

A GIS MODELING METHOD APPLIED TO PREDICTING FOREST SONGBIRD HABITAT

RANDY DETTMERS¹ AND JONATHAN BART²

*Ohio Cooperative Fish and Wildlife Research Unit, Biological Resources Division, U.S. Geological Survey, and
Department of Zoology, Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210 USA*

Abstract. We have developed an approach for using “presence” data to construct habitat models. Presence data are those that indicate locations where the target organism is observed to occur, but that cannot be used to define locations where the organism does not occur. Surveys of highly mobile vertebrates often yield these kinds of data. Models developed through our approach yield predictions of the amount and the spatial distribution of good-quality habitat for the target species. This approach was developed primarily for use in a GIS context; thus, the models are spatially explicit and have the potential to be applied over large areas. Our method consists of two primary steps. In the first step, we identify an optimal range of values for each habitat variable to be used as a predictor in the model. To find these ranges, we employ the concept of maximizing the difference between cumulative distribution functions of (1) the values of a habitat variable at the observed presence locations of the target organism, and (2) the values of that habitat variable for all locations across a study area. In the second step, multivariate models of good habitat are constructed by combining these ranges of values, using the Boolean operators “and” and “or.” We use an approach similar to forward stepwise regression to select the best overall model.

We demonstrate the use of this method by developing species-specific habitat models for nine forest-breeding songbirds (e.g., Cerulean Warbler, Scarlet Tanager, Wood Thrush) studied in southern Ohio. These models are based on species’ microhabitat preferences for moisture and vegetation characteristics that can be predicted primarily through the use of abiotic variables. We use slope, land surface morphology, land surface curvature, water flow accumulation downhill, and an integrated moisture index, in conjunction with a land-cover classification that identifies forest/nonforest, to develop these models.

The performance of these models was evaluated with an independent data set. Our tests showed that the models performed better than random at identifying where the birds occurred and provided useful information for predicting the amount and spatial distribution of good habitat for the birds we studied. In addition, we generally found positive correlations between the amount of habitat, as predicted by the models, and the number of territories within a given area. This added component provides the possibility, ultimately, of being able to estimate population sizes. Our models represent useful tools for resource managers who are interested in assessing the impacts of alternative management plans that could alter or remove habitat for these birds.

Key words: abiotic variables; Cerulean Warbler; cumulative distribution function; forest songbirds; GIS; microhabitat preferences; moisture gradients; presence data; Scarlet Tanager; spatially explicit habitat models; topography; Wood Thrush.

INTRODUCTION

Predictive habitat models that can be applied over large geographic areas have broad applicability in conservation biology and wildlife management, including such subdisciplines as ecosystem management and landscape ecology (Hunter 1996). Assessing the

amount or location of habitat over large areas in a manner that is neither labor intensive nor prohibitively time-consuming holds many benefits for managers and researchers working on large-scale issues.

The wildlife-habitat relationships (WHR) system (Verner and Boss 1980) and the habitat suitability index (HSI), developed by the U.S. Fish and Wildlife Service (1981), were among the first attempts to predict wildlife presence or relative abundance across large areas. These models, however, were based mainly on literature reviews, generally did not pertain to well-defined populations, and were not based on statistical models or procedures. Many of the models were not field tested, or performed poorly when they were tested (Bart

Manuscript received 30 July 1997; revised 2 March 1998; accepted 4 March 1998; final version received 2 April 1998.

¹ Present address: Department of Forestry, Wildlife, and Fisheries, University of Tennessee, Knoxville, Tennessee 37901 USA.

² Present address: Raptor Research Technical Assistance Center, USGS/BRD, Boise State University, 970 Lusk St., Boise, Idaho 83706 USA.

et al. 1984, Dedon et al. 1986, Laymon and Barrett 1986, Raphael and Marcot 1986, Johnson et al. 1989, Robel et al. 1993). Many authors (e.g., Van Horn and Wiens 1991) have recommended against the use of untested WHR and HSI models.

Several statistical methods can be used to develop predictive habitat models for large areas. These methods, and examples of their use, include linear regression (Morrison et al. 1987, Rice et al. 1993, Puttock et al. 1996), logistic regression (Straw et al. 1986, Nadeau et al. 1995, Pausas et al. 1995, St. Georges et al. 1995), discriminant analysis (Mosher et al. 1986, Livingston et al. 1990, Fielding and Haworth 1995), principal component analysis (Debinski and Brussard 1994), canonical correlation analysis (Andries et al. 1994), and classification and regression tree (CART) analysis (O'Connor et al. 1996).

These methods generally require that counts be made of how many individuals of the target species are present on plots. The sampling unit is thus a plot, and the variables are "number of animals present" and one or more descriptors of the habitat. In this context, zero means "none present," not simply "none recorded." When the "number" of animals is simply recorded as either 0 or 1, the data are often referred to as presence-absence data. Many wildlife surveys, however, do not provide this type of information (i.e., count data or presence-absence data). Instead, data are only collected from locations at which animals were observed. We refer to data of this sort as "presence data" to distinguish them from presence-absence data. Presence data are particularly common in surveys of highly mobile vertebrates that might use a given plot when the observer is not present. In such cases, it is almost impossible to declare that a given plot is never used by the target species. Instead, observers collect information about plots that were used.

Clark et al. (1993) and Knick and Dyer (1997) have developed a modeling approach tailored specifically for presence data. Their method is based on the Mahalanobis statistic (e.g., Johnson and Wichern 1988). In principle, this method defines "optimum" habitat as a multivariate vector of the means of the habitat variables, with the means calculated from the presence data. Habitat quality for each plot in the study area is based on the "distance" (i.e., similarity) to that multivariate mean, as measured by the Mahalanobis statistic. Calculations are based on the inverse of the covariance matrix and standardized variables and, thus, are more complex than our simplified description of the method. Knick and Dyer (1997) developed and tested a model for jackrabbits based on this approach. They found good correlation between observed locations, taken from a new sample, and habitat quality, as estimated by their model. For example, they constructed the cumulative distribution of habitat scores and found that the first 20% of the distribution contained 80% of the sightings in their validation sample. Clark et al.

(1993) found that their Mahalanobis model of black bear habitat reflected the same characterizations of habitat use as did tests of habitat selection using their habitat variables individually.

We have also developed an approach for using presence data to develop models that predict habitat quality. Our methods were developed independently of the approach just described, and were primarily for use in a GIS context. Models developed from our approach yield predictions of the amount and spatial distribution of good-quality habitat for the target species. In this paper, we describe our approach, provide examples of bird-habitat models developed using this approach, evaluate the performance of these models, and compare our approach with the method based on the Mahalanobis statistic.

METHODS

Description of the analytic method

The initial procedure in our model development is to partition a study area into "plots" (e.g., pixels in a GIS analysis) and to define a series of habitat predictor variables. The value of every variable must be known for every plot (not just plots sampled for presence of the target organism). Thus, this method generally is based on GIS methods, although, in principle, the "plots" might be X-Y locations that could be obtained without having GIS layers. Next, we obtain a sample of n "presence" observations of the target organism. We know in which plot each observation occurred and, thus, the values of the habitat variables associated with each observation.

Then we construct a quantitative definition of "good habitat" using the habitat variables. Rules for constructing the definition will be described. This definition is used to classify every plot in the study area as "good" or "otherwise." Throughout this article, we use "good habitat" simply to mean the area delineated by the model. No assumption is made about use in relation to availability, or about how the animal decides which areas to use. Let P_s be the proportion of a study area that is delineated as good habitat by a particular model. The quantity P_s is a known parameter, not a random variable, because its value for every plot in the study area can be calculated. We use uppercase letters for parameters and lowercase letters for their estimates. We can also classify each observation of the target species as being in good habitat or otherwise. Let p_o be the proportion of the n presence observations that are in good habitat. The quantity p_o is a random variable and an estimate of the proportion of observations that would be in good habitat with an indefinitely large sample size.

Our modeling procedure is an attempt to find the definition of good habitat that maximizes the quantity $(p_o - P_s)$, subject to the constraint that p_o is greater

than some threshold, t . We thus search for the definition of good habitat that yields

$$\max(p_o - P_s \mid p_o > t).$$

The rationale for using this definition may be explained as follows. If we could find a definition such that 90% of the observations are in good habitat ($p_o = 0.9$), but good habitat covers only 20% of the study area ($P_s = 0.2$), then we would feel fairly confident of having identified habitat that is highly attractive to the organism. In contrast, if a definition of good habitat captures only 60% of the observations and good habitat covers 50% of the study area, then such a definition would not describe highly attractive habitat, because a definition constructed at random would perform nearly this well (i.e., with a randomly constructed definition, the expected value of $p_o - P_s$ is 0.0). A threshold, t , is established because we wish the definition of good habitat to capture a large fraction of the observations and, thus, to be applicable to a large proportion of the target population. For example, we might require that the definition of good habitat capture at least 70% of the observations (i.e., $t \geq 0.7$). Different values of t can be used, depending on the context and intended application of the model.

Many approaches for defining good habitat might be imagined. Our approach involves two primary steps. First, we identify an "optimal" lower to upper range of values, X_{jL} to X_{jU} , for each habitat variable j , $j = 1, \dots, k$. Then in the second step, good habitat is defined by combining these ranges using the Boolean operators "and" and "or." For example, with three habitat variables, the definition of good habitat might be

$$[(X_{1L} < x_{1i} < X_{1U}) \text{ or } (X_{2L} < x_{2i} < X_{2U})] \text{ and } (X_{3L} < x_{3i} < X_{3U}) \quad (1)$$

where x_{1i} , x_{2i} , and x_{3i} are the values of habitat variables 1, 2, and 3 in plot i . This form allows such definitions as [(vegetation of type 1 or type 2) and close to water] and, thus, is much more flexible than a definition using only linear combinations of the habitat variables (e.g., $c_1x_{1i} + \dots + c_kx_{ki}$).

One problem with this approach is that the number of possible definitions for good habitat is extremely large (infinitely large if any of the variables is continuous, because an infinite number of choices for X_{jL} and X_{jU} then exists). An algorithm is thus needed to choose the rule set for defining good habitat. Our two-step modeling approach provides such an algorithm, which we now describe. In the first step, we consider each variable separately. For variable j , $j = 1, \dots, k$, we search for the range of values that yields

$$\max(p_{jo} - P_{js} \mid p_{jo} > t)$$

where p_{jo} is the proportion of observations falling within the range X_{jL} to X_{jU} and P_{js} is the proportion of the study area falling in the range X_{jL} to X_{jU} . We again

employ the constraint of $p_{jo} > t$ to ensure that we have accounted for a reasonably large proportion of the observations. To find this range, we use cumulative distribution functions (CDFs) of (1) the values of habitat variable j associated with the presence observations and (2) the values of habitat variable j for all the plots in a study area. Our calculations follow the equations described in Perry and Smith (1994) for CDFs without stratifying by study area. Using the CDFs, we can easily calculate: $F_o(X_j)$, the proportion of presence observations for which the value of habitat variable j is $\leq X$; and $F_s(X_j)$, the proportion of the study area in which the value of habitat variable j is $\leq X$. If the value of one of these two functions is always greater than the other over the entire range of values for variable j , then it is easy to show that the value of X_j that maximizes $F_o(X_j) - F_s(X_j)$, given $F_o(X_j) > t$, also maximizes $p_{jo} - P_{js}$, given $p_{jo} > t$. Because the condition of one function always being greater than the other over all values of a variable does not always hold, and because the CDFs can have multiple modes, we used the following, more complex, algorithm for finding the range X_{jL} to X_{jU} that maximizes $p_{jo} - P_{js}$.

If we set $p_{jo} = F_o(X_{jU}) - F_o(X_{jL})$ and $P_{js} = F_s(X_{jU}) - F_s(X_{jL})$, where U and L indicate upper and lower values of X_j , then we can represent the expression $p_{jo} - P_{js}$ as

$$[F_o(X_{jU}) - F_s(X_{jU})] + [F_s(X_{jL}) - F_o(X_{jL})]. \quad (2)$$

Thus, for cases in which the CDFs were unimodally distributed and $F_o(X_j) > F_s(X_j)$ for all or nearly all values of X_j , then X_{jU} was found as the value of X_j that maximized $F_o(X_j) - F_s(X_j)$, given $F_o(X_j) > t$. The corresponding X_{jL} was then found as the X_j that maximized $F_s(X_j) - F_o(X_j)$ given $p_{jo} > t$ and $X_{jL} < X_{jU}$. If $F_o(X_{jL})$ was never $< F_s(X_{jL})$ for any value of variable $j < X_{jU}$, then X_{jL} was set to the minimum possible value of variable j (see Fig. 1A). When the CDFs were unimodally distributed and $F_s(X_j) > F_o(X_j)$ for all or nearly all values of X_j , then X_{jL} was found as the X_j that maximized $F_s(X_{jU}) - F_o(X_{jU})$, given $F_o(X_j) < 1 - t$. The corresponding X_{jU} was found as the X_j that maximized $F_o(X_{jL}) - F_s(X_{jL})$, given $p_{jo} > t$ and $X_{jU} > X_{jL}$. If $F_o(X_j)$ was never $> F_s(X_j)$ for any value of variable $j > X_{jL}$, then X_{jU} was set to the maximum value of variable j (see Fig. 1B). For cases in which the CDFs were bimodal, we modified Expression 2 to contain two of each term in this expression, one representing each mode, and then applied the appropriate algorithm to each mode. More complex distributions could be handled in a similar fashion.

Using the range chosen for each variable, we constructed multivariable models by using the Boolean operators and parentheses (see Expression 1). Models were compared based on the value of $p_o - P_s$. All possible single-, two-, and three-variable models were evaluated. For higher order models, we used an approach similar in principle to forward stepwise regression to select the best model at each stage. For

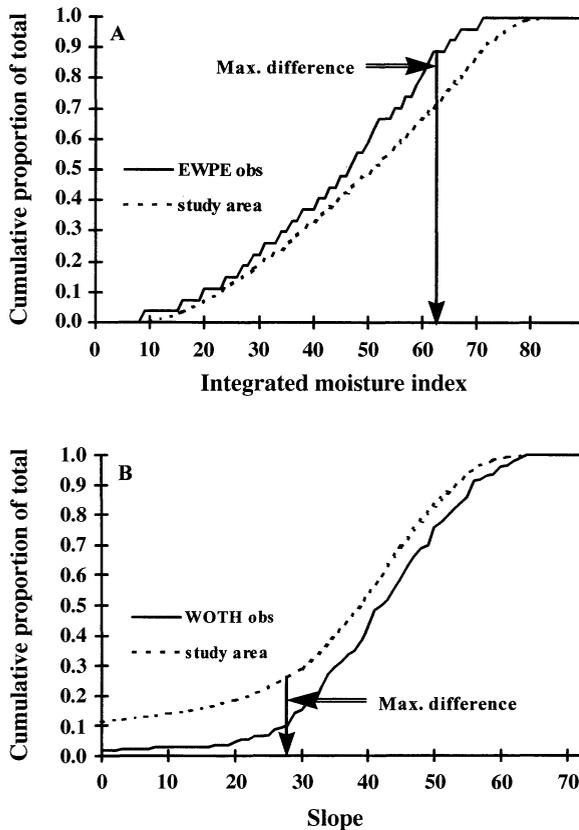


FIG. 1. Examples of how cumulative distribution functions (CDFs) were used in defining ranges of values for the habitat variables. As described in *Description of the analytic method*, the point where the maximum difference occurs between the two functions indicates one of the cutoff values for the range of that variable. In Example (A), which uses the integrated moisture index and observations for Eastern Wood-Pewee (EWPE), the range for slope would be defined as <62 . In (B), which uses slope and observations for Wood Thrush (WOTH), the CDF for the observations is less than that for the study area, so the range would be defined as >27 .

example, all possible four-variable models that could be constructed using the variables from the best three-variable model were considered, and the best one was selected, and so on. We selected the overall best model as the one that produced the maximum value of $p_o - P_s$. Like stepwise regression, this approach does not examine all possible models and does not necessarily find the “best” model, but it is tractable and seems likely to come at least fairly close to finding the optimal definition of good habitat. Knowledge of a species and its habitat was used in deciding how to combine variables (i.e., how to use parentheses and the operators “and” and “or”). In addition, a general rule of thumb was used in deciding which Boolean operator should be used. The “and” operator was typically used for cases in which at least one of the variables or groups of variables to be joined produced a relatively large value (>0.85) of p_o when in the model by itself. The “or” operator was typically used otherwise.

We used this analytic method to develop GIS-based models for nine bird species. Six habitat variables were used as predictors, and we used $t \geq 0.7$. All models were developed and tested at a 30-m resolution.

Study area and species

Our study was conducted on the Wayne National Forest and Raccoon Ecological Management Area in southeastern Ohio, United States, during 1994–1995 (Fig. 2). The predominant habitat type in this region is mature forest (60–100 yr old) dominated by mixed oak (*Quercus* spp.) and mixed hickory (*Carya* spp.). This is the vegetation type in which our study areas were established. Sugar maple (*Acer saccharum*), red maple (*Acer rubrum*), American beech (*Fagus grandifolia*), and tulip poplar (*Liriodendron tulipifera*) were other species commonly contributing to the overstory tree component of our study areas.

Models were developed to delineate preferred habitat for the nine most common forest songbirds in this area: Acadian Flycatcher (*Empidonax vireescens*, ACFL), Cerulean Warbler (*Dendroica cerulea*, CERW), Eastern Wood-Pewee (*Contopus virens*, EWPE), Hooded Warbler (*Wilsonia citrina*, HOWA), Ovenbird (*Seiurus aurocapillus*, OVEN), Red-eyed Vireo (*Vireo olivaceus*, REVI), Scarlet Tanager (*Piranga olivacea*, SCTA), Wood Thrush (*Hylocichla mustelina*, WOTH; see Plate 1), and Worm-eating Warbler (*Helmitheros vermivorus*, WEWA). These were the species for which we were able to collect a sufficient number of observations across all study areas for building and testing models.



PLATE 1. Typical Wood Thrush nesting habitat in southeastern Ohio. Photo credit: George H. Harrison.

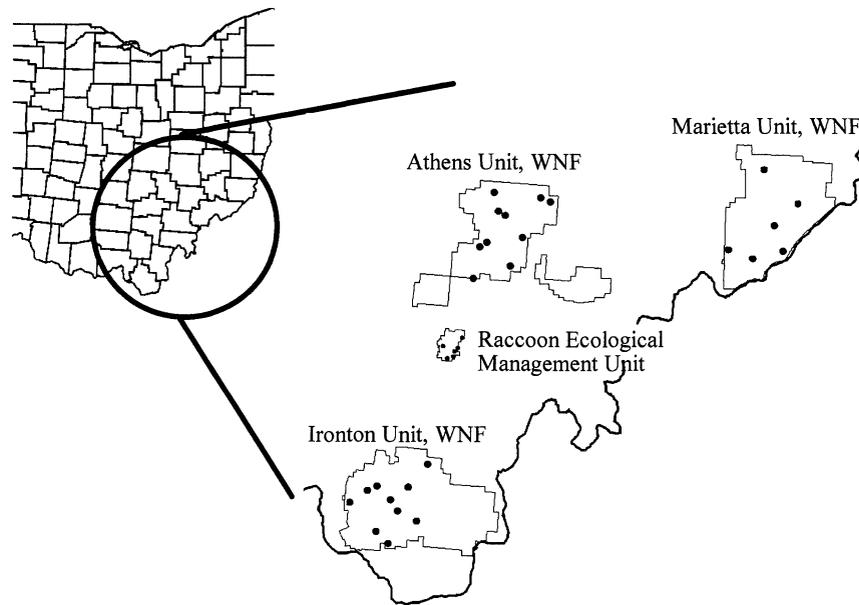


FIG. 2. Location of the 32 areas (solid dots) surveyed in 1995 to collect data for testing the bird-habitat models.

Bird surveys

Surveys were conducted during 0600 to 1200 and followed methods for territory mapping (Robbins 1970). Prior to the surveys, observers were trained for 2–3 d to insure competence in identifying and plotting birds. Each person covered ~30 ha/d. Bird observations were plotted on topographic maps and were later digitized using Arc/Info version 7.0 (Environmental Systems Research Institute 1995). Because we were unable to visit each study area a sufficient number of times, we did not estimate territory boundaries from the survey data, but used only the plotted observations. Thus, the survey data used to develop and evaluate our models consisted of observed points indicating “presence” locations, not polygons delineating territory boundaries.

In 1994, surveys were conducted on five 90-ha study areas, two in the Raccoon Ecological Management Area and three in the Ironton Unit. Each area was surveyed 2–3 times during 20 May–1 July. Data from these areas were used to develop the habitat models.

In 1995, surveys were conducted on 32 30-ha study areas. These areas were selected throughout the Wayne National Forest using a stratified random design. Strata and sample sizes were: Athens Unit, 10 study areas; Ironton Unit, 10; Marietta Unit, 6; and Raccoon Ecological Management Area, 6 (Fig. 2). Areas were surveyed on 3–4 consecutive days during 15 May–15 July. We felt that this level of effort was sufficient for estimating the *number* of territories, but not the location of territory boundaries, on each area. Data from these study areas were used to evaluate the performance (see *Model evaluation*) of habitat models developed using

the 1994 data. These data were also used to calculate our estimates of territory density.

Habitat variables

We obtained GIS layers for the following habitat variables: land cover (forest vs. nonforest), an “integrated moisture index” (Iverson et al. 1997), slope, surface curvature of the landscape, surface morphology (i.e., concave vs. convex land surfaces), and flow accumulation of water downslope. We used an existing land cover classification, prepared from 1986 Thematic Mapper data by the Ohio Department of Natural Resources (Yi et al. 1994), to distinguish forest from nonforest. The great majority of forest cover in southeastern Ohio can be characterized as an oak-hickory type (Zhu and Evans 1994), which we assumed to be an appropriate general habitat type for the species we modeled. The integrated moisture index was a relative rating (from 0 to 100) of environmental moisture, and was designed to indicate increasing levels of moisture available for plant growth. The index was a function of solar radiation, relative slope position, surface curvature, and water holding capacity of the soil. Calculation of the index is explained in Iverson et al. (1997). The other variables were calculated from 7.5-minute digital elevation model (DEM) data at a 1:24 000 scale and 30-m resolution using standard Arc/GRID commands (Environmental Systems Research Institute 1995).

Model evaluation

For consistency with earlier notation, we regarded each of the 32 areas used to evaluate model perfor-

mance as a separate "study area." Model performance was evaluated using the following quantities:

p_{oi} = proportion of the observations from study area i in good habitat, $i = 1, \dots, 32$

\underline{p}_o = mean of the p_{oi}

P_{si} = proportion of study area i in good habitat, $i = 1, \dots, 32$

\underline{P}_s = mean of the P_{si} .

Model performance was evaluated in four ways, each of which is discussed.

Satisfying the constraint.—The constraint used in model development was $p_o \geq 0.7$. It was expected that, occasionally, $p_o < 0.7$ on a new set of plots, but a general tendency for $p_o < 0.7$ would indicate poor performance of the model. We therefore tested whether \underline{p}_o was significantly less than 0.7, using a one-tailed t test. Failure to reject constituted support for the model.

Comparison with a random model.—A model constructed at random would tend to produce $\underline{p}_o - \underline{P}_s = 0.0$. We therefore determined whether the values $p_{oi} - P_{si}$ tended to be greater than 0.0 in our sample. Specifically, we used a one-tailed t test to determine whether $\underline{p}_o - \underline{P}_s$ was significantly greater than 0.0. This test showed whether the models had any predictive power (compared to a random model). The distribution of the differences and their mean values provided a description of how much (if at all) better than chance our models performed.

Distances to good habitat.—If the model were to delineate the spatial distribution of good habitat fairly well, then we might expect that observations outside good habitat would tend to be closer to good habitat than points randomly distributed in "other" habitat. We tested this hypothesis by calculating:

d_{oi} = for study area i , the average distance from observations in "other" habitat to the edge of the closest patch of good habitat, $i = 1, \dots, 32$

\underline{d}_o = mean of the d_{oi}

D_{si} = for study area i , the average distance from randomly selected points in "other" habitat to the edge of the closest patch of good habitat, $i = 1, \dots, 32$

\underline{D}_s = mean of the D_{si} .

The quantity D_{si} was estimated using 1000 randomly selected points for each study area. We used a one-tailed t test to determine whether $\underline{d}_o - \underline{D}_s$ was significantly less than 0.0.

Correlation between number of territories and amount of good habitat.—The 1995 surveys yielded an estimate of the number of territories present in each study area. If the habitat models were performing well, then these estimates should show significant positive

correlations with the amount of good habitat in the areas. The 30-ha areas were fairly uniform, so we divided each one into four 7.5-ha subplots and calculated correlations based on the resulting 128 subplots. When territories extended across the subplot boundaries, we estimated the proportion of the territory within a subplot to the closest fourth and recorded the "number present" accordingly (i.e., allowable values were multiples of 0.25). The subplots were not independent, so we used a bootstrap (Efron and Tibshirani 1993: Section 14.4) based on the 32 primary sampling units (i.e., study areas) to obtain standard errors of the correlation coefficients. A one-tailed t test was then used to determine whether the correlation between number of territories and proportion of the area delineated as good habitat was greater than 0.0.

RESULTS

Model development

The habitat models developed from the 1994 data indicated that Acadian Flycatchers (Fig. 3A) and Worm-eating Warblers primarily utilized stream bottoms and ravines, which have concave land surfaces and high moisture levels (Table 1). These two species were quite restricted in the microhabitats they used and could be considered habitat "specialists" at this level. We use the term "microhabitat" to indicate finer scale areas (e.g., territories or home ranges) than that of a general vegetation association or habitat type, equivalent to Johnson's (1980) second-order habitat. Ovenbirds, Red-eyed Vireos (Fig. 3B), and Wood Thrushes were widely distributed, commonly occurred on hill-sides, and were the most general in their microhabitat preferences. Typically, they were found in mid-slope areas where slopes were relatively steep and moisture levels were intermediate (Table 1). Hooded Warblers and Eastern Wood-Pewees (Fig. 3C) were characterized as hilltop species associated with the dry moisture conditions and convex land forms of ridges (Table 1). Cerulean Warblers (Fig. 3D) and Scarlet Tanagers were also associated with convex land surfaces, although they were not commonly found on the very tops of ridges. Scarlet Tanagers were typically in slightly more moist and less steep areas, whereas Cerulean Warblers tended to occupy drier and steeper sites (Table 1).

Model performance

Satisfying the constraint.—For six of the nine species, $\underline{p}_o > 0.7$; for three species, \underline{p}_o was slightly, but not significantly ($P > 0.05$), less than 0.7 (Fig. 4). The models thus performed well in satisfying the constraint that \underline{p}_o should not be significantly less than 0.7.

Comparison with a random model.—The difference $\underline{p}_o - \underline{P}_o$ was significantly ($P < 0.05$) greater than 0.0 for all nine species (Fig. 4). For all of the species except Ovenbird and Red-eyed Vireo, the difference was greater than 0.10. The difference for the Acadian Fly-

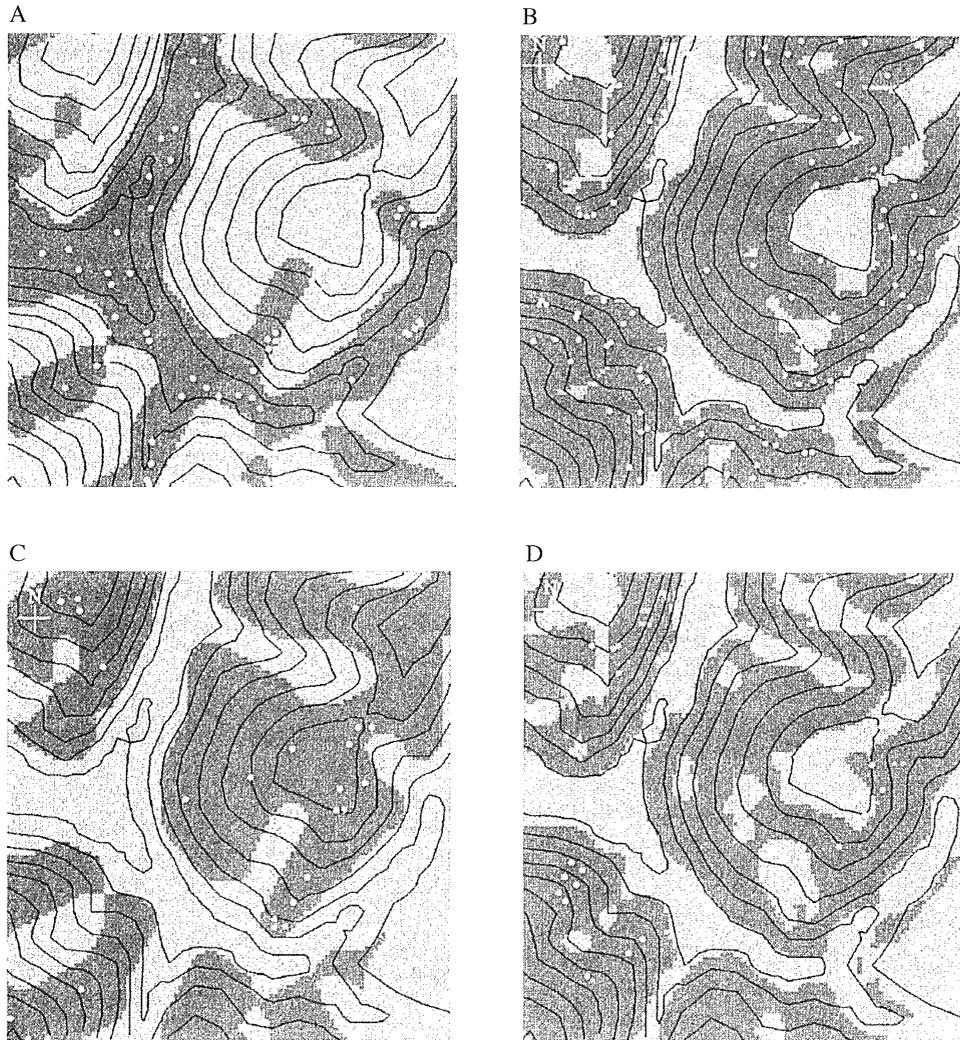


FIG. 3. Models of Acadian Flycatcher (A), Red-eyed Vireo (B), Eastern Wood-Pewee (C), and Cerulean Warbler (D) habitat for one of the test plots in the Raccoon Ecological Management Area. Dark gray shading indicates good habitat, as delineated by the model. White dots indicate locations of individuals of the respective species observed during territory mapping surveys. The figures represent standard topographic maps, with black lines marking 20-foot (6.1-m) elevation contours as generated from digital elevation model data. The white crosses indicate cardinal directions, with north noted by N.

TABLE 1. Definitions of good habitat for the nine bird models developed in this study.

Species	Definition [†]
Acadian Flycatcher	$(29 \leq imi \leq 72$ or $curv < 1$ or $slope < 40$) and $morph = concave$ and $landcov = forest$
Cerulean Warbler	$((morph = convex$ and $27 > imi > 52)$ or $35 \leq slope \leq 61$) and $(9 \leq fa \leq 45$ or $-3 \leq curv \leq 1)$ and $landcov = forest$
Eastern Wood-Pewee	$(slope < 46$ or $-2 \leq curv \leq 5)$ and $morph = convex$ and $(morph = convex$ or $imi < 62)$ and $landcov = forest$
Hooded Warbler	$(morph = convex$ and $curv > -1)$ or $slope > 50$) and $landcov = forest$
Ovenbird	$(imi < 42$ or $(-1 \leq curv \leq 11$ and $slope > 30))$ and $landcov = forest$
Red-eyed Vireo	$31 \leq slope \leq 59$ and $-1 \leq curv \leq 4$ and $9 \leq fa \leq 90$ and $landcov = forest$
Scarlet Tanager	$(slope < 38$ or $imi > 52)$ and $-1 \leq curv \leq 7$ and $landcov = forest$
Wood Thrush	$(9 \leq fa \leq 45$ or $38 \leq imi \leq 64)$ and $slope > 27$ and $landcov = forest$
Worm-eating Warbler	$((35 \leq slope \leq 44$ and $curv < 1$ and $9 \leq imi \leq 61)$ or $morph = concave$) and $landcov = forest$

[†] Abbreviations are as follows: imi, integrated moisture index score; curv, percentage surface curvature; morph, surface morphology (concave or convex); landcov, land cover type (forest or nonforest); fa, water flow accumulation.

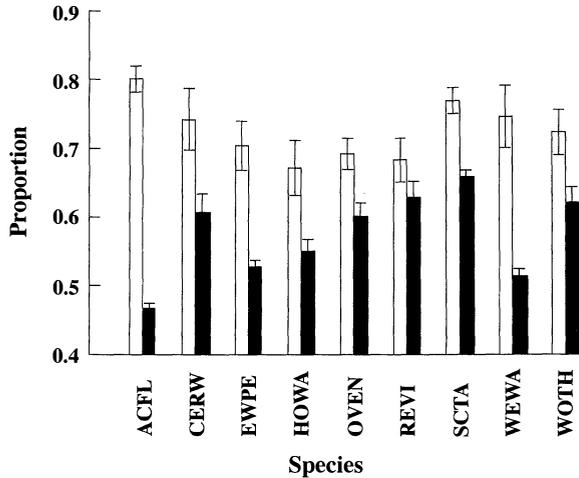


FIG. 4. Comparison of the proportion (mean \pm 1 SE) of the bird observations occurring in good habitat (open bars) with the proportion (mean \pm 1 SE) of a study area delineated as good habitat by the model for each species (solid bars). Proportions differed ($P < 0.05$) for all species. Species codes are given in *Methods: Study area and species*.

catcher and Worm-eating Warbler models was greater than 0.20, indicating that these models were particularly effective. Development of highly predictive models was much easier for these microhabitat specialists than for other species, such as Red-eyed Vireo, that were not as restrictive in the microhabitats that they used.

Distances to preferred habitat.—The difference $d_o - D_s$ (see *Methods, Distances to good habitat*) was significantly ($P < 0.05$) less than 0.0 for all of the species except Wood Thrush (Table 2).

Correlation between number of territories and amount of good habitat.—The correlation between number of territories and proportion of good habitat in a study area was positive for all species and significantly ($P < 0.05$) greater than 0.0 for all species except Hooded Warbler, for which it was marginally significant (Table 3).

DISCUSSION

Model evaluation

The models developed in this study passed several tests designed to evaluate model performance. Good

habitat, as defined by the model, included $>70\%$ of the bird observations. The proportion of observations in good habitat was larger than the proportion of the study area delineated as good habitat. Observations outside of good habitat, as delineated by the model, tended to be closer to good habitat than would occur by chance. The correlation between number of territories and amount of good habitat was generally positive. The models thus provided more information than, for example, use of the survey data alone without habitat information, or use of a randomly constructed habitat model.

Three products of practical use can be obtained from the models: (1) maps delineating good habitat for the species, (2) an estimate of the amount of good habitat within the area of interest, and (3) an estimate of the number of territories within that area. These products could be used in evaluating alternative management plans by determining which of the plans might result in the elimination of important habitat or an unacceptably large number of breeding territories. The territory density estimates from our models would not be highly precise, as indicated by the low correlations for some of the models, and would thus have to be used with extreme caution. Models involving many parameters, such as the ones developed in this study, should not be applied to areas outside of southeastern Ohio without testing the models on a data set from the new area. In many cases, however, approximately the same model may suffice, so that model development might be relatively easy.

Although our models clearly performed better than a random model, they generally did not identify small proportions of the study area that contained large proportions of the target species. Also, the correlations between number of territories and amount of good habitat were quite low for some of the species, indicating that only a small amount of the variability in territory densities was associated with the predicted amount of good habitat. Many factors not accounted for in our models (e.g., food availability, interspecific competition, predator abundance) are likely to affect avian density (Wiens 1989), and thus it is not surprising that our correlations were low. This has been true in many other studies. For example, Morrison et al. (1987) developed

TABLE 2. Distance in meters (mean \pm 1 SE) from observed locations and random points to the nearest patch of good habitat. A one-sided *t* test was used to determine whether the mean for the observations was significantly less than the mean for the random points.

Species	Observations	Random points	Difference	df	<i>P</i>
Acadian Flycatcher	12.3 (0.9)	22.4 (0.9)	-10.1	27	<0.01
Cerulean Warbler	15.0 (1.4)	18.0 (0.4)	-3.0	30	<0.01
Eastern Wood-Pewee	17.3 (1.8)	22.4 (0.4)	-5.1	23	<0.01
Hooded Warbler	17.2 (1.1)	20.9 (0.4)	-3.7	31	<0.01
Ovenbird	15.5 (1.1)	17.5 (0.6)	-2.0	29	<0.05
Red-eyed Vireo	15.8 (1.0)	17.4 (1.0)	-1.6	31	<0.05
Scarlet Tanager	11.2 (0.7)	12.5 (0.2)	-1.3	34	<0.05
Worm-eating Warbler	16.4 (1.5)	19.7 (1.0)	-3.3	45	<0.01
Wood Thrush	17.9 (1.5)	17.8 (1.1)	0.1	28	>0.05

TABLE 3. Correlation (r) between predicted amount of good habitat and observed number of territories. P values are for one-sided t tests used to determine whether the correlation coefficients were greater than zero ($n = 32$ for all). Standard errors were calculated through a bootstrap analysis.

Species	r	1 SE of r	P
Acadian Flycatcher	0.412	0.084	<0.001
Cerulean Warbler	0.332	0.100	0.001
Eastern Wood-Pewee	0.294	0.092	0.002
Hooded Warbler	0.155	0.092	0.051
Ovenbird	0.376	0.105	<0.001
Red-eyed Vireo	0.373	0.101	<0.001
Scarlet Tanager	0.241	0.103	0.013
Worm-eating Warbler	0.175	0.098	0.042
Wood Thrush	0.415	0.098	<0.001

regression models to predict avian density and found that, in 15 of 21 models, the r^2 values were ≤ 0.15 . On the other hand, models are sometimes much more successful at identifying small areas in which the target species are concentrated. For example, Knick and Dyer (1997) used their habitat model to identify 20% of the study area that contained 80% of the observed individuals.

At least three reasons might account for the failure of a model to delineate a small fraction of the study area containing a large proportion of the animals. First, the most important variables may not be identified, or they may not be combined in the best manner. This is clearly an inadequacy of the model. We doubt that any simple way of investigating this possibility exists, but consideration of the species' natural history will often provide some indication of how reasonable the models are. Second, the sample size of observations may be too small. With a larger sample, a better model might be derived. This is essentially a design problem, although one that may be unavoidable because of practical and logistical constraints. Precision of the models in relation to sample size could probably be investigated with bootstrapping techniques, although we did not do this. Third, the animals may occur more or less uniformly throughout most of the study area. This situation makes it impossible to delineate small fractions of the study area that include large fractions of the animals. For example, the models that we presented in this paper were attempts to delineate microhabitats within an appropriate general habitat type covering large portions of the study areas. Thus, it was not surprising that species without highly specific microhabitat preferences were fairly uniformly distributed throughout the study areas. When such a situation occurs, we suggest that if a model seems reasonable and sample sizes seem large enough, then results indicating that a large proportion of the study area is required to account for a large proportion of the observations should be viewed as supporting a null hypothesis that most of the area really is good habitat. In the absence of falsification by future modeling work, the model should not be viewed as having performed poorly, but

rather as having revealed the useful finding that equally good habitat occurs throughout much of the study area. The management implication of such a finding is that no particular portions of the area are considerably better than others. In our study, we would reach this tentative conclusion for Ovenbird, Red-eyed Vireo, Scarlet Tanager, and Wood Thrush.

Ecological implications of the study

In addition to providing products of direct practical use, the modeling approach developed here may reveal new information about ways in which environmental variables affect habitat quality for the target species. In our study, variables describing land cover, topography, and moisture provided the basis for modeling microhabitats for several species of forest-breeding songbirds. For forest-breeding birds in the lower Midwest and southern Appalachians, topography and moisture are important in determining ecological conditions critical to defining microhabitat preferences of these species. Topography and moisture gradients can strongly influence the composition and structure of the plant community in a particular location (Carmean 1965, Host et al. 1987) and are useful in demarcating ecologically meaningful boundaries important in determining plant distributions and forest composition (Franklin 1995, Iverson et al. 1997). The structure and composition of the plant community can, in turn, be a critical factor in defining where preferred microhabitat occurs for a particular bird species (MacArthur and MacArthur 1961, Recher 1969, Rotenberry and Wiens 1980).

Some authors suggest that abiotic environmental variables, in addition to vegetation structure and composition, are likely to improve the accuracy of wildlife habitat descriptions (MacArthur and Wilson 1967, Willson 1974, Roth 1976). Environmental moisture has been suggested as an important direct determinant of microhabitat preferences of some birds (Odum 1950, Bertin 1977, Kendeigh and Fawver 1981) and of bird species diversity in some communities (Smith 1977, Swift et al. 1984, Petit et al. 1985). Thus, the habitat components represented by the abiotic variables used in this study might directly influence habitat preference by providing some of the cues that birds use for selecting preferred habitats. These abiotic variables also have a critical and strong indirect influence through their importance in defining plant community composition and structure. This is very important when detailed, remotely sensed data on specific vegetation types are not available, as was the case for our study.

Another implication of our study is that detailed data on plant species composition or structure may not always be needed to construct useful predictive models. This is fortunate, because such detailed information on the vegetation is seldom available for large areas, and models based on detailed vegetation information usually cannot be extrapolated to regional levels.

Comments on the analytic approach

Three features of the approach used in this study may warrant consideration. First, the general criterion for defining good habitat was maximization of the difference between proportion of the observations in good habitat and proportion of the study area delineated as good habitat. This criterion is somewhat similar to criteria used in some multivariate methods (e.g., discriminant analysis, which maximizes "separation" of two groups), but it is by no means identical to the criteria used in these methods. None of these multivariate methods could be used simply, without modification, to carry out our analysis.

Second, in defining good habitat, we used ranges for each underlying variable along with the Boolean operators "and" and "or." If we had used only "and," then the form of the definition would have been a linear combination in which the coefficients were 1 if the habitat variable was in the specified range and 0 otherwise. Use of the operator "or" and parentheses to form groups seems to preclude any simple mathematical method for identifying the best definition of good habitat, which is unfortunate. However, our method of construction does enable the definition to capture relevant ecological information, particularly pertaining to resources that substitute for one another, in ways that the linear combinations generally used in statistical methods cannot.

Third, we used one particular algorithm to search for the best model, but many others can be imagined. A more formalized algorithm, soundly based on mathematical theory, would provide an improvement in this area of the analysis.

A comparison of our method with the approach based on the Mahalanobis distance may be of interest, because the Mahalanobis method is the most commonly used of only a few modeling methods that, to our knowledge, are applicable for use with presence data. The Mahalanobis method has been used to provide a continuous measure of relative habitat quality, rather than to delineate good habitat. Good habitat could easily be delineated from this method, however, simply by establishing a threshold value. Our method utilized a threshold value and delineated good habitat, but it did not provide a continuous measure of habitat quality. We are unsure exactly how a continuous measure should be derived from our method. A single value could easily be obtained for each habitat variable (e.g., the midpoint of the ranges we used), but use of the "or" operator in defining good habitat is integral to our method and does not lend itself very well to a single description of optimal habitat. An approach could certainly be devised, but might be complex. Thus, the Mahalanobis method may be more general than the approach we developed.

The Mahalanobis method pre-selects a definition of optimal habitat (a multivariate vector of the means for the habitat variables), whereas our approach permits

great flexibility in defining good habitat. One aspect of the flexibility in our method that we like is the ability to identify resources that substitute for each other. Neither the Mahalanobis method, nor any of the multivariate methods that we are aware of, allows this degree of flexibility in defining good habitat. We also believe that the flexibility in defining the $p_{jo} > t$ threshold is a benefit in terms of developing models to suite different applications. Determining the amount of habitat or the proportion of a population to be targeted for management or conservation requires consideration of numerous complex issues (e.g., biology, economics, societal value judgments), which will vary between situations (Shaffer 1987). The value of t should thus be selected to fit the context of the situation for which the model is being developed (e.g., resource planning, conservation applications) and the biology of the target species.

Our method does provide a means for comparing the relative "fit" of different models through the quantity ($p_{jo} - P_{js}$). This measure is not statistically formalized, but it does provide a means of evaluating model performance based on the original data set. We are unaware of any methods for assessing the relative fit of different models developed through the Mahalanobis distance method. Additionally, our method of comparing CDFs to identify the range in values to be used for a given variable might be more appropriate for cases in which the distributions of the predictor variables are not unimodal or normal. Our method seeks to maximize the quantity ($p_{jo} - P_{js}$) for each variable, regardless of the variable's distribution. Mahalanobis distance models are a function of the differences between the means and observed values of the variables in the model as well as the covariance matrix for the variables. Thus, oddly shaped (e.g., bimodal, highly skewed) distributions could cause the Mahalanobis method to inappropriately categorize habitat quality, as the properties of the Mahalanobis function are best known under conditions of multivariate normality.

Development of habitat models that use presence data is at an early stage. Other methods can probably be produced. In particular, to our knowledge, no statistician has yet published on the issue, so it is possible that methods based more extensively on existing multivariate techniques can be developed. Developing methods that make optimal use of biological insights, but that also take full advantage of statistical methods, is probably the next step in this line of inquiry.

ACKNOWLEDGMENTS

Funding was provided by Wayne National Forest and the Northeast Forest Experiment Station (USDA Forest Service), the Global Change Program (BBIRD) of USGS/BRD, and the nongame bird program of Region 3, U.S. Fish and Wildlife Service. Our gratitude to those who assisted in collecting data: Vanessa Artman, Ron Beatty, Sandy Bloomfield, Ann Puthoff, Shay Garriock, Randy Harrison, and Tom Nealen. Special thanks to Elaine Sutherland, Ralph Miller, Louis Iverson, Lynda Andrews, Kathy Flagel, and Steve Lewis for their help and support. Louis Iverson, members of the Ohio

CFWRU, and two anonymous reviewers provided helpful comments on earlier drafts.

LITERATURE CITED

- Andries, A. M., H. Gulinck, and M. Herremans. 1994. Spatial modelling of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. *Ecography* **17**: 278–287.
- Bart, J., D. Petit, and G. Linscombe. 1984. Evaluation of two habitat models developed using the Habitat Evaluation Procedures approach. *Transactions of the North American Wildlife and Natural Resources Conference* **49**:489–499.
- Bertin, R. I. 1977. Breeding habitats of the Wood Thrush and Veery. *Condor* **79**:303–311.
- Carmean, W. H. 1965. Black oak site quality in relation to soil topography in southeastern Ohio. *Soil Science Society of America Proceedings* **29**:308–312.
- Clark, J. D., J. E. Dunn, and K. G. Smith. 1993. A multivariate model of female black bear habitat use for a geographic information system. *Journal of Wildlife Management* **57**:519–526.
- Debinski, D. M., and P. F. Brussard. 1994. Using biodiversity data to assess species–habitat relationships in Glacier National Park, Montana. *Ecological Applications* **4**:833–843.
- Dedon, M. F., S. A. Laymon, and R. H. Barrett. 1986. Evaluating models of wildlife–habitat relationships of birds in black oak and mixed-conifer habitats. Pages 115–119 in J. Verner, M. Morrison, and C. J. Ralph, editors. *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates*. University of Wisconsin Press, Madison, Wisconsin, USA.
- Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap*. Chapman, New York, USA.
- Environmental Systems Research Institute. 1995. Arc/Info command references and users guides 7.0, the geographic information system software. Redlands, California, USA.
- Fielding, A. H., and P. F. Haworth. 1995. Testing the generality of bird habitat models. *Conservation Biology* **9**: 1466–1481.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* **19**:474–499.
- Host, G. E., K. S. Pregitzer, C. W. Ramm, J. B. Hart, and D. T. Cleland. 1987. Landform-mediated differences in successional pathways among upland forest ecosystems in northwestern lower Michigan. *Forest Science* **33**:445–457.
- Hunter, M. L. 1996. *Fundamentals of conservation biology*. Blackwell Science, Cambridge, Massachusetts, USA.
- Iverson, L. R., M. E. Dale, C. T. Scott, and A. Prasad. 1997. A GIS-derived integrated moisture index to predict forest composition and productivity in Ohio forests. *Landscape Ecology* **12**:331–348.
- Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* **61**:65–71.
- Johnson, D. H., M. C. Hammond, T. L. McDonald, C. L. Nustad, and M. D. Schwartz. 1989. Breeding canvasbacks: a test of a habitat model. *Prairie Naturalist* **21**:193–202.
- Johnson, R. A., and D. W. Wichern. 1988. *Applied multivariate statistical analyses*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- Kendeigh, S. C., and B. J. Fawver. 1981. Breeding bird populations in the Great Smokey Mountains, Tennessee and North Carolina. *Wilson Bulletin* **93**:218–242.
- Knick, S. T., and D. L. Dyer. 1997. Distribution of black-tailed jackrabbit habitat determined by GIS in southwestern Idaho. *Journal of Wildlife Management* **61**:75–85.
- Laymon, S. A., and R. H. Barrett. 1986. Developing and testing habitat-capability models: pitfalls and recommendations. Pages 87–91 in J. Verner, M. Morrison, and C. J. Ralph, editors. *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates*. University of Wisconsin Press, Madison, Wisconsin, USA.
- Livingston, S. A., C. S. Todd, W. B. Krohn, and R. B. Owen, Jr. 1990. Habitat model for nesting Bald Eagles in Maine, USA. *Journal of Wildlife Management* **54**:644–653.
- MacArthur, R. H., and J. W. MacArthur. 1961. On bird species diversity. *Ecology* **42**:594–598.
- MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Monographs in Population Biology 1, Princeton University Press, Princeton, New Jersey, USA.
- Morrison, M. L., I. C. Timossi, and K. A. With. 1987. Development and testing of linear regression models predicting bird–habitat relationships. *Journal of Wildlife Management* **51**:247–253.
- Mosher, J. A., K. Titus, and M. R. Fuller. 1986. Developing a practical model to predict nesting habitat of woodland hawks. Pages 31–35 in J. Verner, M. Morrison, and C. J. Ralph, editors. *Wildlife 2000: modeling habitat relationships of terrestrial vertebrates*. University of Wisconsin Press, Madison, Wisconsin, USA.
- Nadeau, S., R. Decarie, D. Lambert, and M. St. Georges. 1995. Nonlinear modeling of muskrat habitat. *Journal of Wildlife Management* **59**:110–117.
- O'Connor, R. J., M. T. Jones, D. White, C. Hunsaker, T. Loveland, B. Jones, and E. Preston. 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. *Biodiversity Letters* **3**:97–110.
- Odum, E. P. 1950. Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology* **31**:587–605.
- Pausas, J. G., L. W. Braithwaite, and M. P. Austin. 1995. Modelling habitat quality for arboreal marsupials in the south coastal forests of New South Wales, Australia. *Forest Ecology and Management* **78**:39–49.
- Perry, R. I., and S. J. Smith. 1994. Identifying habitat associations of marine fishes using survey data: an application to the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences* **51**:589–602.
- Petit, D. R., K. E. Petit, and T. C. Grubb, Jr. 1985. On atmospheric moisture as a factor influencing distribution of breeding birds in temperate deciduous forest. *Wilson Bulletin* **97**:88–96.
- Puttock, G. D., P. Shatkatko, and J. G. Rasaputra. 1996. An empirical model for moose, *Alces alces*, in Algonquin Park, Ontario. *Forest Ecology and Management* **81**:169–178.
- Raphael, M. G., and B. G. Marcot. 1986. Validation of a wildlife–habitat relationship model: vertebrates in a Douglas-fir sere. Pages 129–138 in J. Verner, M. Morrison, and C. J. Ralph, editors. *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates*. University of Wisconsin Press, Madison, Wisconsin, USA.
- Recher, H. 1969. Bird species diversity and habitat in Australia and North America. *American Naturalist* **103**:75–80.
- Rice, S. M., F. S. Guthery, G. S. Spears, S. J. Demaso, and B. H. Koerth. 1993. A precipitation–habitat model for northern bobwhites on semiarid rangeland. *Journal of Wildlife Management* **57**:92–102.
- Robbins, C. S. 1970. Recommendations for an international standard for a mapping method in bird census work. *Audubon Field Notes* **24**:723–726.
- Robel, R. J., L. B. Fox, and K. E. Kemp. 1993. Relationship between habitat suitability index values and ground counts of beaver colonies in Kansas. *Wildlife Society Bulletin* **21**: 415–421.
- Rotenberry, J. T., and J. A. Wiens. 1980. Habitat structure, patchiness, and avian communities in North America steppe vegetation: a multivariate analysis. *Ecology* **61**:1228–1250.
- Roth, R. R. 1976. Spatial heterogeneity and bird species diversity. *Ecology* **57**:773–782.
- St. Georges, M., S. Nadeau, D. Lambert, and R. Decarie.

1995. Winter habitat use by ptarmigan, snowshoe hares, red foxes, and river otters in the boreal forest:tundra transition zone of western Quebec. *Canadian Journal of Zoology* **73**:755-764.
- Shaffer, M. 1987. Minimum viable populations: coping with uncertainty. *In* M. E. Soulé, editor. *Viable populations for conservation*. Cambridge University Press, Cambridge, UK.
- Smith, K. G. 1977. Distribution of summer birds along a forest moisture gradient in an Ozark watershed. *Ecology* **58**:810-819.
- Straw, J. A., Jr., J. S. Wakely, and J. E. Hudgins. 1986. A model for management of diurnal habitat for American woodcock in Pennsylvania. *Journal of Wildlife Management* **50**:378-383.
- Swift, B. L., J. S. Larson, and R. M. DeGraaf. 1984. Relationship of breeding bird density and diversity to habitat variables in forested wetlands. *Wilson Bulletin* **96**:48-59.
- U.S. Fish and Wildlife Service. 1981. Standards for the development of suitability index models. *Ecological services manual* 103. U.S. Department of Interior, Fish and Wildlife Service, Division of Ecological Services. Government Printing Office, Washington, D.C., USA.
- Van Horne, B., and J. A. Wiens. 1991. Forest bird habitat suitability models and the development of general habitat models. U.S. Fish and Wildlife Service, Fish and Wildlife Research 8.
- Verner, J., and A. S. Boss. 1980. California wildlife and their habitats: western Sierra Nevada. U.S. Department of Agriculture, Forest Service, General Technical Report PSW-37. Pacific Southwest Forest and Range Experiment Station, Berkeley, California, USA.
- Wiens, J. A. 1989. The ecology of bird communities. Volumes 1 and 2. Cambridge University Press, Cambridge, UK.
- Willson, M. F. 1974. Avian community organization and habitat structure. *Ecology* **55**:1017-1029.
- Yi, G.-C., D. Risley, and M. Koneff. 1994. Development of Ohio's GIS-based wetlands inventory. *Journal of Soil and Water Conservation* **49**:23-28.
- Zhu, Z., and D. L. Evans. 1994. U.S. forest types and predicted percent forest cover from AVHRR data. *Photogrammetric Engineering and Remote Sensing* **60**:525-531.