Model Selection for Geostatistical Models*

Jennifer A. Hoeting, Richard A. Davis, Andrew A. Merton
Colorado State University

Sandra E. Thompson
Pacific Northwest National Lab

March 25, 2004

Abstract

We consider the problem of model selection for geospatial data. Spatial correlation is typically ignored in the selection of explanatory variables and this can influence model selection results. For example, the inclusion or exclusion 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 approach of ignoring spatial correlation in the selection of explanatory variables. These ideas are further demonstrated via a model for lizard abundance. We also employ 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. S-Plus and R software to implement the geostatistical model selection methods described in this paper is available at www.stat.colostate.edu/∼jah.

KEYWORDS: *geospatial data, AIC, MDL, kriging, matern autocorrelation function, orange-throated whiptail lizard abundance*

1 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 geo-referenced sample locations. As a result, models for spatially correlated data are becoming increasingly important. We consider these two problems together, spatial modeling and model selection. 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.

*Research 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 and Davis). EPA does not endorse any products or commercial services mentioned here. The views expressed here are solely those of the authors.

We adopt a geostatistical model (Cressie, 1993) which can be used to predict a response at unobserved locations. This approach, also referred to as kriging, involves the fitting of an autocorrelation function which describes the relationship between observations based on the distance between the observations. To improve predictions, we adopt a model that incorporates explanatory variables observed at the sample locations.

Spatial correlation is typically ignored in the selection of explanatory variables. Ignoring the autocorrelation structure in the data can influence model selection results. For example, the importance of particular explanatory variables may be 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 standard approach of ignoring spatial correlation in the selection of explanatory variables. The principle of minimum description length (MDL) applied to the variable selection problem is also investigated and simulation results are provided for comparison. We further demonstrate these ideas via a model for the abundance of the orange-throated whiptail lizard found in southern California.

Our paper proceeds as follows. In Section 2 we describe a geostatistical model and methods for parameter estimation. In Section 3 we develop the Akaike Information Criterion as applied to a geostatistical model and discuss spatial model fitting and other model selection issues. Simulation results in Section 4 and an example in Section 5 underscore the importance of accounting for spatial correlation in the selection of explanatory variables. The effect of sampling design on the selection of explanatory variables for geostatistical models is also considered in Section 4. S-Plus and R software to implement the methods described in this paper will be made freely available on the internet.

2 The Geostatistical Model

Let $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))'$ be a partial realization of a random field $\mathbf{Z}(s)$, where $s \in D$ and D is a fixed finite area under study. A model for the random field at any location s is given by

$$Z(s) = \mathbf{X}'(s)\boldsymbol{\beta} + \delta(s), \tag{1}$$

where $\mathbf{X}(s) = (1, X_1(s), \ldots, X_{p-1}(s))'$ is a vector consisting of the constant 1 and $p-1$ explanatory variables observed at location s, $\boldsymbol{\beta}$ is a p vector of unknown coefficients and $\delta(s)$ is the unobserved "regression" error at location s. We assume that the error process $\delta(s)$ is a stationary, isotropic Gaussian process with mean zero and covariance function $\mathrm{Cov}(\delta(s), \delta(t)) = \sigma^2 \rho_{\boldsymbol{\theta}}(||s-t||)$, where σ^2 is the variance of the process, $\rho_{\boldsymbol{\theta}}(\cdot)$ is a family of autocorrelation functions with a parameter vector $\boldsymbol{\theta}$, and $||\cdot||$ denotes Euclidean distance.

This modeling framework allows for a number of standard autocorrelation functions. For modeling, we use the Matern family of autocorrelation functions (Handcock and Stein, 1993; Stein, 1999). The Matern autocorrelation function has the general form

$$\rho_{\boldsymbol{\theta}}(d) = \frac{1}{2^{\theta_2-1}\Gamma\left(\theta_2\right)} \left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{2d\sqrt{\theta_2}}{\theta_1}\right), \qquad \theta_1 > 0,\ \theta_2 > 0, \tag{2}$$

where $\mathcal{K}_{\theta_2}(\cdot)$ 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. The parameter θ_2 controls the smoothness of the random field. The Matern class includes the exponential autocorrelation function when $\theta_2 = .5$ and the Gaussian autocorrelation function as a limiting case when $\theta_2 \to \infty$. The Matern class is very flexible, being able to to strike a balance between these two extremes, thus making it well suited for modeling isotropic random fields.

The autocorrelation function given in (2) can be 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).

2.1 Estimation

The model in (1) is often referred to as a geostatistical model or a universal kriging model. Estimation of the parameters of this model can proceed using one of several likelihood based approaches (Cressie, 1993; Haining, 1990; 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.

The log-likelihood of the parameters in equation (1), $\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2\right)$, based on the data, $\boldsymbol{Z}$, is given by

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

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 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}\left(\boldsymbol{\theta}\right) = \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z}$ and $\hat{\sigma}^2 = \hat{\sigma}^2\left(\boldsymbol{\theta}\right) = \left(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)/n$. The resulting log profile likelihood is

$$\ell_{profile}\left(\boldsymbol{\theta}; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \boldsymbol{Z}\right) = -\frac{1}{2}\log\left|\boldsymbol{\Omega}\right| - \frac{n}{2}\log\left(\hat{\sigma}^2\right) - \frac{n}{2}. \tag{3}$$

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

3 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. Section 3.1 develops the Akaike Information Criterion (AIC) for spatial models of the form (1) while Section 3.2 discusses spatial model fitting. Section 3.3 contains a brief discussion on Minimal Description Length (MDL) and further remarks on model selection issues.

3.1 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. An alternate modeling perspective, which seems particularly well suited for ecological data, is that the underlying *true* model is infinite dimensional and we have no hope of identifying all the requisite factors that go into the process under study. Instead, 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 $\boldsymbol{Z}$ be an n-dimensional random vector with *true* probability density function f_T and consider a family $\{f(\cdot;\psi),\ \psi \in \boldsymbol{\Psi}\}$ of candidate probability density functions. The Kullback-Leibler information between $f(\cdot;\psi)$ and f_T is defined as

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

Applying Jensen's inequality, we see that

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

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

By treating $I(\psi)$ as the information loss associated with $f(\cdot;\psi)$, the idea is to minimize $I(\psi)$ over all candidate models $\psi \in \boldsymbol{\Psi}$. 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;

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

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

To give a heuristic derivation of the AIC statistic in the spatial model setup of (1), we follow the development in Brockwell and Davis (1991), p. 303. Suppose $\boldsymbol{Z} = (Z_1,\ldots,Z_n)'$ and $\boldsymbol{Y} = (Y_1,\ldots,Y_n)'$ are two independent realizations from model (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 $\boldsymbol{\psi} = (\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)'$. Then by the independence of $\boldsymbol{Y}$ and $\boldsymbol{Z}$,

$$E_\psi\left[\Delta(\hat{\boldsymbol{\psi}})\right] = E_\psi\left[E_\psi\left\{-2\log L_Y(\hat{\boldsymbol{\psi}})|\boldsymbol{Z}\right\}\right] = E_\psi\left[-2\log L_Y(\hat{\boldsymbol{\psi}})\right],$$

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

$$-2\log L_Y(\hat{\boldsymbol{\psi}}) = -2\log L_Z(\hat{\boldsymbol{\psi}}) + \hat{\sigma}^{-2}S_Y(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}}) - n, \tag{6}$$

where $S_Y(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta}) = (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol\beta})'\hat{\boldsymbol\Omega}^{-1}(\hat{\boldsymbol\theta})(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol\beta})$. The goal is find an unbiased approximation for $E_\psi\left[\hat\sigma^{-2}S_Y(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})\right]$ of equation (6).

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

$$\begin{aligned} S_Y(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta}) \simeq S_Y(\boldsymbol\beta,\boldsymbol\theta) &+ \Big((\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})-(\boldsymbol\beta,\boldsymbol\theta)\Big)'\frac{\partial S_Y(\boldsymbol\beta,\boldsymbol\theta)}{\partial(\boldsymbol\beta,\boldsymbol\theta)} \\ &+\frac{1}{2}\Big((\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})-(\boldsymbol\beta,\boldsymbol\theta)\Big)'\frac{\partial^2 S_Y(\boldsymbol\beta,\boldsymbol\theta)}{\partial(\boldsymbol\beta,\boldsymbol\theta)\partial(\boldsymbol\beta,\boldsymbol\theta)'}\Big((\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})-(\boldsymbol\beta,\boldsymbol\theta)\Big). \end{aligned} \tag{7}$$

To evaluate the expected value of the terms in (7), 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

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

(iii) for large n, $n\hat\sigma^2 = S_Z(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})$ 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 $\boldsymbol{Y}$ and $\boldsymbol{Z}$, we find that

$$\begin{aligned} E_\psi S_Y(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta}) &\simeq E_\psi S_Y(\boldsymbol\beta,\boldsymbol\theta) + \Big((\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})-(\boldsymbol\beta,\boldsymbol\theta)\Big)'\Big\{V(\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})\Big\}^{-1}\Big((\hat{\boldsymbol\beta},\hat{\boldsymbol\theta})-(\boldsymbol\beta,\boldsymbol\theta)\Big) \\ &\simeq \sigma^2 n + \sigma^2(p+k). \end{aligned}$$

Hence, from the last two terms of (6), we have

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

The quantity

$$AICC = -2\log L_Z(\hat{\boldsymbol\psi}) + 2n\frac{p+k+1}{n-p-k-2} \tag{8}$$

is an approximately unbiased estimate of the expected Kullback-Leibler information evaluated at $\hat{\boldsymbol\psi}$. This version is known as the corrected AIC (AICC) 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

$$AIC = -2\log L_Z(\hat{\boldsymbol\psi}) + 2(p+k+1).$$

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

The argument given above for AICC relied on the validity of standard asymptotic theory for the maximum likelihood estimates of the parameters in the spatial model (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.

The principle of AIC is to select a combination of explanatory variables and models for the covariance function which minimize either AICC or AIC. It is worth remarking that in many classical situations, such as linear regression or time series modeling, AICC 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 AICC 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.

3.2 Spatial Model Fitting

Traditionally, the fitting of the model (1) is accomplished in two steps (see, for example, Venables and Ripley (1999), p. 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 (AICC) (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 is implemented under the working assumption of independence of the residuals (Cressie, 1993; Haining, 1990).

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 which 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 Section 4.

Model selection techniques for spatial models need to include the correlation structure in determining the best set of predictors. By computing the AICC statistic described in Section 3.1 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 with small scale variability as modeled through the correlation in the residuals.

3.3 Other considerations

In Section 3.1 the derivation of the AICC statistic for the geostatistical model (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 AICC 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.*,

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

Here $L(\text{``}fitted\ model\text{''})$ might be interpreted as the code length of the model parameters and $L(\text{``}data\ given\ fitted\ model\text{''})$ 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.

Roughly speaking, the code length of the fitted model based on the MLE, $\hat{\boldsymbol{\psi}}$, can be approximated by $L(\text{“}\textit{fitted model}\text{''}) \simeq \frac{1}{2}(p+k+1)\log_2 n$. 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

$$\text{MDL} = \frac{1}{2}\left(-2\log(L_Z(\hat{\boldsymbol{\psi}})) + \log(n)(p+k+1)\right).$$

The only difference between the value of AICC (using the spatial AICC method) and 2·MDL is the magnitude of the penalty term coefficient. For AICC, the leading coefficient is of order 2 compared to $\log(n)$ for 2·MDL. For sample sizes greater than 8, 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 (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.

4 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 described in Section 3.2. 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 Figure 2, 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.

Table 1: Model Selection Results for the Random Pattern. Independent AICC, Spatial AICC, and MDL report the percentage of simulations that each model was selected. Of the 32 possible models, the results given here include only those with 10% or more support for one of the models.

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

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

For a given sampling pattern, the data were simulated from the model

$$\boldsymbol{Z} = 2 + 0.75\boldsymbol{X}_1 + 0.50\boldsymbol{X}_2 + 0.25\boldsymbol{X}_3 + \boldsymbol{\delta}, \tag{9}$$

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 to 1000 replicates were simulated with a new Gaussian random field generated for each replication.

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 AICC statistic for all 32 possible models. For the traditional method, the AICC statistic was calculated using (8) with $k = 0$. We call this the independent AICC approach. The spatial AICC results were calculated using (8) as well with $k = 2$. More details on the simulation set-up and additional simulation results are given in Thompson (2001).

4.1 General Simulation Results

Table 1 compares the models selected by the spatial AICC and independent AICC approaches. When independence is assumed, the AICC statistic selects the true model ($\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$) 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 AICC independence approach selected models that included both explanatory variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ 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 AICC has superior model selection performance as compared to the independence 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 AICC 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.

Figure 1 illustrates the necessity of including spatial correlation during model selection. The first panel lists the models from smallest to largest *average* AICC 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 AICC is the *true* model $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$. All of the first 16 models listed include X_1, while the first eight models also include 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 ($\boldsymbol{X}_1$ and $\boldsymbol{X}_2$) 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 AICC strategy, we performed additional simulations using different *true* values of the Matern correlation function parameters. The first experiment fixed the range parameter, $\theta_1 = 4$, and varied the smoothness parameter, $\theta_2 = (0.50, 0.75, 1.00, 4.00)$. For the second experiment, the smoothness parameter was fixed, $\theta_2 = 1.00$, and the range parameter was varied, $\theta_1 = (2, 4, 6, 8)$. For each case, 100 sets of random observations were generated according to (1) using the same sampling locations in the previous simulation study, and the two model selection techniques were compared.

Table 2 illustrates how varying the smoothness parameter, θ_2, influenced model selection. As θ_2 was increased from 0.5, which corresponds to an exponential autocorrelation function, the AIC for the spatial model approach tended to pick the *true* model more frequently. In addition, the tendency of over fitting was enhanced. Note that this second result is not undesirable. In contrast,

Table 2: Model Selection Results for Matern Family with Varied Smoothness Parameters. For all simulations the range parameter is fixed at $\theta_1 = 4$. Listed is the percentage of simulations that each model was selected using the Spatial AICC and Independent AICC methods. Of the 32 possible models, the results given here include only those with 10% or more support for one of the models.

Variables in Model	$\theta_2 = 0.50$		$\theta_2 = 0.75$		$\theta_2 = 1.00$		$\theta_2 = 4.00$	
	spat	ind	spat	ind	spat	ind	spat	ind
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$	15	3	35	1	56	2	62	3
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_5$	1	0	3	0	14	0	14	0
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4$	3	0	7	0	11	0	18	1
$\boldsymbol{X}_1, \boldsymbol{X}_2$	22	7	20	4	10	8	0	8
Intercept only	4	30	0	32	0	27	0	22
$\boldsymbol{X}_1$	17	17	12	20	0	14	0	20
$\boldsymbol{X}_2$	4	12	0	11	0	14	0	8

assuming independence during model selection tended to lead to under fitting, often with one or no explanatory variables selected. In fact, the traditional approach appeared to be invariant over all values of θ_2. Table 3 shows similar results when the range parameter, θ_1, was varied. Inclusion of spatial correlation lead to the correct model or an over fit being selected as θ_1 was increased while ignoring spatial dependence during model selection tended to under fit the model (often the intercept-only model was selected).

4.2 Impact of Sampling Pattern

The advantages of using spatial AICC 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 Figure 2. The models selected using spatial AICC for the five sampling patterns are given in Table 4. The highly and lightly clustered patterns select the true model in over 65% of the simulations. For this simulation set-up, 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 the correct model was only selected in 16% of the simulations.

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

For these simulations, the AICC independence method tends to select models with very few

Table 3: Model Selection Results for Matern Family with Varied Range Parameters. For all simulations the range parameter is fixed at $\theta_2 = 1.00$. Listed is the percentage of simulations that each model was selected using the Spatial AICC and Independent AICC methods. Of the 32 possible models, the results given here include only those with 10% or more support for one of the models.

Variables in Model	$\theta_1 = 2$		$\theta_1 = 4$		$\theta_1 = 6$		$\theta_1 = 8$	
	spat	ind	spat	ind	spat	ind	spat	ind
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$	25	2	56	2	66	5	71	4
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_5$	6	0	14	0	11	0	10	0
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4$	7	0	11	0	19	2	14	1
$\boldsymbol{X}_1, \boldsymbol{X}_2$	22	8	10	8	0	14	0	21
Intercept only	0	22	0	27	0	25	0	12
$\boldsymbol{X}_1$	14	20	0	14	0	15	0	23
$\boldsymbol{X}_2$	1	8	0	14	0	11	0	9

Table 4: Spatial AICC Model Selection Results for Five Different Sampling Patterns. Each column reports the percentage of simulations that 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.

Variables in Model	Highly Clustered	Lightly Clustered	Random	Regular Pattern	Grid Design
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$	73	65	46	43	16
$\boldsymbol{X}_1, \boldsymbol{X}_2$	0	2	18	21	35
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4$	12	13	8	8	3
$\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_5$	10	13	11	7	7

explanatory variables, and does a poor job of selecting models that contain the true parameters. The spatial AICC method does very well in selecting the true model, over a variety of sampling designs. The spatial AICC 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 Figure 2.

4.3 Prediction

To evaluate the predictive performance of including spatial correlation in the explanatory variable selection process, we introduce the mean square prediction error (MSPE). MSPE is a measure of average squared difference between the actual and predicted values at a series of locations such that

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

Here $\hat{Z}_j$ is the *universal kriging* predictor for the j^{th} 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.

First, 100 locations were randomly selected over the 10×10 grid. For each simulation in Section 4.1 a new set of observations was generated over the new grid using (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. Figure 3 illustrates the improvement made by incorporating spatial correlation into the model selection process. The mean MSPE for the spatial AICC method was 4.57 compared to 5.50 for the independent AICC 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, the mean MSPE was 39.6.

5 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. Figure 4 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 1.374×10^{11} total models. To make the analysis tractable, the number of explanatory variables was reduced to six. See Thompson (2001) for further details about preliminary explanatory variable selection.

The subset of explanatory variables used in the analysis were Crematogaster ant abundance (3 categories - low, medium, and high), log percent sandy soils, elevation, an bare rock indicator, percent cover, and log percent chapparal plants. This reduced the total number of possible models from approximately 2^{37} to 160 unique models. Note that the presence of categorical variables augments the total number of possible models in a non-trivial manner.

All 160 unique models were fit to the data using the strategy outlined in Section 3.2. 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 5 summarizes the top 3 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 AICC 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 AICC method.

6 Software

Software to perform the model selection strategy for geostatistical models described in Section 3.1 is available at www.stat.colostate.edu/∼jah and at Statlib (lib.stat.cmu.edu). The software is compatible with both S-plus and R statistical packages, the latter which is also freely available at Statlib. The software implements the Matern covariance function (2), but other covariance functions can easily be adopted.

7 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

Table 5: Model selection results for the whiptail lizard data set. Listed are 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.

Predictors	AICC	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, barerock, % chaparral	92.0	41	1
Ant_1, Ant_2, % sand, elevation, barerock, % chaparral	95.3	33	2
Ant_1, % sand, % cover, elevation, barerock, % chaparral	95.6	38	3

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

Appendix

For completeness, we present the working equations used for computing the corrected AIC (AICC) and the minimum description length (MDL) for the variable selection methods described in Section 3. All three strategies (spatial AICC, independent AICC, and MDL) require the evaluation of the log-likelihood. For the observed data $\boldsymbol{Z}$ generated by the random field defined by (1), the likelihood equation is

$$L_Z(\boldsymbol{\psi}) = \frac{(2\pi)^{-n/2}}{|\sigma^2\boldsymbol{\Omega}|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Z}-\boldsymbol{X\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X\beta})\right\},$$

where $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)'$. Thus the log-likelihood is

$$\log(L_Z(\boldsymbol{\psi})) \propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma^2}(\boldsymbol{Z}-\boldsymbol{X\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{Z}-\boldsymbol{X\beta}).$$

For the spatial AICC variable selection method, AICC $= -2\log(L_Z(\boldsymbol{\psi})) + 2n(p+k+1)(n-p-k-2)^{-1}$. Replacing $\boldsymbol{\psi}$ with the maximum likelihood estimate $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2)'$ leads to

$$\text{AICC}_{spat} = n\log\hat{\sigma}^2 + \log|\hat{\boldsymbol{\Omega}}| + n + 2n\frac{(p+k+1)}{(n-p-k-2)},$$

where $\hat{\sigma}^2 = (\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/n$. Under the independence method, computation of AICC is simplified because, by assumption, we assume that there is no correlation between observations, *i.e.*, $\boldsymbol{\Omega} \equiv I_n$ and $k \equiv 0$. Therefore,

$$\text{AICC}_{ind} = n\log\hat{\sigma}^2 + n + 2n\frac{(p+1)}{(n-p-2)}.$$

Finally, MDL $= -\log(L_Z(\boldsymbol{\psi})) + \frac{1}{2}\log(n)(p+k+1)$. Replacing $\boldsymbol{\psi}$ with $\hat{\boldsymbol{\psi}}$ and multiplying by 2, we have

$$2 \times \text{MDL} = n\log\hat{\sigma}^2 + \log|\hat{\boldsymbol{\Omega}}| + n + \log(n)(p+k+1).$$

References

Abramowitz, M. and Stegun, I. A., editors (1965). *Handbook of Mathematical Functions*. Dover: New York.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrox, B. and Caski, F., editors, *Second International Symposium on Information Theory*, page 267.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag.

Burnham, K. P. and Anderson, D. R. (1998). *Model Selection and Inference A Practical Information Theoretic Approach*. Springer: New York.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Inference A Practical Information Theoretic Approach*. Springer: New York, 2nd edition.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, revised edition*. Wiley: New York.

Haining, R. (1990). *Spatial data analysis in the social and enviromental sciences*. Cambridge University Press: Cambridge.

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

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

Hollander, A. D., Davis, F. W., and Stoms, D. M. (1994). Hierarchical representations of species distributions using maps, images and signting data. In Miller, R. I., editor, *Mapping the diversity of Nature*. Chapman and Hall: London.

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

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

McQuarrie, A. D. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific: New Jersey.

Smith, R. L. (2000). Spatial statistics in environmental science. In *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press.

Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer: New York.

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*. PhD thesis, Colorado State University.

Venables, W. N. and Ripley, B. D. (1999). *Statistics and Computing*. Springer: New York, third edition.

Ver Hoef, J. M., Cressie, N., Fisher, R. N., and Case, T. J. (2001). Uncertainty and spatial linear models for ecological data. In Hunsaker, C., Goodchild, M., Friedl, M., and Case, T., editors, *Spatial Uncertainty for Ecology: Implications for Remote Sensing and GIS Applications*, pages 214–237. Springer-Verlag, New York, NY.

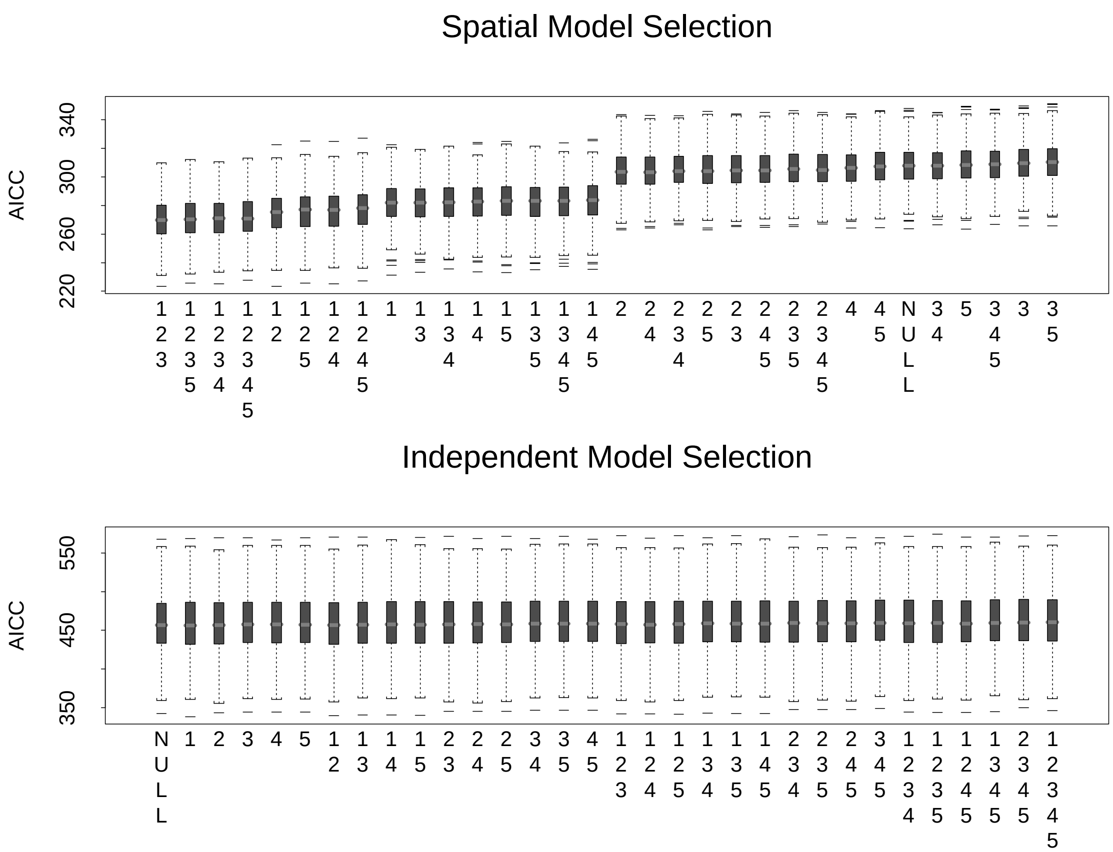

Figure 1: AICC values for the spatial AICC and independent AICC selection strategies. Note that the models for the spatial AICC method have been ordered from smallest to largest *average* AICC over all 500 simulations. The horizontal axis lists the variables included in each model where NULL refers to the intercept-only model.

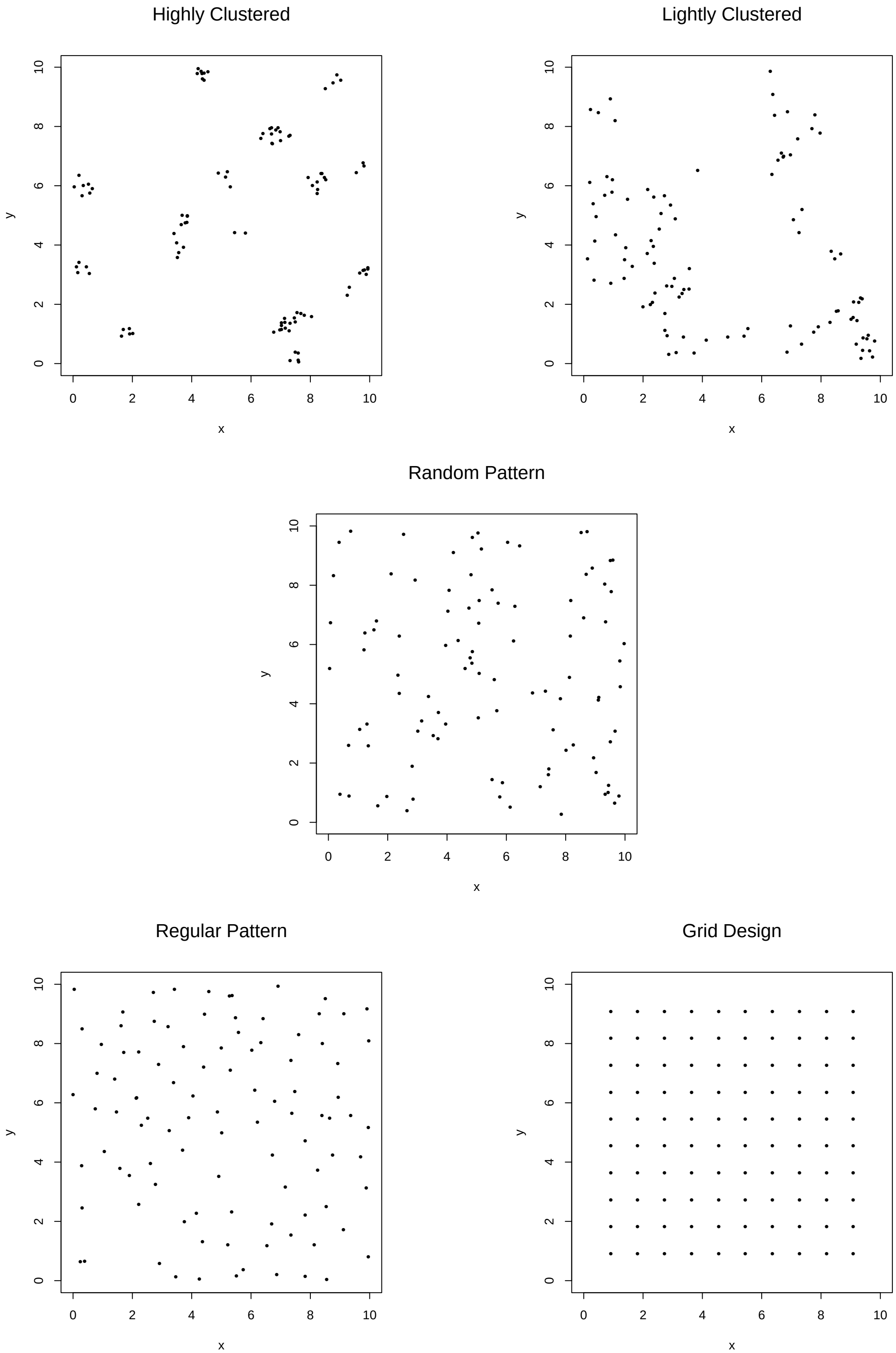

Figure 2: Five Sampling Patterns

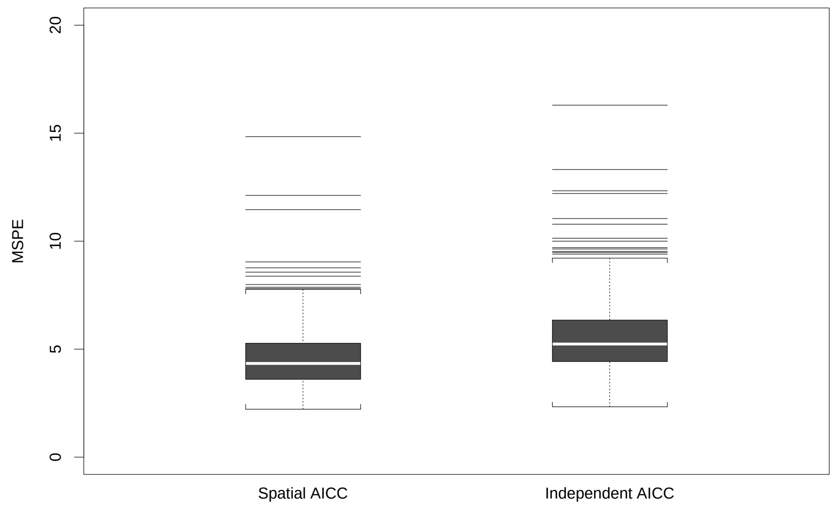

Figure 3: Mean Squared Prediction Error (MSPE) for the two model selection methods based on 500 simulations.

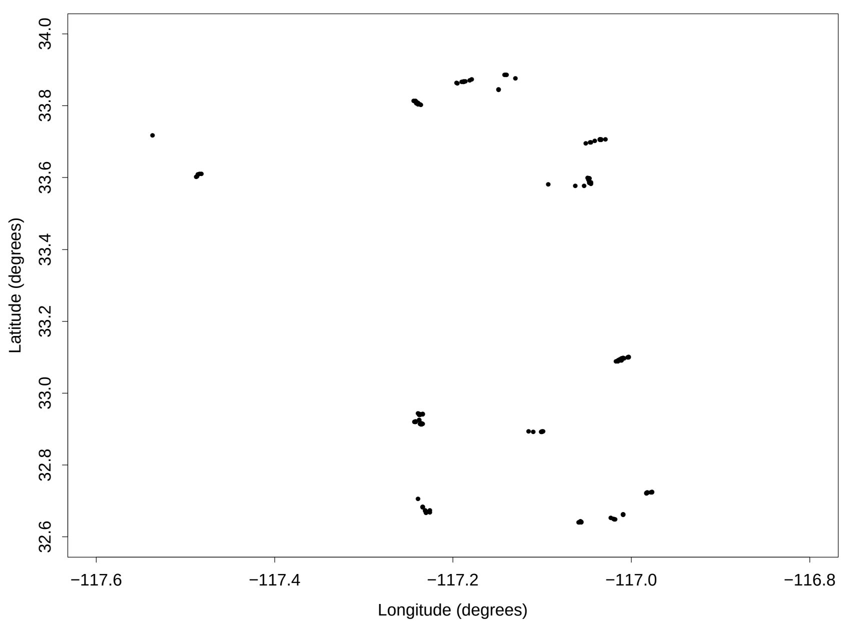

Figure 4: Locations in southern California where the whiptail lizard was observed ($n = 148$).