

MODELING SPECIES–HABITAT RELATIONSHIPS WITH SPATIALLY AUTOCORRELATED OBSERVATION DATA

B. A. WINTLE^{1,3} AND D. C. BARDOS²

¹*Environmental Science Laboratory, School of Botany, University of Melbourne, Victoria 3010, Australia*

²*School of Physics, University of Melbourne, Victoria 3010, Australia*

Abstract. Spatial autocorrelation in wildlife observation data arises when extrinsic environmental processes and patterns that influence the spatial distribution of wildlife are themselves spatially structured, or when species are subject to intrinsic population processes, causing contagion or dispersion effects. Territoriality, Allee effects, dispersal limitations, and social clustering are examples of intrinsic processes. Both forms of autocorrelation can violate the assumptions of generalized linear regression models, resulting in biased estimation of model coefficients and diminished predictive performance. Such consequences may be avoided for extrinsic autocorrelation when autocorrelated environmental variables are available for use as model covariates, whereas intrinsic spatial autocorrelation requires an alternative modeling approach. The autologistic model provides an approach suited to the binary observations often obtained in wildlife surveys, but its performance has not been tested across widely varying sampling intensities or strengths of intrinsic spatial structure.

Here we use simulated data to test the autologistic model under a range of sampling conditions. The autologistic model obtains better fits and substantially better predictive performance than the standard logistic regression model over the full range of sampling designs and intensities tested. We provide a simple Bayesian implementation of the autologistic model, which until now has not been achieved with standard statistical software alone. A step-by-step procedure is given for characterizing and modeling spatial autocorrelation in binary observation data, along with computer code for fitting autologistic models in WinBUGS, a freeware Bayesian analysis package. This approach avoids normal approximations to the pseudo-likelihood, in contrast to previous Bayesian applications of the autologistic model. We provide a sample application of the autologistic model, fitted to survey data for a gliding marsupial in southeastern Australia.

Key words: *autologistic; greater glider; habitat modeling; logistic regression; Petauroides australis; Petauroides volans; spatial autocorrelation; virtual glider; WinBUGS; yellow-bellied glider.*

INTRODUCTION

Autocorrelation in the spatial distribution of wildlife may arise from community (or demographic) processes such as territoriality, competition, and constraints on dispersal. This may be viewed as “intrinsic” spatial structure (Anselin 1992) because its form is specific to the species or the population of interest. Alternatively, environmental “forcing” (Legendre 1993) is a form of “extrinsic” spatial structuring, whereby the environmental variables that determine the “niche” of a species are themselves spatially structured. Wildlife–habitat regression models (usually linear or generalized linear) are commonly used to describe extrinsic processes; such models have utility in conservation planning because they allow prediction of likely spatial distributions of a species across unsurveyed parts of the landscape (Pereira and Itami 1991, Guisan and Zimmerman 2000, Ferrier et al. 2002). Community and population spatial processes are

usually ignored in model building because they are difficult to characterize and incorporate, although some are considered in specific ecological models (Hanski 1992, Lindenmayer et al. 1999, Ferrier et al. 2002, Moilanen and Nieminen 2002) that are yet to be integrated into standard habitat modeling methods. Failure to properly account for either form of spatial structuring when constructing species–habitat regression models brings about non-independence in residual errors, violating a basic assumption of generalized linear models (GLMs) (McCullagh and Nelder 1989). This results in biased standard errors (Johnston 1972) and parameter estimates, resulting in diminished predictive performance (Legendre 1993). In practice, many (or perhaps most) applications of GLMs in habitat analyses breach the independent-errors assumption, because demographic or community processes are ignored in model building or because spatially structured habitat variables are not included in the model. Consequently, these habitat models are likely to provide biased inference and prediction.

Many methods for modeling autocorrelated data are restricted to continuous rather than categorical dependent variables. This derives from the mathematical

Manuscript received 9 November 2004; revised 21 October 2005; accepted 23 December 2005; final version received 9 March 2006. Corresponding Editor: N. T. Hobbs.

³ E-mail: brendanw@unimelb.edu.au

intractability of likelihood functions for discrete-variable auto-models, in contrast to that of, for example, the auto-normal model, for which a maximum likelihood approach is relatively straightforward (Besag 1974). Wildlife observation data are necessarily categorical: either binary (presence/absence) or abundance (number of individuals at a given location). This paper focuses on binary data, common in wildlife-habitat studies. Binary wildlife habitat models are easily fitted using logistic regression (McCullagh and Nelder 1989) and may be used to create habitat maps. Like all linear and generalized linear models, logistic regression assumes independence of errors, and is thus incompatible with spatially autocorrelated observations. Of at least five possible general approaches for modeling autocorrelation in binary data (see next section), many remain largely untested and examples of practical implementations are difficult to find. The most commonly applied is the autologistic model pioneered by Besag (1972, 1974). Standard maximum likelihood estimation (MLE) methods used in logistic regression are untenable for autologistic models except for very small data sets, although MLE for the autologistic model has been achieved using Markov Chain Monte Carlo (MCMC) methods (Huffer and Wu 1998). However, the method is numerically unstable, particularly for strong intrinsic spatial autocorrelation (Huffer and Wu 1998), and no reliable cross-platform implementation appears to be available. Feasible approximations have been sought using the "maximum pseudo-likelihood" estimation (MPLE) method (Besag 1975, Wolfinger and O'Connell 1993), and the penalized quasi-likelihood method (Breslow and Clayton 1993), of which the MPLE is most commonly used in biological examples. Bayesian implementations using the full autologistic likelihood (Heikkinen and Hogmander 1994, Hoeting et al. 2000) have relied on normal approximations to the pseudo-likelihood for conditional distributions of model parameters (see Ryden and Titterton (1998) for an evaluation of this approach). Although the autologistic MPLE performs better than standard logistic regression under certain conditions (Augustin et al. 1996), the method is likely to be unstable when data exhibit strong intrinsic spatial structure (Wu and Huffer 1997).

Biological field surveys often employ randomization and spatial stratification in sampling design to avoid biases induced by unmeasured influences, including spatial autocorrelation (Legendre and Fortin 1989). This provides "representative" samples of the important environmental gradients influencing dependent variables of interest, e.g., the probability of occupancy by a species (Austin and Heyligers 1989). Most statistical methods (e.g., GLMs) require the biologist to obtain a set of "independent" samples (De Grijter and Ter-Braak 1990, Legendre 1993). Indeed, if inference about the relationship between the probability of occupancy by a species and a set of habitat variables is the primary objective of data collection, then it is reasonable to

attempt a minimally spatially autocorrelated sampling design. However, a spatially uniform sample (i.e., where observations are separated by some minimum distance) may *not* be useful if the biologist wishes to characterize the spatial structure in a species distribution, including fine-scale structure due to community processes. If such processes are thought to be acting on individuals in a population, and the goal is to improve prediction of species occupancy at unsurveyed locations, we might consider a sampling design that stratifies sites across environmental gradients (Austin and Heyligers 1989) as well as ensuring a sufficient number of short inter-site separation distances. This would enable fine-scale intrinsic spatial structure to be analyzed.

In reviewing the literature, we noted four important issues that motivated the work presented here. (1) A lack of clear guidance on correct procedures for fitting the full autologistic model and for determining an appropriate neighborhood over which spatial autocorrelation should be modeled partly explains the lack of examples of its use in the ecological literature. (2) Until the advent of WinBUGS, no standard statistical packages were capable of implementing this model. (3) The robustness of the autologistic model to a variety of strengths of spatial autocorrelation in wildlife observation data and to a range of survey sampling intensities has not been tested. (4) The utility of random survey sampling for estimating the autologistic model (and characterizing fine-scale spatial processes) remains untested against alternative strategies that ensure a range of spatial separations between sites. Many of these issues are best addressed by simulation, as it is likely to be difficult, if not impossible, to find "real" data of known spatial structure with which to test such issues.

Here, we will detail the technical issues associated with the autologistic model and will then introduce a simulated "virtual glider" distribution. This distribution will be used throughout the paper to demonstrate and test the model under a range of scenarios, with varying underlying spatial autocorrelation and a range of survey sampling intensities and designs. We will demonstrate that WinBUGS, a freely available interface for Bayesian inference using Gibbs sampling, can be used to simultaneously estimate the autologistic model with covariates and make predictions about unsurveyed parts of the virtual glider distribution. In concluding, we discuss a synthesis of the results obtained from simulations, develop a set of recommendations for implementing the autologistic model, and identify some future research imperatives. A sample application of the autologistic model, fitted to survey data for a gliding marsupial in southeastern Australia, is provided (Appendix D).

THEORETICAL BACKGROUND: THE AUTOLOGISTIC MODEL

The logistic regression model is straightforward to fit using standard maximum likelihood methods (Collett 1999), due to its assumption of spatial independence.

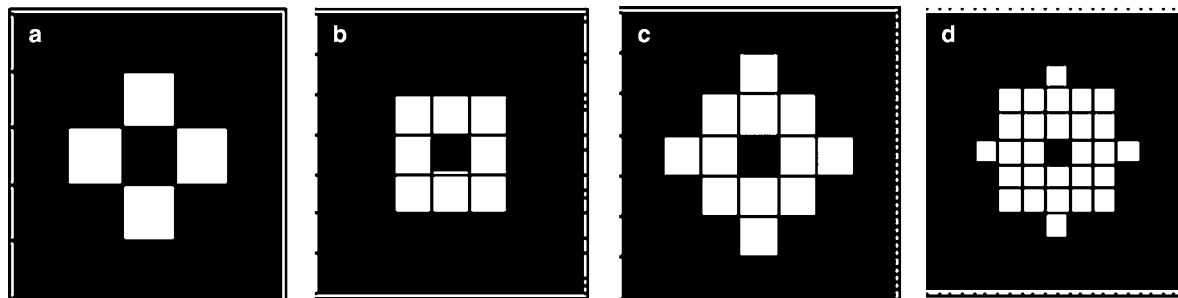


FIG. 1. The form of the (a) first-order, (b) second-order, (c) third-order, and (d) radius-3 neighborhoods used in the autologistic model. The white cells represent those included in the neighborhood of the central (black) cell.

Each site i has an associated binary observation y_i , a true (but unknown) probability of success p_i , and a covariate vector \mathbf{X}_i . Each p_i is independent of all other cells, depending only on the value of the covariates at cell i , such that

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \boldsymbol{\beta}_{\text{cov}} \cdot \mathbf{X}_i \quad (1)$$

(Collett 1999), where α is a regression parameter and $\boldsymbol{\beta}_{\text{cov}}$ is a vector of regression parameters associated with the covariates. The drawback of this simplicity is that no spatial structure is incorporated in the model aside from that of the covariates. The spatial arrangement of the sampled sites is irrelevant. Thus, in the logistic model, the sites can simply be treated as a list:

$$\{\{y_1, p_1\}, \{y_2, p_2\}, \dots, \{y_i, p_i\}, \dots, \{y_N, p_N\}\} \quad (2)$$

that can be arbitrarily reordered without altering the logistic fit. The likelihood for N sites is simply the product of the likelihoods for each cell:

$$\prod_{i=1}^N p_i^{y_i} (1-p_i)^{1-y_i}. \quad (3)$$

The autologistic model, in common with other auto-models (Besag 1974), relaxes the independent-sites restriction, allowing for more realistic modeling of spatial processes with the expense of drastically greater mathematical and computational complexity.

An *autocovariate* at site i is defined as a weighted sum of observations in neighboring sites:

$$\text{autocov}_i = \sum_{j \in N_i} w_{ij} y_j \quad (4)$$

where N_i defines the *neighborhood* region around cell i , and the weights must obey $w_{ij} = w_{ji}$. Because N_i defines those cells that are called *neighbors* of cell i , it is required that if N_i contains cell j , then N_j contains cell i . This mutuality relationship between neighborhoods is easily seen to hold in the most commonly used arrangements (Fig. 1), referred to as nearest neighbors (first-order neighborhood) or next-to-nearest neighbors (second-

order neighborhood), and the like. Any neighborhood definition based on intercell distance will be suitable. The term “radius m neighborhood” is used here to mean all cells on or within a radius of m from a given cell, where distance is measured between the cell centers. The radius-3 neighborhood is illustrated in Fig. 1.

The autocovariate term is added to the logistic regression for p_i , multiplied by an *autocovariate coefficient* β_{auto} , resulting in the *autologistic model* (Besag 1972, 1974):

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \boldsymbol{\beta}_{\text{cov}} \cdot \mathbf{X}_i + \beta_{\text{auto}} \text{autocov}_i \quad (5)$$

which is closely related to the Ising model in statistical physics (Besag 1972). The autocovariate coefficient is then fitted in conjunction with the usual covariate coefficients. The process of carrying out this parameter estimation is much more difficult than in the logistic case.

Models with this form of local probability structure fall within a class known as Markov Random Fields (MRFs), or sometimes more elaborately as LDMRFs (Locally Dependent MRFs). This is a generalization of the concept of a Markov chain to two or more dimensions, which are, for the purposes of this study, spatial dimensions. In a “first-order” Markov chain, the distribution of states x_n at step n depends on the states at all other points solely through the states x_{n-1} and x_{n+1} (Geman and Geman 1984). Analogously, in a first-order MRF, the state at site i depends only on the states of sites “one step” away. In the two-dimensional case, for a square lattice, this means the four nearest neighbors.

The full likelihood for the autologistic model, i.e., the joint probability distribution for all observations y_i of presence or absence, is known analytically only to within a normalization constant (Huffer and Wu 1998):

$$\Pr(y) = \frac{\exp\left(\alpha \sum_i y_i + \boldsymbol{\beta}_{\text{cov}} \cdot \sum_i \mathbf{X}_i y_i + \frac{1}{2} \beta_{\text{auto}} \sum_i y_i \text{autocov}_i\right)}{k(\alpha, \boldsymbol{\beta}_{\text{cov}}, \beta_{\text{auto}})} \quad (6)$$

It is the “constant”

$$k(\alpha, \beta_{\text{cov}}, \beta_{\text{auto}}) = \sum_{\{y_i\}} \exp\left(\alpha \sum_i y_i + \beta_{\text{cov}} \cdot \sum_i \mathbf{X}_i y_i + \frac{1}{2} \beta_{\text{auto}} \sum_i y_i \text{autocov}_i\right) \quad (7)$$

that is the source of difficulty in fitting the model, because it is an analytically intractable function of the regression parameters.

This motivated the early development (Besag 1975) and continuing application (Augustin et al. 1996) of an approximation scheme known as the pseudo-likelihood method, in which the autocovariate term is included in the regression, but the sites are assumed to be independent for the purpose of constructing the likelihood function. That is, the simple product form of the likelihood is used when fitting the autologistic model, allowing standard logistic fitting software to be used. The estimate of model parameters so obtained is known as the Maximum Pseudo-Likelihood Estimate (MPLE). Evidently, this procedure cannot entirely capture the full intrinsic spatial structure of the model, and is therefore expected to lose accuracy with increasing intrinsic spatial autocorrelation (i.e., as the true underlying value of β_{auto} increases). Inaccuracy in the MPLE has been observed by Geyer (1991), Geyer and Thompson (1992), and Wu and Huffer (1997); results from the latter indicate that the MPLE tends to overestimate intrinsic autocorrelation in data containing strong intrinsic structure.

The Maximum Likelihood Estimation (MLE) approach used to fit the logistic model (Eq. 1), cannot be immediately applied in the autologistic case due to the intractability of the likelihood (i.e., the distribution of the data, conditional on the parameters), as discussed above. The Markov Chain Monte Carlo (MCMC) technique was developed in statistical physics (Metropolis et al. 1953) to deal with difficult high-dimensional integration problems, of which a non-normalized distribution is an example.

MCMC methods have been applied to the full autologistic model by Huffer and Wu (1998), employing the MCMC MLE method developed by Geyer and Thompson (1992). The essence of this method is that trial starting values α^s , β_{cov}^s , and β_{auto}^s are chosen for the regression parameters. MCMC sampling of the distribution proportional to the likelihood is then carried out using Gibbs sampling in the outcome space $\{y\}$, conditional on the trial parameters. The resulting sample, in conjunction with the vector of observed responses $\{y_{\text{obs}}\}$, enables the construction of an approximate likelihood function in the parameter space surrounding the trial vector. An MLE is then obtained for this approximate likelihood function, the location of which can then be used as a new trial state, and the process iterated. Convergence occurs if the original trial state is sufficiently close to the true MLE.

The existence of a non-normalized distribution is a familiar situation in Bayesian inference: most of the recent work on the autologistic model (e.g., Heikkinen and Hogmander 1994, Hoeting et al. 2000) has been carried out within a Bayesian framework, employing the MCMC methodology pioneered by Geman and Geman (1984). A non-normalized posterior distribution is constructed for the unknown regression parameters and any missing observations (which are themselves treated as unknown parameters). This distribution, which is conditional on the observations y_{obs} , is then sampled according to one of the various MCMC procedures, such as Gibbs sampling, the Metropolis algorithm, or more generally, the Metropolis-Hastings algorithm. All of these procedures exploit theorems guaranteeing that certain sequences of dependent samples (i.e., Markov chains), whose construction requires only ratios of the target distribution, will converge to that distribution. In the case of Bayesian analysis, the target is the posterior distribution.

An important difference between the Bayesian and MLE approaches is that, in the former case, the normalization “constant” is a function of the data actually observed, y_{obs} , but is independent of the regression parameters, since the posterior distribution is normalized with respect to them. Only ratios of the target distribution, evaluated at different parameter values, are required by MCMC algorithms; because the normalization constant is independent of these parameters, MCMC algorithms allow sampling from the non-normalized posterior distribution for the regression parameters. Gibbs sampling has the advantage that the required ratios reduce to the conditional distributions for the probability of presence at each site, which are readily available for the autologistic model (Eq. 5). These conditional distributions simply represent the probability of a given site being occupied, *conditional* on the occupancy of all other sites and all model parameters.

When Gibbs sampling is applied to the autologistic model, a Markov chain is constructed, with each element consisting of a presence/absence map and a vector of regression parameters. The chain is generated by sampling in turn from the conditional distribution for each parameter, until all the sites and regression parameters have been updated. The “sweep” is then repeated until convergence to a stationary distribution. This allows the simple MRF structure of the model to be exploited, as the conditional distributions for the p_i are available from (Eq. 5). It is important to distinguish the MRF, which is the underlying model assumed for the spatial system, from the Markov chain, which is constructed purely as a means for parameter estimation.

A practical issue arising in the case of missing data (or unsurveyed locations), which does not arise for the simple logistic model, is that the p_i depend on the autocovariates, which cannot all be evaluated without a complete observation vector y . In the Bayesian ap-

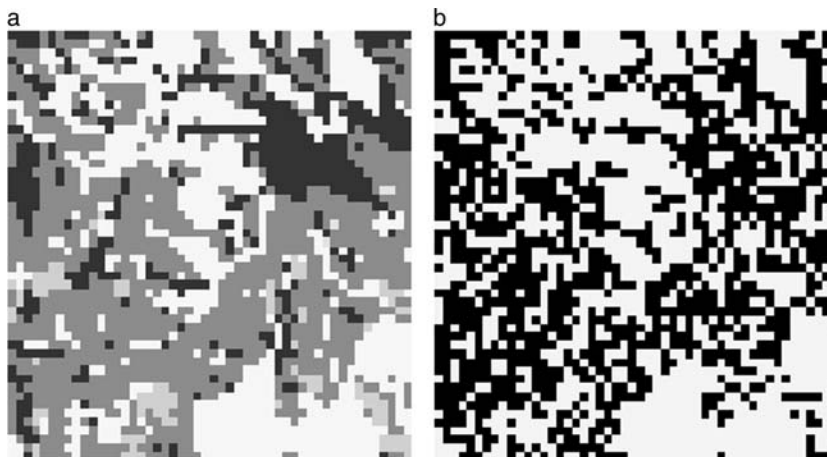


FIG. 2. (a) A virtual glider habitat of 50×50 cells, where each cell is 100×100 m. Cells of the darkest tone have probability of glider occupancy $p_i > 0.75$, the lightest tone represents $p_i < 0.25$, and there are two intermediate tones. (b) A sample realization of binary observations (presence or absence) of the glider obtained by independent Bernoulli trials from the cell probabilities.

proach, this is handled naturally via initialization of the Markov chain. All parameters, whether they are regression or missing data parameters, must be assigned starting values in order to initiate the chain. The effect of “poor” choice of starting values will be reflected in a longer burn-in phase.

In practice, if the degree of intrinsic spatial structure is moderate, it is reasonable to first fit an ordinary logistic model to the data, then sample the “missing values” from the probabilities predicted at unobserved sites, and use this sample as the initial state for the Markov chain. In the case of strong autocorrelation, it may be just as reasonable to initialize the missing values by assuming that $p_i = 0.5$ or some other arbitrary assignment. In any event, the MCMC procedures are robust with respect to the starting state.

The Gibbs sampling approach for obtaining the posterior distribution of the full autologistic model requires specialized software and is beyond the scope of standard statistical packages. The WinBUGS freeware package (Spiegelhalter et al. 2003) provides a flexible solution for Bayesian analysis and is capable of handling spatial neighborhood models (Spiegelhalter et al. 1996), of which the autologistic model is an example. A variety of neighborhood structures and weightings may be implemented, and the stability and optimality properties of the Bayes posterior estimate are available in this approach.

METHODS

The virtual glider

Model performance was evaluated against “truth,” rarely available in practical applications. Simulated observations are derived from a known “virtual glider” (sensu Hirzel et al. 2001) spatial distribution. The latter is generated by a nonparametric habitat suitability model of glider site occupancy probabilities for 2500 1-

ha cells. The model predicts occupancy probability for each cell as a function of various mapped environmental variables representing nutritional, shelter, and metabolic constraints on the species. The probabilities p_i^v of virtual glider occurrence at cells $i = 1, 2, \dots, 2500$ form a probability surface (Fig. 2a) from which binary realizations of glider presence/absence are obtained (Fig. 2b). The superscript “v” distinguishes the virtual glider probabilities from those estimated by the models presented later. Binary realizations of the probability surface, and an extension where intrinsic spatial autocorrelation is included, are used throughout the paper to test spatial modeling methods.

The process of constructing a hypothetical distribution is analogous to building a habitat suitability index model (HSI; Crance 1987, Rand and Newman 1998). A HSI model for a given species and area of land represents a conceptual model that relates each measurable variable of the environment to the suitability of a site for the species, scaled from 0 (unsuitable habitat) to 1 (optimum conditions) (Burgman et al. 2001). Each variable is represented by a single suitability index (SI). SIs are linked within the model by additive, multiplicative, or logical functions that reflect relationships among the variables (Burgman et al. 2001). In constructing a hypothetical distribution for the purpose of a simulation study, the functional relationship between the species and each predictor may be arbitrarily defined, although this approach may bias simulations toward a particular outcome (Austin et al. 1995). We opted to base the habitat requirements of our virtual glider on plausible relationships reflecting current ecological knowledge about greater glider (*Petauroides volans*) habitat preferences and behavior. These were gleaned from a combination of expert knowledge, published studies, and modeling (Kavanagh and Lambert 1990, Lindenmayer et al. 1990, 1995, National Parks and Wildlife Service 1998, Kavanagh 2000).

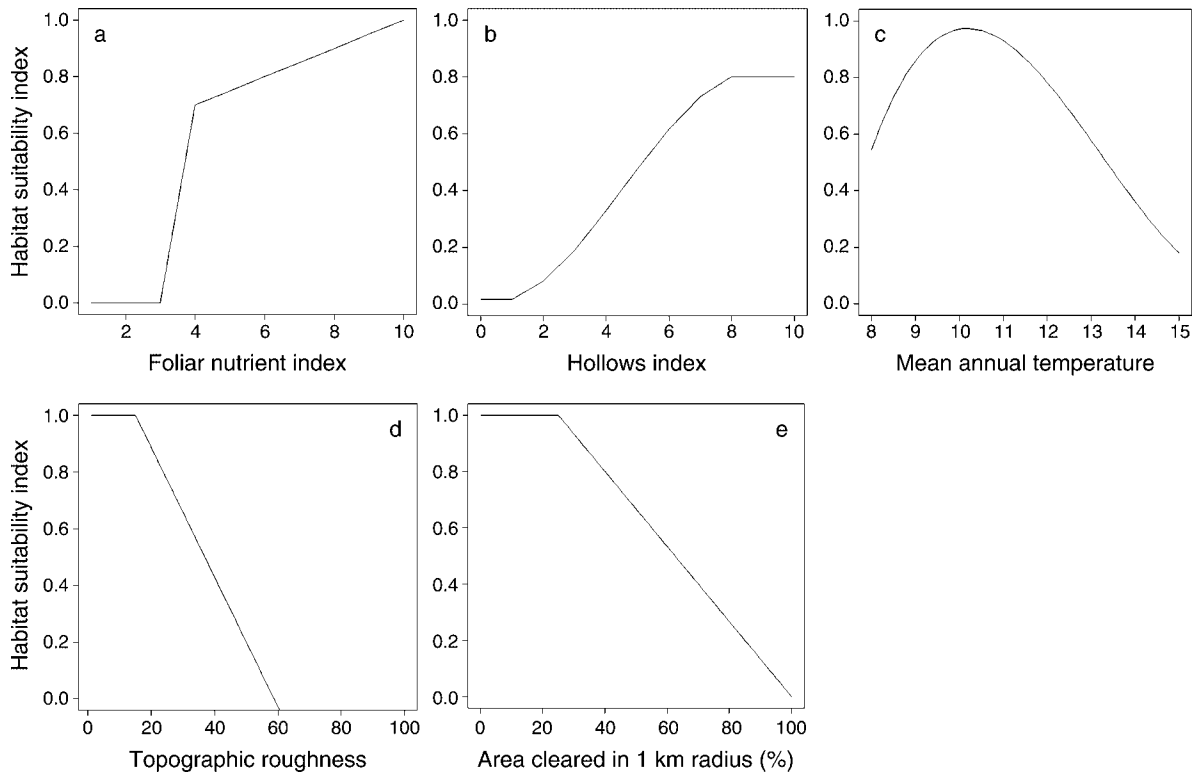


FIG. 3. Direct predictors of virtual glider habitat defined by virtual glider habitat preferences. The suitability functions show how habitat quality varies across the range of each predictor. Predictor names and weights are: (a) foliar nutrient (fol, weight = 4); (b) hollows index (holl, weight = 4); (c) mean annual temperature (temp, weight = 2); (d) topographic roughness in 250-m radius (rough, weight = 1); (e) percentage of area cleared in 1 km radius (clear, weight = 1).

Five separate environmental gradients defined the “true” habitat preferences of the virtual glider (Fig. 3), representing its denning, nutritional, thermal, and mobility constraints. Here, “fol” is a foliar-nutrient index (National Parks and Wildlife Service 1998); “holl” is a tree hollow incidence index (National Parks and Wildlife Service 1998); “temp” is average mean annual temperature derived from a bioclimatic model (Hutchinson et al. 1999), representing the metabolic constraints on the species; “clear 1 km” describes the proportion of cells in a 1 km radius that are devoid of forest cover (National Parks and Wildlife Service 1998); and “rough” is the standard deviation of the cell elevations in a 250 m radius around the cell, indicating local topography roughness that may constrain animal mobility. Gradients were based on five mapped environmental variables (Fig. 3).

Weights were assigned to each predictor to reflect its relative importance in defining the habitat of the virtual glider (Burgman and Lindenmayer 1998), and individual SIs were constructed that define the relationships between each predictor and the probability of finding a glider (Fig. 3). Two critical habitat predictors were identified, relating to nutrition and denning. These predictors were combined into a partial HSI (HSI_1), as

the weighted geometric mean of SI_{fol} for foliar nutrient and SI_{holl} for hollows. The three remaining SIs represent dependence on mean annual temperature (a physiological constraint), topographic roughness within 250 m (a physiological and mobility constraint), and the amount of clearing in 1 km (representing exposure to predation by feral animals and general impacts of anthropogenic disturbance). These were combined into partial HSI_2 as a weighted arithmetic mean. Finally, a weighted geometric mean of HSI_1 and HSI_2 was calculated to create a representation of the intrinsic habitat preferences of the virtual glider. The use of the geometric mean to combine habitat indices in our study implies that the absence of either suitable denning or nutrition at any location results in a habitat value equal to zero for that location.

This underlying model of virtual glider habitat preference was used to create a map of the probability of occurrence of the virtual glider. Computations were undertaken in ArcInfo 7.2 (ESRI 1997).

Parametric approximations to this “true” model were used as the basis for autologistic modeling of spatially autocorrelated observation data. A parametric approximation to the true generating models is the following:

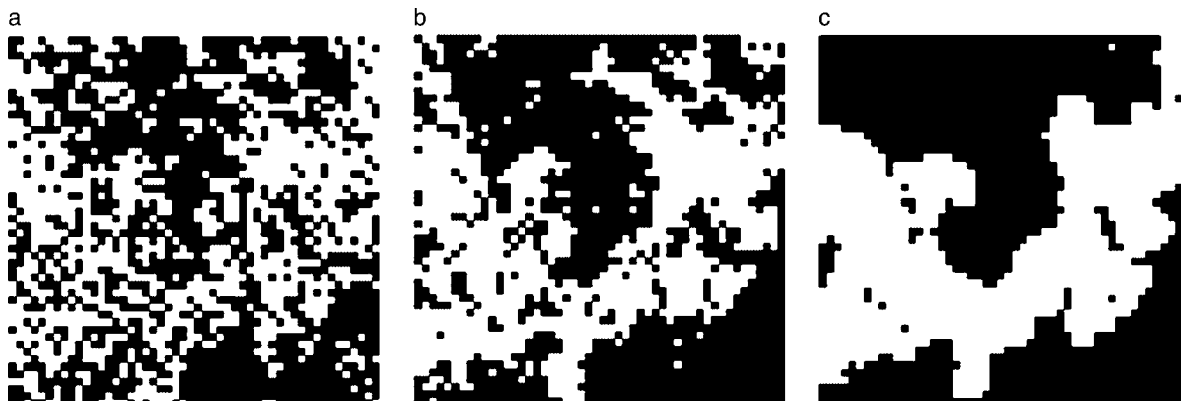


FIG. 4. Binary realizations of three virtual glider distributions representing (a) weak, (b) medium, and (c) strong intrinsic spatial autocorrelation. See Fig. 1 for a realization of the virtual glider distribution without intrinsic spatial autocorrelation. White squares indicate where the virtual glider is present in the particular realization; black squares indicate absence.

$$\begin{aligned} \Pr(Y = 1) = & \alpha + \beta_1 \times \text{fol} + \beta_2 \times \text{fol}^2 + \beta_3 \times \text{fol}^3 + \beta_4 \\ & \times \text{holl} + \beta_5 \times \text{temp} + \beta_6 \times \text{clear1km} + \beta_7 \\ & \times \text{clear1km}^2 + \beta_8 \times \text{rough} + \beta_9 \times (\text{rough})^2. \end{aligned} \tag{8}$$

The parametric approximation (Eq. 8) to the nonparametric generating model explained 87% of the variation in habitat quality values across the 5×5 km virtual glider habitat. No demographic processes or historical disturbance events were incorporated in the simulated glider data. Consequently, any spatial structure in binary observations realized from it was due solely to variables in the generating model.

Incorporating intrinsic spatial autocorrelation in the virtual glider distribution

A spatial aggregation term was incorporated in the virtual glider model to simulate spatial autocorrelation originating from intrinsic demographic or behavioral processes. The virtual glider probability surface introduced above, denoted p_i^y for cell i , gives the probability of occupancy in the absence of spatial aggregation. Spatial autocorrelation was injected with strength β_{auto} , using a third-order neighborhood (see Fig. 1) with a weighting structure that is inversely proportional to the intercell distance:

$$w_{ji} \equiv w_{j_x, j_y, i_x, i_y} = \frac{1}{\sqrt{(j_x - i_x)^2 + (j_y - i_y)^2}} \tag{9}$$

where the Cartesian cell coordinates are written explicitly. If this autocovariate term were simply added to the logistic model (Eq. 1) corresponding to p_i^y , the result for high β_{auto} would be a flooding of the scene with occupied cells. Accordingly, a compensatory term χ , constant across all cells, was subtracted from the model. The value of χ was adjusted for each β_{auto} to ensure that the average number of occupied cells occurring in realizations of the model was similar for all β_{auto} values.

A virtual glider scene with intrinsic autocorrelation strength β_{auto} was generated using a procedure comprising outer and inner iterations. A trial value was chosen for the compensatory parameter $\chi > 0$, whose adjustment constitutes the outer iterative procedure. The inner iteration generated scenes for the current value of χ . For each scene, a binary realization was first obtained from the virtual glider habitat *without* intrinsic spatial autocorrelation, using the probabilities of presence p_i^y previously introduced. The resulting presence/absence map was used to generate initial values for the autocovariates, where autocov_i was determined by Eq. 4 and the weights w_{ij} were given, in this case, by Eq. 9. A cell “ j ” was chosen at random and a new observation for that cell was sampled with probability p_j , given by

$$\log\left(\frac{p_j}{1 - p_j}\right) = \log\left(\frac{p_j^y}{1 - p_j^y}\right) + \beta_{\text{auto}} \text{autocov}_j - \chi. \tag{10}$$

The autocovariate map was updated and this Monte Carlo sampling step was repeated numerous times (e.g., 10^5), resulting in a scene characteristic of β_{auto} and χ . Numerous scenes (e.g., 10) were generated in this manner and the overall average cell occupancy rate was calculated.

The value of χ was then adjusted (increased if the occupancy rate was excessive, decreased if too few cells were occupied) and the entire procedure was repeated until the occupancy rate closely approximated that of the original virtual glider.

Any of the scenes generated in the final pass could then be used as a realization of the intrinsically spatially autocorrelated virtual glider. The realizations used here are depicted in Fig. 4, for weak ($\beta_{\text{auto}} = 0.1$, $\chi = 0.46$), medium ($\beta_{\text{auto}} = 0.7$, $\chi = 3.15$), and strong ($\beta_{\text{auto}} = 2.0$, $\chi = 8.23$) intrinsic spatial autocorrelation.

Simulating observational sampling of the virtual glider

To test the influence of data availability on the predictive performance of the standard logistic and

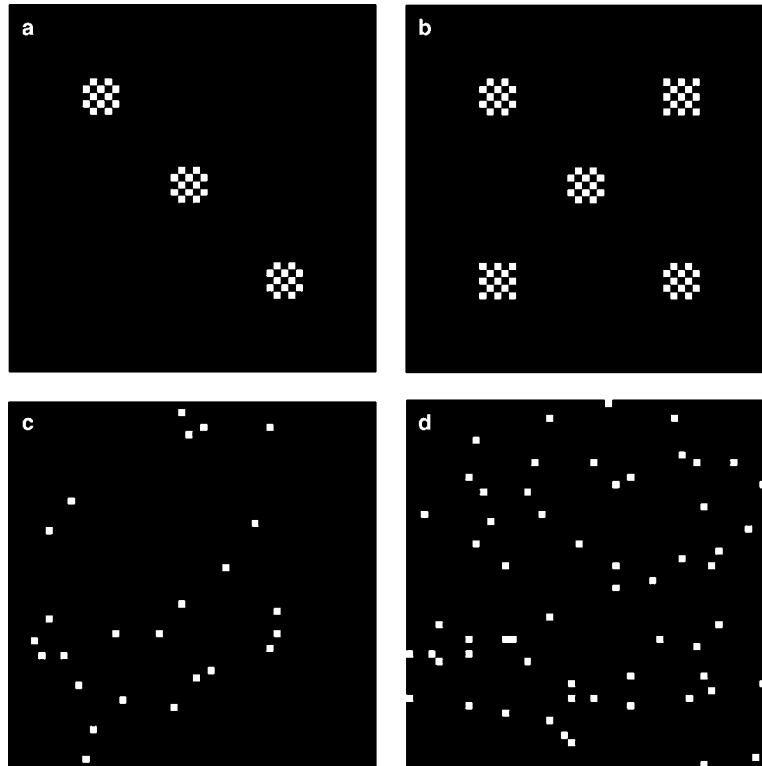


FIG. 5. A clumped sampling pattern utilized to sample (a) 1% and (b) 2% of the virtual glider distribution while sampling numerous first- and second-order neighborhood relationships. Squares in white are sampled, and the remaining squares (in black) are not sampled. Two random samples of (c) 1% and (d) 2% are provided for the sake of comparison.

autologistic models, random samples comprising 20%, 10%, 5%, 2%, and 1% of the virtual glider distribution were taken and used for model fitting. The utility of clumped sampling patterns (Fig. 5), designed to ensure that sufficient neighbors existed at all shorter lag distances in 1% and 2% sampling intensity scenarios, was also tested. Autologistic models were fitted to the data derived from clumped sampling of the virtual glider distribution at medium spatial autocorrelation ($\beta_{\text{auto}} = 0.7$).

*Creating an autocovariate for use
in the autologistic model*

The autologistic model requires the specification of an autocovariate term (see Eq. 4) with a given neighborhood size and with weights w_{ij} assigned to each neighborhood element. Larger neighborhoods require more MCMC computation to fit the model, and the designation of w_{ij} can become complicated. In fact, the exact specification of w_{ij} can have a large influence on the inference and predictions derived from autoregression models (Cressie 1991). Most autologistic model implementations have relied on first-, second-, and third-order neighborhoods (e.g., Augustin et al. 1996, Gumpertz et al. 1997, Huffer and Wu 1998, Hoeting et al. 2000). Preliminary analysis of residual spatial structure from models fitted on virtual glider data implies that most of the residual structure is limited to

lag distances of up to 400 m. This supports limiting neighborhood size to radius 3 or smaller. All neighborhoods used here are illustrated in Fig. 1.

Uniform weighting within the neighborhood implies that spatial autocorrelation is uniform across all lag distances; a difficult assumption to justify both biologically and with reference to residual correlograms (presented later). The use of residual correlograms and variograms (Legendre and Fortin 1989) for parameterizing the weights in the autocovariate term has precedents in the literature (e.g., Albert and McShane 1995, Gotway and Stroup 1995, Bio et al. 2002, Klute et al. 2002). In this study, standard logistic models were fitted to each observational sample and model residuals were used to construct correlograms. Residuals were obtained by subtracting predictions (i.e., fitted values \hat{p}_i) derived from the standard logistic regression model, from observations of the virtual glider at the randomly selected survey locations. Correlograms were examined for each survey sample to identify both an appropriate neighborhood size and the weights appropriate for all inter-site lag distances within the neighborhood. All model testing was undertaken with a third-order neighborhood.

Fitting the autologistic model in WinBUGS

WinBUGS 1.4 was used to fit standard logistic models and to develop an autologistic model of sufficient

generality to model spatially autocorrelated data on square lattices, with minimal alteration to the code for different neighborhoods. The general form of the model is given in Eqs. 4 and 5, where p_i is the probability that cell i is occupied by a glider, \mathbf{X}_i is a vector of environmental variables for the cell, and autocov_i is a weighted average of the number of occupied cells in the neighborhood of cell i . The weight w_{ij} given to cell j is proportional to the level of spatial autocorrelation observed in the correlogram of the residuals of Eq. 1 (the conventional logistic model) at lag distances h_{ij} , where h_{ij} is the Euclidean distance between cells i and j .

The full parametric approximation to the generating model (Eq. 8) was used to fit the standard logistic model. This avoids introducing structure from missing variables into residual errors and thus allows assessment of the impact of “intrinsic” spatial structure on the predictive performance of models.

The model code appearing in the Supplement corresponds to the WinBUGS Bayesian model:

$$\text{logit}(p_i) = \alpha + \beta \mathbf{X}_i + \beta_{\text{auto}} \text{autocov}_i \quad y_i \sim \text{Bernoulli}(p_i) \quad (11)$$

with uninformative prior distributions implemented using normal distributions

$$\alpha \sim \mathcal{N}(0, \tau_\alpha) \quad \beta \sim \mathcal{N}(0, \tau_\beta) \quad \beta_{\text{auto}} \sim \mathcal{N}(0, \tau_{\text{auto}}) \quad (12)$$

where \sim indicates that the variable on the left-hand side is a random variable derived from a distribution specified on the right-hand side; $\tau = 1/\sigma^2$ is the precision and we have typically used values of order $\tau = 10^{-6}$. For illustration, Appendix A contains a code fragment for the particular case of a second-order (Fig. 1), equally weighted neighborhood.

WinBUGS automatically generates conditional distributions for Gibbs sampling, by applying rules (see Gilks et al. 1994) for analyzing the overall model structure, given the submodels appearing in the model specification. In the case of the model specified in Eqs. 11 and 12, these rules permit the direct specification of conditional distribution for each cell, i.e., the p_i specified in Eq. 12, while for the regression parameters α and β_j , the rules lead to construction of the pseudo-likelihood, multiplied by the relevant prior from Eq. 12. The ability of WinBUGS to sample from a wide range of distributions allows it to sample directly from the pseudo-likelihood. Thus, in contrast to Heikkinen and Hogmänder (1994) and Hoeting et al. (2000), we do not rely on a normal approximation to the pseudo-likelihood, nor require the nonlinear maximization step employed to obtain such approximations.

Model evaluation

Models were evaluated both for their ability to capture spatial aggregation in the virtual glider distribution and for predictive performance. Residuals correlograms were used to assess model performance

in terms of the reduction in residual spatial structure achieved.

Model predictive performance was assessed using the area under the receiver operating characteristics (ROC) curve (Swets 1988) as an index of model predictive performance at the cells in the virtual glider range that were *not* used to fit the model. The Area Under the ROC Curve (AUC) evaluates performance across all possible thresholds (0,1), rather than relying on an arbitrary threshold value, and has become a widely accepted measure of binary model predictive performance (Miller and Hui 1991, Fielding and Bell 1997, Pearce and Ferrier 2000). The AUC ranges between 0 and 1, where $\text{AUC} = 1$ implies perfect prediction and $\text{AUC} = 0.5$ implies prediction no better than a random guess.

RESULTS

Random sampling at 20% survey intensity

We examined the performance of the autologistic model under the three scenarios representing weak, intermediate, and strong intrinsic spatial autocorrelation (Fig. 4), with the standard logistic model as a reference point. The autologistic model was stable under all three autocorrelation strengths. Residuals correlograms (see Fig. 6) reveal that although the autologistic model could not explain all spatial structure, residual structure is substantially reduced from that of the standard logistic model.

In terms of predictive performance, the results were conclusive and compelling. The autologistic model provided demonstrably better predictions of the virtual glider distribution when intrinsic spatial structure, beyond that attributable to spatial structure in habitat variables, was present in the observation data. The relative improvement in model predictions by the autologistic model over standard logistic regression increases approximately linearly with the strength of the intrinsic spatial structure added to the data (Appendix B). This result supports the use of autologistic models when spatial structure is apparent in survey data.

Sensitivity to sampling intensity and design

The autologistic model was stable down to 5% random sampling. Very little difference in the predictive performance was evident between models built on 20%, 10%, and 5% sampling intensity (Appendix C). A substantial drop in model stability was evident between the 5% and 2% sampling scenarios. At 2% sampling intensity, the autologistic model became less stable, with posterior distributions of coefficient estimates being highly non-normal, diffuse, and failing to converge to a stationary distribution in some simulations. However, the mean predictive performance of the autologistic model at only 2% sampling was surprisingly good, with mean predictions (\hat{p}_i) giving areas under the ROC curve equivalent to those obtained by models fitted on substantially more data [0.87,0.93]. At 1% sampling

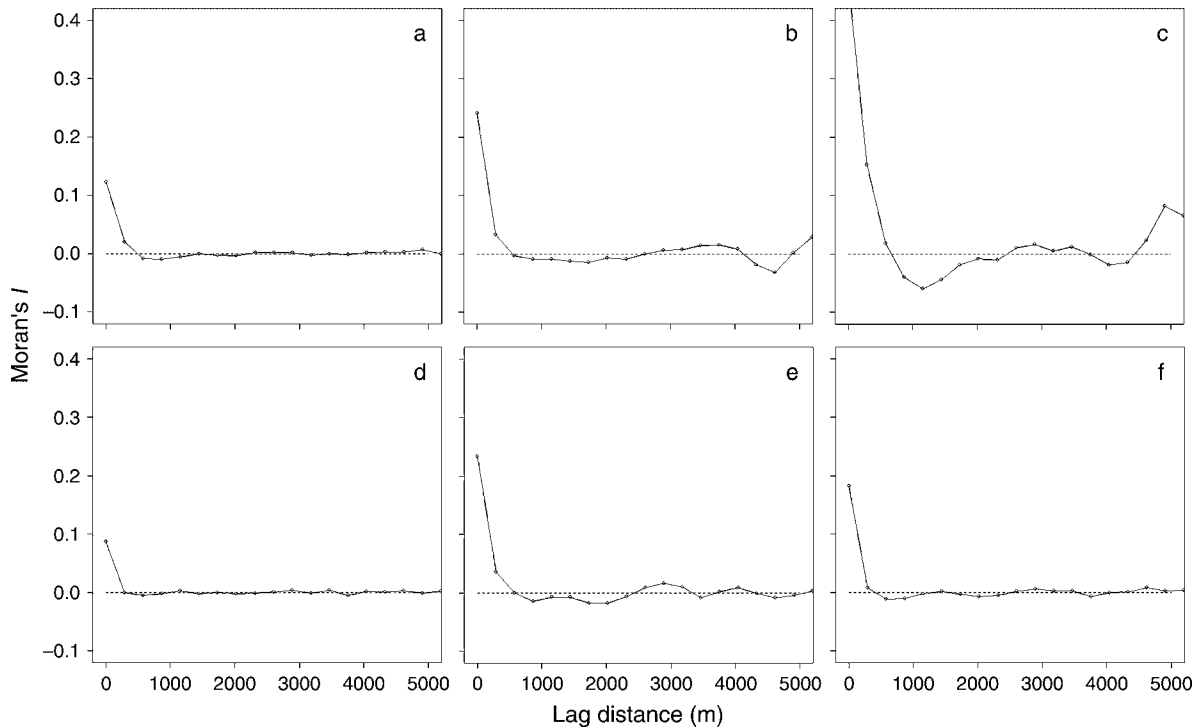


FIG. 6. Residuals correlograms for the standard logistic (upper row, a–c) and autologistic (lower row, d–f) models, fitted on observations obtained by random sampling of 20% of the virtual glider distribution with weak, medium, and strong (ordered left to right) intrinsic spatial autocorrelation.

intensity, the autologistic model failed to converge on coefficient estimates in any simulation.

Nonrandom sampling for autocovariate fitting

The utility of the autologistic model will be limited in data-poor situations, due to its failure to estimate the autocovariate coefficients at 1% sampling intensity and its apparent instability at 2% sampling intensity. A nonrandom sampling strategy was tested (see *Methods*) for utility in extending model performance in such circumstances. Models fitted to the 2% clumped observation data were stable and provided lower coefficient standard errors than those obtained from the autologistic model fitted to randomly sampled data at 2% sampling intensity. The mean predictive performance of the clumped 2% model was not, however, substantially better than that of models derived from randomly sampled observations at the same sampling intensity. Models fitted on 1% observation data were unstable under both sampling strategies, failing to converge on posterior distributions for any of the coefficients.

There was substantial overlap in the area under ROC curves (Fig. 7) between the two types of sampling strategies, indicating minimal impact of the sampling strategy on model predictive performance. Clear separation is evident between predictive performance, over all simulations, of the autologistic and the standard logistic models at 2% sampling intensity (Fig. 7).

DISCUSSION

As far as we are aware, this paper presents the first account of the performance of the autologistic model (Besag 1974) under a range of underlying strengths of intrinsic spatial structure and sampling intensities. This study was warranted because the autologistic model is gaining in popularity in the conservation literature (e.g., Gumpertz et al. 1997, Ferrier et al. 2002, Klute et al. 2002), despite a lack of rigorous testing in the literature to justify its general application. Augustin et al. (1996) and Hoeting et al. (2000) provide comparisons of autologistic model results to those obtained by standard logistic regression, but do not explore the implications of sampling intensity or the possible limitations to its use.

It was anticipated that the performance of the autologistic model must decline as sample sizes become smaller. However, it was unclear at exactly what point there would be insufficient data to reliably estimate the autocovariate coefficient, or indeed whether it would remain stable down to the minimum sample sizes required to reliably estimate the coefficients of a standard logistic regression model. Prior to this study, we speculated that the autocovariate would become unstable when sampling intensity became too low to ensure approximately 20 first-order neighbors within the sample. In addition, we speculated that smaller sample sizes would yield stable estimates of the autocovariate coefficient if they were collected in a nonrandom manner

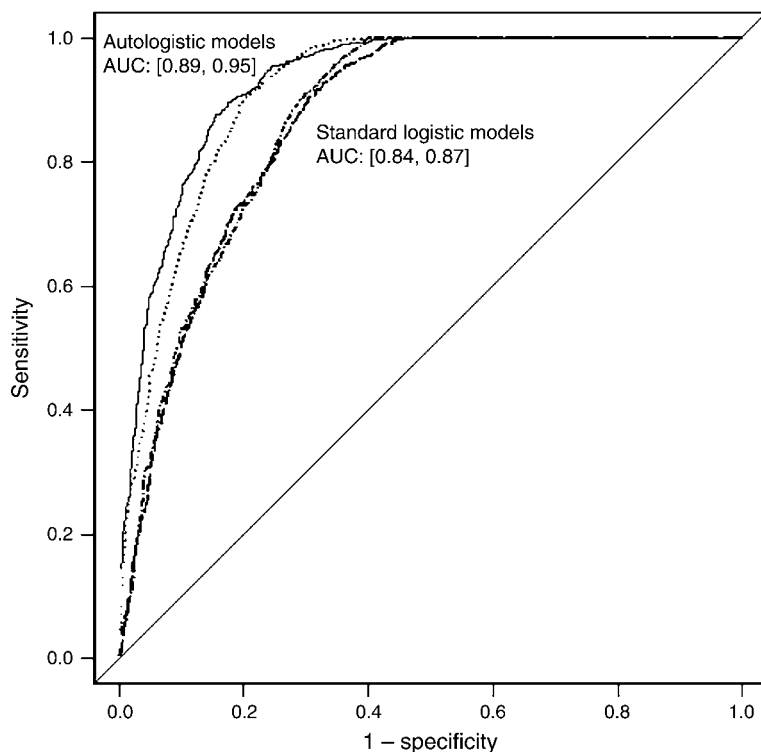


FIG. 7. ROC (receiver operating characteristics) curves representing the predictive performance of the autologistic model at 2% sampling intensity, under clumped (solid black line) and random (dotted line) sampling of the virtual glider distribution, with medium-strength intrinsic spatial autocorrelation. The two inner (dashed) curves show results for the standard logistic regression models produced under clumped and random sampling. The range of AUC (area under curve) values from 10 simulations is given in square brackets. Sensitivity is the proportion of predicted presences that were true presences; specificity is the proportion of predicted absences that were true absences.

that ensured that sufficient nearest neighbor relationships were obtained. Hoeting et al. (2000) employed such a strategy by sampling 5.8% of their total lattice in nine clumps of 16 cells. In that case, they were able to obtain stable autocovariate coefficient estimates. Taking this idea to the extreme would involve deliberately sampling in one large cluster such that all sites had immediately adjacent neighbors. However, such a strategy would commonly lead to a “nonrepresentative” sample of the environment and may result in inefficient estimation of other habitat covariate coefficients.

The autologistic model was found to be robust to high intrinsic spatial autocorrelation, even when the neighborhood employed was smaller than that suggested by the correlogram analysis. It is evident from Fig. 3 that the highest level of spatial autocorrelation tested here is indeed strong compared with the level of aggregation usually observed in species distributions.

More surprising was the finding that the autologistic model was robust at relatively low sampling intensities, maintaining stability in parameter estimation down to 5% of the total lattice sampled and maintaining good predictive performance with as little as 2% sampling intensity. Its failure to converge on stable coefficient estimates under any sampling strategy at 1% sampling

intensity is of concern, although we have tested only one aggregation strategy, and we note that 1% of the 50×50 virtual glider grid represents a small sample size (25 observations) for any modeling study. Results presented in the case study in Appendix D provide some evidence that the autologistic model is robust when less than 1% of the landscape is surveyed, especially if the overall sample size is reasonable (in that case, 133 sites were sampled). Testing the generality of this result is a worthwhile goal for future research. Altering the sampling design, from random to spatially aggregated, had little or no effect on the predictive performance of models in the simulations. However, aggregated sampling did seem to improve the stability of the autologistic model at low sampling intensities. In principle, there must be an optimal trade-off between obtaining a representative sample of the environment and obtaining enough information about the influence of sites on each other. We note that an exhaustive array of aggregation patterns was not tested and this idea therefore cannot be discounted on the basis of the results given here. An investigation of such a trade-off should be a future research priority.

WinBUGS provided a stable, user-friendly (and free) environment within which to estimate the autologistic

model by Gibbs sampling. The simplicity with which the autologistic model is implemented in WinBUGS is evident in the very small amount of coding required to run the model. The most complicated model that we estimated took only 650 seconds to obtain enough posterior samples to construct stable marginal distributions of all parameters. The flexibility provided by WinBUGS for fitting a wide variety of models makes it an increasingly popular tool for implementing Bayesian solutions to complicated models that were previously intractable. Evidence of its increasing popularity can be found in a number of recent publications that utilize WinBUGS, including several in the conservation and ecology literature (e.g., Millar and Meyer 2000, Sæther et al. 2000, Fleishman et al. 2001, Link et al. 2002). Numerous detailed examples of the implementation of BUGS statistical analyses are provided by a book by Congdon (2001) and by the WinBUGS online user manual (*available online*).⁴

Based on our findings, we recommend that the following general procedure for analyzing spatially autocorrelated wildlife survey data will provide better predictive performance than the standard logistic regression modeling that is widely employed at present:

1) Choose a sampling strategy based on a trade-off between obtaining information about covariates vs. intrinsic demographic processes.

2) Characterize the spatial structure in the sample using Moran's *I* correlograms of the residuals from a standard logistic regression model. Methods for fitting and evaluating standard logistic regression models are well established (Harrell 2001, Burnham and Anderson 2002).

3) Assign neighborhood size and weightings based on structure from step 2.

4) Fit autologistic model with covariates (see Supplement for WinBugs code).

5) Examine the residuals correlogram from the autologistic fit.

6) Given the results of step 5, consider repeating steps 3 to 5 until residuals are sufficiently clear of spatial structure.

In this work, we have not focused on the Bayesian aspects of using WinBUGS. In all models tested, we used highly uninformative priors. Assuming that no pertinent information is available, this provides what might be considered an "objective Bayesian" (Link et al. 2002) analysis. Although we have not stressed the advantages of the Bayesian approach, we note that the Bayesian autologistic model has technical advantages over the MLE approach because the evaluation of posterior distributions avoids the instability brought about by seeking the mode of the likelihood function, a necessary component of the MLE method (Wu and Huffer 1997). In addition, WinBUGS provides the

flexibility to incorporate prior information about any parameter where it is available and thus conduct a more traditional Bayesian analysis.

Because spatial autocorrelation is likely to be present in almost all ecological studies, robust methods for incorporating it in ecological inference and prediction are particularly important. This study presents a practical method for incorporating spatial autocorrelation in inference about species habitat relationships. It demonstrates the robustness of the method under a range of realistic conditions.

This work is intended not only to encourage ecologists to consider spatial autocorrelation routinely in habitat analyses, but also to investigate WinBUGS as a tool for statistical inference and prediction.

ACKNOWLEDGMENTS

We thank Julian Fox for sharing insights into spatial autocorrelation and the construction of residual correlograms, Rob Reeves (Queensland University of Technology) for providing the preliminary WinBUGS code from which our autologistic code was developed, Michael McCarthy (University of Melbourne) for WinBUGS coding advice, and Prema Lucas (University of Melbourne) and Terry Walshe (University of Western Australia) for editorial assistance. This research was partly funded by State Forests of NSW and Queensland's Environment Protection Agency (Forest Resources) and by the Australian Research Council (LP0347473).

LITERATURE CITED

- Albert, P. S., and L. M. McShane. 1995. A generalized estimating equations approach for spatially correlated binary data: with an application to the analysis of neuroimaging data. *Biometrics* **51**:627–638.
- Anselin, L. 1992. Spatial data analysis with GIS: An introduction to applications in the social sciences. Technical Report 92–10, University of California, Santa Barbara, California, USA.
- Augustin, N. H., M. A. Muggleston, and S. T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* **33**:339–347.
- Austin, M. P., and P. C. Heyligers. 1989. Vegetation survey design for conservation: gradsect sampling of forests in north-eastern NSW. *Biological Conservation* **50**:13–32.
- Austin, M. P., J. A. Meyers, L. Belbin, and M. D. Doherty. 1995. Modelling of landscape patterns and processes using biological data. Subproject 5: Simulated data case study. Division of Wildlife and Ecology, CSIRO, Canberra, Australia.
- Besag, J. E. 1972. Nearest-neighbour systems and the autologistic model for binary data. *Journal of the Royal Statistical Society, B* **34**:75–83.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society* **2**:192–236.
- Besag, J. 1975. Statistical analysis of non-lattice data. *Statistician* **24**:179–195.
- Bio, A. M. F., P. DeBecker, E. DeBie, W. Huybrechts, and M. Wassen. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biological Conservation* **11**:2189–2216.
- Breslow, N. E., and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**:9–25.
- Burgman, M. A., D. R. Breininger, B. W. Duncan, and S. Ferson. 2001. Setting reliability bounds on Habitat Suitability Indices. *Ecological Applications* **11**:70–78.

⁴ (<http://www.mrc-bsu.cam.ac.uk/bugs/>)

- Burgman, M. A., and D. B. Lindenmayer. 1998. Conservation biology for the Australian environment. First edition. Surrey Beatty and Sons, Chipping Norton, New South Wales, Australia.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.
- Collett, D. 1999. Modelling binary data. First edition. Chapman and Hall/CRC, Boca Raton, Florida, USA.
- Congdon, P. 2001. Bayesian statistical modelling, 1st edition. John Wiley and Sons, Chichester, UK.
- Crance, J. H. 1987. Guidelines for using the Delphi technique to develop habitat suitability index curves. Biological Report 82 10134, U.S. Fish and Wildlife Service, Department of the Interior, Washington, D.C., USA.
- Cressie, N. A. C. 1991. Statistics for spatial data. First edition. John Wiley and Sons, New York, New York, USA.
- De Groot, J. J., and C. J. J. TerBraak. 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology* **22**:407–415.
- ESRI. 1997. ArcInfo. Version 7.2. Environmental Systems Research Institute, Redlands, California, USA.
- Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales. I. Species level modelling. *Biodiversity and Conservation* **11**:2275–2307.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**:38–49.
- Fleishman, E., R. Macnally, J. P. Fay, and D. D. Murphy. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology* **15**:1674–1685.
- Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**:721–741.
- Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood. Pages 156–163 in E. M. Keramidas and S. M. Kaufman, editors. *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*. Interface Foundation of North America, Seattle, Washington, USA.
- Geyer, C. J., and E. A. Thompson. 1992. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: B* **54**:657–699.
- Gilks, W. R., A. Thomas, and D. J. Spiegelhalter. 1994. A language and program for complex Bayesian modelling. *The Statistician* **43**:169–178.
- Gotway, C. A., and W. W. Stroup. 1995. A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological and Environmental Statistics* **2**:157–178.
- Guisan, A., and N. E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**:147–186.
- Gumpertz, M. L., J. M. Graham, and J. B. Ristraino. 1997. Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper. *Journal of agricultural, biological and environmental statistics* **93**:1099–1111.
- Hanski, I. 1992. Inferences from ecological incidence functions. *American Naturalist* **139**:657–662.
- Harrell, F. E. 2001. Regression modeling strategies: with application to linear models, logistic regression, and survival analysis. First edition. Springer-Verlag, New York, New York, USA.
- Heikkinen, J., and H. Hogmander. 1994. Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics* **43**:569–582.
- Hirzel, A. H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* **145**:111–121.
- Hoeting, J. A., M. Leecaster, and D. Bowden. 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics* **5**:102–114.
- Huffer, F. W., and H. Wu. 1998. Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* **54**:509–524.
- Hutchinson, M. F., H. A. Nix, D. J. Houlder, and J. P. McMahon. 1999. ANUCLIM; A software package for systematic interrogation of climate surface coefficient files, as created by the ANUSPLIN package, for biophysical applications. Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.
- Johnston, J. 1972. Econometric methods. Second edition. McGraw-Hill, London, UK.
- Kavanagh, R. P. 2000. Effects of variable-intensity logging and the influence of habitat variables on the distribution of the greater glider *Petauroides volans* in montane forest, south-eastern New South Wales. *Pacific Conservation Biology* **6**:18–30.
- Kavanagh, R. P., and M. J. Lambert. 1990. Food selection by the greater glider, *Petaurus volans*: is foliar nitrogen a determinant of habitat quality? *Australian Wildlife Research* **17**:285–299.
- Klute, D. S., M. J. Lovallo, and W. M. Tzilkowski. 2002. Autologistic regression modeling of American woodcock habitat use with spatially dependent data. Pages 335–343 in J. M. Scott, editor. *Predicting species occurrences; issues of accuracy and scale*. Island Press, Washington, D.C., USA.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**:1659–1673.
- Legendre, P., and M. J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* **80**:107–138.
- Lindenmayer, D. B., R. B. Cunningham, M. T. Tanton, A. P. Smith, and H. A. Nix. 1990. Habitat requirements of the mountain brushtail possum and the greater glider in the montane ash-type eucalypt forests of the Central Highlands of Victoria, Australia. *Australian Wildlife Research* **17**:467–478.
- Lindenmayer, D. B., M. A. McCarthy, and M. L. Pope. 1999. Arboreal marsupial incidence in eucalypt patches in south-eastern Australia: a test of Hanski's incidence function metapopulation model for patch occupancy. *Oikos* **84**:99–109.
- Lindenmayer, D. B., K. Ritman, R. B. Cunningham, J. D. B. Smith, and D. Horvath. 1995. A method for predicting the spatial distribution of arboreal marsupials. *Wildlife Research* **22**:445–456.
- Link, W. A., E. Cam, J. D. Nichols, and E. G. Cooch. 2002. Of bugs and birds: Markov Chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management* **66**:277–291.
- McCullagh, P., and J. A. Nelder. 1989. Generalised linear models. Second edition. Chapman and Hall, London, UK.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machine. *Journal of Chemical Physics* **21**:1087–1091.
- Millar, R. B., and R. Meyer. 2000. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Canadian Journal of Fisheries and Aquatic Sciences* **57**:43–50.
- Miller, M. E., and S. L. Hui. 1991. Validation techniques for logistic regression models. *Statistics in Medicine* **10**:1213–1226.
- Moilanen, A., and M. Nieminen. 2002. Simple connectivity measures in spatial ecology. *Ecology* **83**:1131–1145.

- National Parks and Wildlife Service. 1998. Eden fauna modelling. New South Wales Regional Forest Agreement Steering Committee, Canberra, Australia.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**:225–245.
- Pereira, J. M. C., and R. M. Itami. 1991. GIS-based modelling using logistic multiple regression: a case study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing* **57**:1475–1486.
- Rand, G. M., and J. R. Newman. 1998. The applicability of habitat evaluation methodologies in ecological risk assessment. *Human and Ecological Risk Assessment* **4**:905–929.
- Ryden, T., and D. M. Titterton. 1998. Computational Bayesian analysis of hidden Markov models. *Journal of Computational and Graphical Statistics* **7**:194–211.
- Sæther, B. E., J. Tufto, S. Engen, K. Jerstad, O. W. Rostad, and J. E. Skatan. 2000. Population dynamical consequences of climate change for a small temperate songbird. *Science* **287**:854–856.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Gilks. 1996. BUGS 0.5: Bayesian inference using Gibbs sampling. Technical report, version 2. Medical Research Council, Biostatistics Unit, Institute of Public Health, Cambridge, UK.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. 2003. WinBUGS version 1.4 user manual. Medical Research Council Biostatistics Unit, London, UK.
- Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**:1285–1293.
- Wolfinger, R., and M. O'Connell. 1993. Generalised linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**:233–243.
- Wu, H., and F. W. Huffer. 1997. Modeling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics* **4**:49–64.

APPENDIX A

Sample (abbreviated) of WinBUGS code for modeling spatial autocorrelation in observation data with habitat covariates influencing the probability of presence (*Ecological Archives* A016-066-A1).

APPENDIX B

ROC curves for the standard and autologistic models under varying strengths of injected spatial autocorrelation (*Ecological Archives* A016-066-A2).

APPENDIX C

ROC curves for the autologistic model with varying observation efforts (*Ecological Archives* A016-066-A3).

APPENDIX D

Sample application of the autologistic model: modeling the habitat preferences of the yellow-bellied glider in the Eden region of southeastern Australia (*Ecological Archives* A016-066-A4).

SUPPLEMENT

WinBUGS code for estimating the autologistic model for the yellow-bellied glider described in Appendix D (*Ecological Archives* A016-066-S1).