The purpose of this thesis was to develop and evaluate predictive models of avian species occurrence across two years and three mountain ranges in the Great Basin of central Nevada. I modeled eighteen species’ occurrences and evaluated them using independent data and four different measures of accuracy; correct classification rate (CCR), sensitivity, specificity, and the area under a receiver-operating characteristic (ROC) curve (AUC). Species occurrence rate affected all accuracy measures except AUC. Temporal variation in species distributions had a greater effect on model accuracy than spatial variation, however this was likely due to the extreme annual differences in occurrence rates in this study. Furthermore, no life history characters examined showed significant variation in model performance except for levels of species commonness. This occurred mainly for the accuracy measures that were affected by occurrence rate, and for models developed and evaluated using the temporal data sets exhibiting markedly different occurrence rates.
AN EXPLORATION OF ACCURACY ISSUES REGARDING PREDICTIVE MODELS OF AVIAN OCCURRENCE IN THE CENTRAL GREAT BASIN

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Science
Department of Zoology

by
Lesley Brooke Penfield
Miami University
Oxford, OH
2003

Advisor________________
(Dr. Robert Blair)

Reader_________________
(Dr. Thomas Crist)
TABLE OF CONTENTS

LIST OF TABLES AND FIGURES iii
ACKNOWLEDGMENTS viii

CHAPTER 1: INTRODUCTION TO THESIS 1
Literature Cited 4

CHAPTER 2: THE EFFECTS OF SPATIAL AND TEMPORAL VARIATION IN SPECIES OCCURRENCE RATE ON THE ACCURACY OF OCCURRENCE MODELS
Abstract 6
Introduction 7
Methods 11
Results 14
Discussion 17
Acknowledgements 23
Literature Cited 24
Appendix 43

CHAPTER 3: DOES MODEL ACCURACY DIFFER ACROSS SPECIES ECOLOGICAL CHARACTERISTICS?
Abstract 45
Introduction 46
Methods 48
Results 53
Discussion 55
Acknowledgements 59
Literature Cited 60

CHAPTER 4: CONCLUSION TO THESIS 75
Literature Cited 78
CHAPTER 2

TABLE 1. Two-by two classification table of the observed and predicted occurrence of a species. Each value represents a number of observations such that A+B+C+D is equal to the total number of observations in the evaluation data set. The three measures of accuracy derived from this table and discussed in this chapter are: Correct classification rate (CCR), (A+D)/(A+B+C+D); Sensitivity, A/(A+C); and Specificity, D/(B+D).

TABLE 2. The ten data sets used to build and independently evaluate models of species occurrence.

TABLE 3. Independent variables derived from grids within a GIS. Asterisk means that quadratic terms were also used in the model building as potential predictor variables.

TABLE 4. Accuracy measures across all data sets. Rankings for “poor, good and excellent” are taken from Pearce and Ferrier (2000) (see text). Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima; TQ, Toquima; SH, Shoshone; and TY, Toiyabe.

TABLE 5. Mean and standard deviation for each species’ occurrence rate and model accuracy. Values for each species are averaged across the five data sets. See text for description of categories for discrimination ability.

TABLE 6. Frequency of each predictor in the 89 models constructed across five datasets for 18 species of breeding birds.

TABLE 7. Observed species occurrence rates in each individual mountain range during both the 2001 and 2002 breeding seasons.

FIGURE 1. Location of the Shoshone Mountains, Toquima and Toiyabe Ranges (black rectangle, see inset) in the Great Basin (irregular shape with black border, see inset). The inventoried canyons within the three mountain ranges are illustrated with black lines. Three pairs of canyons
In the Toquima Range and one pair of canyons in the Toiyabe Range connect at the crest of these ranges.

**FIGURE 2.** Point count sites within the five canyons that incise the east and west slopes of the Shoshone Mountain Range. These points are shown over a Digital Elevation Model (DEM).

**FIGURE 3.** Vegetation sampling arrangement at each 75-meter point count. The three outer circles represent the 11.3-meter sampling circles. Each circle is 30 meters from the point count center (sun symbol), and 120 degrees from each other. The direction of the first circle from the center was chosen randomly (Martin et al. 1997). “Transects” used for ocular tube readings are defined as the 30-meter path from the point count center to the sampling circles.

**FIGURE 4.** Effects of observed species occurrence rate on model accuracy in the model-building and evaluation data sets.

**FIGURE 5.** Effects of observed occurrence rate on sensitivity and specificity in the model-building and evaluation data sets.

**FIGURE 6.** Difference in observed species occurrence rates between all model-building and evaluation data sets: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.

**FIGURE 7.** Comparison of the distributions of accuracy measures for models assessing the effects of both spatial and temporal variation. Notice that the scale for sensitivity and specificity are different from the scale for AUC and CCR. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.

**APPENDIX.** Predictor variables that entered logistic regression models for each species in each data set and whether the relationship was positive (+) or negative (-). Abbreviations for variables
are listed in Table 6. Species names from the 43rd supplement to the AOU check-list of North American Birds (Banks et al. 2002).
CHAPTER 3

TABLE 1. Independent variables derived from grids within a GIS. Asterisk means that quadratic terms were also used in the model building as potential predictor variables.

TABLE 2. The ten datasets used to build and independently evaluate models of species occurrence.

TABLE 3. Two-by two classification table of the observed and predicted occurrence of a species. Each value represents a number of observations such that A+B+C+D is equal to the total number of observations in the evaluation data set. The three measures of accuracy derived from this table and discussed in this chapter are: Correct classification rate (CCR), (A+D)/(A+B+C+D); Sensitivity, A/(A+C); and Specificity, D/(B+D).

TABLE 4. Ecological characters and the observed mean occurrence rate for each species modeled. Occurrence rates were averaged across 2001 and 2002. (short distance migrant, SDM; Neotropical migrant bird, NTMB; Resident, RES; woodland breeding species, WL; scrub-successional species, SS; and urban breeding species, URB)

TABLE 5. Accuracy measures across all data sets. Rankings for “poor, good and excellent” are taken from Pearce and Ferrier (2000) (see text). Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima; TQ, Toquima; SH, Shoshone; and TY, Toiyabe.

TABLE 6. Mean and standard deviation for each species’ occurrence rate and model accuracy. Values for each species are averaged across the five data sets. See text for description of categories for discrimination ability.

TABLE 7. Results for ANOVA test of whether some levels of species commonness were modeled more accurately than others. F and P values are displayed. Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. Values in bold were found to be significantly different using a Bonferroni adjusted p value where p < 0.0025 indicates a significant difference (see text).
FIGURE 1. Location of the Shoshone Mountains, Toquima and Toiyabe Ranges (*black rectangle, see inset*) in the Great Basin (*irregular shape with black border, see inset*). The inventoried canyons within the three mountain ranges are illustrated with black lines. Three pairs of canyons in the Toquima Range and one pair of canyons in the Toiyabe Range connect at the crest of these ranges.

FIGURE 2. Vegetation sampling arrangement at each 75-meter point count. The three outer circles represent the 11.3-meter sampling circles. Each circle is 30 meters from the point count center (sun symbol), and 120 degrees from each other. The direction of the first circle from the center was chosen randomly (Martin et al. 1997). “Transects” used for ocular tube readings are defined as the 30-meter path from the point count center to the sampling circles.

FIGURE 3. Differences in accuracy across levels of species commonness in the 2001 data sets. The letters above each individual graph show which groups are significantly different (p < 0.0025). The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.

FIGURE 4. Difference in species occurrence rates between all model-building and evaluation data sets: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.
I am thankful to my advisor, Dr. Robert Blair, for his encouragement and guidance in focusing my ideas into a manageable project as well as his general support throughout the last two years. I am also thankful to my other committee members, Drs. Tom Crist and Robert Schaefer, for keeping me honest with their statistical and procedural help. I would also like to thank Dr. Erica Fleishman for her continued feedback, expertise and encouragement at all stages of this project; John Fay for providing the GIS grids used to derive predictor variables; Avram Primack for general GIS assistance; Chris Betrus for his extremely valuable help getting ready for fieldwork in the Great Basin; Dr. Eric Porter for his time spent discussing conceptual development and management implications; Jason Bulluck for his editing assistance, long hours spent in the field measuring dbh, moral support and for sharing his passion for applied ecology and conservation with me; and to my family, Helen, James, Clint and Elizabeth, for their acceptance and support of the path that I have chosen in life. Financial assistance for this project was provided by Miami University; Stanford University; the Nevada Biodiversity Research and Conservation Initiative; and the Joint Fire Sciences Program via the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture.
CHAPTER 1:  
INTRODUCTION TO THESIS

The popularity of predictive models of species occurrence has increased in recent decades, but their origin extends back much further. Prior to the 1950s, researchers focused on recording detailed natural history and qualitative relationships. It was not until the mid to late 1950s that quantitative ecology surfaced as a discipline (Stauffer 2002). It has been suggested that McArthur and McArthur (1961) presented an early analogue of such a model when they described a correlation between foliage height diversity and bird species diversity (Stauffer 2002). In 1984, a symposium was held to bring together those involved with and to report the status of species-habitat modeling. This meeting resulted in the publication of *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates* (Verner et al. 1986), and brought forth issues regarding the need for increased communication between managers and researchers (Stauffer 2002) and for the use of accuracy assessments. In 1999, another symposium was held that resulted in another publication, this time with a goal of assessing the advancements of habitat models since the last symposium. The most prevalent topics were those regarding spatial scale and accuracy assessment; hence the name of the resulting publication: *Predicting species occurrences: Issues of accuracy and scale* (Scott et al. 2002). In addition to accuracy, the need for increased communication between researchers and managers was another recurring theme from nearly two decades before.

Recent literature (Scott et al. 2002) on species occurrence models has focused on issues of accuracy and scale as a result of advancements in Geographic Information Systems (GIS) and sophisticated statistical procedures. GIS has allowed researchers to not only create spatially explicit models, but also to ask novel questions regarding the various levels at which species may be selecting habitat (Mitchell et al. 2001). An entirely new set of predictor variables are now available including, but not limited to, land use and vegetation characteristics derived from remote sensing and various topographic measures derived from Digital Elevation Models (DEM). Maps of such predictor variables illustrate the degree of spatial variation in environmental variables and allow for tests of how this variation affects species distributions and model accuracy. Furthermore, GIS has made it a fairly easy task to create detailed maps that not
only enable the development of model predictor variables, but also communicate spatial trends better than words alone.

The great majority of species-habitat models developed today use GIS in some aspect, and its popularity has some researchers concerned (Van Horne 2002). Technology is undoubtedly advancing at an unprecedented rate and along with new technology comes an increased likelihood for error. Van Horne (2002) states several concerns that need to be addressed “before we leap wholesale into this new, spatially explicit world” including the possible tendency of researchers to assume that relationships between species and vegetation types are correct and therefore avoid field verification. These concerns are real and practitioners should use caution when drawing conclusions from spatial analyses using GIS or remotely sensed data.

The use of sophisticated statistical procedures has enabled researchers to ask new questions and have greater precision when modeling species distributions. Recent publications involving species occurrence models employ a great number of modeling and evaluation techniques and some studies that compare the performance of several modeling procedures (Dettmers et al. 2002) and accuracy measures (Fielding 2002, Pearce and Ferrier 2000). Two examples of recent advancements in model development techniques include the incorporation of spatial-autocorrelation in order to increase predictive success (Augustin et al. 1996, Klute et al. 2002) and the use of Bayesian statistics due to its proposed benefits when adaptive management is desirable (Bergerud and Reed 1998). Advancements in assessment of model accuracy include the need to evaluate models using data independent from that used in model development (Fielding and Bell 1997, Fielding 2002) and to use accuracy measures that are not affected by species occurrence rates (Pearce and Ferrier 2000, Manel et al. 2001, Fielding 2002). The concern regarding evaluating models with independent data appears to have been adopted by many practitioners. However, the most common accuracy measure presented by researchers is the correct classification rate (CCR), or the proportion of sites classified correctly, and many have demonstrated this measure to be greatly affected by species occurrence rates (Fielding and Bell 1997, Pearce and Ferrier 2000, Manel et al. 2001). These same researchers suggest the use of the area under a Receiver Operating Characteristic (ROC) curve (AUC: area under the curve) as a better measure that is not affected by species prevalence. ROC curves have previously been used in medical research to determine the accuracy of medical treatments (Hanley and McNeil...
1982, Swets 1998, Zweig and Campbell 1993), and have only recently emerged as a tool in ecological studies (Fielding and Bell 1997).

Based on the current status of species occurrence models, I chose to address issues pertaining to model accuracy for my thesis. I wanted to determine if the accuracy of avian occurrence models in central Nevada was affected by spatial and temporal variation in species distributions. Central Nevada is in the Great Basin region, which is comprised of numerous nearly parallel mountain ranges with riparian canyons that provide habitat for many breeding birds. Such a region provides a unique opportunity to assess the effects of spatial variation in species distributions on model accuracy. Furthermore, I used a GIS in this study to allow for incorporation of topographic as well as vegetation predictor variables, and to enable quantification of spatial variation in environmental variables. I also assessed the effects of species occurrence rate and the number of predictor variables on four different accuracy measures. In most studies of model accuracy, the authors show that some species are modeled better than others. Some even go so far as to group species by life history characters and describe a certain group of species as having higher accuracy than others (Mitchell et al. 2001, Pearce et al. 2001, Hepinstall et al. 2002). Therefore, I compared model performance across five ecological characters to determine if model performance could be generalized among similar species.
LITERATURE CITED


CHAPTER 2:
THE EFFECTS OF SPATIAL AND TEMPORAL VARIATION IN SPECIES OCCURRENCE RATE ON THE ACCURACY OF OCCURRENCE MODELS

ABSTRACT

Predictive models of species occurrence have great potential for those trying to prioritize areas for competing land uses, especially when attempting to manage for a certain species. Before widespread application, however, it is necessary that models be thoroughly evaluated using independent data and an effective measure of accuracy. I predicted the distributions of 18 species of breeding birds in three adjacent mountain ranges in the Great Basin. For each species, I built separate models using five different data sets—one set of data for each of two years, and one set for each of the three possible pairs of mountain ranges. I then evaluated each model with independent data using four different measures of accuracy. The four measures were discrimination ability [presented as the area under a receiver operating characteristic plot (AUC)], correct classification rate (CCR), the proportion of presences correctly classified (sensitivity), and the proportion of absences correctly classified (specificity). The objectives of this study were (1) to compare the effects of occurrence rate (proportion of locations occupied) on the four accuracy measures, (2) to assess the effects of spatial and temporal variation in occurrence rate on model accuracy and (3) to determine if the number of predictor variables in a model affects accuracy. Discrimination ability was not affected by occurrence rate, whereas the other three measures of accuracy were significantly affected. The effect of occurrence rate on CCR, sensitivity, and specificity in the evaluation data sets was much greater than in model-building data sets. Temporal variation in species occurrence appears to have a greater impact on model accuracy than does spatial variation. However, the data presented here are from two consecutive years where the occurrence rates were quite different. Discrimination ability was the only accuracy measure that was affected by the number of variables in a model; however, model development was carried out in such a way that overfitting of data was minimized. Overall, models performed well when evaluated with independent data. Those using predictive models of occurrence in a management setting should assess the relative costs of omission and commission errors and collect long-term census data when temporal variation in species distributions is expected to be great.
INTRODUCTION

The accuracy of predictive models of species distributions has become a fundamental concern for both ecologists and land managers. Interest in these models is due, in part, to the numerous direct applications that stem from being able to predict where a species is likely to be present and absent. Of particular interest are those applications directed at conservation. The occurrence of rare and endangered species can be modeled with hopes of determining whether their distributions coincide with current land-use practices that are suitable for their protection (Scott et al. 1993, Caicco et al. 1995), or in order to prioritize lands for different land uses (Peterson et al. 2000). Predictive models of occurrence can also be utilized alongside restoration ecology plans to predict how certain species might respond to various revegetation scenarios (Fleishman and Mac Nally 2004). Other applications of modeling species occurrences include predicting where invasive species are likely to colonize (Wadsworth et al. 2000), and predicting shifts in species distributions due to climate change (Aspinall and Matthews 1994).

Predictive models have also become popular due to the need for efficiency in the design and implementation of conservation management. Traditionally, conservation research has been based on collection of detailed natural history data (Stauffer 2002). Quantitative ecology recently has emerged as a discipline that has the potential to increase efficiency in the land-use decision-making process. The use of Geographic Information Systems (GIS) and sophisticated statistical procedures has facilitated this change. GIS allows for management of broad-scale data sets from statewide or regional to continental scales (Boone and Krohn 2000, Peterson 2001), and thus affords the ability to build and evaluate models of species distributions at multiple levels. Landscape and regional scenarios appear to be the most useful to managers because land-use decisions are often made at these spatial extents. To make land-use decisions based on models of species distributions, one must assess the degree of spatial and temporal variation in a system and thereby understand the limits of a model’s usefulness. A model built using data from one time or place may not be suitable for use in another time or place. However, if model predictions are focused on a location or period of time similar to that used to develop the model, and the species being modeled are not extremely plastic in their habitat requirements, models of occurrence can be quite reliable.

Predictive models are generally static, having no intrinsic ability to account for variation in the distribution of species or environmental variables in space and time (Guisan and
Zimmerman 2000). In contrast, ecosystems and the organisms inhabiting them are remarkably
dynamic, such that prediction errors are inevitable (Fielding 2002). Therefore, we should avoid
the assumption that models are accurate representations of reality (Starfield 1997). Rather, we
should assess their usefulness in terms of relative reliability across space and time. Furthermore,
selection and presentation of accuracy measures for predictive models should convey the
understanding that models are estimations of reality and are affected by natural variation.

Despite the widespread use of species distribution models, many researchers provide little
to no evaluation of model performance in a predictive capacity (Manel et al. 2001). The few
evaluations often simply state a proportion of sites in which the presence or absence of the
species is predicted correctly. Such a measure of accuracy is sometimes called a correct
classification rate (CCR) and is calculated from a two-by-two classification table (Table 1).
Measures of CCR can be misleading in that they are affected by the relative occurrence rate of
the species being modeled (Pearce and Ferrier 2000b, Manel et al. 2001). In addition, a species
is predicted to be present or absent at a given site based on an arbitrary probability threshold
selected by the researcher. For example, if the probability threshold is set at 0.5, then a species is
considered to be present whenever the predicted probability of occurrence is greater than 0.5 and
to be absent whenever the predicted probability of occurrence is less than 0.5. This confounds
assessment of model performance because every threshold is associated with a different level of
accuracy (Pearce et al. 2002, Fleishman et al. 2003). Regardless of the probability threshold
chosen, it can be helpful for a researcher to separately assess a model’s ability to predict
presences and absences. This can be done by calculating sensitivity and specificity from the
values in a classification table for a particular probability threshold. Sensitivity is the proportion
of true positives or sites correctly predicted as occupied (Table 1). Specificity is the proportion
of true negatives or sites correctly classified as being unoccupied (Table 1).

Another measure of model performance is the area under a receiver operating
characteristic (ROC) curve. ROC curves have primarily been used in medical applications, but
have recently been used in ecological studies (Manel et al. 2001). ROC curves are an excellent
accuracy measure because they are not based on arbitrary thresholds and they are less sensitive
to occurrence rate than other commonly used evaluation methods (Pearce and Ferrier 2000b,
Manel et al. 2001, Pearce et al. 2002). The area under an ROC curve (AUC) provides a measure
of discrimination ability (i.e. how well a model can distinguish between occupied and vacant
sites) (Pearce and Ferrier 2000a). The ROC curve plots sensitivity against the proportion of false negatives (1-specificity) (Pearce and Ferrier 2000b) for several probability thresholds. AUC values range from 0.5 (no discrimination ability) to 1.0 (perfect discrimination). Values greater than 0.9 indicate excellent discrimination ability; values between 0.7 and 0.9 indicate a usable, reasonable level of discrimination; and values between 0.5 and 0.7 indicate poor discrimination ability (Pearce and Ferrier 2000b). Hanley and McNeil (1982) demonstrated that AUC values could be interpreted as a probability that a model will distinguish between an occupied and an unoccupied site.

Topographic variation present in the Great Basin of western North America makes the system useful for building and evaluating models of species distribution. The Great Basin includes more than 200 parallel mountain ranges, most arranged in a north-south orientation. These ranges were isolated from each other and the intervening valleys as the region became warmer and drier following the Pleistocene. Hence, these mountain ranges represent distinct habitat islands in the Great Basin landscape (Brown 1978, Grayson 1993). The numerous canyons that drain the east and west slopes of these mountain ranges also may act as habitat islands for some species due to the riparian vegetation and resources they harbor relative to the intervening uplands (McDonald and Brown 1992).

Three adjacent mountain ranges, the Toquima Range, Toiyabe Range, and Shoshone Mountains (Lander and Nye counties, Nevada), were the focus of my study (Figure 1). These ranges are in the same biogeographic subregion (Austin and Murphy 1987, Grayson 1993) and therefore generally contain similar biotic communities. However, differences in elevation and precipitation among mountain ranges lead to among-range differences in vegetation composition and abundance (Grayson 1993). For example, the Toiyabe Range has the highest mean elevation and receives more precipitation annually than the other two ranges. Consequently, the canyons that drain its east and west slopes have more riparian vegetation [e.g. aspen (*Populus tremuloides*), willow (*Salix* spp.), birch (*Betula occidentalis*) and rose (*Rosa woodsii*)]. The relatively dry canyons of the Toquima Range, by contrast, are dominated by piñon (*Pinus monophylla*), juniper (*Juniperus osteosperma*), sagebrush (*Artemisia tridentata* spp.) and rabbitbrush (*Chrysothamnus* spp.), with comparatively little riparian vegetation. The Shoshone Mountains are more mesic than the Toquima Range, but drier than the Toiyabe Range. Consequently, several of the Shoshone canyons have a high proportion of piñon-juniper woodlands and sagebrush with occasional patches
of riparian vegetation. Thus, these three mountain ranges offer an opportunity to test hypotheses regarding how predictive models of species occurrence perform when faced with spatial variation in the distribution of habitats.

Fleishman et al. (2001, 2003) constructed butterfly occurrence models using data from the Toquima Range and evaluated them in the Shoshone Mountains. These butterfly models predicted both presences and absences fairly well (Fleishman et al. 2001, 2003), suggesting that the spatial applicability of the models extended beyond the mountain range in which they were built. Predictions of absence consistently had higher success rates than predictions of presence, which is often the case for species with low occurrence rates (< 0.5). I expected that models for breeding birds with occurrence rates similar to butterflies, constructed and evaluated in the same region, would also yield good results. I anticipated this success for two reasons. First, the species composition of Great Basin birds exhibits a distinct stratification across an elevational gradient (Medin et al. 2000), making their distributions relatively easy to predict based on topographic variables in addition to vegetation. Second, other researchers successfully have modeled bird distributions (Dettmers and Bart 1999, Manel et al. 2001, Pearce et al. 2001, Pearce et al. 2002).

In this study, I built and evaluated predictive models of breeding bird distributions in the Toquima Range, Toiyabe Range, and Shoshone Mountains. For each of 18 species I constructed predictive models using five different subsets of the available data, two temporal (2001 and 2002) and three spatial (Toiyabe / Toquima, Toiyabe / Shoshone, and Toquima / Shoshone), and then evaluated each model using independent data (i.e., the data not used to build the model) (Table 2). Although it may not seem rational to use 2002 data to predict species occurrences in 2001, these two data sets describe species distributions in individual years and therefore provide the opportunity to test hypotheses regarding how models perform when faced with two different scenarios involving temporal variation. Similarly, due to the distinct differences in mountain ranges mentioned above, it may not seem appropriate to use dry and mesic mountain ranges to predict species occurrences in a wet canyon, or any other combination of wet or dry to predict the other. However, land managers are rarely afforded the opportunity to collect a large data set spanning an entire region that allows for random selection of model-building and evaluation data sets. The framework presented here is similar to a real world scenario where data is collected in one or two management units (e.g. mountain ranges) with the goal of modeling species in a nearby
management unit. Furthermore, sample size for species presences are sometimes too small in one mountain range to build a stable model in just one range.

I used models constructed for 18 species across five data sets (Table 2) to address three main issues. First, I compared four different measures of model accuracy and determined whether they were correlated and whether they were affected by occurrence rate. The four accuracy measures were discrimination ability [the area under a Receiver Operating Characteristic curve (AUC)], correct classification rate (CCR), sensitivity and specificity. Second, I assessed whether the accuracy of predictive models of species distributions were affected by both spatial and temporal variation in occurrence rates. Third, I determined whether the number of variables in a model affected each measure of model accuracy.

METHODS

Field methods

Avian censuses were conducted throughout five canyons in the Shoshone Mountains (Figure 2) and Toiyabe Range, and six canyons in the Toquima Range, during the breeding seasons of 2001 and 2002 using 75m-fixed-radius point counts (Ralph et al. 1995, Dobkin and Rich 1998). Two individuals conducted the point counts; one in 2001 and another in 2002. Within each canyon, we placed point counts at least 200 meters apart in each of the dominant vegetation types. During each of the two years, we visited sites three times throughout the breeding season (May 27-June 30) between sunrise (approximately 0530) and 3 hrs after sunrise (0830). During each 5 min visit, we recorded all birds seen or heard using terrestrial habitat. We sampled each site both very early as well as later in the morning. This enabled us to account for temporal variation in detection probability; most birds tend to sing less as the daily temperature increases and different birds tend to sing at different intensities throughout the morning. We allowed for at least ten days between each visit to each site to account for temporal variation throughout the breeding season.

Three researchers collected vegetation data during the breeding season of 2002. All vegetation data were collected within three 11.3m-radius circles (0.04 hectares) near the center of each 75m-radius point count site (Figure 3) (Martin et al. 1997). We gathered vegetation density readings at each cardinal direction in each circle using a density board (Noon 1981). We also measured height and diameter at breast height (dbh) of all trees within each 11.3m-radius circle.
We collected 22 ocular tube readings from each point count site (Figure 3) and used these to calculate percent cover of canopy, shrub, and ground vegetation (Noon 1981). We also recorded the presence, length (or height) and dbh of dead wood (logs and snags) and the presence, mean width and type of water (i.e. seep, flowing, and standing) present at each point count site.

Model development

I used multiple logistic regression to construct predictive models of species occurrence (presence/absence). Potential explanatory variables included environmental predictors measured with a GIS and vegetation data collected at each census site. Using a GIS, I derived environmental values from several grids within a 90-meter buffer around the center of each point (Table 3). The grid for mean annual precipitation was derived from PRISM (Daly et al. 1994). Grids representing distance to water, elevation, slope, aspect as described by mean eastness and mean northness, and topographic exposure were derived using the spatial analyst extension in ArcView 3.2 (ESRI 1996). I considered quadratic terms for precipitation, elevation topographic exposure and slope during model construction (see below). This subset of variables were selected based on the ecology of the species and the possibility of a non-linear relationship between occurrence and these explanatory terms. Finally, I used a measure of primary productivity, the Normalized Difference Vegetation Index (NDVI), which was obtained from a satellite image taken on June 2, 2000. Seto et al. (in review) demonstrated NDVI to have a significant relationship with avian species richness in the same mountain ranges used in this study. All landscape grids had a cell resolution of 30m except for the PRISM grid that had 4km cell resolution. Several sampling points often occurred within one of these PRISM grid cells, and I consider this as several points with the same value for mean annual precipitation. Due to a lack of weather stations in this region, this is the closest approximation to rainfall available.

I built models of species occurrence using five different data sets (two temporal and three spatial) and validated each of the five sets of predictions using an independent set of data not used to build the model (see Table 2). Due to the large number of potential independent variables, I took measures to avoid multi-collinearity. Prior to model construction for each species, I conducted Pearson correlation analyses and defined any combination of two independent variables as highly correlated if \( r > 0.40 \). I then conducted univariate logistic regression analyses with each independent variable and the occurrence of the modeled species. From these analyses, I retained
all variables with log-likelihood $p < 0.25$ as potential variables for the final, multiple regression model (Hosmer and Lemeshow 1989). If two independent variables were correlated in the Pearson correlation analyses, I retained the one with the lower $p$-value. However, if the two $p$-values were the same, I ran a multivariate model with each separately.

I built multiple logistic models for 18 species for which the number of candidate predictor variables was less than the number of sites in which they were present. Across all five model-building data sets, the number of presences of these species ranged from three to 130 (mean 38.2). I used a stepwise selection procedure followed by a best subset selection during model construction. In order to obtain a potential model for each significance level, I used stepwise selection (SAS V8.0) with high alpha values (0.9) (Shtatland 2001). I then used the smallest Akaike’s Information Criterion (AIC) value to select the optimal model (Burnham and Anderson 1998). AIC is a model selection criterion that attempts to balance model complexity (overfitting) and bias (underfitting) using the principle of parsimony (Burnham and Anderson 2001).

Next, I conducted a best subset regression with all models containing one predictor variable more or less than the AIC-optimal model from the stepwise procedure (Shtatland 2001). For example, if the model with the smallest AIC value from the stepwise procedure had three predictor variables, I ran a best subset regression with all models having two, three, and four predictor variables. I again selected the model with the lowest AIC value and used this as the final predictive model for each species in each model-building data set. Model hierarchy was maintained such that models with a quadratic term were considered only if the corresponding linear term for that variable also entered the model.

**Model evaluation and comparisons**

For each data set, I calculated sensitivity (the proportion of true positives, Table 1) and specificity (proportion of true negatives, Table 1) as well as an overall correct classification rate (CCR) for a probability threshold of 0.7 and compared the results to the area under a Receiver Operating Characteristic curve (AUC) (SAS V8.0). I used a probability threshold of 0.7 because I preferred to have fewer overall predictions of higher quality (Fleishman et al 2003). I performed a simple linear regression between each accuracy measure (AUC, CCR, sensitivity and specificity), to explore the association between each. I also used simple linear regression to assess the relationship between each accuracy measure and species occurrence rates in both the model-
building and evaluation data sets. To compare mean values of the different accuracy measures across model type (spatial versus temporal) I conducted two-tailed t-tests. In addition, I used simple linear regression to test whether the number of variables in a model affected the four measures of model accuracy.

RESULTS

Overall model performance

I built models for 18 species of breeding birds across the five data sets (two temporal and three spatial) for a total of 90 models. I present the results of 89 models evaluated with independent data; a model for Sage Thrasher in Toquima/Shoshone was omitted due to the small number of occurrences in the evaluation data set.

The mean discrimination ability of each of the five sets of models (i.e., models constructed using different sets of data) was good (AUC > 0.7) when evaluated with independent data (Table 4). Furthermore, more than 55% of models within each data set had good discrimination (Table 4). Only 3% of all models were excellent (AUC > 0.9), and nearly 40% of the models were poor (AUC < 0.7). Model discrimination in the evaluation data sets (when tested on independent data) was not correlated with original model discrimination (when tested on data used to build the models) (R² = 0.001, p = 0.70). In other words, explanatory and predictive ability of the models was not correlated.

The correct classification rate (CCR, probability threshold = 0.7) was greater than 70% for all data sets when evaluated with independent data (Table 4). In other words, more than 70% of the predictions were correct when the predicted probability of presence was equal to 0.7. Four data sets had a mean CCR greater than 77%, and one data set (2002) had a mean CCR of 85%. In each of the data sets, specificity was consistently much greater than sensitivity, meaning that species absences were predicted more accurately than species presences (Table 4). Indeed, there was a negative relationship between sensitivity and specificity across all models. Similarly, there was a negative relationship between sensitivity and CCR. In contrast, there was a positive relationship between a model’s ability to predict absences (specificity) and CCR, and a positive relationship between a model’s ability to predict presences (sensitivity) and AUC.
Effect of occurrence rate

The observed species occurrence rate (proportion of sites occupied) was positively correlated between each of the model-building and evaluation data sets ($R^2 = 0.37$, $p < 0.0001$). This suggests that within this study area, if a species was present in a high proportion of sites in the model-building data set, it was also likely to be present in a high proportion of sites in the evaluation data set. Despite this correlation, there was a relationship between model accuracy and the observed occurrence rate in the model-building data and between model accuracy and the observed occurrence rate in the evaluation data. For example, there was a positive quadratic relationship between CCR and occurrence rate in the model-building data set ($R^2 = 0.22$, $p < 0.0001$) and between CCR and occurrence rate in the evaluation data set ($R^2 = 0.80$, $p < 0.0001$). Occurrence rate in the evaluation data set explained 80% of the variation in CCR (squared), while occurrence rate in the model-building data set explained only 22% (Figure 4). In general, CCR values were consistently quite high. There was a positive linear relationship between sensitivity and occurrence rate in the model-building data sets ($R^2 = 0.293$, $p < 0.0001$), and between sensitivity and occurrence rate in the evaluation data sets ($R^2 = 0.81$, $p < 0.0001$). In general, a model’s ability to predict presences increased when species had higher rates of occurrence (Figure 4). However, there was a negative quadratic relationship between model specificity and occurrence in both the model-building and evaluation data sets ($R^2 = 0.334$, $p < 0.0001$, $R^2 = 0.84$, $p < 0.0001$) with more variance in specificity explained by occurrence rate in the evaluation data sets (Figure 4). The differential effects of occurrence rate in the model-building and evaluation data sets on specificity and sensitivity are illustrated in Figure 5. Nonetheless, there was more variation in specificity and sensitivity for species with high rates of occurrence in the model-building data set (Figure 5). Table 5 shows the values for mean occurrence rate and mean accuracy for each species modeled. Neither the relationship between model discrimination (AUC) and species occurrence rate in the model-building stage ($R^2 = 0.03$, $p < 0.11$) nor the relationship between AUC and occurrence rate in the evaluation stage ($R^2 = 0.01$, $p = 0.31$) was significant (Figure 4).

Although the observed occurrence rates for the model-building and evaluation data sets were correlated, certain species often had higher rates of occurrence in one of those data sets. The difference in species occurrence rates between model-building and evaluation data sets therefore helps to explain variability in model accuracy across space or time (i.e. in spatial or
temporal models, respectively). It should be noted that because species were present in more sites in 2002 than in 2001, all differences between the 2001 model-building and 2002 evaluation data sets were negative (Figure 6). Indeed, the differences in occurrence rates between these temporal data sets are mirror images of one another, while the differences in occurrence rates between spatial build and evaluation data sets are not mirror images due, in part, to differing sample sizes (Table 2).

Across all models, AUC did not change as the absolute value of this difference in occurrence rates increased ($R^2 = 0.03, p = 0.10$). However, when each data set was considered independently, the difference in occurrence rates between the 2001 model-building and 2002 evaluation data sets had a significant positive relationship with discrimination ability ($R^2 = 0.25, p = 0.04$). Across all models there was no relationship between the difference in occurrence rates between the model building and evaluation data sets and CCR ($R^2 = 0.04, p = 0.08$). However, when each data set was considered individually, the difference in species occurrence rates between 2001 model-building data and 2002 evaluation data showed a significant negative relationship with CCR ($R^2 = 0.5, p = 0.001$) as well as the difference in occurrence rates between 2002 build and 2001 evaluation data sets and CCR ($R^2 = 0.41, p = 0.003$). As the difference in occurrence rates between the model-building and evaluation data sets increased, there was no change in a model’s ability to predict presences (sensitivity) ($R^2 = 0.03, p = 0.08$) or in its ability to predict absences (specificity) ($R^2 = 0.01, p = 0.27$), even when each data set was assessed individually.

Effect of model predictor variables

Of the 18 predictor variables used during model construction, 11 were derived from a GIS and seven were collected in the field at each point count site (Table 6). Four of the five predictors that entered most frequently were GIS-derived variables (elevation, precipitation, NDVI, and slope). Elevation and canopy cover were the most common predictors in species occurrence models; elevation entered 31% of the models while canopy cover entered 29% of the models (see Appendix). Quadratic terms for elevation, slope and precipitation entered 9%, 4% and 6% of the models, respectively. Six models (6.7%) had only predictors that were collected in the field and 17 models (19.1%) had only GIS derived predictors.
The number of predictors per model ranged from one to seven. Most models had three or four predictors (31% and 28%, respectively). In general, as the number of predictor variables in a model increased, model discrimination (AUC) increased as well ($R^2 = 0.21$, $p < 0.0001$). This was true for the models built with the 2001 data set and evaluated with 2002 data and for the models built with the Toiyabe / Toquima data set and tested with Shoshone data. Similarly, there was a positive relationship between sensitivity and the number of variables in a model ($R^2 = 0.06$, $p = 0.017$), but when I considered each data set individually, only the models developed with 2001 data (and evaluated with the 2002 data) showed this relationship. There was no relationship between the number of predictor variables and CCR or specificity.

Spatial and temporal variation

Whether a model predicted species occurrence in space (Toiyabe / Toquima, Toiyabe / Shoshone, and Toquima / Shoshone) or time (2001 and 2002) had no effect on model discrimination. The mean AUC for spatial models (0.73) was no different than that for temporal models (0.73) (Figure 7). However, temporal models displayed a higher average CCR and a higher average specificity relative to spatial models; i.e., temporal models classified absences better than spatial models. In contrast, spatial models exhibited higher average sensitivity than temporal models; i.e., spatial models classified species presences better than temporal models (Figure 7). Overall, the variation in model accuracy was greater in spatial models than in temporal models (Figure 7).

DISCUSSION

Overall model performance

The avian occurrence models presented here appear to perform well when evaluated with independent data from another time or place. The spatial and temporal extent across which these models can be applied makes their use quite practical when one considers the scales of most management decisions. Within a region, it is realistic to assume that land-use decisions might be based on occurrence models derived from census data collected in a previous year or years and within the local landscape [e.g. nearby mountain range(s)]. Therefore, the framework presented in this paper for the development and evaluation of species occurrence models can be applied in any region and potentially for a variety of taxonomic groups.
My results illustrate the importance of calculating more than one measure of model accuracy. Fielding and Bell (1997) and Manel et al. (1999) have also suggested that multiple accuracy measures are quite helpful when trying to discern the overall value of predictive occurrence models. Manel et al. (1999) used the phrase “deconstruction of overall predictive success” when reporting both the accuracy of predictions of presences (sensitivity) and that of predictions of absences (specificity). Several other accuracy measures can be easily calculated from the classification table (Table 1). Forbes (1995), Fielding and Bell (1997) and Manel et al. (2001) discuss calculation, advantages, and disadvantages of these other measures in detail.

There are two major drawbacks associated with measures derived from classification tables. They are often based on an arbitrarily chosen probability threshold value and their accuracy can be greatly affected by occurrence rate, as seen in three of the four accuracy measures presented here. However, it is still informative to calculate these values separately rather than as an overall measure of accuracy to assess the ability of a model to predict presences and absences.

**Effect of occurrence rate**

It is interesting that species occurrence rates in the evaluation data sets had a much stronger relationship with CCR, sensitivity and specificity than did occurrence rates in the model-building data sets (Figures 4 and 5). A high percentage of the variation in the latter three accuracy measures was explained by species occurrence rates in evaluation data sets. This makes sense because accuracy measures describe how well models perform when evaluated on independent data. Therefore, the observed occurrence rates in the evaluation data sets will affect the accuracy measures more so than the occurrence rates in the model-building stage. Unfortunately, one cannot forecast the occurrence rate of a species in the future unless several years of survey data exist. While multiple years of sampling data allow for a better understanding of the temporal variation in species distributions, these data also make the use of predictive models unnecessary. Because yearly censuses are not always financially feasible, I suggest that surveys occur on a regular basis every 3-4 years to ensure that occurrence models are still accurate.

Other studies corroborate my findings that species absences were predicted more accurately than species presences (i.e. specificity was greater than sensitivity) even when the overall predictive success (CCR) was considered to be good (Manel et al. 1999, Fleishman and
One explanation is that occurrence rate has considerable influence on accuracy measures. The average species occurrence rate in this study was 0.28 and ranged from 0.01 to 0.85. The trend presented here and commonly seen in the literature - that low occurrence rate often leads to low model sensitivity - highlights the difficulty in predicting the presence of rare species. This is important considering that rare species are often the focus of conservation planning and land management in general. Correctly predicted absences are helpful in determining where not to prioritize species conservation over land uses that might compete with that objective. However, knowledge of where a species is absent is not sufficient for delineating conservation areas. Obviously, when rare and endangered species are the focus of management decisions, extra care must be taken to fully understand those species’ habitat needs. Species occurrence models should not be considered a panacea in these situations.

The decision threshold used in this study is another explanation for why the ability of models to predict absences was consistently much greater than their ability to predict absences. I used a threshold of 0.7, where a species was considered to be present when the probability of occurrence was > 0.7, as opposed to a threshold of 0.5. It has been demonstrated that stricter classification rules (i.e. decision thresholds) may produce fewer overall predictions of higher quality compared with more lenient classification rules (Fleishman et al. 2003). Zweig and Campbell (1993) suggest that researchers determine the optimal threshold value for each individual circumstance based on “costs” assigned to the two types of possible errors common to binary models. Specifically, in occurrence models, commission error is when a species is predicted to be present when it is in fact absent (false positive), and omission error is when a species is predicted as absent when in is in fact present (false negative). Depending on the purpose of any one occurrence model, the relative “costs” associated with commission and omission errors can be determined (Zweig and Campbell 1993, Fielding 2002). For example, greater economic and ecological costs can be attributed to omission error when using an occurrence model to define areas in need of protection (Fielding 2002). Indeed, conservation of an endangered species will not prove successful if a certain area is not protected because an occurrence model erroneously predicted that the species was absent. Rates of commission and omission error change, as do sensitivity and specificity values, when the probability threshold changes. A probability threshold of 0.5 assumes that the costs of commission and omission
errors are equal, whereas a threshold of 0.7 assumes that omission errors are more serious. Increasing this threshold tends to decrease sensitivity and increase specificity.

An optimal threshold considers both occurrence rate as well as the relative costs of commission and omission for the situation at hand (Zweig and Campbell 1993, Fielding 2002). This can be done by calculating the slope of a line tangent to an ROC curve that incorporates occurrence rates and costs associated with the errors (Zweig and Campbell 1993). Because the ROC curve is comprised of paired sensitivity (true positive fraction) and 1-specificity (true negative fraction) values along a continuum of probability thresholds from 0 to 1, the point where this line intersects the ROC curve provides a pair of sensitivity and specificity values and the consequent optimal error rates associated with these values. The fact that the area under an ROC curve (AUC) is not affected by occurrence rate, coupled with the ROC curve’s usefulness in determining the optimal probability threshold for individual management scenarios, makes it an excellent measure to utilize when evaluating species occurrence models. Therefore, using multiple measures to evaluate model performance may sometimes be valuable, yet Pearce and Ferrier (2000b) state that discrimination ability is a sufficient measure when management relies solely on a rank of sites according to their likelihood of being occupied by a species. Logistic regression models provide this measure of likelihood. However, if a probability threshold is used with logistic regression models to prioritize competing land uses, then knowledge of model calibration, spread and bias in addition to model discrimination are necessary (Pearce and Ferrier 2000b).

Effect of model predictor variables

Although discrimination ability is a useful measure of model accuracy, it is significantly correlated with the number of variables (k) in a model. Including a large number of variables in a model can lead to overfitting such that a model’s temporal or spatial generality is limited. In this study, Akaike’s Information Criterion (AIC) was used in model selection (see methods). This criterion attempts to balance model fit (ability to explain observed variation in the dependent variable) and model complexity (number of independent variables). Therefore, it is unlikely that the relationship between AUC and k would persist for a higher number of variables. It is interesting to note that a positive relationship between k and model discrimination existed for models developed with data from 2001 and from the Toiyabe and Toquima ranges. These
two data sets exhibited species occurrence rates that were always greater in the evaluation data set (Figure 5). Thus, erring toward inclusion of a larger number of variables may be preferable for species with relatively high rates of occurrence.

Effects of spatial and temporal variation

The differences in accuracy between temporal and spatial models may also be attributed to species occurrence rates. There was no significant difference in discrimination ability between the two model types, whereas mean CCR and specificity were significantly higher in temporal models and sensitivity was significantly lower in temporal models. This may be explained by the distinct and consistent differences in species occurrence rates between 2001 and 2002; there was less variation in occurrence rates in space (among mountain ranges). Table 7 illustrates the observed occurrence rates for the 18 species modeled in each mountain range during both 2001 and 2002. The mean occurrence rate in 2002 for the 18 species modeled (0.27) was nearly twice the rate in 2001 (0.15). Given that there were two different observers, one in 2001 and another in 2002, there was potential for some observer bias. However, I am confident that there were real annual differences in species occurrences because presence/absence data were used in model development rather than abundance data, which is more likely be affected by observer differences. Furthermore, both observers received very similar training in bird identification in this region.

Temporal models consistently predicted absences well while predicting presences poorly. Spatial models, while also classifying absences better than presences, had a much wider range of values for these accuracy measures; some presences were classified well and some absences were classified poorly (Figure 4). In this system, variation in occurrence rates of breeding birds from year to year may be greater, and consequently may constrain accuracy of predicted presences, compared with the variation among mountain ranges. Other researchers have documented considerable temporal variation in avian occurrence in desert systems (Johnson 1995, Gutzwiller and Barrow 2001). In systems in which large annual fluctuations in occurrence are the norm, it is advantageous to collect multiple years of occurrence data in order to better document these fluctuations and prioritize locations for different land uses.
Management implications

If management plans are to be based on predictive models of occurrence, then follow-up censuses across the landscape are essential to determine whether management is indeed sufficient to protect species of concern, and to trigger changes in land use if current management is not achieving ecological objectives. Such an adaptive management framework ideally should be applied to all management plans to ensure that land use and available funding are being appropriately assigned. Ecosystems and the species that inhabit them are inherently dynamic and no model is universally applicable across space and time. In fact, Fielding and Haworth (1995) suggest that biologists be cautious when applying predictive models because we may be attempting to predict “systems that are inherently unpredictable.” Nonetheless, there is a need to understand the distributions of species in order to protect them and their habitats. Predictive models of occurrence, and more importantly their subsequent evaluation using independent data, provide substantial insight into the spatial and temporal variation in species occurrences. Such insight can only better equip ecologists and land managers to make responsible decisions.
I am thankful to my advisor, Dr. Robert Blair, for his encouragement and guidance in focusing my ideas into a manageable project as well as his general support throughout the last two years. I am also thankful to my other committee members, Drs. Tom Crist and Robert Schaefer, for keeping me honest with their statistical and procedural help. I would also like to thank Dr. Erica Fleishman for her continued feedback, expertise and encouragement at all stages of this project; John Fay for providing the GIS grids used to derive predictor variables; Avram Primack for general GIS assistance; Chris Betrus for his extremely valuable help getting ready for fieldwork in the Great Basin; Dr. Eric Porter for his time spent discussing conceptual development and management implications; Jason Bulluck for his editing assistance, long hours spent in the field measuring dbh, moral support and for sharing his passion for applied ecology and conservation with me; and to my family, Helen, James, Clint and Elizabeth, for their acceptance and support of the path that I have chosen in life. Financial assistance for this project was provided by Miami University; Stanford University; the Nevada Biodiversity Research and Conservation Initiative; and the Joint Fire Sciences Program via the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture.
LITERATURE CITED


TABLE 1. Two-by two classification table of the observed and predicted occurrence of a species. Each value represents a number of observations such that $A+B+C+D$ is equal to the total number of observations in the evaluation data set. The three measures of accuracy derived from this table and discussed in this chapter are: Correct classification rate (CCR), $(A+D)/(A+B+C+D)$; Sensitivity, $A/(A+C)$; and Specificity, $D/(B+D)$.

<table>
<thead>
<tr>
<th></th>
<th>Observed present</th>
<th>Observed absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted present</td>
<td>A (True Positive)</td>
<td>B (False Positive)</td>
</tr>
<tr>
<td>Predicted absent</td>
<td>C (False Negative)</td>
<td>D (True Negative)</td>
</tr>
</tbody>
</table>
TABLE 2. The ten data sets used to build and independently evaluate models of species occurrence.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Model-building data set</th>
<th>Number of sites</th>
<th>Evaluation data set</th>
<th>Number of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal</strong></td>
<td>2001</td>
<td>176</td>
<td>2002</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>176</td>
<td>2001</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>Toiyabe and Toquima Ranges</td>
<td>125</td>
<td>Shoshone Mountains</td>
<td>51</td>
</tr>
<tr>
<td><strong>Spatial</strong></td>
<td>Toiyabe Range and Shoshone Mountains</td>
<td>120</td>
<td>Toquima Range</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Toquima Range and Shoshone Mountains</td>
<td>107</td>
<td>Toiyabe Range</td>
<td>69</td>
</tr>
</tbody>
</table>
TABLE 3. Independent variables derived from grids within a GIS. Asterisk means that quadratic terms were also used in the model building as potential predictor variables.

<table>
<thead>
<tr>
<th>Grid</th>
<th>Value derived within a 90 meter buffer around each point count center</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Digital elevation model (DEM)</td>
<td>Mean elevation</td>
</tr>
<tr>
<td>*Slope</td>
<td>Mean slope</td>
</tr>
<tr>
<td>Northness</td>
<td>Mean northness on a scale from -100 (south facing) to 100 (north facing)</td>
</tr>
<tr>
<td>Eastness</td>
<td>Mean eastness on a scale from –100 (west facing) to 100 (east facing)</td>
</tr>
<tr>
<td>Distance to water</td>
<td>Mean distance to water (running or standing) in meters</td>
</tr>
<tr>
<td>Normalized difference vegetation index (NDVI)</td>
<td>Mean NDVI</td>
</tr>
<tr>
<td>Topographic exposure</td>
<td>Mean topographic exposure where negative numbers represent valleys and positive values represent ridge tops</td>
</tr>
<tr>
<td>*PRISM (Daly et al. 1994)</td>
<td>Mean annual precipitation</td>
</tr>
</tbody>
</table>
TABLE 4. Accuracy measures across all data sets. Rankings for “poor, good and excellent” are taken from Pearce and Ferrier (2000) (see text). Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima; TQ, Toquima; SH, Shoshone; and TY, Toiyabe.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Discrimination ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean CCR</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
</tr>
<tr>
<td>2001 2002</td>
<td>78.6</td>
</tr>
<tr>
<td>2002 2001</td>
<td>85.1</td>
</tr>
<tr>
<td>TYSH TQ</td>
<td>77.7</td>
</tr>
<tr>
<td><strong>Spatial</strong></td>
<td></td>
</tr>
<tr>
<td>TYTQ SH</td>
<td>70.6</td>
</tr>
<tr>
<td>TQSH TY</td>
<td>77.2</td>
</tr>
<tr>
<td><strong>All models</strong></td>
<td></td>
</tr>
<tr>
<td>NA NA</td>
<td>77.8</td>
</tr>
</tbody>
</table>
TABLE 5. Mean and standard deviation for each species’ occurrence rate and model accuracy. Values for each species are averaged across the five data sets. See text for description of categories for discrimination ability.

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean occurrence rate</th>
<th>Mean AUC</th>
<th>Mean CCR</th>
<th>Mean Sensitivity</th>
<th>Mean Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>American robin</td>
<td>0.376 (0.16)</td>
<td>0.687 (0.10)</td>
<td>63.9 (19.8)</td>
<td>0.09 (0.14)</td>
<td>0.90 (0.19)</td>
</tr>
<tr>
<td>Blue-grey gnatcatcher</td>
<td>0.154 (0.07)</td>
<td>0.793 (0.04)</td>
<td>85.2 (7.69)</td>
<td>0.06 (0.14)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td>Brewer’s sparrow</td>
<td>0.476 (0.11)</td>
<td>0.791 (0.10)</td>
<td>64.9 (12.6)</td>
<td>0.36 (0.21)</td>
<td>0.83 (0.24)</td>
</tr>
<tr>
<td>Broad-tailed hummingbird</td>
<td>0.241 (0.11)</td>
<td>0.762 (0.07)</td>
<td>77.4 (12.4)</td>
<td>0.06 (0.10)</td>
<td>0.97 (0.05)</td>
</tr>
<tr>
<td>Black-throated grey warbler</td>
<td>0.470 (0.20)</td>
<td>0.820 (0.06)</td>
<td>76.9 (3.98)</td>
<td>0.41 (0.40)</td>
<td>0.86 (0.14)</td>
</tr>
<tr>
<td>Bushtit</td>
<td>0.102 (0.05)</td>
<td>0.640 (0.10)</td>
<td>89.5 (7.10)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Cassin’s finch</td>
<td>0.194 (0.08)</td>
<td>0.680 (0.10)</td>
<td>79.6 (12.3)</td>
<td>0.04 (0.06)</td>
<td>0.98 (0.03)</td>
</tr>
<tr>
<td>Chipping sparrow</td>
<td>0.230 (0.11)</td>
<td>0.730 (0.08)</td>
<td>76.2 (14.7)</td>
<td>0.04 (0.06)</td>
<td>0.97 (0.06)</td>
</tr>
<tr>
<td>Dark-eyed junco</td>
<td>0.097 (0.05)</td>
<td>0.785 (0.08)</td>
<td>90.6 (4.70)</td>
<td>0.10 (0.10)</td>
<td>0.99 (0.02)</td>
</tr>
<tr>
<td>Green-tailed towhee</td>
<td>0.748 (0.15)</td>
<td>0.743 (0.09)</td>
<td>74.2 (13.2)</td>
<td>0.69 (0.39)</td>
<td>0.53 (0.37)</td>
</tr>
<tr>
<td>McGillivray’s warbler</td>
<td>0.426 (0.16)</td>
<td>0.803 (0.05)</td>
<td>71.6 (14.3)</td>
<td>0.30 (0.30)</td>
<td>0.87 (0.14)</td>
</tr>
<tr>
<td>Mountain bluebird</td>
<td>0.108 (0.06)</td>
<td>0.726 (0.10)</td>
<td>88.5 (8.40)</td>
<td>0 (0)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td>Mountain chickadee</td>
<td>0.349 (0.12)</td>
<td>0.724 (0.05)</td>
<td>67.3 (13.9)</td>
<td>0.17 (0.16)</td>
<td>0.90 (0.12)</td>
</tr>
<tr>
<td>Mourning dove</td>
<td>0.108 (0.04)</td>
<td>0.610 (0.07)</td>
<td>88.9 (7.09)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Red-shafted flicker</td>
<td>0.210 (0.08)</td>
<td>0.685 (0.04)</td>
<td>78.3 (13.1)</td>
<td>0.01 (0.01)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td>Sage thrasher</td>
<td>0.085 (0.04)</td>
<td>0.840 (0.10)</td>
<td>90.4 (6.04)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Spotted towhee</td>
<td>0.505 (0.13)</td>
<td>0.699 (0.15)</td>
<td>59.8 (14.5)</td>
<td>0.24 (0.31)</td>
<td>0.91 (0.10)</td>
</tr>
<tr>
<td>Western tanager</td>
<td>0.191 (0.07)</td>
<td>0.615 (0.05)</td>
<td>80.4 (12.2)</td>
<td>0 (0)</td>
<td>0.99 (0.01)</td>
</tr>
</tbody>
</table>
TABLE 6. Frequency of each predictor in the 89 models constructed across five data sets for 18 species of breeding birds.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Abbreviation</th>
<th>Number of total models</th>
<th>Proportion of total models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean eastness</td>
<td>EastX</td>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean slope squared</td>
<td>slope2</td>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean northness</td>
<td>North</td>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>Percent sage cover</td>
<td>Sagecov</td>
<td>5</td>
<td>0.06</td>
</tr>
<tr>
<td>Mean annual precipitation squared</td>
<td>prism2</td>
<td>5</td>
<td>0.06</td>
</tr>
<tr>
<td>Basal area of dead wood</td>
<td>DeadBA</td>
<td>7</td>
<td>0.08</td>
</tr>
<tr>
<td>Mean distance to water</td>
<td>H2Odist</td>
<td>7</td>
<td>0.08</td>
</tr>
<tr>
<td>Mean elevation squared</td>
<td>elev2</td>
<td>8</td>
<td>0.09</td>
</tr>
<tr>
<td>Mean topographic exposure</td>
<td>ExpoX</td>
<td>10</td>
<td>0.11</td>
</tr>
<tr>
<td>Percent shrub cover</td>
<td>Shrbcov</td>
<td>11</td>
<td>0.12</td>
</tr>
<tr>
<td>Percent ground cover</td>
<td>Grndcov</td>
<td>13</td>
<td>0.15</td>
</tr>
<tr>
<td>Percent mahogany cover</td>
<td>mahogcov</td>
<td>13</td>
<td>0.15</td>
</tr>
<tr>
<td>Width of water</td>
<td>H2Owidth</td>
<td>13</td>
<td>0.15</td>
</tr>
<tr>
<td>Mean slope</td>
<td>Slope</td>
<td>14</td>
<td>0.16</td>
</tr>
<tr>
<td>Mean NDVI</td>
<td>Ndvi</td>
<td>15</td>
<td>0.17</td>
</tr>
<tr>
<td>Mean annual precipitation</td>
<td>Prism</td>
<td>17</td>
<td>0.19</td>
</tr>
<tr>
<td>Percent canopy cover</td>
<td>Cancov</td>
<td>26</td>
<td>0.29</td>
</tr>
<tr>
<td>Mean elevation</td>
<td>Elev</td>
<td>28</td>
<td>0.31</td>
</tr>
</tbody>
</table>
### TABLE 7. Observed species occurrence rates in each individual mountain range during both the 2001 and 2002 breeding seasons.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>American robin</td>
<td>0.216</td>
<td>0.261</td>
<td>0.036</td>
<td>0.373</td>
<td>0.536</td>
<td>0.107</td>
</tr>
<tr>
<td>Blue-grey gnatcatcher</td>
<td>0.078</td>
<td>0.014</td>
<td>0.089</td>
<td>0.255</td>
<td>0.072</td>
<td>0.179</td>
</tr>
<tr>
<td>Brewer’s sparrow</td>
<td>0.176</td>
<td>0.304</td>
<td>0.482</td>
<td>0.373</td>
<td>0.420</td>
<td>0.464</td>
</tr>
<tr>
<td>Broad-tailed hummingbird</td>
<td>0.039</td>
<td>0.246</td>
<td>0</td>
<td>0.157</td>
<td>0.391</td>
<td>0.143</td>
</tr>
<tr>
<td>Black-throated grey warbler</td>
<td>0.333</td>
<td>0.058</td>
<td>0.286</td>
<td>0.725</td>
<td>0.145</td>
<td>0.75</td>
</tr>
<tr>
<td>Bushtit</td>
<td>0.020</td>
<td>0</td>
<td>0.054</td>
<td>0.216</td>
<td>0.072</td>
<td>0.143</td>
</tr>
<tr>
<td>Cassin’s finch</td>
<td>0.176</td>
<td>0.058</td>
<td>0.036</td>
<td>0.255</td>
<td>0.072</td>
<td>0.232</td>
</tr>
<tr>
<td>Chipping sparrow</td>
<td>0.098</td>
<td>0.029</td>
<td>0.071</td>
<td>0.392</td>
<td>0.116</td>
<td>0.304</td>
</tr>
<tr>
<td>Dark-eyed junco</td>
<td>0.039</td>
<td>0</td>
<td>0.018</td>
<td>0.157</td>
<td>0.087</td>
<td>0.089</td>
</tr>
<tr>
<td>Green-tailed towhee</td>
<td>0.568</td>
<td>0.348</td>
<td>0.589</td>
<td>0.824</td>
<td>0.681</td>
<td>0.732</td>
</tr>
<tr>
<td>McGillivray’s warbler</td>
<td>0.314</td>
<td>0.391</td>
<td>0.036</td>
<td>0.451</td>
<td>0.710</td>
<td>0.054</td>
</tr>
<tr>
<td>Mountain bluebird</td>
<td>0.039</td>
<td>0.014</td>
<td>0.036</td>
<td>0.137</td>
<td>0.029</td>
<td>0.196</td>
</tr>
<tr>
<td>Mountain chickadee</td>
<td>0.294</td>
<td>0.072</td>
<td>0.214</td>
<td>0.451</td>
<td>0.174</td>
<td>0.429</td>
</tr>
<tr>
<td>Mourning dove</td>
<td>0.118</td>
<td>0.029</td>
<td>0.071</td>
<td>0.137</td>
<td>0.058</td>
<td>0.036</td>
</tr>
<tr>
<td>Red-shafted flicker</td>
<td>0.235</td>
<td>0.116</td>
<td>0.018</td>
<td>0.216</td>
<td>0.203</td>
<td>0.071</td>
</tr>
<tr>
<td>Sage thrasher</td>
<td>0.039</td>
<td>0.014</td>
<td>0.054</td>
<td>0.059</td>
<td>0.087</td>
<td>0.107</td>
</tr>
<tr>
<td>Spotted towhee</td>
<td>0.333</td>
<td>0.290</td>
<td>0.232</td>
<td>0.059</td>
<td>0.478</td>
<td>0.482</td>
</tr>
<tr>
<td>Western tanager</td>
<td>0.137</td>
<td>0.130</td>
<td>0.054</td>
<td>0.549</td>
<td>0.174</td>
<td>0.018</td>
</tr>
</tbody>
</table>
FIGURE 1. Location of the Shoshone Mountains, Toquima and Toiyabe Ranges (black rectangle, see inset) in the Great Basin (irregular shape with black border, see inset). The inventoried canyons within the three mountain ranges are illustrated with black lines. Three pairs of canyons in the Toquima Range and one pair of canyons in the Toiyabe Range connect at the crest of these ranges.
FIGURE 2. Point count sites within the five canyons that incise the east and west slopes of the Shoshone Mountain Range. These points are shown over a Digital Elevation Model (DEM).
FIGURE 3. Vegetation sampling arrangement at each 75-meter point count. The three outer circles represent the 11.3-meter sampling circles. Each circle is 30 meters from the point count center (sun symbol), and 120 degrees from each other. The direction of the first circle from the center was chosen randomly (Martin et al. 1997). “Transects” used for ocular tube readings are defined as the 30-meter path from the point count center to the sampling circles.
FIGURE 4. Effects of observed species occurrence rate on model accuracy in the model-building and evaluation data sets.
FIGURE 5. Effects of observed species occurrence rate on sensitivity and specificity in the model-building and evaluation data sets.
FIGURE 6. Difference in observed species occurrence rates between all model-building and evaluation data sets: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.
FIGURE 7. Comparison of the distributions of accuracy measures for models assessing the effects of both spatial and temporal variation. Notice that the scale for sensitivity and specificity are different from the scale for AUC and CCR. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.
APPENDIX. Predictor variables that entered logistic regression models for each species in each data set and whether the relationship was positive (+) or negative (-). Abbreviations for variables are listed in Table 6. Species names from the 43rd supplement to the AOU check-list of North American Birds (Banks et al. 2002).

<table>
<thead>
<tr>
<th>Species</th>
<th>Species</th>
<th>2001</th>
<th>2002</th>
<th>TYTQ</th>
<th>TYSH</th>
<th>TQSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>American robin <em>Turdus migratorius</em></td>
<td></td>
<td>H2Owidth +</td>
<td>NDVI+</td>
<td>expo</td>
<td>north</td>
<td>north</td>
</tr>
<tr>
<td></td>
<td></td>
<td>expo -</td>
<td>expo -</td>
<td>slope</td>
<td>north</td>
<td>NDVI+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>deadBA -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue-grey gnatcatcher <em>Polioptila caerulea</em></td>
<td></td>
<td>elev -</td>
<td>elev -</td>
<td>slope</td>
<td>cancov</td>
<td>cancov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>slope -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>elev2 -</td>
<td>PRISM -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brewer’s sparrow <em>Spizella breweri</em></td>
<td></td>
<td>cancov -</td>
<td>cancov -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>elev +</td>
<td>elev2 +</td>
<td>cancov</td>
<td>expo</td>
<td>H2Owidth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H2Owidth -</td>
<td>slope -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRISM +</td>
<td>PRISM2 +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>north +</td>
<td>elev -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>expo -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad-tailed hummingbird <em>Selasphorus platycercus</em></td>
<td></td>
<td>H2Owidth +</td>
<td>shrbcov +</td>
<td></td>
<td></td>
<td>PRISM2 +</td>
</tr>
<tr>
<td></td>
<td></td>
<td>shrbcov +</td>
<td>PRISM +</td>
<td></td>
<td></td>
<td>shrbcov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRISM -</td>
<td>PRISM2 +</td>
<td></td>
<td></td>
<td>shrbcov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>north +</td>
<td>elev -</td>
<td></td>
<td></td>
<td>shrbcov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>expo -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H2Owidth +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRISM -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cancov -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black-throated grey warbler <em>Dendroica nigrescens</em></td>
<td></td>
<td>cancov +</td>
<td>groundcov -</td>
<td></td>
<td></td>
<td>sagecov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRISM -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H2Owidth +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bushtit <em>Psaltriparus inimus</em></td>
<td></td>
<td>cancov +</td>
<td>PRISM -</td>
<td>cancov</td>
<td>H2Owidth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>slope -</td>
<td></td>
<td></td>
<td>east +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>shrbcov +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassin’s finch <em>Carpodacus cassini</em></td>
<td></td>
<td>shrbcov +</td>
<td>cancov +</td>
<td></td>
<td></td>
<td>cancov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mahogcov -</td>
<td>PRISM -</td>
<td>cancov</td>
<td>shrbcov</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H2Odist -</td>
<td></td>
<td></td>
<td>mahogcov</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mahogcov -</td>
<td></td>
<td></td>
<td>shrbcov</td>
<td></td>
</tr>
<tr>
<td>Chipping sparrow <em>Spizella passerina</em></td>
<td></td>
<td>cancov +</td>
<td>cancov +</td>
<td>sagecov</td>
<td></td>
<td>cancov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>slope -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>east -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRISM -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>slope -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>shrbcov -</td>
<td>cancov +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>slope2 -</td>
<td></td>
<td></td>
<td>cancov</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>east -</td>
<td></td>
<td></td>
<td>cancov</td>
<td></td>
</tr>
</tbody>
</table>

43
<table>
<thead>
<tr>
<th>Species</th>
<th>2001</th>
<th>2002</th>
<th>TYTQ</th>
<th>TYSH</th>
<th>TQSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark-eyed junco <em>Junco hyemalis</em></td>
<td>grndcov -</td>
<td>mahogcov +</td>
<td>mahogcov +</td>
<td>mahogcov +</td>
<td>mahogcov +</td>
</tr>
<tr>
<td></td>
<td>mahogcov +</td>
<td>deadBA +</td>
<td>grndcov -</td>
<td>PRISM +</td>
<td></td>
</tr>
<tr>
<td>Green-tailed towhee <em>Pipilo chlorurus</em></td>
<td>PRISM2 -</td>
<td>elev +</td>
<td>elev +</td>
<td>deadBA -</td>
<td>deadBA -</td>
</tr>
<tr>
<td></td>
<td>elev +</td>
<td>cancov -</td>
<td>slope +</td>
<td>slope -</td>
<td>slope -</td>
</tr>
<tr>
<td></td>
<td>north -</td>
<td>slope2 -</td>
<td>cancov -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McGillivray’s warbler <em>Oporornis tolmiei</em></td>
<td>NDVI +</td>
<td>slope +</td>
<td>NDVI +</td>
<td>H2Odist -</td>
<td>shrbcov +</td>
</tr>
<tr>
<td></td>
<td>shrbcov +</td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td>shrbcov +</td>
<td>expo -</td>
</tr>
<tr>
<td></td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>H2Odist +</td>
</tr>
<tr>
<td></td>
<td>expos -</td>
<td>NDVI -</td>
<td>slope2 +</td>
<td>expo +</td>
<td>PRISM -</td>
</tr>
<tr>
<td></td>
<td>slope2 -</td>
<td>NDVI -</td>
<td>PRISM -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mountain bluebird <em>Sialia currucoides</em></td>
<td>expo +</td>
<td>NDVI -</td>
<td>NDVI -</td>
<td>NDVI -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NDVI -</td>
<td>slope +</td>
<td>slope2 +</td>
<td>expo +</td>
<td>PRISM -</td>
</tr>
<tr>
<td></td>
<td>slope2 -</td>
<td>NDVI -</td>
<td>PRISM -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mountain chickadee <em>Poecile gambeli</em></td>
<td>deadBA +</td>
<td>cancov +</td>
<td>cancov +</td>
<td>cancov +</td>
<td>cancov +</td>
</tr>
<tr>
<td></td>
<td>PRISM -</td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td>slope +</td>
<td>slope +</td>
</tr>
<tr>
<td></td>
<td>grndcov -</td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td>shrbcov +</td>
<td>shrbcov +</td>
</tr>
<tr>
<td></td>
<td>east -</td>
<td>cancov -</td>
<td>cancov -</td>
<td>NDVI +</td>
<td>H2Owidth +</td>
</tr>
<tr>
<td></td>
<td>H2Owidth -</td>
<td>cancov -</td>
<td>cancov -</td>
<td>H2Odist -</td>
<td></td>
</tr>
<tr>
<td>Mourning dove <em>Zenaida macroura</em></td>
<td>elev2 -</td>
<td>cancov -</td>
<td>elev -</td>
<td>PRISM -</td>
<td>mahogcov -</td>
</tr>
<tr>
<td></td>
<td>elev +</td>
<td>expo -</td>
<td>cancov -</td>
<td>expo -</td>
<td>PRISM -</td>
</tr>
<tr>
<td></td>
<td>east +</td>
<td>PRISM -</td>
<td>cancov -</td>
<td>PRISM -</td>
<td>H2Owidth -</td>
</tr>
<tr>
<td></td>
<td>H2Owidth -</td>
<td>cancov -</td>
<td>cancov -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red-shafted flicker <em>Colaptes auratus</em></td>
<td>deadBA +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
</tr>
<tr>
<td></td>
<td>PRISM -</td>
<td>H2Odist -</td>
<td>NDVI +</td>
<td>H2Owidth -</td>
<td>slope -</td>
</tr>
<tr>
<td></td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>H2Odist -</td>
<td>H2O width +</td>
</tr>
<tr>
<td></td>
<td>elev +</td>
<td>cancov -</td>
<td>elev2 -</td>
<td>elev +</td>
<td>elev +</td>
</tr>
<tr>
<td></td>
<td>grndcov +</td>
<td>slope -</td>
<td>mahogcov -</td>
<td>cancov +</td>
<td>cancov -</td>
</tr>
<tr>
<td></td>
<td>NDVI -</td>
<td>elev +</td>
<td>cancov -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td></td>
<td>elev +</td>
<td>cancov -</td>
<td>cancov -</td>
<td>slope2 -</td>
<td>slope2 -</td>
</tr>
<tr>
<td></td>
<td>north -</td>
<td>cancov -</td>
<td>cancell -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sage thrasher <em>Oreoscoptes montanus</em></td>
<td>elev +</td>
<td>elev2 -</td>
<td>elev +</td>
<td>elev +</td>
<td>elev +</td>
</tr>
<tr>
<td></td>
<td>cancov -</td>
<td>slope -</td>
<td>mahogcov -</td>
<td>cancov -</td>
<td>cancov -</td>
</tr>
<tr>
<td></td>
<td>north -</td>
<td>elev2 -</td>
<td>cancov -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td></td>
<td>north -</td>
<td>cancov -</td>
<td>cancov -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td></td>
<td>mahogcov -</td>
<td>north -</td>
<td>cancov -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td></td>
<td>mahogcov -</td>
<td>north -</td>
<td>cancov -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td>Spotted towhee <em>Pipilo maculatus</em></td>
<td>elev -</td>
<td>elev -</td>
<td>elev -</td>
<td>elev -</td>
<td>cancov +</td>
</tr>
<tr>
<td></td>
<td>west -</td>
<td>cancov -</td>
<td>cancov +</td>
<td>cancov +</td>
<td>H2Odist -</td>
</tr>
<tr>
<td></td>
<td>grndcov -</td>
<td>grndcov -</td>
<td>grndcov -</td>
<td>H2Odist -</td>
<td>H2O dist +</td>
</tr>
<tr>
<td></td>
<td>north -</td>
<td>north -</td>
<td>north -</td>
<td>north -</td>
<td>north -</td>
</tr>
<tr>
<td>Western tanager <em>Piranga ludoviciana</em></td>
<td>north -</td>
<td>PRISM -</td>
<td>mahogcov +</td>
<td>H2Odist -</td>
<td>PRISM -</td>
</tr>
<tr>
<td></td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>NDVI +</td>
<td>mahogcov +</td>
<td>PRISM -</td>
</tr>
<tr>
<td></td>
<td>expo -</td>
<td>expo -</td>
<td>expo -</td>
<td>mahogcov +</td>
<td>expo -</td>
</tr>
<tr>
<td></td>
<td>mahogcov -</td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td>mahogcov +</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mahogcov -</td>
<td>PRISM -</td>
<td>PRISM -</td>
<td>mahogcov +</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td>H2Odist -</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H2O width -</td>
<td>H2O width -</td>
<td>H2O width -</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3:  

DOES MODEL ACCURACY DIFFER ACROSS SPECIES ECOLOGICAL CHARACTERISTICS?

ABSTRACT

Models of species occurrence attempt to predict which habitats are most likely to be occupied based on habitat characteristics. These models have become increasingly popular in the literature and in their use by land managers due to the rapid loss of habitat and the need for effective and efficient management plans. Species occurrence models have potential to be directly applied to reserve design and prioritization of conservation efforts. I built avian occurrence models within adjacent mountain ranges in the Great Basin of central Nevada for 18 species across five data sets (two temporal and three spatial) for a total of 90 models. Models were evaluated using data independent of that used for model construction. I present four different measures of model accuracy; the area under an ROC curve (AUC), correct classification rate (CCR), sensitivity and specificity, and then compare these accuracy measures across five ecological characters; riparian dependence, migration habit, breeding habitat type, niche width and species commonness. The objective of this study was to determine whether accuracy varies across species ecological characters. If accuracy does vary in a predictable manner across similar groups of species, we may be better able to understand and generalize model accuracy. None of these ecological characters, except for levels of species commonness, showed differences in model accuracy. Specifically, in this data set, models developed for common species had a greater mean sensitivity than rare and medium species models while rare species models had a higher mean CCR and specificity compared with models developed for common and medium species. These results further support the findings of others suggesting that species occurrence rate greatly affects some measures of accuracy more than others.
INTRODUCTION

Models of species occurrence, or habitat suitability models, attempt to predict which habitats are most likely to be occupied based on habitat characteristics such as vegetation and topography. Accurate prediction of species occurrences is desirable if floral and faunal conservation is to be both effective and efficient. Occurrence models have become a practical conservation tool due to the expense of comprehensive biological surveys and the increased availability of broad-scale landscape predictors. Applications of predictive occurrence models are numerous when developed for virtually any taxonomic group. For example, one can model the distribution of a rare and threatened species to assess their level of protection and how that might be altered by changing land-use practices (Scott et al. 1993, Caicco et al. 1995). Furthermore, occurrence models can be used to predict shifting species distributions associated with climate change (Aspinall and Matthews 1994) as well as to predict the expansion of exotic species (Wadsworth et al. 2000).

Before model implementation, issues of model accuracy and generality should be addressed and empirically tested in a given system. There has been increased interest in the way that accuracy is measured and reported for species distribution models (Fielding and Bell 1997, Boone and Krohn 1999, Pearce and Ferrier 2000a, Pearce et al. 2001). A common practice is to use the same data set to both develop and evaluate occurrence models and then simply present a percentage of sites correctly classified as a measure of accuracy. Recently however, many researchers recommended testing models with independent data (Fielding and Bell 1997, Fielding 2002) and reporting several different accuracy measures (Manel et al. 2001). The recommendation for evaluating models with independent data arises from the fact that models can be biased toward the data sets used in model development such that their use in another time or place may not be very successful. Furthermore, the need for more than one accuracy measure is due to the fact that each different accuracy measure reports some distinct aspect of a model’s accuracy, such as discrimination ability, calibration, correct classification rate (CCR), sensitivity or specificity (Pearce and Ferrier 2000b, Manel et al. 2001). More than one accuracy measure is also beneficial because some common measures (e.g. CCR) can be greatly affected by species occurrence rate (Manel et al. 2001, Pearce and Ferrier 2000a, Fielding 2002).
One way of assessing the accuracy and applicability of species occurrence models is to group species according to various life-history categories in an attempt to understand why some species models have higher accuracy than others. Differences in model accuracy have been documented based on species occurrence rate (Pearce and Ferrier 2000b, Karl et al. 2002), niche breadth (i.e. habitat specialists versus generalists) (Hepinstall et al. 2002) and mobility (Pearce et al. 2001). Occurrence rate is defined as the proportion of sites where a species is recorded as present. Pearce et al. (2001) presented the results of occurrence models for several taxa in north-east New South Wales, and model accuracy was often shown to be greater for common species. However, this effect was not consistent across biological group; specifically, rare plants demonstrated greater accuracy than common plants. Niche breadth refers to the diversity of vegetative communities in which a given species can thrive and reproduce (Hepinstall et al. 2002). In other words, if a species always occupies one habitat, it is considered to be a specialist with a narrow niche breadth, and the opposite is true for generalist species that can occupy several different habitat types. Pearce and Ferrier (2000b), Mitchell et al. (2001), and Hepinstall et al. (2002) suggest that, in general, there is greater model accuracy for specialist species. Pearce et al. (2001) also found that species with less mobility, such as vascular plants and small reptiles, had greater model accuracy compared with vagile species such as diurnal birds and marsupials. In general, it appears that species with more homogenous distributions (i.e. less temporal and spatial variability) may be modeled more accurately.

Consideration of species characteristics besides species commonness and evenness may also help to interpret why some species models have greater accuracy than others. In the context of this study of avian occurrence models, I categorize species based on five life-history traits that may shed light on the discrepancies in model accuracy between certain avian species. Birds are ideal study organisms because their conspicuous nature makes them easy to census and much is known about their life history. Furthermore, birds have been proposed by many as potential indicator species for a variety of ecosystem features (Swengel and Swengel 1999, Ulicza and Angelstam 2000, Jansen and Robertson 2001, Tingley et al. 2002). Accurate occurrence models for avian species may therefore prove to be very useful management tools. Many researchers have demonstrated the success of avian occurrence models (Dettmers and Bart 1999, Manel et al. 2001, Pearce et al. 2001, Pearce et al. 2002); however, there is still much to understand about the differential success of these models relating to species ecological characters.
Riparian species constitute the majority of the non-game avifauna in the western region of North America despite the fact that riparian habitats occupy less than 1% of this region (Knopf and Samson 1994). Riparian habitats provide relatively high productivity in an otherwise xeric landscape. The vegetation structure in these habitats is dense and complex, thereby providing shelter from predators and breeding sites that are unavailable in the surrounding drier areas (Ryser 1985). The Great Basin region is comprised of numerous north-south oriented mountain ranges that act as habitat islands separated by expansive valleys that became warmer and drier following the Pleistocene (Brown 1978, Grayson 1993). The numerous canyons that drain the east and west slopes of the mountain ranges also may act as habitat islands for some species due to the riparian vegetation and resources they harbor relative to the intervening uplands (McDonald and Brown 1992). The habitat present in such riparian canyons, and the relatively under-studied piñon-juniper woodlands of the intervening uplands, supports a high diversity of birds along an elevational gradient (Medin et al. 2000). Neotropical migrant species make up 30 to 50% of the avian breeding community in piñon-juniper woodlands (Balda and Masters 1980) and up to 60% of the breeding landbird community in riparian areas (Johnson et al. 1977). Conservation of western riparian species, and Neotropical migrants in general, is a topic of concern since many species’ populations are being negatively affected by grazing (Saab et al. 1995) and by habitat destruction in both their wintering and breeding grounds (Sherry and Holmes 1995).

One objective of this study was to present a general framework for developing and evaluating species occurrence models that may be extended to any geographic region. In addition, I assessed the differential accuracy of avian models across various life-history characters such as migration habit, riparian dependence, breeding habitat type, commonness and habitat specificity (i.e. niche breadth). I aimed to determine whether these functional groupings of species provided a basis for understanding the mechanism by which model accuracy varies across species.

METHODS

Study area

Avian census data was collected in the Great Basin of central Nevada (Nye and Lander counties) in 16 canyons within three adjacent mountain ranges: the Toquima Range, the Toiyabe Range and the Shoshone Mountains (Figure 1). These ranges have similar biogeographic histories
and therefore generally contain similar biotic communities. However, differences in average elevation and precipitation among mountain ranges leads to differences in abundance of various vegetation communities. For example, the Toiyabe Range has a slightly higher mean elevation and receives much snowfall annually. Consequently, the canyons that drain its east and west slopes have more riparian vegetation (e.g. aspens (*Populus tremuloides*), willows (*Salix spp.*), birch (*Betula occidentalis*) and rose (*Rosa woodsii*)) than the other two ranges. On the other hand, the canyons in the Toquima Range are relatively dry and therefore have mostly piñon (*Pinus monophylla*), juniper (*Juniperus osteosperma*), sage (*Artemisia tridentata* ssp.), and rabbit brush (*Chrysothamnus* ssp.), with very little riparian vegetation. The Shoshone Mountains are wetter than the Toquima Range, but drier than the Toiyabe Range. Consequently, several of the canyons consist mostly of piñon/juniper woodlands and sagebrush with occasional riparian areas. These differences and similarities among mountain ranges constitute a landscape with realistic variation and thus a great opportunity to develop and separately evaluate predictive models of occurrence.

*Field methods*

We conducted avian censuses throughout five canyons in the Shoshone Mountains and Toiyabe Range, and six canyons in the Toquima Range during the breeding seasons of 2001 and 2002 using 75m fixed-radius point counts (Ralph et al. 1995, Dobkin and Rich 1998). Two individuals conducted the avian censuses, one in 2001 and another in 2002. Within each canyon, we placed point counts at least 200m apart in each of the dominant vegetation types. During each of the two years, we visited sites three times between sunrise (approximately 0530) and 3 hrs after sunrise (0830) during the breeding season (May 27-June 30) and recorded each bird that was seen or heard actively using terrestrial habitat within the point for 5 min. We sampled each site both very early as well as later in the morning. This enabled us to account for temporal variation in detection probability; most birds tend to sing less as the daily temperature increases and different birds tend to sing at different intensities throughout the morning.

Vegetation data was sampled by three individuals in three, 11.3m-radius circles (0.04 hectares) nested within each 75m-radius point count at each site (Figure 2) (Martin et al. 1997). We gathered vegetation density readings at each cardinal direction in each circle using a density board (Noon 1981). We also measured height and diameter at breast height (dbh) of all trees within each 11.3m circle. We collected 22 ocular tube readings from each point count site (Figure
2) to calculate percent canopy, shrub and ground covers (Noon 1981). We also recorded the presence and dbh of dead wood (logs and snags) and the presence, mean width and type of water (i.e. seep, flowing, and standing).

**Model development**

I used multiple logistic regression models to construct predictive models of species occurrence (presence/absence). Potential explanatory variables included environmental predictors measured with a GIS and fine-scale vegetation data collected at each census site. I derived predictor values measured with a GIS from several grids within a 90-meter buffer around each point count center (Table 1). The grid for mean annual precipitation was derived from PRISM data (Daly et al. 1994). Grids representing distance to water, elevation, slope, aspect as described by mean eastness and mean northness, and topographic exposure were derived using the spatial analyst extension in ArcView 3.2 (ESRI 1996). Finally, I used a measure of primary productivity, called the Normalized Difference Vegetation Index (NDVI), which was obtained from satellite imagery collected on June 2, 2000. All grids have a cell resolution of 30m except for the PRISM grid, which has 4km cell resolution. I considered quadratic terms for precipitation (PRISM), elevation and slope data during model construction (see below). This subset of predictor variables was selected based on the ecology of the species and the possibility of a non-linear relationship between occurrence and these explanatory terms.

I modeled species occurrences using five different data sets (three spatial and two temporal) and validated each on an independent data set (see Table 2). Prior to model construction for each species, I conducted correlation analyses and defined any combination of two independent variables as highly correlated if \( r > 0.40 \). I then conducted univariate analyses with each independent variable and the occurrence of the model species. I retained all variables with log-likelihood p-values < 0.25 as potential variables for the final, multivariate model (Hosmer and Lemeshow 1998). One of two highly correlated independent variables were retained if it had the higher p-value. If the two p-values were similar for the highly correlated variables, I ran a multivariate model with each separately.

I built multiple logistic models for 18 species whose sample sizes were large enough to build and validate statically stable models. These species had sample sizes ranging from 3 to 130 (mean 38.2) occurrences in the five build data sets. I used a stepwise selection procedure followed
by a best subset selection during model construction. To obtain a potential model for each significance level, I used stepwise selection (SAS v8.0) with very high alpha values (0.9) (Shtatland 2001). I then used the smallest Akaike’s Information Criterion (AIC) value to select the optimal model (Burnham and Anderson 1998). AIC is a model selection criterion that attempts to balance model complexity (overfitting) and bias (underfitting) using the principle of parsimony (Burnham and Anderson 2001).

Next, I conducted a best subset regression with all models containing one predictor more or less than the AIC-optimal model from the stepwise procedure (Shtatland 2001). For example, when the model with the smallest AIC value from the stepwise procedure had three predictor variables, then I ran a best subset regression with all models having two, three and four predictor variables. Again, I selected the model with the lowest AIC value and used this as the final predictive model for each species in each build data set. Model hierarchy was maintained such that models with a quadratic term were considered only if its corresponding linear term entered as well.

Model evaluation

Despite the pervasive use of logistic regression in constructing wildlife distribution models, many researchers provide no evaluation of model performance and those who do simply state a percentage or fraction of sites predicted correctly (Manel et al. 2001). Such a measure of accuracy is sometimes called a “correct classification rate” and is calculated from a two-by-two classification table (Table 3). This measure of accuracy can be misleading in that it is affected by the relative frequency (occurrence rate) of the species being modeled (Pearce and Ferrier 2000 and Manel et al. 2001). In addition, a species was predicted to be present or absent at a given site based on an arbitrary probability threshold. For example, when the probability threshold is set at 0.5, a species was considered to be present whenever the probability of occurrence is greater than 0.5. This is a problem because every threshold gives a different evaluation of model accuracy (Pearce et al. 2002). Despite these problems, it can be helpful to separately assess a model’s ability to predict presences and absences. I did this by calculating sensitivity (the proportion of true positives or A /(A+C) (from Table 3) and specificity (proportion of true negatives or D /(B+D) (from Table 3) from the values in a classification table at a particular threshold. I calculated these values as well as an overall correct classification rate (CCR) for a probability threshold of 0.7 and
compared the results to the area under a Receiver Operating Characteristic (ROC) curve. I chose a probability threshold of 0.7 because I preferred to have fewer overall predictions of higher quality (Fleishman et al. 2003)

The area under a ROC curve provides a measure of discrimination ability (i.e. how well a model can distinguish between occupied and unoccupied sites) (Pearce and Ferrier 2000). An ROC curve plots sensitivity versus the proportion of false negatives (1-specificity) (Pearce and Ferrier 2000) for several possible probability thresholds. Values greater than 0.9 for the area under the ROC curve (AUC) indicate excellent discrimination ability, values between 0.7 and 0.9 indicate a usable, reasonable level of discrimination while values between 0.5 and 0.7 indicate poor discrimination ability (Pearce and Ferrier 2000). ROC curves avoid the use of arbitrary thresholds and show less sensitivity to the frequency of species occurrence than other commonly used evaluation methods (Pearce and Ferrier 2000, Manel et al. 2001, Pearce et al. 2002).

Comparison of model performance

For species comparisons, I classified ecological characters such as riparian dependence, migration status, breeding habitat type, commonness and niche breadth to determine whether certain species models had greater accuracy than others. Dobkin and Wilcox (1986) define riparian dependence of Great Basin birds as either non-riparian, intermediately riparian (those species that can occupy both riparian and non-riparian habitats), or riparian. I used Peterjohn and Sauer’s (1993) classifications for a species’ migration status and breeding habitat. They describe migration status as Neotropical migrant, short-distance migrant, or resident species. Breeding habitat type was defined as woodland, sage scrub [succesional scrub (Peterjohn and Sauer 1993)] or urban. For commonness as an ecological character, a species was defined as common, moderate or rare based on where a species’ occurrence rate fell on the frequency distribution of all species’ occurrence rates. If it was within the highest quartile, then I considered the species common, in the lowest quartile, rare, and in the middle two quartiles, moderate. Finally, I define species niche breadth by the number of habitat types in which it was recorded. In general, there were three main habitat types in the region being sampled: sagebrush, piñon-juniper woodlands and riparian. I gave species a rating of one, two or three depending on the number of different habitat types in which it was recorded. For a list of species and their ecological characters, see Table 4.
I compared the mean accuracy of models for each ecological trait within each data set using a one-way ANOVA. If the F statistic was significant, suggesting that one of the pairs of means is statistically different, I then used a Bonferroni adjusted, all-pairs comparison to determine which pairs were significantly different.

**RESULTS**

*Overall model performance*

Overall, the models performed well in terms of their discrimination ability when validated on independent data from another time (year) or place (mountain range) (mean ± standard deviation AUC = 0.729 ± 0.101, n = 89): 50 models displayed good discrimination while three were considered excellent, 10 were considered marginal, and 26 considered poor (Table 5). Similarly, models performed well in terms of CCR with probability threshold = 0.7 (mean ± standard deviation CCR = 77.8 ± 14.4, n = 89): 22 had correct classification rates between 70 and 80%, 28 had rates between 80 and 90%, and 18 models classified greater than 90% of the sites correctly. Twenty-one models classified less than 70% of the sites correctly and 5 of these classified less than 50% of the sites correctly. Most models (about 92%) classified species absences (mean ± standard deviation of specificity = 0.926 ± 0.16, n = 89) better than species presences (mean ± standard deviation of sensitivity = 0.145 ± 0.25, n = 89) (Table 5).

I built 5 models for each species and determined mean accuracy for each species by averaging AUC, CCR, sensitivity and specificity across these five models (Table 6): 11 of the 18 species had good mean discrimination, and 7 had poor mean discrimination ability (AUC) (Table 6). None of the species had excellent discrimination when averaged across the five data sets. Overall, CCR was fairly high when averaged across the five models for each species (Table 6): 2 species had a mean CCR greater than 90% meaning that, on average, more than 90% of sites were correctly classified in each data set for those species; 5 species had a mean CCR between 80 and 90%, and 7 species had a mean CCR between 70 and 80%. Only 4 species had a mean CCR less than 70%.

As mentioned earlier, most models predicted absences better than presences, and this trend was apparent at the species level as well. In fact, 5 species had a mean sensitivity equal to zero (i.e. no presences predicted correctly), 6 species had a mean sensitivity between zero and 0.1, and 7 species had mean sensitivity greater than or equal to 0.1 (Table 6). On the other hand,
3 species had a mean specificity equal to 1 (i.e. all absences correctly predicted), 11 species had a mean specificity between 0.9 and 1, and 3 species had a mean specificity between 0.8 and 0.9. Contrary to this trend of poorly predicted presences relative to absences, the most prevalent species, the green-tailed towhee (GTTO), had a mean specificity of 0.53 and a mean sensitivity of 0.63 (Table 6).

Model performance across group classifications

For all data sets (i.e. 2001, 2002, TYTQ, TYSH, and TQSH) there was no difference in model accuracy based on a species’ migration habit, riparian dependence, niche width, or breeding habitat (Table 7). The Bonferroni adjustment resulted in a conservative test where \( p < 0.05/20 = 0.0025 \) indicated a significant difference between factor levels. Therefore, the only significant differences in model accuracy were seen for different levels of species commonness (Table 7). For this ecological character, CCR and sensitivity demonstrated significant differences. When models were developed using 2001 data, CCR and sensitivity showed significant differences for differing levels of species commonness (Table 7 and Figure 3). Likewise, CCR was significantly greater for rare species compared with common and medium species when models were developed in both 2001 and 2002 data sets. In contrast, common species had significantly higher sensitivity values compared with rare and medium species (Figure 3). This was true for models developed in 2001 and TYTQ data sets (Table 7).

Other differences that were marginally significant are noteworthy due to the conservative nature of the Bonferroni adjustment. Specificity was greater for rare and medium species compared with common species when models were built in 2001 and evaluated in 2002. Models for rare species in the TYTQ data set showed higher CCR than medium species. In this same data set, models for rare species had higher specificity than models for common species. In the TYSH data set, there were slight, but non-significant differences in specificity and sensitivity across levels of species commonness. There appears to be a slight difference in specificity values for species with varying levels of riparian dependence, but only in the TQSH data set (Table 7). Models developed for riparian species in the TQSH mountain ranges had lower specificity than intermediately riparian species. In terms of breeding habitat, discrimination ability seems to differ in the 2001 and TQSH data sets (Table 7). Models developed for sage scrub breeding species in the 2001 data set, had higher discrimination abilities than woodland breeding species, and models
developed for sage-scrub species in the TQSH data set had higher discrimination abilities than urban breeding species.

To understand what may be driving the differences in model accuracy among degrees of species commonness, I compared the differences in species occurrence rates between development and evaluation data sets. For a given species, this difference provides a measure of the spatial and temporal variation in its distribution. The average species occurrence rate in the development and evaluation data sets was 0.28 and 0.18, respectively. The average difference in occurrence rates across all species and models was 0.151 (± 0.12). Rare species tended to have smaller differences in occurrence rates (0.07 ± 0.04) between data sets than did medium (0.15 ± 0.09) and common species (0.21 ± 0.15). However, this difference was not significant except in the temporal data sets (i.e. those developed and cross-evaluated in 2001 and 2002). Specifically, models developed for rare species in both the 2001 and 2002 data sets had significantly lower differences in occurrence rate than did common species (F = 7.88, p = 0.005). It should be noted that, in general, more birds were recorded in 2002 compared with 2001 such that all differences in species occurrence rates were negative for species modeled in 2001 and evaluated in 2002 (Figure 4).

DISCUSSION

The framework presented here for developing and evaluating predictive models of avian occurrence provides a management tool for use in the Great Basin as well as other regions. The lack of a relationship between accuracy and the various groups of similar species illustrates the robust nature of predictive models with respect to differences in ecologies. These data suggest that species commonness is the only species characteristic to significantly affect CCR, sensitivity and specificity (Table 3). Others have also shown species occurrence rates to affect accuracy, particularly noting that the area under an ROC curve (AUC) is not affected by species occurrence rate and is therefore a superior measure of accuracy (Fielding and Bell 1997, Pearce and Ferrier 2000b, Manel et al. 2001, Pearce et al. 2002). In this study, AUC values demonstrated this same benefit. Based on the predominance of significant differences in accuracy in the 2001 data set, annual variation in species occurrence appears to affect accuracy more than spatial variation in occurrence (seen in models developed and evaluated across mountain ranges). However, the lack of significant differences in models developed with 2002 data suggests that it is not annual
variation that is the problem per se, but when species occurrence rate is much greater in the evaluation data set compared with the data set used in model development (Figure 4).

Both contrary and supportive of other findings, the results presented here suggest that either rare or common species can be modeled more precisely depending on which accuracy measure is used. For example, common species presences were more often classified correctly (i.e. high sensitivity values) compared with rare species. However, rare species had higher overall correct classification rate (CCR), and rare species’ absences were more often correctly classified (i.e. high specificity values) than that for common species. This trend was present in all data sets, but was most pronounced and significant when models were developed and evaluated in the 2001 and 2002 data sets, respectively. Even AUC values, despite the fact that they are not usually affected by species occurrence rate (Fielding and Bell 1997, Pearce and Ferrier 2000b, Manel et al. 2001, Fielding 2002), were marginally higher for common species whose occurrence models were developed in 2001. AUC and sensitivity were positively correlated (see Chapter 2), and as mentioned earlier, sensitivity was higher for common species as well. The implications of this for land management are significant since the level of annual variation and rates of species occurrences in this study are quite realistic. Rare species are most often the focus of management efforts, yet we cannot predict their presences with confidence. The ability to correctly predict species presences is often of greater utility than to correctly predict absences from a particular area, especially if that area is to be designated as protected. Therefore, sensitivity appears to be a more valuable measure in this context, and I conclude that the occurrence of common species can be modeled more accurately than rare species. Manel et al. (1999) also made this conclusion and suggested that the most constructive use of occurrence models may be to model common species that also act as biological indicators.

The lack of a relationship between accuracy and niche breadth or riparian dependence is somewhat surprising. Specialists species, whether they are described as having a narrow niche or as riparian obligates, would seem to be less difficult to model than those that have broad niches or are intermediately riparian. In fact, others have found that specialist species are modeled with higher accuracy than generalist species (Pearce and Ferrier 2000b, Mitchell et al. 2001, Hepinstall 2002). I would expect this trend to be consistent across studies, especially when several of the predictor variables used in these models actually described whether a site was comprised of riparian habitat. Species may be selecting habitat based on characteristics not used as predictors in
this study, or they may be selecting habitat at a different spatial scale than the predictors derived from the GIS grids and collected in the field. MacFaden and Capen (2002) demonstrated that avian habitat selection occurs at multiple scales suggesting that the incorporation of both coarse- and fine-grained predictor variables in occurrence models. In contrast, others have provided evidence that coarse-grained predictors alone perform as well as fine-grained predictors at developing accurate avian occurrence models (Mitchell et al. 2001, Suarez-Seoane et al. 2002).

The lack of association of accuracy with breeding habitat and migration habit is not as surprising. Whether a species is a Neotropical migrant, a short distance migrant, or a resident species does not directly provide an indicator of model accuracy. Mitchell et al. (2001), however, showed that Neotropical migrants and short distance migrants had higher model accuracy than models developed for resident species. Great Basin birds in this study showed this same trend, with models developed for Neotropical migrants displaying the highest accuracy, but this was not a significant trend. A confounding factor may be due to the fact that the Neotropical migrant species modeled in this study were classified as either common or medium in their occurrence, and as mentioned previously, these levels of occurrence were often modeled with higher accuracy than rare species.

Similarly, whether a species breeds in a woodland, sage scrub [succesional scrub (Peterjohn and Sauer 1993)], or urban habitat may not be a character that allows us to foresee how well they will be modeled. Models for successional scrub species did have higher discrimination ability compared to species occupying other breeding habitat types, but this difference was not significant. This may be due to the fact that this group of species on average had more predictors per model than the other species groups, and the number of predictors per model is positively correlated with discrimination ability (see Chapter 2).

Generalization of habitat models to understand likely responses associated with groups of similar species was suggested by Sisk et al. (2002) as a more useful goal than attempting to involve all the details of each unique species-habitat relationship in a model. These researchers eloquently remind us of the “pronounced trade-off between model utility and model complexity.” Undoubtedly, complex, species-specific models are necessary when management of a threatened and/or endangered species is the goal, yet management plans often include a suite of species that require similar types of habitat. If model accuracy for certain types of organisms is understood a priori, then models can be adjusted to account for such variations. For example, Boone and Krohn
(1999) introduce an a priori ranking of species associated with their likelihood of occurrence in future censuses. This ranking, called "Likelihood of Occurrence Rank" or LOORs, is based on both species and population attributes such as niche breadth, body mass, spatial and temporal stability, and changes within a species geographic range. Boone and Krohn (1999) found that species with high LOORs values were more often modeled correctly than species with low LOORs values. This finding is similar to the general conclusion presented here and by Pearce and Ferrier (2000b) that common species are often more accurately modeled (i.e. those species that would seem to have a high likelihood of occurrence or LOORs) than rare species.

**Management implications**

As researchers and land managers, we cannot expect models to be representations of truth (Starfield 1997); rather, we should attempt to develop models that allow for realistic spatial and temporal variation and models whose use extends beyond that for the single species. These data have not provided any evidence that model accuracy can be generalized, except when dealing with species commonness. Further research is needed to assess how accuracy varies across different ecological characters while incorporating different eco-regions and habitat types. Variation in model accuracy across species is extremely difficult to fully understand as it encompasses all levels of habitat selection from species interactions to physiological constraints and is intimately associated with variation in these factors across space, time and individuals. When considered in this light, generalization of model performance seems futile. Yet, to generalize based on empirical knowledge of species-habitat relationships that are indeed similar for like species may elicit further understanding while at the same time allowing for ease of application.
ACKNOWLEDGMENTS

I am thankful to my advisor, Dr. Robert Blair, for his encouragement and guidance in focusing my ideas into a manageable project as well as his general support throughout the last two years. I am also thankful to my other committee members, Drs. Tom Crist and Robert Schaefer, for keeping me honest with their statistical and procedural help. I would also like to thank Dr. Erica Fleishman for her continued feedback, expertise and encouragement at all stages of this project; John Fay for providing the GIS grids used to derive predictor variables; Avram Primack for general GIS assistance; Chris Betrus for his extremely valuable help getting ready for fieldwork in the Great Basin; Dr. Eric Porter for his time spent discussing conceptual development and management implications; Jason Bulluck for his editing assistance, long hours spent in the field measuring dbh, moral support and for sharing his passion for applied ecology and conservation with me; and to my family, Helen, James, Clint and Elizabeth, for their acceptance and support of the path that I have chosen in life. Financial assistance for this project was provided by Miami University; Stanford University; the Nevada Biodiversity Research and Conservation Initiative; and the Joint Fire Sciences Program via the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture.


**TABLE 1.** Independent variables derived from grids within a GIS. Asterisk means that quadratic terms were also used in the model building as potential predictor variables.

<table>
<thead>
<tr>
<th>Grid</th>
<th>Value derived within a 90 meter buffer around each point count center</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Digital elevation model (DEM)</td>
<td>Mean elevation</td>
</tr>
<tr>
<td>*Slope</td>
<td>Mean slope</td>
</tr>
<tr>
<td>Northness</td>
<td>Mean northnness on a scale from -100 (south facing) to 100 (north facing)</td>
</tr>
<tr>
<td>Eastness</td>
<td>Mean eastness on a scale from -100 (west facing) to 100 (east facing)</td>
</tr>
<tr>
<td>Distance to water</td>
<td>Mean distance to water (running or standing) in meters</td>
</tr>
<tr>
<td>Normalized difference vegetation index (NDVI)</td>
<td></td>
</tr>
<tr>
<td>Topographic exposure</td>
<td>Mean topographic exposure where negative numbers represent valleys and positive values represent ridge tops</td>
</tr>
<tr>
<td>*PRISM (Daly et al. 1994)</td>
<td>Mean annual precipitation</td>
</tr>
</tbody>
</table>
TABLE 2. The ten data sets used to build and independently evaluate models of species occurrence.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Model-building data set</th>
<th>Number of sites</th>
<th>Evaluation data set</th>
<th>Number of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001</td>
<td>176</td>
<td>2002</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>176</td>
<td>2001</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>Toiyabe and Toquima Ranges</td>
<td>125</td>
<td>Shoshone Mountains</td>
<td>51</td>
</tr>
<tr>
<td><strong>Spatial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Toiyabe Range and Shoshone Mountains</td>
<td>120</td>
<td>Toquima Range</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Toquima Range and Shoshone Mountains</td>
<td>107</td>
<td>Toiyabe Range</td>
<td>69</td>
</tr>
</tbody>
</table>
TABLE 3. Two-by two classification table of the observed and predicted occurrence of a species. Each value represents a number of observations such that \( A+B+C+D \) is equal to the total number of observations in the evaluation data set. The three measures of accuracy derived from this table and discussed in this chapter are: Correct classification rate (CCR), \( \frac{A+D}{A+B+C+D} \); Sensitivity, \( \frac{A}{A+C} \); and Specificity, \( \frac{D}{B+D} \).

<table>
<thead>
<tr>
<th></th>
<th>Observed present</th>
<th>Observed absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted present</td>
<td>A (True Positive)</td>
<td>B (False Positive)</td>
</tr>
<tr>
<td>Predicted absent</td>
<td>C (False Negative)</td>
<td>D (True Negative)</td>
</tr>
</tbody>
</table>
TABLE 4. Ecological characters and the observed mean occurrence rate for each species modeled. Occurrence rates were averaged across 2001 and 2002. (short distance migrant, SDM; Neotropical migrant bird, NTMB; Resident, RES; woodland breeding species, WL; scrub-successional species, SS; and urban breeding species, URB)

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean occurrence</th>
<th>Commonness</th>
<th>Riparian dependence</th>
<th>Migration status</th>
<th>Breeding habitat</th>
<th>Niche breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td>American robin</td>
<td>0.376</td>
<td>Medium</td>
<td>Riparian</td>
<td>SDM</td>
<td>Urban</td>
<td>1</td>
</tr>
<tr>
<td>Blue-grey gnatcatcher</td>
<td>0.154</td>
<td>Medium</td>
<td>Intermediate</td>
<td>NTMB</td>
<td>WL</td>
<td>3</td>
</tr>
<tr>
<td>Brewer’s sparrow</td>
<td>0.476</td>
<td>Common</td>
<td>Non</td>
<td>NTMB</td>
<td>SS</td>
<td>1</td>
</tr>
<tr>
<td>Broad-tailed hummingbird</td>
<td>0.241</td>
<td>Medium</td>
<td>Riparian</td>
<td>NTMB</td>
<td>WL</td>
<td>2</td>
</tr>
<tr>
<td>Black-throated grey warbler</td>
<td>0.470</td>
<td>Common</td>
<td>Non</td>
<td>NTMB</td>
<td>SS</td>
<td>1</td>
</tr>
<tr>
<td>Bushtit</td>
<td>0.102</td>
<td>Rare</td>
<td>Intermediate</td>
<td>RES</td>
<td>SS</td>
<td>3</td>
</tr>
<tr>
<td>Cassin’s finch</td>
<td>0.194</td>
<td>Medium</td>
<td>Intermediate</td>
<td>SDM</td>
<td>WL</td>
<td>1</td>
</tr>
<tr>
<td>Chipping sparrow</td>
<td>0.230</td>
<td>Medium</td>
<td>Non</td>
<td>NTMB</td>
<td>URB</td>
<td>2</td>
</tr>
<tr>
<td>Dark-eyed junco</td>
<td>0.097</td>
<td>Rare</td>
<td>Non</td>
<td>SDM</td>
<td>WL</td>
<td>2</td>
</tr>
<tr>
<td>Green-tailed towhee</td>
<td>0.748</td>
<td>Common</td>
<td>Non</td>
<td>NTMB</td>
<td>SS</td>
<td>2</td>
</tr>
<tr>
<td>McGillivray’s warbler</td>
<td>0.426</td>
<td>Common</td>
<td>Riparian</td>
<td>NTMB</td>
<td>SS</td>
<td>2</td>
</tr>
<tr>
<td>Mountain bluebird</td>
<td>0.108</td>
<td>Rare</td>
<td>Non</td>
<td>SDM</td>
<td>WL</td>
<td>3</td>
</tr>
<tr>
<td>Mountain chickadee</td>
<td>0.349</td>
<td>Medium</td>
<td>Intermediate</td>
<td>RES</td>
<td>WL</td>
<td>1</td>
</tr>
<tr>
<td>Mourning dove</td>
<td>0.108</td>
<td>Rare</td>
<td>Non</td>
<td>SDM</td>
<td>URB</td>
<td>3</td>
</tr>
<tr>
<td>Red-shafted flicker</td>
<td>0.210</td>
<td>Medium</td>
<td>Intermediate</td>
<td>SDM</td>
<td>SS</td>
<td>2</td>
</tr>
<tr>
<td>Sage thrasher</td>
<td>0.085</td>
<td>Rare</td>
<td>Non</td>
<td>SDM</td>
<td>SS</td>
<td>1</td>
</tr>
<tr>
<td>Spotted towhee</td>
<td>0.505</td>
<td>Common</td>
<td>Non</td>
<td>SDM</td>
<td>SS</td>
<td>2</td>
</tr>
<tr>
<td>Western tanager</td>
<td>0.191</td>
<td>Medium</td>
<td>Intermediate</td>
<td>NTMB</td>
<td>WL</td>
<td>1</td>
</tr>
</tbody>
</table>
TABLE 5. Accuracy measures across all data sets. Rankings for “poor, good and excellent” are taken from Pearce and Ferrier (2000) (see text). Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima; TQ, Toquima; SH, Shoshone; and TY, Toiyabe.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Model-Building data set</th>
<th>Evaluate data set</th>
<th>Mean CCR</th>
<th>Mean Sensitivity</th>
<th>Mean Specificity</th>
<th>Mean AUC</th>
<th>Poor (AUC&lt;0.7)</th>
<th>Good 0.7&lt;AUC&lt;0.9</th>
<th>Excellent AUC&gt;0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal</td>
<td>2001 2002</td>
<td>78.6 0.148</td>
<td>0.963</td>
<td>0.721</td>
<td>8 (44.4%)</td>
<td>10 (55.6%)</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002 2001</td>
<td>85.1 0.001</td>
<td>0.996</td>
<td>0.735</td>
<td>8 (44.4%)</td>
<td>10 (55.6%)</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TYSH TQ</td>
<td>77.7 0.158</td>
<td>0.902</td>
<td>0.735</td>
<td>7 (38.9%)</td>
<td>10 (55.6%)</td>
<td>1 (5.6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td>TYTQ SH</td>
<td>70.6 0.239</td>
<td>0.899</td>
<td>0.731</td>
<td>6 (33.3%)</td>
<td>10 (55.6%)</td>
<td>2 (11.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TQSH TY</td>
<td>77.2 0.183</td>
<td>0.867</td>
<td>0.718</td>
<td>7 (41.2%)</td>
<td>10 (58.8%)</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All models</td>
<td>NA NA</td>
<td>77.8 0.145</td>
<td>0.926</td>
<td>0.729</td>
<td>36 (40.4%)</td>
<td>50 (56.2%)</td>
<td>3 (3.3%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 6. Mean and standard deviation for each species’ occurrence rate and model accuracy. Values for each species are averaged across the five data sets. See text for description of categories for discrimination ability.

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean occurrence rate</th>
<th>Mean AUC</th>
<th>Mean CCR</th>
<th>Mean Sensitivity</th>
<th>Mean Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>American robin</td>
<td>0.376 (0.16)</td>
<td>0.687 (0.10)</td>
<td>63.9 (19.8)</td>
<td>0.09 (0.14)</td>
<td>0.90 (0.19)</td>
</tr>
<tr>
<td>Blue-grey gnatcatcher</td>
<td>0.154 (0.07)</td>
<td>0.793 (0.04)</td>
<td>85.2 (7.69)</td>
<td>0.06 (0.14)</td>
<td>0.90 (0.01)</td>
</tr>
<tr>
<td>Brewer’s sparrow</td>
<td>0.476 (0.11)</td>
<td>0.791 (0.10)</td>
<td>64.9 (12.6)</td>
<td>0.36 (0.21)</td>
<td>0.83 (0.24)</td>
</tr>
<tr>
<td>Broad-tailed hummingbird</td>
<td>0.241 (0.11)</td>
<td>0.762 (0.07)</td>
<td>77.4 (12.4)</td>
<td>0.06 (0.10)</td>
<td>0.97 (0.05)</td>
</tr>
<tr>
<td>Black-throated grey warbler</td>
<td>0.470 (0.20)</td>
<td>0.820 (0.06)</td>
<td>76.9 (3.98)</td>
<td>0.41 (0.40)</td>
<td>0.86 (0.14)</td>
</tr>
<tr>
<td>Bushtit</td>
<td>0.102 (0.05)</td>
<td>0.640 (0.10)</td>
<td>89.5 (7.10)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Cassin’s finch</td>
<td>0.194 (0.08)</td>
<td>0.680 (0.10)</td>
<td>79.6 (12.3)</td>
<td>0.04 (0.06)</td>
<td>0.98 (0.03)</td>
</tr>
<tr>
<td>Chipping sparrow</td>
<td>0.230 (0.11)</td>
<td>0.730 (0.08)</td>
<td>76.2 (14.7)</td>
<td>0.04 (0.06)</td>
<td>0.97 (0.06)</td>
</tr>
<tr>
<td>Dark-eyed junco</td>
<td>0.097 (0.05)</td>
<td>0.785 (0.08)</td>
<td>90.6 (4.70)</td>
<td>0.10 (0.10)</td>
<td>0.99 (0.02)</td>
</tr>
<tr>
<td>Green-tailed towhee</td>
<td>0.748 (0.15)</td>
<td>0.743 (0.09)</td>
<td>74.2 (13.2)</td>
<td>0.69 (0.39)</td>
<td>0.53 (0.37)</td>
</tr>
<tr>
<td>McGillivray’s warbler</td>
<td>0.426 (0.16)</td>
<td>0.803 (0.05)</td>
<td>71.6 (14.3)</td>
<td>0.30 (0.30)</td>
<td>0.87 (0.14)</td>
</tr>
<tr>
<td>Mountain bluebird</td>
<td>0.108 (0.06)</td>
<td>0.726 (0.10)</td>
<td>88.5 (8.40)</td>
<td>0 (0)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td>Mountain chickadee</td>
<td>0.349 (0.12)</td>
<td>0.724 (0.05)</td>
<td>67.3 (13.9)</td>
<td>0.17 (0.16)</td>
<td>0.90 (0.12)</td>
</tr>
<tr>
<td>Mourning dove</td>
<td>0.108 (0.04)</td>
<td>0.610 (0.07)</td>
<td>88.9 (7.09)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Red-shafted flicker</td>
<td>0.210 (0.08)</td>
<td>0.685 (0.04)</td>
<td>78.3 (13.1)</td>
<td>0.01 (0.01)</td>
<td>0.99 (0.01)</td>
</tr>
<tr>
<td>Sage thrasher</td>
<td>0.085 (0.04)</td>
<td>0.840 (0.10)</td>
<td>90.4 (6.04)</td>
<td>0 (0)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Spotted towhee</td>
<td>0.505 (0.13)</td>
<td>0.699 (0.15)</td>
<td>59.8 (14.5)</td>
<td>0.24 (0.31)</td>
<td>0.91 (0.10)</td>
</tr>
<tr>
<td>Western tanager</td>
<td>0.191 (0.07)</td>
<td>0.615 (0.05)</td>
<td>80.4 (12.2)</td>
<td>0 (0)</td>
<td>0.99 (0.01)</td>
</tr>
</tbody>
</table>
TABLE 7. Results for ANOVA test of whether some levels of species commonness were modeled more accurately than others. F and P values are displayed. Data sets are abbreviated as follows: TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. Values in bold were found to be significantly different using a Bonferroni adjusted p value where p < 0.0025 indicates a significant difference (see text).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>4.30</td>
<td>0.033</td>
<td>0.19</td>
<td>0.832</td>
<td>1.95</td>
<td>0.177</td>
<td>0.16</td>
<td>0.857</td>
<td>2.38</td>
<td>0.129</td>
</tr>
<tr>
<td>CCR</td>
<td>12.5</td>
<td><strong>0.001</strong></td>
<td>28.8</td>
<td><strong>0.0001</strong></td>
<td>4.63</td>
<td>0.027</td>
<td>2.36</td>
<td>0.129</td>
<td>3.24</td>
<td>0.070</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>26.0</td>
<td><strong>0.0001</strong></td>
<td>1.35</td>
<td>0.288</td>
<td><strong>21.6</strong></td>
<td><strong>0.0001</strong></td>
<td>4.40</td>
<td>0.031</td>
<td>6.56</td>
<td>0.010</td>
</tr>
<tr>
<td>Specificity</td>
<td>8.86</td>
<td>0.003</td>
<td>2.53</td>
<td>0.113</td>
<td>4.71</td>
<td>0.026</td>
<td>4.63</td>
<td>0.027</td>
<td>3.14</td>
<td>0.075</td>
</tr>
</tbody>
</table>
FIGURE 1. Location of the Shoshone Mountains, Toquima and Toiyabe Ranges (black rectangle, see inset) in the Great Basin (irregular shape with black border, see inset). The inventoried canyons within the three mountain ranges are illustrated with black lines. Three pairs of canyons in the Toquima Range and one pair of canyons in the Toiyabe Range connect at the crest of these ranges.
FIGURE 2. Vegetation sampling arrangement at each 75-meter point count. The three outer circles represent the 11.3-meter sampling circles. Each circle is 30 meters from the point count center (sun symbol), and 120 degrees from each other. The direction of the first circle from the center was chosen randomly (Martin et al. 1997). “Transects” used for ocular tube readings are defined as the 30-meter path from the point count center to the sampling circles.
FIGURE 3. Differences in accuracy across levels of species commonness in the 2001 data sets. The letters above each individual graph show which groups are significantly different (p < 0.0025). The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.
FIGURE 4. Difference in species occurrence rates between all model-building and evaluation data sets; TQSH, Toquima / Shoshone; TYSH, Toiyabe / Shoshone, TYTQ, Toiyabe / Toquima. The whiskers indicate the largest and smallest values except in the case where there are outliers. Outliers are defined as values that are more than 1.5 times the inter-quartile range.
CHAPTER 4:
CONCLUSION TO THESIS

This thesis had one main objective, and four minor objectives. The overall goal was to develop and evaluate models of avian species occurrence in three adjacent mountain ranges in central Nevada. In doing so, I aimed to determine whether (1) species occurrence rates, (2) spatial and temporal variation in species distributions, and (3) the number of predictor variables per model affected four different measures of accuracy. My fourth minor objective was to determine how model accuracy varied across groups of similar species based on life-history characters.

Several of the findings presented in this thesis further support those illustrated by the work of others. First, I demonstrated in both Chapters 2 and 3 that species occurrence rates affect measures of model accuracy derived from a two-by-two classification matrix, but not the area under a Receiver operating characteristic (ROC) curve (AUC). Furthermore, I discovered that for those measures affected by occurrence rate, there was a stronger relationship between occurrence rate in the evaluation data set and model performance than between occurrence rate in the model-development stage and the same measures of accuracy (i.e. more variation in a particular accuracy measure was explained by the occurrence rates in the evaluation data set than in the build data set. Second, the data presented here demonstrate that when there is a large difference in occurrence rate between the evaluation and model-building data sets as seen in the temporal models there is likely to be less variation in accuracy measures. Chapter 3 of this thesis also demonstrates how accuracy does not appear to differ across groups of similar species, except for species commonness, and then only in temporal data sets and for those accuracy measures calculated from a two-by-two classification table. In general, common species were modeled more accurately than rare species. In conclusion, species commonness is really just another measure for occurrence rate, thereby making this result not unique from those just stated.

Overall model accuracy was good, but when I consider the repercussions involved with incorrectly classified sites, I strongly suggest further strides be taken to increase accuracy. By increased accuracy, I do not think that we necessarily need to seek out new measures of accuracy because good measures already exist. Rather we should continually look for ways in which to develop models that decrease our overall error by incorporating variability that is inherent in all
biological systems. For example, the a priori inclusion of model terms that address such issues as the likelihood of species occurrence (Boone and Krohn 1999), spatial autocorrelation (Klute et al. 2002), and seasonal or annual fluctuations in distributions need more attention.

Based on the general results from Chapter 2 and more specifically from Chapter 3, I feel that Boone and Krohn’s (1999) suggestion for an a priori consideration of the likelihood of occurrence in a particular species model is merited. There is overwhelming evidence that common species with somewhat specific habitat characteristics are modeled more accurately than those that are rare and/or have more general habitat requirements (Pearce and Ferrier 2000, Pearce et al. 2001, Hepinstall et al. 2002). These are characteristics that any researcher can find for most study organism at the onset of model development. Therefore, this is a relatively easy step to take, especially when one considers the increased accuracy associated with such a measure.

Autocorrelation is essentially a lack of independence between adjacent sampling sites that results in increased prediction error (Legendre 1993). The problems associated with spatial autocorrelation are well known and documented (Legendre 1993, Fielding and Bell 1997, Hiebler 2000, Fielding 2002), though most species occurrence models do not consider them explicitly. It is easier to assume spatial independence of one’s data because most statistical procedures rely on this assumption. However, when autocorrelation is considered, models should better describe variation in species occurrence and consequently have greater accuracy. Klute et al. (2002) showed that models of American Woodcock habitat use that consider spatial autocorrelation are more accurate than those that do not. Due to the sampling design of this study, with census routes following canyons, I feel that there may be some spatial dependence between species occurrence in nearby sites. Regardless, this is an issue that should be addressed explicitly in any system.

Similar to the incorporation of spatial terms into an occurrence model, there is also a need to address issues of temporal variation in species distributions. As seen in the previous two chapters, when the occurrence rates of species differs dramatically between model development and evaluation data sets, some accuracy measures are greatly affected. Use of an accuracy measure that is not affected by occurrence rate is one thing that can prevent this, but collection of multiple years of data and the incorporation of a term for lag effects associated with the previous years weather may also help to decrease error. Multiple years of sampling can be expensive,
especially for taxonomic groups that are not as conspicuous as birds, but there is no better way to gain an understand of how a particular group of organisms reacts to seasonal and annual fluxes in climate and food. Once there is an understanding of this, models can be developed that incorporate such fluxes thereby increasing accuracy.

Each of the above suggestions involves the costs of learning new, sometimes sophisticated, statistical procedures and a larger amount of initial data collection with a benefit of increased accuracy and/or decreased error. Whether the benefits outweigh the costs, especially in the land-management arena, is yet to be seen. The paradox of utility versus complexity is apparent in this situation and should not be ignored. New technologies and increased data can obviously further our understanding of the dynamic nature of biological systems and increase our confidence when making land-use decisions. However, proper use of these new technologies cannot be expected by those with minimal to no training, and sufficient funding and time are rarely allotted to those desiring several years of census data before land-use decisions are made. I feel confident that the advancements in new statistical procedures have the ability to increase model performance and I urge researchers involved with these types of studies to communicate their findings and skills to those who may need them in order to make informed land-use decisions.
LITERATURE CITED


