbf{L}' = (\mathbf{L}\mathbf{A}')(\mathbf{A}\mathbf{L}') = \mathbf{L}_*\mathbf{L}_*'$, where the columns of $\mathbf{A}$ are orthonormal. A SAR model can be developed as readily for $\mathbf{L}_*$ as for $\mathbf{L}$. In fact, if we think of $\mathbf{A}$ as coordinate vectors, any rotation of the matrix $\mathbf{A}$ will create yet another SAR model, so there are an infinite number of them.

With this appendix, we also wish to correct an error that has been perpetuating in the literature, beginning with Haining (1990, p. 89), and we also found it in Schabenberger and Gotway (2005) and Dormann et al. (2007). The authors state that, suppose in (B.1) that $\mathbf{M} = \mathbf{I}$ and $\mathbf{\Lambda} = \mathbf{I}$, so that $\mathbf{C}$ is symmetric. That implies that $(\mathbf{I}-\mathbf{C})^{-1} = [(\mathbf{I}-\mathbf{B})(\mathbf{I}-\mathbf{B}')]^{-1} = (\mathbf{I}-\mathbf{B}-\mathbf{B}'+\mathbf{B}\mathbf{B}')^{-1}$. So, while $\mathbf{C} = \mathbf{B}+\mathbf{B}'-\mathbf{B}\mathbf{B}'$ is sufficient for equality, it is not necessary, and it lacks a critical component; that $\mathbf{B}+\mathbf{B}-\mathbf{B}\mathbf{B}'$ must have zeros on the diagonal for $\mathbf{C}$ to be a CAR model. The proper way to proceed was outlined above, by letting $\mathbf{C} = \mathbf{D}^{-1}\mathbf{Q}$ in the equality

$$(\mathbf{I}-\mathbf{D}^{-1}\mathbf{Q})^{-1}\mathbf{D}^{-1} = (\mathbf{I}-\mathbf{B}-\mathbf{B}'+\mathbf{B}\mathbf{B}')^{-1},$$

where the left-hand side is uniquely determined from the right-hand size, and $\mathbf{C}$ satisfies the requirement for a CAR model. As we demonstrated in the previous paragraph, it is not possible to go uniquely from the left-hand side to the right-hand side without additional constraints.

3 APPENDIX C: 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 variable, $\mathbf{X}$ 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, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, 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 loglikelihood 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,

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

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 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

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

Substituting (C.2) back into (C.1) yields the -2loglikelihood as a function of $\boldsymbol{\rho}$ only,

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

Equation (C.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 $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}_{\hat{\boldsymbol{\rho}}}^{-1} \mathbf{X})^{-1} \mathbf{X} \mathbf{V}_{\hat{\boldsymbol{\rho}}}^{-1} \mathbf{y}$.

We developed the inverse covariance matrix $\boldsymbol{\Sigma}_A^{-1} = \text{diag}(\mathbf{W1}) - \rho \mathbf{W})$, 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, (C.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 $\boldsymbol{\Sigma}_A = (\boldsymbol{\Sigma}_A^{-1})^{-1}$, and then obtain $\mathbf{V}_{\boldsymbol{\rho}} = \boldsymbol{\Sigma}[\mathbf{i}, \mathbf{i}]$, where $\mathbf{i}$ is a vector of indicators that subsets the rows and columns of $\boldsymbol{\Sigma}$ to only those for sampled locations. Then, a third step is a second inverse to find $\mathbf{V}_{\boldsymbol{\rho}}^{-1}$. This is computationally expensive. A faster way uses results from partitioned matrices and Schur complements. In general, let the square matrix $\boldsymbol{\Sigma}$ with dimensions $(m+n) \times (n+m)$ be partitioned into block submatrices,

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

with dimensions given below each matrix. Assume $\mathbf{A}$ and $\mathbf{D}$ are nonsingular. Then define the matrix function $\mathbf{S}(\boldsymbol{\Sigma}, \mathbf{A}) = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ as the Schur complement of $\boldsymbol{\Sigma}$ 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, which has dimension less than $\mathbf{\Sigma}$, and only one inverse is required. So, if we let $\mathbf{A}$ correspond to the rows and columns of the observed locations, and $\mathbf{D}$ correspond to the 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 $n \times n$, the number of missing data.

4 APPENDIX D: Prediction and Smoothing

Here, we give the formulas used in creating Fig. 7. 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 C. 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. 7a contains the observed values plus the predicted values at nodes with missing values, while Fig. 7c 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,

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

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

For the IAR model smoothing in Fig. 7d, 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. 7d.