This is to certify that the
dissertation entitled

IMPROVING WILDLIFE HABITAT MODEL PERFORMANCE:
SENSITIVITY TO THE SCALE AND DETAIL OF
VEGETATION MEASUREMENTS

presented by

Lance Jay Roberts Jr.

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Fisheries and Wildlife

Brian A Maurer
Major Professor's Signature

4/30/2009
Date

IMPROVING WILDLIFE HABITAT MODEL PERFORMANCE: SENSITIVITY TO THE SCALE AND DETAIL OF VEGETATION MEASUREMENTS

By

Lance Jay Roberts Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Fisheries and Wildlife

2009

ABSTRACT

IMPROVING WILDLIFE HABITAT MODEL PERFORMANCE: SENSITIVITY TO THE SCALE AND DETAIL OF VEGETATION MEASUREMENTS

By

Lance Jay Roberts Jr.

Monitoring the impacts of resource use and landscape change on wildlife habitat over large areas is a daunting assignment. Forest land managers could benefit from linking the frequent decisions of resource use (timber harvesting) with a system of wildlife habitat accounting, but to date these tools are not widely available. I examined aspects of wildlife habitat modeling that: (in Chapter 2) could potentially lead to the establishment of wildlife habitat accounting within a resource decision support tool, (in Chapter 3) improve our theoretical understanding and methods to interpret the accuracy of wildlife habitat models, (in Chapter 4) explore the effects of vegetation classification systems on wildlife habitat model results, and (in Chapter 5) show that forest structural estimates from satellite imagery can improve potential habitat distribution models (GAP) for forest bird species.

The majority of the analyses in this dissertation were done using a forest resource inventory developed by the State of Michigan (IFMAP). Paired with field vegetation and bird samples from sites across the lower peninsula of Michigan, we compared the relative accuracy of wildlife habitat relationship models built with plot-scale vegetation samples and stand-scale forest inventory maps. Recursive partitioning trees were used to build wildlife habitat models for 30 bird species. The habitat distribution maps from the Michigan Gap Analysis (MIGAP) were used as a baseline for comparison of model

accuracy results. Both the plot and stand-scale measurements achieved high accuracy and there were few large differences between plot and stand-scale models for any individual species. Where the plot and stand-scale models were different, they tended to be species associated with mixed habitats. This may be evidence that scale of vegetation measurement has a larger influence on species associated with edges and ecotones. Habitat models that were built solely with land cover data were less accurate than models that included detailed vegetation composition and structure information. This result was supported in multiple analyses, including forest structural estimates generated from satellite imagery.

There are distinct patterns of model accuracy and especially commission and omission errors that are linked to species ecological traits and method of error calculation. These patterns are illustrated with figures that relate the model results to a conceptual relationship between a species' probability of presence at a given location and the suitability of the habitat at that location. The correct application of accuracy assessment is key to correctly understanding the utility of a model and to avoid discounting a model as useless when it is in fact informative. I also compared the relative accuracy of wildlife habitat relationship models built with three different hierarchical vegetation classifications. Despite major differences in the distribution of field sites among the classes, there was little difference in terms of bird habitat model accuracy between the classifications at any given level. The number of classes (level of the hierarchy) appeared to be more important to bird habitat model accuracy than did the nature of the classification itself.

ACKNOWLEDGEMENTS

My committee and especially my advisor Dr. Brian Maurer were of great assistance in the development of this dissertation and I owe them a great deal of thanks. I also wish to thank Mike Donovan of the Michigan Department of Natural Resources for project support, guidance, and funding. I am very grateful to have had a Plant Science Fellowship during my first two years at MSU. Co-workers in the Maurer Lab were wonderful to spend time with and bounce ideas off of. I am very grateful for my supportive and talented family for encouraging me to complete this endeavor, and for putting up with all the vagaries that come with graduate school and academia.

TABLE OF CONTENTS

CHAPTER 2 – ASSESSING THE UTILITY OF A FOREST RESOURCE INVENTORY DATABASE FOR USE IN MONITORING WILDLIFE HABITAT

CHAPTER 3 – PATTERNS OF WILDLIFE HABITAT MODEL PERFORMANCE IN RELATION TO THE HABITAT SPECIFICITY AND PREVALENCE OF SPECIES

CHAPTER 4 – INFLUENCE OF VEGETATION CLASSIFICATIONS ON WILDLIFE HABITAT MODEL PERFORMANCE

CHAPTER 5 – MAPPING FOREST STRUCTURE FROM SATELLITE IMAGERY FOR WILDLIFE HABITAT DISTRIBUTION MODELS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO THE USE OF WILDLIFE HABITAT MODELS IN CONSERVATION AND MANAGEMENT

A model is an abstraction or simplified representation of reality. We use models to simplify concepts and to help us understand systems that are too complex to measure directly. The abundance, distribution, and dynamics of wildlife populations are examples of system states that are far too complicated to measure with any confidence or without unfeasible levels of effort. We can, however, measure the abundance, distribution, and dynamics of *vegetation* across a landscape with much less effort and much higher confidence. And since the distribution of wildlife populations depends in large part on the distribution of their habitats (e.g. Grinnell 1917), we can use these data to produce simplified representations (i.e. models) of where we expect wildlife to be distributed. These simplified representations of wildlife habitat distributions can be used, within certain limits (Guisan and Thuiller 2005), to guide our use of natural resources.

Wildlife habitat models are vital to managers who must plan and perform conservation activities, as well as anticipate landscape and climatic changes, all with very limited information. The task of monitoring and maintaining biological diversity (and especially species of special concern) creates a monumental task for public land managers who, by law, are required to complete these difficult mandates (Manley et al. 2004). Wildlife habitat modeling is also an active branch of ecology and conservation, and ecologists strive to improve the quality of these models through a wide variety of means (Austin 2007).

As habitat modeling and other conservation projects are implemented there are a multitude of choices that must be made as to the features that will be included in the models, and the sources of these data. These choices go hand in hand with the limitations provided by research budgets and the difficulty (cost) of acquiring more detailed and accurate data. Typically the independent variables will consist of environmental data that may include categorical habitat classes, vegetation or substrate measurements, and climate or other abiotic features. These can be generated in any number of ways, from classified satellite imagery to intensive field samples. Dependent variables consist of species occurrence records, and these data also require many methodological choices involving tradeoffs in accuracy, information content, and cost to acquire. Alternatively, when species occurrence data are not available, expert opinions can be used to draw the links between species and the habitats they are associated with. But with expert opinion at least some objectivity is lost. These options should be taken very seriously in the design of a wildlife habitat modeling project, and managers and scientists alike should weigh the costs of acquiring data against the incremental benefits that result from including more detail in a model.

Knowledge of the ecology of the species is vitally important in building wildlife habitat models (McPherson and Jetz 2007). Species that are abundant, prevalent (present on a large proportion of the study area), conspicuous, have high site fidelity (return to same breeding locations every year), and have close associations with measureable vegetation characteristics (habitat specific) are likely to be more accurately modeled than species that are rare, nomadic, or not closely associated with vegetation characteristics (Seoane et al. 2005b). As long as the time and resources are available, each species should be considered a unique case and knowledge of the species' life history taken into account in model development (meaning both the type of model and the input data), as opposed to a single modeling method applied to all species. Knowledge of a species' life history can also aid users of wildlife habitat models in their interpretation of the results. Care should especially be taken to understand the assumptions that were used in model construction process, and to apply the model outputs at the correct scale and level of certainty.

I have crafted four projects based on wildlife habitat models in order to illustrate some of the potential uses, limitations, and advances that are achievable in the field. In Chapter 2 I use the data generated by a forest inventory database for state lands in Michigan to evaluate the potential for its use in wildlife habitat modeling. This database is continuously updated and managed as a resource use decision support tool. If wildlife habitat can be accounted using the same system, it will allow state land managers to incorporate wildlife habitat into resource use decisions, and to forecast wildlife habitat conditions into the future. Chapter 3 explores in more detail some of the effects of ecological traits that are particular to each species, and how we can use this knowledge in constructing and interpreting models, model results, and model accuracy calculations. Chapter 4 examines the influence that vegetation classification systems (the most common environmental variable input) have on wildlife habitat model use and accuracy. Chapter 5 describes a method of using satellite imagery to build estimates of forest structure that can be used with estimates of land cover built at the same scale and using the same imagery, and describes the magnitude of improvement from including such data. And I close in Chapter 6 with a summary of the important conclusions and overall lessons that one can take home from this dissertation.

CHAPTER 2

ASSESSING THE UTILITY OF A FOREST RESOURCE INVENTORY DATABASE FOR USE IN MONITORING WILDLIFE HABITAT

Introduction

One major task of state and federal land managers is to provide accurate information to policy makers on the impacts of various land use decisions on non-consumptive natural resources like wildlife habitat. Monitoring the impacts of resource use and landscape change on wildlife habitat throughout a natural area, state, or region is a daunting assignment. There are, however, significant assets available to managers and researchers interested in accomplishing this task. Many national forests, wildlife preserves, and state land management agencies keep detailed accounts of vegetation resources in spatially explicit and regularly updated databases. These data have been limited in their application for tasks such as monitoring wildlife habitat due in large part to the general complexity of wildlife habitat modeling.

Michigan Department of Natural Resources (MDNR) personnel are seeking to implement a system of habitat accounting for all species, not just the important game species or rare species that have been monitored in the past. To this end, I assessed the utility of Michigan's Integrated Forest Monitoring, Assessment, and Prescription (IFMAP) database as a tool for tracking statewide quantities of wildlife habitat. I used vegetation and bird data from field sites to build wildlife habitat models. The results are used to: 1) show the potential magnitude of improvement available when detailed vegetation data are used in comparison to land cover data that has been relied upon in habitat models to date, and 2) investigate the relative accuracy of models built with vegetation measurements recorded at different scales (plot vs. stand).

Wildlife-habitat models can be a useful component of ecosystem management and play a critical role in determining conservation priorities and making land management decisions. The effects of forest management on wildlife populations are numerous and varied, including: removing individuals, interrupting dispersal between populations, changing the patterns of movement and migration, and altering abiotic conditions (Wigley and Roberts 1997). All of these effects can have considerable influence on population vital rates. Forest management creates a dynamic mosaic of habitat on the landscape, altering plant species composition and especially age distributions. These fluid conditions require wildlife populations to continually adjust through changes in abundance, movements, and persistence (Villard et al. 1999, Donovan and Flather 2002, Gu et al. 2002, Thompson et al. 2003, Hanley et al. 2005). Natural resource managers are actively seeking new and innovative ways to account for wildlife-habitat dynamics, in large part through modeling and landscape-level assessments of vegetation conditions.

In some cases natural resource managers do not incorporate available forest inventory data, instead relying solely on land cover type maps to assess habitat distributions (Lawler et al. 2004, Seoane et al. 2004a). This can be because vegetation information is not available at a sufficient resolution, accuracy, and/or sampling intensity that would make predictive wildlife habitat models accurate enough to be useful. Or, it can be a result of the complexity of building wildlife-habitat models. As a result, landscape-scale models of potential wildlife habitat like GAP are frequently relied upon for local conservation projects. Systematically collected forest inventory data can have significant value in developing wildlife habitat models (Karl et al. 1999, Welsh et al. 2006), but it is still uncommon to include these data in models of wildlife habitat distribution (Flather et al. 1992, Imhoff et al. 1997, He et al. 1998, Osborne et al. 2001, Heikkinen et al. 2004, Seoane et al. 2004b).

As inventory technology and data resources become more available, evaluation and refinement of these emerging assets can ensure that they are efficiently translated to conservation research and management applications. This study highlights the potential benefits of applying systematic forest resource inventories for modeling wildlife-habitat distributions, and their use in local monitoring and decision-making. By utilizing local and regional database resources, modelers and managers can apply more detailed vegetation inventory to generate more accurate spatial habitat assessments to "step down" regional habitat assessments like GAP to local applications and provide information for tactical resource management decisions (Noon et al. 2003, Gottschalk et al. 2005, Austin 2007).

Natural resource managers are burdened with the fact that resource use decisions affect wildlife populations not just in the immediate area at the present time, but over larger spans of space and time. Combining natural resource management (especially timber harvesting decisions) with regularly maintained and accessible wildlife habitat information creates near real-time opportunities for adaptive management. The projection of future conditions in vegetation and subsequent assessment of wildlife population abundance is inevitably associated with large levels of uncertainty. This is an inherent, and sometimes overlooked, part of species distribution modeling (Whittaker et al. 2005).

Apart from the uncertainty that arises in the absence of detailed vegetation information, wildlife habitat models are susceptible to many sources of error that must be carefully considered and accounted for. The habitat resources that limit the occurrence of a species can vary across its range, and its absence in a location can be due to many factors (in addition to habitat associations) including; competition (Herzog and Kessler 2006), population abundance (Linder et al. 2000), dispersal (Mortberg 2001), and more (McPherson and Jetz 2007). Species location data is typically sparse, and where these data are available, perhaps only a small portion of the species sampled are abundant enough to be useful for statistical modeling (Araujo and Guisan 2006). This supports the idea that for conservation of all species, especially the rare ones, expert-based descriptions of habitat associations may be necessary (Hernandez et al. 2008), but see also Seoane (2005a).

Often, habitat resources are evaluated for only a limited number of economically important species, such as game animals or endangered species (Hansen et al. 1999, Karl et al. 1999). In Michigan, the initial approach has been to institute a Gap Analysis Program (MIGAP) for the state (Donovan et al. 2004), following the National GAP protocol (Scott et al. 1993). The GAP protocol relies on a state or region-wide land cover map derived from Landsat satellite imagery (MDNR 2001), and expert-based assessments of habitat associations to build potential habitat distribution maps (Edwards et al. 1996). GAP was not designed to inform local resource use decisions, but instead to coordinate conservation efforts between management groups. Nevertheless they are often inappropriately applied in local and tactical-level conservation projects. The proper use of GAP is to identify locations of potential conservation value, and then use more detailed approaches to convert the potential habitat distributions into maps that more accurately represent habitat suitability for each species (Edwards et al. 1996, Edwards et al. 1998, Peterson 2005).

In a previous study I found that MIGAP models overestimate the amount of available habitat for most species (unpublished report). When treated as a prediction of presence/absence, the MIGAP models result in a high rate of commission error (predicted present but not detected) but low omission error rates (predicted absent when actually detected). This result was also shown by Petersen and Kluza (2003). Of the many possible reasons for this pattern of errors, two are most likely. First, the landscape-level land cover maps derived from most satellite image classifications (MDNR 2001) do not have the spatial accuracy or vegetation description detail necessary for revealing an accurate distribution of habitats on the ground (Roloff et al. 2008), so GAP models typically err on the side of including areas with even a very small chance of species occurrence. Second, published wildlife-habitat relationships are in many cases not refined enough to describe the specific vegetation elements that drive habitat associations, nor are they detailed enough to compensate for the geographical differences in habitat associations across a species' range. Both of these issues result in the inclusion of more locations (as *potential* habitat) than each species would actually occupy.

The Michigan DNR instituted a resource inventory called the Integrated Forest Monitoring, Assessment, and Prescription (IFMAP) program. IFMAP is a Geographic Decision Support System (DSS) that tracks the stand-scale forest composition and structure for state-owned lands throughout Michigan, and contains detailed vegetation information on non-forested areas. Given its detail, IFMAP appears to be ideal for purposes of wildlife habitat evaluation. The goals of this project are: 1) to determine the amount of improvement (if any) in prediction accuracy of wildlife habitat distribution models when forest inventory data is included in the habitat descriptions, and 2) to determine whether it would be appropriate to build a wildlife-habitat modeling component into the IFMAP DSS so that tactical-level wildlife habitat evaluations can occur simultaneously with resource management decisions. The future of wildlife decision support in Michigan is a refined modeling protocol that can reduce the error rates inherent in the currently available tools and can track the changes in area and distribution of wildlife habitat with each management action on state-owned lands. These data could be valuable for making resource use decisions, especially in a DSS environment like Michigan's IFMAP program.

Methods

The study area is located in the Lower Peninsula of Michigan, which is separated into two ecoregional divisions (Albert 1995). At approximately the midpoint north-south there is a border between the Laurentian Mixed Forest Province to the north, and the Eastern Deciduous Forest Province to the south. The northern landscape is primarily forested, with a wide variety of coniferous and deciduous species present, and the southern landscape is primarily an agricultural matrix with pockets of deciduous forest, largely in riparian and wet areas not suitable for agriculture (MDNR 2001). The landscape of Michigan has changed significantly since presettlement, including the near complete elimination of dominant old growth hemlock/hardwood forests in exchange for second growth hardwoods and conifers (White and Mladenoff 1994).

In 2005, a survey crew visited five locations in the northern Lower Peninsula consisting of separate management units of the Pere Marquette State Forest in Grand Traverse, Benzie, and Manistee counties. In 2006 and 2007, six State Game Areas in the southern Lower Peninsula were sampled. The southern sites cover a wide variety of habitats including forested, lowland, and agricultural land cover types, the northern units were all primarily forested. These 2000-3000 acre units were selected to match locations that IFMAP stand-scale surveys had been completed by MDNR personnel.

In each unit, thirty randomly distributed plots were selected from a larger set of randomly generated coordinates, and stratified according to the relative abundance of different land cover types. GPS units were used to locate the field sites where bird and vegetation surveys were performed within a 50m radius of the plot center. Bird and vegetation surveys were conducted between late May and early July. On average 26 of the 30 sites per compartment were surveyed, the remaining sites were excluded due to access restrictions or time constraints in the field. Results were calculated for 393 sites in total.

Field vegetation surveys were based on the IFMAP protocol used by MDNR personnel (MDNR 2005). This method relies on visual estimates of canopy closure, canopy and sub-canopy species cover, average height, ground cover type and density, and measurements of basal area, and diameter. Basal area measurements were carried out using a 10 basal area factor prism, and diameter measurements were taken with a diameter tape. The dominant habitat type at each survey site was also classified into a hierarchical vegetation cover class following IFMAP classification rules.

Songbird surveys were conducted between mid May and early July in response to the return of migrants and onset of the breeding season. A regionally standard bird survey protocol (Ralph et al. 1995, Howe et al. 1997) was used for conducting the songbird surveys: point counts were carried out by identifying bird species and their individual locations from the centers of the survey circles using sight and sound within a 10 minute time interval. The surveys were conducted between 6:00AM and 11:00AM. No surveys were conducted during rain or strong wind.

At each of the plot coordinates I sampled the IFMAP GIS stand inventory maps and calculated a set of habitat conditions (Table 2.1) including: vegetation cover class, average basal area, average diameter of all canopy trees, canopy closure, proportion of canopy cover from deciduous trees, canopy species richness, canopy species diversity (Simpson's Reciprocal Index), subcanopy cover, subcanopy species diversity, overall size, upland/lowland (binary), plantation (binary), and location (binary – North/South). All maps were built on the same resolution (30m x 30m cells) and extent as the original 2001 MIGAP/IFMAP land cover dataset. This same set of habitat variables was created from the plot-scale vegetation survey data.

The IFMAP forest inventory contains a four level hierarchical classification of land cover for each stand. In the 393 field plot samples there are 75 level four land cover classes. These 75 classes are unevenly distributed among the sites, dominated by 11 classes which were assigned to over 50% of the sites, while 54 classes were represented by five or fewer sites (<2% of the total). I included only level three classes in the statistical models, which have a much more even distribution of sites among the classes.

The IFMAP level three classes are similar in number and description to the MIGAP land cover types (MDNR 2001).

The list of species in this analysis included only the bird species that are likely to be detected in field surveys, i.e. eliminating nocturnal, non-vocal, and rare birds. The majority of species in the overall sample were not present in large enough numbers to build prevalence-based habitat-association models; so all analyses have been conducted on a set of thirty of the most prevalent bird species that represent a variety of upland, lowland, forest, and non-forest habitats (Table 2.2). I simplified the recorded abundance of each species at each site into the binary variable of presence/absence (detected/not detected). The size of this sample does not support the use of abundance for these statistical models.

I constructed three 'phases' of models to assess the relative differences in accuracy as a result of adding either more detailed (e.g. additional vegetation structure and composition information) or higher spatial resolution vegetation data (e.g. stand-scale measurements to plot-scale estimates). The accuracy of predicted bird species distributions at each model phase was assessed against field survey data using three statistical criteria; omission/commission error, kappa, and area under the curve of receiver-operator characteristic plots (ROC/AUC). The accuracy measures were averaged over all species to evaluate overall patterns. The purpose of this phased, multiple accuracy-test scenario was to illustrate the applicability of wildlife habitat models given the restrictions of different input data sources.

Two of the three phases of wildlife-habitat models were generated with a statistical model known as recursive partitioning trees (Feldesman 2002), also known as classification and regression trees or CART (De'ath and Fabricius 2000). Recursive partitioning models were generated with the 'RPART' module in R (Atkinson and Therneau 2000), and accuracy measures were calculated with the 'PresenceAbsence' package (Freeman and Moisen 2008a). Recursive partitioning is a statistical classifier that iteratively divides the samples into increasingly homogeneous groups based on a cutoff value for a single independent variable, and is similar to (non parametric) discriminant analysis. Recursive partitioning performs well in comparison to most other statistical wildlife habitat relationship models and provides a flexible and easily interpreted method for linking vegetation data with species occurrences (Segurado and Araujo 2004, Prasad et al. 2006). The structure of expert-based models is similar to those generated with recursive partitioning (i.e. a set of logical conditions or rules defining vegetation classes and cutoffs in structure or composition variables). I used recursive partitioning to predict each species' probability of presence at each sample location, and compared these predictions to the field observations. I sought to keep the complexity of the recursive partitioning models low so I used a relatively large complexity parameter (value used to decide whether to include a new split) and limited the number of splitting levels to four in the trees to prevent over fitting of the independent data (Anderson and Burnham 2002). Unsupported splits and branches were pruned with a leave-one-out cross-validation routine. The resulting models have a maximum of 15 splits resulting in 16 classified groups (end nodes of the classification tree).

For the MIGAP models, I overlaid the habitat distribution maps on the field survey plot locations to identify the sites where appropriate habitat was predicted to be available. I treated this list of sites as predicted presences, and compared them with the observed locations in the field surveys. The accuracy of each species' model was assessed using 2x2 error matrices (actual presence/absence vs. predicted presence/absence) to calculate commission error (sites where the species was incorrectly predicted to be present), omission error (sites where the species was incorrectly predicted to be absent), and kappa. Kappa accounts for large differences in the number of sites in the present and absent categories (Karl et al. 2000, Manel et al. 2001). The construction of error matrices for RPART models requires that a response threshold (probability of occurrence value that separates presence from absence) be set so that sites could be classified into the binary presence/absence categories. The threshold used in this study is the value that sets the predicted prevalence equal to the observed prevalence of each species, a method supported by Freeman and Moisen (2008b). With this technique the threshold for common species will be higher than for less prevalent species, preventing artificially inflated omission errors for less prevalent species (Chapter 3). For the MIGAP models, both the predicted and observed values are binary so no threshold is necessary.

To provide an additional measure of accuracy, and a comparison to kappa, I used ROC/AUC (Fielding and Bell 1997, McPherson et al. 2004, Allouche et al. 2006). ROC/AUC provides an accuracy measurement that is independent of the response value threshold. In general, kappa and ROC/AUC are highly correlated, but ROC/AUC is more apt to represent the accuracy of models built for less prevalent species.

For the first model phase I used the MIGAP potential habitat distribution models (Donovan et al. 2004). MIGAP will be used as a baseline for comparison against the statistical and expert wildlife-habitat models (described below). The MIGAP models consist of cross-walking the vegetation cover classes identified in a database of wildlife habitat associations to the MIGAP land cover classes (MDNR 2001), then producing a binary (present/absent) map output for each of 327 bird, mammal, reptile, and amphibian species. The habitat distribution maps were then clipped to each species' range extent. These potential habitat distribution maps were overlaid on the field site locations to determine the predicted presence or absence for the thirty species of interest (Table 2.2) and tested with field detections. I expected the models described below to achieve higher accuracy than the MIGAP models.

The second phase of habitat models used RPART to determine the set of vegetation cover classes (but not structure or composition) that best accounted for the detection of individuals among the surveys. Any improvement from the MIGAP models to the more detailed IFMAP stand-scale land cover models (Phase 2a) or the plot-scale land cover models (Phase 2b) could be the result of two differences: 1) a more refined selection of appropriate land cover types in the statistical models vs. the original MIGAP (expert-based) habitat list, and/or 2) a more accurate depiction of the spatial arrangement of cover types in the Phase 2 models vs. the MIGAP satellite land cover classification. In the case of the IFMAP GIS database (Phase 2a) the maps are derived from aerial imagery and field surveys, and in the case of the field plots (Phase 2b) the vegetation and bird samples were conducted on the same 50m radius plots. A third difference between the models has the potential for influencing different accuracy results between phase 1 and 2 models. There are slightly different land cover class definitions between the two data sets. Some of the classes that are unique to one land cover map may be important habitat descriptors for certain species.

In addition to assigning a vegetation cover class to each stand, IFMAP inventories collect data on canopy and sub-canopy cover, species composition, and related forest structural variables (Table 2.1). The same set of survey data that is included in IFMAP inventories was gathered on a 50-meter radius sample plot by field technicians skilled in plant identification and measurement. The third phase of models includes these additional variables in statistical (recursive partitioning) habitat association models. If the accuracy of models improves between Phase 2a and 3a, and/or 2b and 3b, this can be attributed to the additional vegetation information accounting for more of the variation in the wildlife field samples, but not because of the spatial arrangement of the vegetation. Since the Phase 3a and 3b models were built with the same set of variables, but recorded on a different scale, any differences in accuracy can be attributed to the differences in scale (plot vs. stand) of the vegetation measurements.

Results

The overall accuracy of the three model phases ranked in an expected order (Figure 2.1). Kappa values measure the departure from randomness, 0.0 being no different from random, 1.0 representing perfect prediction. The average kappa values for the original MIGAP models (Phase 1, kappa=0.09) are lower than for recursive partitioning stand and plot-scale cover type only models (Phase 2a and 2b, kappa=0.29, 0.31). When vegetation composition and structure data are included, the average kappa values increase (kappa=0.39 and 0.40, respectively for Phase 3a and 3b models). The majority of the Phase 3a and 3b species models (25/30 and 27/30, respectively) scored 0.3 or better vs. half of the Phase 2a and 2b models (16/30 and 15/30 respectively) and very few of the Phase 1 models (2/30). Overall, the differences between phases 2 and 3 are significant but the differences between the stand and plot level models are not (Figure 2.1).

ROC/AUC results show a similar difference in accuracy between Phases 2 and 3 (Figure 2.2), and the trend of higher accuracy for Phase 3 vs. Phase 2 is repeated. These results (Figures 2.1 and 2.2) show that the scale at which these vegetation measurements are recorded (Phase 2a vs. 2b, and 3a vs. 3b) accounts for less of a difference in model accuracy than does the addition of detailed vegetation characteristics (Phase 2a vs. 3a, and 2b vs. 3b). This pattern is consistent for both kappa and ROC/AUC, which are correlated in these data (Phase 3a: $R^2 = 0.55$, 3b: $R^2 = 0.47$).

The MIGAP models (Phase 1) revealed higher rates of commission error, but lower omission errors, when compared with the other model phases (Figure 2.3). All of the recursive partitioning statistical models (Phase 2 and 3) showed similar omission and commission error rates, but Phase 3 models trended lower. The differences between phase 2 and 3 omission and commission errors are highly significant ($p<0.01$), but the differences between the stand and plot scale models are not.

The accuracy of the stand and plot-scale models (Phase 3a and 3b) are similar for most species. The exceptions to this pattern (with a difference between stand and plot-scale models of kappa >= 0.2) were seen for Field Sparrow, Tufted Titmouse, and Northern Flicker (all of which are associated with mixed habitats, Table 2.2). Other species that showed relatively large differences between stand and plot-scale models were also mostly from the mixed habitat guild. ROC/AUC supports this result, the largest differences in AUC values between Phase 3a and 3b models were for Cedar

Waxwing (mixed habitat guild) and Northern Flicker. For Field Sparrow, Northern Flicker, and Cedar Waxwing the plot-scale models were more accurate, but for Tufted Titmouse it was the stand-scale. For a more detailed examination of these results, including correlation between model accuracy and species prevalence, and results aggregated by habitat guild see Chapter 3.

There was a wide range of variables included in the Phase 3 models. Every model included cover type at least once (average = 1.6, Table 2.3), the majority of the models (3a: 21/30, 3b: 24/30) included cover type at the root node (first split). Location (north or south) was the next most common first splitting variable (3a: 6/30, 3b: 6/30). Diameter, basal area, subcanopy cover, and canopy diversity were the next most common variables included in the recursive partitioning models. Site descriptors such as overall size, upland/lowland, and plantation were included only rarely. On average each model included approximately five variables and six splits (out of a maximum of 15).

Discussion

These results establish that forest resource databases like IFMAP can be as useful as intensive plot-scale field samples in monitoring wildlife habitat, and suggest that a wildlife habitat resource module could be successfully implemented into forest resource decision support tools. This will make it possible to track changes in wildlife habitat resources that result from each timber resource management action. There are numerous practical and technological hurdles that need to be accounted for in this process. Since appropriate statewide wildlife survey data from which to fit statistical models rarely exist, I advise following a strategy similar to gap analysis; creating expert-based models for each species based on published habitat accounts and local habitat associations. While the accuracy shown by expert models is likely to be lower than models fit with statistical algorithms, they could prove to be a dramatic improvement over GAP models, assuming continued efforts to refine habitat definitions as additional wildlife location data become available (Seoane et al. 2005a).

One of the most important results from this study is that the resolution at which these vegetation measurements were recorded (small plot measurements vs. entire stand summaries) is less vital to model accuracy than is the addition of detailed vegetation characteristics (i.e. vegetation structure and composition vs. land cover types). When results are averaged over many species there is very little difference in accuracy between these two scales of habitat measurement (Figures 2.1 and 2.2). However, looking further into this comparison reveals some interesting differences. Those species that show a large difference in accuracy between stand and plot-scale models belong to the mixed/edge habitat guild (Table 2.2). The majority of forest and other habitat guild birds have a small difference in accuracy between their plot and stand-scale models. It appears that the hard stand-edge delineations in the IFMAP GIS database may mask some of the important ecotonal features of wildlife habitat, and thus lead to lower accuracy for mixed/edge habitat species. In contrast, the plot-scale vegetation samples are more likely to accurately depict these edge and mixed conditions. Still, at least one mixed habitat species was more accurately modeled with stand data (Tufted Titmouse). An additional explanation could be that the selection of habitat variables (Table 2.1) is biased towards aboveground vegetation characteristics. Spatial pattern metrics (e.g. edge indices) could provide some important additional information to stand level vegetation information in this context.

In addition to being a rich source of habitat information, the nature of the stand level IFMAP GIS inventory data lends itself to the calculation of landscape pattern and context variables that could add important habitat information for many species. A small set of spatial variables (Riitters et al. 1995, Gustafson 1998) such as edge density, patch size, patch shape, distance to important landscape features such as water or roads, and other metrics could account for a significant amount of variation in bird habitat locations and improve habitat model predictions even further (MacFaden and Capen 2002). A major restriction to the use of landscape pattern metrics with the IFMAP stand data, however, is the prevalence of artificial edges created by ownership and jurisdictional boundaries. It was for this reason that I did not include landscape pattern metrics in this analysis.

There was a wide variety of vegetation structural and composition variables that were included in the statistical models (Table 2.3), and most species were associated with at least two of these. Cover types from the IFMAP ecological classification system were included in every species' model, often at more than one level in the classification tree (Table 2.3), which indicates that the vegetation community plays an important role in wildlife habitat models. Cover type was also the most common first splitting variable among all the species, followed by location (north/south). All of these provide evidence of a broad to fine scale hierarchy of habitat selection cues.

The MIGAP models (Phase 1) are unique in this study in having many commission errors and few omissions. This is consistent with their intended purpose, to track the distribution of potential habitat and identify those species which do not have an appropriate amount of habitat under the protection of reserves (Rodriguez et al. 2007). For this reason omission errors could be much more damaging to conservation purposes than commission errors. The rest of the statistical models show rates of omission and commission error that are nearly equal (Figure 2.3). I included MIGAP in this study as a predictive habitat model despite knowing that this application is inconsistent with its desired purpose.

The automatic selection of thresholds for calculating error matrices for each species appears to have found a balance between omission and commission error that maximizes model quality (kappa). Depending on the choice of threshold, there can be positive or negative relationships between commission/omission errors and prevalence, and these choices should be part of the a priori strategy of the habitat modeling project (see Chapter 3).

It is important to note the effects that the method of accuracy assessment has on these results. Calculating the accuracy of binary models (habitat or not) is relatively simple. At any given location the model predicts either the presence or absence of each species, and field wildlife samples can confirm this. A 2x2 error matrix is then built, and % correctly classified (PCC), omission/commission errors, and kappa can easily be calculated. Statistical models like recursive partitioning are seemingly more refined in that at each location a continuous (0.0 – 1.0) probability of presence is generated. This continuous scale, in theory, more accurately fits the actual probability that a species will be present at that location. In other words, for any given location there is some probability between 0 and 1 that a species will actually be present there. For common species, and in very appropriate habitats this value might be close to one, but in inappropriate habitats might be low or even zero.

This phenomenon can be described as a probability of occurrence function (Karl et al. 2000) where the probability of presence can be graphed against a gradient of different habitats, each of which is unique to that species (Chapter 3). Rarer species are likely to show either a lower probability of occurrence across their habitat suitability gradient or a narrower range of habitats with a high probability of occurrence, and thus may be less likely to be accurately modeled by binary predictions. Accuracy measures calculated from a error matrix (kappa, omission/commission error) suffer from the same limitation as GAP models, essentially binning a continuous prediction scale into a binary one. ROC/AUC on the other hand is calculated over all possible thresholds in the probability of presence value and is therefore an attractive accuracy measure (Freeman and Moisen 2008b). The kappa statistic accounts for large differences in the numbers of presences and absences in a sample, but requires the careful choice of thresholds to adequately represent model quality. ROC/AUC is independent of threshold and may therefore be a less biased measure of model quality (Manel et al. 2001, Allouche et al. 2006).

Despite the steps I took to prevent over fitting of the RPART models (namely limiting the number of branching levels to four, and using a relatively high complexity parameter), there were instances where careful examination of the classification tree revealed splitting definitions that do not make sense in terms of current knowledge of species-habitat associations. These issues and others can be eliminated over time with careful examination and continuous refinement of the wildlife habitat models and definitions.

All of these models, and especially GAP, are prone to inherently large commission errors. As the ultimate goal of the work described in this paper is the conservation of wildlife habitat, we should seek to minimize the omission error rate even at the expense of increasing the commission error rate. The reason for this would be to preserve as much *potential* habitat for each species as possible, as any increase in omission errors associated with wildlife habitat models will lead to neglecting potential habitat for that species. The omission error rate can be minimized in relation to the commission error by choice of a different threshold for the probability of presence value (see Chapter 3 and Freeman and Moisen 2008b).

Only half of published studies evaluate model performance, many fewer used statistics designed to account for abundance/prevalence (Manel et al. 2001). The models produced in this project are not built with separate training and validation samples. I did not have a large enough sample size, especially in the case of the least abundant species included in this project. I believe it was justified in this case because I am not currently using these data to predict the locations for any real applications. The results shown here are simply used to describe the relative utility of the IFMAP forest inventory database in relation to other sources like field plot vegetation samples.

CHAPTER 3

PATTERNS OF WILDLIFE HABITAT MODEL PERFORMANCE IN RELATION TO THE HABITAT SPECIFICITY AND PREVALENCE OF SPECIES

Introduction

Wildlife habitat models are an important component of ecosystem management and often play a critical role in determining conservation priorities and making land management decisions. They are vital to land managers who must perform conservation activities with limited information. The accuracy of wildlife habitat models is a popular area of study, and ecologists strive to improve the quality of these models by improving statistical methods (Elith et al. 2006, Hernandez et al. 2008), using more detailed environmental predictors (Gottschalk et al. 2005, Bergen et al. 2007), using better methods for testing model quality (Manel et al. 2001, Vaughan and Ormerod 2005), optimizing the spatial scale of vegetation samples (Karl et al. 2000, Lawler and Edwards 2006), accounting for spatial artifacts (Segurado et al. 2006, Bahn and McGill 2007), and more (Araujo and Guisan 2006).

The success of many conservation activities is closely tied to the proper application of wildlife habitat models (Fahrig 2001, Jetz et al. 2008). Successful application of models can be disrupted by poor-quality input data, a lack of understanding of species ecology leading to incorrect model design, or by inappropriate conclusions from model results (Austin 2007). For example, using a model to predict the distribution of birds during the breeding season may give good results for abundant species, but the same technique might fail (result in very inaccurate predictions) for rare species. The optimal set of input data may be very different for different species (e.g. habitat generalists and specialists). Sometimes, models built for use as a coarse-filter description of potential habitat distribution (like GAP) are incorrectly applied as a prediction of species occurrence (Scott et al. 1993). Problems such as these can be avoided if we refine our understanding of the relationship between wildlife habitat model accuracy and species traits such as prevalence, habitat specificity, and detectability (Seoane et al. 2005b). A better understanding of the inherent relationships between species ecological traits and model performance would provide a basis for properly (and more successfully) implementing wildlife habitat models for use in monitoring and decision-making (McPherson and Jetz 2007).

There are many examples that illustrate the relationship between species ecology and the ability to predict distribution of habitats and species occurrence. Rarer species tend to result in models with lower accuracy than abundant species for many reasons, in part because of uncertainty due to small sample size (Karl et al. 2002), and in part due to ecological reasons like the more frequent local extinctions associated with metapopulation dynamics (Storch and Sizling 2002). Species that have greater specialization with measureable environmental characteristics are more accurately modeled than generalists because statistical models are able to discriminate between used and unused sites (Seoane et al. 2005b, Tsoar et al. 2007). Generally, the more environmental variables included, the better the model performance (sometimes due to the inclusion of an important limiting resource, sometimes due to a serendipitous correlation unrelated to species ecology). Ideally, model variables should be chosen to reflect specific habitat cues that are important for the group of species included in the study. It is hard to know which variables are important, and it is impossible to include everything, even expert opinion may not provide useful information (Seoane et al. 2004b, 2005a).

In addition to questions about model construction and application, predictive wildlife habitat models are susceptible to many sources of error and uncertainty that must be carefully considered and accounted for. The habitat resources that limit the occurrence of a species can vary in different parts of its range, and its absence in a location can be due to many factors (in addition to habitat associations) including; inter and intra-specific competition (Whittaker and Levin 1975, Herzog and Kessler 2006), population abundance or conservation status (Linder et al. 2000, Hepinstall et al. 2002), dispersal and site fidelity (Knick and Rotenberry 2000, Pulliam 2000, Mortberg 2001), and more (Guisan and Thuiller 2005). Species location data is typically sparse, leading to a high rate of sampling error and the inability of statistical methods to fit models to sparse data (Araujo and Guisan 2006). In predictive habitat mapping science there is a large population of statistical models to choose from. The various statistical approaches have been evaluated in numerous studies (e.g. Segurado and Araujo 2004, Elith et al. 2006, Austin 2007), and no method has emerged as the single best one (Seoane et al. 2005b). In fact, model quality appears to be more dependent on characteristics of the species being modeled and especially the detail of the environmental input data than choice of algorithm (Guisan et al. 2007). In this paper, I used recursive partitioning (RPART), which performs well in comparison to most other statistical wildlife habitat models and provides a flexible and easily interpretable method for linking vegetation data with species occurrences (Segurado and Araujo 2004, Prasad et al. 2006). Recursive partitioning models are transparent and easy to interpret, and the structure of their classification rules are similar to expert-based habitat definitions.

Most wildlife habitat models assign a likelihood of occurrence (or detection or presence) for each species to each site in a field sample. When sites (or habitat classes or a gradient in some other vegetation characteristic) are ranked and plotted against the probability of occurrence, the result is a declining function representing the likelihood of the species' presence on each habitat type (dotted line in Figure 3.1). The shape of this function differs for each species, and in reality is unknown. The purpose of wildlife habitat models is to estimate this unknown relationship. For the most common species there may be a high probability of occurrence across a large portion of the habitats (as in Figure 3.1a). Less common species may show a lower probability of occurrence on a smaller proportion of the habitats (as in Figure 3.1b), or alternatively a high probability of occurrence on a very small proportion of the habitats. The prevalence of each species is shown by the area under the curve of the likelihood of occurrence function (dotted lines in Figure 3.1). Habitat specialists will be associated with a narrower range of habitats than generalists, and in combination with high or low prevalence these ecological characteristics of each species will largely determine the success of habitat models.

Wildlife habitat models attempt to fit the likelihood of occurrence function as closely as possible, based on the habitat information (vegetation and environmental measurements) that is provided to them. The simplest models are binary (like GAP). The shaded area in Figure 3.1 represents the set of habitats that the model (in this case a hypothetical one) identifies as appropriate for that species. The area of the plot where the prediction surface (shaded area) overlaps the occurrence function reveals the correct presences, while the area above the curve reveals the incorrect presence predictions (commissions). The breadth of the model prediction surface has a large effect on the proportion of omissions and commissions. If a larger threshold in probability of occurrence is chosen (for example 0.5: Figure 3.1c, 3.1d), this defines the boundary of the prediction surface and can have a large effect on the number of omissions and commissions. Comparing Figures 3.1a and 3.1c, the narrower prediction surface (defined by the probability of occurrence threshold = 0.5) has the result of decreasing the number of commissions, but increasing the omissions. Figures 3.1b and 3.1d show a similar situation for a rare species. In this case, the use of the 0.5 threshold results in a very large proportion of all the habitats occupied by this species to be predicted as absent. Despite being a very commonly used (i.e. default) threshold value, 0.5 is rarely appropriate for use with a large set of species with varying prevalence and ecological traits. A threshold tied to each species' prevalence is a better approach (Freeman and Moisen 2008b).

Most wildlife habitat models give a continuous probability of occurrence value to each habitat instead of binary as shown in Figure 3.1. These models have the potential to fit the probability of occurrence function more closely. But the use of a probability threshold is still necessary to calculate many model accuracy measurements – like percent correctly classified (PCC), kappa, omission and commission error, sensitivity, and specificity – that are calculated from an error matrix (Table 3.1). Kappa and some other metrics [like the true skill statistic (Allouche et al. 2006)], are designed to correct for bias induced by large differences in the number of presences and absences between species, but these measures are still very sensitive to the choice of a threshold in probability of occurrence that defines the boundary between presence and absence. For this reason, threshold independent measures, like area under the receiver-operator characteristic curve (ROC/AUC) have become popular, but still are susceptible to problems (Lobo et al. 2008).

No accuracy measure is best in all situations. For a successful conservation project, wildlife habitat model users need to be aware of not only the weaknesses of each accuracy measurement, but also of the types of errors that are likely to result from building models for species with particular ecological traits and prevalence. In this study, I attempt to link wildlife habitat model performance with the specificity of species' habitat associations, and prevalence. The results should help modelers know what to expect, in terms of model quality and accuracy, from their particular data.

Methods

The study area is located in the Lower Peninsula of Michigan, which is separated into two ecoregional divisions (Albert 1995). At approximately the midpoint north-south there is a border between the Laurentian Mixed Forest Province to the north, and the Eastern Deciduous Forest Province to the south. In 2005 a survey crew visited five locations in the northern Lower Peninsula, and in 2006 and 2007 six locations in the southern Lower Peninsula were sampled. In each unit (~2000-3000 acres), thirty randomly distributed plots per year were sampled for birds and vegetation within a 50m radius of the plot center. The complete dataset consists of 393 locations where both vegetation and birds were sampled. At each of these locations I also have stand-scale vegetation measurements from a statewide forest resource database. A more detailed description of the study area can be found in Chapter 2.

At each site I calculated a set of habitat conditions (Table 3.2) from both the plot and stand-scale vegetation measurements, and these data are compared in this study to examine the effect of resolution on the habitat association models. The choice of variables to include in this dataset was large. In this particular resource database (MDNR 2005) each measurement unit included the size and cover of each canopy species; the size, density, and height of sub-canopy species; dominant ground cover; and stand-scale variables such as basal area, presence of slash, overall size, land cover type/vegetation cover class, management type (plantation, even or uneven aged), upland or lowland, and canopy closure.

The list of species included in this analysis was reduced to include only those species that are likely to be observed in field surveys, i.e. eliminating nocturnal and non-vocal birds, and with high enough prevalence to produce statistical habitat models (Table 3.2). These species are found in a variety of upland, lowland, forest, and non-forest habitats. I simplified the recorded abundance of each species at each site into the binary variable of presence/absence.

Wildlife habitat models were generated with a statistical algorithm known as recursive partitioning (Feldesman 2002), also known as classification and regression trees or CART (De'ath and Fabricius 2000). Recursive partitioning models were run using the 'RPART' module (Atkinson and Therneau 2000) in R version 2.8.0. Recursive partitioning is a classifier that iteratively divides the samples by selecting a cutoff value for a single variable that separates samples into increasingly homogeneous groups (Segurado and Araujo 2004, Prasad et al. 2006). I used recursive partitioning to predict each species' probability of presence at each sample location, and compared these predictions to the field observations. I sought to restrict the complexity of the models to prevent overfitting the habitat measurements by limiting the number of splitting levels to four and I used leave-one-out cross-validation to prune unsupported branches (Anderson and Burnham 2002). This resulted in trees that have a maximum of 16 habitat groups (defined by recursive partitioning tree end nodes).

Each model was evaluated using multiple statistical criteria (Table 3.1). I show the results for omission/commission error, kappa, and area under the curve of the receiver-operator characteristic plot (ROC/AUC). All the accuracy measures except ROC/AUC require using a 2x2 error matrix (actual presence/absence vs. predicted presence/absence). The construction of error matrices requires that a response value cutoff (threshold in probability of occurrence that separates presence from absence) be set so that sites are binned into the binary presence/absence categories. I used two thresholds, the first sets the threshold equal to each species' prevalence. Species with lower prevalence (rarer species) will therefore have a lower threshold and will include relatively more sites in the predicted 'present' category (and potentially also more commission errors). The second threshold is set at whatever value makes the model's predicted prevalence equal to the observed prevalence for each species, a method supported by Freeman and Moisen (2008b). In this case, the quality of the model has more to do with threshold value than does species prevalence.

Kappa accounts for large differences in the number of sites in the present and absent categories (Karl et al. 2000, Manel et al. 2001) and reflects the improvement over a random distribution among the categories. But kappa is calculated from the error matrix, and therefore still relies on choice of threshold values. To provide a threshold independent measure of accuracy I used ROC/AUC (Fielding and Bell 1997, McPherson et al. 2004). In general, kappa and ROC/AUC are highly correlated, but ROC/AUC is more apt to represent the accuracy of models built for less prevalent species (Allouche et al. 2006).

Two models were built for each species, one with plot-scale vegetation measurements and one with stand-scale data. The accuracy measures were averaged over all 30 species to evaluate overall patterns, and detailed model results are shown for four individual species that differ in prevalence and habitat specificity (Table 3.3). Ovenbird is the most prevalent bird in this sample, American Robin is a prevalent bird that is a decicuous forest generalist, Black-throated Green Warbler is a low-prevalence bird in this sample that has specific habitat associations (deciduous forest with conifer understory), Yellow-billed Cuckoo is a low-prevalence bird associated with mixed (especially edge) habitats. For each of these species I show detailed model results, and link the model to a conceptual diagram that displays model structure in relation to measures of accuracy.

Results

When the probability of occurrence threshold for binning a site in the present category is equal to each species' prevalence, the average kappa values were similar for both the stand and plot-scale models (kappa=0.37 and 0.39, respectively). The majority of the species models in each set (22/30 for stand-scale and 27/30 for the plot-scale data) scored 0.3 or better (Figure 3.2a and 3.2b). The only species that showed a large difference between the stand and plot-scale models (defined as a difference between the models of kappa >= 0.15) were mostly mixed habitat species, including Field Sparrow (mixed habitat guild), Yellow-billed Cuckoo (mixed habitat guild), Nashville Warbler (forest habitat guild, but is also found in shrub habitats), Gray Catbird (mixed habitat guild), and Northern Flicker (mixed habitat guild). ROC/AUC supports this result, the largest differences in AUC values between stand and plot-scale models were for Cedar Waxwing (mixed habitat guild) and Northern Flicker. For all of these species the plot-scale models were more accurate.

When the threshold that sets predicted prevalence equal to actual prevalence is used, the results are similar (Figures 3.2c and 3.2d). The overall difference between stand and plot-scale models is slightly smaller (kappa = 0.39 and 0.40 respectively), and there are fewer species that show a large difference between the stand and plot-scale models (Northern Flicker, Tufted Titmouse, and Field Sparrow), but again all were of the mixed habitat guild.

The association between species prevalence and model prediction accuracy shows mixed results. With both of the thresholds, kappa shows no correlation for either the stand-scale vegetation models (Figures 3.2a and 3.2c, $R^2 < 0.05$) or the plot-scale models (Figure 3.2b and 3.2d, $R^2 < 0.01$). ROC/AUC, however, appears to show an increase in model accuracy with decreasing species prevalence for both the stand-scale (Figure 3a, $R^2 = 0.10$) and the plot-scale vegetation data (Figure 3.3b, $R^2 = 0.27$) although the correlations are weak.

Threshold choice had a minimal effect on kappa (see above), but a large effect on commission and omission error rates. With the threshold equal to prevalence for each species (Figures 3.4a and 3.4b), commission errors increase consistently with declining prevalence for both the stand-scale (Figure 3.4a, $R^2 = 0.47$) and plot-scale (Figure 3.4b,

$R^2 = 0.57$) models. Omission errors decrease slightly as prevalence declines, with no correlation ($R^2 = 0.02$ for stand-scale models, $R^2 = 0.04$ for plot-scale). Nearly all of the models showed lower omission error rates than commission error. With the threshold set where the predicted prevalence equals the actual prevalence for each species (Figures 3.4c and 3.4d), the trends associated with prevalence are much less distinct. In addition, there is less of a difference overall between commission and omission error rates. When the 0.5 threshold in probability of occurrence is used, the results change drastically due to some low prevalence species showing 0% commission and 100% omission error rates (i.e. the models predicted no presences).

Looking at kappa and ROC/AUC, Ovenbird and Black-throated Green Warbler had relatively accurate models for both the stand and plot-scale vegetation data, while American Robin and Yellow-billed Cuckoo had relatively inaccurate models (Table 3.4). This result fit with groups of species based on habitat specificity. Specialists showed a significantly higher accuracy with both kappa and ROC/AUC for the stand-scale vegetation measurements, but ROC/AUC was not significant for plot-scale models. Choice of threshold value had a small effect on kappa (exceptions to this are the stand-scale models for Yellow-billed Cuckoo, and the plot-scale models for Black-throated Green Warbler). ROC/AUC is threshold independent, but there are relatively large differences in commission and omission errors between the two threshold choices (Tables 3.4a and 3.4b). Commission error is higher when threshold is equal to species prevalence (Table 3.4a), and omission errors are higher when threshold is the value that sets the model's predicted prevalence equal to the actual prevalence of each species. There did not appear to be a relationship between habitat specificity and omission and commission errors (the differences between generalists and specialists are not significant). Commission errors appear to be more correlated with species prevalence (Figures 3.4a-d) than with habitat specificity.

Recursive partitioning models (using plot-scale vegetation measurements) for each of the four species are represented in Figure 3.5 (in a manner similar to Figure 3.1). Each vertical box shows one habitat class (group of sample sites with similar habitat features) generated by the recursive partitioning tree, and the width of each box is proportional to the number of samples included in that class. The classes are ranked along the horizontal axis in order of decreasing predicted presence value, shown on the vertical axis. Dark shaded portions of each box represent the predicted presences (which is equal to the proportion of presences observed in the data), and light shaded areas represent absences. Threshold values for generating the error matrix are the same as described previously, and are shown graphically by using dashed and dotted lines. The first is set at the prevalence of each species (along the vertical axis, dotted line), the second at the point where observed prevalence equals predicted prevalence for each species (along the top horizontal, dashed lines).

The more accurate models (Ovenbird [Figure 3.5a] and Black-throated Green Warbler [Figure 3.5d]) have a smaller proportion of their habitat classes near intermediate (0.5) values in predicted presence probability, while the less accurate models (American Robin [Figure 3.5b] and Yellow-billed Cuckoo [Figure 3.5c]) have a relatively large proportion of predictions at intermediate values. The two methods of selecting threshold values result in a large difference for the less prevalent species while the more prevalent species have thresholds at similar values (as noted previously and shown in Table 3.4). For American Robin (Figure 3.5b) both thresholds give exactly the same model results, but for Yellow-billed Cuckoo (Figure 3.5c) and Black-throated Green Warbler (Figure 3.5d) the two thresholds result in drastically different prediction surface. The prediction surface is only slightly different for Ovenbird (Figure 3.5a) with the two thresholds.

The second method of selecting threshold values (where the model's predicted prevalence = the actual prevalence of each species) appears to set the number of omissions and commissions very close to equal, while the first method (threshold = species prevalence) tends to minimize omissions (Table 3.4, Figure 3.4). With the threshold = species prevalence, the increase in commission error rates with less prevalent species (Figures 3.4a and 3.4b) can be explained by the fact that there tend to be more predicted presences for rare species than with the second threshold, and a larger proportion of these are incorrect ("b" in Table 3.1).

Discussion

It is important for users of wildlife habitat models to fully understand the methods of accuracy assessment they are using in conservation projects. But it is difficult to grasp the meaning of accuracy measurements without a solid understanding of how the outputs of a statistical model relate to the calculations inherent in the accuracy assessments. I have attempted to show this relationship graphically in Figure 3.1 (hypothetical binary models) and Figure 3.5 (actual recursive partitioning model output). In both figures, the horizontal axis represents a ranked list of habitats, arranged in order of highest (at the left) to lowest (at the right) quality. In Figure 3.1 the horizontal axis is gradient made up of all possible sample sites, in Figure 3.5 these are groups of sample sites with similar vegetation measurements (binned by the recursive partitioning model). The vertical axis refers to the probability that a species will be present. In Figure 3.1 it is the theoretical probability of occurrence which takes into account the influence of population size, growth, dispersal ability, fidelity, and site history. In Figure 3.5, the scale of the vertical axis is the predicted probability of presence for each group of sample sites output by the recursive partitioning model. This probability is the actual proportion of samples within each group where the species was recorded as present, and therefore includes all of the same influences as Figure 3.1 (population vital rates, fidelity, site history) plus detectability of the species. Detection probabilities vary quite drastically by species, and add a significant amount of uncertainty to the probability of occurrence (or predicted presence probability) values.

What some have considered relatively unimportant details or arbitrary assumptions (e.g. the choice of threshold value for converting continuous predictions into a binary error matrix) can have large effects on the results and/or interpretation. Kappa, omission error, and commission error are all calculated by selecting a threshold probability of occurrence value (for example 0.5, Figures 3.1c and 3.1d) that determines where the prediction surface (shaded areas in Figure 3.1) ends. This essentially forces a continuous prediction scale into a binary one, with the result of treating all the predictions of probability of occurrence less than 0.5 as 'absent' (see shaded areas and the threshold [dashed line] in Figure 3.1c and 3.1d). An effective method is to set this threshold dependent on each species' prevalence (2008b), as has been done in this study, and to understand that species' ecological characteristics may affect the quality of wildlife habitat models.

I have summarized some patterns of wildlife habitat model utility and have given specific examples. Wildlife habitat models for habitat specialists can be inherently more accurate than generalists because statistics can more easily define habitat classes that clearly delineate appropriate from poor habitats (given the limitations of the resolution, extent, and accuracy of vegetation resources measurements included in the model). The detail and scale of habitat model inputs plays a large role in the ability to accurately predict species locations. Stand-scale vegetation measurements (in comparison to plot-scale measurements) may not be as appropriate for describing edge and mixed habitat associations, but tend to be well suited to other species and have the added advantage of providing the possibility of calculating spatial pattern information (not included in this study).

The correlation between model accuracy and prevalence is typically positive (e.g. Vaughan and Ormerod 2005) since rarer species are simply less likely to be present on any given location of appropriate habitat (Manel et al. 2001). The challenge in building wildlife habitat models is to predict appropriate habitat sites with a high probability of presence, but this can be difficult because of the fact that rare species may actually have a low probability of being detected on even the best sites. Accuracy measures fail to take this into account and therefore may not reflect an actual measure of model quality, but instead an inability to account for uncertainty in species occurrence. The uncertainty in species occurrences can be due to a myriad of ecological reasons unrelated to habitat

(Storch and Sizling 2002), as well as detectability (MacKenzie et al. 2005, Royle et al. 2005).

Most common accuracy measures, including kappa when a 0.5 threshold is used, are relatively inflexible to the species' ecological characteristics (such as habitat specificity and prevalence). Kappa, however, can be used effectively when the choice of threshold in probability of occurrence is flexible, and tied to each species' prevalence (2008b). In this study, the relationship between prevalence and model accuracy was weak (kappa, Figure 3.2) or even negative (ROC/AUC, Figure 3.3).

For wildlife habitat models there is a tradeoff between sensitivity and specificity (Allouche et al. 2006). Sensitivity is the probability of correctly classifying a presence, specificity is the probability of correctly classifying an absence (Table 3.1). By changing the probability of occurrence threshold to increase one, the other declines. ROC/AUC assesses model accuracy across all values of sensitivity/specificity and therefore is no affected by threshold choice, but kappa can change drastically (Allouche et al. 2006, Freeman and Moisen 2008b). The proper choice of threshold values can optimize the specificity vs. sensitivity tradeoff, even when using kappa, but see Manel et al. (2001). The two threshold values used in this study had relatively small effects on kappa (Figure 3.2, Table 3.4), but large effects on commission and omission error rates (Figure 3.3, Table 3.4).

All of these models, except for species with very specific habitat associations, are prone to inherently large commission error rates. As the ultimate goal of all the work described in this paper is the conservation of wildlife habitat, it may be desirable to minimize the omission error rate even at the expense of increasing the commission error rate. The reason for this would be to preserve as much *potential* habitat for each species as possible, as any increase in omission errors associated with wildlife habitat models will lead to neglecting potential habitat for that species (Wilson et al. 2005). If this is a desirable condition of a wildlife habitat modeling project for a large set of species, then the threshold for considering a location as appropriate habitat should be equal to the prevalence of the species (the first threshold used in this study). When this is the case, commission error rates increased with less prevalent species, but omission errors were low across all species (Figures 3.4a and 3.4b). However, when resources are limited and only a small set of locations can be targeted for conservation, then a different approach may be necessary so that the most important locations are protected.

When a model performs poorly, it can be due to many factors, and a poor quality model may not in fact be useless. Some species have a low prevalence across their range and are not likely to have a high probability of occurrence at any given location. For example, species that show low site fidelity and instead are nomadic or focused on a spatially patchy/irruptive food resource may be recorded in many different locations over different years, but still within similar habitats. In this case a model that does not incorporate the food resource will be unable to predict those locations year to year. Accuracy measures will show that this model performs poorly, but in fact it does a very good job of describing the distribution of appropriate habitat. This model, with very poor accuracy, may in fact be very useful. Other examples of important vegetation and environmental measurements may be immeasurable and therefore will not be included in model construction, but if this is known beforehand the model may still be useful depending on the application. Often, higher-level effects such as community interactions, predation, and inter-specific competition can add variation to species distributions that habitat models cannot track. Species undergoing rapid population increases or declines can be difficult to model accurately, but if these patterns are understood then steps can be taken to make interpretation of model results more feasible.

CHAPTER 4

INFLUENCE OF VEGETATION CLASSIFICATIONS ON WILDLIFE HABITAT MODEL PERFORMANCE

Introduction

Wildlife habitat models are an important component of ecosystem management and often play a critical role in determining conservation priorities and making management decisions. They are vital to managers who must perform conservation activities with limited information. The accuracy of wildlife habitat models is a popular area of study, and ecologists strive to improve the quality of these models by improving statistical methods (Elith et al. 2006, Hernandez et al. 2008), using more detailed environmental predictors (Gottschalk et al. 2005, Bergen et al. 2007), using better methods for testing model quality (Manel et al. 2001, Vaughan and Ormerod 2005), optimizing the spatial scale of vegetation samples (Karl et al. 2000, Lawler and Edwards 2006), accounting for spatial artifacts (Segurado et al. 2006, Bahn and McGill 2007), and more (Araujo and Guisan 2006).

Frequently the availability, rather than the suitability, of environmental and habitat information is the determining factor as to whether a predictor variable will be included in a wildlife habitat model (Roloff et al. 2008). Often, the only habitat data that are available for modeling wildlife distributions are spatial classifications of vegetation or land cover (from aerial or satellite imagery). The number and type of classes in these data are not necessarily determined by their appropriateness for wildlife habitat modeling (i.e. the perceived differences in habitat types by each species), but instead by the limitations of a satellite image classification technique and/or the perception of land management professionals. The effects of vegetation classification system design (specifically the number and type of classes) on wildlife habitat models is an underrepresented topic in the large volume of literature on wildlife habitat modeling (Scott et al. 2002). As an artifact of the statistics used to build wildlife habitat models, more classes will often lead to better model fit, but it is difficult to determine the effects of altering the arrangement of samples among classes (i.e. changing the class definitions) on model results.

Terrestrial vegetation classification has a long history in ecological theory (Watt 1947, Kuchler 1951, Daubenmire 1952, Grime 1974). Classifications have taken many forms, from *a posteriori* statistical analysis of field measurements (Bray and Curtis 1957, Greig-Smith et al. 1967), to *a priori* potential climax vegetation community states (Pfister and Arno 1980, Cook 1996), to large scale ecoregional assessments (Bailey 1983). If the environmental gradients, disturbance dynamics, and management goals are properly weighed, then environmental classifications can be very useful in a wide variety of conservation projects (Bourgeron 1988). Otherwise classifications can suffer from a lack of rigor (they are not valid outside a limited area), or they will not represent real ecological processes and landscape dynamics like successional trends and the distribution of species along the ecological gradients in a particular location (Grossman et al. 1999).

A posteriori classifications may be more susceptible to these issues because they are dependent on the quality of the field samples used in their construction. Statistical clustering algorithms also tend to minimize the within-class variability and maximize the between-class variability, which may mask true ecological processes that represent themselves as very fine patterns. It is possible that *a priori* classification systems could identify these less obvious patterns and provide a more accurate representation of landscape dynamics. Modern vegetation classification systems are often a mix of *a priori* class selection and *a posteriori* statistical analyses (Grossman et al. 1999). Even if a classification system achieves all of the requirements listed above and is an accurate representation of the environmental gradients and landscape dynamics of the region, the vegetation classes and spatial patterns of their distribution on a landscape may not be the same as those perceived by wildlife species (i.e. may not reflect the factors that represent limiting resources (O'Connor 2002)).

As wildlife habitat modeling and other conservation projects are implemented there are a multitude of choices that must be made as to the specific components that will be included in the models, not to mention the sources of these data. These choices go hand in hand with the limitations provided by research budgets and the difficulty (cost) of acquiring more detailed and accurate data. Typically the independent variables will consist of environmental data that may include categorical habitat classes, vegetation or substrate measurements, and climate or other abiotic features. These can be generated in any number of ways, from classified satellite imagery to intensive field samples.

Although it seems clear that systematically collected forest inventory data, including both vegetation composition and structure, can have significant value in developing wildlife habitat models (Karl et al. 1999, Welsh et al. 2006), it is uncommon to include these data in models of wildlife habitat distribution (Flather et al. 1992, Imhoff et al. 1997, He et al. 1998, Osborne et al. 2001, Heikkinen et al. 2004, Seoane et al. 2004b). In many cases, the detailed vegetation information that would improve the accuracy of predictive wildlife habitat models are simply not available without intensive field sampling, and so modelers rely solely on vegetation classifications. Depending on the level of detail included in the classification this could be an appropriate strategy, but it depends on the biological characteristics of the species in question and the purpose of the modeling project. For regional assessments of potential habitat distribution (like GAP), a simple land cover classification may be appropriate. But land cover based habitat assessments like GAP are frequently, and inappropriately, applied to local conservation projects or resource management decisions (Noon et al. 2003).

In a previous study examining the accuracy of GAP models in Michigan (MIGAP, Donovan et al. 2004) I found that MIGAP models overestimated the amount of available habitat for most species. When treated as a prediction of presence/absence, the MIGAP models result in high rates of commission error (predicted present but not detected) but low omission error rates (predicted absent when actually detected). Of the many possible reasons for these errors, two are most likely. First, the landscape-level land-cover maps derived from satellite image classifications (MDNR 2001) contain relatively broad vegetation cover classes with no categories for mixed deciduous and conifer forest, which are abundant in the western Great Lakes landscape. Therefore, land cover maps may not have the spatial accuracy or vegetation description detail necessary for revealing an accurate distribution of habitats for many species, so MIGAP models typically err on the side of including areas with even a small chance of species occurrence. Second, published accounts of wildlife-habitat relationships are in many cases not refined enough to describe specific vegetation elements that drive habitat associations, nor are they detailed enough to compensate for the geographical differences in habitat associations across a species' range. Both of these issues result in the inclusion of more locations (as potential habitat) than each species regularly occupies.

The particular class definitions in a classification are important to the accuracy of a wildlife habitat model (Roloff et al. 2008). If the classes are such that used vs. unused habitats are clearly divided for a particular species, then a statistical model will be very accurate. However, a classification with a large number of classes will be likely to predict species presence more accurately than one with fewer classes, simply by chance and the ability of because statistical algorithms. It may be difficult therefore to determine whether it is the quality of the class definitions or the number of classes that leads to differences in wildlife habitat model accuracy between two vegetation classifications.

There are many examples that illustrate the relationship between species ecology and the ability to predict distribution of habitats and species occurrence. Species that have greater specialization on measureable environmental characteristics are more accurately modeled that generalists because statistical models are better able to discriminate between used and unused sites (Seoane et al. 2005b, Tsoar et al. 2007). Rarer species are typically associated with less accurate habitat distribution models than are abundant species. This pattern can result from sampling issues (Karl et al. 2002), or for ecological reasons like the more frequent local extinctions associated with the metapopulation dynamics of less common species (Storch and Sizling 2002). Generally, the more variables included, the better the model performance, but care should be taken to avoid spurious relationships due to chance. Ideally, these variables should be chosen to reflect specific habitat and environmental cues (potentially limiting resources) that are important for the group of species included in the study. But it is difficult to avoid inclusion of characteristics that are in fact arbitrary, even expert opinion may not provide useful information (Seoane et al. 2004b, 2005a).

In this study, I compare the relative accuracy of wildlife habitat relationship models built with three different hierarchical vegetation classifications. The first classification was developed for a statewide forest resource inventory database, the second is a statistically fit set using the first as training data, and the third comes from an unsupervised clustering routine. I compare the effects of the classification system, in particular the definition and number of classes, on the success of wildlife habitat models.

Methods

The study area is located in the Lower Peninsula of Michigan, which is separated into two ecoregional divisions (Albert 1995). At approximately the midpoint north-south there is a border between the Laurentian Mixed Forest Province to the north, and the Eastern Deciduous Forest Province to the south. The northern landscape is primarily forested, with a wide variety of coniferous and deciduous species present, and the southern landscape is primarily an agricultural matrix with pockets of deciduous forest, largely in riparian and wet areas not suitable for agriculture (MDNR 2001). In 2005 a survey crew visited five locations in the northern Lower Peninsula, and in 2006 and 2007 six locations in the southern Lower Peninsula were sampled. In each unit (~2000-3000 acres), up to thirty randomly distributed plots were sampled each year for birds and vegetation within a 50m radius of the plot center. The complete dataset consists of 460 locations where both vegetation and birds were sampled. A more detailed description of the study area and methods can be found in Chapter 2.

To determine the potential influence of habitat classification type and number of classes, I developed three different classifications, each with three levels (defined by the number of classes). The baseline vegetation classification was developed *a priori* by the Integrated Forest Monitoring, Assessment, and Prescription (IFMAP) program in Michigan as a result of a process involving foresters and ecologists (MDNR 2004, 2005). IFMAP is a geographic decision support system (DSS) that tracks stand-level forest composition and structure for state-owned lands throughout Michigan, and contains detailed vegetation information on non-forested areas. The vegetation classification system is a combined physiognomic and floristic hierarchical design with each level separating finer classes, similar to Anderson et al. (1976). As it is designed primarily for forestry purposes, there are more forest vegetation classes (70+) than open land or wetland classes (~40) at the finest level of the classification. The logical structure of the classification consists of a series of many 'IF-THEN-ELSE' decisions which bin every location into a single class. Each decision is made based on the presence or amount of abiotic features or plant cover.

I manually assigned one hierarchical class value to each of the 460 field plots based on the IFMAP decision rules as applied to the field vegetation measurements. These samples resulted in 52 level-four, 16 level-three, and 9 level-two classes (Table 3.1). The distribution of sites among the classes is moderately skewed, with the largest five level-four classes comprising 25% of all the sites, and 22 classes made up of only five sites or fewer (average number of sites per class = 8.9, s.d. = 7.6).

The predicted classes were generated with a statistical algorithm known as recursive partitioning (Feldesman 2002), also known as classification and regression trees or CART (De'ath and Fabricius 2000). Recursive partitioning models were run using the 'RPART' module in R (Atkinson and Therneau 2000). One of the advantages of recursive partitioning is the similarity between the decision rules used to classify sites in the IFMAP database and those generated with recursive partitioning. The training data for the predicted vegetation classes were the field plot class assignments as the dependent variable and a large set of vegetation measurements as the independent variables. The vegetation measurements included three site descriptors, eleven calculated composition and structure variables, and the percent cover of each canopy tree species within the plot boundary (47 species were recorded over all the field plots). The recursive partitioning models were then used to predict the vegetation class for each site, and this set of predictions (one for each level from 2 through 4) is used as the second classification in wildlife habitat model comparisons.

The third vegetation classification used in the wildlife habitat model comparisons is a set of classes assigned by a 'partitioning around medoids' cluster analysis, calculated with the 'Cluster' module in R (Maechler 2008). This approach allows the user to choose a number of desired clusters (k), then the algorithm chooses the k representative samples (medoids) that minimize the stress of the final clustering based on a dissimilarity matrix of all the plots. The dissimilarity matrix was generated with the Bray-Curtis distance measure. I assigned all field plots to a 9, 16, and 52 group classification to compare with the actual and predicted level-2, 3, and 4 classifications.

The list of bird species included in this analysis was reduced to include only those species that are likely to be observed in field surveys, i.e. eliminating nocturnal and non-vocal birds, and abundant enough to plausibly calculate a statistical habitat model (present at > 5% of sites). These 30 species represent a variety of upland, lowland, forest, and non-forest habitats. I simplified the recorded abundance of each species at each site into the binary variable of presence/absence. Recursive partitioning was used again to predict each species' probability of presence at each sample location, and accuracy measures were calculated by comparing these predictions to the field (presence/absence) observations. A more detailed description of the model construction methods is covered in Chapter 2.

I show the results for omission/commission error, kappa, and area under the curve of the receiver-operator characteristic plot (ROC/AUC). All the accuracy measures, except ROC/AUC, require using a 2x2 error matrix (actual presence/absence vs. predicted presence/absence). The construction of these error matrices required that a response value cutoff (probability level that separates presence from absence) be set so that the sites were classified into the binary presence/absence categories. I used a unique threshold for each species that sets the predicted prevalence of the recursive partitioning model equal to the observed prevalence for that species. This method is supported by Freeman and Moisen (2008b). Kappa accounts for large differences in the number of sites in the present and absent categories (Karl et al. 2000, Manel et al. 2001) and reflects the improvement over a random distribution among the categories. To provide a threshold independent measure of accuracy I used ROC/AUC (Fielding and Bell 1997, McPherson et al. 2004). In general, kappa and ROC/AUC are highly correlated, but ROC/AUC is more apt to represent the accuracy of models built for less prevalent species (Allouche et al. 2006). I averaged error and accuracy measures over all 30 species and tested for significance between means with a paired t-test (species defined the pairings).

Even with large within group variation, a paired t-test can produce a significant result by a consistent in crease or decrease in value for each pair between two samples.

Results

The IFMAP vegetation classification system is a detailed hierarchical assembly of over 115 land cover and vegetation classes (Table 4.1, part 1). Of these, more than 70 can be characterized as forest vegetation classes. When the IFMAP classes are compared to the recursive partitioning predictions, there is little support for the number of classes, especially at levels 3 and 4. 7 of 9 level-2 classes, 10 of 16 level-3 classes, and 21 of 52 level-4 classes were retained in the recursive partitioning classification (Table 4.1, part 2). The disagreement between these two classifications is not dominated by any one general (level-1) vegetation or land-cover type. In fact, all of the level-1 classes show a similar reduction in the number of predicted classes from the IFMAP training data. The classes that tend to be eliminated in the predicted classification are the least frequent IFMAP classes in the dataset.

Despite low levels of class representation in the predicted classification, the recursive partitioning predictions were fairly accurate. 85% of the sites were classified correctly at level-2 (kappa = 0.83), 75% at level-3 (kappa = 0.73), and 53% at level-4 (kappa = 0.51). Given that many of the classes were not retained between the original IFMAP classification and the prediction classification, and that the kappa values are quite high, the recursive partitioning classification seems to support (at least the major classes of) the IFMAP classification system.

There is no direct way to compare the IFMAP classification with the cluster analysis classes, but some summary statistics are revealing. The moderately skewed distribution of sites among the IFMAP classes contrasts with a relatively even distribution of sites among the cluster analysis classes for level-2, while levels 3 and 4 show a skewed distribution of sites among classes, very similar to the IFMAP classification. The mean number of sites per class is the same in both classifications at each level, but the standard deviation is much larger in level-2 for the IFMAP classification than the cluster analysis classification, and similar for levels 3 and 4 (Table 4.2).

When looking for agreement between the IFMAP classification and the clusters, only three classes in the level-2 cluster classification have 50% or more of their sites within a single IFMAP classification (herbaceous agriculture, upland shrub, and upland coniferous forest). All of the other IFMAP classes have sites spread out over many clusters. Similarly for levels 3 and 4, many of the IFMAP classes have sites distributed across a wide variety of cluster classes. Six out of 16 level-3 IFMAP classes share more than 50% of their sites with a single cluster class. These are: agricultural crops, oak deciduous forest, aspen deciduous forest, planted pine forest, natural pine forest, and upland mixed forest. The latter three of these were all associated with a single cluster class. At level-4, 16 out of 52 IFMAP classes shared 50% or more of their sites with a single cluster. These were spread out over a wide variety of upland, lowland, forest, and non-forest classes.

There were small differences in overall accuracy of the bird habitat models between the IFMAP and cluster classifications. Within each classification type, the higher levels (more classes) were significantly more accurate over all 30 species (paired t-test, $p<0.05$) than the lower levels (fewer classes). The IFMAP classification resulted in the highest accuracy at every level. The predicted classes showed lower accuracy at level-4 (kappa, Figure 4.1 and AUC, Figure 4.2) but not at level-2 or 3. There were no significant differences in commission or omission errors (Figure 4.3a) between any of the classifications at any hierarchical level. The differences in accuracy between the IFMAP classes and the predicted classes were only significant at levels 3 and 4 (paired t-test, $p<0.05$). The predicted classification had only 21 classes at level-4, compared to 52 in both the IFMAP and cluster classifications (Table 4.1). At level-4, there were more species with relatively accurate models (kappa>0.2, AUC>0.75) using the IFMAP classes than with the predicted or cluster classes (29 for IFMAP vs. 26 for both the predicted and cluster classes). The difference was even larger at level-3 (26 for IFMAP vs. 21 for both the predicted and cluster classes), but there was little difference at level-2 (16 vs. 14 for predicted and 15 for cluster classes).

Discussion

Comparing the two alternate classifications to the IFMAP classification shows a declining level of agreement with increasing number of classes. The disagreement increased both in the number of sites assigned to different classes, and (comparing IFMAP to the predicted class) in the number of classes themselves. Despite this disagreement, the accuracy of bird habitat models increased with higher levels (more classes) of each classification, indicating that number of classes does in fact increase the ability of statistical wildlife habitat models to fit sample data. However, there were significant differences between the classifications (within each level) that indicate the quality and format of the classification can also influence wildlife habitat model performance.

These results show that the IFMAP habitat classification system is as useful, or better than, an *a posteriori* statistical clustering classification for modeling habitat associations of a large suite of bird species. The detail of the forested habitat classes (at level-3 or above) appears to be adequate for describing habitat types used by a set of bird species in the Midwest. The IFMAP data and variables selected for this study are biased towards canopy and forest measurements. If the IFMAP resource database fails to be appropriate for any particular species it would most likely be for open habitat (e.g. grassland and wetland), or mixed habitat and edge species where the compartment based GIS data fail to adequately describe complex ecotonal conditions (Chapter 1).

The major differences between the IFMAP classification system and the two alternatives presented in this study are the number of less frequent classes, and/or the composition (habitat type definition) of the classes themselves. When comparing the IFMAP classification to the predicted classes, the identity and composition of the classes are roughly the same but the least frequent classes are absent in the predicted set. The bird habitat model results show that removing the least frequent classes does have a significant negative effect, but the magnitude of this effect appears to be small (IFMAP vs. predicted, Figures 4.1 and 4.2).

When comparing the IFMAP classification to the cluster classes, the number of classes is the same but the identity of the classes (i.e. distribution of sites among the classes) differs. Looking at the level-3 classifications, there is a significant difference in accuracy (kappa and AUC, Figures 4.1 and 4.2) between the model results, but the difference at level-4 is not significant. This implies that the effect of the distribution of sites among classes on wildlife habitat models accuracy (i.e. using different classification systems) can be offset if you: 1) include a large enough number of classes, and 2) make sure the classes represent some portion of the ecological processes the lead to vegetation community formation (i.e. the classes represent reality).

Whether these results support the use of *a priori*/expert models (like IFMAP) or *a posteriori*/statistical habitat classifications (like the cluster analysis) is not clear. But since the IFMAP classification system led to higher accuracy at every level of the classification and resulted in nearly every species showing more accurate models than with either the predicted or cluster analysis classifications, I feel its use is warranted. Given the long history of study in community ecology, perhaps *a priori* and expert models are more likely to represent the fine scale landscape processes that might be missed with more objective statistical (*a posteriori*) methods of ecological classification. It is difficult to determine whether the IFMAP level-3 classification (16 classes) would be preferable to the level-4 (52 classes) in this context. The notable improvement (kappa = 0.31 for level-3, kappa = 0.36 for level-4) could be the result of the more accurate depiction of vegetation communities across this landscape, thus leading to more accurate predictions, or it could be a statistical artifact of dividing the sites up into arbitrarily small groups thus over-fitting the sample data.

CHAPTER 5

MAPPING FOREST STRUCTURE FROM SATELLITE IMAGERY FOR WILDLIFE HABITAT DISTRIBUTION MODELS

Introduction

Land cover maps derived from satellite imagery are useful in coarse-filter approaches to identifying the distribution of wildlife habitat (Boone and Krohn 2000). The gap analysis program (GAP) is an example of a widely used wildlife habitat monitoring program that has been implemented at a nearly continental extent. The GAP protocol (Scott et al. 1993) relies on a statewide land cover map derived from Landsat satellite imagery (MDNR 2001), and expert-based descriptions of wildlife habitats that define a set of land cover types that are preferred by each species (Edwards et al. 1996). GAP wildlife habitat distribution maps have been used to identify local land units that should receive priority in conservation inventory or management projects (e.g. Rodriguez et al. 2007), and to allocate funding for a given species or habitat (e.g. Kiester et al. 1996). The use of GAP maps is limited to showing the distribution of habitat potential at a coarse (landscape to regional) scale (Edwards et al. 1996). They are not reliable as predictive models of species locations at a local scale, but when they are used as such they are susceptible to large rates of commission error (Peterson and Kluza 2003 and Chapters 2 and 3).

Since land cover maps bin all locations into a set of vegetation or land-cover classes, there is the potential for large variation of habitat characteristics within the classes. Forest vegetation classes would be particularly prone to this variation due to the three-dimensional properties of a tree-dominated plant community. Many wildlife species are dependent on these three-dimensional characteristics of forest habitats for nest locations and feeding platforms, so forest structure measurements could be valuable data for modeling efforts (Karl et al. 1999). The addition of forest structural characteristics to wildlife habitat models based on land cover (like GAP) could lead to more accurate habitat distribution estimates and more successful conservation planning.

In contrast to land-cover data, field-based forest inventory programs often collect numerous plot-level measurements valuable for a wider range of applications. For example, the USDA Forest Service has systematically inventoried forests nationwide since the 1930s under the Forest Inventory and Analysis (FIA) program (Hansen et al. 1992). The information generated from forest inventories forms the basis for developing management policies, habitat protection strategies, and resource utilization decisions. Forest inventory programs like FIA monitor many forest conditions (e.g. timber volume, age distribution), but not in a spatially explicit format (GIS-based compartment and stand records). Managers rely on these data even though spatial conditions such as adjacency can have a significant impact on resource utilization decisions (Borges and Hoganson 1999). There are a few forest inventories maintained as spatially explicit GIS databases by state and federal natural resource agencies. These datasets include several timber production centered parameters that provide useful information for characterizing forest composition and structural conditions within patches (stands), but their spatial extent is limited by jurisdictional boundaries.

Combining plot-level forest inventories and satellite imagery through classification can extrapolate the detail of forest plot surveys across the entire spatial coverage of a satellite image scene, and spatial patterns of forest resources can therefore be assessed. These data could lead to more accurate areal summaries than a randomized plot-level survey, and would allow managers to strategically plan for the spatial distribution of successional and development stages of the forest across the landscape. Despite the fact that the most successful remote sensing examples have been cover type classifications (e.g. Wolter et al. 1995), techniques designed to assess forest structure have been increasingly successful (Wulder 1998, Scarth and Phinn 2000, Moisen and Frescino 2002, Moisen et al. 2006).

One of the largest advantages of remote sensing for forest inventory is that satellite remote sensors are not restricted by jurisdictional boundaries, and therefore can provide a more inclusive estimate of forest resources than institutional inventories. The diversity of spatial scales, temporal reproducibility, and the wealth of information that is possible to glean from remotely sensed imagery make these data very attractive for conservation and wildlife management projects. However, the technical expertise required to develop these data is limiting, as is the quality of the training data required (number and detail of reference plots). In relatively simple and homogeneous forests (e.g. boreal conifers) where classification of forest structure is most successful, there is often a significant correlation between vegetation density and the structure variable being classified (e.g. Turner et al. 1999, Cohen et al. 2003). In other words, "greenness" is proportional to basal area, stem density, or biomass. In the complex and mixed forests that are present in the western Great Lakes, however, relatively simple classification techniques, like regression, may not be effective. Fortunately, many classification algorithms are designed to identify minute spectral differences between land-cover classes and exploit these responses to generate accurate classifications of continuous-scale forest attributes (Wulder 1998, Moisen and Frescino 2002).

Optical and infrared sensors record only the electromagnetic signal that is reflected and emitted from the sum of all targets in a pixel, and cannot penetrate the surface of many targets. In closed-canopy deciduous forests, for example, little of the recorded radiance per pixel is contributed by sub-canopy elements like tree branches, stems, understory plants, and ground cover. Due to this phenomenon, it is unlikely that forest structural characteristics like the diameter of tree trunks, or the height of the canopy will contribute identifiable spectral patterns to satellite image pixels. For example, a closed-canopy *young* maple forest will look very similar to a closed-canopy *mature* maple forest even though structural aspects such as basal area, average stem diameter, biomass, and canopy height may be very different. However, in temperate forests deciduous tree species lose their leaves during the fall and develop anew in spring. Timing the acquisition of satellite imagery during these times allows the sub-canopy elements to contribute to the radiance signal that is recorded in a satellite image (Wolter et al. 1995). The spectral characteristics of many deciduous species' leaves change due to changes in leaf chemistry, and these patterns also help to discern forest community types with satellite imagery (Dymond et al. 2002). In some situations, forest structural characteristics can be strongly associated with particular spectral bands of imagery, like aboveground biomass and NIR (Zheng et al. 2004). Lu et al. (2004) have shown that there are correlations between Landsat TM spectral values and measured forest structure (average stand diameter, average stand height, basal area, and aboveground biomass) in deciduous South American forests.

For assessments of forest cover or other stand-level information, a grain size of 10-100 meters appears to be ideal for aggregating the spectral qualities of tree crowns, canopy gaps, and sub-canopy elements (Wulder and Franklin 2003). Grains smaller than 10 meters are susceptible to being dominated by any individual portion of the target (like shadow, background, or canopy) which means that these targets must be surveyed individually and provided to the classifier. Fine-grained imagery will not adequately reflect the continuous nature of forest stands, but may be more useful for classifying variables related to individual trees, while imagery with pixel dimensions larger than 100 meters has the potential to aggregate the features of interest (Wulder et al. 2004).

Techniques designed to assess forest vertical structure from reflected spectral signatures have seen mixed success (Scarth and Phinn 2000, Hansen et al. 2001, Xian et al. 2002, Cohen et al. 2003, Zheng et al. 2004). However, the non-parametric k-Nearest Neighbors (kNN, Denoeux 1995) technique has been successful in mapping landscape-scale assessments of forest structure and cover from medium resolution satellite imagery. Researchers in Minnesota used kNN to classify cover type, basal area, and diameter (Franco-Lopez et al. 2001, Haapanen et al. 2004). In Sweden researchers mapped wood volume, age, and biomass (Reese et al. 2002). And kNN has been employed extensively to map diameter, height, age, basal area, and volume in Finland (Maltamo and Kangas 1998, Tomppo et al. 2002, Tuominen et al. 2003). Liu et al. (2003) compared kNN to other cover type classification methods including traditional parametric classifiers and artificial neural network models. They found that kNN equaled the accuracy of the neural network models (greater than 90% overall accuracy for six classes) despite its much simpler implementation. Both the kNN and neural network models performed significantly better than traditional classifiers.

With kNN, pixel-level errors are generally large for continuous-scale forest structural classifications. Accuracy measures are typically reported in RMSE (root mean square error), or RMSE as a percentage of the mean of the reference sample plot values (relative RMSE or %RMSE). Often the %RMSE values are as much as 50-100% (Reese et al. 2002, Makela and Pekkarinen 2004). However, when estimates are aggregated over a larger area, e.g. within a patch or stand, the estimates are often as good as field measurements (Trotter et al. 1997, Holmstrom et al. 2001, Reese et al. 2002). The difficulty in assessing the accuracy of continuous scale variables and the natural variance in structural measurements in temperate mixed deciduous forests create a large amount of uncertainty in these data, and add to the apparent error in accuracy calculations.

The goal of this study is to classify a set of forest structure measurements for a heavily forested area in northern Michigan, and assess their utility for wildlife habitat modeling. As noted, there are numerous examples of mapping land-cover from satellite imagery, but relatively few examples of mapping forest structural variables. If cover type and forest structure maps can be produced in combination and at a useful resolution and level of accuracy, it would represent a valuable tool to monitor wildlife habitat resources at a minimum of effort and cost. Other wildlife habitat monitoring programs require many person-hours to gather both wildlife occurrence data and vegetation measurements. With both land cover and vegetation structure mapped on the same spatial extent and resolution, we may be able to more effectively (and efficiently) monitor the habitat resources of many species across jurisdictional boundaries.

Methods

The study area for this project lies in the eastern Upper Peninsula of Michigan (UP). This approximately 2 million acre (800,000 hectare) landscape is 80% forested by area, and ownership is split into four major groups; National Forest (35%), state forest (35%), industrial and non-industrial private (25%), and protected wilderness (5%). The major vegetation types include northern hardwood forests, white and red pine forests, jack pine barrens, aspen monocultures, mixed hardwood-conifer forests, conifer swamps, and bogs (Albert 1995) but there is considerable overlap in species composition. Forests in the Upper Great Lakes region are managed primarily for timber production, and this anthropogenic influence is the primary form of disturbance, replacing fire in many locations (White and Mladenoff 1994).

I classified forest structural characteristics using the kNN algorithm and multi-temporal Landsat 7 Landsat Enhanced Thematic Mapper imagery (Franco-Lopez et al. 2001, Haapanen and Ek 2001). Up to five dates of imagery (Table 5.1) were included for each of two scenes: Row 28, Paths 22 and 23. All imagery was acquired between 2000 and 2003, and has been georectified in the UTM coordinate system (spheroid GRS80, datum NAD83, zone 16) to less than 1/3 of a pixel using nearest-neighbor resampling. I corrected diffuse haze on each image using the Haze-optimized Transformation (Zhang et al. 2002, Zhang and Guindon 2003), and combined all the bands into a multi-temporal raw digital number (DN) composite image for each scene. Over 1000 Forest Inventory and Analysis (FIA) (Hansen et al. 1992) survey plots for each scene were used as ground control points to train and test the kNN classifier. I created maps for five vegetation measurements; basal area (units: square feet per acre), average diameter at breast height (units: inches), total biomass (units: tons per acre), canopy height (units: feet), and stem density (units: number of stems per acre).

The FIA database represents the most detailed and extensive forest inventory in the United States. I accessed FIA sample data and plot coordinates through a cooperative agreement with the USDA Forest Service North Central Research Station in St. Paul, MN. All image classifications and accuracy assessments were performed by Forest Service staff on Forest Service computers. Classified image products were altered to assure that FIA plot locations cannot be identified from published FIA sample data. The FIA program began their sixth forest inventory cycle in Michigan in 2000, and these training data include samples from the first three sub-cycles (2000 through 2003).

FIA plots are arranged in a four-subplot array, and I tested the effectiveness of both the plot and subplot level aggregation of vegetation measurements for training the kNN image classifier. The subplots are 8 meters in width and three of the subplots are arranged around a center subplot at 36.6m center to center. The spatial resolution at which the FIA survey data are gathered is smaller than a single Landsat pixel, and it is unlikely that two sub-plots will be associated with a single pixel (Haapanen et al. 2004).

I used two different methods for aggregating FIA plot measurements to use as training data for satellite image classifications, aggregating the four subplots together into a single sample, or treating each of the four subplots as a separate sample. When the four subplot measurements were aggregated over the entire plot, the vegetation information was associated with a 3x3 mean filtered image pixel for classifications. When each sub-plot was used separately as a ground control point, the 8-meter diameter sub-plot information was matched to a single 30-meter Landsat pixel. The sub-plot aggregation, therefore, results in four times as many reference plots (4000+ per image) as the plot-level aggregation. Correlations between structural measurements and spectral values were calculated to examine the information content of the imagery. For each classification, 90% of the plots were used in training, and 10% were held out for accuracy calculations.

I examined a set of inputs and parameters to determine the optimal conditions that maximize kNN classification accuracy. These included altering the spectral band combinations of the imagery, kNN classification parameters, training data aggregation (plot vs. subplot), removing FIA plots with large variation between subplot measurements, and post processing of classified imagery. Parameters of the kNN classifier included k (the number of reference plots used to calculate unknown pixel predictions), geographical distance weighting, and stratifying the classifications by upland/lowland or general forest cover type (upland conifer/lowland conifer/hardwood/aspen). I tested transformations to lower the dimensionality of the spectral data inputs, including NDVI, principal components analysis, and Tasseled Cap. I also removed the least correlated spectral bands, and various combinations of transformed and untransformed imagery in an attempt to optimize the spectral inputs. Finally, I used post-classification processing of the imagery with a 3x3 mean filter. All the accuracy measures were calculated for path 22 and those results are shown below. Final structure variable classifications for both paths were developed based on the optimal set of parameters and inputs as determined from the path 22 image classification trials. Accuracy was calculated with root mean square error (RMSE) of the predicted values vs. actual values of the test plots (10% of all the FIA plots), and R^2 of the plotted actual vs. predicted values. I also calculated RMSE of the difference between the classified images on the region of overlap between the scenes.

The resulting classified images were used as landscape-level habitat assessments in wildlife habitat models. I used the habitat distribution maps developed by the Michigan Gap Analysis Program (MIGAP) (Scott et al. 1993, Donovan et al. 2004) as a baseline to examine whether accuracy could be improved with the addition of kNN classified structure maps. In a previous study I found that the MIGAP models overestimate the amount of available habitat for most species. When treated as a prediction of presence/absence, the MIGAP models result in a high rate of commission error (predicted present but not detected) but low omission error rates (predicted absent when actually detected).

I examined the published habitat descriptions (Brewer et al. 1991, Donovan et al. 2004) that were used to develop the MIGAP models and identified species that had specific forest structural associations that were likely to be ignored in the original MIGAP models. I then used 169 locations from the Hiawatha National Forest Breeding Bird Survey where I assessed the presence/absence of five of these species: Scarlet Tanager, Eastern Wood-pewee, Chipping Sparrow, Black-throated Blue Warbler, and Pine Warbler. Each of these species was present at 15-40% of the survey sites, and also had MIGAP models that resulted in a better than random prediction (kappa = 0.0 is random, all the models for the species listed above had a kappa >= 0.15). These conditions were used to see if an already successful model could be improved with the addition of one or two simple structural characteristics to the vegetation cover data that were used to build the original MIGAP model.

I overlaid the MIGAP habitat distribution maps and the kNN structure classifications on the field survey plot locations to identify the sites where the combination of maps predicted the location of appropriate habitat. For each of the species listed above, I selected up to two of the structural variables and subset the classified map at the average, or average +/- 0.5 standard deviations (Table 5.5). The cutoff value was selected to make the predicted prevalence close to, but not less than, the actual prevalence from the Hiawatha National Forest bird survey records (Table 5.7). I treated this list of sites as predicted presences, and compared them with the observed locations from the field surveys. The accuracy of each set of models was assessed using 2x2 error matrices (actual presence/absence vs. predicted presence/absence) to calculate the number of commission errors (incorrect presence predictions), omission errors (incorrect absence predictions), percent correctly classified sites (PCC), and kappa. Kappa accounts for large differences in the number of sites in the present and absent categories (Karl et al. 2000, Manel et al. 2001).

Results

The five structural variables used in this analysis show patterns of high correlation with each other (Table 5.2). Height, diameter, and biomass all show correlations (R) above 0.50 when aggregated at the subplot level. Similarly, basal area and stem density are highly correlated, as well as basal area and biomass. Stem density and canopy height, as well as stem density and diameter are negatively correlated. The two groups of variables linked by the largest correlations could be summarized as 'size' (height, diameter, and biomass) and 'density' (basal area and stem density), although basal area and biomass are also highly correlated.

Patterns of correlation between forest structure measurements and spectral values differ by type of measurement, season, and spectral band. Overall, the association between spectral values and structure are rather weak. All of the measured maximum correlation (R^2) values for subplot-level aggregations of FIA data are below 0.22, and average less than 0.13 (Table 5.3a). Plot-level correlations are consistently higher than subplot-level (Table 5.3b), but the values are still low. There is a large amount of variation in correlation values across spectral bands. ETM+ bands 3, 5, and 7 tend to have the highest correlations. High correlations were seen in the summer and early fall images (Table 5.3), but there were high correlations for all image dates with the exception of the early spring (April) image. The highest average correlations over all image dates and spectral bands are seen for the structural variables stem density and basal area (note that these are also highly correlated in the FIA subplots). Stem density and basal area show relatively high correlation with spectral bands across all the seasons. Average diameter, canopy height, and total biomass show weaker correlations with spectral values, limited mainly to the summer and early fall (July, August, and September) images. Overall, the lowest correlations were seen in ETM+ band 4 (near infrared), while the highest correlations were seen in ETM+ bands 5 and 7 (the longer wavelength infrared bands).

The plot-level reference data used to classify 3x3 mean-filtered spectral data lead to higher classified map accuracy than do subplot values with the unfiltered imagery, matching the correlation results. The lowest RMSEs were generated with higher values for k in the classification parameters, but higher values of k also lowered R^2 values by shrinking the range of the predicted values. The tradeoff between R^2 and RMSE takes place in the k=3 to k=5 range. None of the other kNN parameters (geographical distance weighting or stratifying the classifications by upland/lowland or general forest cover type) improved the accuracy of the predictions.

Optimizing the spectral band inputs by statistical transformations (PCA and tasseled cap) or removing uncorrelated bands did not improve, and usually decreased, the accuracy of classifications. Removing the plots with large variability between subplot measurements also lowered accuracy. Post processing the classified images with a 3x3 mean filter did improve the accuracy calculations significantly. Therefore, it appears that the spatial aggregation of these data (3x3 mean filtered spectral inputs, FIA plot-level measurements, and post-classification smoothing) have a larger effect on the quality of the map outputs than do altering the spectral band combinations of the input data, or other parameters of the classifications.

Table 5.4 shows the classification accuracy results for all five structure variables, from classifications with k=5, 3x3 mean filtered spectral values, no band selection or transformations, and plot-level aggregation of FIA measurements. Diameter showed the lowest R^2 and highest %RMSE (root mean square error as a percentage of the average reference plot value), but also had the lowest range of any of the five structure variables I classified. Height and stem density had the lowest %RMSE values, but intermediate R^2. Basal area and biomass showed the highest R^2, but had intermediate %RMSE values. The range of predicted values is reduced in comparison to the reference samples. On the overlap region between paths 22 and 23, RMSE values calculated from the difference between image pixels were very similar to within-image RMSE (Table 5.4). The exceptions were height, which actually showed much lower RMSE between image paths than within path 22, and stem density, which had the lowest %RMSE within the path 22 image, showed a much larger RMSE on the overlap region.

A visual evaluation of the classified images reveals a very close association with aerial photography and the Hiawatha National Forest GIS inventory database. Stand boundaries that correspond to breaks in structure values on the classified imagery are frequently visible, and other patterns (e.g. linear features like roads, and open areas like lakes) are readily recognizable. When the pixel values for the basal area kNN predictions are averaged within national forest stands and compared to the GIS inventory database basal area measurement records, there is a high correlation ($R^2 = 0.61$). Despite the high correlation, the range of the Hiawatha inventory basal area measurements (0 – 270 square feet per acre) is larger than the kNN predictions (0 – 180 square feet per acre). Even with respect to the high pixel-level RMSE values, and decreased range of predicted values, these data are clearly are representative of forest structural conditions on the ground.

Adding the classified structure maps to the MIGAP models showed marked improvement for the five species included in this analysis. In every case, both kappa and percent correctly classified (PCC) increased when the structure elements were added (Table 5.6). The apparent source of improvement was the reduction of commission errors, at the expense of a smaller increase in omission errors (Table 5.7). The largest improvements were seen for Chipping Sparrow, Black-throated Blue Warbler, Eastern Wood-pewee, and Scarlet Tanager. Pine Warbler showed only a small improvement.

Discussion

These results show that simple forest structural estimates built at the same resolution as currently available land cover distribution maps can significantly improve the accuracy of wildlife habitat distribution models. The technical challenge in developing these data is not as great as many may assume, but the success of classifying continuous-scale forest structural elements depends largely on the quality of the ground control data (i.e. the forest structural measurement used for reference samples). FIA represents an excellent and highly valuable source of information for this type of analysis, but access to these data (specifically the plot locations) is restricted, and few sources of reference data like it are available.

Accuracy assessment of continuous-scale vegetation structure maps is difficult and perhaps even uninformative when classified at the pixel level and using point samples. The natural variation of the forest structural measurements used in this study appears to be at a scale larger than an 8-meter subplot, or potentially even a 30-meter pixel. There is a clear potential for scaling discontinuities in the use of 8 meter diameter FIA subplot surveys to classify 30 meter Landsat ETM pixels, and I suspect that is a main reason why the plot level aggregation of FIA data returned higher correlations with spectral values (smoothed with a 3x3 mean filter) than subplot level aggregation (and unfiltered spectral data). Because per pixel errors (RMSE) are high, it may give the impression that the classifications were unsuccessful. For continuous-scale classifications such as these, I feel that it is necessary to include assessments that reveal accuracy at a scale larger than a single pixel, e.g. averaged over a forest stand or other type of patch. Evidence for this is shown by the visual evaluation of the structural maps in comparison to aerial photography and a national forest GIS inventory database. Perhaps another method to assess the pattern of association over space would be more useful than the per pixel methods I used (RMSE and R^2). RMSE may be a highly misleading accuracy measure because it is dependent on the range of each variable. Also, different results were obtained when calculated against the test plots vs. the image path overlap areas, where at least two of the classified variables revealed opposite trends.

The kNN estimator is a flexible method for imputing continuous scale forest structural values to unknown pixels (Meng et al. 2007). I have found it to be data hungry, both in terms of the number of spectral layers and reference points, supporting the findings of other users of this classification technique (Franco-Lopez et al. 2001, Holmstrom and Fransson 2003, Budreski et al. 2007, Koukal et al. 2007). Any efforts to optimize the spectral inputs resulted in either no effect or a decrease in accuracy. I also found that it was not just the number of reference points supplied to the classifier, but the spatial scale (see Barth et al. 2009 for a discussion of the impacts of scale on kNN results) at which the measurements were assembled that increased the accuracy of classifications (plot-level aggregation resulted in 25% as many reference points as subplot aggregation, yet achieved greater accuracy). The number of reference plots used in the pixel value calculations (k) has a significant effect on the results. I used k=5 and achieved relatively high accuracy numbers but a shrinking of the range of predicted values. But in my trials, k=3 was nearly as accurate and would have resulted in a smaller reduction of predicted variable ranges (Franco-Lopez et al. 2001).

GAP models were not intended for use as predictive models of species occurrence, rather they are intended to show a coarse assessment of the distribution of potential habitat. Still, combining land cover and forest structural information would be valuable to create potential habitat distributions with less uncertainty, thereby reducing the risk of overestimating the amount of habitat resources available on the landscape for any particular species. Overestimating the available habitat for species of management concern has deep implications given the importance of thresholds in habitat amount (Fahrig 2001), and disproportionately large effects of landscape patterns (Donovan and Flather 2002).

The five species selected for this study were expected to show improvements with the addition of structure information. These are not intended to be final products or to show conclusions about the ecology of these species, but instead to show that the potential for improvement of GAP models with simple forest structural elements exists and the magnitude of improvement that might be possible. Adding simple forest structural information could represent the highest potential for increased accuracy in large scale wildlife habitat models at the lowest cost per effort. Assembling wildlife occurrence data will always be effort intensive, but using remote sensors shows continued potential for reducing the effort and cost necessary to monitor the distribution of wildlife habitat and to identify gaps in our conservation networks.

CHAPTER 6

CONCLUSION AND SYNTHESIS

I have summarized some patterns of wildlife habitat model utility and have given specific examples. Habitat specialists are more likely to produce accurate models than generalists because statistics can more easily define habitat classes that clearly delineate appropriate from poor habitats for these species. Scale plays a large role also, as coarser data may not be as appropriate for describing edge and mixed habitat associations as higher resolution data. The correlation between model accuracy and prevalence is typically positive (e.g. Vaughan and Ormerod 2005) since rarer species are simply less likely to be present on any given location of appropriate habitat (Manel et al. 2001). The challenge in building wildlife habitat models is to predict appropriate habitat sites where there is a high probability of presence, but this can be difficult because of the fact that rare species may actually have a low probability of being detected on even the best sites. Accuracy measures fail to take this into account and therefore may not reflect an actual measure of model quality, but instead an inability to account for uncertainty in species occurrence. The uncertainty in species occurrences can be due to a myriad of ecological reasons unrelated to habitat (Storch and Sizling 2002), as well as detectability (MacKenzie et al. 2005, Royle et al. 2005).

Most common accuracy measures, including kappa when a 0.5 threshold is used, are relatively inflexible to the species' ecological characteristics (such as habitat specificity and prevalence). Kappa, however, can be used effectively when the choice of threshold in probability of occurrence is flexible, and tied to each species' prevalence

(Freeman and Moisen 2008b). In this study, the relationship between prevalence and model accuracy was weak (kappa, Figure 3.2) or even negative (ROC/AUC, Figure 3.3).

For wildlife habitat models there is a tradeoff between sensitivity and specificity (Allouche et al. 2006). Sensitivity is the probability of correctly classifying a presence, specificity is the probability of correctly classifying an absence (Table 3.1). By changing the probability of occurrence threshold to increase one, the other declines. ROC/AUC assesses model accuracy across all values of sensitivity/specificity and therefore is not affected by threshold choice, but kappa can change drastically (Allouche et al. 2006, Freeman and Moisen 2008b). The proper choice of threshold values can optimize the specificity vs. sensitivity tradeoff, even when using kappa, but see Manel et al. (2001). The two threshold values used in this study had relatively small effects on kappa (Figure 3.2, Table 3.4), but large effects on commission and omission error rates (Figure 3.3, Table 3.4).

All of the models I developed, except for species with very specific habitat associations, were prone to inherently large commission error rates. As the ultimate goal of all the work described in this paper is the conservation of wildlife habitat, it may be desirable to minimize the omission error rate even at the expense of increasing the commission error rate. The reason for this would be to preserve as much *potential* habitat for each species as possible, as any increase in omission errors associated with wildlife habitat models will lead to neglecting potential habitat for that species (Wilson et al. 2005). If this is a desirable condition of a wildlife habitat modeling project for a large set of species, then the threshold for considering a location as appropriate habitat should be equal to the prevalence of the species (the first threshold used in this study). When this is the case, commission error rates increased with less prevalent species, but omission errors were low across all species (Figures 3.4a and 3.4b). However, when resources are limited and only a small set of locations can be targeted for conservation, then a different approach may be necessary so that the most important locations are protected.

When a model performs poorly, it can be due to many factors, and a poor quality model may not in fact be useless. Some species have a low prevalence across their range and are not likely to have a high probability of occurrence at any given location (Seoane et al. 2005b). For example, species that show low site fidelity and instead are nomadic or focused on a spatially patchy/irruptive food resource may be recorded in many different locations over different years, but still within similar habitats. In this case a model that does not incorporate the food resource will be unable to predict those locations year to year. Accuracy measures will show that this model performs poorly, when in fact it does a very good job of describing the distribution of appropriate habitat. This model, with very poor accuracy, may in fact be very useful for conservation activities. Other examples of important vegetation and environmental measurements may be immeasurable and therefore will not be included in model construction, but if this is known beforehand the model may still be useful depending on the application. Often, higher-level effects such as community interactions, predation, and inter-specific competition can add variation to species distributions that habitat models cannot track. Species undergoing rapid population increases or declines can be difficult to model accurately, but if these patterns are understood then steps can be taken to make interpretation of model results more feasible.

Low resolution landscape and regional-scale models (like GAP) were not intended for use as predictive models of species occurrence, rather they are intended to show the distribution of potential habitat. But models such as these can also be improved (have less uncertainty) by combining land cover and forest structural information. Adding estimates of forest vertical structure and composition can reduce the risk of overestimating the amount of habitat resources available for species that are associated with forest habitats. This was one of the primary goals in Chapters 2 and 5. Overestimating the available habitat for species of management concern has deep implications given the importance of thresholds in habitat amount (Fahrig 2001), and disproportionately large effects of landscape patterns (Donovan and Flather 2002). Though I didn't specifically test it, I suspect that improving the detail and optimizing the scale of the vegetation data used as inputs to wildlife habitat models would have a much larger effect upon the utility of models than using newer and more elaborate statistical methods, despite the many efforts devoted to the latter.

APPENDICES

TABLES

Table 2.1: List of habitat variables included in each modeling phase. The number and detail of vegetation cover classes are comparable to Level 3 in the hierarchical ecological classification system developed by Anderson et al. (1976). Phases 1-3 are ordered from less to more vegetation information and/or lower to higher spatial resolution. The number of vegetation classes varies between phases. MIGAP: 19 classes (11 forest types); Stand-scale: 20 classes (8 forest types); Plot-scale: 19 classes (9 forest types).

Habitat model variable	**MIGAP and cover class (1/2a/2b)**	**Stand and plot-scale vegetation models (3a/3b)**
Vegetation cover class	variable	Variable
Basal area	N/A	Average of three measurements per stand
Diameter at breast height	N/A	Proportional average for all species in stand
Canopy closure	N/A	Visual estimate (four 25% categories)
Deciduous canopy cover	N/A	Sum of deciduous cover divided by total
Canopy species richness	N/A	Count of canopy species
Canopy species diversity	N/A	Simpson's (1/P) diversity of canopy species cover
Subcanopy cover	N/A	Sum of individual species cover
Deciduous subcanopy cover	N/A	Sum of deciduous cover divided by total
Subcanopy richness	N/A	Count of subcanopy species
Subcanopy species diversity	N/A	Simpson's (1/P) diversity of subcanopy species cover
Overall size of canopy trees	N/A	Average size of dominant trees (sap, pole, log)
Upland or lowland	N/A	Binary marker
Plantation	N/A	Binary marker
Location	Inherent in MIGAP maps, or binary	Binary (North/South)

Table 2.2: List of bird species included in models. Prevalence lists the proportion of survey sites at which each species was present (out of 393 total). Most (17) of the species are associated with forest habitats, some (9) are associated with mixed (edge) habitats, and fewer are wetland (3) and grassland (1) species (Peterjohn and Sauer 1993).

Common Name	**Scientific Name**	**Prevalence**	**Habitat**
Ovenbird	*Seiurus aurocapillus*	0.55	Forest
Red-eyed Vireo	*Vireo olivaceus*	0.55	Forest
American Goldfinch	*Carduelis tristis*	0.41	Grassland
Blue Jay	*Cyanocitta cristata*	0.42	Forest
Common Yellowthroat	*Geothlypis trichas*	0.32	Wetland
Black-capped Chickadee	*Poecile atricapillus*	0.40	Forest
American Robin	*Turdus migratorius*	0.34	Mixed
Rose-breasted Grosbeak	*Pheucticus ludovicianus*	0.31	Forest
Red-winged Blackbird	*Agelaius phoeniceus*	0.20	Wetland
Chipping Sparrow	*Spizella passerina*	0.30	Mixed
Eastern Wood-Pewee	*Contopus virens*	0.31	Forest
Veery	*Catharus fuscescens*	0.25	Forest
American Redstart	*Setophaga ruticilla*	0.22	Forest
Gray Catbird	*Dumetella carolinensis*	0.25	Mixed
Scarlet Tanager	*Piranga olivacea*	0.24	Forest
Indigo Bunting	*Passerina cyanea*	0.24	Mixed
Wood Thrush	*Hylocichla mustelina*	0.19	Forest
Great Crested Flycatcher	*Myiarchus crinitus*	0.21	Forest
Eastern Tufted Titmouse	*Baeolophus bicolor*	0.19	Mixed
Eastern Towhee	*Pipilo erythrophthalmus*	0.18	Forest
Field Sparrow	*Spizella pusilla*	0.17	Mixed
Hermit Thrush	*Catharus guttatus*	0.13	Forest
White-breasted Nuthatch	*Sitta carolinensis*	0.15	Forest
Northern Flicker	*Colaptes auratus*	0.16	Mixed
Yellow-billed Cuckoo	*Coccyzus americanus*	0.13	Mixed
Cedar Waxwing	*Bombycilla cedrorum*	0.13	Mixed
Nashville Warbler	*Vermivora ruficapilla*	0.07	Forest
Pine Warbler	*Dendroica pinus*	0.09	Forest
Alder Flycatcher	*Empidonax alnorum*	0.05	Wetland
Black-throated Green Warbler	*Dendroica virens*	0.07	Forest

Table 2.3: Inclusion rate of habitat variables in the stand and plot-scale statistical models (phases 3a and 3b). The RPART algorithm fits a recursive partitioning tree to the vegetation data that best accounts for the presence and absence of each species. At each node one variable is selected and used to split the sites into two groups. Numbers reveal the average number of times each variable was included per model.

Variable	**Phase 3a**	**Phase 3b**
Cover class	1.60	1.60
Basal area	0.53	0.47
Diameter	0.50	0.60
Canopy closure	0.17	0.27
Canopy % deciduous	0.20	0.43
Canopy richness	0.33	0.20
Canopy diversity	0.43	0.63
Subcanopy cover	0.57	0.47
Subcanopy % deciduous	0.23	0.43
Subcanopy richness	0.23	0.20
Subcanopy diversity	0.43	0.37
Overall size	0.17	0.13
Upland or lowland	0.10	0.03
Plantation	0.03	0.00
Location	0.33	0.47

Table 3.1: Error matrix used to calculate kappa, omission and commission error rates, sensitivity and specificity, and other accuracy measures (but not ROC/AUC). Cells ‘a’ and ‘d’ are the number of correct presence and absence predictions, respectively. Cell ‘b’ is the number of incorrect presence predictions, and cell ‘c’ is the number of incorrect absence predictions.

		Observations	
		Presence	absence
Predictions	presence	a	b
	absence	c	d

Accuracy measure equations:

(total number of samples = n = a + b + c + d)

Kappa = $$\frac{\left(\frac{(a+d)}{n}\right)-\left(\frac{(a+b)(a+c)+(c+d)(b+d)}{n^2}\right)}{1-\frac{((a+b)(a+c)+(c+d)(d+b))}{n^2}}$$

Commission error = $\frac{b}{a+b}$ Omission error = $\frac{c}{a+c}$

Sensitivity = $\frac{a}{a+c}$ Specificity = $\frac{b}{b+d}$

Percent correctly classified (PCC) = $\frac{a+d}{n}$

User’s accuracy = 1 – commission error = $\frac{a}{a+b}$

Producer’s accuracy = 1 – omission error = $\frac{a}{a+c}$ = sensitivity

Table 3.2: List of habitat variables included in the recursive partitioning models. The number and detail of vegetation cover classes are comparable to Level 3 in the hierarchical ecological classification system developed by Anderson et al. (1976).

Habitat variable	Vegetation measurements
Vegetation cover class	20/19 classes, 8/9 forest types (stand/field)
Basal area	Average of three measurements per stand
Diameter at breast height	Proportional average for all species in stand
Canopy closure	Visual estimate (four 25% categories)
Proportion of deciduous canopy cover	Sum of deciduous cover divided by total
Canopy species richness	Count of canopy species
Canopy species diversity	Simpson's (1/P) diversity of canopy species cover
Subcanopy cover	Sum of individual species cover
Proportion of deciduous subcanopy cover	Sum of deciduous cover divided by total
Subcanopy richness	Count of subcanopy species
Subcanopy species diversity	Simpson's (1/P) diversity of subcanopy species cover
Overall size of canopy trees	Average size of dominant trees (sap, pole, log)
Upland or lowland	Binary marker
Plantation	Binary marker
Location	North/South

Table 3.3: List of bird species included in models. Prevalence lists the proportion of survey sites at which each species was present (out of 393 total). Most (17) of the species are associated with forest habitats, some (9) are associated with mixed (edge) habitats, and fewer are wetland (3) and grassland (1) species (Peterjohn and Sauer 1993). Prevalence rank shows the order that species are listed in Figures 3.2-3.4.

Common Name	Scientific Name	Prevalence	Habitat	Specificity
Ovenbird	***Seiurus aurocapillus***	**0.55**	**Forest**	**Specialist**
Red-eyed Vireo	*Vireo olivaceus*	0.55	Forest	Specialist
Black-capped Chickadee	*Poecile atricapillus*	0.40	Forest	Generalist
American Robin	***Turdus migratorius***	**0.34**	**Mixed**	**Generalist**
Blue Jay	*Cyanocitta cristata*	0.42	Forest	Specialist
Rose-breasted Grosbeak	*Pheucticus ludovicianus*	0.31	Forest	Specialist
Common Yellowthroat	*Geothlypis trichas*	0.32	Wetland	Specialist
American Goldfinch	*Carduelis tristis*	0.41	Grassland	Specialist
Chipping Sparrow	*Spizella passerina*	0.30	Mixed	Generalist
Eastern Wood-Pewee	*Contopus virens*	0.31	Forest	Specialist
Veery	*Catharus fuscescens*	0.25	Forest	Specialist
American Redstart	*Setophaga ruticilla*	0.22	Forest	Specialist
Scarlet Tanager	*Piranga olivacea*	0.24	Forest	Specialist
Indigo Bunting	*Passerina cyanea*	0.24	Mixed	Generalist
Hermit Thrush	*Catharus guttatus*	0.13	Forest	Specialist
Great Crested Flycatcher	*Myiarchus crinitus*	0.21	Forest	Specialist
Red-winged Blackbird	*Agelaius phoeniceus*	0.20	Wetland	Specialist
Northern Flicker	*Colaptes auratus*	0.16	Mixed	Generalist
Gray Catbird	*Dumetella carolinensis*	0.25	Mixed	Generalist
Wood Thrush	*Hylocichla mustelina*	0.19	Forest	Specialist
Cedar Waxwing	*Bombycilla cedrorum*	0.13	Mixed	Generalist
Eastern Towhee	*Pipilo erythrophthalmus*	0.18	Forest	Specialist
White-breasted Nuthatch	*Sitta carolinensis*	0.15	Forest	Specialist
Nashville Warbler	*Vermivora ruficapilla*	0.07	Forest	Specialist
Eastern Tufted Titmouse	*Baeolophus bicolor*	0.19	Mixed	Generalist
Yellow-billed Cuckoo	***Coccyzus americanus***	**0.13**	**Mixed**	**Generalist**
Field Sparrow	*Spizella pusilla*	0.17	Mixed	Generalist
Black-throated Green Warbler	***Dendroica virens***	**0.07**	**Forest**	**Specialist**
Pine Warbler	*Dendroica pinus*	0.09	Forest	Specialist
Alder Flycatcher	*Empidonax alnorum*	0.05	Wetland	Specialist

Table 3.4: Results of model accuracy measurements for the four species targeted for detailed examination and averaged for the 10 habitat generalists and 20 habitat specialists included in this analysis. Table 3.4a shows kappa and commission/omission error calculated with the threshold = prevalence, Table 3.4b uses the threshold where predicted prevalence = actual prevalence. One asterisk indicates that the average values for generalists and specialists are significantly different from each other at P=0.1 (two asterisks for P=0.05). Significance calculated with an independent groups T-test.

Threshold = prevalence

Table 3.4a	Kappa		AUC		% Commission/Omission	
	Stand	Plot	Stand	Plot	Stand	Plot
Ovenbird	0.61	0.61	0.86	0.84	16.3/19.8	17.2/16.0
American Robin	0.35	0.33	0.72	0.70	43.5/41.8	48.7/35.1
Yellow-billed Cuckoo	0.14	0.36	0.78	0.76	80.6/2.0	61.4/44.3
Black-throated Green Warbler	0.54	0.49	0.89	0.85	52.3/25.0	56.8/29.0
Generalists	0.30**	0.35*	0.74**	0.77	56.1/28.2	56.8/29.6
Specialists	0.40**	0.41*	0.80**	0.81	51.4/25.8	50.6/26.1

Threshold where predicted prevalence = actual

Table 3.4b	Kappa		AUC		% Commission/Omission	
	Stand	Plot	Stand	Plot	Stand	Plot
Ovenbird	0.60	0.61	0.86	0.84	18.9/16.1	17.2/16.0
American Robin	0.35	0.33	0.72	0.70	43.5/41.8	48.7/35.1
Yellow-billed Cuckoo	0.31	0.32	0.78	0.76	59.6/62.0	59.4/57.4
Black-throated Green Warbler	0.59	0.59	0.89	0.85	42.4/32.1	6.7/54.8
Generalists	0.31**	0.34**	0.74**	0.77	38.0/59.7	44.0/50.5
Specialists	0.44**	0.43**	0.80**	0.81	41.4/42.1	37.9/45.7

Table 4.1: List of vegetation and cover classes defined in the hierarchical IFMAP classification system (used as a baseline in this study). Top table shows the number of all classes defined in the classification, the lower table shows the number of classes sampled in this study (460 total field plots).

Complete list			
		# classes	
Level-1 class descriptions	**Level-2**	**Level-3**	**Level-4**
Urban	2	4	4
Agricultural	2	4	7
Upland Openland	4	4	14
Upland Forest	3	10	47
Water	1	1	1
Wetlands	2	7	38
Sparsely vegetated	1	4	4
Total	15	34	115

Field sample totals						
	# classes			# classes retained in predicted classification		
Level-1 classes	**Level-2**	**Level-3**	**Level-4**	**Level-2**	**Level-3**	**Level-4**
Agricultural	1	1	3	0	0	0
Upland Openland	3	3	7	3	2	2
Upland Forest	3	8	27	2	5	13
Wetlands	2	4	15	2	3	6
Total	9	16	52	7	10	21

Table 4.2: Comparison of the distribution of sites among classes, and the number of classes at each level for each classification.

	# sites/class (std. dev.)			# classes		
Classification	Level			Level		
	2	3	4	2	3	4
IFMAP	51.33 (54.4)	28.88 (19.1)	8.88 (7.6)	9	16	52
Predictions	66 (54.2)	46.2 (20.6)	22 (19.7)	7	10	21
Clusters	51.33 (21.2)	28.88 (17.3)	8.88 (6.2)	9	16	52

Table 5.1: Image dates and phenology information. The selection of imagery was targeted at providing a range of (snow free) leaf-off and leaf-on images across the phenological range of tree species in the northern Great Lakes region.

Path/Row		
P23/R28	P22/R28	Phenology
n/a	April 26, 2000	Early spring, leaf-off
May 19, 2000	May 21, 2003	Mid-spring, early leaves
July 28, 2003	August 03, 2001	Mid-summer, full leaves
September 9, 2000	September 17, 2000	Early-fall, beginning senescence
October 10, 2000	October 19, 2000	Late fall, complete senescence

Table 5.2: Correlations (R) between FIA vegetation measurements aggregated at the subplot level for path 22 reference plots.

	Diameter	Height	Stem density	Biomass	Basal area
Diameter	1	0.6690	-0.2646	0.5768	0.2890
Height		1	-0.1432	0.6164	0.2310
Stem density			1	0.3366	0.6029
Biomass				1	0.6705
Basal area					1

Table 5.3: Correlation (R^2) between satellite spectral values and vegetation measurements aggregated at the (5.3a) sub-plot level, and (5.3b) plot level summarized across all image dates for path 22 imagery. Sub-plot measurements are matched with raw spectral data. Plot values are matched with 3x3 mean filtered spectral values. The three bands showing the largest correlations are also listed.

Table 5.3a

	Max	Avg.	St Dev	Max Band IDs
Basal area	0.179	0.093	0.053	SeptETM7, SeptETM5, JulyETM5
Biomass	0.129	0.049	0.042	JulyETM3, SeptETM3, SeptETM7
Height	0.180	0.067	0.060	JulyETM3, SeptETM7, JulyETM2
Diameter	0.212	0.089	0.063	SeptETM7, SeptETM3, JulyETM3
Stem density	0.216	0.122	0.069	SeptETM5, JulyETM5, SeptETM7

Table 5.3b

	Max	Avg.	St Dev	Max Band IDs
Basal area	0.340	0.174	0.104	SeptETM7, SeptETM5, SeptETM3
Biomass	0.213	0.071	0.065	JulyETM3, SeptETM3, SeptETM7
Height	0.224	0.075	0.070	SeptETM7, SeptETM3, JulyETM3
Diameter	0.274	0.106	0.081	SeptETM7, SeptETM3, JulyETM3
Stem density	0.301	0.172	0.098	SeptETM5, JulyETM5, SeptETM7

Table 5.4: Accuracy of the kNN classifications for each of the five structural measurement maps for path 22 imagery. The inputs for these maps were plot-level FIA measurements, and a 3x3 mean filtered image composite of raw DN spectral values. The maps were generated with a 90% build set of over 1000 FIA plots, and accuracy (R^2 and RMSE) were calculated with the remaining 10% of the plots. %RMSE is calculated from RMSE as a percentage of the mean for each structure variable. Overlap RMSE shows the RMSE for approximately 3.5 million pixels in the region of overlap between paths 22 and 23.

	Mean	RMSE	%RMSE	R^2	Overlap RMSE
Basal area	74.9	19.8	26.4	0.40	26.7
Biomass	29.7	12.0	40.6	0.43	11.9
Height	45.3	11.2	24.7	0.30	9.7
Diameter	6.9	4.4	64.1	0.10	1.5
Stem density	179.2	31.1	17.4	0.33	64.9

Table 5.5: Bird species and habitat model descriptions. Habitat descriptions were taken from the species habitat rules and descriptions in the MIGAP habitat decision rules (Brewer et al. 1991, Donovan et al. 2004). The strategy for building structure models was to identify up to two structure variables from the MIGAP habitat descriptions and choose a cutoff at the average value or average +/- units of 0.5 x standard deviation, while keeping the predicted prevalence similar to (but not less than) the Hiawatha National Forest Bird Survey recorded prevalence (Table 7). Forest structural associations from the MIGAP habitat descriptions are highlighted in bold.

Species	Scientific name	MIGAP habitat description	Structure model
Black-throated Blue Warbler	*Dendroica caerulescens*	**Mature** hardwood forest with **dense** undergrowth	MIGAP + high stem density + high height
Chipping Sparrow	*Spizella passerina*	**Open** mixed forest and savannah	MIGAP + low biomass
Eastern Wood-pewee	*Contopus virens*	**Mature, open** deciduous woodlands	MIGAP + large diameter + low stem density
Scarlet Tanager	*Piranga olivacea*	**Tall, mature** mixed hardwood forest	MIGAP + high height + large diameter
Pine Warbler	*Dendroica pinus*	**Widely spaced,** mature pine forest	MIGAP + low basal area

Table 5.6: Bird habitat model results are shown for each species (percent correctly classified [PCC] and kappa) comparing the original MIGAP models with the MIGAP plus structure models. On average, both PCC and kappa are higher for the MIGAP plus structure models. The difference between the averages for both PCC and kappa is significant at $p<0.05$ (paired t-test).

	PCC		Kappa	
Species	MIGAP	Structure	MIGAP	Structure
Black-throated Blue Warbler	59.8	78.1	0.30	0.56
Chipping Sparrow	66.9	75.1	0.38	0.50
Eastern Wood-pewee	54.4	65.7	0.13	0.31
Scarlet Tanager	45.6	73.4	0.14	0.40
Pine Warbler	75.1	82.8	0.33	0.34
Average	60.4	75.0	0.26	0.42

Table 5.7: Bird habitat model results are shown for each species (commission and omission errors) comparing the original MIGAP models with the MIGAP plus structure models. The addition of structure elements in the models has the effect of reducing the number of commissions more than the increase in omissions.

	#records	# predictions		# commissions		# omissions	
Species	HNF	MIGAP	Struct.	MIGAP	Struct.	MIGAP	Struct.
Black-throated Blue Warbler	58	124	81	67	30	1	7
Chipping Sparrow	65	111	79	51	28	5	14
Eastern Wood-pewee	76	121	78	61	30	16	28
Scarlet Tanager	39	127	66	90	36	2	9
Pine Warbler	20	56	31	39	20	3	9
Average	51.6	107.8	67.0	61.6	28.8	5.4	13.4

FIGURES

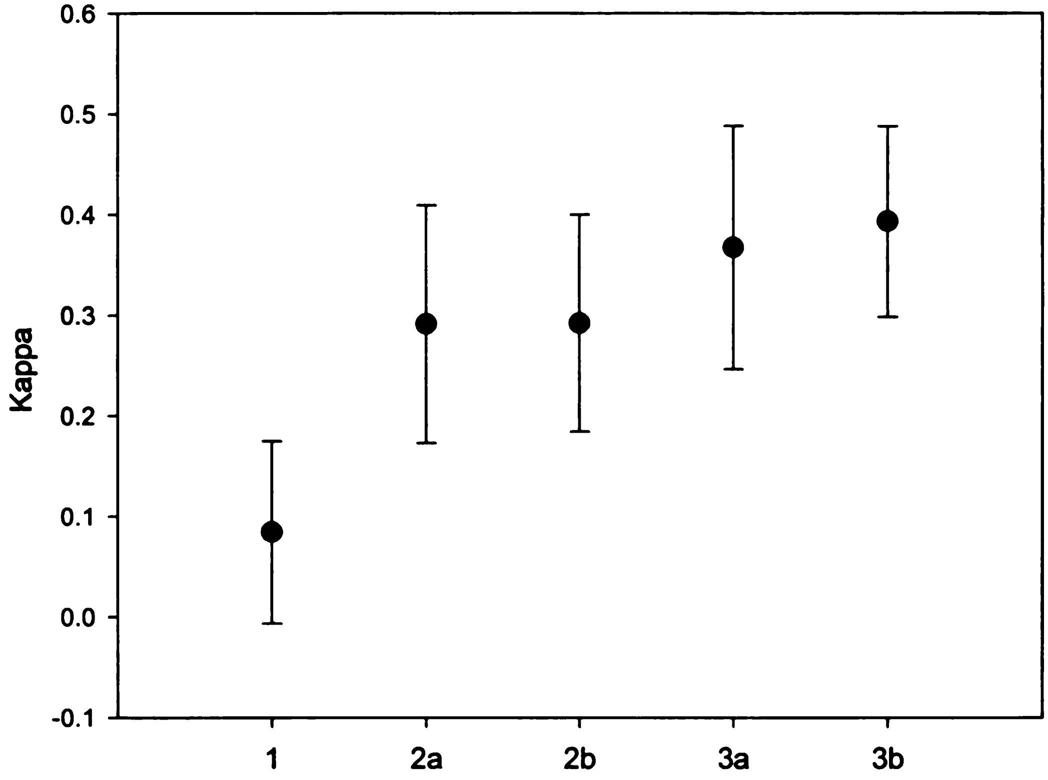

Figure 2.1: Model accuracy (kappa, scale -1 to +1) by model type, results averaged for all 30 species. Label codes refer to the Phases: 1 – MIGAP, 2 – vegetation cover classes only, 3 – full set of vegetation measurements and cover classes, a – stand-scale measurements, b – plot-scale measurements. Error bars show 1 standard deviation. Paired t-tests reveal that between Phases 1, 2, and 3 the accuracy increases significantly ($p<0.05$), but not between scales of vegetation measurements (a and b).

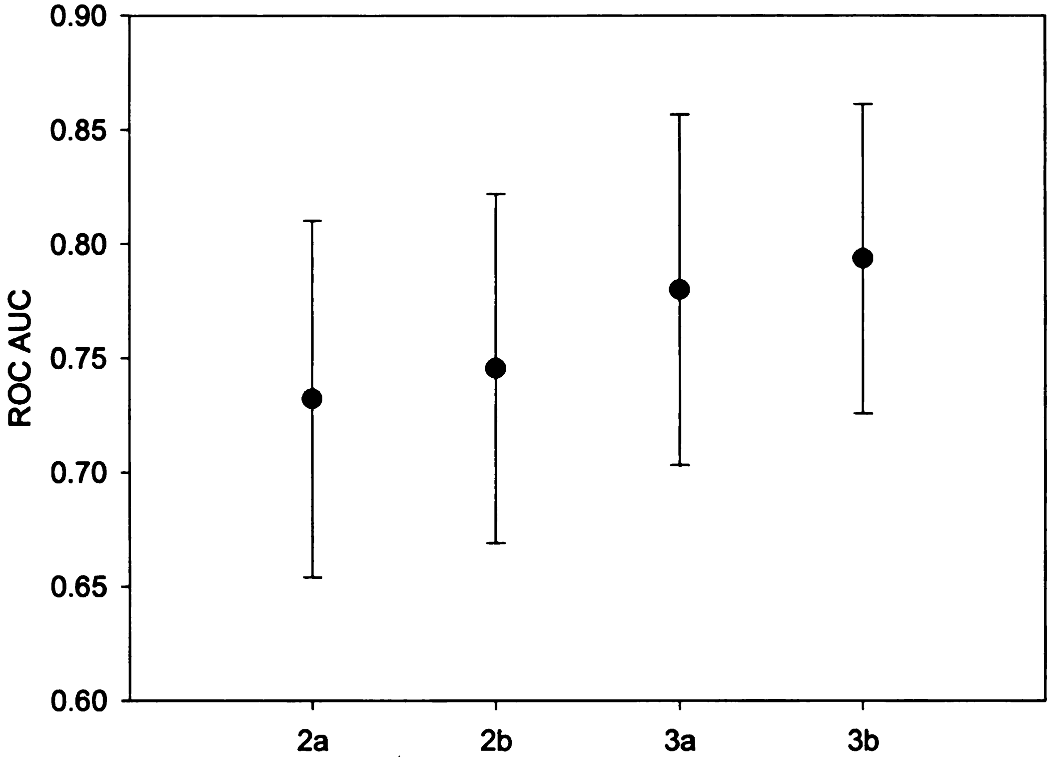

Figure 2.2: Model accuracy (ROC/AUC, scale 0.5 to 1.0) by model type, results averaged for all 30 species. Label codes refer to the Phases: 1 – MIGAP, 2 – vegetation cover classes only, 3 – full set of vegetation measurements and cover classes, a – stand-scale measurements, b – plot-scale measurements. The difference between the cover type models (Phase 2) and the full vegetation measurements models (Phase 3) is similar with ROC/AUC and Kappa (Figure 2). The MIGAP models are not included because they are binary models so there is no way to calculate ROC/AUC. Error bars show 1 standard deviation. Paired t-tests reveal that between Phases 2 and 3 the accuracy increases significantly ($p<0.05$), but not between scales of vegetation measurements (a and b).

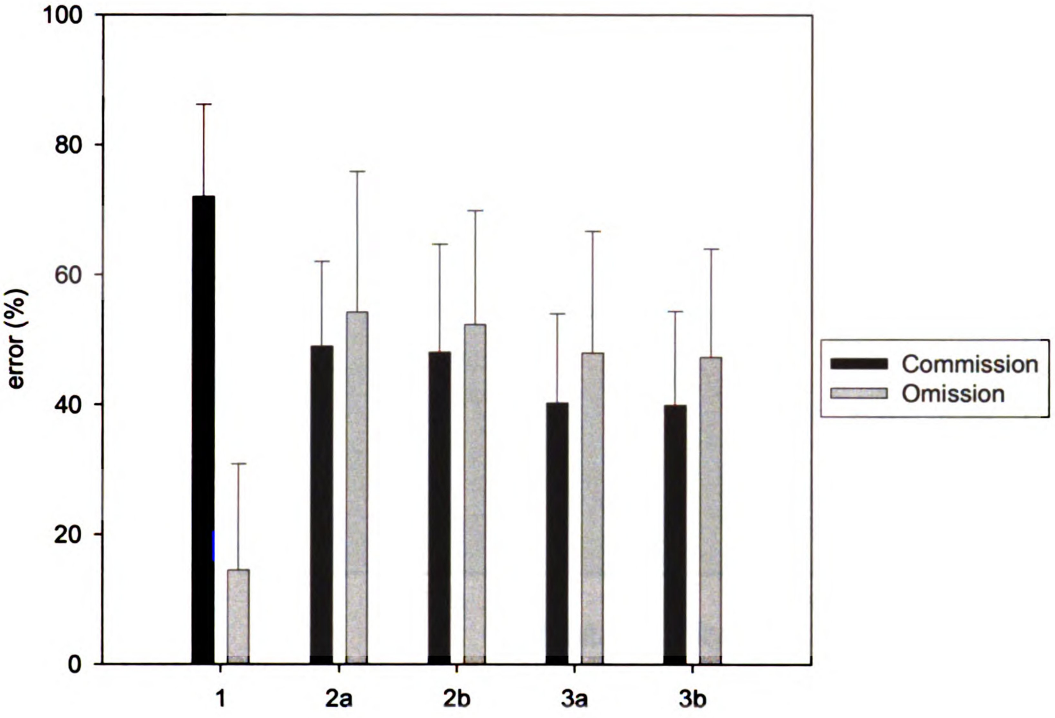

Figure 2.3: Model accuracy (commission and omission error rates) by model type, results averaged for all 30 species. Label codes refer to the following: 1 – MIGAP, 2 – vegetation cover classes only, 3 – full set of vegetation measurements and cover classes, a – stand-scale measurements, b – plot-scale measurements. Commission error is the percentage of sites incorrectly predicted as present, omission error is the percentage of absent sites that were incorrectly predicted. In comparison to the statistical models (Phases 2 and 3) MIGAP models show relatively high commission error rates, while keeping omission errors rates lower. Error bars show 1 standard deviation. The differences between 2a and 3a, and 2b and 3b are significant at $p<0.01$ for both omission and commission error. The differences between 2a and 2b, and 3a and 3b are not significant.

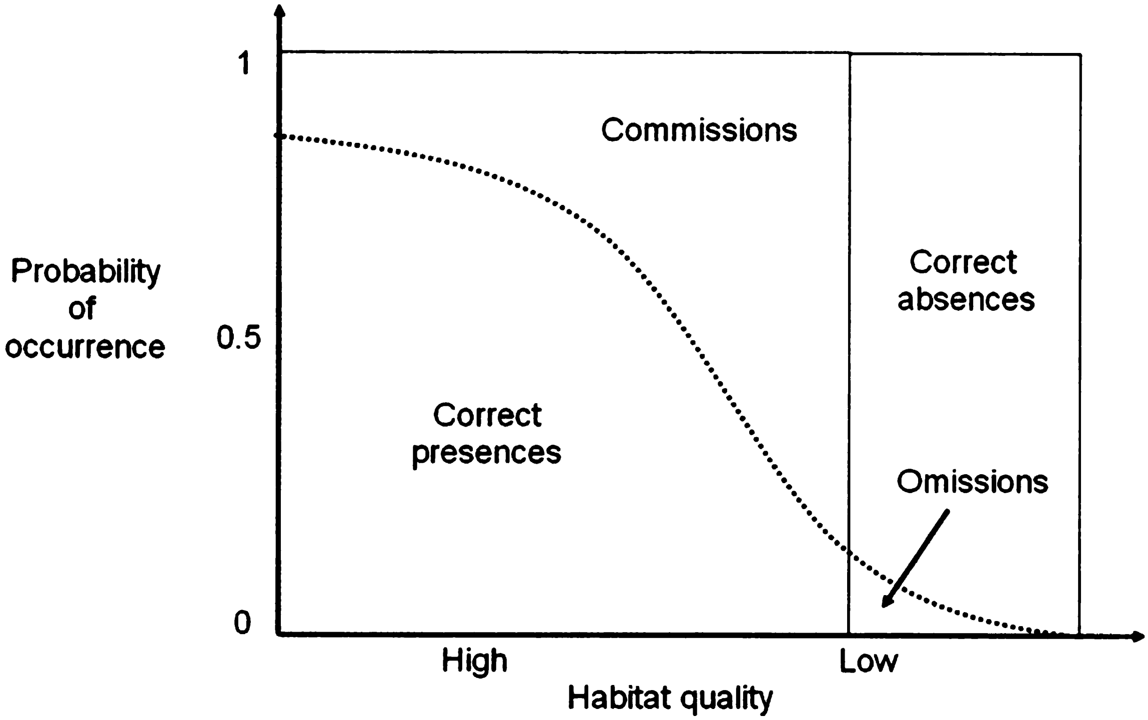

Figure 3.1a: A binary model (e.g. GAP potential habitat) for an abundant habitat generalist includes nearly all of the habitats that are potentially used by this species. Omission errors are low because a large proportion of all the sites are predicted as present in the model (shaded area), and commission errors are high because this species is present with a relatively low probability in many of the locations where the model predicts its presence.

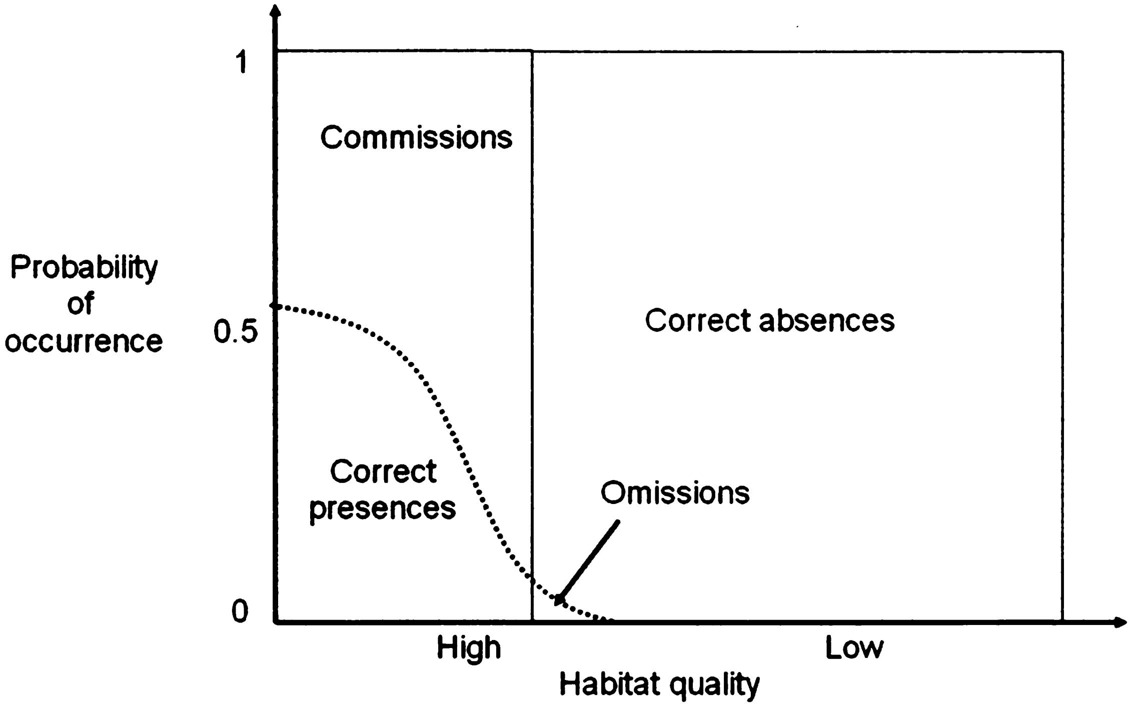

Figure 3.1b: A binary model (e.g. GAP potential habitat) for a rare habitat specialist includes nearly all of the habitats that are potentially used by this species. Omission errors are relatively low because a large proportion of all the habitats used by this species are included, and commission errors are high because this species has a relatively low probability of presence on the sites predicted as present in the model (shaded area), even on the highest quality sites.

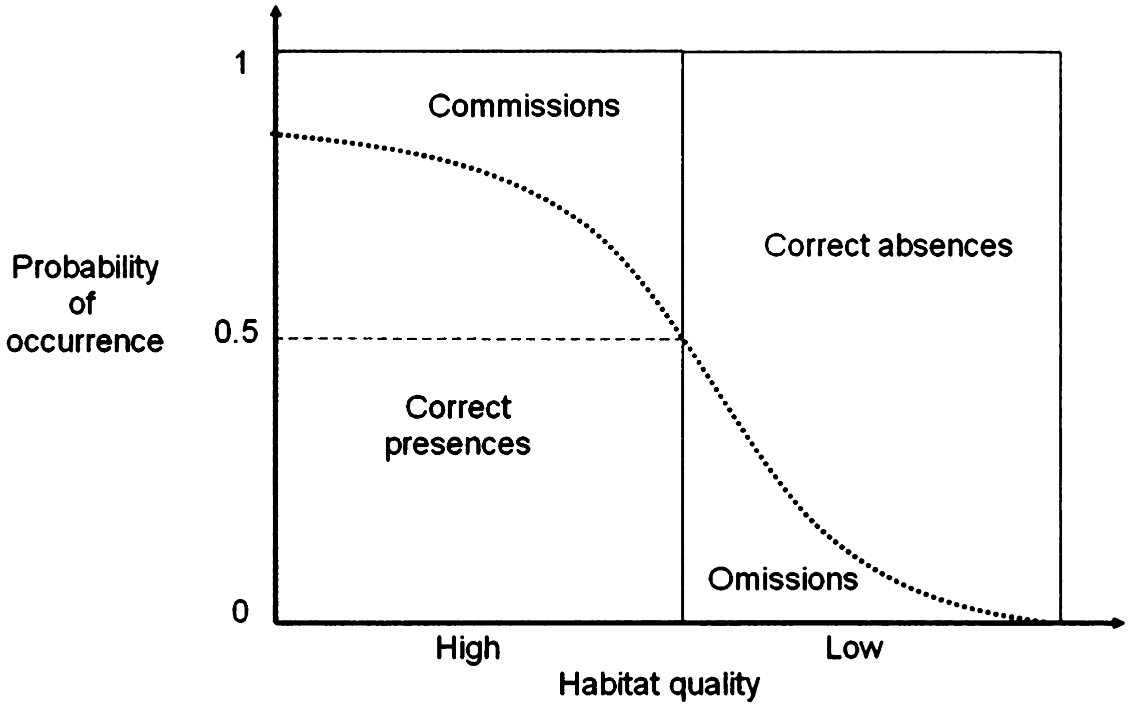

Figure 3.1c: A binary model for an abundant habitat generalist with a threshold for probability of occurrence set at 0.5. The predicted present area includes most of the habitats that are potentially used by this species, but fewer than in Figure 1a. Omission errors have increased, but commission errors are lower.

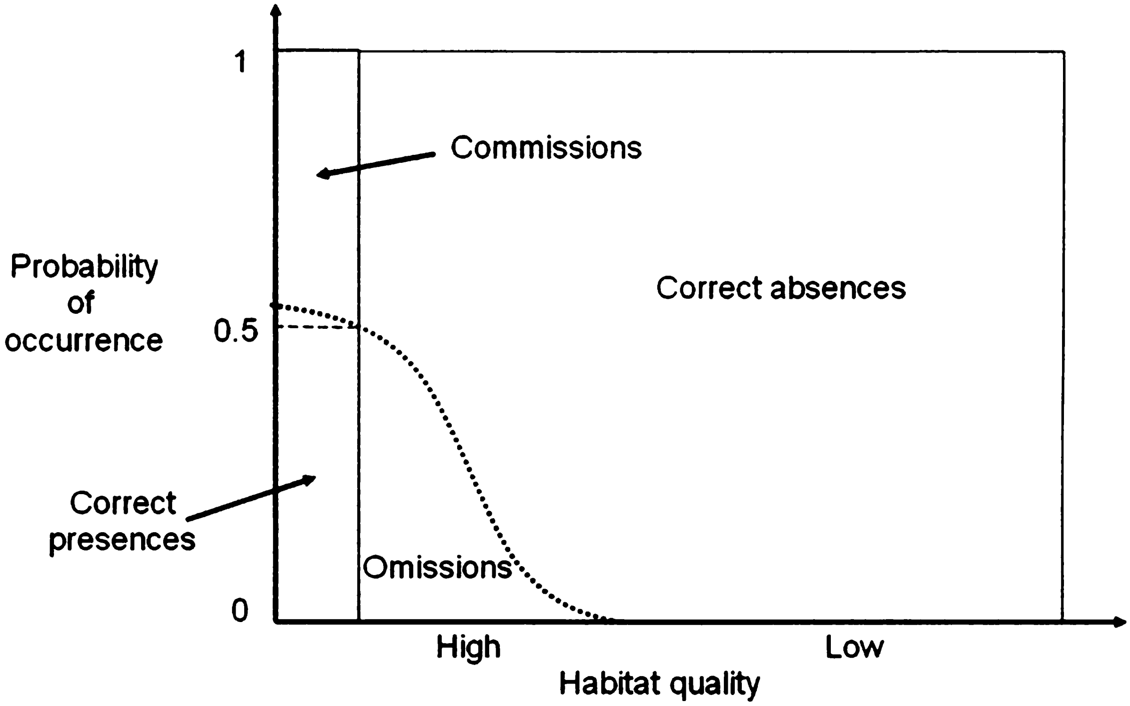

Figure 3.1d: A binary model for a less prevalent habitat specialist with a threshold for probability of occurrence set at 0.5. The predicted presence area includes a relatively small portion of the habitats that are potentially used by this species because the required probability of occurrence (0.5) is met at only a few of the sites. Omission errors are high because a large proportion of all the habitats used by this species are not included in the predicted presence set, and yet commission errors are still high because this species is present with a low probability, even on its most appropriate habitats.

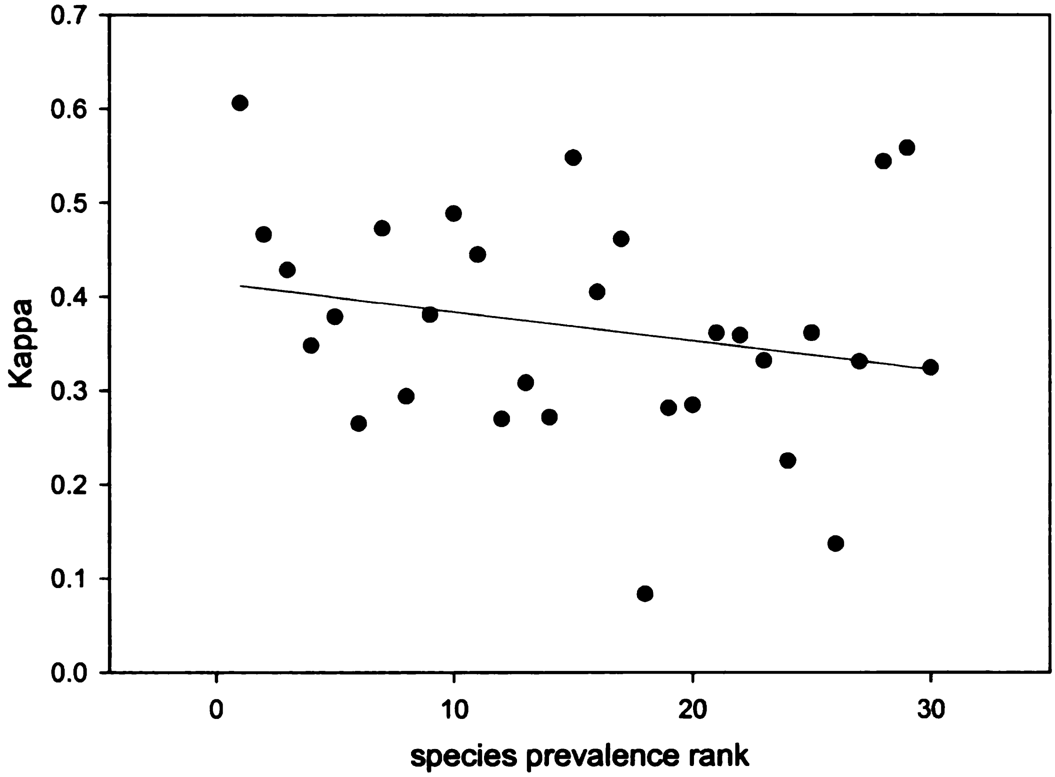

Figure 3.2a: Stand-scale model accuracy (kappa) for all 30 species as a function of species prevalence rank for stand-scale vegetation models (solid line is a linear regression, $R^2 = 0.05$). Threshold = prevalence for each species.

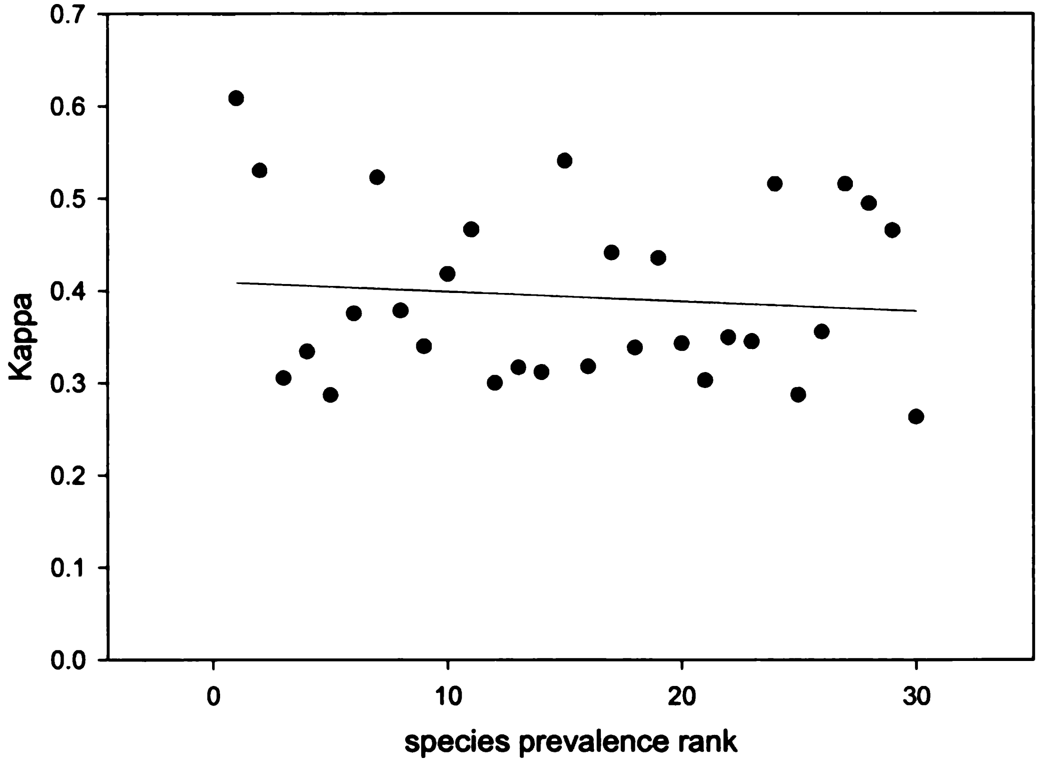

Figure 3.2b: Plot-scale model accuracy (kappa) for all 30 species as a function of species prevalence rank for plot-scale vegetation models (solid line is a linear regression, $R^2 = 0.01$). Threshold = prevalence for each species.

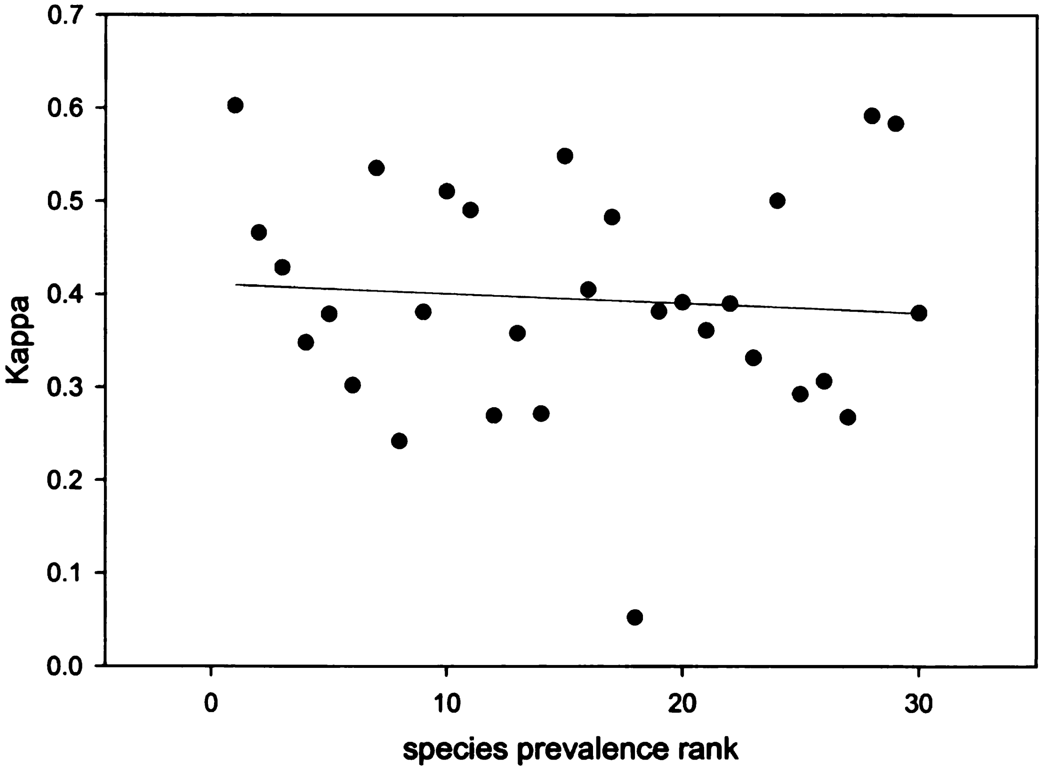

Figure 3.2c: Stand-scale model accuracy (kappa) for all 30 species as a function of species prevalence rank for stand-scale vegetation models (solid line is a linear regression, $R^2 = 0.01$). Threshold is set where the predicted prevalence of the model = actual prevalence for each species.

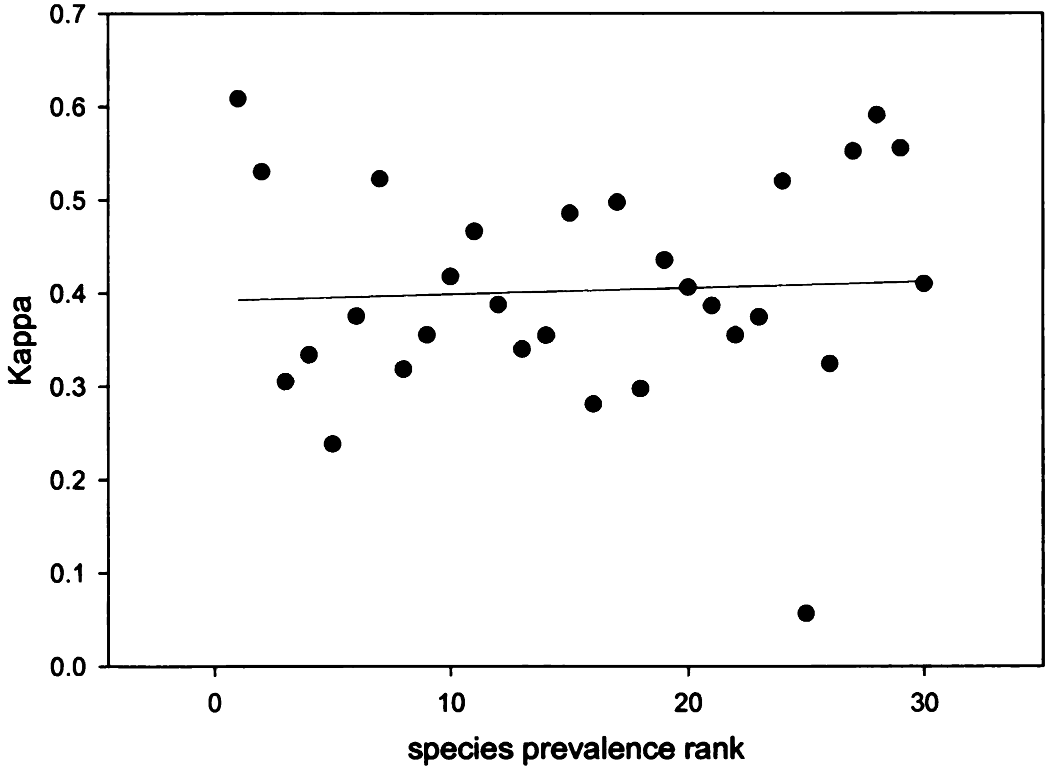

Figure 3.2d: Plot-scale model accuracy (kappa) for all 30 species as a function of species prevalence rank for plot-scale vegetation models (solid line is a linear regression, R^2 = 0.00). Threshold is set where the predicted prevalence of the model = actual prevalence for each species.

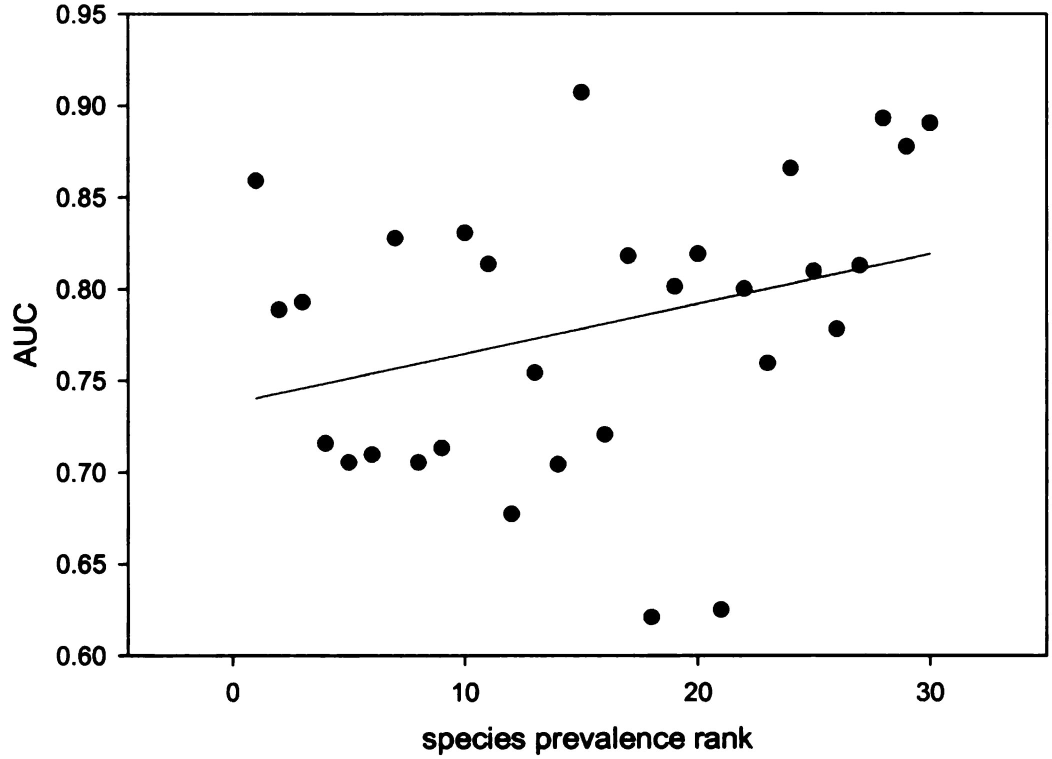

Figure 3.3a: Model accuracy (ROC/AUC) for all 30 species as a function of species prevalence for stand-scale vegetation models (solid line is linear regression, $R^2 = 0.10$).

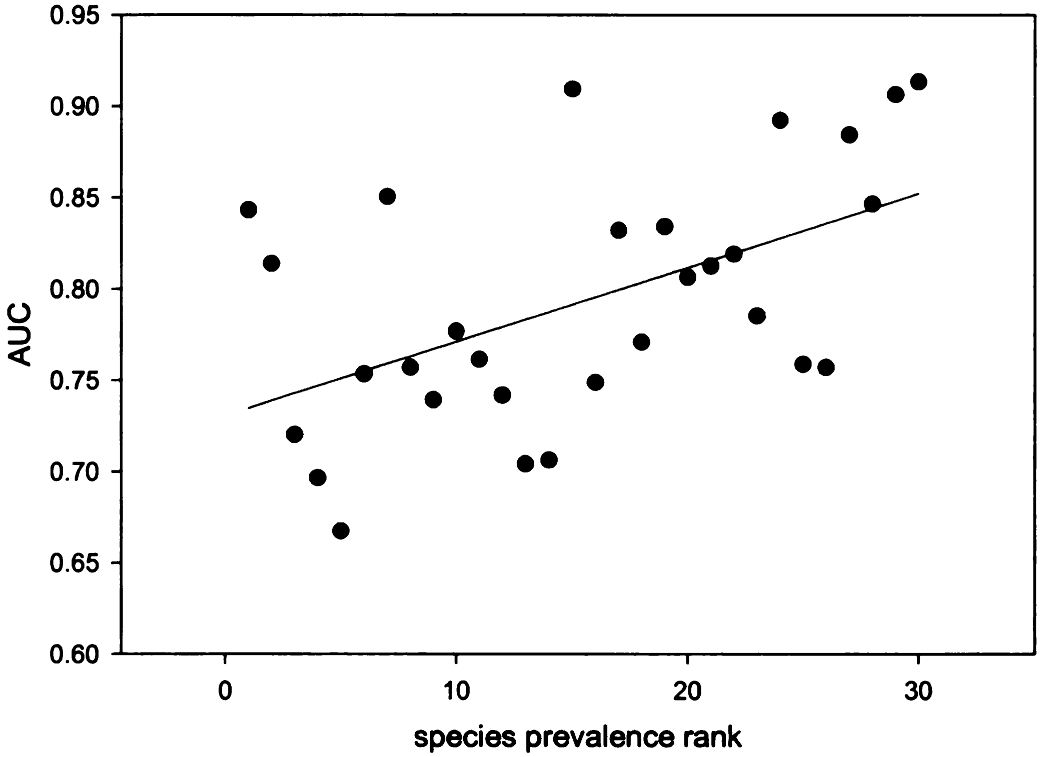

Figure 3.3b: Model accuracy (ROC/AUC) for all 30 species as a function of species prevalence for plot-scale vegetation models (solid line is linear regression, $R^2 = 0.28$).

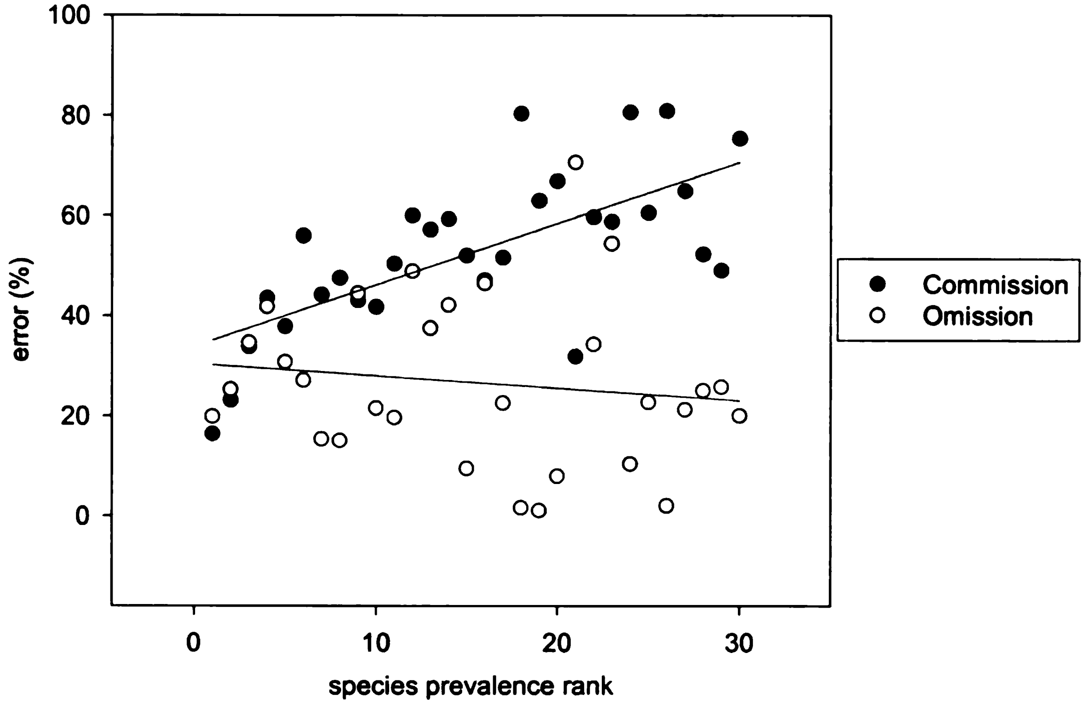

Figure 3.4a: Omission and commission error rates for all 30 species as a function of species prevalence for stand-scale vegetation models (solid lines are linear regressions, commission error $R^2 = 0.47$, omission error $R^2 = 0.02$). Threshold = prevalence for each species.

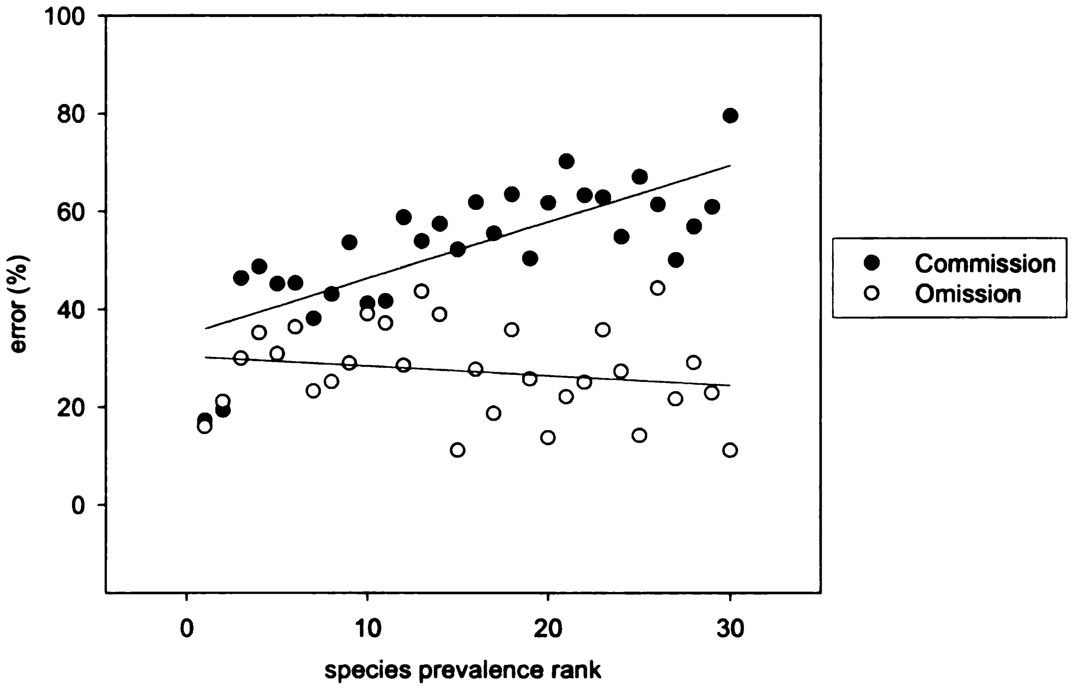

Figure 3.4b: Omission and commission error rates for all 30 species as a function of species prevalence for plot-scale vegetation models (solid lines are linear regressions, commission error $R^2 = 0.57$, omission error $R^2 = 0.04$). Threshold = prevalence for each species.

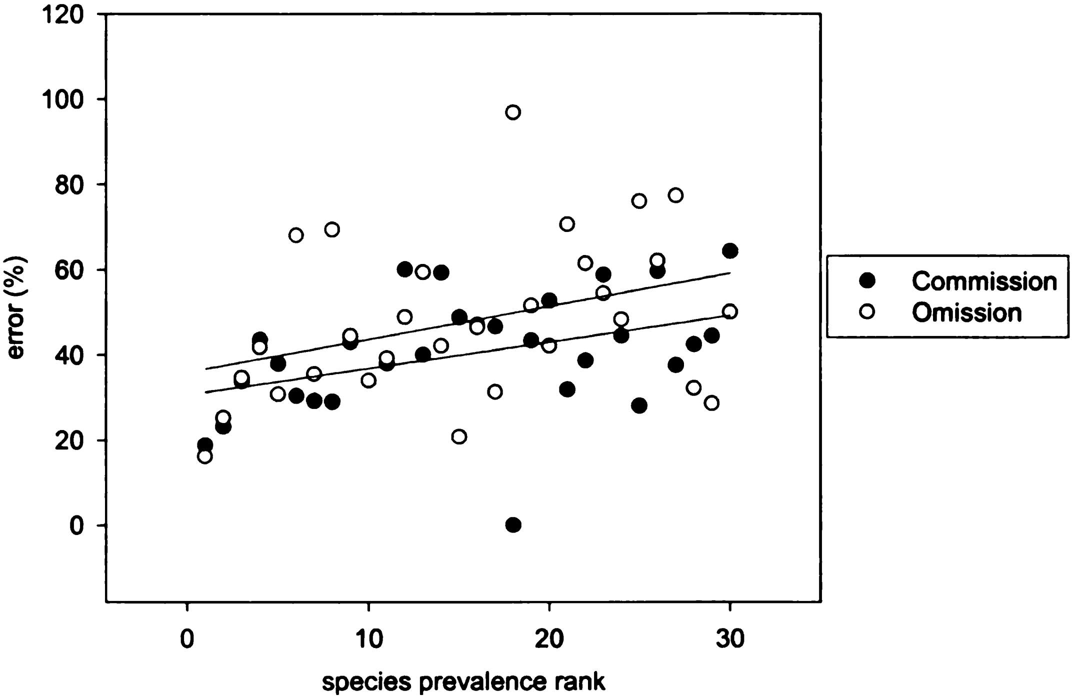

Figure 3.4c: Omission and commission error rates for all 30 species as a function of species prevalence for stand-scale vegetation models (solid lines are linear regressions, commission error $R^2 = 0.16$, omission error $R^2 = 0.13$). Threshold is set where the predicted prevalence = actual prevalence for each species.

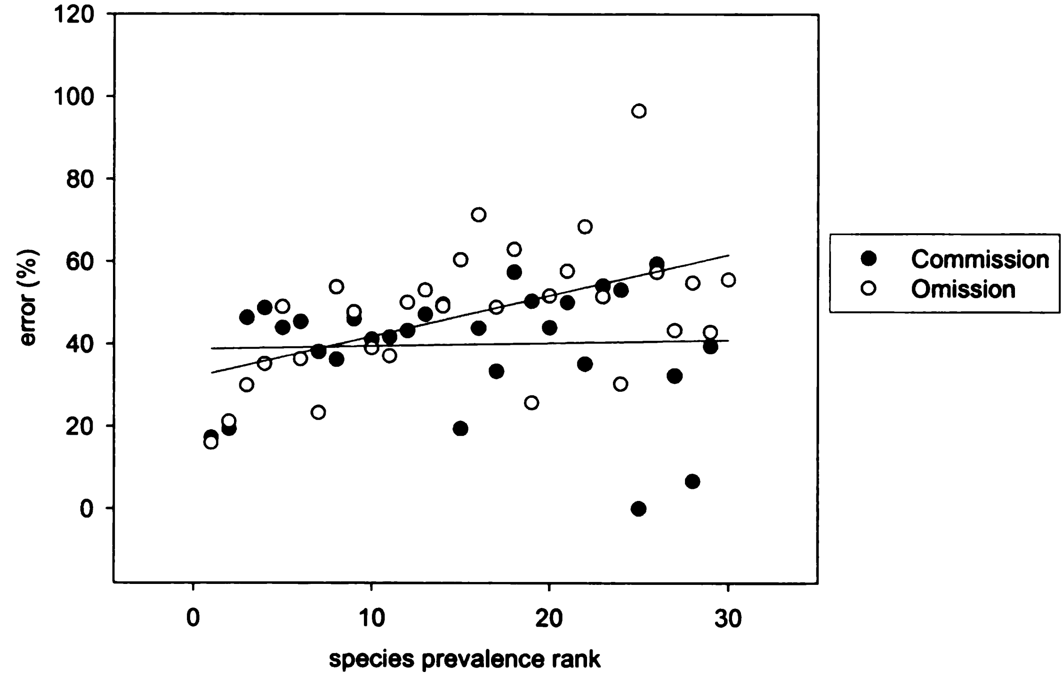

Figure 3.4d: Omission and commission error rates for all 30 species as a function of species prevalence for plot-scale vegetation models (solid lines are linear regressions, commission error $R^2 = 0.00$, omission error $R^2 = 0.27$). Threshold is set where the predicted prevalence = actual prevalence for each species.

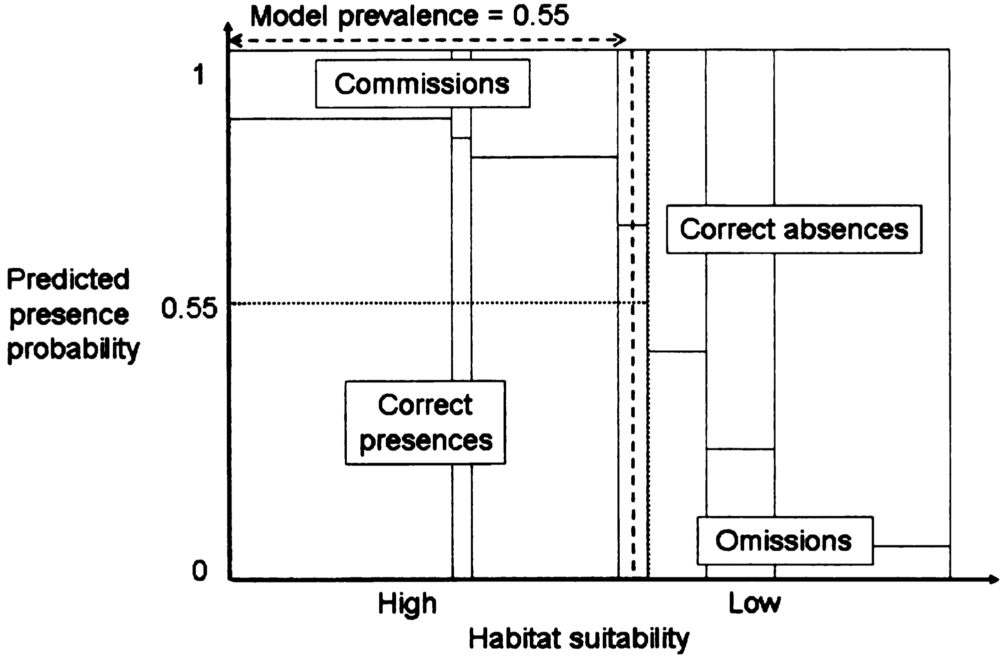

Figure 3.5a: Graphical representation of the Ovenbird plot-scale recursive partitioning model. The height of the dark shaded boxes represent the predicted presence probability for a group of sites, the width of each box represents the proportion of all sites that fall into that group. The threshold values for calculating accuracy measures are shown by the dotted and dashed lines (see text for details), and in this case both thresholds result in the same error matrix values.

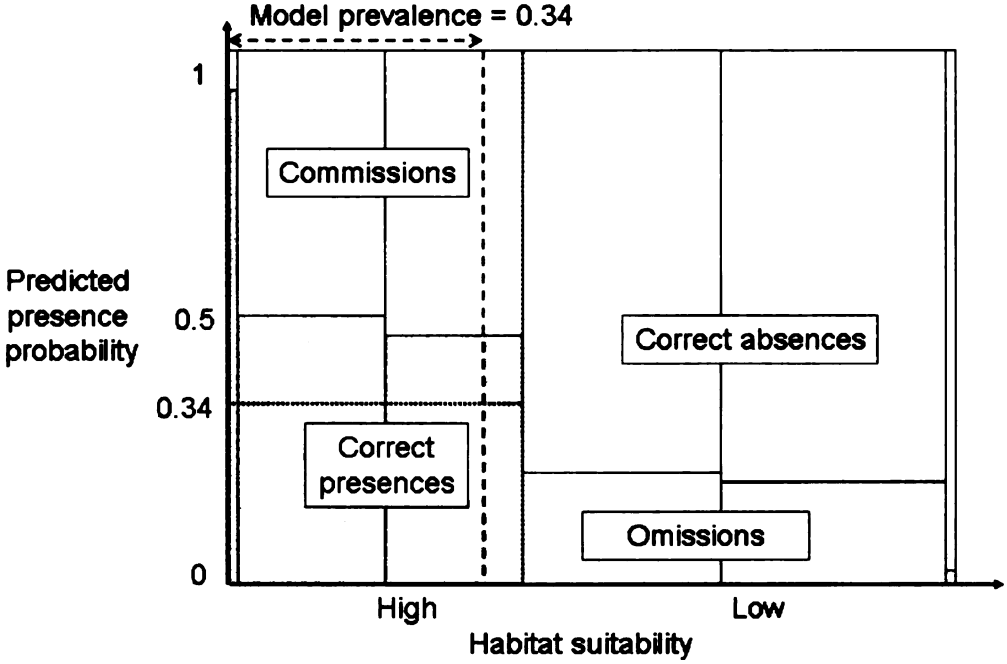

Figure 3.5b: Graphical representation of the American Robin plot-scale recursive partitioning model. The height of the dark shaded boxes represent the predicted presence probability for a group of sites, the width of each box represents the proportion of all sites that fall into that group. The threshold values for calculating accuracy measures are shown by the dotted and dashed lines (see text for details), and in this case both thresholds result in the same error matrix values.

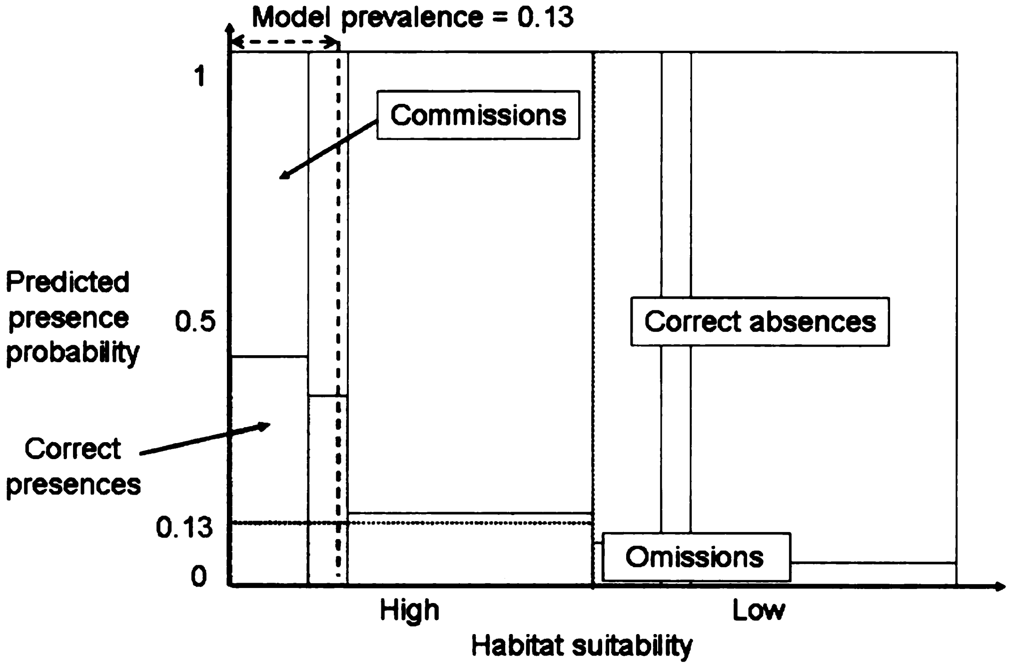

Figure 3.5c: Graphical representation of the Yellow-billed Cuckoo plot-scale recursive partitioning model. The height of the dark shaded boxes represent the predicted presence probability for a group of sites, the width of each box represents the proportion of all sites that fall into that group. The threshold values for calculating accuracy measures are shown by the dotted and dashed lines (see text for details), and in this case the two thresholds result in different error matrix values.

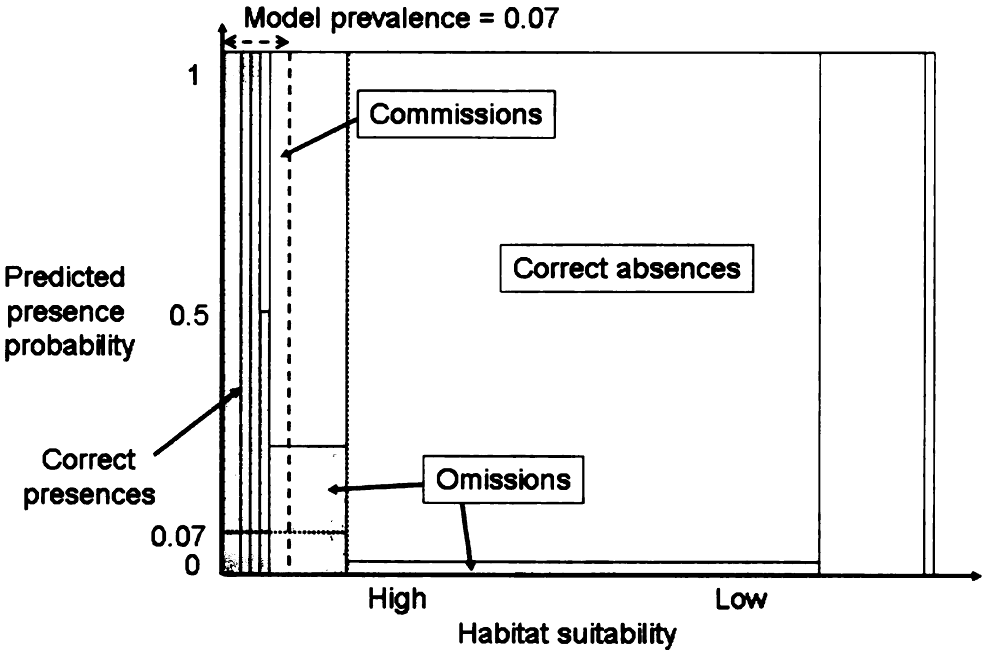

Figure 3.5d: Graphical representation of the Black-throated Green Warbler plot-scale recursive partitioning model. The height of the dark shaded boxes represent the predicted presence probability for a group of sites, the width of each box represents the proportion of all sites that fall into that group. The threshold values for calculating accuracy measures are shown by the dotted and dashed lines (see text for details), and in this case the two thresholds result in different error matrix values.

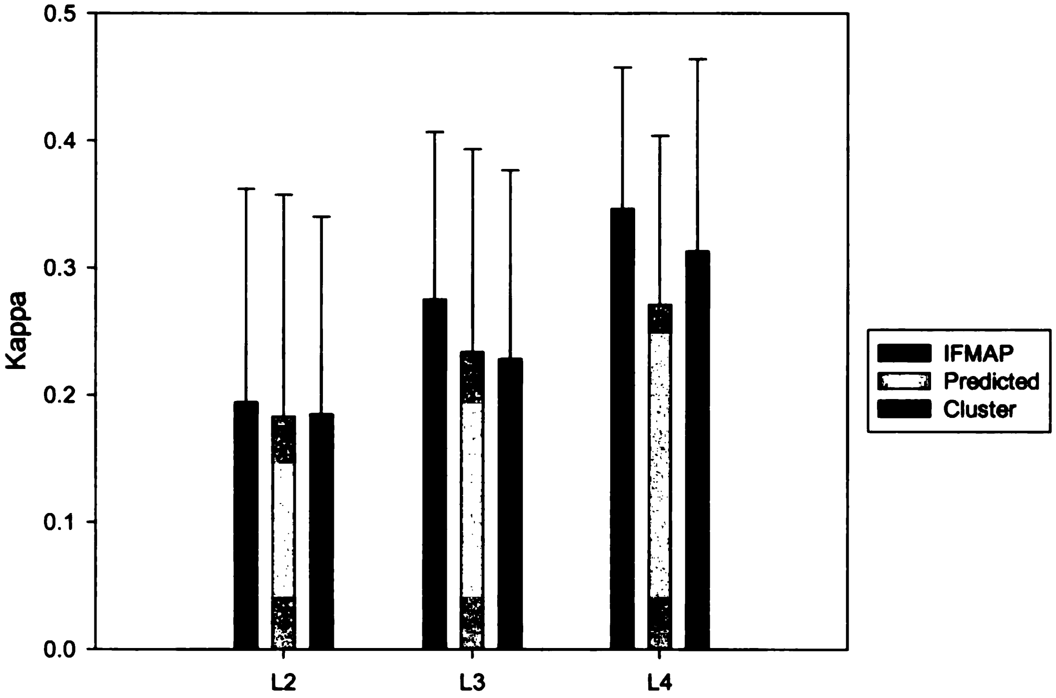

Figure 4.1: Accuracy (kappa) averaged over all 30 bird species, with error bars showing one standard deviation. Level-2 models are not significantly different from each other (paired t-test). Within level-3, the difference between IFMAP and the predicted classification models are significantly different ($p < 0.1$, paired t-test), as are IFMAP and cluster models ($p < 0.05$). At level-4, the IFMAP and predicted classifications are significantly different ($p < 0.05$), but the IFMAP and cluster classifications are not. All of the between level differences (within the same classification) are significant ($p<0.05$).

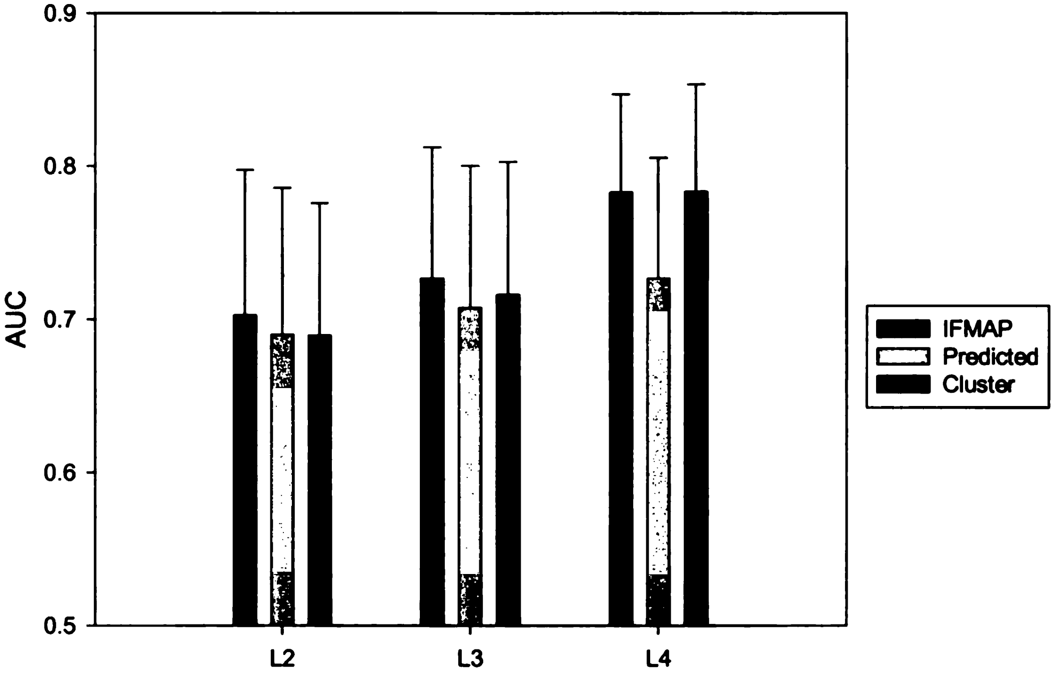

Figure 4.2: Accuracy (AUC) averaged over all 30 bird species, with error bars showing one standard deviation. None of the classifications within a given level are significantly different ($p < 0.05$, paired t-test), except for the IFMAP and cluster vs. predicted classifications in level-4. Only the between level-3 and level-4 differences for IFMAP and cluster classifications are significant ($p<0.05$).

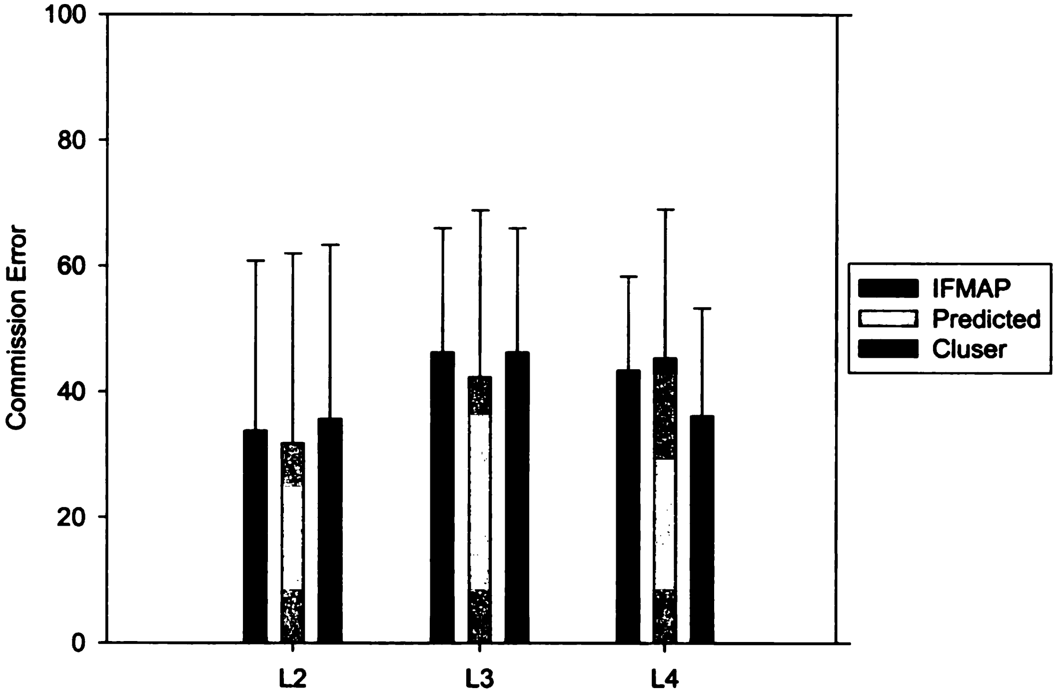

Figure 4.3a: Rates of commission error averaged over all 30 bird species, with error bars showing one standard deviation. None of the classifications within or between levels are significantly different ($p < 0.05$, paired t-test).

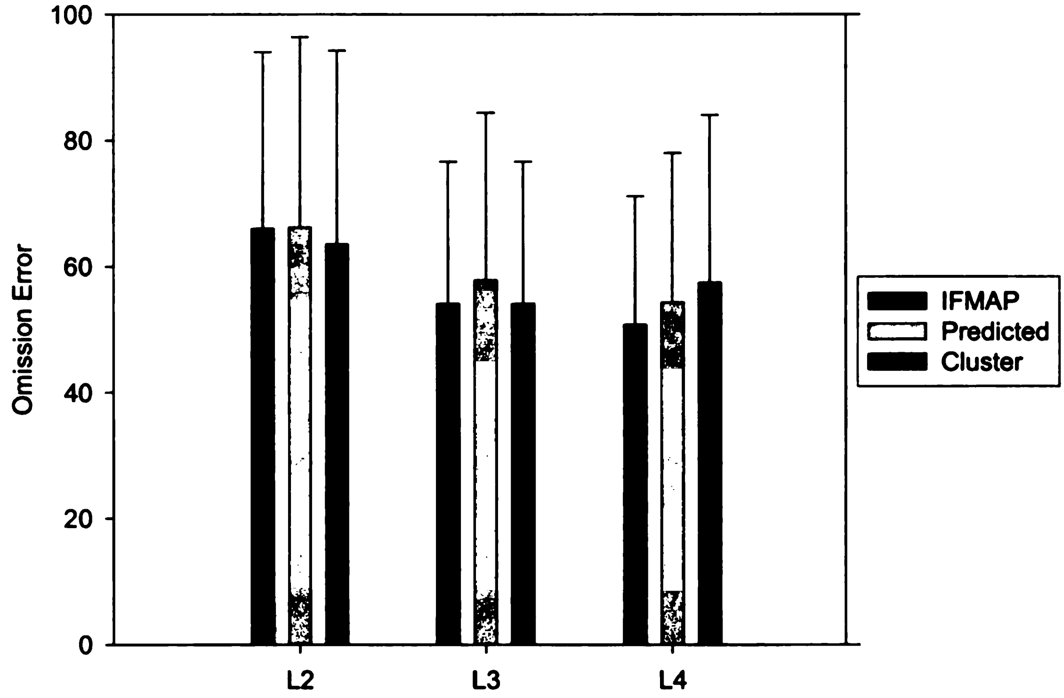

Figure 4.3b: Rates of omission error averaged over all 30 bird species, with error bars showing one standard deviation. None of the classifications within or between levels are significantly different ($p < 0.05$, paired t-test).

BIBLIOGRAPHY

Albert, D. A. 1995. Regional landscape ecosystems of Michigan, Minnesota, and Wisconsin: a working map and classification. *in*. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology **43**:1223-1232.

Anderson, D. R., and K. P. Burnham. 2002. Avoiding Pitfalls When Using Information-Theoretic Methods. Journal of Wildlife Management **66**:912-918.

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

Araujo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography **33**:1677-1688.

Atkinson, E. J., and T. M. Therneau. 2000. An Introduction to Recursive Partitioning Using RPART Routines. Mayo Foundation.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling **200**:1-19.

Bahn, V., and B. J. McGill. 2007. Can niche-based distribution models outperform spatial interpolation? Global Ecology and Biogeography **16**:733-742.

Bailey, R. G. 1983. Delineation of ecosystem regions. Environmental Management **7**:365-373.

Barth, A., J. Wallerman, and G. Stahl. 2009. Spatially consistent nearest neighbor imputation of forest stand data. Remote Sensing of Environment **113**:546-553.

Bergen, K. M., A. M. Gilboy, and D. G. Brown. 2007. Multi-dimensional vegetation structure in modeling avian habitat. Ecological Informatics **2**:9-22.

Boone, R. B., and W. B. Krohn. 2000. Predicting broad-scale occurrences of vertebrates in patchy landscapes. Landscape Ecology **15**:63-74.

Borges, J. G., and H. M. Hoganson. 1999. Assessing the impact of management unit design and adjacency constraints on forestwide spatial conditions and timber revenues. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **29**:1764-1774.

Bourgeron, P. S. 1988. Advantages and Limitations of Ecological Classification for the Protection of Ecosystems. Conservation Biology **2**:218-220.

Bray, J. R., and J. T. Curtis. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs **27**:326-349.

Brewer, R., G. A. McPeek, and R. J. Adams, editors. 1991. The atlas of breeding birds of Michigan. Michigan State University Press, East Lansing, MI.

Budreski, K. A., R. H. Wynne, J. O. Browder, and J. B. Campbell. 2007. Comparison of segment and pixel-based non-parametric land cover classification in the Brazilian Amazon using multitemporal landsat TM/ETM+ imagery. Photogrammetric Engineering and Remote Sensing **73**:813-827.

Cohen, W. B., T. K. Maiersperger, S. T. Gower, and D. P. Turner. 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sensing of Environment **84**:561-571.

Cook, J. E. 1996. Implications of modern successional theory for habitat typing: A review. Forest Science **42**:67-75.

Daubenmire, R. 1952. Forest Vegetation of Northern Idaho and Adjacent Washington, and Its Bearing on Concepts of Vegetation Classification. Ecological Monographs **22**:301-330.

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

Denoeux, T. 1995. A K-Nearest Neighbor Classification Rule-Based on Dempster-Shafer Theory. Ieee Transactions on Systems Man and Cybernetics **25**:804-813.

Donovan, M. L., G. M. Nesslage, J. J. Skillen, and B. A. Maurer. 2004. The Michigan Gap Analysis Project Final Report. Michigan Department of Natural Resources - Wildlife Division, Lansing, MI.

Donovan, T. M., and C. H. Flather. 2002. Relationships among north American songbird trends, habitat fragmentation, and landscape occupancy. Ecological Applications **12**:364-374.

Dymond, C. C., D. J. Mladenoff, and V. C. Radeloff. 2002. Phenological differences in Tasseled Cap indices improve deciduous forest classification. Remote Sensing of Environment **80**:460-472.

Edwards, T. C., E. T. Deshler, D. Foster, and G. G. Moisen. 1996. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conservation Biology **10**:263-270.

Edwards, T. C., G. G. Moisen, and D. R. Cutler. 1998. Assessing map accuracy in a remotely sensed, ecoregion-scale cover map. Remote Sensing of Environment **63**:73-83.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Fahrig, L. 2001. How much habitat is enough? Biological Conservation **100**:65-74.

Feldesman, M. R. 2002. Classification trees as an alternative to linear discriminant analysis. American Journal of Physical Anthropology **119**:257-275.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation **24**:38-49.

Flather, C. H., S. J. Brady, and D. B. Inkley. 1992. Regional Habitat Appraisals of Wildlife Communities - a Landscape-Level Evaluation of a Resource Planning-Model Using Avian Distribution Data. Landscape Ecology **7**:137-147.

Franco-Lopez, H., A. R. Ek, and M. E. Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment **77**:251-274.

Freeman, E. A., and G. Moisen. 2008a. PresenceAbsence: An R package for presence absence analysis. Journal of Statistical Software **23**.

Freeman, E. A., and G. G. Moisen. 2008b. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modelling **217**:48-58.

Gottschalk, T. K., F. Huettmann, and M. Ehlers. 2005. Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. International Journal of Remote Sensing **26**:2631-2656.

Greig-Smith, P., M. P. Austin, and T. C. Whitmore. 1967. Application of Quantitative Methods to Vegetation Survey .I. Association-Analysis and Principal Component Ordination of Rain Forest. Journal of Ecology **55**:483-&.

Grime, J. P. 1974. Vegetation classification by reference to strategies. Nature **250**:26-31.

Grinnell, J. 1917. The Niche-Relationships of the California Thrasher. Auk **34**:427-433.

Grossman, D. H., P. S. Bourgeron, W.-D. N. Busch, D. T. Cleland, W. Platts, G. C. Ray, C. R. Robbins, and G. J. Roloff. 1999. Principles for Ecological Classification. Pages 353-393 *in* R. C. Szaro, N. C. Johnson, W. T. Sexton, and A. J. Malk, editors. Ecological Stewardship: A Common Reference for Ecosystem Management. Elsevier Science Ltd., Oxford, UK.

Gu, W. D., R. Heikkila, and I. Hanski. 2002. Estimating the consequences of habitat fragmentation on extinction risk in dynamic landscapes. Landscape Ecology **17**:699-710.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? Ecological Monographs **77**:615-630.

Gustafson, E. J. 1998. Quantifying landscape spatial pattern: What is the state of the art? Ecosystems **1**:143-156.

Haapanen, R., and A. Ek. 2001. Software and Instructions for kNN Applications in Forest Resources Description and Estimation. *in* Staff Paper Series in the Department of Forest Resources at the University of Minnesota, St. Paul, MN.

Haapanen, R., A. R. Ek, M. E. Bauer, and A. O. Finley. 2004. Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment **89**:265-271.

Hanley, T. A., W. P. Smith, and S. M. Gende. 2005. Maintaining wildlife habitat in southeastern Alaska: implications of new knowledge for forest management and research. Landscape and Urban Planning **72**:113-133.

Hansen, A. J., J. J. Rotella, M. P. V. Kraska, and D. Brown. 1999. Dynamic habitat and population analysis: An approach to resolve the biodiversity manager's dilemma. Ecological Applications **9**:1459-1476.

Hansen, M. H., T. Frieswyk, J. F. Glover, and J. F. Kelly. 1992. The Eastwide forest inventory data base: users manual. Gen. Tech. Rep. NC-151, U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Hansen, M. J., S. E. Franklin, C. Woudsma, and M. Peterson. 2001. Forest structure classification in the North Columbia mountains using the Landsat TM Tasseled Cap wetness component. Canadian Journal of Remote Sensing **27**:20-32.

He, H. S., D. J. Mladenoff, V. C. Radeloff, and T. R. Crow. 1998. Integration of gis data and classified satellite imagery for regional forest assessment. Ecological Applications **8**:1072-1083.

Heikkinen, R. K., M. Luoto, R. Virkkala, and K. Rainio. 2004. Effects of habitat cover, landscape structure and spatial variables on the abundance of birds in an agricultural-forest mosaic. Journal of Applied Ecology **41**:824-835.

Hepinstall, J. A., W. B. Krohn, and S. A. Sader. 2002. Effects of Niche Width on the Performance and Agreement of Avian Habitat Models. Pages 868 *in* J. M. Scott, P. J. Heglund, and M. L. Morrison, editors. Predicting Species Occurrences. Island Press.

Hernandez, P. A., I. Franke, S. K. Herzog, V. Pacheco, L. Paniagua, H. L. Quintana, A. Soto, J. J. Swenson, C. Tovar, T. H. Valqui, J. Vargas, and B. E. Young. 2008. Predicting species distributions in poorly-studied landscapes. Biodiversity and Conservation **17**:1353-1366.

Herzog, S. K., and M. Kessler. 2006. Local vs. regional control on species richness: a new approach to test for competitive exclusion at the community level. Global Ecology and Biogeography **15**:163-172.

Holmstrom, H., and J. E. S. Fransson. 2003. Combining remotely sensed optical and radar data in kNN-estimation of forest variables. Forest Science **49**:409-418.

Holmstrom, H., M. Nilsson, and G. Stahl. 2001. Simultaneous estimations of forest parameters using aerial photograph interpreted data and the k nearest neighbour method. Scandinavian Journal of Forest Research **16**:67-78.

Howe, R. W., G. J. Niemi, S. J. Lewis, and D. A. Welsh. 1997. A Standard Method for Measuring Songbird Populations in the Great Lakes Region. The Passenger Pigeon **59**:183-194.

Imhoff, M. L., T. D. Sisk, A. Milne, G. Morgan, and T. Orr. 1997. Remotely sensed indicators of habitat heterogeneity: Use of synthetic aperture radar in mapping vegetation structure and bird habitat. Remote Sensing of Environment **60**:217-227.

Jetz, W., C. H. Sekercioglu, and J. E. M. Watson. 2008. Ecological correlates and conservation implications of overestimating species geographic ranges. Conservation Biology **22**:110-119.

Karl, J. W., P. J. Heglund, E. O. Garton, J. M. Scott, N. M. Wright, and R. L. Hutto. 2000. Sensitivity of species habitat-relationship model performance to factors of scale. Ecological Applications **10**:1690-1705.

Karl, J. W., L. K. Svancara, P. J. Heglund, N. M. Wright, and J. M. Scott. 2002. Species Commonness and the Accuracy of Habiat-relationship Models. Pages 868 *in* J. M. Scott, P. J. Heglund, and M. L. Morrison, editors. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press.

Karl, J. W., N. M. Wright, P. J. Heglund, and J. M. Scott. 1999. Obtaining environmental measures to facilitate vertebrate habitat modeling. Wildlife Society Bulletin **27**:357-365.

Kiester, A. R., J. M. Scott, B. Csuti, R. F. Noss, B. Butterfield, K. Sahr, and D. White. 1996. Conservation prioritization using GAP data. Conservation Biology **10**:1332-1342.

Knick, S. T., and J. T. Rotenberry. 2000. Ghosts of habitats past: Contribution of landscape change to current habitats used by shrubland birds. Ecology **81**:220-227.

Koukal, T., F. Suppan, and W. Schneider. 2007. The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. Remote Sensing of Environment **110**:431-437.

Kuchler, A. W. 1951. The Relation between Classifying and Mapping Vegetation. Ecology **32**:275-283.

Lawler, J. J., and T. C. Edwards. 2006. A variance-decomposition approach to investigating multiscale habitat associations. Condor **108**:47-58.

Lawler, J. J., R. J. O'Connor, C. T. Hunsaker, K. B. Jones, T. R. Loveland, and D. White. 2004. The effects of habitat resolution on models of avian diversity and distributions: a comparison of two land-cover classifications. Landscape Ecology **19**:515-530.

Linder, E. T., M. A. Villard, B. A. Maurer, and E. V. Schmidt. 2000. Geographic range structure in North American landbrids: variation with migratory strategy, trophic level, and breeding habitat. Ecography **23**:678-686.

Liu, C. M., L. J. Zhang, C. J. Davis, D. S. Solomon, T. B. Brann, and L. E. Caldwell. 2003. Comparison of neural networks and statistical methods in classification of ecological habitats using FIA data. Forest Science **49**:619-631.

Lobo, J. M., A. Jimenez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography **17**:145-151.

Lu, D. S., P. Mausel, E. Brondizio, and E. Moran. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management **198**:149-167.

MacFaden, S. W., and D. E. Capen. 2002. Avian habitat relationships at multiple scales in a New England forest. Forest Science **48**:243-253.

MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving inferences in popoulation studies of rare species that are detected imperfectly. Ecology **86**:1101-1113.

Maechler, M. 2008. Cluster Analysis Extended Rousseauw et al. 1.11.11.

Makela, H., and A. Pekkarinen. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecology and Management **196**:245-255.

Maltamo, M., and A. Kangas. 1998. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **28**:1107-1115.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology **38**:921-931.

Manley, P. N., W. J. Zielinski, M. D. Schlesinger, and S. R. Mori. 2004. Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. Ecological Applications **14**:296-310.

McPherson, J. M., and W. Jetz. 2007. Effects of species' ecology on the accuracy of distribution models. Ecography **30**:135-151.

McPherson, J. M., W. Jetz, and D. J. Rogers. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Journal of Applied Ecology **41**:811-823.

MDNR. 2001. MIGAP Land Cover. *in*. Michigan Department of Natural Resources - Forest, Mineral and Fire Management Division, Lansing, MI.

MDNR. 2004. Review of Remote Sensing Technologies used in the IFMAP Project: Final Report. Space Imaging and Michigan Department of Natural Resources, Ann Arbor, MI.

MDNR. 2005. IFMAP Field Manual. Field Manual Michigan Department of Natural Resources, Lansing, MI.

Meng, Q. M., C. J. Cieszewski, M. Madden, and B. E. Borders. 2007. K nearest neighbor method for forest inventory using remote sensing data. Giscience & Remote Sensing **44**:149-165.

Moisen, G. G., E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards. 2006. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological Modelling **199**:176-187.

Moisen, G. G., and T. S. Frescino. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling **157**:209-225.

Mortberg, U. M. 2001. Resident bird species in urban forest remnants; landscape and habitat perspectives. Landscape Ecology **16**:193-203.

Noon, B. R., D. D. Murphy, S. R. Beissinger, M. L. Shaffer, and D. Dellasala. 2003. Conservation planning for US National Forests: Conducting comprehensive biodiversity assessments. Bioscience **53**:1217-1220.

Osborne, P. E., J. C. Alonso, and R. G. Bryant. 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. Journal of Applied Ecology **38**:458-471.

Peterjohn, B. G., and J. R. Sauer. 1993. North American Breeding Bird Survey annual summary 1990-1991. Bird Populations **1**:1-24.

Peterson, A. T. 2005. Kansas Gap Analysis: The importance of validating distributional models before using them. Southwestern Naturalist **50**:230-236.

Peterson, A. T., and D. A. Kluza. 2003. New distributional modelling approaches for gap analysis. Animal Conservation **6**:47-54.

Pfister, R. D., and S. F. Arno. 1980. Classifying Forest Habitat Types Based on Potential Climax Vegetation. Forest Science **26**:52-70.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems **9**:181-199.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters **3**:349-361.

Ralph, C. J., J. R. Sauer, and S. Droege, editors. 1995. Monitoring Bird Populations by Point Counts. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, . Albany, CA.

Reese, H., M. Nilsson, P. Sandstrom, and H. Olsson. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. Computers and Electronics in Agriculture **37**:37-55.

Riitters, K. H., R. V. O'Neill, C. T. Hunsaker, J. D. Wickham, D. H. Yankee, S. P. Timmins, K. B. Jones, and B. L. Jackson. 1995. A factor analysis of landscape pattern and structure metrics. Landscape Ecology **10**:23-39.

Rodriguez, J. P., L. Brotons, J. Bustamante, and J. Seoane. 2007. The application of predictive modelling of species distribution to biodiversity conservation. Diversity and Distributions **13**:243-251.

Roloff, G. J., M. L. Donovan, D. W. Linden, and M. L. Strong. 2008. Lessons Learned from Using GIS to Model Landscape-Level Wildlife Habitat. Pages 287-320 *in* J. J. Millspaugh and F. R. Thompson, editors. Models for Planning Wildlife Conservation in Large Landscapes. Elsevier, Burlington, MA.

Royle, J. A., J. D. Nichols, and M. Kery. 2005. Modelling occurrence and abundance of species when detection is imperfect. Oikos **110**:353-359.

Scarth, P., and S. Phinn. 2000. Determining forest structural attributes using an inverted geometric-optical model in mixed eucalypt forests, Southeast Queensland, Australia. Remote Sensing of Environment **71**:141-157.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. Derchia, T. C. Edwards, J. Ulliman, and R. G. Wright. 1993. Gap Analysis - a Geographic Approach to Protection of Biological Diversity. Wildlife Monographs:1-41.

Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson. 2002. Introduction. *in* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting Species Occurrences - Issues of Accuracy and Scale. Island Press, Washington.

Segurado, P., and M. B. Araujo. 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography **31**:1555-1568.

Segurado, P., M. B. Araujo, and W. E. Kunin. 2006. Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology **43**:433-444.

Seoane, J., J. Bustamante, and R. Diaz-Delgado. 2004a. Are existing vegetation maps adequate to predict bird distributions? Ecological Modelling **175**:137-149.

Seoane, J., J. Bustamante, and R. Diaz-Delgado. 2004b. Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. Ecological Modelling **171**:209-222.

Seoane, J., J. Bustamante, and R. Diaz-Delgado. 2005a. Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conservation Biology **19**:512-522.

Seoane, J., L. M. Carrascal, C. L. Alonso, and D. Palomino. 2005b. Species-specific traits associated to prediction errors in bird habitat suitability modelling. Ecological Modelling **185**:299-308.

Storch, D., and A. L. Sizling. 2002. Patterns of commonness and rarity in central European birds: reliability of the core-satellite hypothesis within a large scale. Ecography **25**:405-416.

Thompson, I. D., J. A. Baker, and M. Ter-Mikaelian. 2003. A review of the long-term effects of post-harvest silviculture on vertebrate wildlife, and predictive models, with an emphasis on boreal forests in Ontario, Canada. Forest Ecology and Management **177**:441-469.

Tomppo, E., M. Nilsson, M. Rosengren, P. Aalto, and P. Kennedy. 2002. Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. Remote Sensing of Environment **82**:156-171.

Trotter, C. M., J. R. Dymond, and C. J. Goulding. 1997. Estimation of timber volume in a coniferous plantation forest using Landsat TM. International Journal of Remote Sensing **18**:2209-2223.

Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distributions **13**:397-405.

Tuominen, S., S. Fish, and S. Poso. 2003. Combining remote sensing, data from earlier inventories, and geostatistical interpolation in multisource forest inventory. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **33**:624-634.

Turner, D. P., W. B. Cohen, R. E. Kennedy, K. S. Fassnacht, and J. M. Briggs. 1999. Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. Remote Sensing of Environment **70**:52-68.

Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology **42**:720-730.

Villard, M. A., M. K. Trzcinski, and G. Merriam. 1999. Fragmentation effects on forest birds: Relative influence of woodland cover and configuration on landscape occupancy. Conservation Biology **13**:774-783.

Watt, A. S. 1947. Pattern and Process in the Plant Community. Journal of Ecology **35**:1-22.

Welsh, H. H., J. R. Dunk, and W. J. Zielinski. 2006. Developing and applying habitat models using forest inventory data: An example using a terrestrial salamander. Journal of Wildlife Management **70**:671-681.

White, M. A., and D. J. Mladenoff. 1994. Old-Growth Forest Landscape Transitions from Pre-European Settlement to Present. Landscape Ecology **9**:191-205.

Whittaker, R. H., and S. A. Levin, editors. 1975. Niche: theory and application. Dowden, Hutchinson and Ross, Stroudsburg, Pennsylvania.

Whittaker, R. J., M. B. Araujo, J. Paul, R. J. Ladle, J. E. M. Watson, and K. J. Willis. 2005. Conservation Biogeography: assessment and prospect. Diversity and Distributions **11**:3-23.

Wigley, T. B., and T. H. Roberts. 1997. Landscape-level effects of forest management on faunal diversity in bottomland hardwoods. Forest Ecology and Management **90**:141-154.

Wilson, K. A., M. I. Westphal, H. P. Possingham, and J. Elith. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation **122**:99-112.

Wolter, P. T., D. J. Mladenoff, G. E. Host, and T. R. Crow. 1995. Improved Forest Classification in the Northern Lake-States Using Multitemporal Landsat Imagery. Photogrammetric Engineering and Remote Sensing **61**:1129-1143.

Wulder, M. 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Progress in Physical Geography **22**:449-476.

Wulder, M. A., and S. E. Franklin, editors. 2003. Remote Sensing of Forest Environments: Concepts and Case Studies. Kluwer Academic Publishers, Boston.

Wulder, M. A., R. J. Hall, N. C. Coops, and S. E. Franklin. 2004. High spatial resolution remotely sensed data for ecosystem characterization. Bioscience **54**:511-521.

Xian, G., Z. Zhu, M. Hoppus, and M. Fleming. 2002. Application of Decision-Tree Techniques to Forest Group and Basal Area Mapping Using Satellite Imagery and Forest Inventory Data. FIEOS 2002 Conference Proceedings.

Zhang, Y., and B. Guindon. 2003. Quantitative assessment of a haze suppression methodology for satellite imagery: Effect on land cover classification performance. Ieee Transactions on Geoscience and Remote Sensing **41**:1082-1089.

Zhang, Y., B. Guindon, and J. Cihlar. 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. Remote Sensing of Environment **82**:173-187.

Zheng, D. L., J. Rademacher, J. Q. Chen, T. Crow, M. Bresee, J. le Moine, and S. R. Ryu. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. Remote Sensing of Environment **93**:402-411.