Research article

The effects of habitat resolution on models of avian diversity and distributions: a comparison of two land-cover classifications

Joshua J. Lawler[1,2,7,*], Raymond. J. O'Connor[2], Carolyn T. Hunsaker[3], K. Bruce Jones[4], Thomas R. Loveland[5] and Denis White[6]

[1]*Margaret Chase Smith Center for Public Policy, University of Maine, Orono, Maine 04469, USA;* [2]*Department of Wildlife Ecology, University of Maine, Orono, Maine 04469-5755, USA;* [3]*USDA Forest Service, Pacific Southwest Research Station, Fresno, CA 93710, USA;* [4]*US Environmental Protection Agency, Las Vegas, NV 89193-3478, USA;* [5]*EROS Data Center, USGS, Sioux Falls, South Dakota 57198, USA;* [6]*US Environmental Protection Agency, Corvallis, Oregon 97333, USA;* [7]*Current address: US Environmental Protection Agency 200 SW 35th St. Corvallis, OR 97333, USA;* *Author for correspondence (e-mail: lawler.joshua@epa.gov)*

Abstract

Quantifying patterns is a key element of landscape analysis. One aspect of this quantification of particular importance to landscape ecologists is the classification of continuous variables to produce categorical variables such as land-cover type or elevation stratum. Although landscape ecologists are fully aware of the importance of spatial resolution in ecological investigations, the potential importance of the resolution of classifications has received little attention. Here we demonstrate the effects of using two different land-cover classifications to predict avian species richness and the occurrences of six individual species across the conterminous United States. We compared models built with a data set based on 14 coarsely resolved land-cover variables to models built with a data set based on 160 finely resolved land-cover variables. In general, comparable models built with the two data sets fit the data to similar degrees, but often produced strikingly different predictions in various parts of the country. By comparing the predictions made by pairs of models, we determined in which regions of the US predictions were most sensitive to differences in land-cover classification. In general, these sensitive areas were different for four of the individual species and for predictions of species richness, indicating that alternate classifications will have different effects in the analyses of different ecological phenomena and that these effects will likely vary geographically. Our results lead us to emphasize the importance of the resolution to which continuous variables are classified in the design of ecological studies.

Introduction

Landscape ecology has been defined, in part, as the study of the effects of pattern on process (Turner 1989). One of the key elements of landscape ecology, therefore, is the appropriate measurement of pattern. At the outset of any ecological investigation, the re-searcher must make a set of assumptions that guide the design of the study. Landscape ecologists are perhaps most familiar with the importance of spatial scale in the design of ecological studies. Matching the grain and the extent of the measurement of pattern with the scales at which the processes in question operate is often critical to understanding the system un-

der study (Wiens 1989). It is also well recognized that the patterns we choose to measure (Anderson and Shugart 1974) and the specific methods we use to quantify them (Hargis et al. 1997) are crucial aspects of any investigation.

One critical choice often made at the outset of a study involves the resolution with which classifications are made. This issue is of particular importance to landscape ecologists who are often forced to summarize complex landscapes by classifying them into a finite number of land-cover types or vegetation classes, but it applies to the classification of any continuous variable. Evidence for the importance of the resolution of classifications can be seen in the classic foliage-height diversity studies of MacArthur et al. (1966) and Recher (1969) which showed that avian species diversity was best explained by foliage height diversity classified into two layers in Puerto Rico, three layers in North America, three layers in Australia, and four layers in Panama. Absent a marked difference in forest heights in the four regions, these studies show that birds potentially discriminate vegetation structure more finely in Panama than in Puerto Rico.

Wiens (1989) pointed out that, for logistical reasons, expanding the extent of a study also usually entails enlarging the grain, resulting in a loss of spatial resolution. The same logistical constraints typically also entail a reduction in the resolution of a classification. In fact, many studies conducted at multiple spatial scales (e.g., Gutzwiller and Anderson 1987; Bergin 1992; Saab 1999) change the resolution at which the attributes are described as the scale changes. Analyzing habitat associations in this way assumes that animals perceive different types of patterns at different spatial scales–generally that they distinguish finer detail in habitat structure or heterogeneity at finer spatial scales (Hildén 1965; Hutto 1985). Testing for the pure effects of spatial scale must be done using a consistent set of variables and holding the resolution of any classification constant (e.g., O'Neill et al. 1991; Qi and Wu 1996; Steffan-Dewenter et al. 2002; Williams et al. 2002).

The objective of this study was to demonstrate the importance of the resolution of classifications–here the level of detail with which land-cover types were categorized–for building spatially extensive predictive models. Specifically, we asked whether the level of the classification of land-cover data affected the structure of models of bird species richness and individual bird species distributions across the contermi-

nous United States. We compared models built with data sets based on 14 land-cover classes to those built with data sets based on 160 land-cover classes. We document major effects in the magnitude and geographical distribution of model sensitivity to a change in the level of the land-cover classification used.

Methods

Modeling approach

We built models of avian species richness and individual species distributions using a set of predictor variables representing climate, topography, road density, land ownership, land-cover type, and landscape pattern, factors which have been found to be associated with avian distributions or species richness (e.g., O'Connor et al. 1996, 1999; Jones et al. 2000; Githaiga-Mwicigi et al. 2002). We built two models predicting species richness and two models each for predicting the individual distributions of six bird species. All models used the same climate, topography, road-density and land-ownership data, but for each of the seven response variables (species richness and the incidence of each of six species), the two models used a different set of land-cover and landscape-pattern data. One model used coarsely discriminated land-cover data consisting of 14 classes (hereafter coarse-cover models) and a second used finely discriminated land-cover data consisting of 160 classes (hereafter fine-cover models). A sampling grid of 12,518 hexagonal cells covering the conterminous United States served as the sampling frame for our models (White et al. 1992). Each hexagon had an area of approximately 640 km^2 and a center-to-center spacing of about 27 km. Each of the variables used in the models applied to the area of a hexagon. A more detailed description of the data sets used in the present study is provided by O'Connor et al. (1996).

Bird species richness

Data on avian species richness were obtained from the Breeding Bird Survey (BBS), an extensive roadside survey program conducted by the United States Geological Survey's (USGS) Patuxent Wildlife Research Center and the Canadian Wildlife Service. The program maintains 4100 census routes, approximately 3000 of which are surveyed each summer. The gen-

eral locations of routes are distributed randomly first within 1-degree blocks of latitude and longitude across the US and Canada and secondarily within BBS-defined physiographic regions within these blocks. Each route is 39.4 km long, follows secondary roads, and consists of 50 point counts spaced 0.8 km apart. All birds either heard or seen within 0.4 km of each point are recorded over a 3-min period. The routes are surveyed once a year between June and July, depending on the latitude of the route. Because the BBS is a roadside survey conducted during daylight hours, birds that avoid roads, that are nocturnal, or that are otherwise undetectable with the above survey methods are under-represented (Robbins et al. 1986) and birds that are attracted to roadsides (e.g., kingbirds) are likely to be over-represented.

Because tallies of species richness on a given route increased with the number of years the route was surveyed, we used a subset of 1189 BBS routes consisting of those routes in the conterminous US that had data for at least 7 years from 1981 to 1990. Having at least 7 years of data allowed us to reliably compute a standardized estimate for the 10-year species tally, an effective estimate of long-term richness on the route (Jones and O'Connor, unpublished). We chose the years between 1981 and 1990 to best coincide with the 1990 Advanced Very High Resolution Radiometer (AVHRR) data from which the land-cover classifications and land-cover pattern metrics were generated.

Individual species occurrences

We used BBS data from the same set of routes and the same 10-year period to model the distributions of six individual species. By selecting a set of birds with geographic ranges covering relatively large potions of the United States, we were able to investigate the geographic aspects of the effects of the two land-cover classifications. We chose relatively common passerines so that we would have adequate data for building the models. The six species we chose (House Wren *Troglodytes aedon*, Savannah Sparrow *Passerculus sandwichensis*, Pine Siskin *Carduelis pinus*, Ovenbird *Seiurus aurocapillus*, Red-eyed Vireo *Vireo olivaceus*, and Black-capped Chickadee *Poecile atricapilla*) represented six different families. Our response variables for these species consisted of a measure of incidence calculated as the proportion of years from 1981 to 1990 that a species was recorded on a route.

Climate, topography, road density, and land ownership

We obtained climate data from the Historical Climate Network database (1996). Climate variables included the mean, the maximum, and the minimum levels of precipitation, temperature for both July and January, and seasonality (the difference between July and January temperatures). The elevation correction method of Marks (1990) was used to model temperature data to 1-km resolution. The precipitation data, which had originally been modeled to 10-km resolution by Daly et al. (1994), were likewise interpolated to 1 km with a linear model.

The topographic data used in the models were obtained from the USGS Digital Elevation Models (DEM) and included both elevation (mean, maximum, minimum, and the range of elevation for each hexagon) and a quantification of rivers (the length of large, perennial, intermittent, and braided rivers, respectively in each hexagon). Other variables included were road density (km of highway and km of secondary roads) and the proportion of the hexagon on federal land.

Land-cover type and landscape pattern

We used two sets of land-cover data and corresponding landscape-pattern data. We used the Loveland et al. (1991, 1995) classification of 1.1-km resolution AVHRR data from the National Oceanic and Atmospheric Administration (NOAA) operational meteorological satellite program. Using a maximum vegetation index compositing technique, Loveland et al. classified the vegetation in the conterminous US into 159 distinct land-cover types. O'Connor et al. (1996) added an urban class to this data set using the Digital Chart of the World (Danko 1992). For the detailed habitat descriptor data set we used these 160 land-cover classes. For the coarse habitat descriptor data set we aggregated these 160 classes to a smaller set of 14 land-cover types that approximated those described by Anderson et al. (1976) as level II categories (Table 1). For example, this aggregation involved combining several detailed land-cover classes, such as those defined as 'Ponderosa Pine and Lodgepole Pine Forest' and 'Douglas Fir, Sitka Spruce, and Western Hemlock Forest' into one general class labeled 'Conifer Forest'. A detailed description of the finely resolved land-cover classification can be found in Loveland et al. (1991). For both the coarse and the

Table 1. Aggregation of 160 finely resolved land-cover classes into 14 coarsely resolved classes. The numbers in the right-hand column correspond to the 159 classes defined by Loveland et al. (1991) and an additional urban class.

Coarsely resolved land-cover classes	Finely resolved land-cover classes
Crops and pastures	1-34
Grasslands and crops	35-39
Woodlands and crops	40-54
Grasslands	55-65, 86-89
Shrub-dominated rangeland	66-71
Grass-shrub rangeland	72-85
Deciduous forest	90-97
Coniferous forest	98-132
Mixed forest	133-148
Water	149
Coastal Wetlands	150-154
Barren	155
Alpine tundra	156-159
Urban areas	160

detailed land-cover type data sets, the distribution of 1-km^2 pixels was used to calculate the fraction of each hexagon covered by each land-cover type.

We used several metrics to represent the landscape patterns associated with both sets of land-cover type variables (160 cover-types, and 14 cover-types). For each hexagon, individual patches were identified using the techniques described in Timmins and Hunsaker (1995). We calculated the average patch-size for each land-cover class, as well as the scaled mean patch size and the scaled mean patch perimeter for each land-cover class. In addition, we calculated a series of 24 landscape pattern metrics such as juxtaposition index, fractal dimension, and Simpson's diversity index (Hunsaker et al. 1994).

Statistical models and comparisons

We used regression tree analyses (Breiman et al. 1984; Clark and Pregibon 1992; Venables and Ripley 1994) to build all 14 models. Regression trees have several advantages over the more traditional, equivalent parametric analysis techniques (e.g., linear, logistic, and polytomous regression) (De'ath and Fabricius 2000). Regression tree analysis is a non-parametric method (and thus less dependent on the distribution of the data) and is particularly good at modeling non-additive interactions between variables (Clark and Pregibon 1992). The ability to model such complex interactions was important in a large-scale study such as ours, in which the relationships between variables might be different in different parts

of the United States. Regression trees, and their categorical counterpart, classification trees, have been used to successfully model species richness patterns (O'Connor et al 1996; Rathert et al. 1999), as well as individual species distributions (Lawler and Edwards 2002; Olden and Jackson 2002; Iverson and Prasad 1998).

Regression trees work by way of recursive binary partitioning of data into increasingly homogenous groups with respect to the response variable. First, the one explanatory variable and the one splitting threshold along the distribution of that variable that best splits the data into two groups is chosen. This is accomplished by trying all possible variables and all possible splitting thresholds until the ones that create the two most homogenous groups of data with respect to the response variable are chosen. Each of the new groups is then split in the same fashion until each group is indivisible (i.e., it contains one observation or multiple observations at a single tied value), or until some minimum criteria are met. The resulting model is a tree-like structure consisting of a series of nodes (see Figure 1). The nodes represent points at which splits were made or splitting was stopped. Regression tree models are deliberately over-fit and then pruned back on the basis of some form of cross-validation (Miller 1994). We reduced tree models using a ten-fold cross-validated pruning technique described in Venables and Ripley (1994).

All models were constructed using the 1189 hexagons for which reliable survey data existed for the years in question. We assessed model performance
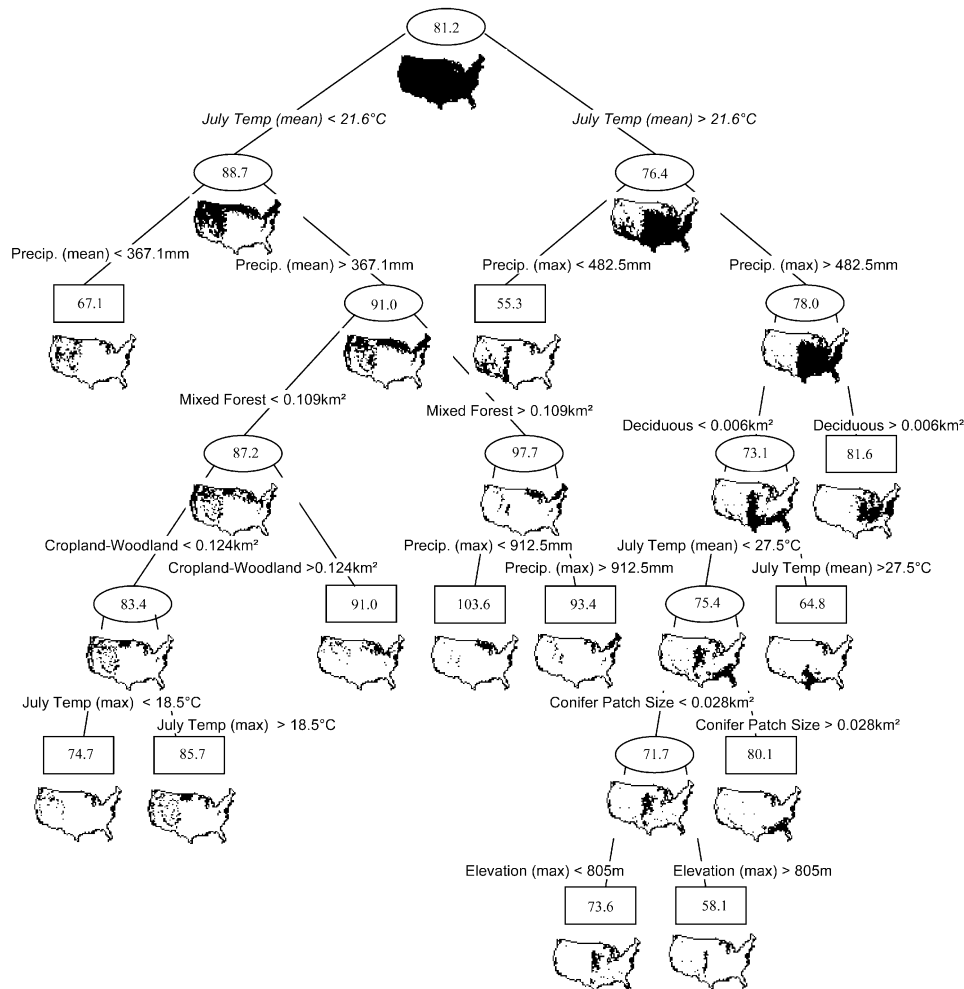
Figure 1. Regression tree model of avian species richness across the United States using coarsely discriminated land-cover variables (coarse-cover model). The numbers in the ovals (intermediate nodes) and rectangles (terminal nodes) represent the average number of bird species identified in BBS routes in hexagons represented by that node. The variables and the values used in each split are located along the branches of the tree at each split. The area of the United States defined by each node is depicted in black on the map below the node.

using cross-validation. We built models using a randomly selected 90% of the data and then applied the models to the remaining 10% and calculated the percentage of the deviance in the test data set explained by the model. We averaged the results of 100 such cross-validation analyses for each of the richness and individual species distribution models. After assessing performance, we used the models to make predictions for all 12,518 hexagons in the US. Applying the tree models predictively entailed plugging each new data point into the model and determining into which of the 'terminal nodes' it was classified. Each group of data in an end node was then mapped to demonstrate its spatial distribution (Figure 1).

The data sets used in many macroecological studies often exhibit attributes such as spatial autocorrelation that can be problematic for certain statistical analyses by producing biased estimators and hence affecting statistical hypothesis tests. It is important to note, however, that spatial autocorrelation is not always problematic and is often explained by real relationships between spatially structured explanatory and response variables (Diniz-Filho et al. 2003). Due to the spatial autocorrelation inherent in our data sets, we have restricted our statistical analyses to measures of the deviance explained by the regression trees and the individual variables (measures akin to R^2 values in a linear regression model). We do not emphasize

the relative importance of particular variables within a model, but instead concentrate on the differences in the predictions of pairs of models.

We investigated the geographic sensitivity of continental predictions of species richness and individual species distributions to the level of detail of land-cover classifications. The degree to which model predictions differed between the coarse-cover and the fine-cover models provided one measure of the importance of the detail of the classifications to the modeling of continental patterns. We determined the degree to which different geographic regions were sensitive to the detail of classification by analyzing the degree of spatial overlap of the predicted groups of hexagons produced by each pair of models. We used an index based on one developed by Dufrêne and Legendre (1997),

$$I = 100\left(1 - \left(\frac{N_{AB}}{N_A}\right)\left(\frac{N_{AB}}{N_B}\right)\right)$$

where N_A and N_B are the number of hexagons in a region (an end node) predicted by model A and model B respectively, and N_{AB} is the number of hexagons that the two regions have in common. The index I takes values from 0 (complete overlap) to 100 (no overlap). Perfect one-to-one correspondence of a cluster of hexagons across models would yield an index value of 0.

Results

Species richness

The coarse-cover model of species richness used a total of nine variables, including various aspects of climate, elevation, land-cover, and landscape pattern and classified the data into 12 groups (Figure 1). These groups are represented in Figure 1 by terminal nodes (rectangles) containing mean species richness values for that set of hexagons. The internal nodes (ovals) represent groups of data that were subsequently split into smaller groups. The ovals also display the mean species richness of the hexagons they represent. The geographic locations of the hexagons in each group are represented on small maps below each node.

The coarse-cover model first split the data into two groups, those hexagons with average July tempera-

tures greater and less than 21.6 °C (Figure 1). Note that the hexagons in areas with lower mean July temperatures had slightly higher (88.7 vs. 76.4) average species richness, and were in the northern and western portions of the US. The model then divided these two groups of data using mean precipitation and maximum precipitation for hexagons with lower and higher average July temperatures, respectively. Drier areas (those with lower average July temperatures and lower mean precipitation and those with higher July temperatures and lower maximum precipitation) needed no other variables to predict bird species richness. These drier hexagons were found in the highest two terminal nodes of the regression tree and had relatively low average species richness values (67.1 and 55.3). Wetter hexagons required additional variables to explain bird species richness and the tree proliferated further for these hexagons. The coarse-cover model explained 47% of the deviance in the data. Climate variables accounted for most of this deviance (36% of the total), whereas land-cover and landscape pattern accounted for 10% of the deviance. Elevation explained the remaining 1%.

The fine-cover model similarly divided the data into 12 groups using climate, land-cover, and landscape pattern but not elevation data (Figure 2). A striking similarity in two models was the large influence of the climate variables. The first three splits in both models were identical and were based on July temperatures and precipitation levels. Overall, the fit of the fine-cover model (deviance explained = 46%) was not very different from that of the coarse-cover model. The cross-validation analyses confirmed this similarity in performance (Table 2). As in the coarse-cover model, climate played a major role in the fine-cover model, accounting for 37% of the deviance.

Despite the similarities in model structure due to climate variables, the models differed markedly as to the land-cover variables used to classify hexagons into different groups. For example, the group of hexagons represented by the right-most internal node in both trees (mean species richness = 78.0) were classified quite differently by the two models (Figure 1, Figure 2). The coarse-cover model used the area of deciduous forest, mean July temperature, the size of conifer patches, and elevation, whereas the fine-cover model used mean July temperature, the area of savanna, and the size of patches of southeastern deciduous forest. With one or two exceptions, the corresponding geographic areas delineated by these splits in each of the models were different.
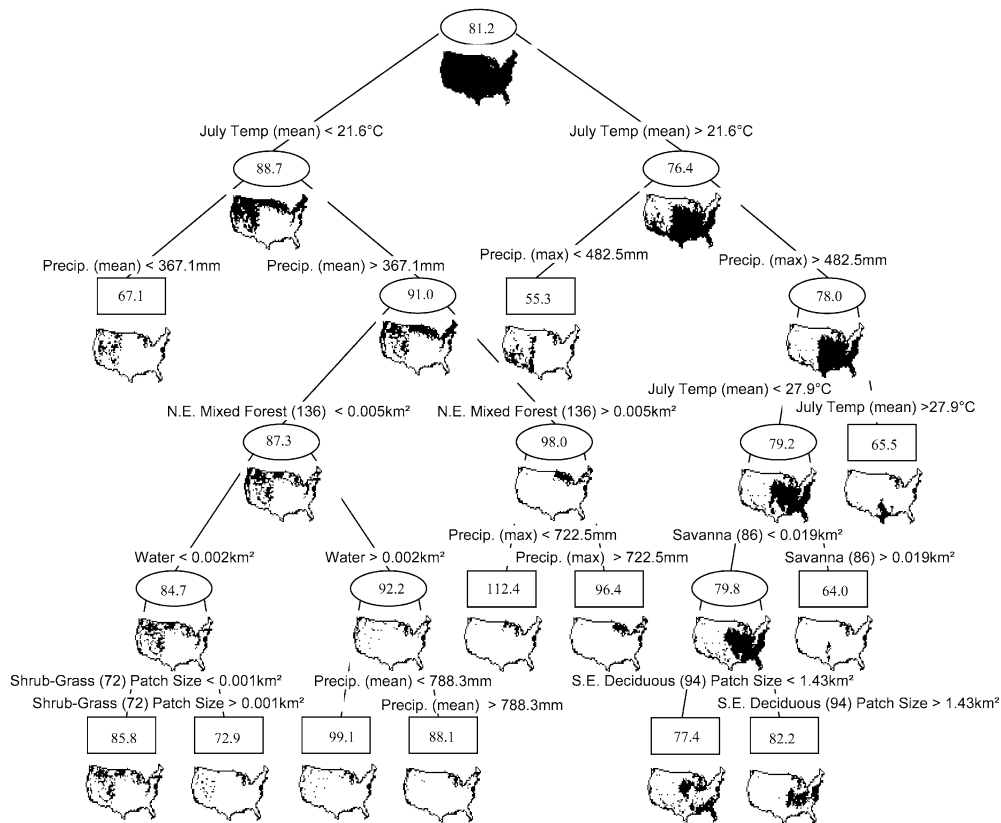
Figure 2. Regression tree model of avian species richness across the United States using finely discriminated land-cover variables (fine-cover model). The numbers in parentheses refer to land-cover classes as described by Loveland et al. (1991). See Figure 1 for a complete description of the figure.

Table 2. Performance of two sets of models built to predict avian species richness and the distributions of each of four species. Performance was assessed with a cross-validation technique and is reported here as the average percentage of the deviance explained by the models for 100 test data sets.

Response variable	Cross-validated model performance (mean % deviance explained ± SD)	
	Coarsely resolved land-cover models	Finely resolved land-cover models
Species richness	34.4 ± 10.2	36.1 ± 10.8
Pine Siskin	48.6 ± 15.6	51.2 ± 13.3
Ovenbird	60.9 ± 10.0	65.6 ± 7.8
Red-eyed Vireo	60.4 ± 6.8	64.6 ± 7.1
Black-capped Chickadee	66.5 ± 7.5	66.4 ± 6.7

In general, the two models predicted similar patterns of species richness across the United States (Figure 3). The northeast and north-central US were predicted to be highest in species richness whereas areas across much of the central and southwestern US were predicted to have relatively few species. However, closer examination of the predictions from the two models reveals substantial differences in specific regions. For example, the fine-cover model predicts more detailed patterns of species richness in the northeast and north-central US. In contrast, the
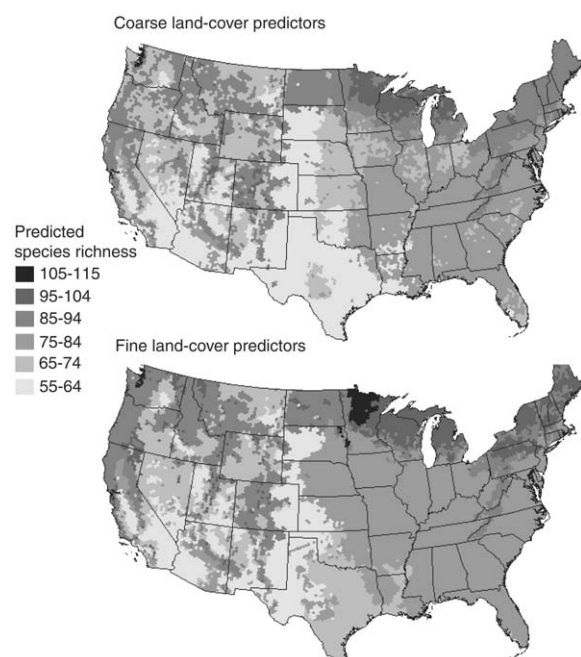
Coarse land-cover predictors

Predicted
species richness
- ■ 105-115
- ■ 95-104
- ■ 85-94
- ■ 75-84
- ■ 65-74
- ■ 55-64

Fine land-cover predictors

Figure 3. Avian species richness as predicted by two regression tree models (see Figure 1, Figure 2) built with variables derived from coarsely and finely resolved land-cover classifications respectively. Because the predictions have been binned into six categories, the shaded regions on the maps do not always show a one-to-one correspondence with the terminal nodes of the regression tree models.

coarse-cover model provides more detail in the patterns of richness in the northwestern US.

Geographic sensitivity of the prediction of species richness

As the two maps of predicted species richness indicate, the coarse- and fine-cover models of species richness generated some very similar predictions for some groups of hexagons, whereas other groups were classified quite differently by the two models, such that hexagons clustered in a particular node of the fine-cover model were spread across multiple terminal nodes of the coarse-cover model, or vice versa. This resulted in high values of the sensitivity index. Groups of hexagons that were classified in a similar way by the two models, on the other hand, had low index values and could be considered to be less sensitive to the detail of land-cover classification. Mapping the index values allowed us to see the geographic distribution of model sensitivity to the detail of the land-cover classification (Figure 4a). The darker regions of the map in Figure 4a depict areas

where the prediction of species richness was highly dependent on the particular land-cover classification used. These areas included many of the mountainous regions, the Pacific Northwest, the Midwest, and portions of the northeastern US. Areas in which the models explained species richness with climate variables only were, obviously, insensitive to the detail of land-cover data and appear in white on the map.

Because of the striking shapes of the more sensitive areas on the map in Figure 4a, we compared their distributions to a map of ecoregions of the US (Omernik 1987). We found that the more sensitive regions had a high degree of overlap with specific ecoregions (Figure 4b). The mountainous ecoregions, in particular, were sensitive to the detail of land-cover data used in the models. These included the North Cascades, Cascades, Eastern Cascades Slopes and Foothills, Northern Rockies, Middle Rockies, Wasatch and Uintas, and the Northeastern Highlands. In addition, the explanation of bird species richness in many of the northeastern and north central forested ecoregions–the Northern Lakes and Forests, North Central Hardwood Forests, Laurentian Plains and Hills, and the Northeastern Coastal Zone–was strongly affected by the detail of land-cover data used in the models.

Individual species predictions

The coarse- and fine-cover individual species models built for the House Wren and the Savanna Sparrow used only climate and elevation variables. For both of these species, the models built with the two different data sets were identical. The House Wren models explained 63% of the deviance in the data and used a combination of both July and January temperatures and maximum elevation values to predict occurrence. The models built for the Savannah Sparrow explained 56% of the deviance in the data and used a similar combination of variables. The models built for the other four species all incorporated land-cover variables and the coarse- and fine-cover models for all of these species differed (Table 3). For example, the coarse-cover model for the Pine Siskin included the mean perimeter of patches of grasslands and shrublands and a measure of land-cover. In contrast, the fine-cover model included pine forest and maximum precipitation. Despite the differences in the variables used in each pair of models, the coarse- and fine-cover models built for each species generally explained a similar portion of the deviance. The largest
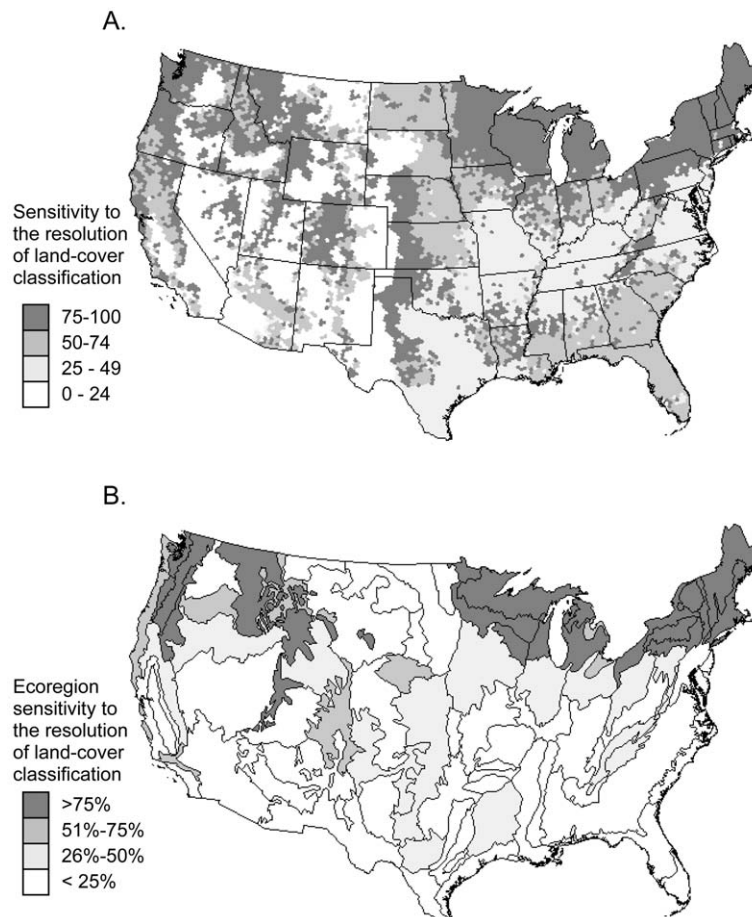
Figure 4. Maps of the degree to which the prediction of avian species richness is sensitive to the level of discrimination of land-cover data. In map A, sensitivity was estimated using an index that ranged from a low of 0 to a high of 100. See the text for a complete description of the index. Map B shows the coincidence of the areas of high sensitivity (index value greater than 75) with ecoregions of the conterminous US (Omernik 1987). The map shows the percentage of sample units (hexagons) in each ecoregion with an index value greater than 75.

difference was seen in the Red-eyed Vireo models, which explained 64% (coarse-cover model) and 68% (fine-cover model) of the deviance in the data. The cross-validation analyses confirmed these relatively small differences in model performance (Table 2). The largest differences were in the performance of the Ovenbird models, which explained 60.9% (SD = 10.0%) and 65.6% (SD = 7.8%) of the test-set deviance (coarse- and fine-cover models, respectively) and the Red-eyed Vireo models, which explained 60.4% (SD = 6.8%) and 64.6% (SD = 7.1%) of the test-set deviance (coarse- and fine-cover models, respectively).

Although the models generally fit the data to a similar degree, the predictions produced by the coarse and fine-cover models for the Pine Siskin, Ovenbird,

Red-eyed Vireo, and Black-capped Chickadee showed some striking differences (Figure 5). Both the models built for the Pine Siskin predicted occurrences throughout much of the mountainous western US (Figure 5). The fine-cover model predicted a more restricted distribution, whereas the coarse-cover model generally predicted a lower probability of occurrence across larger areas. Both models predicted relatively low probabilities of occurrence in the northern states east of Montana, where the species is also known to occur. Furthermore, the coarse-cover model predicted slightly higher probabilities of occurrence in parts of Florida and Georgia in the southeastern US where the Pine Siskin does not occur.

Both the Ovenbird and the Red-eyed Vireo models showed the most striking differences between coarse-

Table 3. Variables in eight regression tree models built to predict the presence of four bird species across the conterminous United States. Two models were built for each species: one with coarsely resolved land-cover classes (second column from the left) and one with finely resolved land-cover classes (second column from the right). Numbers in parentheses represent Loveland et al. (1991) land-cover classes.

Species	Coarsely resolved land-cover models		Finely resolved land-cover models	
	Variable	% deviance explained	Variable*	% deviance explained
Pine Siskin				
	Mean July temperature	42	Pine Forest (109)	47
	Grass-shrub patch perimeter	6	Mean July temperature	9
	Conifer patch size	5	Pine, fir patch size (116)	3
	Conifer patch perimeter	4	Maximum precipitation	3
	Land-cover diversity	3		
	Total	*60*	*Total*	*62*
Ovenbird				
	Deciduous forest	31	Beach, maple, spruce, fir forest (137)	33
	Minimum July temperature	15	Oak, hickory forest (94)	13
	Mixed forest	7	Oak, maple, pine patch size (136)	8
	Mean elevation	6	Mean July temperature	7
	Maximum January temperature	4	Range in elevation	4
	Conifer forest	3	Minimum elevation	3
	Crops and pastures	2	Beach, oak, pasture (93)	2
	Total	*68*	*Total*	*70*
Red-eyed Vireo				
	Deciduous forest	39	Minimum precipitation	35
	Mixed forest	14	Beach, oak, pasture (93)	9
	Crops and pastures	6	Pine, oak, poplar (141)	8
	Maximum elevation	3	Beach, maple, spruce, fir forest (137)	5
	Mean July temperature	2	Oak, maple, pine forest (136)	4
			Soy, corn, oak, pine patch size (45)	3
			Oak, maple, pine patch size (136)	2
			Oak, hickory, walnut forest (95)	2
	Total	*64*	*Total*	*68*
Black-capped Chickadee				
	Mean January temperature	47	Mean January temperature	44
	Minimum precipitation	15	Forage crops, hay, oak, maple (52)	16
	Crops and pastures, patch size	7	Mean precipitation	3
	Mean July temperature	2	Corn, soybeans (17)	3
			Western fir, spruce, patch size (122)	3
	Total	*71*	*Total*	*69*

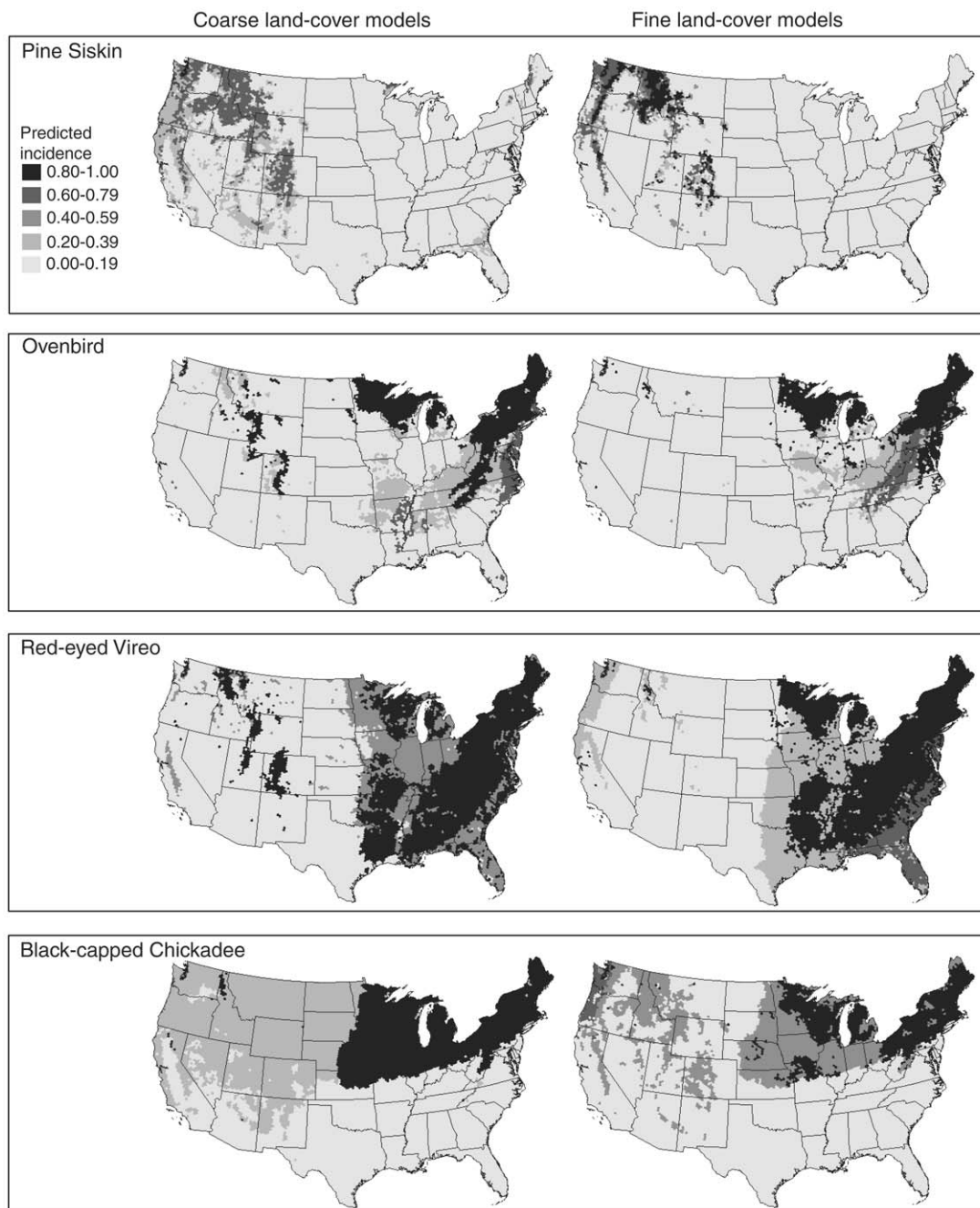*Land-cover variable names are approximate (see Loveland et al. [1991] for full names).

Figure 5. Maps of the probabilities of occurrence of four bird species as predicted by each of two models built with land-cover variables from coarsely and finely resolved classifications, respectively.

cover and fine-cover predictions in the western part of the country (Figure 5). The coarse-cover models predicted occurrences throughout portions of the Rocky Mountains, whereas the fine-cover models for these two species predicted only limited occurrences

in these areas. Given the known distributions of both species, the coarse-cover models appear to be over-predicting occurrences in many of these mountains. The coarse-cover model for the Ovenbird further differed from the fine-cover model by predicting occur-

rences in the southeastern US throughout portions of the states of Louisiana, Arkansas, and Mississippi. The fine-cover model for the Red-eyed Vireo further differed from the coarse-cover model by predicting occurrences throughout western Washington and Oregon in the northwestern corner of the country.

The Black-capped Chickadee models showed strong differences in predictions across much of the country (Figure 5). In the northeastern US, the coarse-cover model consistently predicted a high probability of occurrence (0.90) compared to the fine-cover model that predicted probabilities of 0.92 in some portions of the northeast and 0.55 in others. Similarly, the coarse-cover model predicted probabilities of occurrence between 0.25 and 0.40 across much of the western US compared to the fine-cover model that predicted probabilities of 0.48 and 0.55 in several mountainous regions and probabilities of 0.15 and 0.03 throughout the rest of the west.

Geographic sensitivity of the prediction of individual species occurrences

Geographic patterns in the sensitivity of predictions of individual species models to the detail of land-cover classifications differed across species (Figure 6). For the Pine Siskin, the areas most sensitive to land-cover classifications were in the western and far northeastern US. In contrast, the most sensitive areas for the prediction of Ovenbird occurrences were in the eastern and north-central US. The most sensitive areas for the Red-eyed Vireo were spread across much of the eastern part of the country and those for the Black-capped Chickadee were concentrated in the north-central US and in the mountainous regions of the west. The only substantial similarities in the maps of the sensitive areas for these four species were in the Rocky Mountains stretching from northern Idaho south through Colorado, in which predictions for all four species were relatively highly sensitive to the land-cover classification used, and in the arid western US, where none of the model predictions were particularly sensitive.

Discussion

The spatial scale at which ecological relationships are investigated can have a profound effect on the results and conclusions of a study (O'Neill et al. 1986, Wiens 1989). Rahbek and Graves (2001) demonstrated the
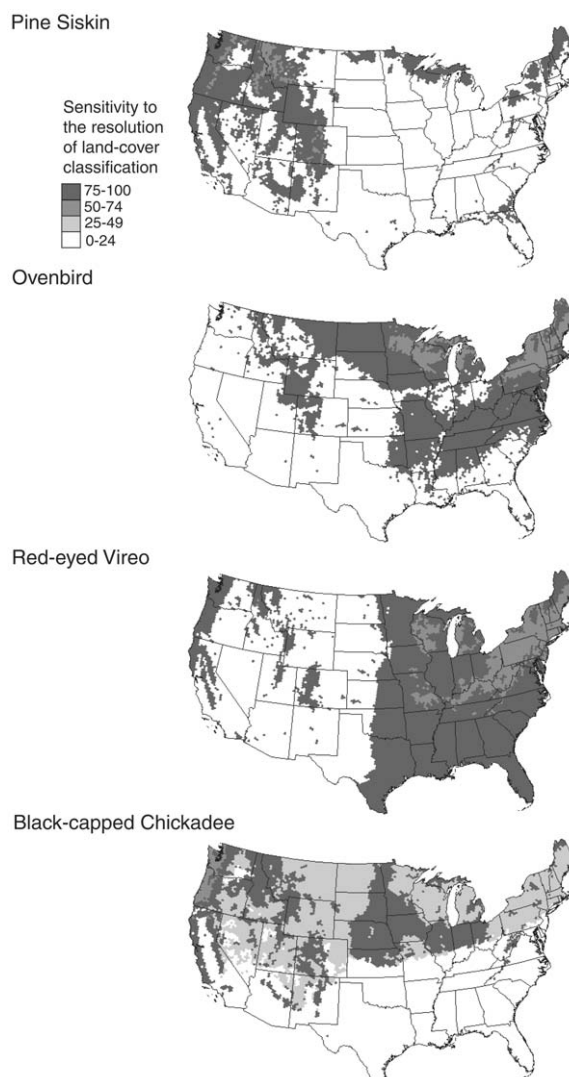

Figure 6. Maps of the degree to which the predictions of species occurrences are sensitive to the level of discrimination of land-cover data. Sensitivity was estimated using an index that ranged from a low of 0 to a high of 100. See the text for a complete description of the index.

effect of grain size on models of avian species richness across South America. Their analyses showed that increases in quadrat size generally resulted in increases in model fit. In addition, although models built at different spatial resolutions generally used the same set of explanatory variables, the ranking of the importance of the variables changed with quadrat size (Rahbek and Graves 2001). Our analyses indicate that the resolution with which categorical explanatory variables such as land cover are classified may affect

both the composition of the group of variables included in a model and the relative importance of those variables.

Furthermore, our results indicate that the effects of the level of detail of classifications are not limited to the prediction of species richness values or the modeling of an isolated species. Because our findings applied to both richness values and four out of six individual species models, we conclude that the choice of land-cover classifications is likely to affect many different ecological analyses over large spatial extents. With two possible exceptions, our results also indicate that the differential effects of using the two land-cover classifications are not limited to particular areas of the country or specific land-cover types. We found little similarity across species in the areas identified as being particularly sensitive to the land-cover classification used. This implies that the differences we see are driven by specific relationships between individual species and land cover and not by general attributes of the landscape or attributes of specific differences in the two classifications. That is to say, the fact that there are more species of trees in the eastern US than in the western US, and consequently more finely resolved deciduous and mixed forest classes in the 160-class land-cover data, may have affected some of the models, but it did not affect them in similar ways.

The two possible exceptions to this conclusion are evidenced by the two less obvious similarities in the maps depicting relative sensitivity to the level of detail of land-cover classifications (Figure 4, Figure 6). The predictions of the models of four individual species, as well as those of the models of species richness, were all sensitive to the detail of land-cover classifications in the mountainous regions of the western US – specifically the Rocky Mountains – and least sensitive in the arid west. In the case of the mountainous regions, it is well known that both vegetation and avifaunas change rapidly with elevation (Whittaker 1960; Terborgh 1977). In these areas of high species diversity and high habitat heterogeneity, coarse land-cover classifications that aggregate different habitats could potentially obscure the detail necessary to accurately predict the presence of individual species and, at the same time, reduce the ability to predict these areas of high species richness by classifying the vegetation in a similar way to other areas with much lower habitat heterogeneity. In contrast to the mountainous regions which might be expected to require finely resolved land-cover for modeling avian

species richness, the fact that models used only climatic variables in the arid west might indicate that even coarser measures of vegetation reflecting crudely defined regions of differential moisture and temperature would better predict avian diversity and occurrence than finely resolved vegetation classes. Species diversity is known to be related to moisture gradients, particularly in dry regions (Brown and Davidson 1977; Abramsky and Rosenzweig 1983; Currie 1991; van Rensberg et al. 2002). We found that species richness in these regions was predicted exclusively by climatic factors (July temperatures and annual precipitation). Further investigation revealed that the regions in which only climatic variables were needed to predict species richness had lower land-cover diversity and larger patches of land cover than the rest of the country. Consequently, areas with such low landscape diversity and heterogeneity may have lacked the variation needed for land-cover variables to add to the explanatory power of the climate variables.

There are some obvious trade-offs associated with using different resolutions of land-cover classifications in modeling species distributions. Coarsely resolved land-cover classifications generally have higher classification accuracy than more finely resolved classifications, but the classes themselves are necessarily more heterogeneous. In contrast, finely resolved land-cover classifications have more homogenous classes, but generally have higher levels of classification error (e.g., Stehman et al. 2003). The most appropriate classification for modeling a specific species depends on the species' life history and specific habitat requirements. For example, consider a small bird whose diet consists largely of small seeds from grain crops. If the bird's distribution is modeled against the distribution of individual agricultural crops (e.g., oats, barley, and wheat) the resulting model will have a poor fit because no single grain crop matches the species' distribution – the bird can be present where oats are absent if wheat is present, present where wheat is absent if barley is present, and so on. Conversely, modeling the species distribution against the distribution of all agricultural grain crops also fails because the species is absent from areas where large grains such as corn are present. Only if the agricultural grain resource is differentiated to the level of 'small grains' will we detect the match of species and food resource.

Because we compared only two classifications, it is quite possible that we failed to match the habitat

preferences of some of the six species we modeled with one or more land-cover variables. The fact that no land-cover variables were incorporated in the models built for the House Wren and the Savannah Sparrow could be interpreted as a failure of the land-cover classifications to adequately capture the habitat preferences of these birds. As in the example of the 'small grain' consumer given above, accurate models for some individual species will require the definition of species-specific land-cover classifications designed to address specific habitat requirements (e.g., Mc-Comb et al. 2002; Luoto et al. 2002).

Both of our models of species richness highlight the well-known associations between climate and species richness (Pianka 1966; Schall and Pianka 1978; Currie 1991; Gaston 2000). We found that both precipitation and July temperatures were important predictors of avian species richness regardless of the land-cover classification used in the model. In general, species richness was higher in areas with more rainfall, potentially supporting the hypothesis that diversity increases with productivity (Currie et al. 1999). However, although given the opportunity, our models did not incorporate seasonality or winter temperatures, two factors that have played major roles in other studies of species richness (Pianka 1967; Herrera 1978; Ricklefs 1980), nor did temperature play the role predicted by the diversity-productivity hypothesis. This inverse relationship with temperature, reflecting somewhat of an inverse latitudinal gradient at a sub-continental scale has been noted by others (e.g., Hawkins et al. 2003).

Individual bird species may be associated with different habitats throughout their geographic ranges (Cody 1978; Collins 1983). These differences may be driven by the availability of alternative resources in different areas or by interspecific interactions that vary regionally (Cody 1978). For example, a species that is associated with a broad range of coniferous forest types in the eastern US may be limited to a few specific forest types in the western US by a specific food resource that does not occur in all western conifer forests or by competition with western species that are not found in the eastern part of the species' range. Thus, it is not surprising that our results imply that different levels of land-cover classification may be required to accurately predict species distributions in different regions.

Our findings have substantial implications for large-scale efforts to model and map biodiversity, such as the National Gap Assessment Program (GAP)

(Scott et al. 1993). The fact that models built with different land-cover classifications produced different predictions in different regions of the US implies that any standardized national or continental approach to land-cover classification may fail to accurately predict a wide range of species over large areas.

Modeling efforts will therefore likely need to be tailored to specific areas. Thus, the regional and state approaches of the National GAP may more adeptly capture differences in variable resolution than a standardized approach that attempts to model all of the United States as a whole. Whether this is actually an advantage of the regional GAP approach would need to be determined with an analysis similar to ours involving the specific data sets used in the National GAP.

The findings reported here lead to two important conclusions. If the detail with which sample attributes are discriminated can affect the outcome of ecological studies, determining the proper resolution of any classification needs to be incorporated into the design of studies with the same care that recent research has shown must be paid to the effects of scale. Secondly, analyses that cover large spatial extents should take into account the fact that the effects of using a particular classification may have a strong geographic component. Thus, different classifications of the same continuous variable may be required to represent habitat in different areas. Our results show that it is not defensible to adopt arbitrary categories of habitat and to assume their geographic invariance. We conclude that the resolution with which continuous variables are classified, like scale and the selection of the variables themselves, is a critical aspect of ecological analyses.

Acknowledgments

References

Abramsky Z. and Rosenzweig M.L. 1983. Tilman's predicted productivity-diversity relationship shown by desert rodents. Nature 309: 150–151.

Anderson J.R., Hardy E.E., Roach J.T. and Witmer R.E. 1976. A land use and land cover classification system for use with remote sensor data. US Geological Survey Professional Paper 964. Washington, DC, USA.

Anderson S.H. and Shugart H.H. 1974. Habitat selection of breeding birds in an east Tennessee deciduous forest. Ecology 55: 828–837.

Bergin T.M. 1992. Habitat selection by the Western Kingbird in western Nebraska: a hierarchical analysis. Condor 94: 903–911.

Breiman L., Friedman J.H., Olshen R.A., and Stone C.J. 1984. Classification and regression trees. Wadsworth, Belmont, California, USA.

Brown J.H. and Davidson D.W. 1977. Competition between seed-eating rodents and ants in desert ecosystems. Science 196: 880–882.

Clark L.A. and Pregibon D. 1992. Tree-based models. In: Chambers J.M. and Hastie T.J. (eds), Statistical models in S. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, California, USA, pp. 377–419.

Cody M.L. 1978. Habitat selection and interspecific territoriality among the sylviid warblers of England and Sweden. Ecological Monographs 48: 351–396.

Collins S.L. 1983. Geographic variation in habitat structure of the Black-throated Green Warbler (Dendroica virens). Auk 100: 382–389.

Currie D.J. 1991. Energy and large-scale patterns of animal- and plant- species richness. American Naturalist 137: 27–49.

Currie D.J., Francis A.P. and Kerr J.T. 1999. Some general propositions about the study of spatial patterns of species richness. Ecoscience 6: 392–399.

Daly C., Neilson R.P. and Phillips D.L. 1994. A statistical topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology 33: 140–158.

Danko D.M. 1992. The digital chart of the world. GeoInfo Systems 2: 29–36.

De'ath G. and Fabricius K.E. and 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81: 3178–3192.

Diniz-Filho J.A.F., Bini L.M., and Hawkins B.A. 2003. Spatial autocorrelation and red herrings in geographical ecology. Global ecology and Biogeography 12: 53–64.

Dufrêne M., and Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67: 345–366.

Gaston K.J. 2000. Global patterns in biodiversity. Nature 405: 220–227.

Githaiga-Mwicigi J.M.W., Fairbanks D.H.K. and Midgley G. 2002. Hierarchical processes define spatial pattern of avian assemblages restricted and endemic to the arid Karoo, South Africa. Journal of Biogeography 29: 1067–1087.

Gutzwiller K.J. and Anderson S.H. 1987. Multiscale associations between cavity-nesting birds and features of Wyoming streamside woodlands. Condor 89: 534–548.

Hargis C.D., Bissonette J.A. and David J.L. 1997. Understanding measures of landscape pattern. In: Bissonette J.A. (ed.), Wildlife and landscape ecology: effects of pattern and scale. Springer, New York, New York, USA, pp. 231–261.

Hawkins B.A., Porter E.E., Diniz-Filho J.A.F. 2003. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. Ecology 84: 1608–1623.

HCN – Historical Climate Network. 1996. Monthly precipitation and temperature data. US Department of Energy-Oak Ridge National Laboratory and National Ocean and Atmospheric Administration-National Climate Data Center. (URL: http://cdiac.esd.ornl.gov/cdiac/r3d/ushcn/ushcn.html#TEXT)

Herrera C.M. 1978. On the breeding distribution pattern of European migrant birds: MacArthur's theme reexamined. Auk 95: 496–509.

Hildén O. 1965. Habitat selection in birds. Annales Zoologici Fennici 2: 53–75.

Hunsaker C.T., O'Neill R.V., Timmons S.P., Jackson B.L., Levine D.A. and Norton D.J. 1994. Sampling to characterize landscape pattern. Landscape Ecology 9: 207–226.

Hutto R.L. 1985. Habitat selection by nonbreeding, migratory, land birds. In: Cody M.L. (ed.), Habitat selection in birds. Academic Press, San Diego, California, USA, pp. 455–476.

Iverson L.R., and Prasad A.M. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs 68: 465–485.

Jones K.B., Neale A.C., Nash M.S., Riitters K.H., Wickham J.D., O'Neill R.V. and van Remortel R.D. 2000. Landscape correlates of breeding bird richness across the United States mid-Atlantic region. Environmental Monitoring and Assessment 63: 159–174.

Lawler J.J. and Edwards T.C. Jr. 2002. Landscape patterns as habitat predictors: building and testing models for cavity-nesting birds in the Uinta Mountains of Utah, USA. Landscape Ecology 17: 233–245.

Loveland T.R., Merchant J.W., Ohlen D.J. and Brown J.F. 1991. Development of a land-cover characteristics data base for the conterminous United States. Photogrammatic Engineering and Remote Sensing 57: 1453–1463.

Loveland T.R., Merchant J.W., Reed B.C., Brown J.F., Ohlen D.J., Olson P. and Hutchinson J. 1995. Seasonal land-cover regions of the United States. Annals of the Association of American Geographers. 85: 339–355.

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

530

MacArthur R.H., Recher H. and Cody M.L. 1966. On the relation between habitat selection and species diversity. American Naturalist 100: 319–332.

Marks D. 1990. The sensitivity of potential evapotranspiration to climate change over the continental United States. In: Guciniski H., Marks D. and Turner D.P. (eds), Biospheric feedbacks to climate change: the sensitivity of regional trace gas emissions, evapotranspiration, and energy balance to vegetation redistribution. EPA/600/3-90/078. US Environmental Protection Agency, pp. IV.1-IV.31.

McComb W.C., McGRath M.T., Spies T.A. and Vesely D. 2002. Models for mapping potential habitat at landscape scales: and example using Northern Spotted Owls. Forest Science 48: 203–216.

Miller T.W. 1994. Model selection in tree-structured regression. In: Proceedings of the Statistical Computing Section, American Statistical Association, pp. 158-163.

O'Connor R.J., Jones M.T., White D., Hunsaker C., Loveland T., Jones B., and Preston E. 1996. Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States. Biodiverstiy Letters 3: 97–110.

O'Connor R.J., Boone R.B., Jones M.T., and Lauber T.B. 1999. Linking continental climate and land use patterns with grassland bird distribution in the conterminous United States. Studies in Avian Biology 19: 45–59.

Olden J.D., and Jackson D.A. 2002. A comparison of statistical approaches for modeling fish species distributions. Freshwater Biology 47: 1976–1995.

Omernik J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,000). Annals of the American Association of Geographers 77: 118–125.

O'Neill R.V., Turner S.J., Cullinan V.I., Coffin D.P., Cook T., Conley W., Brunt J., Thomas J.M., Conley M.R. and Gosz J. 1991. Multiple landscape scales: an intersite comparison. Landscape Ecology 5: 137–144.

O'Neill R.V., DeAngelis D.L., Allen T.F.H. and Waide J.B. 1986. A hierarchical concept of ecosystems. Monographs in Population Biology 23. Princeton University Press, Princeton, New Jersey, USA.

Pianka E.R. 1966. Latitudinal gradients in species diversity: a review of concepts. American Naturalist 100: 33–46.

Pianka E.R. 1967. On lizard species diversity: North American flatland deserts. Ecology 48: 333–51.

Qi Y. and Wu J. 1996. Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices. Landscape Ecology 11: 39–49.

Rahbek C., and Graves G.R. 2001. Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences USA 98: 4534-4539.

Rathert D., White D., Sifneos J.C., and Hughes R.M. 1999. Environmental correlates of species richness for native fish in Oregon, USA. Journal of Biogeography 26: 257–273.

Recher H. 1969. Bird species diversity and habitat diversity in Australia and North America. American Naturalist 103: 75–80.

Ricklefs R.E. 1980. Geographical variation in clutch size among passerine birds. Ashmole's hypothesis. Auk 97: 38–49.

Robbins C.S., Bystrak D., Geissler P.H. 1986. The breeding bird survey: its first fifteen years. Fish and Wildlife Service Resource Publication 157, Washington, DC, USA.

Saab V. 1999. Importance of spatial scale to habitat use by breeding birds in riparian forests: a hierarchical analysis. Ecological Applications 9: 135–151.

Schall J.J. and Pianka E.R. 1978. Geographic trends in the number of species. Science 201: 679–686.

Scott M.J., Davis F., Csuti B., Noss R., Butterfield B., Groves C., Anderson H., Caicco S., D'Errchia F., Edwards Jr. T.C., Ulliman J. and Wright R.G. 1993. Gap Analysis: a geographic approach to protection of biodiversity. Wildlife Monographs 57: 1–41.

Steffan-Dewenter I., Münzenberg U., Bürger C., Carsten T., and Tscharntke T. 2002. Scale-dependent effects of landscape context on three pollinator guilds. Ecology 83: 1421–1432.

Stehman S.V., Wickham J.D., Smith J.H., Yang L. 2003. Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: statistical methodology and regional results. Remote Sensing of Environment 86: 500–516.

Terborgh J. 1977. Bird species diversity on an Andean elevation gradient. Ecology 58: 1007–1019.

Timmons S.P. and Hunsaker C.T. 1995. Tools for visualizing landscape pattern for large geographic areas. In: Goodchild M.F. (ed.), GIS and environmental modeling: progress and research issues. GIS World Books, Fort Collins, Colorado, USA, pp. 473–477.

Turner M.G. 1989. Landscape ecology: the effect of pattern on process. Annual Review of Ecology and Systematics 20: 171–197.

van Rensberg B.J., Chown S.L. and Gaston K.J. 2002. Species richness, environmental correlates, and spatial scale: a test using South African birds. American Naturalist 159: 566–577.

Venables W.N. and Ripley B.D. 1994. Modern applied statistics with S-Plus. Springer-Verlag, New York, New York, USA.

White D., Kimmerling J. and Overton W.S. 1992. Cartographic and geographic components of a global design for environmental monitoring. Cartography and Geographic Information Systems. 19: 5–22.

Whittaker R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30: 279–338.

Wiens J.A. 1989. Spatial scaling in ecology. Functional Ecology 3: 385–387.

Williams S.E., Marsh H., and Winter J. 2002. Spatial Scale, species diversity, and habitat structure: small mammals in Australian tropical rain forest. Ecology 83: 1317–1329.