The Effectiveness of Predicting Breeding Bird Distributions Using Probabilistic Models

Karen H. Beard, *Utah State University*
N. Hengarten
D. K. Skelly

Available at: https://works.bepress.com/karenh_beard/34/
The Effectiveness of Predicting Breeding Bird Distributions

Using Probabilistic Models

KAREN H. BEARD ¹

NICOLAS HENGARTNER ²

DAVID K. SKELLY ³

¹ School of Forestry and Environmental Studies, Yale University, 205 Prospect Street, New Haven, Connecticut 06511-2104, U.S.A.

² Department of Statistics, Yale University, P.O. Box 208290, New Haven, Connecticut 06520-8290, U.S.A.

³ School of Forestry and Environmental Studies, Department of Ecology and Evolutionary Biology, Yale University, 370 Prospect Street, New Haven, Connecticut 06511-2104, U.S.A.

Running Head = Predictive Modeling of Bird Distributions

4100 Words

¹ Person to whom correspondence should be sent
phone 203 432-5692; fax 203 432-5556
e-mail address = karen.beard@yale.edu
Abstract: Conservation biologists need to be able to predict species distributions based on easily collected data available at regional scales. We quantified the effectiveness of different types of data for predicting bird distributions for the state of Idaho. We developed probabilistic models to evaluate the ability of vegetation, climate, spatial autocorrelation data, combinations of these variables and a null model to predict species presence/absence for 40 bird species. We determined the probability of correctly predicted presences and absences for each species using a training-testing sample methodology. This method involves splitting the data set into two portions: one portion was used to “fit” the models using maximum likelihood and the second portion was then predicted from the fitted models. The predicted number of observed birds from the second portion of the data was then compared to actual presence/absence values to assess model performance. Overall, differences in average performance among the parameterized models were small. Vegetation, climate and spatial models each predicted approximately 60% of the presences correctly. Models employing a combination of these factors consistently improved model performance, but only slightly (ca. 4% improvement). By contrast, the null model correctly predicted just 35% of the presences. Our results suggest that (1) parameterized models are a substantial improvement over a null model but still make frequent mistakes in predicting species presence/absence, and (2) data availability may be the most important factor in determining which variables to use to predict species presence/absence. In some cases, available and relatively inexpensive climate data or incomplete distributional information may be the preferred data option.
Introduction

The study of distributional patterns of species at large spatial scales has long been of interest to ecologists (Andrewartha & Birch 1954; Krebs 1972; Brown 1995). From a conservation perspective, knowledge of species distributions at regional scales is crucial to effectively manage biodiversity and protect endangered species (Noss 1983). The reliability and accuracy of this knowledge, however, is contingent on the quality of the surveys generating the data. Such data are difficult to obtain (Burbidge 1991), and distributional data for most species at this scale are still lacking.

To minimize the problems with inadequate distribution data, there have been repeated calls in the conservation literature to increase monitoring efforts (e.g., Ehrlich 1996; Blaustein et al. 1994), and a number of government programs have been designed to meet this need (e.g., USDA Forest Service Forest Inventory and Analysis Program, USGS Biological Resources Division, Natural Heritage Program). Unfortunately, even with greatly increased monitoring efforts, sampling will remain unevenly distributed across space, time, and taxa.

Therefore, an inescapable feature of large-scale distributional data is their patchiness and incompleteness. Nonetheless, conservation biologists must increasingly rely on these data to help guide planning efforts and, in particular, to build predictions about where they expect to find targeted species or high levels of biodiversity at statewide scales. We contrasted different conceptual hypotheses for making the best of inadequate distributional data using geospatial information that is often available at this scale of study, specifically vegetation, climate, and local species’ survey data. We chose to analyze species presence/absence because this type of distributional data is the most practical and widely available (Hulbert 1971; Whittaker 1972; Peet 1974; Steinberg & Kareiva 1997).
The hypotheses we compared were: (1) species presence/absence can best be predicted by establishing statistical relationships between vegetation types and existing distribution data, (2) species presence/absence can best be predicted by establishing statistical relationships between climatic variables and existing distribution data, (3) species presence/absence can best be predicted by the spatial pattern of presence and absence in nearby locations (i.e., the statistical relationship between the effect of neighboring observations on an observation may be used to fill in gaps in the data), and (4) species presence/absence can best be predicted using some combination of these factors.

Each of these approaches has been used in the past (e.g., Hogmonder & Moller 1996; Augustin et al. 1996; Buckland & Elston 1993; Buckland et al. 1996); however, the most popular conceptual model seems to be that vegetation best predicts vertebrate presence/absence (Austin 1991; Caicco et al. 1995). For instance, in the United States, the Gap Analysis Program (GAP) has received a great amount of attention. GAP is based on the assumption that species are habitat specialists with respect to vegetation types (Scott et al. 1993). Although it is easy to provide evidence that many species are distributed with respect to vegetation (or climate or space), it is not as easy to quantify the relative effectiveness of different predictive variables at this scale. We developed a method for quantifying the effectiveness of these different factors at a coarse spatial resolution over large areas and identified which factors would likely provide us with the most effective method for predicting species presence/absence using relatively available data.

To quantify and compare the effectiveness of these different variables, we used data from the North American Breeding Bird Survey (BBS) for the state of Idaho. The data were used to construct a map of species presence/absence at 290 locations for 40 species. For each of these locations we quantified three additional variables: vegetation type, climate, and the spatial pattern of neighboring presence/absence of bird species. Vegetation type was described using
the Idaho GAP digital vegetation map (Caicco et al. 1995). Climate was characterized using the National Climatic Data Center (NCDC) weather station data averaged over 30-year (1961-1990). The spatial pattern of species presence/absence using the BBS data was described mathematically. We then developed probabilistic models for each of the variables and combinations of variables as well as a null model. Using the model outputs, we compared the effectiveness of the conceptual approaches described above for predicting species distributions.

Methods

Data Collection

Every spring volunteers participating in BBS collect data from over 2,000, 39.4-km (25-mile) routes across North America (Robbins et al. 1986). We used data collected from BBS routes in the state of Idaho during the eight seasons from 1988 to 1995. We focused our study on data collected during this time interval because (1) the number of routes dramatically increased in the state of Idaho in 1988, (2) a relatively consistent number of routes were surveyed during this period, and (3) it corresponded well with the time interval over which the vegetation data were most relevant. Volunteers collected data from 58 routes over this time interval.

Volunteers collect BBS route data during 50, 3-minute stops located 0.8 km (0.5 miles) apart. The observer records all birds seen or heard within a distance of 200 m. For the purposes of this study, we compressed breeding bird stop data and used data describing five equal, 8-km (5-mile) segments for each route; each segment contained data from 10, 3-minute stops. The segment locations were defined spatially by the latitude and longitude of their centers.

Counts of birds observed within segments were converted to species presence or absence using the following algorithm: a species qualified as present if it was observed within a segment for 2 or more years of the 8. This resulted in 290 segments describing presence/absence for each
species. The resulting map of species presence/absence was used to parameterize the statistical models for each species (Fig. 1).

Species were excluded from the study if they were present at less than 10% or more than 90% of the sites. This constraint reduced the pool of species observed in Idaho over this time interval from 238 to 89. From this subset, we randomly selected 40 species for the analysis (Table 1). The most widely distributed species included in this study, the Yellow Warbler (Dendroica petechia), was present at 56% of the sites.

Vegetation types were collected from the original Idaho Gap Analysis Program vegetation map (Caicco et al. 1995). The original map contained 119 vegetation types which, for the purposes of this study, were consolidated into 33 vegetation types using a standard algorithm (Scott et al. 1993). We entered each of the 8-km route segments into ARC/INFO. The vegetation types along each segment and their relative percent coverage of the segment were recorded. Vegetation type was described by the category that represented the majority of the length of each segment. This resulted in the use of 22 vegetation types to describe the primary vegetation for the 290 segments.

We described climatic conditions using three variables: (1) average annual precipitation, (2) average January temperature and (3) average July temperature. We assumed that the 30-year, averaged NCDC weather station data (Ownby & Ezell 1992) from the weather station nearest to the center of each route adequately described the climate along the route. Therefore, we assumed that climatic conditions were constant along each route, and all five segments of each route were assigned the same climate variables. There are 105 weather stations in Idaho and 45 were used to describe the routes.
Model descriptions

Vegetation Model

The vegetation model assumed that the presence/absence of a species from site to site was independent and the probability $p$ of observing a presence at each site only depended on the major vegetation type describing the segment. We let $p_v$ be the probability of a presence when vegetation type $v$ was dominant and let $v(s)$ be the major vegetation type at site $s$. The joint distribution then was

$$P[Z(1) = z_1, \ldots, Z(290) = z_{290}] = \prod_{x=1}^{290} \left( p_{v(s)} (1 - p) \right)^{z_s} = \prod_{v=1}^{22} \left( p_v^{N(v)} (1 - p_v) \right)^{T(v) - N(v)},$$

where $T(v)$ was the total number of sites with vegetation type $v$ and $N(v)$ the number of sites with vegetation type $v$ where the species was present. The maximum likelihood estimator was, therefore, the relative frequency of presences in each vegetation type (i.e., $\hat{p} = N(v) / T(v)$).

Not all vegetation types were present within the training sample for many species. If a vegetation type was excluded from the training sample, the probability of observing the species within the testing sample was estimated as the “global” probability of observing the species in the state of Idaho. The probability assigned to such a site would be similar to the probability assigned to the site by the null model (described below). Three extremely rare vegetation types had only one occurrence in the 290 segments and were excluded from the training sample in approximately 20% of the model simulations. Five vegetation types were relatively rare and excluded from the training sample in approximately 1-4% of the simulations. The other 14 vegetation types were excluded from the training sample infrequently (in less than 1% of the model runs).
Climate Model

For the climate model, we assumed that the probability of a presence was independent across the segments and that it depended only on the climate of the segment. All five route segments were characterized with the climate data recorded by the weather station nearest to the center of the route: average annual precipitation (AP), average January temperature (JaT) and average July temperature (JuT). Specifically, the logarithm of the odds ratio $\ln \frac{p_s}{1- p_s}$, where $p_s$ is the probability at site $s$, was modeled by an additive function of the three climate variables:

$$\ln \frac{p_s}{1- p_s} = \mu + f_1(\text{AP}_s) + f_2(\text{JaT}_s) + f_3(\text{JuT}_s).$$

To model the effect of climate variables, an additional assumption was used, namely, that the probability of a presence varied as a smooth function of the climate variables. Each of the functions $f_1$, $f_2$, and $f_3$ were modeled by natural cubic splines with seven degrees of freedom as opposed to a linear fit for added flexibility. We chose 22 degrees of freedom for this model to enable equitable comparisons between the climate model and vegetation model which also had this number of degrees of freedom. The climate model was then fit using standard logistic regression techniques. For about half the species considered, however, the model was over-parameterized, resulting in failure of the fitting algorithm to converge. To remedy this problem, the degrees of freedom were reduced until convergence was reached. The resultant number of degrees of freedom ranged from three to seven.

Spatial Model
The spatial model was designed to capture the contextual pattern of species distributions. Unlike the other models, for this model we assumed that the presences and absences of birds at neighboring segments influenced the probability \( p(s) \) of a presence at a segment \( s \). The model specified the conditional distributions of the random variables as

\[
Z(s) = \begin{cases} 
1 & \text{the species is present at site } s \\
0 & \text{otherwise.}
\end{cases}
\]

Given the values of \( Z(j) = z_j, j \neq s \) because

\[
P[Z(s) = z_s \mid Z(j) = z_j, j \neq s] = \frac{\exp[(\mu + \sum_{j \neq s} \omega(s, j) z_j) z_s]}{1 + \exp[(\mu + \sum_{j \neq s} \omega(s, j) z_j) z_s]}.
\]

Therefore, we modeled the conditional probability of a presence at segment \( s \) where \( \mu \) measures the global intensity of presences as a baseline and \( \omega(s, j) \) measures the effect of a presence at site \( j \) on the probability of a presence at site \( s \). A positive \( \omega(s, j) \) implies a positive association between presences, whereas a negative \( \omega(s, j) \) implies a negative association.

We assumed that \( \omega(s, j) \) was strictly a smooth function of the Euclidean distance \( d = d(s, j) \) between the centers of segment \( j \) and segment \( s \). We chose to model this relationship using a spline function with seven degrees of freedom because it optimized model fit. Preliminary data exploration strongly suggested that spatial dependence was localized well within a 1° latitude-longitude (or approximately 80-km) radius. Therefore, \( f(d(s, j)) \) was set to zero when the distance between the segments was >1°. Clearly, the possible dependence between sites was limited to the state of Idaho.

Unlike the vegetation and climate models, the likelihood of observing the species at a given location using the spatial model was practically impossible to evaluate numerically due to
the difficulties in calculating the normalization constant. Therefore, maximum likelihood estimation was not an option. Instead, the maximization of the pseudo-loglikelihood with respect to the unknown parameters was used. Cressie (1993) and Besag (1975) define the pseudo-loglikelihood to be the sum of the logarithm of the conditional probabilities. To create models that combined vegetation, climate, and spatial autocorrelation data, we simply added the variables in the models.

Null Model

We created a null model to assess the relative performances of the other models. The model assumed the probability of a presence was independent across sites and the probability of a presence for each segment only depended on the global probability $p$ of the species presence in the state of Idaho. The maximum likelihood estimate for the relative frequency $p$ was the observed relative frequency: $\hat{p} = N/T$, where $N$ was the total number of sites where the species was present in the state of Idaho and where $T$ was the total number of possible sites in the state of Idaho or the 290 route segments. Therefore, the null model contained information about the commonness of the species. Because the null model assumed the common probability of a species presence did not depend on environmental variables, the performance difference between the null model and the parameterized models was used to quantify the predictive power of the environmental factors.

Modeling Approach

We tested the predictive ability of the models using a training-testing methodology commonly used in statistical classification (Devroye et al. 1996). To use
this methodology, sites were randomly divided into two samples, a training sample and a testing sample. The training sample contained 232 sites (80% of the data) and the testing sample contained the remaining 58 sites (20% of the data). The presence/absence data of the training sample were used to fit the models. The fitted models were then used to predict the probability of a presence, \( p \), at each of the sites in the testing sample.

To convert the probabilistic model responses for each of the testing locations to species presence/absence a weighted coin with probability \( p \) of observing the species in each given location was tossed. If the bird was actually present in a site, the model would either (1) correctly predict the bird was present in the site or (2) commit an omission error (incorrectly predict the bird was absent from the site). If the bird was actually absent from the site, the model would either (1) correctly predict the bird was absent from the site or (2) commit a commission error (incorrectly predict the bird was present in the site). On average, the expected results from the weighted coin toss would be as follows: if bird is present, 100 x \( p \)% are correct and 100 x (1-\( p \))% are omission errors, and if the bird is absent, 100 x \( p \)% are commission errors and 100 x (1-\( p \))% are correct. Therefore, we determined the expected number of correctly assigned presences, correctly assigned absences, number of commission errors, and number of omission errors in the testing sample based on the probability of a presence assigned to each site.

The entire process, fitting the models with the training sample and evaluating the models with the testing sample, was repeated 100 times for each model. For each of the 100 runs of the models a different randomly selected subset of sites was designated as the training and testing samples. For the eight models the same randomly selected sites were employed for each of the 100 runs. This process was repeated for each of the 40 bird species. We analyzed both the number of sites predicted correctly where the species was actually present and the overall
number of correct predictions. Because most species are rare, the number of correctly predicted
presences is important because it pinpoints areas of concern rather than vast areas where the
species is absent.

Results

On average the parameterized models correctly assigned approximately 60% of
the presences (Fig. 2). The number of correctly predicted presences varied among the
models ($F_{7,31992}=222.87, p<0.0001$), but differences among parameterized models were
quantitatively small. Multiple-factor models outperformed all other models by a small
margin (Tukey test, $p<0.05$), but were not different from one another. The spatial model
outperformed the other single factor models by a small margin (Tukey test, $p<0.05$). By
contrast, all parameterized models substantially outperformed the null model which
correctly assigned presences just 34.5% of the time (Tukey test, $p<0.05$).

In the process of predicting presences correctly, the models also incorrectly assigned
presences to approximately as many sites as they incorrectly assigned absences. In other words,
the rates of omission and commission errors were roughly equal (Fig. 3). Unlike any other
model, the null model committed more commission and omission errors than it assigned
correctly predicted presences.

We also analyzed the ability of the models to predict presences and absences collectively.
On average, the parameterized models predicted 77.5% of the sites correctly. The number of
correct predictions also varied among the models ($F_{7,31992}=1379.88, p<0.0001$). The
vegetation/spatial and vegetation/climate/spatial models slightly outperformed the other models
(Tukey test, $p<0.05$). The climate/spatial model outperformed the vegetation/climate model, and
all combination models outperformed the other models (Tukey test, $p<0.05$). The spatial model
outperformed the other two single factor models (Tukey test, \( p<0.05 \)). The climate and vegetation models did not differ (Tukey test, \( p>0.05 \)). All differences among the parameterized models were small. However, all models substantially outperformed the null model which predicted 62.3% of the sites correctly (Tukey test, \( p<0.05 \)).

Finally, model performance was strongly correlated with both species abundance (Fig. 4) and taxonomic association (Burleigh 1972; DeSante & Pyle 1986; Ehrlich et al. 1988, Fig. 5). For the taxonomic analysis, we used scientific orders to differentiate species outside the order Passeriformes and (super) families to differentiate among Passeriforms. The number of correctly predicted presences increased with the numbers of presences (most species were present at fewer than half the sites) and for particular taxa. In order of descending model performance, taxa were as follows: (1) Passeroidea, (2) Tyrannidae, (3) Muscicapoidae, (4) Sylvioidea, (5), Corvoidea (6) Falconiformes, (7) Charadriiformes, (8) Piciformes, and (9) Caprimulgiformes. The pattern of performance among taxa is not explained by the average abundance of the species within each taxon (Pearson's product-moment correlation: \( \text{df}=7, t=-0.2775, p=0.79 \)).

**Discussion**

Probabilistic models parameterized using commonly available information on vegetation, climate, and spatial distribution can be used to effectively predict species presence at regional scales. In relative terms, the parameterized models predicted species presences about twice as well as a null model. In addition, the parameterized models performed remarkably evenly. Using any of the three single factor models, approximately the same number of occurrences could be predicted correctly for a species or taxon.

Similarity in the predictive ability of the models might be explained by the causal linkages among the evaluated factors. For example, it is well known that vegetation and climate
are related to each other as well as to factors such as elevation, latitude, and topography (Begon et al. 1986; Daubenmire 1959; Woodward 1987). There are also reasons to expect both climate and vegetation to be spatially autocorrelated (Legendre 1993; Brown 1995). Therefore, because we expect these potential predictors to be correlated with each other and in space, we might have expected the factors to predict equally well.

We found, however, that adding a second variable, especially climate or spatial data to the vegetation model, slightly but consistently improved the predictive ability of the models (model predictions improved by approximately 4%). Therefore, even though these parameters are causally linked, they are not completely redundant in terms of predictive ability. This result was surprising based on the single-factor model results. We expected that adding related factors would cause model overfit and jeopardize the generalizability of the model. We did find that a model that combined all three variables became overfit and was unable to outperform the models that combined only two variables.

Errors and Scaling

Coarse-grained approaches to predict species distributions have been criticized for their inability to predict without large sources of error (Short et al. 1995; Williams 1996; Conroy & Noon 1996). Clearly, uncertainties in map reference data and variations in model assumptions influence the patterns of predicted occurrences of species and the amount of errors (Stoms et al. 1992; Conroy & Noon 1996). The errors in our output were due to both cartographic errors and generalization as well as to the imperfect relationship between species and these different variables (Stoms et al. 1992; Williams 1996). We quantified the errors in using the coarse-grained approach, but did not attempt to identify the error sources. We attempted to minimize
errors in our analysis by using the data itself to fit the models and using the same scale of resolution in both the fitted and predicted models.

Our results, therefore, in part reflect the quality of the data we employed in parameterizing the models. We deliberately chose to use widely available data because we believe these data are most likely to be used in conservation efforts given time and funding constraints. It is probable that more refined climate, vegetation, and spatial information could improve the performance of the models we developed. Such changes could affect our conclusions about the comparability of different factors. Lastly, our analysis only included bird species and different taxa are likely to respond to the environment at different spatial scales (McNab 1963; Harris 1984).

Management Implications

Using readily available data sources we found we could correctly locate up to 65% of the sites a species occupies. Although, all of our parameterized models significantly outperformed the null model, all of the parameterized models also made many errors. We conclude that conservation biologists attempting to indirectly estimate species distributions and using comparable information sources should evaluate their plans with the knowledge that, with significant frequency, they will miss species presences and will mistakenly assign presences to locations without a species presence. In the latter case, the model may actually point out potentially valuable, unoccupied habitat. As has become clear from metapopulation theory, the potential importance of such habitat can be great (Hanski & Gilpin 1991).

The broader point is that even though models predicting species distributions are not perfect, our results suggest that using different parameters does not greatly affect model performance. The lack of distinction among the models in this regard has great practical
significance because there are often important tradeoffs among variables, including data collection costs and availability (Burbidge 1991). For example, the production of vegetation maps from remotely sensed information is an expensive, time-consuming process requiring highly trained professionals (Austin 1991). By contrast, climate data are often readily available. Spatial or contextual information is simply a by-product of organized sampling. Our results suggest that the marginal benefit of paying for coarse-scale data of one type when another type is already available may be small or nil (Margules & Austin 1991).

In particular, the results of these analyses could have implications for GAP. GAP distribution maps are based on the careful documentation of the distribution of terrestrial vegetation. GAP uses known associations between vegetation and habitat to predict the distribution of species presence/absence (Scott et al. 1993). Our results indicate that mistakes based on such an assumption may be fairly common. In addition, we found no evidence that vegetation is a superior predictive variable compared with climate and contextual distribution information. In fact, it may be that in some cases climate information provides a much cheaper but equally effective basis for predicting species distributions. This conclusion deserves an important caveat. In the state of Idaho, large regions remain covered with native vegetation. In regions where vegetation cover has been heavily modified (e.g., midwestern, eastern or coastal regions of the U.S.), land-use or land cover may be more effective and climate less effective as a predictor variable for species presence/absence.

In other situations existing, but patchily collected, distribution information may provide an effective basis for predicting presence/absence at unsampled locations. Spatial autocorrelation or “filling in the blanks” has been suggested as a predictive strategy in some instances (Legendre 1993) but see Koenig (1998). In one sense, using distributional context makes the fewest assumptions because the model does not assume the distribution varies with
just one variable (or set of variables), and thus, it captures the relevant factors affecting a species within a local region without assuming the best predictive variable is known. Our results suggest that biologists and managers should not disregard available distributional information if it is incomplete within a focal region. Finally, the high spatial autocorrelation of species presences and high error rates of predictive models are good arguments for expanding programs like the Breeding Bird Survey to include other taxa.

**Acknowledgments**

We thank P. Kareiva for inspiring this study and for helpful comments, A. Carter and M. Pascual for their suggestions on modeling, D. Veit for ornithological advice, K. Johnston for GIS assistance, C. Flores for translating and G. Booth for technical support. We would also like to thank J. Kiesecker, L. Freidenberg, A. Beckerman, K. Rothley, P. Hambach, F. Davis, and two anonymous reviewers for their contributions to previous versions of this manuscript. DKS was supported by a NSF Postdoctoral Fellowship in Environmental Biology.
Literature Cited:


Table 1. Common names, scientific names and the frequency of observing 40 bird species on 290 8-km BBS route segments in Idaho from 1988-1995.

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Scientific Name</th>
<th>Frequency of Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>California Gull</td>
<td><em>Larus californicus</em></td>
<td>0.114</td>
</tr>
<tr>
<td>Common Snipe</td>
<td><em>Gallinago gallinago</em></td>
<td>0.297</td>
</tr>
<tr>
<td>Spotted Sandpiper</td>
<td><em>Actitis macularia</em></td>
<td>0.217</td>
</tr>
<tr>
<td>Long-billed Curlew</td>
<td><em>Numenius americanus</em></td>
<td>0.124</td>
</tr>
<tr>
<td>Killdeer</td>
<td><em>Charadrius vociferus</em></td>
<td>0.434</td>
</tr>
<tr>
<td>Northern Harrier</td>
<td><em>Circus cyaneus</em></td>
<td>0.207</td>
</tr>
<tr>
<td>Red-tailed Hawk</td>
<td><em>Buteo jamaicensis</em></td>
<td>0.376</td>
</tr>
<tr>
<td>American Kestrel</td>
<td><em>Falco sparverius</em></td>
<td>0.397</td>
</tr>
<tr>
<td>Red-naped Sapsucker</td>
<td><em>Sphyrapicus nuchalis</em></td>
<td>0.159</td>
</tr>
<tr>
<td>Common Nighthawk</td>
<td><em>Cordeiles minor</em></td>
<td>0.207</td>
</tr>
<tr>
<td>Hammond’s Flycatcher</td>
<td><em>Empidonax hammondii</em></td>
<td>0.259</td>
</tr>
<tr>
<td>Dusky Flycatcher</td>
<td><em>Empidonax oberholseri</em></td>
<td>0.266</td>
</tr>
<tr>
<td>Horned Lark</td>
<td><em>Eremophila alpestris</em></td>
<td>0.424</td>
</tr>
<tr>
<td>Black-billed Magpie</td>
<td><em>Pica pica</em></td>
<td>0.400</td>
</tr>
<tr>
<td>Gray Jay</td>
<td><em>Perisoreus canadensis</em></td>
<td>0.107</td>
</tr>
<tr>
<td>American Crow</td>
<td><em>Corvus brachyrhynchos</em></td>
<td>0.331</td>
</tr>
<tr>
<td>European Starling</td>
<td><em>Sturnus vulgaris</em></td>
<td>0.417</td>
</tr>
<tr>
<td>Yellow-headed Blackbird</td>
<td><em>Xanthocephalus xanthocephalus</em></td>
<td>0.155</td>
</tr>
<tr>
<td>Red-winged Blackbird</td>
<td><em>Agelaius phoeniceus</em></td>
<td>0.490</td>
</tr>
<tr>
<td>Species</td>
<td>Binomial Name</td>
<td>Probability</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Brewer’s Blackbird</td>
<td>Euphagus cyanocephalus</td>
<td>0.545</td>
</tr>
<tr>
<td>Cassin’s Finch</td>
<td>Carpodacus cassini</td>
<td>0.248</td>
</tr>
<tr>
<td>Savannah Sparrow</td>
<td>Passerculus sandwichensis</td>
<td>0.269</td>
</tr>
<tr>
<td>Lark Sparrow</td>
<td>Chondestes grammacus</td>
<td>0.197</td>
</tr>
<tr>
<td>Chipping Sparrow</td>
<td>Spizella passerina</td>
<td>0.448</td>
</tr>
<tr>
<td>Brewer’s Sparrow</td>
<td>Spizella breweri</td>
<td>0.403</td>
</tr>
<tr>
<td>Sage Sparrow</td>
<td>Amphispiza belli</td>
<td>0.114</td>
</tr>
<tr>
<td>Song Sparrow</td>
<td>Melospiza melodia</td>
<td>0.462</td>
</tr>
<tr>
<td>Fox Sparrow</td>
<td>Passerella iliaca</td>
<td>0.121</td>
</tr>
<tr>
<td>Tree Swallow</td>
<td>Tachycineta bicolor</td>
<td>0.203</td>
</tr>
<tr>
<td>Rough-winged Swallow</td>
<td>Stelgidopteryx serripennis</td>
<td>0.224</td>
</tr>
<tr>
<td>Cedar Waxwing</td>
<td>Bombycilla cedrorum</td>
<td>0.110</td>
</tr>
<tr>
<td>Red-eyed Vireo</td>
<td>Vireo olivaceus</td>
<td>0.103</td>
</tr>
<tr>
<td>Yellow Warbler</td>
<td>Dendroica petechia</td>
<td>0.562</td>
</tr>
<tr>
<td>Audubon’s Warbler</td>
<td>Dendroica coronata</td>
<td>0.438</td>
</tr>
<tr>
<td>MacGillivray’s Warbler</td>
<td>Oporornis tolmiei</td>
<td>0.386</td>
</tr>
<tr>
<td>Sage Thrasher</td>
<td>Oreoscoptes montanus</td>
<td>0.331</td>
</tr>
<tr>
<td>Winter Wren</td>
<td>Troglodytes troglodytes</td>
<td>0.159</td>
</tr>
<tr>
<td>Red-breasted Nuthatch</td>
<td>Sitta canadensis</td>
<td>0.303</td>
</tr>
<tr>
<td>Black-capped Chickadee</td>
<td>Parus atricapillus</td>
<td>0.228</td>
</tr>
<tr>
<td>Mountain Bluebird</td>
<td>Sialia currucoides</td>
<td>0.141</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Map of the state of Idaho with BBS route segments denoted by their geographic centers. Presence/absence map for MacGillivray’s Warbler (x, presence; o, absence).

Figure 2. Mean percentage (± 1 SE) of sites that predicted species presence if it was present.

Figure 3. Mean percentage of sites that correctly predicted species presence, correctly predicted species absence, committed a commission error or committed an omission error.

Figure 4. Mean percentage of sites correctly predicted species presence plotted against commonness of 40 bird species.

Figure 5. Mean percentage of correctly predicted presences for 9 bird taxa. For each taxon, responses for the 4 single-factor models are presented. The taxa are arranged, left to right, from least to most abundant (average number of sites occupied).