Ecological Applications, 16(1), 2006, pp. 87–98
© 2006 by the Ecological Society of America

MODEL SELECTION FOR GEOSTATISTICAL MODELS

Jennifer A. Hoeting,[1,3] Richard A. Davis,[1] Andrew A. Merton,[1] and Sandra E. Thompson[2]

[1]*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523-1877 USA*
[2]*P.O. Box 999, K5–12, Richland, Washington 99352 USA*

Abstract. We consider the problem of model selection for geospatial data. Spatial correlation is often ignored in the selection of explanatory variables, and this can influence model selection results. For example, the importance of particular explanatory variables may not be apparent when spatial correlation is ignored. To address this problem, we consider the Akaike Information Criterion (AIC) as applied to a geostatistical model. We offer a heuristic derivation of the AIC in this context and provide simulation results that show that using AIC for a geostatistical model is superior to the often-used traditional approach of ignoring spatial correlation in the selection of explanatory variables. These ideas are further demonstrated via a model for lizard abundance. We also apply the principle of minimum description length (MDL) to variable selection for the geostatistical model. The effect of sampling design on the selection of explanatory covariates is also explored. R software to implement the geostatistical model selection methods described in this paper is available in the Supplement.

Key words: *AIC; geospatial data; kriging; Matern autocorrelation function; MDL; orange-throated whiptail lizard abundance.*

Introduction

Ecologists and scientists in other fields typically consider a number of plausible models in statistical applications. Formal consideration of model selection in ecological applications has dramatically increased in recent years, perhaps in part due to the publication of the book by Burnham and Anderson (1998, 2002). Concurrently, the wide availability of inexpensive global positioning systems and other advances in technology have allowed for the collection of vast quantities of data with georeferenced sample locations. As a result, models for spatially correlated data are becoming increasingly important. We consider these two problems of spatial modeling and model selection together. The importance of accounting for spatial correlation has been discussed in other contexts (Cressie 1993), but the effect of spatial correlation on model selection has not been fully explored.

Our general philosophy for choosing a model is that we would like to incorporate information that we believe influences the response variable while acknowledging that we do not know everything associated with the response. These unknowns could be quantities that we did not (or could not) measure, complex variable interactions, heterogeneity, etc. Thus, an error process is often included in the model that ''accounts'' for these unknowns. For example, we may suspect that the abundance of a certain species is dependent on the availability of a certain type of vegetation and the predator-to-prey ratio. But we must acknowledge that other variables are likely to play an important role, such as the abundance of fresh water or the prevalence of a certain disease. Thus, the model that we construct must account for these unknown influences. This is the main role of any error term in any modeling exercise. The problem becomes more complicated when we consider that there may be competing models each using a different subset of known variables. For example, perhaps there are two types of vegetation that the species will eat. Is either vegetation species a better predictor of abundance or is some combination of the two the best predictor? In other words, which subset of explanatory variables and error structure together provides the best model? To attempt to answer this question, we adopt a geostatistical model (Cressie 1993) that can be used to predict a response at unobserved locations. This approach, also referred to as kriging, involves the fitting of an autocorrelation function that describes the relationship between observations based on the distance between the observations. This method allows for any number of the explanatory variables observed at the sample locations to be included in the model to improve the overall predictions.

Typically, spatial correlation is ignored in the selection of explanatory variables. Ignoring the autocorrelation structure in the data can influence model selec-

Manuscript received 26 March 2004; revised 14 July 2004; accepted 27 August 2004; final version received 27 September 2004. Corresponding Editor: D. S. Schimel. For reprints of this Invited Feature, see footnote 1, p. 3.

[3] E-mail: jah@lamar.colostate.edu

tion results. For example, the importance of particular explanatory variables may not be apparent when spatial correlation is ignored. To address this problem, we consider the Akaike Information Criterion (AIC) as applied to a geostatistical model. We provide simulation results that show that using AIC for a geostatistical model is superior to the standard approach of ignoring spatial correlation in the selection of explanatory variables. We also consider the impact of the sampling pattern on the model selection. We further demonstrate these ideas via a model for the abundance of the orange-throated whiptail lizard found in southern California. The principle of minimum description length (MDL) applied to the variable selection problem is also investigated and simulation results are provided for comparison.

The Geostatistical Model

Suppose we are interested in the abundance of the orange-throated whiptail lizard in a specific region in southern California (analysis results of this data set are given in *Example*). Assume that we have collected information at each of 150 sites spread across the area of interest. Our data set consists of the average number of lizards observed per day, the percentage of vegetation coverage, the abundance of ants (a primary food source), and a georeference for each site, such as latitude and longitude. It is not feasible to collect data at all possible locations, thus we are assuming that these 150 sites are representative of the entire area of interest. Let $Z(s_i)$ denote the average abundance of lizards at site i where $i = 1, \ldots, 150$. Thus the vector $\mathbf{Z} = (Z(s_1), \ldots, Z(s_{150}))'$ is a partial realization of the continuous random field over this finite area, D. In other words, we are assuming that at any given site s within the domain D, the average abundance of the lizards is a function of a specific set of variables that can be observed along with some random noise.

A model for the continuous random field at any location $s \in D$ is given by

$$\begin{aligned} Z(s) &= \beta_0 + \beta_1 X_1(s) + \cdots + \beta_{p-1} X_{p-1}(s) + \delta(s) \\ &= \mathbf{X}'(s)\boldsymbol{\beta} + \delta(s) \qquad (1) \end{aligned}$$

where $\mathbf{X}(s) = (1, X_1(s), \ldots, X_{p-1}(s))'$ is a p-vector consisting of the constant 1 and $p - 1$ explanatory variables observed at location s, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})'$ is a p-vector of the unknown model coefficients, and $\delta(s)$ is the unobserved "regression" error at location s. For example, $X_1(s)$ and $X_2(s)$ may be the percentage of vegetation coverage and the abundance of ants at location s, respectively. For computational ease, we will assume that the error process $\delta(s)$ is a stationary, isotropic Gaussian process with mean zero and covariance function $\mathrm{Cov}(\delta(s_i), \delta(s_j)) = \sigma^2 \rho_{\boldsymbol{\theta}}(\|s_i - s_j\|)$. Here, σ^2 is the variance of the process, $\rho_{\boldsymbol{\theta}}(\|\cdot\|)$ is a family of autocorrelation functions with a parameter vector $\boldsymbol{\theta}$ of length k, and $\|\cdot\|$ denotes the Euclidean distance between two sites. Thus, we assume that the correlation between any two sites is only a function of the distance between them. In deciding among the covariates, we must also choose an appropriate autocorrelation function. As we will demonstrate, these two issues are inextricably linked.

The autocorrelation function must satisfy certain mathematical conditions in order to be valid. This restricts our selection to one of a number of standard autocorrelation families. Most readers should be familiar with the independent error process associated with multilinear regression. In this case, one is assuming that the errors are identically distributed and independent of one another and of location. For geospatial data, it is reasonable to assume that observations that are nearby will have similar response values, so we seek to model this relationship via the autocorrelation function. A rich family of autocorrelation functions is the Matern family (Handcock and Stein 1993, Stein 1999). The Matern autocorrelation function has the general form

$$\rho_{\boldsymbol{\theta}}(d) = \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)}\left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right)^{\theta_2} K_{\theta_2}\left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right)$$

$$\theta_1 > 0 \qquad \theta_2 > 0 \qquad (2)$$

where $K_{\theta_2}(\cdot)$ is the modified Bessel function of order θ_2 (Abramowitz and Stegun 1965). The "range" parameter θ_1 controls the rate of decay of the correlation between observations as distance increases. Large values of θ_1 indicate that sites that are relatively far from one another are moderately (positively) correlated. The "smoothness" parameter θ_2 can be described as controlling behavior of the autocorrelation function for observations that are separated by small distances. The Matern class includes the exponential autocorrelation function when $\theta_2 = 0.5$ and the Gaussian autocorrelation function as a limiting case when $\theta_2 \to \infty$. The Matern class is very flexible, being able to strike a balance between these two extremes, thus making it well suited for a variety of applications. Figs. 1 and 2 illustrate the flexibility of the Matern autocorrelation function. Notice that, for small distances, the correlation between sites is large and decreases as distance increases.

The autocorrelation function given in Eq. 2 can be further adapted to include the possibility of measurement error, called "nugget" in many spatial contexts. A mixture model that incorporates measurement error in these spatial models is considered in Thompson (2001). To minimize the complexity of the current discussion, we have chosen not to include a nugget effect in our simulations or analysis of the lizard data example. It should be noted that selection of the form of the autocorrelation function can be easily incorporated into the model selection process. For example, one

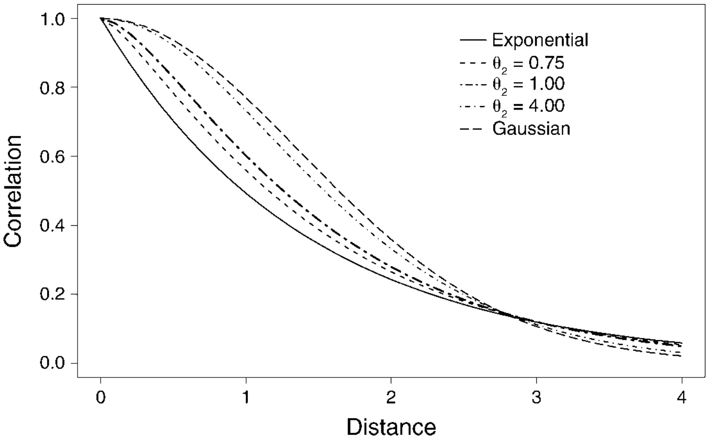

FIG. 1. Matern autocorrelation function for several parameter values. The horizontal axis is the distance between points, and the vertical axis is the correlation between two points at a given distance. We used a fixed range parameter, $\theta_1 = 2.00$, with various smoothness parameter values, θ_2. Note that the exponential autocorrelation is equivalent to the Matern autocorrelation function with $\theta_2 = 0.50$ and that the Gaussian autocorrelation function corresponds to the limiting case such that $\theta_2 \to \infty$.

could assume that the autocorrelation function is Matern but allow the selection process to determine whether or not a nugget should be included.

Estimation

The model in Eq. 1 is often referred to as a geostatistical model or a universal kriging model. For a particular subset of explanatory variables and a structure for the error process, we are now tasked with estimating the parameters $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\theta}$. Estimation of the parameters of this model can proceed using one of several likelihood-based approaches (Haining 1990, Cressie 1993, Smith 2000) or a Bayesian approach (Handcock and Stein 1993, Thompson 2001). Here, we consider the former. Both approaches can be computationally challenging to implement for large sample sizes.

Using the assumption that the error process is Gaussian, the log-likelihood of the parameters in Eq. 1, ($\boldsymbol{\theta}$, $\boldsymbol{\beta}$, σ^2), based on the observed data, $\mathbf{Z}$, is given by

$$\ell(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2; \mathbf{Z}) \propto -\frac{1}{2}\log|\sigma^2\boldsymbol{\Omega}| - \frac{1}{2\sigma^2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$$

where $\boldsymbol{\Omega} = [\rho_{\boldsymbol{\theta}}(\|s_i - s_j\|)]$ represents the matrix of correlations between all pairs of observations, $i, j = 1, \ldots, n$. By concentrating out $\boldsymbol{\beta}$ and σ^2, the profile likelihood can be easily computed, which can often accelerate optimization of the likelihood. That is, by maximizing the likelihood with respect to $\boldsymbol{\beta}$ and σ^2, we obtain $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Z}$ and $\hat{\sigma}^2 = \hat{\sigma}^2(\boldsymbol{\theta}) = (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})/n$. The resulting log profile likelihood is

$$\ell_{\text{profile}}(\boldsymbol{\theta}; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{Z}) \propto -\frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{n}{2}. \quad (3)$$

Maximizing Eq. 3 produces the maximum likelihood estimates for the parameters of the spatial autocorrelation function, $\boldsymbol{\theta}$.

An alternative approach for parameter estimation is the restricted maximum likelihood (REML) approach of Patterson and Thompson (1971). Cressie (1993:93) supports the use of REML over maximum likelihood as a method of estimation when the number of explanatory variables is large. For model selection, most procedures involve a component consisting of the maximized likelihood function. Since REML does not maximize the likelihood, we do not consider REML here further. However, once a model has been selected, the

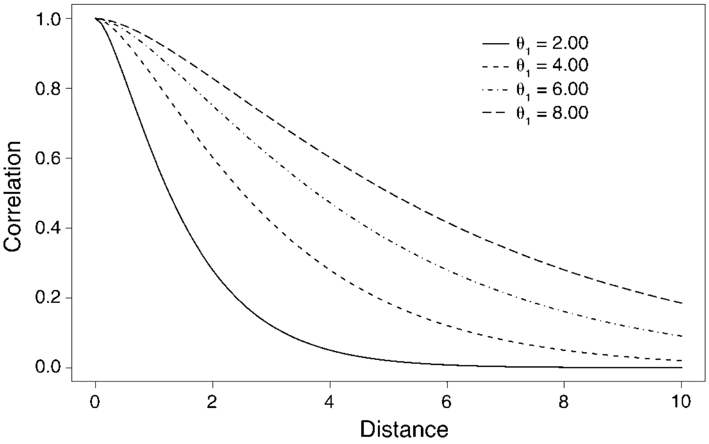

FIG. 2. Matern autocorrelation function for smoothness parameter $\theta_2 = 1.00$ and various range parameter values, θ_1.

researcher is free to re-estimate the model parameters using, for example, REML for parameter estimation.

Model Selection for Geostatistical Models

Model selection is a critical ingredient in nearly any model building exercise. Depending on one's philosophical bent, which is often driven by the modeling objective, there are a myriad of procedures for selecting an optimal model subject to a particular criterion. The introductions in the books by McQuarrie and Tsai (1998) and Burnham and Anderson (2002) give excellent accounts of the various philosophies underpinning model selection. It is important, however, to adopt a model selection paradigm that reflects the ultimate objective of the modeling process. For example, an explanatory model that establishes useful relationships between explanatory and response variables may not necessarily perform as well as a predictive model and vice versa. We first discuss development of the Akaike Information Criterion (AIC) for spatial models of the form in Eq. 1 followed by a discussion of spatial model fitting. The third subsection contains a brief discussion of the concept of minimal description length (MDL) and further remarks on model selection issues. (Appendix B gives the formulas for all three model selection procedures described here.)

Returning to our working example of the whiptail lizard, the current question at hand is which model should be selected. Should we include both of the potential explanatory variables, just one, or perhaps neither? What is required is a quantitative measure of how closely each of the candidate models coincides with the true model. We may also wish to penalize less parsimonious models. We suggest that AIC, extended to spatial models, accomplishes these goals.

AIC for spatial models

There are often two points of view taken in model selection. The first presumes that there exists a true finite-dimensional model from which the data were generated. For example, one might hypothesize the true model to be linear in which there exists an explicit linear relationship between the explanatory variables and the response. In this case, the key modeling objective is to identify the correct set of covariates that comprise the model. The second modeling perspective, which seems particularly well suited for ecological data, is that the "truth" and, consequently, the underlying true model, is essentially infinite dimensional and we have no hope of identifying all the requisite factors that go into the process under study. In other words, reality cannot be expressed as a simple, "true model" because, as Burnham and Anderson (1998) observe, "[Ecological] systems are complex, with many small effects, interactions, individual heterogeneity, and individual and environmental covariates (being mostly unknown to us)." Thus, the goal is to find the best approximating finite dimensional model to this infinite-dimensional problem.

Under the first scenario, consistency should be a minimum requirement of a model selection procedure. That is, as more data are acquired, the model selection procedure should ultimately choose the correct model with probability one. In the second situation, when the true model is infinite dimensional, a model selection procedure ought to choose a finite-dimensional model that is closest to the true model in some sense. The Akaike Information criterion (Akaike 1973) is one procedure that is designed to achieve this second goal.

AIC was developed as an estimator of the Kullback-Leibler information. Roughly speaking, AIC is a measure of the loss of information incurred by fitting an incorrect model to the data. To describe the main idea behind AIC, let $\mathbf{Z}$ be an n-dimensional random vector with true probability density function f_{T} and consider a family $\{f(\cdot;\boldsymbol{\psi}),\ \boldsymbol{\psi}\in\boldsymbol{\Psi}\}$ of candidate probability density functions. The Kullback-Leibler information between $f(\cdot;\boldsymbol{\psi})$ and f_{T} is defined as

$$I(\boldsymbol{\psi}) = \int -2\log\left\{\frac{f(\mathbf{z};\boldsymbol{\psi})}{f_{\mathrm{T}}(\mathbf{z})}\right\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z}. \tag{4}$$

Applying Jensen's inequality, we see that

$$\begin{aligned} I(\boldsymbol{\psi}) &= \int -2\log\left\{\frac{f(\mathbf{z};\boldsymbol{\psi})}{f_{\mathrm{T}}(\mathbf{z})}\right\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z} \\ &\geq -2\log\left\{\int \frac{f(\mathbf{z};\boldsymbol{\psi})}{f_{\mathrm{T}}(\mathbf{z})} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z}\right\} \\ &= -2\log\left\{\int f(\mathbf{z};\boldsymbol{\psi})\,d\mathbf{z}\right\} = 0 \end{aligned}$$

with equality holding if and only if $f(\mathbf{z};\boldsymbol{\psi}) = f_{\mathrm{T}}(\mathbf{z})$ almost everywhere with respect to the true model f_{T}.

By treating $I(\boldsymbol{\psi})$ as the information loss associated with $f(\cdot;\boldsymbol{\psi})$, the idea is to minimize $I(\boldsymbol{\psi})$ over all candidate models $\boldsymbol{\psi}\in\boldsymbol{\Psi}$. Unfortunately, this is not possible without knowing f_{T}, thus we need to adopt a strategy that is not dependent on the unknown density f_{T}.

First rewrite the Kullback-Leibler information in the following manner:

$$\begin{aligned} I(\boldsymbol{\psi}) &= \int -2\log\left\{\frac{f(\mathbf{z};\boldsymbol{\psi})}{f_{\mathrm{T}}(\mathbf{z})}\right\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z} \\ &= \int -2\log\{f(\mathbf{z};\boldsymbol{\psi})\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z} \\ &\quad + \int 2\log\{f_{\mathrm{T}}(\mathbf{z})\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z} \\ &= \Delta(\boldsymbol{\psi}) + \int 2\log\{f_{\mathrm{T}}(\mathbf{z})\} f_{\mathrm{T}}(\mathbf{z})\,d\mathbf{z}. \end{aligned} \tag{5}$$

The first term, defined as the Kullback-Leibler index, can be written as $\Delta(\boldsymbol{\psi}) = E_{\mathrm{T}}\{-2\log L_{\mathbf{Z}}(\boldsymbol{\psi})\}$ where the expectation is taken with respect to the true density and $L_{\mathbf{Z}}(\boldsymbol{\psi})$ is the likelihood based on the candidate model corresponding to using the data $\mathbf{Z}$. Note that the second term in Eq. 5 is a constant and plays no role in the minimization of $I(\boldsymbol{\psi})$. While it is generally not possible to compute either $\Delta(\boldsymbol{\psi})$ or $\Delta(\hat{\boldsymbol{\psi}})$, where $\hat{\boldsymbol{\psi}}$ is the maximum likelihood estimate of $\boldsymbol{\psi}$, we instead strive to find a model that minimizes an unbiased estimate of $E_{\psi}(\Delta(\hat{\boldsymbol{\psi}}))$, where E_{ψ} represents the expectation operator relative to the candidate density $f(\cdot; \boldsymbol{\psi})$.

A heuristic derivation of the AIC statistic in the spatial model setup of Eq. 1 can be found in *AIC derivation.* The quantity

$$\text{AIC}_c = -2 \log\{L_{\mathbf{Z}}(\hat{\boldsymbol{\psi}})\} + 2n\frac{p + k + 1}{n - p - k - 2} \quad (6)$$

is an approximately unbiased estimate of the expected Kullback-Leibler information evaluated at $\hat{\boldsymbol{\psi}}$, where there are p explanatory variables, including an intercept term, k is the number of parameters associated with the autocorrelation function, and n is the number of observed sites. This version is known as the corrected AIC (AIC$_c$) which includes a measure of the quality of fit of the model (first term) and a penalty factor for the introduction of additional parameters into the model (second term). The AIC statistic for this model is

$$\text{AIC} = -2 \log\{L_{\mathbf{Z}}(\hat{\boldsymbol{\psi}})\} + 2(p + k + 1).$$

For large n, the penalty factors, $2n(p + k + 1)/(n - p - k - 2)$ and $2(p + k + 1)$ are nearly equivalent. The AIC$_c$ statistic has a more severe penalty for larger-order models that helps counterbalance the tendency of AIC to overfit models to data.

The principle of AIC is to select a combination of explanatory variables and models for the autocorrelation function that minimize either AIC$_c$ or AIC. It is worth remarking that, in many classical situations, such as linear regression or time-series modeling, AIC$_c$ and AIC are not consistent order selection procedures. In other words, as the sample size increases there is a positive probability that a model selected by AIC$_c$ or AIC does not correspond to the true model. Nevertheless, these statistics should produce good estimates of the Kullback-Leibler information for which they were formulated.

Spatial model fitting

Traditionally, the fitting of the model in Eq. 1 is accomplished in two steps (see, for example, Venables and Ripley [1999:439–444]). In the first step, explanatory variables for modeling the large-scale variation are chosen via a model selection technique such as Akaike's Information corrected criterion (AIC$_c$; Sugiura 1978, Hurvich and Tsai 1989). Second, the residuals from the model are examined for spatial correlation and a suitable family of correlations is chosen. The estimates of the parameters in the trend surface are updated using generalized least squares followed by maximum likelihood estimation of the parameters of the covariance function using the residuals. This two-step estimation process is repeated until some suitable convergence criterion is attained. Since a correlation function is not identified in the selection of the explanatory variables in step 1, AIC$_c$ is implemented under the working assumption of independence of the residuals (Haining 1990, Cressie 1993).

A limitation of the model selection procedure described above is that it ignores potential confounding between explanatory variables and the correlation in the spatial noise process $\{\delta(s)\}$. Although it is extremely convenient to select explanatory variables for the model before fitting a covariance function to the residuals, it is generally not a good idea to separate these two steps. The inclusion of one or more important explanatory variables may remove or reduce the correlation structure of the residuals from the model. For example, Ver Hoef et al. (2001) demonstrate the similarities between a model with independent errors and a linearly decreasing mean and a model with correlated errors and a constant mean. Alternatively, ignoring the autocorrelation structure of the error process may mask explanatory variables that are very important in modeling the mean function. The additional noise in the data can overwhelm the information in the data, resulting in the identification of fewer important explanatory variables. An example of this behavior will be explored in *Simulation*.

Model selection techniques for spatial models need to include the correlation structure in determining the best set of predictors. By computing the AIC$_c$ statistic described in Eq. 6 for all possible sets of explanatory variables and autocorrelation functions, one can find a single ''best'' model or a set of models which fit the data well. This method attempts to strike a balance between the competing forces of large scale variability, as modeled via the explanatory variables, and small scale variability, as modeled through the correlation in the residuals.

Other considerations

The AIC$_c$ statistic in Eq. 6 for the geostatistical model in Eq. 1 required that the true model was a member of the family of candidate models, all of which were finite dimensional. However, in many applications (McQuarrie and Tsai 1998, Burnham and Anderson 2002), the AIC$_c$ selection procedure enjoys additional optimality properties regarding the choice of a finite-dimensional model when the true model is in fact infinite dimensional. This includes the notion of efficiency for prediction in time series models and optimal signal-to-noise ratios for linear models (McQuarrie and Tsai 1998).

AIC and other information-based criteria such as BIC and HQ (Kass and Raftery 1995, McQuarrie and Tsai 1998) have an objective function consisting of two pieces. The first is related to -2(log-likelihood), which is a measure of the quality of fit of a model, and the second is a penalty factor for the introduction of additional parameters into the model. The principle of minimum description length (MDL), an idea developed by Rissanan in the 1980s, also contains two similar pieces, but is motivated by different ideas. MDL attempts to achieve maximum data compression by the fitted model.

The idea behind MDL is to decompose the code length of the ''data'' into two pieces (see the survey paper by Lee [2001] for more details). Roughly speaking, the code length of the ''data'' is the amount of memory required to store the data. Typically, the code length of the data can be decomposed into the sum of the code length of the fitted model and the code length of the data given the fitted model, i.e.,

$$L(\text{data}) = L(\text{fitted model}) + L(\text{data given fitted model}).$$

Here, L(fitted model) might be interpreted as the code length of the model parameters and L(data given fitted model) as the code length of the residuals from the fitted model. It follows that a more complex model is chosen provided there has been a compensating decrease in the code length of the residuals. According to the MDL principle, the best model is the one producing the shortest code length for the data. The attraction of this procedure is that the data is being compressed in the most efficient manner possible and the notion of a true model at any level is not required.

The code length of the fitted model based on the MLE, $\hat{\boldsymbol{\psi}}$, can be approximated by $L(\text{fitted model}) \simeq 1/2(p + k + 1)\log_2 n$. The code length of the data given the model based on $\hat{\boldsymbol{\psi}}$ is approximated by $\log_2 L(\hat{\boldsymbol{\psi}})$. Adding these terms together and rescaling, the minimum description length is defined by

$$\text{MDL} = \frac{1}{2}\{-2\log[L_{\mathbf{Z}}(\hat{\boldsymbol{\psi}})] + \log(n)(p + k + 1)\}.$$

The only difference between the value of AIC$_c$ (using the spatial AIC$_c$ method) and 2×MDL is the magnitude of the penalty term coefficient. For AIC$_c$, the leading coefficient is of order 2 compared to $\log(n)$ for 2×MDL. For sample sizes greater than eight, the penalty for 2×MDL is larger. For example, when $n = 100$, $p = 4$, and $k = 2$ the penalty coefficients are 2 and 4.60, respectively. MDL generally selects more parsimonious models, i.e., models with fewer explanatory variables.

Bayesian model averaging is an alternative approach to model selection and prediction (Hoeting et al. 1999). The idea of Bayesian model averaging is to average across several models instead of selecting one model. In computing the average, each model is weighted by its posterior model probability, a measure of the degree of model support in the data. Empirical and theoretical results over a broad range of model classes indicate that Bayesian model averaging can provide improved out-of-sample predictive performance as compared to single models. For the geostatistical model in Eq. 1, Thompson (2001) showed that Bayesian model averaging can offer improved predictive performance as compared to the single models that are selected when spatial correlation is ignored. However, the gains are modest in the simulations that were explored.

Simulation

To explore the impact of ignoring spatial correlation on model selection, we carried out a simulation comparing the explanatory variables selected using standard independent AIC model selection which ignores spatial correlation to those selected using the spatial AIC approach and the MDL approach. In addition to comparing the impact of accounting for spatial correlation in the selection of a set of explanatory variables, we also explored the impact of sampling pattern on the selection of explanatory variables. We considered five sampling patterns shown in Fig. 3; highly clustered, lightly clustered, random, regular, and a grid design. Finally, we conducted some simulation studies to characterize the strength of the predictive ability when spatial correlation is included in the selection process of explanatory variables.

We simulated five possible explanatory variables: $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_4$, $\mathbf{X}_5$. Each explanatory variable was independently generated from a standardized Student's t distribution with 12 degrees of freedom, $\mathbf{X}_i \sim \sqrt{12/10}(t_{12})$ for $i = 1, \ldots, 5$. The explanatory variables were fixed and identical for all simulations.

For a given sampling pattern of size $n = 100$, the data were simulated from the model

$$\mathbf{Z} = \mathbf{2} + 0.75\mathbf{X}_1 + 0.50\mathbf{X}_2 + 0.25\mathbf{X}_3 + \boldsymbol{\delta} \qquad (7)$$

where $\boldsymbol{\delta}$ is a Gaussian random field with mean zero, $\sigma^2 = 50$, and autocorrelation Matern with parameters $\theta_1 = 4$ and $\theta_2 = 1$. (Results for other values of the Matern parameters are also provided.) For each sampling pattern, 500–1000 replicates were simulated with a new Gaussian random field generated for each replication. The largest signal of Eq. 7 is associated with $\mathbf{X}_1$ which is three times the ''strength'' of $\mathbf{X}_3$. Thus, we expect that the majority of models selected should at least include $\mathbf{X}_1$.

With five possible explanatory variables, there are $2^5 = 32$ possible combinations of explanatory variables, including the intercept-only model. For each realization, we computed the AIC$_c$ statistic for all 32 possible models. For the traditional method, the AIC$_c$ statistic was calculated using Eq. 6 with $k = 0$. We call this the independent AIC$_c$ approach. The spatial AIC$_c$ results were calculated using Eq. 6 as well with $k =$

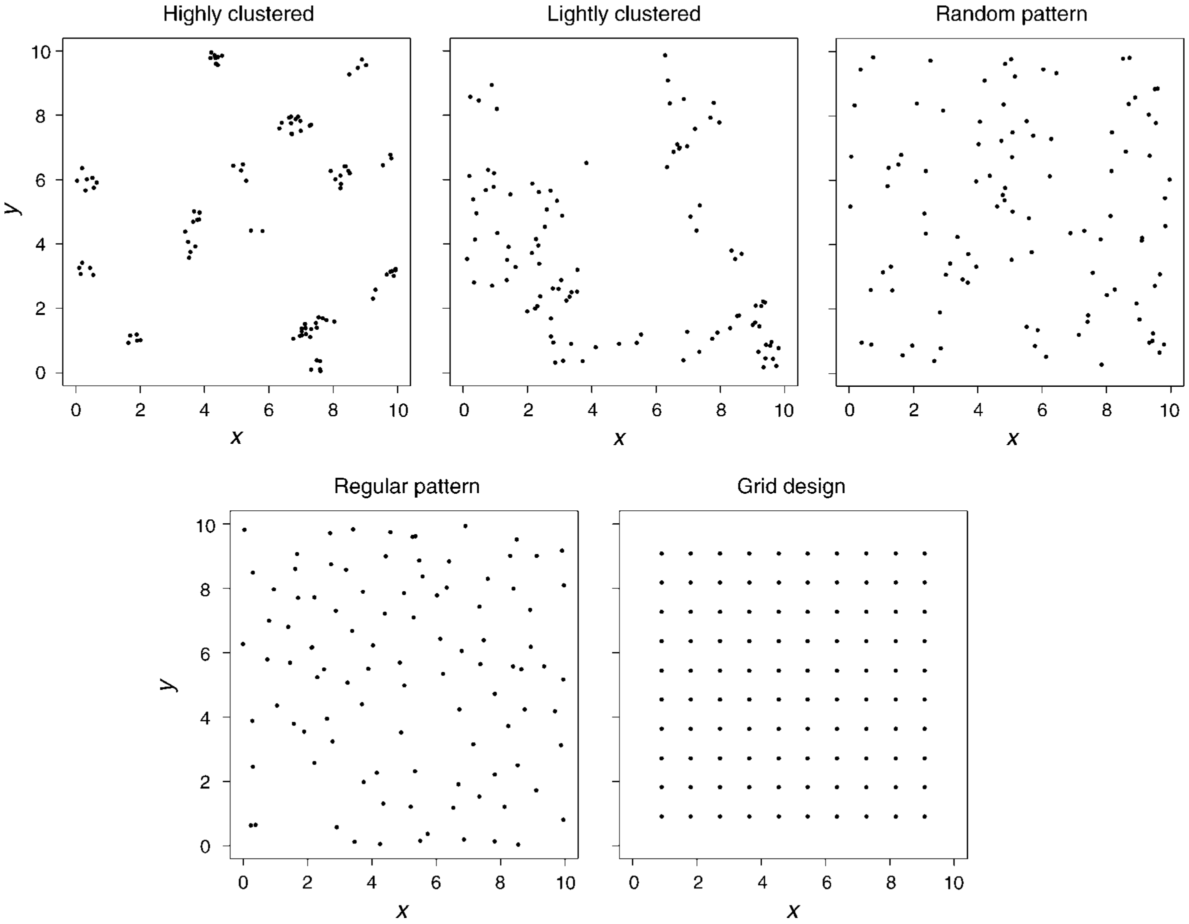

FIG. 3. Five sampling patterns.

2. Further details on the simulation set-up and additional simulation results are given in Thompson (2001).

General simulation results

Table 1 compares the models selected by the spatial AIC$_c$ and independent AIC$_c$ approaches. When independence is assumed, the AIC$_c$ statistic selects the true model ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$) only 12 out of 500 simulations (2.4%) while the intercept-only model is selected in 134 out of 500 simulations (26.8%). Over all 500 simulations, the AIC$_c$ independence approach selected models that included both explanatory variables $\mathbf{X}_1$ and $\mathbf{X}_2$ only 15.8% of the time. These results provide a vivid example of the drawbacks of the standard model selection approach for spatially correlated data. In total, the first explanatory variable is in 40.2% of the selected models, and the second explanatory variable is included in 35.4% of the models.

Spatial AIC$_c$ has superior model selection performance as compared to the independent AIC$_c$ method. The true model is selected in 56.0% of the simulations (Table 1). When the true model is not selected, this method tends to overestimate the number of parameters in the model, selecting models with one or two extra variables (28.4%). In contrast to the AIC$_c$ independence approach, the first explanatory variable is in 100% of the selected models and the second explanatory variable is included in 98.6% of the models.

Fig. 4 illustrates the necessity of including spatial correlation during model selection. The top panel lists

TABLE 1. Model selection results for the random pattern.

Variables in model	Spatial AIC$_c$	Independent AIC$_c$	MDL
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$	56.0	2.4	40.4
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_5$	14.4	0.2	4.2
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_4$	10.8	0.2	0.8
$\mathbf{X}_1$, $\mathbf{X}_2$	10.2	8.4	46.4
Intercept only	0.0	26.8	0.0
$\mathbf{X}_1$	0.4	14.2	1.2
$\mathbf{X}_2$	0.0	13.8	0.2

Notes: Spatial and independent AIC$_c$ (corrected Akaike Information Criterion) and MDL (minimum description length) report the percentage of simulations in which each model was selected. Of the 32 possible models, the results given here include only those with 10% or more support from at least one model selection method.

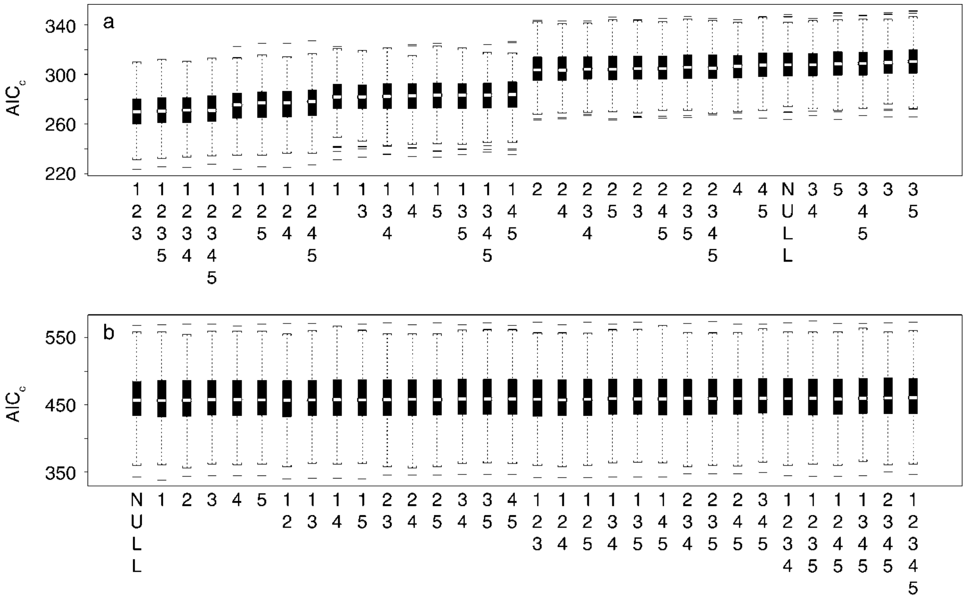

FIG. 4. AIC$_c$ values for (a) the spatial AIC$_c$ and (b) the independent AIC$_c$ selection strategies. Note that the models for the spatial AIC$_c$ method have been ordered from smallest to largest average AIC$_c$ over all 500 simulations. The horizontal axis lists the variables included in each model; NULL refers to the intercept-only model.

the models from smallest to largest average AIC$_c$ over all 500 simulations. The horizontal axis list the variables included in the model where null refers to the intercept-only model. Note that the model with the smallest average AIC$_c$ is the true model ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$). All of the first 16 models listed include $\mathbf{X}_1$, while the first eight models also include $\mathbf{X}_2$. In sharp contrast, the boxplots for the independence assumption during model selection are virtually identical. Although the models are listed from most to least parsimonious, any rearrangement would look nearly identical. The lack of trend in this plot illustrates that ignoring spatial dependence during variable selection may lead to selection of an inappropriate model.

Table 1 also demonstrates MDL's ability to select the appropriate model when spatial correlation is accounted for during variable selection. Although it only selects the ''true'' model for 40.4% of the simulations, it selects the model containing only $\mathbf{X}_1$ and $\mathbf{X}_2$ 46.4% of the time. These results are consistent with the idea that MDL more strongly penalizes models with a large number of explanatory variables and thus tends to select more parsimonious models. Also note that MDL selects one of three models for more than 90% of the simulations.

To further evaluate the performance of the spatial AIC$_c$ strategy, we performed additional simulations using different true values of the Matern correlation function parameters. These results are given in Appendix A. As the range and smoothness parameters increased, the true model was selected with increasing frequency. The result for the range parameter is somewhat surprising and may be a result of the signal-to-noise ratio used in these simulations. The independent AIC$_c$ approach had uniformly poor performance for all parameter values.

Impact of sampling pattern

The advantages of using spatial AIC$_c$ when the data are spatially correlated are enhanced when the sampling pattern includes both some closely spaced and more distant pairs of sample locations. Similar simulations to those described above were performed using the five sampling patterns shown in Fig. 3. The models selected using spatial AIC$_c$ for the five sampling patterns are given in Table 2. The highly and lightly clustered patterns select the true model in over 65% of the simulations. For this simulation setup, as the sampling pattern provides less information at small distances, the selection of the correct explanatory variables becomes more challenging. Indeed, for the grid design,

TABLE 2. Spatial AIC$_c$ model selection results for five different sampling patterns.

Variables in model	Highly clustered	Lightly clustered	Random	Regular pattern	Grid design
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$	73	65	46	43	16
$\mathbf{X}_1$, $\mathbf{X}_2$	0	2	18	21	35
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_4$	12	13	8	8	3
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_5$	10	13	11	7	7

Notes: Each column reports the percentage of simulations in which each model was selected. Of the 32 possible models, the results given here include only those with 10% or more support for at least one of the sampling patterns.

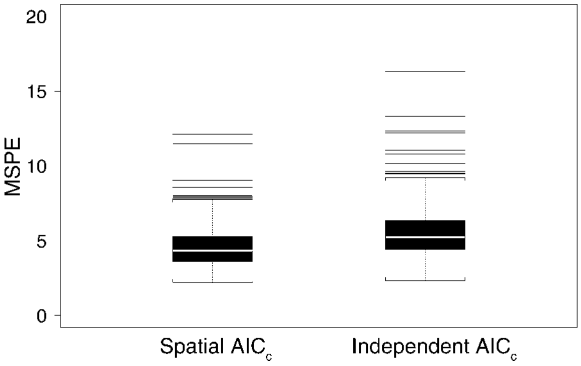

FIG. 5. Mean squared prediction error (MSPE) for the two model selection methods based on 500 simulations.

the correct model was only selected in 16% of the simulations.

For all five sampling patterns, the independent AIC$_c$ approach gave similar results to those for the random pattern given in Table 1. Over all five sampling patterns, the independent AIC$_c$ approach selected the correct model in less than 1% of the simulations and the model with $\mathbf{X}_1$ and $\mathbf{X}_2$ was selected 5% of the simulations.

For these simulations, the AIC$_c$ independence method tends to select models with very few explanatory variables, and does a poor job of selecting models that contain the true parameters. The spatial AIC$_c$ method does very well in selecting the true model, over a variety of sampling designs. The spatial AIC$_c$ approach performs best when the sampling pattern provides sample locations at both close and near distances such as the highly and lightly clustered patterns shown in Fig. 3.

Mean square prediction error

Another measure of the importance of including spatial correlation during model selection is the concept of the mean square prediction error (MSPE). We can evaluate MSPE for the simulated data because we know the true underlying model, Eq. 7. MSPE is the average squared difference between the actual and predicted values at the new series of locations such that

$$\text{MSPE} = \frac{1}{n}\sum_{j=1}^{n}(Z_j - \hat{Z}_j)^2.$$

Here $\hat{Z}_i$ is the universal kriging predictor for the jth prediction location using the maximum likelihood estimate of the parameter vector $\boldsymbol{\psi}$ and Z_j is the true value at location j. Small values of MSPE indicate predicted values are close to the true values on average, where an MSPE of exactly zero corresponds to perfect prediction. What we expect to see is that the MSPE is systematically smaller for spatial AIC$_c$ compared to independent AIC$_c$.

First, 100 locations were randomly selected over the same study area. For each simulation in *General simulation results*, a new set of observations was generated over the new grid using Eq. 1. Next we computed the predicted response at each site using the selected model from each method. Last, MSPE was calculated using both methods for each of the 500 simulations. Fig. 5 illustrates the improvement made by incorporating spatial correlation into the model selection process. The mean MSPE for the spatial AIC$_c$ method was 4.57 compared to 5.50 for the independent AIC$_c$ selection method (an improvement of 16.9%). Over the set of 500 simulations, the two methods selected the same model only 11 times. When these simulations were removed from the data set, the improvement in mean MSPE increases to 17.3%. It should be noted that, when spatial correlation was ignored altogether, i.e., independent error structure, the mean MSPE was 39.6.

EXAMPLE

We applied the model selection strategy to the whiptail lizard data previously analyzed by Hollander et al. (1994) and Ver Hoef et al. (2001). The data set consists of abundance data for the orange-throated whiptail lizard in southern California. A total of 256 locations in 21 regions were used for trapping. Each observation consists of the average number of lizards caught per day at each location. After removing sites where no lizards were caught, a total of 148 observations remained for the abundance analysis. Fig. 6 shows that the pattern of the sites where the lizards were observed was highly clustered. A log transformation was applied to the response, average number of lizards caught per day, to allow for the use of a Gaussian random field.

There are total of 37 explanatory variables available including information on vegetation layers, vegetation types, topographic position, soil types, and abundance of ants. This corresponds to approximately 2^{37} or 1.374 $\times$ 10^{11} total models. To make the analysis tractable, the number of explanatory variables was reduced to six. See Ver Hoef et al. (2001) for further details about preliminary explanatory variable selection.

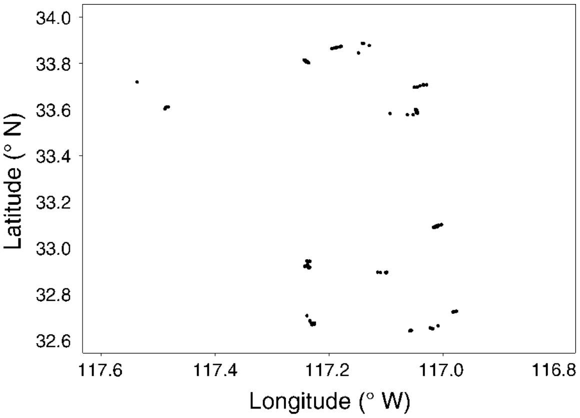

FIG. 6. Locations in southern California where the whiptail lizard was observed (n = 148).

TABLE 3. Model selection results for the whiptail lizard data set.

Predictors	AIC_c	Spatial rank	Independent rank
Ant_1, sand (%)	54.1	1	66
Ant_1, Ant_2, sand (%)	54.8	2	56
Ant_1, sand (%), cover (%)	55.7	3	59
Ant_1, Ant_2, sand (%), cover (%), elevation, bare rock, chaparral (%)	92.0	41	1
Ant_1, Ant_2, sand (%), elevation, bare rock, chaparral (%)	95.3	33	2
Ant_1, sand (%), cover (%), elevation, bare rock, chaparral (%)	95.6	38	3

Notes: The first column lists the explanatory variables selected using AIC as the selection criterion. The rank of the model (by AIC) is provided under both model selection strategies. Ant_1 corresponds to low abundance, and Ant_2 corresponds to medium abundance.

The subset of explanatory variables used in the analysis were *Crematogaster* ant abundance (three categories: low, medium, and high), log(percentage of sandy soils), elevation, a binary indicator variable that described whether or not the rock was bare, percentage of cover, and log(percentage of chapparal plants). Ant abundance is a categorical variable and has five unique modeling subsets. This leads to a total of $5 \times 2^5 = 160$ possible models.

All 160 unique models were fit to the data using the spatial AIC_c and independent AIC_c strategies. We assumed a Matern autocorrelation structure (without nugget) for each model. For comparison, the traditional model selection approach was also applied to the data set. Table 3 summarizes the top three models selected when employing each strategy. For each model, the corresponding rank under the opposing strategy is also listed. The two methods select very different models. When spatial dependence is incorporated into the selection of explanatory variables, very parsimonious models are chosen and are consistent with the results of Ver Hoef et al. (2001). The traditional approach leads to much more complicated models. By initially assuming independent covariates, the selection process is trying to compensate for correlation in the error structure by incorporating too many explanatory variables. In fact, the full model has the smallest AIC_c when the correlation structure is not incorporated into model selection. Finally, the top three models selected by the MDL method exactly matched those selected by the spatial AIC_c method.

SOFTWARE

Software to perform the model selection strategy for geostatistical models is available in the Supplement. The software, written in the R language, implements the Matern covariance function Eq. 2, but can easily be modified to incorporate other covariance functions.

CONCLUSIONS

Our results demonstrate the problems that can be encountered in the selection of an appropriate set of explanatory variables when spatial correlation is ignored. Both the AIC and MDL criteria based on the geostatistical models performed well in the selection of appropriate explanatory variables. Ignoring spatial correlation in the selection of explanatory variables and/or in the modeling of the data can lead to the selection of too few explanatory variables as well as higher prediction errors. In addition, we showed that for the sampling patterns considered here, it is advantageous to consider a clustered type of sampling design that offers observation pairs at both small and larger distances.

We have considered the impact of ignoring spatial correlation on the selection of explanatory variables. We must note that the concept of ''all possible models'' can become intractable quickly when the number of potential explanatory variables becomes large. For this presentation, we have assumed that only the candidate predictors enter the model as main effects with no interactions or higher order terms under consideration. Thus, for a data set with 10 potential explanatory variables there are $2^{10} = 1024$ candidate models if only a single error structure is examined. But a data set with 20 potential explanatory variables leads to 1.04×10^6 candidate models under the same setup. This number will increase further if, for example, we allow interactions or higher-order polynomial fits. Thus there are practical limitations to the proposed model selection method. To overcome this limitation, the researcher has many avenues open to her. She can perform exploratory data analyses to reduce the number of potential explanatory variables, limit the candidate models to a particular class (such as linear), restrict the error structure to a single form, e.g., Matern without nugget, and so forth. Often a researcher can rely on her expertise to further reduce the size of the family of candidate models. Both spatial AIC_c and MDL allow the researcher to restrict the type and class of models that best suits her needs.

Finally, other aspects of model misspecification, such as the appropriateness of the adoption of a Gaussian random field and stationarity autocorrelation function, are also important. Cressie (1993:289) and Smith (2000:94–96) summarize some of the research on these issues.

AIC Derivation

To give a heuristic derivation of the AIC statistic in the spatial model setup of Eq. 1, we follow the development in Brockwell and Davis (1991:303). Suppose $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ are two independent realizations from Eq. 1 at fixed locations $(s_1, \ldots, s_n)$ with true parameter value $\boldsymbol{\psi}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0^2)'$. Let $f(\cdot\,; \boldsymbol{\psi})$ be a candidate Gaussian density function corresponding to the parameter vector $\boldsymbol{\psi}_0 = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)'$. Then, by the independence of $\mathbf{Y}$ and $\mathbf{Z}$,

$$E_\psi[\Delta(\hat{\boldsymbol{\psi}})] = E_\psi(E_\psi\{-2 \log [L_\mathbf{Y}(\hat{\boldsymbol{\psi}})] \,|\, \mathbf{Z}\}) = E_\psi[-2 \log L_\mathbf{Y}(\hat{\boldsymbol{\psi}})]$$

where $L_\mathbf{Y}$ is the likelihood based on $\mathbf{Y}$ and $\hat{\boldsymbol{\psi}}$ is the maximum likelihood estimate of $\boldsymbol{\psi}$ based on $\mathbf{Z}$. Using properties of the Gaussian density function and the representation $\hat\sigma^2 = (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}})(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})/n$, we have

$$-2 \log [L_\mathbf{Y}(\hat{\boldsymbol{\psi}})] = -2 \log [L_\mathbf{Z}(\hat{\boldsymbol{\psi}})] + \hat\sigma^{-2}\, S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - n \qquad (8)$$

where $S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$. The goal is to find an unbiased approximation for $E_\psi[\hat\sigma^{-2}S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})]$ of Eq. 8.

Using a second-order Taylor series to expand $S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ in a neighborhood of $(\boldsymbol{\beta}, \boldsymbol{\theta})$, we obtain

$$\begin{aligned} S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) \simeq{}& S_\mathbf{Y}(\boldsymbol{\beta}, \boldsymbol{\theta}) + [(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{\beta}, \boldsymbol{\theta})]\frac{\partial S_\mathbf{Y}(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial(\boldsymbol{\beta}, \boldsymbol{\theta})} \\ &+ \frac{1}{2}[(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{\beta}, \boldsymbol{\theta})]'\frac{\partial^2 S_\mathbf{Y}(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial(\boldsymbol{\beta}, \boldsymbol{\theta})\partial(\boldsymbol{\beta}, \boldsymbol{\theta})'} \\ &\times [(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{\beta}, \boldsymbol{\theta})]. \end{aligned} \qquad (9)$$

To evaluate the expected value of the terms in Eq. 9, we assume that standard asymptotics hold for the MLE $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat\sigma^2)'$. These are

i) $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})'$ is approximately normal with mean $(\boldsymbol{\beta}, \boldsymbol{\theta})'$ and asymptotic covariance matrix given by the inverse of the Fisher information, I_n.

ii) For large n, I_n^{-1} can be approximated by

$$V(\boldsymbol{\beta}, \boldsymbol{\theta}) := \left\{-\frac{1}{2\hat\sigma^2}E_\psi\left[\frac{\partial^2 S_\mathbf{Y}(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial(\boldsymbol{\beta}, \boldsymbol{\theta})\partial(\boldsymbol{\beta}, \boldsymbol{\theta})'}\right]\right\}^{-1}$$

iii) For large n, $n\hat\sigma^2 = S_\mathbf{Z}\,(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ is distributed as $\sigma^2\chi^2(n - p - k)$ and is independent of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})'$, where k is the dimension of the parameter $\boldsymbol{\theta}$ associated with the correlation function for the noise process $\{\delta(s)\}$.

Using the independence of $\mathbf{Y}$ and $\mathbf{Z}$, we find that

$$\begin{aligned} E_\psi S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) &\simeq E_\psi S_\mathbf{Y}(\boldsymbol{\beta}, \boldsymbol{\theta}) + [(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{\beta}, \boldsymbol{\theta})]' \\ &\quad \times [V(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})]^{-1}[(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - (\boldsymbol{\beta}, \boldsymbol{\theta})] \\ &\simeq \sigma^2 n + \sigma^2(p + k). \end{aligned}$$

Hence, from the last two terms of Eq. 8, we have

$$\begin{aligned} &E_\psi[\hat\sigma^{-2}S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})] - n \\ &\quad = E_\psi(\hat\sigma^{-2})E_\psi[S_\mathbf{Y}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})] - n \\ &\quad \simeq \left(\sigma^2\frac{n - p - k - 2}{n}\right)^{-1}\sigma^2(n + p + k) - n \\ &\quad = 2n\frac{p + k + 1}{n - p - k - 2}. \end{aligned}$$

The quantity

$$\mathrm{AICC} = -2 \log [L_\mathbf{Z}(\hat{\boldsymbol{\psi}})] + 2n\frac{p + k + 1}{n - p - k - 2} \qquad (10)$$

is an approximately unbiased estimate of the expected Kullback-Leibler information evaluated at $\hat{\boldsymbol{\psi}}$.

The argument given above for $\mathrm{AIC_c}$ relied on the validity of standard asymptotic theory for the maximum likelihood estimates of the parameters in the spatial model in Eq. 1. In order for these results to hold, it is likely an increasing sample size that both fills in and expands the domain under study is required. In the statistics literature, this is often referred to as infill and increasing domain asymptotics. Unfortunately, asymptotic theory for maximum likelihood estimates for unequally spaced data is not fully developed. In the case where data are regularly spaced on a lattice, more complete asymptotic results can be obtained.

Acknowledgments

Research was supported in part by National Science Foundation grants DEB-0091961, DMS-9806243 (Hoeting), DMS-0308109 (Davis), and by STAR Research Assistance Agreement CR-829095 awarded to Colorado State University by the U.S. Environmental Protection Agency (Hoeting, Davis, and Merton). EPA does not endorse any products or commercial services mentioned here. The views expressed here are solely those of the authors.

Literature Cited

Abramowitz, M., and I. A. Stegun, editors. 1965. Handbook of mathematical functions. Dover, New York, New York, USA.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* B. Petrox and F. Caski, editors. Proceedings of the Second International Symposium on Information Theory. Akademia Kiado, Budapest, Hungary.

Brockwell, P. J., and R. A. Davis. 1991. Time series: theory and methods. Springer-Verlag, New York, New York, USA.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information theoretic approach. Springer-Verlag, New York, New York, USA.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference: a practical information theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Cressie, N. A. C. 1993. Statistics for spatial data. Revised edition. Wiley, New York, New York, USA.

Haining, R. 1990. Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge, UK.

Handcock, M. S., and M. L. Stein. 1993. A Bayesian analysis of kriging. Technometrics **35**:403–410.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: a tutorial with discussion. Statistical Science **14**:382–417.

Hollander, A. D., F. W. Davis, and D. M. Stoms. 1994. Hierarchical representations of species distributions using maps, images and sighting data. Pages 71–90 *in* R. I. Miller, editor. Mapping the diversity of nature. Chapman and Hall, London, UK.

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika **76**: 297–307.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association **90**:773–795.

McQuarrie, A. D., and C.-L. Tsai. 1998. Regression and time series model selection. World Scientific, Hackensack, New Jersey, USA.

Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. Biometrika **58**:545–554.

Smith, R. L. 2000. Spatial statistics in environmental science. Pages 152–183 *in* W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young, editors. Nonlinear and nonstationary signal processing. Cambridge University Press, Cambridge, UK.

Stein, M. L. 1999. Interpolation of spatial data. Springer, New York, New York, USA.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communications in Statistics Part A—Theory and Methods **7**:13–26.

Thompson, S. E. 2001. Bayesian model averaging and spatial prediction. Dissertation. Colorado State University, Fort Collins, Colorado, USA.

Venables, W. N., and B. D. Ripley. 1999. Statistics and computing. Third edition. Springer, New York, New York, USA.

Ver Hoef, J. M., N. Cressie, R. N. Fisher, and T. J. Case. 2001. Uncertainty and spatial linear models for ecological data. Pages 214–237 *in* C. Hunsaker, M. Goodchild, M. Friedl, and T. Case, editors. Spatial uncertainty for ecology: implications for remote sensing and GIS applications. Springer-Verlag, New York, New York, USA.

APPENDIX A

A comparison of the performance of the spatial and independent corrected Akaike Information Criterion (AIC$_c$) model selection strategies for various parameterizations of the Matern autocovariance function (*Ecological Archives* A016-007-A1).

APPENDIX B

A listing of the working equations used to compute spatial AIC$_c$, independent AIC$_c$, and minimum description length (MDL) (*Ecological Archives* A016-007-A2).

SUPPLEMENT

Software to perform model selection for geostatistical models (*Ecological Archives* A016-007-S1).