Ecological Applications, 15(2), 2005, pp. 554–564

FACTORS AFFECTING SPECIES DISTRIBUTION PREDICTIONS: A SIMULATION MODELING EXPERIMENT

Gordon C. Reese,[1,4] Kenneth R. Wilson,[1] Jennifer A. Hoeting,[2] and Curtis H. Flather[3]

[1]*Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA*
[3]*U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado 80526 USA*

Abstract. Geospatial species sample data (e.g., records with location information from natural history museums or annual surveys) are rarely collected optimally, yet are increasingly used for decisions concerning our biological heritage. Using computer simulations, we examined factors that could affect the performance of autologistic regression (ALR) models that predict species occurrence based on environmental variables and spatially correlated presence/absence data. We used a factorial experiment design to examine the effects of survey design, spatial contiguity, and species detection probability and applied the results of ten replications of each factorial combination to an ALR model. We used additional simulations to assess the effects of sample size and environmental data error on model performance. Predicted distribution maps were compared to simulated distribution maps, considered ''truth,'' and evaluated using several metrics: omission and commission error counts, residual sums of squares (RSS), and areas under receiver operating characteristic curves (AUC). Generally, model performance was better using random and stratified survey designs than when using other designs. Adaptive survey designs were an exception to this generalization under the omission error performance criterion. Surveys using rectangular quadrats, designed to emulate roadside surveys, resulted in models with better performance than those using square quadrats (using AUC, RSS, and omission error metrics) and were most similar in performance to a systematic quadrat design. Larger detection probabilities, larger sample sizes, contiguous distributions, and fewer environmental data errors generally improved model performance. Results suggest that spatially biased sample data, e.g., data collected along roads, could result in model performance near that of systematic quadrat designs even in the presence of potentially confounding factors such as contiguity of distributions, detection probability, sample size, and environmental data error.

Key words: autologistic regression model; detection probability; environmental data error; habitat relationship modeling; prediction accuracy assessment; roadside survey; sample data; sample size; sampling bias; spatial contiguity; species range.

Introduction

Species distribution data are useful for population monitoring (Shaffer et al. 1998), biodiversity mapping (Bojórquez-Tapia et al. 1995), and conservation management (Corsi et al. 1999). Suitable habitats have been described for only a small percentage of species (Garrison et al. 2000); consequently, overlays of geospatial species sample data (e.g., records with location information from museum collections or spatially extensive annual surveys; henceforth sample data) with environmental variables (e.g., elevation, vegetation types, land use) are often used to determine wildlife-habitat relationships and predict distributions (Stoms et al. 1992, Anderson et al. 2003). As pressures on our biological heritage increase the need for quick decisions, modeling procedures often rely mostly or exclusively on existing data and regularly ignore error estimates. Time and money considerations, increased accessibility to sample data (Edwards et al. 2000) and the development of various modeling approaches (e.g., Stockwell and Noble 1992, Augustin et al. 1996, Pearce and Ferrier 2000, Hirzel et al. 2002) make distribution estimates with extant sample data appealing and easy. When the data used for modeling are less than optimal, which is often the case, inherent data errors or biases can manifest and negatively affect predictions (Veregin 1989). Many of these modeling approaches deduce wildlife-habitat relationships from similar databases and are, therefore, each exposed to similar data quality issues.

Use of existing data is often our best option for addressing urgent issues, yet we know little about the effects of common biases. For example, few studies have investigated model performance (e.g., the bias and precision of a prediction) as a function of sample data quantity, quality, or spatial configuration. Hirzel et al. (2001) and Stockwell and Peterson (2002) have shown that larger sample sizes lead to greater accuracy. Also, Hirzel and Guisan (2002) reported differences in prediction accuracy resulting from various ''optimal'' sur-

Manuscript received 17 November 2003; revised 12 June 2004; accepted 19 June 2004; final version received 12 July 2004. Corresponding Editor: W. V. Reid.

[4] E-mail: greese@cnr.colostate.edu

vey designs, but did not evaluate effects from ''biased'' survey designs. Kadmon et al. (2004) found that near-road surveying biases had little effect on prediction accuracy when a bioclimatic model was used.

Accessibility has long been an important consideration when field surveys are conducted (Funk and Richardson 2002, Reddy and Dávalos 2003, Kadmon et al. 2004). For example, roads provide vantage points from which annual surveys such as the North American Breeding Bird Survey are conducted; however, this nonrandom design creates obvious difficulties with respect to inference about the population in question. Roadside habitat can differ from the habitat composition of surrounding areas (Bart et al. 1995, Keller and Scallan 1999) and some analyses are limited when sample data are concentrated along roads or are near accessible sites (e.g., Bojórquez-Tapia et al. 1995).

Additional modeling factors could confound the performance of models that predict species distributions. First, the spatial contiguity of a species geographic range (a measure of whether the distribution is connected throughout or broken into disjunct units) could interact with a survey design to affect model performance. Second, surveys can incorrectly identify a species as present that is actually absent (false presence) or fail to detect a species that is actually present (false absence). Furthermore, because absence data (i.e., survey conducted but species not detected) are often not available, false absences could also occur when geographic areas without confirmation of species presence are treated as ''absences'' (see Corsi et al. 1999, Anderson et al. 2003). In either case, the prevalence of false absence or false presence records will affect empirical attempts to predict distributions based on environmental variables (Tyre et al. 2003). Third, mapped environmental variables from which habitat affinities are derived contain errors (Janssen and van der Wel 1994). It is important to understand how and to what degree these factors affect model performance because as Dean et al. (1997) found, overlaying predicted species distributions with as little as 5% error could considerably alter the estimated distribution of species richness.

Our objective was to examine how survey design, spatial contiguity, detection probability, sample size, and error in environmental data affect the performance of models predicting a species distribution using autologistic regression (ALR) with covariates (Augustin et al. 1996). The ALR model extends basic logistic regression with an extra covariate (i.e., degree of spatial autocorrelation) used to model responses, in this case a grid of observed binary responses. We used computer simulations, which allow testing of numerous scenarios without complications from natural variation, to evaluate these relationships and test our hypothesis that spatially biased survey designs result in models that perform more poorly than do unbiased survey designs. The outcomes from these experiments will provide insights on how various survey designs perform under various ecological and modeling circumstances.

Methods

A flow chart and general overview of our methods are available in Appendix A. Data and simulation programs are available in the Supplement.

Environmental variables

We acquired elevation (DEM) and national land-cover (NLC) raster data for a 2.25-km^2 region of the San Andres Mountains of New Mexico from the U.S. Geological Survey and the New Mexico Resource Geographic Information System Program, respectively. The specific region was selected to reduce computer simulation time and contained 2500 grid cells, (50×50 cells), and relatively few land-cover types, (evergreen forest, shrubland, and grasslands).

Species distribution maps

We generated three hypothetical ''true'' species distribution maps from the environmental variables using ArcView 3.2a (Environmental Systems Research Institute, Redlands, California, USA). The distributions covered 10% of the study area (250 cells) and were created by developing environmental data affinities that associated a defined elevation range with a single land-cover type. Distributions were comprised of one, three, or six patches that defined the high, moderate, or low spatial contiguity treatments, respectively (Fig. 1).

Survey designs

We surveyed the distribution maps with replacement to emulate actual field sites that have received multiple surveys over time. Therefore, survey designs applied equal amounts of effort, but did not necessarily result in equal numbers of sample data. We compared six survey designs that included random, stratified, adaptive, and three quadrat designs: systematic, rectangular, and square. In the stratified design, surveys were equally divided and randomly located amongst the three land cover types.

Under an adaptive design, a design primarily used to survey rare and clumped distributions (see Thompson 1990), the study area was randomly surveyed, with replacement, until a surveyed cell resulted in a presence. At this point, the eight adjacent cells were surveyed, without replacement, until no presences occurred in any adjacent cell. Finally, all surveyed sites were replaced and the procedure returned to the first step (random selection of survey cell).

Three designs (systematic, rectangular, and square) used quadrats of 25 000 m^2 and 20 samples were randomly drawn from within each quadrat. To increase the probability of surveying cells at the edge of the study area, quadrats were permitted to overlay the study area boundary, resulting in partial quadrats. For the systematic design, a random location within a 354×354 m

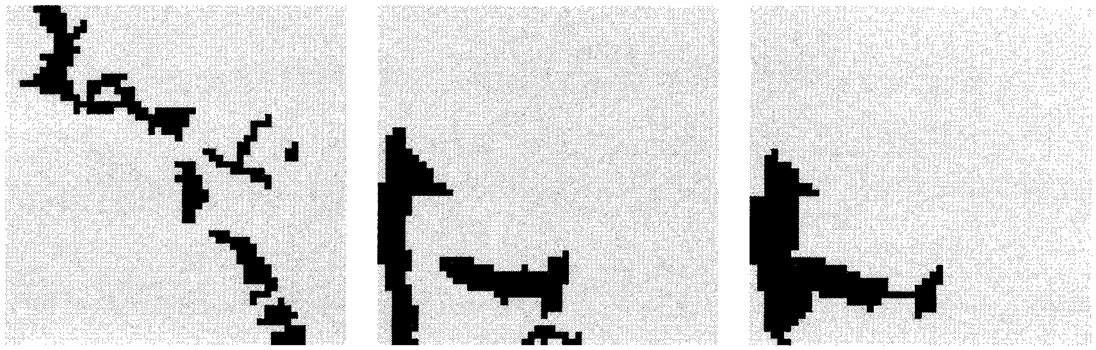

FIG. 1. Hypothetical species distribution (black) maps, occupying 250 of 2500 cells, created by defining habitat associations based on national land cover and elevation data. The three levels of spatial contiguity with six (leftmost panel), three (center panel), or one (rightmost panel) species distribution patches correspond to low, moderate, and high contiguity. The distribution with low spatial contiguity was defined as grassland from 1810–1862 m, the distribution with moderate spatial contiguity was defined as evergreen forest from 1946–2020 m, and the distribution with high spatial contiguity was defined as evergreen forest from 1967–2064 m.

area in the lower left corner of the study area was selected as the center of the first quadrat. Subsequent quadrats were then spaced equidistantly (354 m) in the horizontal and vertical directions. This placement resulted in 16, 20, or 25 quadrats per simulation, depending on whether the first quadrat overlapped 0, 1, or 2 study area edges, respectively. The rectangular and square quadrat designs used at least 18 randomly located, potentially overlapping, rectangular (50 $\times$ 500 m) or square (approximately 158 $\times$ 158 m) quadrats. If fewer than 360 samples were drawn from the 18 quadrats (due to quadrat areas that partially fell outside the study extent), additional quadrats were used. Rectangular quadrats were designed to mimic roadside surveys. We randomly oriented each rectangular quadrat between 0° and 360° and found numerous data dimensions within an unpublished U.S. Forest Service vertebrate database and, therefore, arbitrarily selected the 1:10 width to length ratio (see Appendix B for examples).

Simulation process

We also examined the relationship between survey design and model performance when the potentially confounding factors (spatial contiguity, detection probability, sample size, and environmental error) were varied. For each combination of survey design (six), spatial contiguity (three), and detection probability (three; $p = 0.5$; 0.75; or 1), we simulated 10 independent replications. Surveys resulted in presences only when the surveyed grid cell was part of the distribution and only when a computer-generated uniform (0, 1) variate was less than or equal to p; an absence (in this case, a false absence) occurred in these cells if the uniform variate was greater than p. A cell surveyed as a presence was permanently recorded; however, false absences cells were changed if they were resurveyed and then resulted in a presence. An absence occurred wherever the surveyed cell was not part of the distribution. We did not test effects from false presences, considering these to be much less common an occurrence than false absences.

We also examined effects from sample size and environmental data error. To test sample size effects we doubled the number of samples (720) and used only the distribution map of moderate spatial contiguity and $p = 0.75$. The effects of environmental data error on model performance were assessed by introducing classification error to the NLC data and measurement error to the DEM. Following Dean et al. (1997), we tested error rates of 5% and 20% and additionally tested an extreme case where 50% of the cells contained error. In addition to error rate, we examined how the location of error (randomly or selectively placed) affects model performance. Both random and selective error maps were created from the NLC data at each level of error. For the random error map, randomly selected cells were reclassified and for the selective error map only land-cover class edges were reclassified, i.e., error predominated in areas of transition between adjacent land-cover classes (see Pathirana 1999). Additionally, we assumed a positive relationship between the amount of classification error and the buffer area around cells that define land-cover class edges. For 5% error, the buffer area included only cells adjacent in the cardinal directions, diagonals were added to the buffer area for tests of 20% error, and an additional one-cell separation in the cardinal directions was added to the buffer area for use with 50% error. For both error location types, selected cells were changed to the land-cover class of the nearest neighbor of a different class.

We introduced error only to randomly selected cells in the DEM because we expected that selective error (i.e., larger probability of error near some geographic feature) is less likely in a continuous variable such as elevation. Elevation values were changed to the mean of the surrounding 3 $\times$ 3 neighborhood, resulting in measurement error ranging from -11.0 to 12.3 m. The sample data that resulted from surveying the distribution map with moderate spatial contiguity with $p = 0.75$ (360 samples) were combined with the environmental data containing error to generate new ALR predictions.

Autologistic regression model

Augustin et al. (1996) developed an ALR model to predict a species distribution by modifying logistic regression to model spatial autocorrelation in a grid of binary responses (e.g., presence/absence data) while continuing to model associations with grids of covariates (e.g., environmental data). We used a modified ALR approach (see Hoeting et al. 2000; G. S. Young and J. A. Hoeting, program Autologit.cc, *available online*)[5] with Bayesian parameter estimation.

ALR is a logistic regression model that incorporates an additional term to account for clustering of species occurrences over the landscape. The ALR model estimates the presence/absence of a species across a grid of cells according to the model

$$\Pr(x_i = 1 \mid \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{z}_i, \beta) = \frac{\exp[\theta_0 + \theta_1 z_{i1} + \mathrm{L} + \theta_p z_{ip} + \beta s(x_i)]}{1 + \exp[\theta_0 + \theta_1 z_{i1} + \mathrm{L} + \theta_p z_{ip} + \beta s(x_i)]}$$

where the probability that site i is occupied ($\Pr[x_i = 1]$) is conditioned on the occupancy pattern in some neighborhood of site i ($\mathbf{x}_{-1}$), the estimated parameters ($\theta_0, \ldots, \theta_p$) reflect the effect of the p covariates for site i ($z_{i1}, \ldots, z_{ip}$) on occupancy at site i, and the estimated parameter (β) reflects the effect of the spatial covariate $s(x_i)$ on site occupancy. The spatial covariate, $s(x_i)$, is defined as the total number of cells where a species was detected in the eight cells that surround site i.

In the Bayesian modeling paradigm, the true presence/absence values over the entire region of interest are considered to be parameters to be estimated based on some observed data, which are observed for only a portion of the region of interest. Each survey replication produced four observed data sets: (1) an indicator for whether a cell was surveyed, (2) an indicator for whether a surveyed cell resulted in the detection of a species (together, these two variables define presence/absence sample data), and the environmental variables (3) elevation and (4) land-cover type. These results were used to parameterize an ALR model to predict probabilities of species occurrence in cells not surveyed. Parameter estimation was carried out using Gibbs sampling (further details are listed in the appendices and supplement). Hoeting et al. (2000) provide additional details on the model and the method of estimation.

As implemented here, the model assumes perfect detection; however, to mimic real sample data we chose to violate this assumption, thus some observed data were generated using a detection probability less than one. Additional model assumptions were met with the simulated design of the study, e.g., responses were independent of the success of surveying at other locations and territory size of our hypothetical species was set to be equivalent to the area of one grid cell (see Cressie 1993:section 6.4).

Performance assessment

We used several metrics to evaluate the performance of predicted distribution maps. We compared the true presence/absence to the predicted probability of presence for each grid cell and generally report the mean of our performance metrics over the 10 replications for each factorial combination. However, when we pooled our results across survey designs we report medians as a measure of central tendency because the distributions of performance metrics were skewed (Zar 1996:24). The number of grid cells for the map and the percent of cells occupied never changes, so we report cell counts of omission (species truly present but predicted to be absent) and commission (species not present but predicted to be present) error using the 0.5 threshold commonly used to differentiate a predicted presence from a predicted absence (Fielding and Bell 1997). Different thresholds would change cell counts (e.g., decreasing the threshold would tend to reduce omission error and increase commission error); therefore, two threshold independent metrics, residual sum of squares (RSS) and area under receiver operating characteristic curves (AUC), were also computed (see Fielding and Bell 1997, Pearce and Ferrier 2000).

RSS is computed as

$$\mathrm{RSS} = \frac{1}{10} \sum_{j=1}^{10} \sum_{i=1}^{2500} (\hat{x}_{ij} - x_i)^2$$

where $\hat{x}_{ij}$ is the predicted value for cell i in replication j and x_i is the true presence (or absence) at cell i. Smaller RSS values indicate better performing models.

To investigate the discrimination of each model prediction we generated receiver operating characteristic curves. A receiver operating characteristic curve is one where the true-presence fraction is plotted against the false-presence fraction for a sequence of thresholds of predicted probability of presence. We used 0.01 increments of the threshold between and including 0 and 1 and used the trapezoidal rule to calculate the area under the curve (AUC; Pearce and Ferrier 2000). Larger AUCs indicate better performance, i.e., better discrimination.

RESULTS

Model performance was dependent on survey design and varied between metrics (Table 1). Generally (averaged across all factors), predictions made from random and stratified sample data were better than those from quadrat designs (systematic, square, and rectangular), and were better than adaptive survey designs

[5] ⟨www.stat.colostate.edu/~jah/software/⟩

TABLE 1. Performance, based on four metrics, of predicting a simulated species distribution map using an autologistic regression model, as a function of survey design.

Survey design	Model performance metric				Average rank
	AUC	RSS	Omission	Commission	
Random	0.96 (1)	110 (1)	115 (2)	33 (2)	1.5
Stratified	0.96 (1)	112 (2)	119 (3)	32 (1)	1.75
Adaptive	0.95 (3)	259 (6)	53 (1)	307 (6)	4
Systematic	0.91 (4)	152 (3)	145 (4)	59 (4)	3.75
Rectangular	0.90 (5)	158 (4)	150 (5)	57 (3)	4.25
Square	0.85 (6)	187 (5)	161 (6)	83 (5)	5.5

Notes: The means by survey design, across the factors spatial contiguity level and detection probability, are reported. The rank of each performance metric is given parenthetically. The mean rank is estimated within each survey design. Survey design indicates the method used to generate presence/absence data from a known distribution map. Discrimination was evaluated with the area under receiver operating characteristic curves (AUC), and larger values correspond to better discrimination. Smaller residual sums of squares (RSS) indicate better performing models. Omission and commission errors were calculated with a 0.5 threshold, and the numbers of incorrectly predicted grid cells are reported.

by all measures except omission error. Although adaptive survey designs had the smallest number of omission errors, they produced the greatest RSS values and the largest number of commission errors. Rectangular quadrat designs produced models that performed better than square quadrat designs and only slightly worse than systematic quadrat designs. These are general trends and exceptions can be seen when comparing among the individual survey designs within each of the potentially confounding factors (spatial contiguity, detection probability, sample size, and environmental error; Figs. 2–5).

Using the median across survey designs, prediction performance was positively related to level of contiguity in the distribution maps. This suggested that maps with high spatial contiguity generally were predicted better than were maps with low contiguity. However, the relationship between contiguity and prediction performance was not consistent. We observed numerous deviations from a monotonic increase in performance with increasing levels of contiguity. Notably, the adaptive design had the poorest performance for the high contiguity maps using RSS and commission performance measures (Fig. 2b, d) and the systematic, rectangular, and square designs performed least well under moderate distributional contiguity (Fig. 2a, c).

When detection probability was varied, omission and commission errors showed a consistent reciprocal relationship; for each increase in p, omissions decreased and commissions increased (Fig. 3c, d). AUC and p were positively related for all survey designs except systematic which showed lower performance for $p \geq 0.75$ (Fig. 3a). Under RSS, most designs (random, stratified, systematic, and rectangular) showed improved performance with larger detection probabilities; adaptive and square designs performed worst when species, if present, were always detected (Fig. 3b).

Larger sample sizes almost always yielded better performing predictions of occurrence regardless of the survey design used (Fig. 4). The exceptions to this pattern were restricted to two designs and two performance metrics. Prediction performance was slightly worse (RSS) with square quadrats and commission errors increased with square and rectangular designs at the doubled sample size (Fig. 4b, d). Also of note, increasing the sample size greatly improved predictions based on RSS and commission error under the adaptive survey design (Fig. 4b, d).

Introduction of covariate error, either randomly or selectively, was characterized by an inverse relationship between prediction performance and level of environmental error (Fig. 5). The individual survey designs generally yielded predictions that performed more poorly as the amount of covariate error increased. There were exceptions for all survey designs except the rectangular where models performed slightly better under a 5% error level than 0% error level (e.g., Fig. 5a, b, d, f, g, h). Random and stratified designs yielded the most robust predictions (smallest absolute change between 0% and 50% error) for all but the omission error criterion. The differences in effects on model performance were generally small when comparing selectively to randomly located error.

DISCUSSION

We found that the performance of models predicting species distributions predicted with ALR is dependent on the design used to collect the sample data and the measure used to evaluate model performance. Random and stratified designs performed the best according to AUC, RSS, and the number of commission errors. These designs also appeared to be affected less by environmental data error than other designs (see Fig. 5). Adaptive survey designs consistently produced the smallest number of omission errors which was expected because, unlike the other designs, surveys resulting in a presence directed subsequent survey locations. However, the small number of omission errors associated with adaptive designs came at a cost; adaptive designs

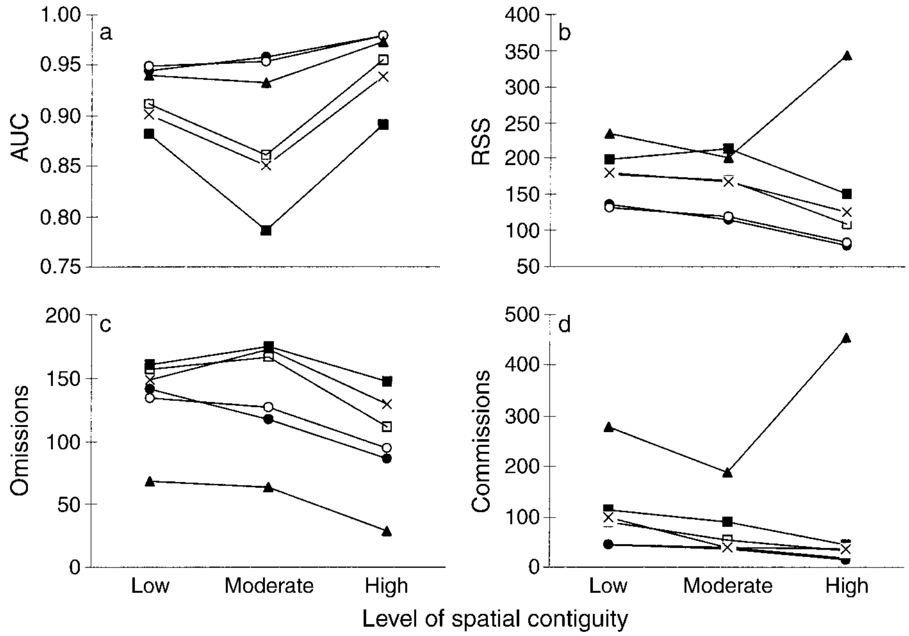

FIG. 2. Performance of survey designs in predicting a simulated species distribution map as a function of three levels of spatial contiguity (low, moderate, or high) using (a) AUC, (b) RSS, (c) number of omissions, and (d) number of commissions as performance measures. Results are the means of the three detection probabilities. The survey designs are random (solid circles), stratified (open circles), adaptive (solid triangles), systematic (open squares), rectangular (×), and square (solid squares).

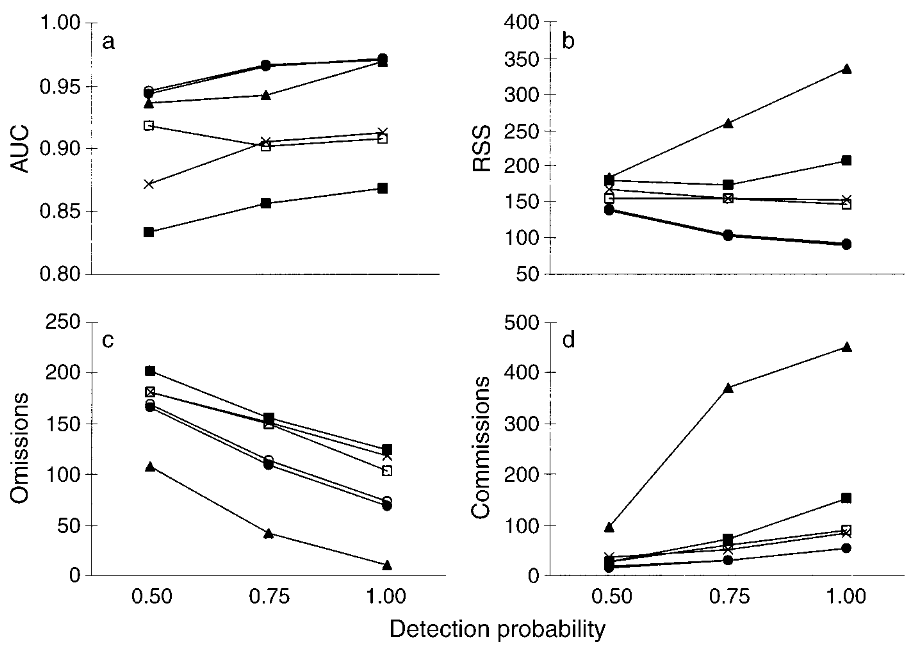

FIG. 3. Performance of survey designs in predicting a simulated species distribution map as a function of three detection probabilities (0.50, 0.75, or 1.00) using (a) AUC, (b) RSS, (c) number of omissions, and (d) number of commissions as performance measures. Results are the means of the three spatial contiguity levels. The survey designs are random (solid circles), stratified (open circles), adaptive (solid triangles), systematic (open squares), rectangular (×), and square (solid squares).

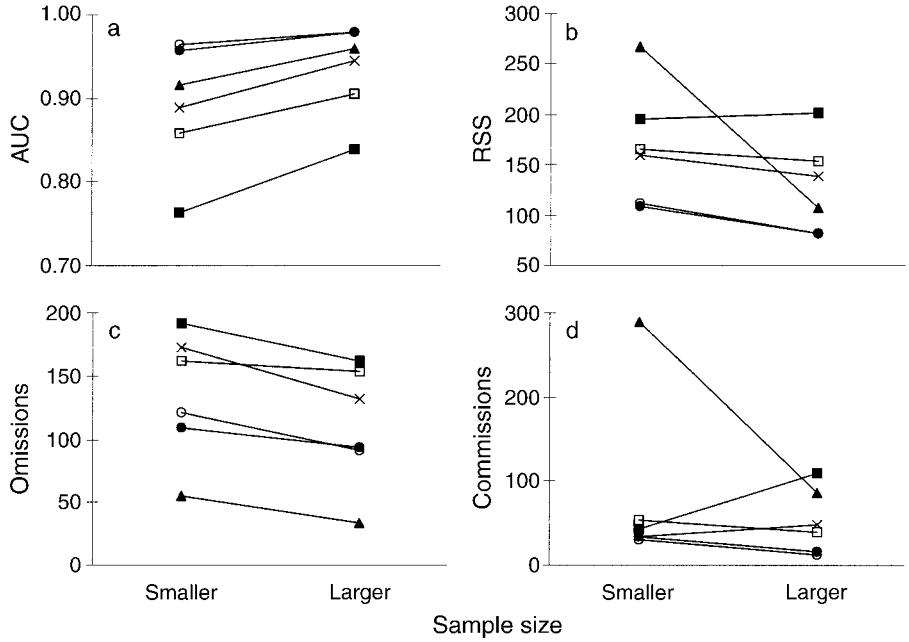

FIG. 4. Performance of survey designs in predicting a simulated species distribution map as a function of smaller (360) or larger (720) sample size (number of samples drawn) using (a) AUC, (b) RSS, (c) number of omissions, and (d) number of commissions as performance measures. Comparisons are for the distribution map with moderate spatial contiguity and with $p = 0.75$, and results are the means of 10 replications for each combination of design × sample size. The survey designs are random (solid circles), stratified (open circles), adaptive (solid triangles), systematic (open squares), rectangular (×), and square (solid squares).

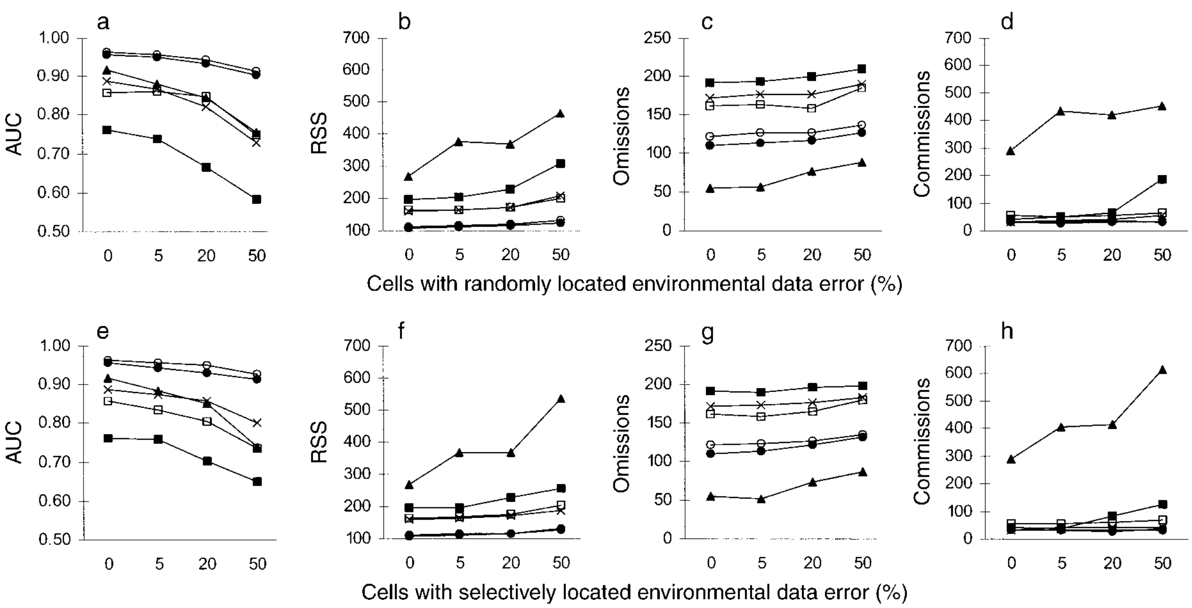

FIG. 5. Performance of survey designs in predicting a simulated species distribution map as a function of four levels of environmental data error (0%, 5%, 20%, or 50%), located either randomly or selectively along land-cover type edges. Performance measures include (a, e) AUC, (b, f) RSS, (c, g) number of omissions, and (d, h) number of commissions. Tests were completed on the distribution map with moderate spatial contiguity and with $p = 0.75$, and results are the means of 10 replications for each combination of design × level of environmental data error. The survey designs are random (solid circles), stratified (open circles), adaptive (solid triangles), systematic (open squares), rectangular (×), and square (solid squares).

regularly resulted in the largest number of commission errors and the largest residual sums of squares.

Our study further showed that models developed from rectangular quadrats performed equally well as models developed from systematic quadrats, which is promising considering the prevalence of systematic surveys in pre-planned designs (Scherrer 1984 as cited in Fortin et al. 1989). Also, predictions from surveys with rectangular quadrats were often better than those from (random) square quadrats. Similarly, Pearson and Ruggiero (2003) found that linear transects were more efficient than square grids in measuring small mammal community parameters. Rectangular quadrats may survey heterogeneity of a landscape better than square quadrats by intersecting a greater number of habitats.

While it has long been recognized that a statistically based survey design is the best approach to reduce bias and, perhaps, maximize detection, much of the available and existing species data were collected using ''designs'' that maximized accessibility. For example, a spatially biased survey design would include data collected predominantly along roads. Our study began with the hypothesis that spatially biased survey designs result in models that perform more poorly than do unbiased survey designs and both our results and those of Kadmon et al. (2004) support this hypothesis (Table 1). Our study design may have exaggerated the difference between spatially biased and spatially unbiased designs by testing equal amounts of effort and not necessarily equal amounts of data. In other words, the quadrat designs (systematic, rectangular, square), by being limited spatially, and the adaptive design, by concentrating effort near presences, were likely to result in more replicate surveys (samples within the same cells) and, therefore, less data (since data was reduced to one sample per cell) than either of the spatially unbiased survey designs (random, stratified). We would expect smaller difference between data collected via a spatially biased and non-spatially biased survey had we surveyed without replacement.

Ranks of the survey designs generally remained unchanged when we varied potentially confounding factors (spatial contiguity, detection probability [p], sample size, environmental error). First, and as expected, models performed better with maps containing more contiguous species distributions than with maps containing less contiguous distributions. More contiguous distributions could increase spatial correlation and improve model performance (e.g., with ALR). Luoto et al. (2002) also found that the occurrence of a butterfly species was better predicted in areas with greater contiguity (note however that the test area with high contiguity was adjacent to the training data, making it difficult to rule out spatial autocorrelation as an alternative explanation for better model performance). Second, p and model performance were generally positively related. A larger p yields more correct data for a given amount of survey effort. By contrast, a smaller p likely results in more false absences and thereby increases the difficulty of modeling habitat associations (Tyre et al. 2003). Including an error estimate for p during statistical analysis could improve model performance (Hoeting et al. 2000). Third, and similar to other findings (Hirzel et al. 2001, Stockwell and Peterson 2002), model performance was generally improved by increasing the number of samples. Larger sample sizes may reduce the effects from data inaccuracies (see White and Garrott 1986). Fourth, there was a negative relationship between model performance and environmental data error, likely the result of weaker habitat associations when sample data coincided with errors in environmental data.

Our tests also returned a number of unexpected results. First, stratified designs yielded models that performed almost as well as those resulting from random designs probably because sample data were almost uniformly distributed among the strata. For example, the areas of the three land-cover types were nearly equal (approximately 36.5, 31.4, and 32.1% for evergreen forest, shrubland, and grassland, respectively). Additionally, Hirzel and Guisan (2002) report that a systematic design yielded more accurate results than a random design. The difference between their results and ours may be due to our restricting systematic designs to quadrats that surveyed only 20% of the landscape, whereas their approach was not based on quadrats and probably produced smaller distances between survey sites.

Second, performance of the adaptive design was very sensitive to sample size. In particular, the number of commission errors resulting from the adaptive design was drastically reduced by doubling the sample size (Fig. 4d). With most of the species distribution already surveyed by the adaptive design at the small sample size, doubling the number of samples moved the ratio of samples in unoccupied to samples in occupied areas closer to 1:1. This balance was found to decrease commission error. The adaptive design, while poor overall (Table 1), could be the best design when conditions allow for a large sample size.

Third, contrary to our expectation we did not find a consistent pattern of increase or decrease in model performance metrics as we systematically altered our confounding factors. For example, we observed a number of instances when model performance was greater at low contiguity than at moderate contiguity (Fig. 2a, b, c) and instances when model performance was greater at moderate contiguity than at high contiguity (Fig. 2b, d). Moreover, there were a few cases where there was no improvement in model performance with larger sample sizes. In particular, the rectangular and square designs showed an increase in the number of commission errors with the larger sample size (see Fig. 4d). We suspect that these results are idiosyncratic; being a function of how our particular experiment was implemented (e.g., our choice to fix the size, shape, and spatial structure of a species distribution). A more definitive understanding of these patterns will require additional investigation.

Other issues important to predicting species occurrence that warrant further investigations include the effects of biased surveys (e.g., roadside surveys) and false absences on model performance. Perhaps more fundamentally, investigations are needed that explicitly explore why different model performance measures vary under different circumstances. Real roads and roadside surveys are likely to differ from those simulated in this study. First, our simulated roads were randomly located, whereas engineering and political decisions make real road placement geographically biased. Habitats never traversed by roads would be without sample data in surveys conducted exclusively along roads (Hanowski and Niemi 1995). Second, in our study the probability of surveying was equal for all locations within a rectangular quadrat; however, survey density might actually be inversely proportional to the distance from a road. Third, habitat types of real roadside areas often differ from the habitat types found in the surrounding landscape (Bart et al. 1995, Keller and Scallan 1999). A more sophisticated experiment will be required before we have a more complete understanding of the influence of biased surveys in general, and roadside surveys in particular, on predictions of species occupancy.

MacKenzie et al. (2002) and Tyre et al. (2003) show how to estimate false-absence rates from repeated surveys; however, when existing data are without measures of effort these procedures may be difficult to implement. Similarly, false absences are likely when sites with no data are randomly selected and used as ''absences'' with statistical programs requiring binary data. The robustness of statistical algorithms to assumed absences (and false absences) in randomly selected sites requires investigation. Rather than selecting ''absences'' randomly, it might be better to select from environmentally weighted locations (Zaniewski et al. 2002). Methods that model the association between presence-only sample data and environmental variables are another option (see Hirzel et al. 2002, Zaniewski et al. 2002).

There is a lack of consensus about which metrics best measure model performance (see Fielding and Bell 1997, Hirzel and Guisan 2002, Zaniewski et al. 2002, Anderson et al. 2003). We used multiple metrics to evaluate model performance for a number of reasons, one of which is that characteristics of the landscape affect measurement. Consider the following hypothetical example. Assume a study area that has 2500 cells and is 90% unoccupied and, therefore, composed mostly of zero values when depicted as a binary map (i.e., 10% of the grid cells would equal one, as is the case in our simulations). Evaluation with a residual sum of squares (RSS) metric for a model that predicts moderate values of probability of presence, e.g., 0.5, in all grid cells would conclude that the model performed poorly compared to a model that predicts small values of probability of presence, e.g., 0.1, in all cells. This conclusion would be drawn even when the probabilities predicted in truly occupied cells are always larger than the probabilities predicted in truly unoccupied cells. Under such an assumed distribution, a model that predicts a probability of presence of 0.1 for every cell would appear to perform better (RSS = 225), despite having zero discrimination (as measured by AUC), than a model that predicts probability of presence of 1 and 0.5 in all truly occupied (x_i = 1) and unoccupied (x_i = 0) cells, respectively (RSS = 563). Also, when calculating the AUC, the denominator of the false-presence fraction (number of unoccupied sites) would be much larger than that of the true-presence fraction (number of occupied sites), thus one commission error would have less effect than one omission error. The number of omission and commission errors is, as always, a direct result of the selected threshold (e.g., 0.5). Accordingly, the behavior of metrics used to evaluate spatial data continues to be an important area of investigation.

Conclusion

We used a simulated study design to test the effects from several factors on the performance of models that predict species distribution maps. Considering that the absolute values of our reported results are contingent upon factors such as the size of the distributions (10%), the simplicity of habitat relationships (only two controlling environmental variables), and sample size, we present this study as an initial comparison of performance among common survey designs.

Species distribution modeling increasingly uses objective methods due, in part, to the development of spatial statistical procedures and the availability of geospatial data. To improve the predictive power of new sample data, several design considerations appear to be important. First, maximize sample size and detection probability. Second, design for the contiguity of the distribution (e.g., increase effort for less contiguous distributions). Third, balance the amount of information collected from suitable and unsuitable habitats. Fourth, try to minimize error in the prediction variables (measurement error, survey error) since it too contributes to a general erosion of model performance. Further, if maximizing model performance based on AUC, RSS, and commission error are important, use a random or stratified design. To minimize omission error, an adaptive design could work best. If, instead, the study uses spatially biased data (most existing data), expect a reduction in model performance. Our results suggest that the predictive power of roadside survey designs is near that of systematic quadrat designs, at least for the ALR model. We argue that this is promising considering the abundance and availability of opportunistically collected sample data. However, we did not include some factors in our simulated road design. Thus, our model performance estimates could exceed the prediction potential of real road-biased data.

Our results, to an unknown degree, depend on the form of the model (e.g., ALR) and the specific spatial structure of our data (environmental variables, distributions, sample data). Therefore, additional research using other species occurrence modeling algorithms and varying spatial patterns would be useful. Ultimately, whether or not the maps created in predicting a species distribution are acceptable will depend on the application. For example, Dean et al. (1997) found that overlaying predicted species distributions containing as little as 5% error could alter the distribution of species richness. However, more prediction error might be allowed when managing for a threatened or endangered species, especially if no other data exists. Increased understanding of factors that affect model performance, such as how to incorporate environmental mechanisms (e.g., competition) into modeling efforts (see Morrison 2001), could further narrow the gap between prediction and truth.

Acknowledgments

We would like to thank D. Bender and S. Shriner for project design suggestions and reviews of earlier drafts. This project was funded by the USDA Forest Service, Southwestern Region (Contract #: 01-JV-11221611-177) and the National Science Foundation (Award #: DBI-0109969). We would also like to thank the two anonymous reviewers for suggestions that improved this article.

Literature Cited

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling **162**:211–232.

Augustin, N. H., M. A. Mugglestone, and S. T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology **33**:339–347.

Bart, J., M. Hofshen, and B. G. Peterjohn. 1995. Reliability of the Breeding Bird Survey: effects of restricting surveys to roads. Auk **112**:758–761.

Bojórquez-Tapia, L. A., I. Azuara, and E. Ezcurra. 1995. Identifying conservation priorities in Mexico through geographic information systems and modeling. Ecological Applications **5**:215–231.

Corsi, F., E. Duprè, and L. Boitani. 1999. A large-scale model of wolf distribution in Italy for conservation planning. Conservation Biology **13**:150–159.

Cressie, N. A. 1993. Statistics for spatial data. John Wiley and Sons, New York, New York, USA.

Dean, D. J., K. R. Wilson, and C. H. Flather. 1997. Spatial error analysis of species richness for a Gap analysis map. Photogrammetric Engineering and Remote Sensing **63**: 1211–1217.

Edwards, J. L., M. A. Lane, and E. S. Nielsen. 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. Science **289**:2312–2314.

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.

Fortin, M. J., P. Drapeau, and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio **83**:209–222.

Funk, V. A., and K. S. Richardson. 2002. Systematic data in biodiversity studies: use it or lose it. Systematic Biology **51**:303–316.

Garrison, B. A., R. A. Erickson, M. A. Patten, and I. C. Timossi. 2000. Accuracy of wildlife model predictions for bird species occurrences in California counties. Wildlife Society Bulletin **28**:667–674.

Hanowski, J. M., and G. J. Niemi. 1995. A comparison of on- and off-road bird counts: do you need to go off-road to count birds accurately? Journal of Field Ornithology **66**: 469–483.

Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling **157**:331–341.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology **83**:2027–2036.

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.

Janssen, L. L. F., and F. J. M. van der Wel. 1994. Accuracy assessment of satellite-derived land-cover data: a review. Photogrammetric Engineering and Remote Sensing **60**: 419–426.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications **14**:401–413.

Keller, C. M. E., and J. T. Scallan. 1999. Potential roadside biases due to habitat changes along breeding bird survey routes. Condor **101**:50–57.

Luoto, M., M. Kuussaari, and T. Toivonen. 2002. Modelling butterfly distribution based on remote sensing data. Journal of Biogeography **29**:1027–1037.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology **83**:2248–2255.

Morrison, M. L. 2001. A proposed research emphasis to overcome the limits of wildlife-habitat relationship studies. Journal of Wildlife Management **65**:613–623.

Pathirana, S. 1999. Distribution of errors in a classified map of satellite data. Geocarto International **14**:70–81.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling **133**:225–245.

Pearson, D. E., and L. F. Ruggiero. 2003. Transect versus grid trapping arrangements for sampling small-mammal communities. Wildlife Society Bulletin **31**:454–459.

Reddy, S., and L. M. Dávalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography **30**:1719–1727.

Scherrer, B. 1984. Biostatistique. Gatan Morin Editeur, Chicoutimi, Québec, Canada.

Shaffer, H. B., R. N. Fisher, and C. Davidson. 1998. The role of natural history collections in documenting species declines. Trends in Ecology and Evolution **13**:27–30.

Stockwell, D. R. B., and I. R. Noble. 1992. Induction of sets of rules from animal distribution data—a robust and informative method of data-analysis. Mathematics and Computers in Simulation **33**:385–390.

Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling **148**:1–13.

Stoms, D. M., F. W. Davis, and C. B. Cogan. 1992. Sensitivity of wildlife habitat models to uncertainties in GIS data. Photogrammetric Engineering and Remote Sensing **58**:843–850.

Thompson, S. K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association **85**:1050–1059.

Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications **13**:1790–1801.

Veregin, H. 1989. A taxonomy of error in spatial databases. Technical paper 89–12. National Center for Geographic Information and Analysis, University of California, Santa Barbara, California, USA.

White, G. C., and R. A. Garrott. 1986. Effects of biotelemetry triangulation error on detecting habitat selection. Journal of Wildlife Management **50**:509–513.

Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modeling **157**:261–280.

Zar, J. H. 1996. Biostatistical analysis. Prentice-Hall, Upper Saddle River, New Jersey, USA.

APPENDIX A

A general description and flow chart of methods used are available in ESA's Electronic Data Archive: *Ecological Archives* A015-014-A1.

APPENDIX B

Examples of the tested survey designs are available in ESA's Electronic Data Archive: *Ecological Archives* A015-014-A2.

SUPPLEMENT

A program containing algorithms to survey square distributions is available in ESA's Electronic Data Archive: *Ecological Archives* A015-014-S1.