

Chapter 5 Insights from ecological theory on temporal dynamics and species distribution modeling

Robert J. Fletcher, Jr.^{1*}, Jock S. Young², Richard L. Hutto², Anna Noson², and Christopher T. Rota^{1,3}

¹Department of Wildlife Ecology and Conservation, PO Box 110430, 110 Newins-Ziegler Hall, University of Florida, Gainesville, FL 32611

²Avian Science Center, Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

³Present address: Department of Fisheries and Wildlife Sciences, University of Missouri, Columbia, MO 65221, USA

*Contact information: e-mail: robert.fletcher@ufl.edu

Abstract

Species distribution models are valuable tools for conservationists, managers, and land planners, yet can have limited effectiveness in predicting distributions locations in space and time. One way to help overcome such limitations is to incorporate perspectives from ecological theory to guide the modeling process, which could improve predictive performance and inference on resource quality. We identify two key concepts gleaned from habitat selection and metapopulation theory regarding the variation in species occurrence over time: density-dependent habitat use with fluctuating population size and explicit consideration of temporal persistence in occurrence across environmental gradients. We incorporate these perspectives into modeling approaches and test them against conventional logistic regression models using a large-scale data for four bird species that vary widely in their temporal variation in occurrence and habitat breadth in Montana and Idaho, 1994-2004. We expected that models for species that exhibit more variability in occurrence over time and use multiple habitats should benefit more from incorporating these concepts than for species that show less variability. Overall, incorporating density-dependent temporal variation in habitat use substantially improved model fit and moderately improved model predictive ability relative to conventional models.

5.1 Introduction

“Theory potentially has value far beyond the satisfaction of intellectual curiosity” Huston (2002)

Understanding species distributions in space and time is essential to ecology, evolution, and conservation biology. There is a growing need for robust habitat models that can adequately predict species distributions across broad spatial scales (Guisan and Thuiller 2005). An invaluable tool for conservation biologists (Norris 2004), species distribution models can be used to evaluate potential management actions, interpret the potential effects of climate change, and maximize biodiversity with reserve selection algorithms (Guisan and Thuiller 2005). Yet the usefulness of such models is limited by a number of factors (Guisan and Thuiller 2005; Araújo and Guisan 2006), including poor incorporation of ecological theory in modeling approaches (Austin 2002; Huston 2002; Guisan et al. 2006). This is unfortunate because many theories in ecology can help guide the model building process, which may not only improve model predictions but may also provide greater inference regarding habitat quality in heterogeneous landscapes.

Our objectives are several-fold. First, we synthesize important, although often overlooked, assumptions and predictions from ecological theory, focusing on habitat selection and metapopulation theory, for variation in species occurrence over time. Second, we describe how these perspectives can be incorporated into models of species distribution with data on the detection/non-detection of species across environmental gradients. Third, as a case study we apply these models to evaluate model fit and predictive ability relative to more conventional modeling approaches using a relatively large-scale and long-term data set of bird populations inhabiting forests across Montana and northern Idaho, USA.

5.2 Management challenge, ecological theory, and statistical framework

Currently, a major limitation of species distribution models and their implementation for management is their lack of dynamic predictions; i.e., models are temporally static in the absence of perturbations of environmental variables (Guisan and Zimmermann 2000; Castellón and Sieving 2006), even though we know that patterns of habitat use vary with population size and other factors (Gill et al. 2001; Sergio and Newton 2003). While other, more mechanistic, modeling approaches exist that capture temporal dynamics (Lischke et al. 2007), such models typically contain numerous parameters that are difficult to parameterize for more than a few species across spatial scales relevant for management purposes. Incorporation of key dynamics into species distribution models could thus be very useful, since a major strength of species distribution models is the ability to fit models to many

species across large spatial scales with relatively sparse data (Guisan and Thuiller 2005; Elith et al. 2006).

A second major limitation is that interpreting habitat quality from patterns of density or occurrence can be misleading (Van Horne 1983). Incorporating deterministic aspects of temporal dynamics into species distribution models may provide one way to address these issues by explicitly focusing on consistency of occupancy over time. To guide our perspective on temporal dynamics, we draw from both habitat selection and metapopulation theory.

5.2.1 Perspectives from habitat selection and metapopulation theory

Habitat selection and metapopulation theory both make general assumptions and predictions regarding occupancy and population persistence within habitats that are relevant to species distribution models. All common ecological theories on habitat selection to date—the ideal free, ideal despotic, and ideal pre-emptive distributions (Fretwell and Lucas 1970; Pulliam and Danielson 1991)—make two key assumptions relevant to developing species distribution models. First, animals are ideal. That is, animals can identify the best quality habitats available, in terms of fitness rewards. Even when animals behave less than ideally, as long as they can gain reliable information about habitat quality from sampling, settlement will approach what is expected of ideal behavior (Pulliam and Danielson 1991; Tyler and Hargrove 1997). This assumption is relevant to species distribution models because it suggests that occupancy can be a measure of habitat quality (Sergio and Newton 2003), as is often implicitly assumed, but it also leads to predictions relevant to temporal variation in occupancy (see below). Thus, while the ideal despotic and pre-emptive distributions assume territoriality/aggressive behaviors and pre-emption by early settlers can influence settlement of some individuals whereas the ideal free distribution assumes individuals are free to settle anywhere, each theory assumes that animals are generally ideal (Fretwell and Lucas 1970; Pulliam and Danielson 1991).

Second, habitat selection theory assumes that the range of habitats used by any given species will vary with population density, where low-quality sites are more likely to be used as population density increases (Fig. 5.1). Because average fitness will decrease with increasing population size, this phenomenon can regulate populations (Brown 1969; cf. Rodenhouse et al. 1997). Density-dependent habitat use has been documented in a wide variety of taxa, from fish (Fraser and Sise 1980) to birds (Gill et al. 2001). Note that density-dependent habitat selection can occur at a variety of spatio-temporal scales, from arrival order of migrants within a breeding season (Shochat et al. 2005), to variation in habitat use among years where population size varies (Gill et al. 2001). Such density-dependent habitat use can have consequences for managing animal populations and can impose

problems in diagnosing causes of population decline (Norris 2004). Variation in the range of habitats used as a function of population density makes conventional habitat use models limited (Wiens 1989). This limitation arises because habitat models built using information from a time of high population density will not be able to distinguish good- from poor-quality habitats, while models built using information from a time of low population density will fail to recognize the range of potentially suitable habitat types. Because the range of habitats used changes with population size (O'Connor 1981; Chamberlain and Fuller 1999; Gill et al. 2001; Sergio and Newton 2003) and land management strategies can influence population size (Newton 1998), accurate forecasting of the dynamics of animal communities will need to link density-dependent variation in habitat use with expected changes in population size that occur from landscape change.

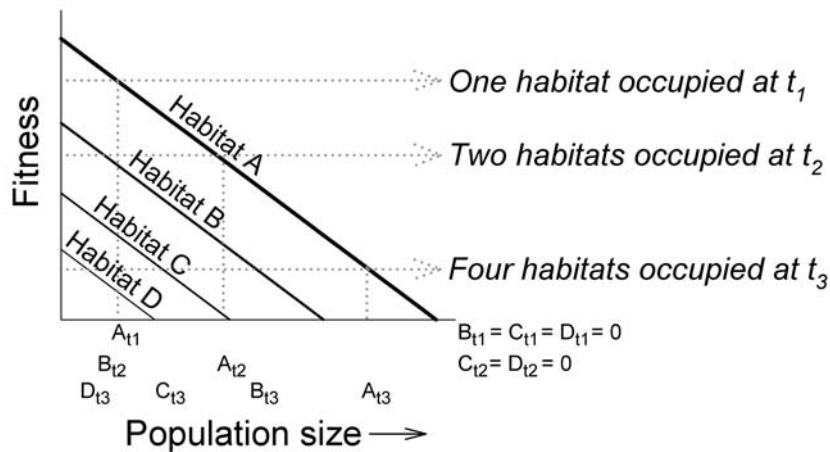


Fig. 5.1. Habitat selection theory generally assumes that as population size increases within a habitat, quality decreases. Organisms begin to use sub-prime habitat only after population size in the high-quality habitat grows to a point where the suitability therein is equal to the suitability in the next best habitat. Shown here is an example drawing from the ideal free distribution (Fretwell and Lucas 1970), where there are four different habitats that vary in quality with population size. Lines are arranged from highest (A) to lowest (D) quality. As population size increases over time, the number of habitats occupied increases. Thus, basic habitat selection theory predicts that observed species-environment relationships (specifically, the breadth of habitats occupied) should change with fluctuations in population size.

Metapopulation and habitat selection theory also make a related prediction that populations will be more likely to persist in habitats of high suitability (e.g., Pulliam 1988; Hames et al. 2001), and thus be occupied more consistently through time. This prediction follows from habitat selection and metapopulation theory for different reasons. Habitat selection theory suggests that low quality

habitat is occupied only at high population densities because animals behave ideally (see above), and temporal variation in habitat use will be greater in low-quality habitats.

In metapopulation theory, persistence in areas of high habitat quality is predicted based on variation in local colonization/extinction dynamics. Much of metapopulation theory uses metrics that reflect variation in local population size (e.g., patch area) or surrounding population size (propagule pressure; e.g., isolation) to predict variation in extinction/colonization dynamics. Metapopulation theory predicts that local extinctions should be less likely, and local colonizations more likely, with increased habitat suitability, all else being equal (Moilanen and Hanski 1998; Thomas et al. 2001; Fleishman et al. 2002; Bonte et al. 2003). Extinction is less likely because higher density populations should occur in areas of higher quality. Accordingly, Hanski and Ovaskainen (2000) suggested that information on habitat quality might actually be used to adjust patch area metrics to reflect changes in local population size that occur via variation in habitat quality (see also Moilanen and Hanski 1998). Colonizations could also be more likely in high-quality, but unoccupied, habitat when animals use some component of habitat quality in their searching (e.g., patch detection) or settlement decisions (Tyler and Hargrove 1997; Fletcher 2006). The net result is that temporal variation in occupancy should lead to high-quality habitat being more frequently occupied through time than low-quality habitat, all else being equal, and thus create a higher likelihood of persistence, than low-quality habitat.

5.2.2 Statistical framework

Incorporating these perspectives may not only improve predictions of species distribution models by providing dynamic predictions, but may also provide a more refined perspective on habitat quality by identifying areas that are consistently occupied (Sergio and Newton 2003). Yet, to date, there has been little attempt to incorporate these perspectives into model building on a wide suite of species at scales relevant to management strategies. Here, we incorporate these ideas into species distribution models that accommodate detection/non-detection (presence-absence) data. Specifically, we (1) incorporate temporal variation in occurrence via changes in population size, and (2) model local indices of persistence rather than patterns of occurrence alone. To do so, we use a generalized linear modeling (GLM) framework because GLMs can address spatio-temporal complexities (e.g., repeated measures, spatial autocorrelation) in data sets, which is essential for appropriately addressing temporal dynamics.

Conventional GLMs for detection/non-detection data come in the form of logistic regression models, where we specify a logit link function and a binomial error distribution. Expanding these models to accommodate multi-year data re-

quires addressing the potential for non-independence (i.e., repeated measures) across years, which can be accomplished either using generalized estimating equations (GEE) or by specifying random effects terms in GLMs. We refer to this model structure as a ‘conventional model’ throughout.

Extending conventional models to incorporate temporal variation in occurrence can be solved by either developing separate models for each time period considered and subsequently identifying if coefficients change consistently with changes in population size, or by including temporal indices of population size directly into multi-year models. Here we focus on the latter approach, because it provides a more parsimonious modeling solution that easily allows for translation into species distribution maps. By incorporating a temporal index of population size and interactions of this index with relevant environmental variables, models can test directly for variation in occurrence and whether observed habitat-relationships change with fluctuating population size.

Modeling indices of persistence can be accomplished using numerous approaches. Markov models that estimate local colonization/extinction events have seen widespread use in metapopulation ecology, but to link such models to predicting the occurrence of species across landscapes and/or regions requires the strong assumption of equilibrium (Moilanen 2000), and predictions represent an ‘equilibrium probability of occupancy’. Because species distribution models are often used to interpret changes in land use or to understand reasons for species declines (Guisan and Thuiller 2005), the assumption of equilibrium may limit the use of metapopulation models in many situations. Alternatively, estimates of the frequency of occurrence through time at a given location can provide another measure of persistence, where locations that are more frequently (consistently) occupied are more likely to persist. With this approach, the response variable is the proportion of time a sampling unit is occupied (e.g., 3 out of 5 years a species is present), and the explanatory variables are the average values through time. Note that this approach assumes the environment is relatively constant through time, which may not be appropriate in some situations. These kinds of data have been analyzed with logistic regression using binomial responses rather than binary responses (Dunford et al. 2002), linear regression (Johnson and Igl 2001) or ordinal logistic regression techniques (Hames et al. 2001). Using binomial responses in logistic regression assumes counts across years are independent Bernoulli trials, which may limit that approach in many cases where species show site fidelity or when there are other reasons to assumed non-independence across years. Linear regression is limited because it will not provide appropriate variance estimates (constrained to the 0-1 interval) and can make predictions beyond observable values (Guisan and Harrell 2000). Consequently, we focus here on ordinal logistic regression as a potential approach for directly modeling indices of persistence.

Ordinal logistic regression, in our case employing the cumulative logit model (also known as the proportion odds model), is an extension of logistic regression where the response variable is a series of ranked values (Allison 1999). For modeling occupancy through time, the dependent variables are categories of the frequency of occurrence. Note that the model makes no assumptions regarding the distances between observed categories (Allison 1999), such that it can accom-

modate datasets where time between sampling periods varies. The cumulative logit model structure is formalized as:

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \alpha_j + \beta_1 x_1 + \dots + \beta_k x_k, \quad j = 1, \dots, J - 1, \quad (5.1)$$

where the response variable represents the log odds of a point being occupied for Y out of j years, J is the total number of categories (time periods sampled), α_j 's are the intercepts, β 's are the habitat coefficients, and k is the number of explanatory variables. In this model, the coefficients of predictor variables are the same for the ordered categories, but the intercepts for each category are different, resulting in J equations.

This approach may be useful for understanding habitat occupancy and habitat quality based on the way that this model is interpreted. As the name implies, a cumulative logit analysis provides the probability of being in an ordered category or lower. Consider a data set where sites were sampled on an annual basis for five years across an environmental gradient. A cumulative logit analysis would provide predicted estimates of the frequency of occurrence through time as a function of the environmental gradient; however, a 60% probability of occupancy at a location four out of five years would actually be interpreted as a 60% chance that the species would occupy the location four or fewer years out of a total of five years. The model can be rearranged to give the likelihood of a minimum probability of occurrence, which could be useful for conservation purposes by providing a close analog to a measure of persistence. Here, a 60% probability of occupancy at a location four out of five years would be interpreted as a 60% chance that the species would occupy the location a minimum of four years out of a total of five years. A model with this structure is:

$$\log\left(\frac{P(Y \geq j)}{1 - P(Y \geq j)}\right) = \alpha_j + \beta_1 x_1 + \dots + \beta_k x_k, \quad j = 1, \dots, J - 1 \quad (5.2)$$

5.2.3 Comparing models

To determine whether the incorporation of these aspects of theory improves model performance, we need ways to compare results among models. Models that incorporate temporal indices of population size can be compared with conventional model selection criteria and validation statistics (Fielding and Bell 1997); however, cumulative logit models have a different currency. Because there are a variety of statistics developed for estimating the robustness of binary models, an intuitive approach would be to convert putative persistence models to a similar scale as conventional models. There are at least two ways to do this for frequency of occurrence data. First, ordinal responses can be recoded to presence-absence predic-

tions, using some sort of cutoff (Guisan and Harrell 2000), similar to cutoffs used for binary models. Second, ordinal responses can be back-transformed to estimate the average probability of occurrence, p_{avg} , at a location for any given year, based on the number of years surveyed:

$$f_0 = (1 - p_{\text{avg}})(1 - p_{\text{avg}}) \dots = (1 - p_{\text{avg}})^t$$
$$p_{\text{avg}} = 1 - (f_0)^{1/t}, \quad (5.3)$$

where t is the number of years surveyed and f_0 is the probability that a location is predicted to be occupied in none of the years (from equation 5.1). Back-transformation allows for direct comparisons with evaluation data sets using the same approaches as for binary data.

5.3 Model and model validation techniques

5.3.1 Modeling database

As a case study, we apply the above modeling approaches to a large database that incorporates multiple years of data across broad spatial scales. The USFS Northern Region Landbird Monitoring Program (NRLMP) involves monitoring all diurnal landbird species that can be detected through a single methodology throughout Montana and northern Idaho (Hutto and Young 2002). The NRLMP surveys birds on a series of transects located using a geographically stratified random sampling design on Forest Service lands across the Northern Region and on lands of several cooperating agencies (Fig. 5.2). The NRLMP includes approximately 350 transects on USFS lands (30 on each National Forest unit) and 200 transects on other lands. Each transect consists of 10 permanently marked points, at which 10-min bird counts are conducted according to a standard point-count protocol (Hutto et al. 1986; Ralph et al. 1995). We truncated detections to include only those ≤ 100 m from the center points.

For the purposes of incorporating predictions from ecological theory, we focus on transects that occurred over a subset of the region with adequate GIS-based information (see below; Fig. 5.2). In this area, we divided the dataset into transects that had been surveyed for at least six years (1995, 1996, 1998, 2000, 2002, 2004) to develop models (6768 point \times year combinations; Fig. 5.2), and used the remaining transects that had been surveyed for 1-5 years to test models (6256 point \times year combinations). We chose this partitioning approach for model development and testing to ensure that we captured potential temporal dynamics

in model development and to test whether the inclusion of population dynamics improved predictions of occurrence.

The NRLMP generates long-term data on bird distribution in association with detailed local-level and GIS-based landscape-level vegetation information. Here we focus solely on GIS-based measures that allow for predictive modeling across the study area (see below). However, the local field measures were used to identify the points with substantial changes in habitat during the monitoring time period. We used this record of habitat changes over time to remove survey points ($n = 296$) that changed substantially in vegetation structure (usually due to timber harvesting activity), because persistence models are less interpretable if environmental changes occur over the time period used for model development.

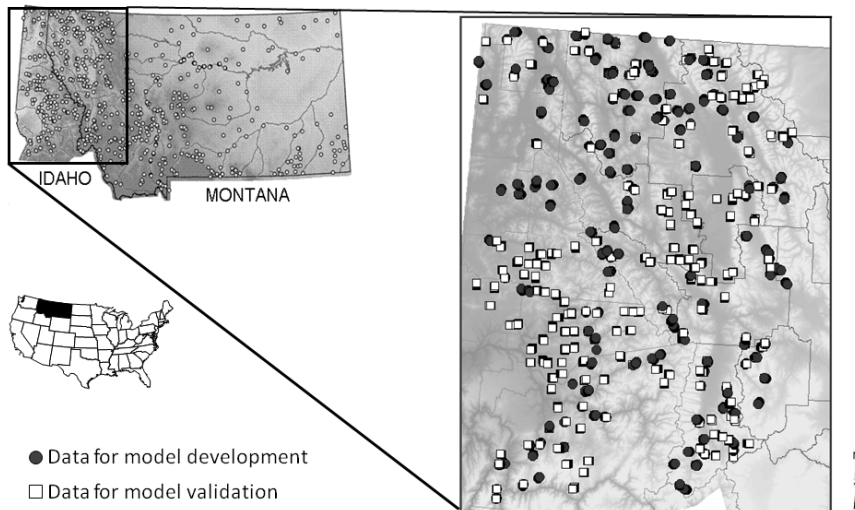


Fig. 5.2. The database used for modeling species distribution was generated from the USFS Northern Region Landbird Monitoring Program, which spans Montana and northern Idaho (inset). We focused modeling efforts on the western part of this area, where relatively accurate GIS-based environmental information was available. In this region, we divided the dataset into model-building and test sets, with model-building data being drawn from all transects that had 6 years of temporal replication. The remaining data were used for model testing.

5.3.2 Focal species

We contrasted models developed for bird species that vary in their use of different habitats and in their variation in occurrence across years. We expected that explicitly incorporating temporal variation in population size into models should improve model performance for species known to use multiple habitats and that show greater temporal variation in occurrence, but should be less useful for those species that are rigid in their associations and less variable in occurrence over time. Similarly, we expected that using measures of persistence would be more useful for species that varied substantially across years in their occurrence than for species with less temporal variation.

To choose focal species, we considered those species that occurred on at least 5% of points throughout the study period. From this list, we calculated the coefficient of variation (CV) in occurrence across the six years available in the database, and we calculated vegetation breadth (Colwell and Futuyma 1971), based on local habitat categories identified in Hutto and Young (1999). We then chose species that represent extremes in their variation in occupancy through time and in their variation in habitats used. Specifically, we ranked the upper and lower quartiles for CV and habitat breadth and factorially selected the four species that represent these extremes (Table 5.1): ruby-crowned kinglet (*Regulus calendula*), hermit thrush (*Catharus guttatus*), dark-eyed junco (*Junco hyemalis*), and dusky flycatcher (*Empidonax oberholseri*). Thus, we expected that persistence models should most improve predictions for the thrush and kinglet, which vary considerably across years in occurrence, with less improvement for the junco and flycatcher, and that temporal models to most improve models for the kinglet and provide least improvement for the flycatcher.

Table 5.1. Differences in temporal variability in occurrence across years and vegetation breadth (CV, vegetation breadth) for focal species used in species distribution modeling.

Coefficient of variation across years		
<i>Vegetation breadth</i>	<i>Low</i>	<i>High</i>
<i>Low</i>	Dusky Flycatcher (0.11, 5.96)	Hermit Thrush (0.42, 6.66)
<i>High</i>	Dark-eyed Junco (0.11, 11.75)	Ruby-crowned Kinglet (0.38, 8.56)

5.3.3 Environmental covariates

GIS-based vegetation measures were derived from a 15-m resolution, digital land-cover map developed by the USFS (Northern Region Vegetation Mapping Project, R1-VMP) based on Landsat TM imagery, aerial photo interpretation, and field data collection from 2001 (Brewer et al. 2004). This is the most accurate GIS database currently available for this region. We used three separate data layers from this database, general cover type (conifer forest, deciduous forest, shrub), successional stage (four tree-diameter categories [dbh]), and percent canopy cover (three categories). We also calculated elevation using a 30-m resolution digital elevation model (USGS 2002) and derived the mean annual precipitation from a Parameter Elevation Regressions on Independent Slopes Model (PRISM), which was based on climate data from 1961-1990 (PRISM Climate Group, Oregon State University, <http://www.prismclimate.org>, 2004). We used this GIS-based information to measure both local- (within 100 m of each point) and landscape-scale (within 1 km of each point) environmental variables. The 1 km landscape scale was chosen on the basis of other investigations in this region that showed strong correlations at this scale (Tewksbury et al. 2006). These variables capture the major biotic and abiotic gradients in the region relevant to avian distribution. For the four successional stage and three percent canopy cover categorical variables, we initially subjected variables to a principal component analysis (PCA), to derive new, independent variables that represent variation in dbh and percent cover information. For both PCAs, two new components were derived (explaining 92% of the variation in each analysis), one of which described a linear gradient in dbh/canopy cover, and the other described non-linearity of each variable (Table 5.2).

Table 5.2. Explanatory variables used in development of models.

Variable	Description
Physical/abiotic environment	
Elevation (km)*	Elevation taken from digital elevation model
Precipitation*	Mean annual precipitation (based on data from 1961-1990)
Latitude*	North-south gradient, in decimal degrees (dd.dddd°)
Longitude*	East-west gradient, in decimal degrees (dd.dddd°)
Local vegetation (100 m)	
CanopyPC1	Principal component reflecting moderate canopy cover
CanopyPC2	Principal component reflecting linear gradient in canopy
DbhPC1	Principal component reflecting linear gradient in dbh
DbhPC2	Principal component reflecting moderate dbh
Conifer*	Percent of conifer
Deciduous	Presence of broadleaf deciduous forest
Shrub	Percent of shrub-dominated areas; often clearcuts

Landscape structure (1km)	
Conifer Forest*	Percent of conifer
Deciduous forest	Percent of deciduous forest
Shrub	Percent of shrub-dominated areas; often clearcuts
Road density	Kilometers of roads within 1 km
Population	
Annual index	Annual index of population size (frequency of points occupied)
Other factors	
Time of year*	Julian date to accommodate potential time of year bias

*Nonlinear effects considered

5.3.4 Model development

We compared conventional modeling approaches involving binary response data with approaches that explicitly incorporate temporal dynamics. To develop models, we used GLMs, considering repeated samples (points) over time as repeated measures using GEE in conventional models. While other approaches exist and are useful for modeling species distributions (e.g., Elith et al. 2006), dealing with potential biases (e.g., detection probability; MacKenzie et al. 2002), and uncertainty in model predictions (e.g., Karanth et al. 2004), we chose this framework to allow for standardized and seamless relative comparisons among the model types described above. Overall, we considered three general types of GLMs: conventional logistic regression (hereafter referred to as conventional models), conventional models with temporal indices of population size (hereafter referred to as temporal models), and cumulative logit models (hereafter referred to as persistence models). Each model type considered the same explanatory variables, except that temporal models also considered a temporal index in population size (i.e., frequency of points occupied each year across the region) and its pair-wise interactions with environmental variables.

We initially screened environmental explanatory variables for strong correlations ($r > 0.7$) and subsequently removed highly correlated variables. Then, we used univariate screening of linear and non-linear (quadratic) relationships of species occurrence and environmental variables (Fletcher and Hutto 2008). Based on this screening, for further model analyses we included only explanatory variables that improved model fit over an intercept-only model, as judged by QAICc (Pan 2001), a model selection criterion for GEE models analogous to Akaike's Information Criteria (AIC; Burnham and Anderson 2002). Given this initial screening, we used a manual backward elimination procedure (see, e.g., Thogmartin et al. 2004), where we determined variables for removal based on the 95% confidence limits of the parameter estimates; we stopped removing variables when QAICc (or

AICc for cumulative logit models) was not lowered with the exclusion of variables (i.e., removal of more variables did not improve model fit).

Using these reduced models, we then tested for predictive accuracy using the independent data set mentioned above. Predictive accuracy was evaluated by calculating the area under the ROC curve (AUC), the Kappa statistic, and the correct classification rate (Fielding and Bell 1997). AUC is a frequently used index that ranges between 0.0-1 and does not require threshold cutoffs to be determined. Kappa measures the proportion of correctly predicted points after the probability of chance agreement has been removed. The correct classification rate, also known as 'overall prediction success' (OPS; Liu et al. 2005) is the proportion of detections/non-detections correctly classified in the test data. For the Kappa statistic and the correct classification rate, we used a threshold cutoff value based on the prevalence of each species in the model building data set (Liu et al. 2005).

5.3.5 Model results

Overall, in the data set used for model building, kinglets occurred on 19.4% of points, flycatchers on 10.3%, thrushes on 6.4% and juncos on 48.3% of points. Models that incorporated temporal indices of population size fit the data for all species much better than conventional models that did not include temporal indices (Table 5.3). The relative model improvement, as evidenced by change in QAICc, was greatest for the kinglet and least for the flycatcher, consistent with our expectations (Table 5.3). In general, density-dependent relationships identified from temporal models suggested that relationships in low population size years tended to be weak, whereas relationships were stronger in high population size years (Fig. 5.3). Note that because persistence models were based on a different type of response variable, model selection criteria could not be compared directly with criteria for conventional and temporal models. Nonetheless, significant predictor variables explaining distributions for each model type tended to be the same, with the main exception being that the addition of the temporal index in population size altered some relationships with occurrence (see, e.g., Fig. 5.3).

Predictive accuracy measures suggested that, overall, temporal models tended to provide modest improvements over conventional models for all species (Fig. 5.4). As we expected on the basis of differences in the temporal variation in their occupancy rates and habitats used among species (Table 5.1), improvements based on AUC were greatest for the kinglet and least for the flycatcher. Persistence models improved accuracy predictions for hermit thrush only, based on the Kappa statistic and OPS statistic. However, OPS was also greater for persistence models for the kinglet and flycatcher relative to conventional and temporal models, even though Kappa and AUC statistics did not indicate improvement of persistence models for those species.

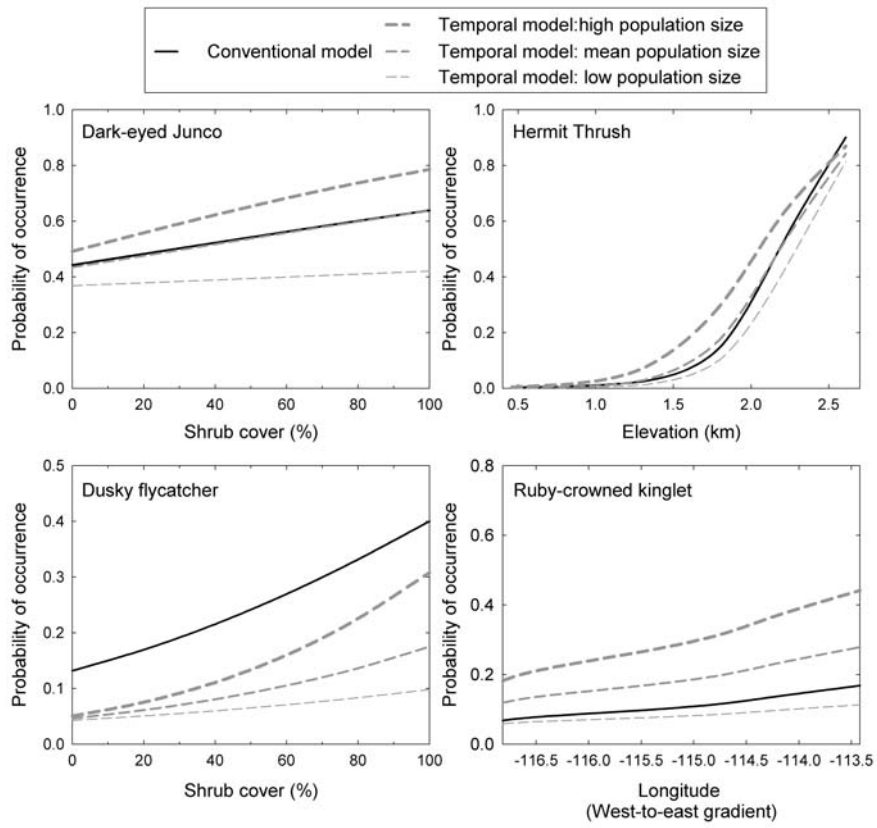


Fig. 5.3. Some examples of how temporal variation alters model predictions for each species relative to conventional models that do not incorporate such variation. Shown are partial relationships that exhibited significant interactions with population size.

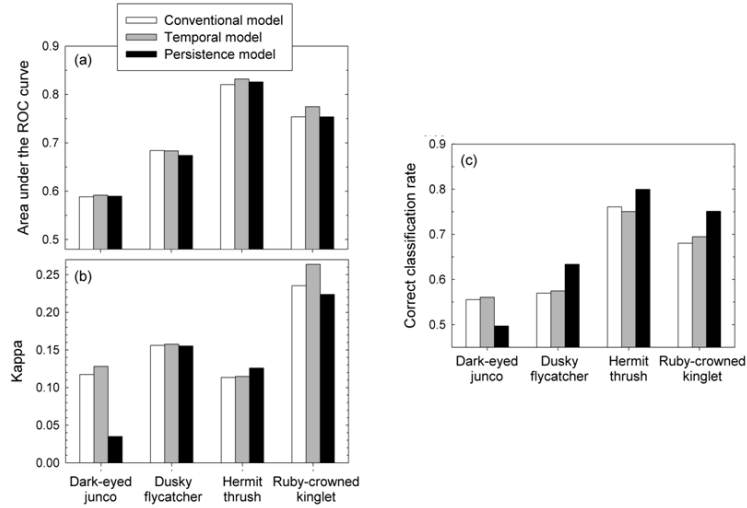


Fig. 5.4. Assessment indices of model performance against test data, using (a) area under the ROC curve, (b) the Kappa statistic, and (c) the correct classification rate (or overall predictive success; OPS). For Kappa and the correct classification rate, we used prevalence values of each species to determine appropriate threshold cutoff values.

Table 5.3. Model fit, based on QAICc, of the best model that includes and excludes temporal variation in population size.

Species/model	K	QAICc	Δ QAICc
Dark-eyed junco			
Best conventional model	10	9075.8	52.9
Best temporal model	14	9022.9	0.0
Dusky flycatcher			
Best conventional model	12	3870.4	10.9
Best temporal model	19	3859.5	0.0
Hermit thrush			
Best conventional model	11	2581.7	72.9
Best temporal model	12	2508.8	0.0
Ruby-crowned kinglet			
Best conventional model	12	5558.6	223.8
Best temporal model	16	5334.8	0.0

5.4 Past, current, and future applications

Aspects of ecological theory are relevant to, though often ignored in, the development of species distribution models. In the past development of models, when theory has been appreciated, only perspectives from niche theory have been emphasized in modeling approaches (Austin 2002). Niche theory reminds us, for example, that species-environment relationships can often be non-linear (Austin 2002, 2007). Here we show that additional perspectives gleaned from habitat selection and metapopulation theory can further improve species distribution modeling efforts, by emphasizing that species-environment relationships should vary in consistent ways as populations vary in size over time. Spatio-temporal variation in population size might help to explain past discrepancies in environmental relationships, such as variation in animal sensitivity to patch size (Johnson and Igl 2001).

The current application of these ideas provides an important step in linking perspectives from theory to species distribution models and evaluating whether such perspectives matter. These perspectives greatly improved model fit and temporal models consistently improved predictive ability, although predictive ability was only moderately increased for all species. There are some notable limitations of the NRLMP database that likely limit predictive ability at the point-count level, regardless of modeling method. For instance, counts were conducted only once over a 1.5 month window each year, and consequently Julian date was a strong predictor of occurrence for all species considered. Nonetheless, the fact that these theoretical perspectives generally increased model fit and predictive ability in the face of data limitations suggests that such perspectives should be considered in other investigations. In addition, Austin (2007) recently argued that tests of predictive ability of models are insufficient and that models must be grounded in ecological realism as well. We believe that by incorporating temporal perspectives in models, and thus grounding models in theoretical predictions that have been documented in a wide range of taxa (e.g., Fraser and Sise 1980; Gill et al. 2001; Sergio and Newton 2003), ecological realism may be greatly improved. Furthermore, while our focus has been on vagile organisms, these perspectives are also relevant for species distribution modeling of more sessile organisms, such as plants. Numerous investigations on plant metapopulation dynamics highlight the need to incorporate temporal measures of persistence (e.g., Verheyen et al. 2004), and spillover effects that can occur in plants (Brudvig et al. 2009) are analogous to density-dependent habitat selection described above.

Future applications of these perspectives are potentially diverse. Appreciation of the variation in species-environment relationships with population size could be incorporated into other species distribution modeling approaches, including Bayesian approaches, regression trees, and various machine-learning techniques. Future applications should consider these perspectives during model development, and perhaps other approaches to modeling persistence and temporal variation in population size could further refine models. For example, while we focused on ordinal logistic regression for modeling persistence (cf. Hames et al. 2001), other approaches, such as Markov-based models, may be more flexible and

provide greater insight for interpreting variation in temporal dynamics (MacKenzie et al. 2003). Finally, our approach allows for models to become more dynamic, which could be very useful in future applications of species distribution models applied to landscape and global change. Models could be used, for instance, to directly predict how species distribution changes with ongoing habitat loss that invariably reduces population size, which could alter predictions for occupancy in remaining habitats relative to predictions from conventional models.

5.5 Data availability and suitability

Many investigations that sample species distribution have at least limited temporal replication that could capture some relevant dynamics, yet temporal variability is often treated as a nuisance rather than as an opportunity (but see Bissonette and Storch 2007). Our modeling efforts suggest that temporal replication could provide opportunities for developing better models of species distribution, such that investigations of species distribution might be better off repeating samples over time rather than sampling new locations each time period. This potential space-for-time tradeoff in sampling for species distribution models is an ongoing issue in monitoring designs and model development.

The utility of repeating samples over time at the expense of spatial replication will undoubtedly be greater for species that have greater temporal variation in occupancy, such as the Ruby-crowned Kinglet modeled here. Temporal replication may also be relatively more useful than spatial replication when models are being developed to interpret ongoing landscape change. However, if spatial extents of data collection do not capture relevant spatial variation in environmental and land use gradients, temporally adequate data may not be sufficient for developing useful models in the first place. Further analysis of the temporal dynamics described here and the tradeoff of capturing such dynamics relative to spatial sampling procedures would be informative.

Indices of temporal variation in population size can be gleaned from survey protocols used in developing models or through use of independent data sets. Here we chose to derive indices using the NRLMP database, because of the broad spatial and temporal extent of the data base. However, independent data sets, such as the ongoing Breeding Bird Survey in the U.S. and in parts of Canada, could also be integrated into modeling efforts by providing temporal indices of population size. In doing so, even when an investigation is completed, models could be useful into some point in the future by altering predictions based on population size indices obtained from ongoing monitoring programs.

As species distribution modeling continues to advance as a field, explicit consideration of ecological theory in sampling designs and in the modeling process should prove helpful for improving accuracy of predictions, increasing insights regarding resource quality, and interpreting landscape change. We contend that while ecological theories might not always be ‘right’, they can still be useful (cf. Box and Draper 1987).

18 Robert J. Fletcher, Jr.^{1*}, Jock S. Young², Richard L. Hutto², Anna Noson², and Christopher T. Rota^{1,3}

Acknowledgments

This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant #2006-55101-17158. The landbird database was created through support from USFS Northern Region (03-CR-11015600-019). We thank two anonymous reviewers who provided valuable comments on a previous version of this manuscript, which improved the ideas presented here.

References

- Allison PD (1999) Logistic regression using the SAS system. SAS Institute, Inc., Cary, NC
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *J Biogeogr.* 33:1677–1688
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157:101–118
- Austin M (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Model.* 200:1–19
- Bissonette JA, Storch I (eds) (2007) Temporal dimensions of landscape ecology: wildlife responses to variable resources. Springer, New York
- Bonte D, Lens L, Maelfait JP, Hoffmann M, Kuijken E (2003) Patch quality and connectivity influence spatial dynamics in a dune wolfspider. *Oecologia* 135:227–233
- Box GEP, Draper NR (1987) Empirical model-building and response surfaces. Wiley, New York
- Brewer CK, Berglund D, Barber JA, Bush R (2004) Northern region vegetation mapping project summary report and spatial datasets, version 42. Northern Region USFS
- Brown JL (1969) The buffer effect and productivity in tit populations. *Am. Nat.* 103:347–354
- Brudvig LA, Damschen EI, Tewksbury JJ, Haddad NM, Levey DJ (2009) Landscape connectivity promotes plant biodiversity spillover into non-target habitats. *Proc. Natl. Acad. Sci.* 106:9328–9332
- Burnham KP, Anderson DR (2002) Model selection and inference: a practical information-theoretic approach. Springer-Verlag, N. York
- Castellón TD, Sieving KE (2006) Landscape history, fragmentation, and patch occupancy: models for a forest bird with limited dispersal. *Ecol. Appl.* 16:2223–2234
- Chamberlain DE, Fuller RJ (1999) Density-dependent habitat distribution in birds: issues of scale, habitat definition and habitat availability. *J Avian Biol.* 30:427–436
- Colwell RK, Futuyma DJ (1971) On the measurement of niche breadth and overlap. *Ecol.* 52:567–576
- Dunford W, Burke DM, Nol E (2002) Assessing edge avoidance and area sensitivity of Red-eyed Vireos in southcentral Ontario. *Wilson Bull.* 114:79–86
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24:38–49
- Fleishman E, Ray C, Sjögren-Gulve P, Boggs CL, Murphy DD (2002) Assessing the roles of patch quality, area, and isolation in predicting metapopulation dynamics. *Conserv. Biol.* 16:706–716
- Fletcher RJ, Jr. (2006) Emergent properties of conspecific attraction in fragmented landscapes. *Am. Nat.* 168:207–219
- Fletcher RJ, Jr., Hutto RL (2008) Partitioning the multi-scale effects of human activity on the occurrence of riparian forest birds. *Landscape Ecol.* 23:727–739
- Fraser DF, Sise TE (1980) Observations on stream minnows in a patchy environment: a test of a theory of habitat distribution. *Ecol.* 61:790–797
- Fretwell SD, Lucas HL, Jr (1970) On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. *Acta Biotheo.* 19:16–36
- Gill JA, Norris K, Potts PM, Gunnarsson TG, Atkinson PW, Sutherland WJ (2001) The buffer effect and large-scale population regulation in migratory birds. *Nature* 412:436–438
- Guisan A, Harrell FE (2000) Ordinal response regression models in ecology. *J Veg. Sci.* 11:617–626

20 Robert J. Fletcher, Jr.1*, Jock S. Young2, Richard L. Hutto2, Anna Noson2, and Christopher T. Rota1,3

- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8:993–1009
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol. Model.* 135:147–186
- Hames RS, Rosenberg KV, Lowe JD, Dhondt AA (2001) Site reoccupation in fragmented landscapes: testing predictions of metapopulation theory. *J Anim. Ecol.* 70:182–190
- Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404:755–758
- Huston MA (2002) Introductory essay: critical issues for improving predictions. In: Scott JM, Heglund PJ, Morrison ML (eds) *Predicting species occurrences: issues of accuracy and scale.* Island Press, Washington DC
- Hutto RL, Young JS (1999) *Habitat relationships of landbirds in the Northern Region, USDA Forest Service. General Technical Report RMRS-GTR-32.* U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT
- Hutto RL, Young JS (2002) Regional landbird monitoring: perspectives from the Northern Rocky Mountains. *Wildl. Soc. Bull.* 30:738–750
- Hutto RL, Pletschet SM, Hendricks P (1986) A fixed-radius point count method for nonbreeding and breeding-season use. *Auk* 103:593–602
- Johnson DH, Igl LD (2001) Area requirements of grassland birds: A regional perspective. *Auk* 118:24–34
- Karanth KU, Nichols JD, Kumar NS, Link WA, Hines JE (2004) Tigers and their prey: predicting carnivore densities from prey abundance. *Proc. Natl. Acad. Sci.* 101:4854–4858
- Lischke H, Bolliger J, Seppelt R (2007) Dynamic spatio-temporal landscape models. In: Kienast F, Wildi O, Ghosh S (eds) *A changing world: challenges for landscape research.* Springer, Dordrecht
- Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecol.* 83:2248–2255
- MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB (2003) Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecol.* 84:2200–2207
- Moilanen A (2000) The equilibrium assumption in estimating the parameters of metapopulation models. *J Anim. Ecol.* 69:143–153
- Moilanen A, Hanski I (1998) Metapopulation dynamics: Effects of habitat quality and landscape structure. *Ecol.* 79:2503–2515
- Newton I (1998) *Population limitation in birds.* Academic Press, Inc., San Diego, CA
- Norris K (2004) Managing threatened species: the ecological toolbox, evolutionary theory and declining-population paradigm. *J Appl. Ecol.* 41:413–426
- O'Connor RJ (1981) Habitat correlates of bird distribution in British census plots. *Stud. Avian Biol.* 6:533–537
- Pan W (2001) Model selection in estimating equations. *Biom.* 57:529–534
- Pulliam HR (1988) Sources, sinks, and population regulation. *Am. Nat.* 132:652–661
- Pulliam HR, Danielson BJ (1991) Sources, sinks, and habitat selection: a landscape perspective on population dynamics. *Am. Nat.* 137:S50–S66
- Ralph CJ, Sauer JR, Droege S (1995) *Monitoring bird populations by point counts.* USDA Forest Service General Technical Report PSW-GTR 149.
- Rodenhouse NL, Sherry TW, Holmes RT (1997) Site-dependent regulation of population size: A new synthesis. *Ecol.* 78:2025–2042
- Sergio F, Newton I (2003) Occupancy as a measure of territory quality. *J Anim. Ecol.* 72:857–865

- Shochat E, Patten MA, Morris DW, Reinking DL, Wolfe DH, Sherrod SK (2005) Ecological traps in isodars: effects of tallgrass prairie management on bird nest success. *Oikos* 111:159–169
- Tewksbury JJ, Garner L, Garner S, Lloyd JD, Saab V, Martin TE (2006) Tests of landscape influence: Nest predation and brood parasitism in fragmented ecosystems. *Ecol.* 87:759–768
- Thogmartin WE, Sauer JR, Knutson MG (2004) A hierarchical spatial model of avian abundance with application to Cerulean Warblers. *Ecol. Appl.* 14:1766–1779
- Thomas JA, Bourn NAD, Clarke RT, Steward KE, Simcox DJ, Pearman GS, Curtis R, Goodger B (2001) The quality and isolation of habitat patches both determine where butterflies persist in fragmented landscapes. *Proc. R Soc. Lond. Ser. B* 268:1791–1796
- Tyler JA, Hargrove WW (1997) Predicting spatial distribution of foragers over large resource landscapes: A modeling analysis of the ideal free distribution. *Oikos* 79:376–386
- Van Horne B (1983) Density as a misleading indicator of habitat quality. *J Wildl. Manag.* 47:893–901
- Verheyen K, Vellend M, Van Calster H, Peterken G, Hermy M (2004) Metapopulation dynamics in changing landscapes: a new spatially realistic model for forest plants. *Ecol.* 85:3302–3312