Evaluating Population-Habitat Relationships of Forest Breeding Birds at Multiple Spatial and Temporal Scales Using Forest Inventory and Analysis Data

Todd M. Fearer

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Fisheries and Wildlife Sciences

Stephen P. Prisley, Co-Chair
Dean F. Stauffer, Co-Chair
Patrick D. Keyser
Carola A. Haas
Richard G. Oderwald

6th October 2006
Blacksburg, Virginia

Keywords: Akaike's Information Criterion, Breeding Bird Survey, classification and regression trees, forest breeding birds, Forest Inventory and Analysis, generalized linear model, landscape, logistic regression, multiscale, wildlife habitat modeling

Evaluating Population-Habitat Relationships of Forest Breeding Birds at Multiple Spatial and Temporal Scales Using Forest Inventory and Analysis Data

Todd M. Fearer

ABSTRACT

Multiple studies have documented declines of forest breeding birds in the eastern United States, but the temporal and spatial scales of most studies limit inference regarding large scale bird-habitat trends. A potential solution to this challenge is integrating existing long-term datasets such as the U.S. Forest Service Forest Inventory and Analysis (FIA) program and U.S. Geological Survey Breeding Bird Survey (BBS) that span large geographic regions. The purposes of this study were to determine if FIA metrics can be related to BBS population indices at multiple spatial and temporal scales and to develop predictive models from these relationships that identify forest conditions favorable to forest songbirds. I accumulated annual route-level BBS data for 4 species guilds (canopy nesting, ground and shrub nesting, cavity nesting, early successional), each containing a minimum of five bird species, from 1966-2004. I developed 41 forest variables describing forest structure at the county level using FIA data from for the 2000 inventory cycle within 5 physiographic regions in 14 states (AL, GA, IL, IN, KY, MD, NC, NY, OH, PA, SC, TN, VA, and WV). I examine spatial relationships between the BBS and FIA data at 3 hierarchical scales: 1) individual BBS routes, 2) FIA units, and 3) and physiographic sections. At the BBS route scale, I buffered each BBS route with a 100m, 1km, and 10km buffer, intersected these buffers with the county boundaries, and developed a weighted average for each forest variable within each buffer, with the weight being a function of the percent of area each county had within a given buffer. I calculated 28 variables describing landscape structure from 1992 NLCD imagery using Fragstats within each buffer size. I developed predictive models relating spatial variations in bird occupancy and abundance to changes in

forest and landscape structure using logistic regression and classification and regression trees (CART). Models were developed for each of the 3 buffer sizes, and I pooled the variables selected for the individual models and used them to develop multiscale models with the BBS route still serving as the sample unit. At the FIA unit and physiographic section scales I calculated average abundance/route for each bird species within each FIA unit and physiographic section and extrapolated the plot-level FIA variables to the FIA unit and physiographic section levels. Landscape variables were recalculated within each unit and section using NCLD imagery resampled to a 400 m pixel size. I used regression trees (FIA unit scale) and general linear models (GLM, physiographic section scale) to relate spatial variations in bird abundance to the forest and landscape variables. I examined temporal relationships between the BBS and FIA data between 1966 and 2000. I developed 13 forest variables from statistical summary reports for 4 FIA inventory cycles (1965, 1975, 1989, and 2000) within NY, PA, MD, and WV. I used linear interpolation to estimate annual values of each FIA variable between successive inventory cycles and GLMs to relate annual variations in bird abundance to the forest variables.

At the BBS route scale, the CART models accounted for > 50% of the variation in bird presence-absence and abundance. The logistic regression models had sensitivity and specificity rates > 0.50. By incorporating the variables selected for the models developed within each buffer (100m, 1km, and 10km) around the BBS routes into a multiscale model, I was able to further improve the performance of many of the models and gain additional insight regarding the contribution of multiscale influences on bird-habitat relationships. The majority of the best CART models tended to be the multiscale models, and many of the multiscale logistic models had greater sensitivity and specificity than their single-scale counter parts. The relatively fine resolution and extensive coverage of the BBS, FIA, and NLCD datasets coupled with the overlapping multiscale approach of these analyses allowed me to incorporate levels of variation in both habitat and bird occurrence and abundance into my models that likely represented a more comprehensive range of ecological variability in the bird-habitat relationships relative to studies conducted at smaller scales and/or using data at coarser resolutions.

At the FIA unit and physiographic section scales, the regression trees accounted for an average of 54.1% of the variability in bird abundance among FIA units, and the GLMs accounted for an average of 66.3% of the variability among physiographic sections. However, increasing the observational and analytical scale to the FIA unit and physiographic section decreased the measurement resolution of the bird abundance and landscape variables. This limits the applicability and interpretive strength of the models developed at these scales, but they may serve as indices to those habitat components exerting the greatest influences on bird abundance at these broader scales.

The GLMs relating average annual bird abundance to annual estimates of forest variables developed using statistical report data from the 1965, '75, '89, and 2000 FIA inventories explained an average of 62.0% of the variability in annual bird abundance estimates. However, these relationships were a function of both the general habitat characteristics and the trends in bird abundance specific to the 4-state region (MD, NY, PA, and WV) used for these analyses and may not be applicable to other states or regions. The small suite of variables available from the FIA statistical reports and multicollinearity among all forest variables further limited the applicability of these models. As with those developed at the FIA unit and physiographic sections scales, these models may serve as general indices to the habitat components exerting the greatest influences on bird abundance trends through time at regional scales.

These results demonstrate that forest variables developed from the FIA, in conjunction with landscape variables, can explain variations in occupancy and abundance estimated from BBS data for forest bird species with a variety of habitat requirements across spatial and temporal scales.

Acknowledgements

I need to start by foremost thanking my wife, Stephanie, and daughter, Libby. Stephanie took on a full time job to financially support our family and allow me to pursue this opportunity. While she enjoyed her job and the people she worked with, it took time away from her second "job" as a wonderful mother to our daughter. That was very difficult for her, and I want her to know how eternally grateful I am for making that sacrifice. Also, she provided tremendous support in putting up with my long work days, many seven day work weeks, and all the other nuances that are part of graduate school. My 2-year old daughter, though she doesn't know it, was a continual source of encouragement and support just in hearing her say, "I love you, daddy". I love you both!

I have had the privilege of being co-advised by two exceptional professors, Dr. Steve Prisley and Dr. Dean Stauffer. They gave me the flexibility and freedom to explore my own ideas, but also knew when and how to step in and provide needed guidance and direction. They were tremendously supportive as I balanced family with work. My professional development, interests, and values are largely a function of their influence. They are both great mentors and good friends. Dean deserves an additional note of commendation because he also served as my masters' advisor. I extend my sincere thanks to him for putting up with me through two graduate degrees.

I extend my sincere thanks to my graduate committee members, Dr. Pat Keyser, Dr. Carola Haas, and Dr. Rich Oderwald. They were a continual source of needed direction, and I sincerely appreciate their constant support through the duration of this project. I also want to thank them for their flexibility in accommodating my family life and demands outside of graduate school. It has been a privilege working with all of them.

I would like to thank the National Council for Air and Stream Improvement, Inc. (NCASI) and the U.S. Forest Service Southern Research Station for providing funding for this project.

I extend my gratitude and appreciation to my fellow graduate students. Their moral support, willingness to act as a sounding board for ideas and problems, and of course humor made my experience here that much more enjoyable. I wish them success with their own projects and endeavors.

I would like to thank the Virginia Tech Department of Fisheries and Wildlife as whole, and all the faculty and staff that make it what it is. I have been involved with this Department on and off for the past 10 years now, and I have always appreciated its professionalism as well as its camaraderie.

I extend my love and admiration to my mother and father. They have continued to be my strongest advocates, and I thank them for their continued love, support, and encouragement.

Finally, I would like to thank my Lord and Savior, Jesus Christ. All things are possible through Him, and His love and grace have helped me through the most difficult trials. I give Him all the glory for this accomplishment, as it is only a testimony to the gifts and blessings He has provided to me.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction and Justification

Forest breeding birds have been intensively studied with a large portion of research focused specifically on their habitat associations and demographics (DeStefano 2002). At forest stand or similar small scales, the effects of microhabitat structure and composition, forest stand conditions, and silvicultural treatments on the demographics of forest breeding birds have been documented extensively (e.g., Thompson et al. 1995, Hunter et al. 2001, Jobes et al. 2004, Hanowski et al. 2006). In some cases, such studies have been experimental rather than correlative in nature, allowing for more rigorous examination of forest bird-habitat relationships and providing the opportunity to assign some causality to the observed relationships (e.g., Yahner 1997, Hanowski et al. 2006). Relationships of forest birds to landscape-level habitat components also have received tremendous attention in recent years. The effects of such features as patch size, patch shape, diversity of cover types in the landscape, edge habitat, and other landscape attributes on forest songbird demographics have been the focus of intensive research and are well documented (e.g., Freemark et al. 1995, McGarigal and McComb 1995, Faaborg et al. 1998, Debinski and Holt 2000, Penhollow and Stauffer 2000).

In spite of this extensive pool of research, many factors continue to confound our understanding of bird-habitat relationships, rendering it incomplete in a variety of ways. Most studies have been limited in both temporal and spatial scale, providing only relative ‘snapshots’ of bird-habitat relationships in space and timeframes relative to the study. Therefore, findings from such studies are not easily adapted to explaining bird-habitat relationships at relatively large spatial (e.g., physiographic regions) and/or temporal (e.g., 20-30 years) scales. Biologists are increasingly recognizing that selection of a certain scale, spatial or temporal, automatically imposes *a priori* constraints on their research, and influences the type of data required and gathered for a study, the patterns evident in that data, and the relationships detected between birds and habitat (Trani 2002, Bissonette 2003). This poses the challenging task of defining the most appropriate scale (e.g. microhabitat or landscape scale) for study (i.e. the one having the greatest influence on a given bird species) and then selecting the most appropriate variables for describing the relationships of bird abundance to habitat at that scale (Hagan and Meehan 2002). A more appropriate approach may be to consider multiple scales simultaneously, given that many bird species respond to habitat variables at a variety of spatial scales (Villard et al. 1998, Hagan and Meehan 2002, MacFaden and Capen 2002, Thogmartin et al. 2004). The necessity to address this multiscale aspect of habitat selection has been recognized as a critical component to understanding wildlife-habitat relationships relative to forest cover characteristics, as well as addressing how natural and anthropomorphic changes to these characteristics impact the distribution and abundance of wildlife species (Maurer 2002, Edwards et al. 2003). However, the very multiscale and spatially overlapping data required for such studies, especially at the microhabitat level, are difficult to obtain and quickly render larger projects logistically and financially prohibitive, potentially limiting their scope and broader applicability.

Declines in Neotropical migrants in the eastern United States had been documented as early as the late 1800s and early 1900s (Peterjohn et al. 1995), but evidence of large-scale declines of many species in recent decades across physiographic regions in the eastern United States has created significant concern among natural resource managers (Peterjohn et al. 1995, Villard and Maurer 1996). These declines are not species or guild specific, but rather have been occurring across a range of species with varying habitat (e.g. early successional or mature forest interior) and life history traits (e.g. neotropical migrants and nonmigrants)(Askins 2000, Sauer et al. 2005). Also, population trends for some species are inconsistent and vary among regions (Sauer et al. 2005), implying that the underlying spatiotemporal factors effecting bird populations also are varying among regions. Natural resource managers recognize the need to address management issues at these physiographic or ecosystem levels (DeStefano 2002). However, the limited temporal and spatial extents of existing research noted above, coupled with the logistical constraints of conducting studies at scales necessary to address these issues severely limits effective evaluations of bird-habitat relationships at these larger scales.

A potential solution to this challenge is to use existing, long-term datasets that have been developed independently of each other for various purposes. The U.S. Forest Service (USFS) Forest Inventory and Analysis (FIA) Program and USGS Breeding Bird Survey (BBS) Program provide such an opportunity. Both programs are national in their scope, with data collection spanning over 4 decades, while data for each project are collected at relatively fine resolutions. The extensive spatial and temporal nature, coupled with the fine spatial resolution, of these datasets provide a unique opportunity to investigate forest habitat relationships of breeding songbirds at multiple spatial and temporal resolutions.

The Forest Inventory and Analysis Program

The FIA is a national program within USFS that was initiated to collect, analyze, and report information on the status and trends of America's forests (Miles et al. 2001). The FIA was initially established by the McSweeney-McNary Forest Research Act of 1928, and has received further refinement from the Resources Planning Act of 1974, the Forest and Rangeland Renewable Resources Research Act of 1978, the National Forest Management Act of 1976 (USDA Forest Service 1992), and the Agricultural Research, Extension, and Education Reform Act of 1998 (USDA Forest Service 1998). The mission of the FIA is to "make and keep current a comprehensive inventory and analysis of the present and prospective conditions of and requirements for the renewable resources of the forest and rangelands of the United States" (USDA Forest Service 2004). The FIA program is delineated into 5 regional units associated with the 5 Forest Service Research Stations (Luppold and McWilliams 2000). Historically, FIA inventory cycles have occurred at approximate 10 year intervals (USDA Forest Service 1992). However, the 1998 Farm Bill mandated that this scheme be modified to a pseudo-annual cycle, with data collected on 20% of the inventory plots within each state annually (Miles et al. 2001).

The sampling methodology used by the FIA program has gone through several modifications since the program's inception, and also has varied among the FIA regions (Frayer and Furnival 1999, Miles et al. 2001). Generally, forest inventories have followed a 2-phase procedure. The first phase consisted of using aerial photography and/or remotely sensed data to characterize the acreage of forest and non-forest land in the U.S.; this phase provides information regarding land use, timber volume classes, and general forest type categories. In Phase 2, a subset of 125,000 permanent ground plots arranged on a 5 km (3 mile) grid were selected and measured to gather forest and tree information with a focus on timber characteristics, yielding more specific information regarding tree and vegetation characteristics at the plot level (Miles et al. 2001, Burkmann 2002*b*). The amount of forest acreage delineated from Phase 1 determined the intensity of sampling in Phase 2, with an average of one plot inventoried per 2428 ha (6000 ac) of forest (Burkmann 2002*a*). Ground plots typically consisted of a cluster of 5-10 fixed or variable radius subplots arranged in assorted patterns designed to cover a 0.4 ha (1 ac) sample area, with tree and vegetation data collected within each subplot (USDA Forest Service 1992, Miles et al. 2001).

In 1995 a standardized plot design was adopted by all FIA regions (Frayer and Furnival 1999), and this design was used in all inventories initiated after 1998 (USDA Forest Service 2006). This design consists of four 7.3 m (24.0 ft) radius subplots, approximately 0.0169 ha (1/24 ac) each, to measure trees at least 12.7 cm (5 in) diameter and four 2.07 m (6.8 ft) radius microplots for measuring smaller trees (Fig. 1, Miles et al. 2001). In 1998 the Forest Health Monitoring (FHM) program was integrated into the FIA. As a result the FIA adopted FHM's method of using hexagons of equal area to establish ground plots systematically across the landscape during Phase 1 of an inventory cycle. These sampling hexagons are 2402.7 ha (5937.2 ac) in size and contain one ground plot each. Use of these hexagons eliminated the sampling discrepancies among FIA regions (Brand 2002). Integration of the FHM into the FIA also added a third ground sampling phase to the inventories. The Phase 3 sampling consists of a subset of Phase 2 sample plots that are measured for a suite of forest health attributes, such as tree crown conditions, lichen community composition, understory vegetation, down woody debris, and soil attributes. There is approximately 1 Phase 3 plot for every 16 Phase 2 plots [i.e., 1 Phase 3 plot per 38,851.2 ha (96,000 ac)]; the Phase 3 measurements occur in all FIA inventories initiated after 2000 (Burkmann 2002*a*).

The data from these inventories are available in 2 forms. The first is comprehensive statistical and analytical reports that are developed at the completion of each inventory cycle by the FIA units for each state in their region (USDA Forest Service 1992). The reports present the inventory data in tabular form at the state and county scales, summarizing such variables as forest area; timber volume; and tree growth, removals, and mortality (e.g., Johnson 1992). In 1986 the eastern experimental stations adopted a set of 25 core tables that are standard for each report across states and subsequent survey cycles (USDA Forest Service 1992).

The second form is the FIA online database (hereafter referred to as the FIADB) that is accessible via the FIA website (http://www.fia.fs.fed.us/; USDA Forest Service 2006). This database contains data from the field inventories as well as additional variables computed from the field measurements (e.g., tree expansion factor, net cubic foot volume)(USDA Forest Service 2006). There are 9 tables within the database organized in a hierarchical fashion: Survey, County, Plot, Subplot, Condition, Tree, Seedling, Sitetree, and Boundary (USDA Forest Service 2006). Data from the most recently completed inventory cycle within each state are available in the database, but the availability of data from earlier inventory cycles varies among states. The U.S. Forest Service implemented efforts to achieve data uniformity across all states for inventories initiated after 1997 in accordance with the adoption of the standardized plot design (Miles et al. 2001). The completeness and consistency of data from inventories prior to 1997 is dependent on the date of the inventory cycle and the region within which the inventory occurred. Miles et al. (2001) and more recent versions of the FIA database user's manual provide comments indicating when data elements differ among FIA units, between successive inventories, or between older inventories and those conducted with the standardized plot design.

The FIA database (hereafter FIADB) contained several variables that may be directly comparable to, or could be used to easily calculate, measures of habitat structure and composition typically collected in forest bird-habitat studies; the potential of using FIA data in landscape level or multi-scale wildlife-habitat studies has been acknowledged by others (Lund et al. 1998). Some work has been done using FIA data to examine forest-wildlife habitat relationships (Chojnacky and Dick 2000, Chojnacky 2001, Welsh et al. 2006, Zielinski et al. 2006). The studies by Chojnacky found only 3 FIA variables useful in modeling habitat characteristics for the Mexican spotted owl (*Strix occodentalis lucida*): diameter at breast height, tree species, and live/dead/cut tree status (Chojnacky 2001). Welsh et al. (2006) developed 29 variables, 6 of which were selected for the final models predicting presence-absence of Del Norte salamanders (*Plethodon elongatus*) on FIA plots. Zielinski et al. (2006) measured 20 variables using USDA Forest Service Region 5 FIA protocol at known fisher (*Martes pennanti*) resting sites and compared these data with those collected at actual FIA inventory plots to develop a model describing fisher resting habitat. While these studies were successful, some focused on specific habitat components (e.g. dead and down woody debris; Chojnacky and Heath 2002), and each considered a single wildlife species. Therefore, the full potential of the FIA data for examining wildlife habitat relationships likely has not been realized. The FIA data provide a potential tool for identifying and mapping many fine-resolution characteristics (e.g. tree density, height, and species diversity) not easily derived from remotely-sensed data that may contribute knowledge to a variety of wildlife-habitat relationships and help further explain variations in songbird abundance within specific forest stands.

The Breeding Bird Survey

The Breeding Bird Survey (BBS) was initiated by the U.S. Fish and Wildlife Service (FWS) Migratory Bird Population Station (now the USGS Patuxent Wildlife Research Center) in cooperation with the Canadian Wildlife Service in 1966 as a continental program for monitoring North America's breeding bird populations (Peterjohn et al. 1995, Sauer 2003*b*). With survey routes initially established in the eastern United States and Canada, the program has continually expanded and now encompasses the entire continental United States, Alaska, and much of Canada. The BBS currently includes greater than 4000 survey routes and new routes are added each year (Sauer et al. 2003).

Each BBS route is permanently located along secondary roads. Routes are 39.4 km (24.5 miles) long with 50 stops at 0.8 km (0.5 mile) intervals (Peterjohn et al. 1995). Routes are surveyed once per year during the peak of the bird breeding season, typically late May (southern states) and June, by trained volunteers skilled in the identification of birds by sight and sound (Peterjohn et al. 1995, Sauer et al. 2005). Observers record all birds detected within 0.4 km (0.25 miles) of each stop during a 3-minute observation period (Peterjohn et al. 1995). Raw survey data at the route level are accessible online via the BBS website (http://www.pwrc.usgs.gov/bbs/) for every state for the duration of the BBS (1966-present). A row within a raw data file corresponds to observation data for a species for a given route-year combination and includes the state ID, route ID, year, species (AOU code), species observations summed by every 10 stops on the route, the number of stops at which that species was recorded on that run of the route, and the total observations for that species for that run of the route.

Several potential biases exist within the BBS data that should be considered before applying them to research and analyses. Not all BBS routes are surveyed every year, with routes located near population centers being more consistently run than those in remote areas, resulting in regional variation in the efficiency of the survey (Peterjohn et al. 1995). The spatial distribution and density of routes also varies, with the highest density occurring in the northeastern U.S. The number of routes within states also has changed, with most states adding routes over time (Peterjohn et al. 1995). Observer effects within the BBS have been documented and addressed in multiple studies (Sauer et al. 1994, Peterjohn et al. 1995, James et al. 1996, Link and Sauer 1998), and they recommended that observers be included as covariates in any analyses using BBS data (Sauer et al. 1994, Link and Sauer 1998). Some bird species with highly specialized habitat requirements (e.g., wetland birds) are poorly sampled by the BBS because routes are rarely occur in their habitat (Sauer et al. 2003). Cover type characteristics and changes along roadsides may not be representative of the surrounding area, and some roads may create edge effects and be avoided or selected by some bird species (Peterjohn et al. 1995). A variety of methods have been suggested for addressing some of these complications, and controversy exists regarding the effects of these complications as well as the best methods for accounting for them in analyses (e.g. Francis et al. 2005, Sauer et al. 2005). However, BBS data have been used extensively in a variety of studies investigating trends in bird abundance and bird-habitat relationships across a variety of spatial and temporal scales (e.g. Royle et al. 2002, Sauer et al. 2003, Thogmartin et al. 2004) and remain the most comprehensive source of bird abundance and trend data in the United States.

Study Goals and Objectives

I had two overall goals with this project: (1) integrate data collected by two independent programs, the FIA and BBS, and determine if forest characteristics derived from the FIA data can be used to explain spatial and temporal variations in bird species abundance as measured by the BBS; and (2) from these relationships develop predictive models that identify forest conditions and landscape configurations most conducive to various species and guilds of forest breeding songbirds. Because the spatial and temporal characteristics of these data differed and their compatibility was unknown, my analyses initially were exploratory in nature. I attempted to determine the spatial and temporal scales at which I felt the 2 databases were most compatible and would yield the greatest information regarding bird-habitat relationships. I specifically wanted to focus on modeling bird-habitat relationships at multiple overlapping spatial scales over broad geographic extents (e.g. physiographic province(s) or species' range), as I viewed such relationships as most critical to the current conservation needs of forest songbirds. Given these goals, I developed the following specific objectives and steps within each:

1) Determine the extent to which the FIA and BBS databases are spatially and temporally compatible.
   a) Assess the spatial and temporal characteristics of the two datasets, examining the completeness, data quality, and data variability across space and time.
   b) Identify spatial and temporal scales at which the databases potentially overlap, integrate the data into a uniform dataset, and attempt to correct for inconsistencies, noting how they may confound further analyses.
2) Determine if FIA data can explain spatial and temporal variations in forest songbird population indices as measured by the BBS.
   a) Determine the relationship between forest variables developed from FIA data to variations in abundance within 4 guilds of forest songbirds based on nesting ecology: mature forest canopy, mature forest ground/shrub, cavity, and early successional.
   b) Determine the relationship between forest variables developed from FIA data to bird abundance variations at multiple spatial scales, with emphasis on the FIA unit, physiographic section, or species range scale (≥ 1 physiographic section).
3) Develop predictive models based on these identified patterns that identify forest conditions and landscape configurations favorable to species & guilds of forest breeding songbirds.
   a) Examine various modeling techniques, seeking approaches that are most appropriate to the distributions and scale of the data available.
   b) Develop models at both single and overlapping spatial scales that explain variability in bird species occupancy and relative abundance across space and time.

4) Identify weaknesses in current FIA and BBS protocols that limit their use as tools for explaining patterns in forest songbird habitat use and population trends and propose remedies for these weaknesses.

Methods

Study Area

My study area encompassed 5 physiographic provinces of the eastern United States, the Alleghany Plateau, Cumberland Plateau, Interior Low Plateau, Ridge and Valley, and Blue Ridge Mountains (Bailey 1980), totaling 61,050,703 ha (150,859,573 ac). The outer boundary of these provinces delineated the limit of my study area and included all or portions of 14 states, AL, GA, IL, IN, KY, MD, NC, NY, OH, PA, SC, TN, VA, and WV (Fig. 2), 52 FIA survey units (Fig. 3), and 16 physiographic sections (Fig. 2b). The major forest types of this region included oak-hickory and mixed mesophytic hardwoods typically dominated by trees from the beech (*Fagus* spp.), birch (*Betula* spp.), hickory (*Carya* spp.), maple (*Acer* spp), oak (*Quercus* spp), poplar (*Liriodendron* spp.), and pine (*Pinus* spp.) genera (Braun 1950).

Data Organization and Variable Development

BBS Data.—I accumulated BBS data for the 5 physiographic provinces for the 4 species guilds: mature forest canopy (MFCan), mature forest ground/shrub (MFGS), cavity, and early successional (EarlySucc) from 1966 to 2004. The end year corresponded to the latest year a state completed the 2000 FIA inventory cycle. When selecting candidate species within each guild, I made an effort to include a mix of those with significantly increasing, decreasing, or no discernable population trends as well as those of high conservation concern (e.g., cerulean warbler, *Dendroica cerula*). I also included a mix of residents, short-distance, and long-distance migrants as well as a gradient of habitat specialists to generalists in an effort to account for the effects of such life history and behavioral characteristics on habitat selection.

I selected the initial suite of candidate species by evaluating bird data within the 2000 FIA inventory cycle to assess their suitability for calculating abundance indices at varying spatial scales during this cycle. Because the BBS has continued to add survey routes since its inception, the data available during the 2000 FIA inventory period corresponded with the most complete BBS dataset. To account for potentially spurious year effects, as well as to best associate the BBS data with the FIA data, I calculated average route-level abundance estimates over a 4-year window, with the end year of the window based on the completion year of the 2000 inventory within individual states, which ranged from 1998 to 2004. I based the 4-year window on the average number of years it took states to complete the 2000 inventory. Because some states completed their inventories in less than 4 years, I felt using a state's completion year of the survey, rather than the start year, ensured that the forest characteristics measured by the FIA were those present when the BBS data were collected. I excluded routes not having acceptable runs as defined by the BBS (Sauer et al. 2005) for all of the 4 years within their given window. I also included 44 routes that were not established using the BBS's stratified random methodology that was developed to increase the precision of long-term trend estimates (Sauer et al. 2005). Because my objectives focused on relating abundance to habitat characteristics and I was considering such a large geographic area, I felt the inclusion of these routes would have no confounding effects on my analyses. Of the 472 BBS routes within my study region, 227 were suitable for analyses.

Using the 4-year moving window introduced a potential year effect in the BBS abundance data, and I tested for this effect using ANOVA. For each possible 4-year window, I calculated route-level estimates of bird abundance within each guild across the entire study region as my response variable, and used the start year of the window as the grouping factor to delineate each window. I conducted these analyses over the entire study area and within entire guilds (versus individual species) in an effort to minimize the effects of confounding spatial factors (e.g. species distributions), assuming that the cause of any short-term temporal variations, such as weather, would be synchronized over most of my study region and have a relatively consistent influence on the entire population (Koenig 2002).

I summarized abundance estimates across the entire study region as well as within individual physiographic sections, states, and FIA units to assess the number of bird species having reliable data within these spatial units. As the spatial unit became smaller, the number of routes per spatial unit decreased with an increasing variability in the number of routes within units. Because the focus of this project was at the species' range or multiple physiographic provinces, I emphasized data quality at these scales while using the smaller scales to further refine species selection. Species had to have a minimum sample size of 14 routes (i.e. observed on ≥ 14 routes) across the study region, the minimum required for a relatively precise trend estimate (Peterjohn et al. 1995), with preference given to those species with a large sample size and relatively small coefficient of variation in abundance to be considered for the study. This allowed for the potential inclusion of species with restricted geographic distributions and/or habitat specialists not present in every region. For species with geographic distributions restricted to portions of my study area, I included BBS route data for all routes falling within the boundary of their geographic distributions plus an additional 50km buffer in an effort to account for uncertainty in estimating a species' true distribution. I used geographic distribution maps available in ArcView shape file format from NatureServe (Ridgley et al. 2005) to define each bird's distribution.

I used the regional trend analyses available on the BBS website (Sauer et al. 2005) to further screen these candidate species and selected those with data adequate for calculating reliable trend estimates at the physiographic region scale for the duration of this study (1966-2004). These analyses estimated trends for each species for a specified time period, and the results included the estimated population trend (in %/year), their statistical significance (stated as a P-value), the number of combined routes on which trends were estimated, the variance of the trend estimate, the weighted regional abundance for the species (average birds/route), and a categorical regional credibility estimate (blue, yellow, red, with blue implying the greatest credibility; Sauer et al 2005). Because the physiographic region delineations used by the BBS differed slightly from those described by Bailey (1980), I assessed species' trends for the Allegheny Plateau, Blue Ridge, Cumberland Plateau, Highland Rim, Lexington Plain, Ohio Hills, and Ridge and Valley (Sauer et al. 2005). Species had to have trend estimates based on ≥14 routes (Peterjohn et al. 1995), even if that trend was not significant, and at least a yellow credibility estimate in at least one region over the 1966-2004 study duration to be considered for further inclusion in the dataset, with preference given to those with reliable trend data in more than 1 region and a relatively small confidence interval for the trend estimate. From this pool I chose 5 species each for the MFCan, EarlySucc, and cavity guilds and 11 species for the MFGS guild (Tables 1 and 2). I included a disproportionate number in the MFGS guild because the habitat requirements of these species incorporated the greatest range of forest cover characteristics and therefore allowed for a more comprehensive assessment of the FIA to capture a range of habitat requirements. Within the remaining 3 guilds, I used Pearson's correlations to assess the relationship of abundance patterns through time among the pool of candidate species and further refine my selection, assuming that species having highly correlated abundance patterns ($r > 0.5$) likely shared similar habitat requirements. In an effort to reduce redundancy in my analyses, I selected those species having the most reliable data in the BBS and lower correlations with other species in the same guild.

FIA Data.—I included data for this study from 4 FIA inventory cycles: 1965 (± 4 years), 1975 (± 4 years), 1989 (± 4 years), and 2000 (± 4 years). I obtained raw inventory data for each state from the FIA online database for the 2000 inventory cycle and organized it within Microsoft (MS) Access databases. I gathered the FIA statistical summary reports for individual FIA survey units (when available) or states within my study area for the 1965 and 1975 inventory cycles. Because report format and content varied among states and between cycles, I identified tables for each state and cycle that matched or most closely matched 1 of the 25 core tables standard to all reports after 1986 (USDA Forest Service 1992) and keyed those data into a second suite of Access databases (hereafter referred to as the summary dataset). I used the FIA Mapmaker tool available on the FIA website (http://www.fia.fs.fed.us/tools-data/tools/) to acquire the core table data specific to my study region from the 1989 and 2000 inventories.

Using the 2000 inventory data, I developed 43 forest variables at the FIA plot level perceived as potentially relevant to the habitat requirements of those bird species included in the guilds (Table 3). I focused on variables that described forest structural attributes as many studies have demonstrated the importance of microhabitat structure to habitat suitability (Hagan and Meehan 2002, Lichstein et al. 2002, MacFaden and Capen 2002). Also, the spatial extent of my study area and resulting variation in tree species composition precluded the development of meaningful plant species-specific variables. I developed these variables based on descriptions of bird habitat requirements available on the NatureServe website and the citations contained therein (NatureServe 2005). Because the FIA raw dataset contains over 300 variables that are either collected or calculated from field data (USDA Forest Service 2006), this process was highly exploratory in nature, and I produced several redundant variables in an effort to best describe a given habitat requirement. I applied these variables to the analyses relating variations in bird abundance data to variations in forest structure across space for a fixed point in time (the 2000 inventory cycle; hereafter referred to as the spatial analyses).

I acquired statistical reports for all states completing inventories during the 1965 and 1975 inventories; VA did not complete a 1965 inventory and IL and IN did not complete a 1975 inventory. However, the data content in each report varied considerably among states and between cycles, with only 4 states (NY, PA, MD, and WV with 11 FIA units total; Fig. 3) having compatible data at the FIA unit scale that spanned the 4 inventory cycles. These data were contained in 6 tables, 3 describing forest area parameters and 3 describing volume parameters. From these data I developed 13 variables describing forest structure across the 11 FIA units for each of the inventory cycles (Table 4). As with the raw dataset, this process was exploratory and produced several redundant variables. I applied these variables to the analyses describing relationships between changes in bird abundance and forest structure from 1966-2000 (hereafter referred to as the temporal analyses).

Landcover Data.—I acquired National Land Cover Data (NLCD, ca 1992) from the USGS for each state in my study area to develop a suite of landscape variables. I chose these data because I felt they would best approximate the landscape patterns present during the 2000 FIA inventory and their landcover classification scheme allowed for easy incorporation into the analyses for this project. Because the focus of my analyses was on natural forest cover, I reclassified each image, combining those classifications in the developed, barren, non-natural woody, and herbaceous planted/cultivated categories (12 total) into a single nonforested category. I used Fragstats v3.3 (McGarigal et al. 2002) to develop 14 landscape variables for my study area for patches within the deciduous and coniferous forest cover classes and 15 variables for the landscape as a whole (Table 5). Fragstats calculates a number of edge and core area variables that consider the amount and contrast of edge in the landscape and effects of these edges on the interior or core areas of a patch. Both the edge contrast (used in edge contrast calculations) and edge depth (used in core area calculations) values are based on a predefined weight file that I developed for these analyses (Appendix A). I considered each possible edge combination for the cover types and assigned each a weight ranging from 0 to 1 and a width value in meters defining the width of a patch's edge (i.e. the buffer along a patch's perimeter not considered part of its core) (Appendix B). High contrast edges, such as those between a forest and nonforest stand were given a high edge weights and large widths. Low contrast edges, such as those between a deciduous and coniferous stand were given a low weights and narrow widths.

Integrating the FIA, BBS, and Landcover Data

I used ArcGIS v9.1 (ESRI 2005) for all GIS related database work and analyses and SYSTAT v.11 (SYSTAT Software, Inc. 2005) for all statistical analyses unless otherwise noted. For the spatial analyses, I attempted to integrate the FIA and BBS databases at their finest resolutions first (individual plots and routes), as this was the finest scale at which I would be performing analyses. This also facilitated efficient summary of the data to coarser resolutions as needed. I overlaid a shapefile of BBS routes (USGS PWRC 2004) and individual FIA inventory plot locations from the 2000 inventory into a GIS database. The FIA plot locations were the uncorrected or 'perturbed and swapped' coordinates provided with the raw data in the FIADB

(USDA Forest Service 2006), and the degree of error this introduced into the plot locations varied regionally and therefore was impossible to quantify, especially in a model-based context (McRoberts et al. 2005). I created 3 buffer layers around each BBS route with radii of 100 m, 1 km, and 10 km to model habitat characteristics at overlapping spatial scales at the route level (Thogmartin et al. 2004). Based on the known home range and breeding territory area requirements of the species included in this study (NatureServe 2005), this range of scales should incorporate most of the ecological requirements of the species included in this study. To associate the values of the plot-level variables developed using the 2000 cycle data to the BBS route buffers, I created a Thiessen polygon layer from the FIA plot locations and intersected this with each of the 3 buffer layers (Fig. 4). Thiessen polygons create a layer of irregularly shaped tiles, with the location and length of the tile boundaries assigned according to the relative placement of the points (FIA plots in this cases) upon which the polygons are created. The placement of a given tile boundary is the minimum perpendicular distance between 2 points (Burrough and McDonnell 1998). Because the uncorrected locations provided for all plots in OH and NY were the county centroids, making the county the smallest sampling unit in these states, they were excluded from the plot-level analyses. I assigned the values of the FIA variables for a given plot to its Thiessen polygon, and then calculated route-level values for each variable within the 3 buffers sizes using an area-weighted average, with the weights calculated as the proportional area a given Thiessen polygon comprised within the buffer. This created values for each FIA variable at 3 overlapping spatial scales at the BBS route level. I repeated this process to associate the county-level FIA variables to the BBS route buffers, with the counties serving as the polygon layer that was intersected with the BBS buffers (Appendix C). I extrapolated the plot-level FIA variables to the county level using the expansion factors and algorithms as described in the FIA database users guide (USDA Forest Service 2006); the FIA variables for OH and NY also were included in county-level analyses.

To incorporate the Fragstats landscape variables at the BBS route level, I overlaid the 3 buffers onto the cover types developed using the reclassified NLCD imagery, clipped the cover types based on the boundary of each buffer, and calculated values for the variables within each buffer size. As with the FIA variables, this created values for each landscape variable at 3 overlapping spatial scales at the BBS route level (Appendix C).

To examine the nature of the relationships between the BBS and FIA data among entire FIA units and physiographic sections, I calculated the average abundance/route for each bird species within each FIA unit and physiographic section, and again extrapolated the plot-level FIA variables to the FIA unit and physiographic section levels using the instructions in the FIA database users manual (USDA Forest Service 2006, Appendix D). For BBS routes that spanned multiple units or sections, I calculated a proportional abundance for each bird species for that section of the route by multiplying the route-level average abundance by the proportion of the route in each unit (or section). I also recalculated the landscape variables at these scales by clipping the reclassified NLCD imagery using the boundaries of the FIA units and physiographic sections and analyzing them with Fragstats. Because the FIA units and physiographic sections each covered relatively large areas, I was unable to develop the landscape variables using the original 30m pixel resolution of the NLCD data due to computational limitations. I resampled the NLCD imagery at a 400m pixel size, the approximate scale of USGS land use/land cover imagery (USGS 1990), using the nearest neighbor algorithm of the Resample tool available in ArcGIS (ESRI 2005) and calculated the landscape variables at this scale (Appendix D). This eliminated the 4 core area variables (Table 5) because this resolution was greater than my maximum edge depth (100 m) for defining core area.

For the temporal analyses that related FIA variables developed from the summary dataset to the BBS data, I calculated average annual abundance values (avg/route) for each bird species from 1966-2000 across the 11 FIA units for which I had data. I used linear interpolation to estimate annual values of each FIA variable between successive inventory cycles (Appendix E). Because the rate of change for variables varied between cycles, I performed these interpolations between each cycle individually rather than across the entire time range. This resulted in annual values (n = 35) of bird abundance and the FIA variables averaged across a region comprising 11 FIA units that had comparable data for each of the 4 FIA cycles.

Assessing Relationships between the BBS and FIA Data

For both the spatial and temporal analyses, I used Pearson's correlations to assess the degree of correlation among all predictor variables at all scales, as well as to initially evaluate the relation of each FIA and landscape variable to the abundance values for each bird species (see Appendices F - N for all correlation matrices used in the spatial analyses). I chose this approach rather than PCA or other multivariate exploratory approaches because I felt it would provide a more precise pairwise assessment of the FIA variables to each other, the landscape variables, and the bird abundance estimates, and allow for a better overall assessment of each variable's performance. I also used this process to identify the FIA and landscape variables most related to changes in bird abundance and thus reduced the full dataset to a subset of potential predictor variables for model development.

For the analyses at the BBS route level (i.e. the BBS route was the sample unit), I examined the correlations of the FIA and landscape variables to the route-level bird abundance values for each species across the entire study area and subsetted those with a $P \leq 0.05$. I then examined the intercorrelations of the variables within the FIA and landscape subsets. For those with high correlations ($r > 0.50$), I removed the variable(s) having the weaker relationship (smaller r value) to the bird abundance values from the subset (Robertsen et al. 2002). This created a candidate pool of both FIA and landscape variables having the strongest initial correlation with variations in bird abundance and minimal intercorrelations with other possible independent variables. I repeated this process for every buffer size around the BBS routes with the FIA variables calculated at the plot and county-level, as the weighted averages of the variables changed within each buffer. I used the same process at the FIA unit and physiographic section scales, with individual units or sections, rather than BBS routes, as my sample unit. For the temporal analysis, year was my sample unit, and I examined the correlations among the annual abundance estimates and FIA variables as outlined above.

Model Development

Spatial analyses at the plot and county level.—For each bird species, I developed models relating bird presence-absence and abundance on a BBS route to the FIA and landscape variables across all routes within the study area. It is important to note that these models examined bird-habitat relationships at the finest resolution (individual BBS routes) but over the broadest extent (all routes within the study area) possible given the data. Modeling both presence-absence and abundance provided an evaluation of which characteristics were minimally necessary for site occupancy as well as which characteristics improved the quality of a site (as indexed by an increase in abundance). I used both stepwise logistic regression (Hosmer and Lemeshow 1989) and classification trees (Breiman et al. 1984, De'ath and Fabricius 2000) for the presence-absence models, and only regression trees (Breiman et al. 1984, De'ath and Fabricius 2000) for the abundance models (classification and regression trees are referred to collectively as CART models). I selected these modeling procedures based on work by Dettmers et al. (2002) and Pearce et al. (2002) who demonstrated successful performance in modeling bird-habitat relationships relative to other common modeling procedures. For the logistic regression, I used both forward and back stepping procedures (α to enter and remove = 0.15) to ensure I had a suite of models ranging from a global model to one containing a single predictor. The models developed by the stepwise procedures served as my *a priori* set, and I used the bias corrected Akaike's Information Criterion (AIC_c; Burnham and Anderson 1998) to select the most appropriate models. I chose this approach for developing my *a priori* set of models because many bird species still had several potential predictor variables (often > 15) even after the initial screening process using the Pearson's correlations, and developing a suite of potential models from such a large pool of predictor variables was impractical. Using the stepwise procedure provided an efficient means of creating a sensible set of probable models that I felt maintained the nature of an *a priori* protocol (Dettmers et al. 2002).

CART models recursively partition the data to find increasingly homogeneous groups as the data are sequentially split to create an explanatory tree model, with the partitions referred to as branches or splits and the resulting data subgroups referred to as nodes. The recursive nature of the CART models, their ability to accommodate correlated predictor variables, and their lack of assumptions regarding data distributions made them especially attractive for these data and the exploratory nature of these analyses (De'ath and Fabricius 2000, Urban 2002). For the classification trees, I used the Gini index as the loss function with a minimum proportional reduction of error (*PRE*) at any split of 0.01 and a minimum of 5 objects allowed in any node (Breiman at al. 1984). The Gini index is a variance estimate based on all comparisons of possible pairs of values and takes the form $1 - \Sigma c^2$, where c = the proportions of each response (presence and absence in this case) in each category (Brieman et al. 1984, De'ath and Fabricius 2002). At each split, the Gini index tends to split the largest category into a separate group. Because the CART procedure tends to overfit models and highlight spurious and nonsensical relationships (Urban 2002), I evaluated each model and 'pruned' it as necessary by increasing the minimum *PRE* to remove terminal nodes containing such relationships. I considered spurious relationships to be only those that explained little variability in bird occupancy or abundance patterns and were contrary to well established habitat preferences for a given species. For the regression trees, I excluded all routes with zero observations for a given species. I chose the least squares method as the loss function, which is the equivalent to that of linear models, with a minimum *PRE* = 0.01 and minimum of 5 objects allowed in any node (Brieman et al. 1984). I pruned overfitted models using the same process as the presence-absence models.

After developing models for the 3 buffer sizes, I pooled the variables selected for the individual models at each buffer size and used them to develop multiscale models with the BBS route still serving as the sample unit (Thogmartin et al. 2004). I used logistic regression and CART, modeling presence-absence and abundance as described above with some modifications. Several variables were selected repeatedly at each buffer size and obviously were highly correlated with each other. Because I wanted to assess the effect of scale on the importance and selection of any given variable, I included all variables as potential predictors regardless of their intercorrelation. To help account for this in the logistic regression models, I set the α to enter and remove at 0.01. If highly correlated predictors ($r > 0.50$) were selected for a model, I eliminated the one having the lowest correlation to bird abundance and then repeated the stepwise procedure. Because the recursive nature of CART allows the data to identify the most appropriate variables, correlated variables can be included as potential predictors. In fact, the tree development can help identify the spatial structuring and any nonlinear additive relationships among predictor and dependent variables (Urban 2002, Thogmartin et al 2004).

Spatial analyses at the FIA unit and physiographic section level.— Because my sample sizes were reduced at these scales (n = 31 FIA units and 16 physiographic sections), I only modeled variations in bird abundance relative to the forest and landscape variables. At the FIA unit scale, I used regression trees, selecting the least squares method as the loss function, with a minimum *PRE* = 0.01 and minimum of 5 objects allowed in any node. I included all possible predictor variables identified from the initial screening using Pearson's correlations (correlation between abundance and predictor with a $P \leq 0.05$) regardless of predictor intercorrelations, and I pruned trees as necessary by increasing the minimum *PRE* to remove terminal nodes containing obviously spurious relationships.

At the physiographic section scale, I chose general linear models (GLM; Vernier et al. 2002) to model variations in bird abundance to the FIA and landscape variables. I felt that the small sample size at this scale precluded the use of the regression tree procedure. However, it is likely that no procedure appropriately accommodated such small sample sizes, and therefore these models serve to simply identify possible trends among physiographic sections. Also, I only included bird species with abundance data in all physiographic sections. This eliminated 4 species from this analysis (black-throated blue warbler, Canada warbler, chestnut-sided warbler, and rose-breasted grosbeak). I log-transformed (ln Y+1) bird abundance estimates to better approximate a normal distribution (Zar 1999). I used a stepwise procedure for model development, including all variables as potential predictors regardless of their intercorrelation, with an α to enter and remove = 0.001. If highly correlated predictors ($r > 0.50$) were selected for a model, I eliminated those having the lowest correlation to bird abundance and then repeated the stepwise procedure. I used the bias-corrected Akaike's Information Criterion (AIC_c; Burnham and Anderson 1998) to select the most appropriate models from those produced using the stepwise procedure.

See Table 7 for a summary of the differing resolutions and modeling methodologies used in the spatial analyses.

Modeling temporal relationships.—I used GLMs to relate annual variations in bird abundance to changes in the FIA variables. I log-transformed (ln Y+1) bird abundance estimates to better approximate a normal distribution. However, 2 factors confounded the model development in these analyses. First, all of the FIA variables developed for these analyses were highly correlated ($|r| \geq 0.592$). Therefore, for each species I chose the variable with the strongest correlation (greatest absolute Pearson *r* value) to annual bird abundance for inclusion in the model. Second, the number of routes sampled in the BBS increased over time, resulting in a nonconstant and increasing sampling effort, which artificially inflated many of the abundance estimates. I included the number of routes surveyed per year as a parameter in each model and used the sequential (type I) sum-of-squares option available in the GLM routine of Minitab v.13 (Minitab 2000) to fit the models in an effort to remove any effect of nonconstant sampling effort on bird abundance before including the habitat parameter in the model.

See Appendix P for a quick-reference outline of the modeling methodology that highlights the differences among the spatial and temporal scales.

Identifying weaknesses in current FIA data as a tool for explaining patterns in forest songbird habitat use and populations trends

The process of identifying weaknesses within the current FIA data as a tool for explaining patterns in forest songbird habitat use and population trends was ongoing throughout the duration of this project. Because the FIA and BBS datasets existed at different temporal and spatial scales and were not created with compatibility in mind, several potential shortcomings for developing wildlife habitat variables were present within the FIA data structure and emerged at different stages throughout the project. I maintained records that identified and documented these weaknesses, and when possible, developed recommendations for countering them and improving the FIA database as a wildlife management tool.

Table 1. Common and scientific names, guild associations, USFWS bird banding laboratory species codes, and abundance data within the study area during the 2000 FIA inventory cycle for the 26 bird species selected for this study. Selected BBS routes were those having acceptable runs (Sauer et al. 2005) during a 4-year window ending on the completion year of each state's 2000 FIA cycle.

Species	Scientific name	Species code	N routes in range[a]	Routes w/ presence[b]	% routes w/ presence	Avg. Count	Min	Max	C.V.
Cavity									
Downy woodpecker	*Picoides pubescens*	DOWO	227	224	98.7	2.9	0	12.5	0.738
Great-crested flycatcher	*Myiarchus crinitus*	GCFL	227	212	93.4	3.0	0	15.5	0.940
Pileated woodpecker	*Dryocopus pileatus*	PIWO	227	200	88.1	2.4	0	17.8	1.246
Red-bellied woodpecker	*Melanerpes carolinus*	RBWO	222	203	91.4	5.0	0	20.8	0.983
Tufted titmouse	*Baeolophus bicolor*	ETTI	227	226	99.6	13.0	0	54.5	0.666
Early Successional									
Chestnut-sided warbler	*Dendroica pensylvanica*	CSWA	169	102	60.3	3.0	0	20.8	1.505
Common yellowthroat	*Geothlypis trichas*	COYE	227	221	97.4	13.5	0	53.3	0.835
Eastern towhee	*Pipilo erythrophthalmus*	RSTO	227	226	99.6	10.9	0	41.3	0.766
Prairie warbler	*Dendroica discolor*	PRAW	223	128	57.4	1.2	0	16.5	1.816
Yellow-breasted chat	*Icteria virens*	YBCH	220	147	66.8	4.3	0	29.5	1.504
Mature forest canopy									
Cerulean warbler	*Dendroica cerulea*	CERW	220	78	35.4	0.7	0	18.3	2.670
Eastern wood-pewee	*Contopus virens*	EAWP	227	226	99.6	5.2	0	25.8	0.864
Pine warbler	*Dendroica pinus*	PIWA	202	80	39.6	1.3	0	14.8	2.295
Red-eyed vireo	*Vireo olivaceus*	REVI	227	226	99.6	30.5	0	130.0	0.829
Wood thrush	*Hylocichla mustelina*	WOTH	227	227	100	12.2	0.25	59.3	0.902
Mature forest ground-shrub									
American redstart	*Setophaga ruticilla*	AMRE	213	150	70.4	3.4	0	23.0	1.367
Black and white warbler	*Mniotilta varia*	BAWW	227	151	66.5	1.5	0	12.0	1.537
Black-billed cuckoo	*Coccyzus erythropthalmus*	BBCU	226	91	40.3	0.3	0	4.5	2.082
Black-throated blue warbler	*Dendroica caerulescens*	BTBW	151	52	34.4	1.1	0	21.3	2.690
Canada warbler	*Wilsonia canadensis*	CAWA	137	42	30.7	0.5	0	15.5	3.432
Hooded warbler	*Wilsonia citrina*	HOWO	227	155	68.3	3.1	0	33.8	1.704
Kentucky warbler	*Oporornis formosus*	KEWA	202	122	60.4	1.4	0	22.8	1.820
Ovenbird	*Seiurus aurocapillus*	OVEN	227	188	82.8	10.7	0	63.8	1.144
Rose-breasted grosbeak	*Pheucticus ludovicianus*	RBGR	174	121	69.5	1.8	0	14.8	1.283
Worm eating warbler	*Helmitheros vermivorus*	WEWA	210	122	58.1	1.3	0	15.0	1.913
Yellow-billed cuckoo	*Coccyzus americanus*	YBCU	227	193	85.0	2.9	0	13.8	1.106

[a] BBS routes falling within the geographic distribution of the species relative to the study area.

[b] Number of routes having ≥1 occurrence of that species during the 4-year window used to estimate the average abundance.

Table 2. Population trend estimates (1966-2004), regional concern data, and average counts within the North American Bird Conservation Initiative (NABCI) Appalachian Bird Conservation Region (BCR 28) for the bird species included in this study. Trend and count data were estimated using the regional trend analyses available on the BBS website (Sauer et al. 2005).

Species	Migratory status	Regional concern	Trend Estimate (%/year)	Trend P-value	N Routes	Trend 95% CI		Avg. Count
						Lower	Upper	
Cavity								
Downy woodpecker	non	N	-0.9	<0.01	331	-1.5	-0.3	2.91
Great-crested flycatcher	long	N	-1.1	<0.01	322	-1.8	-0.4	2.83
Pileated woodpecker	non	N	1.7	<0.01	307	0.9	2.5	1.89
Red-bellied woodpecker	non	N	4.1	<0.01	273	3.3	5.0	3.12
Tufted titmouse	non	N	1.1	<0.01	327	0.7	1.5	11.03
Early Successional								
Chestnut-sided warbler	long	N	1.4	0.02	166	0.2	2.5	1.90
Common yellowthroat	short-long	N	-0.4	0.11	330	-0.9	0.1	12.95
Eastern towhee	non-short	Y	-1.9	<0.01	334	-2.5	-1.3	12.92
Prairie warbler	short-long	Y	-5.1	<0.01	236	-6.6	-3.6	1.77
Yellow-breasted chat	long	Y	-3.7	<0.01	252	-4.5	-3.0	6.94
Mature forest canopy								
Cerulean warbler	long	Y	-3.0	<0.01	143	-3.8	-2.2	1.03
Eastern wood-pewee	long	Y	-3.4	<0.01	330	-4.0	-2.9	5.21
Pine warbler	non-short	N	1.0	0.23	132	-0.6	2.6	1.16
Red-eyed vireo	long	N	0.9	<0.01	331	0.5	1.3	24.61
Wood thrush	long	Y	-1.9	<0.01	333	-2.4	-1.4	15.45
Mature forest ground-shrub								
American redstart	long	N	0.3	0.68	244	-1.2	1.9	2.55
Black and white warbler	long	Y	-2.8	<0.01	250	-4.1	-1.6	1.70
Black-billed cuckoo	long	Y	-2.8	0.01	210	-4.7	-0.8	0.42
Black-thr. blue warbler	long	N	1.9	0.03	90	0.3	3.6	0.55
Canada warbler	long	N	1.7	0.27	73	-1.3	4.8	0.23
Hooded warbler	long	N	1.0	0.16	226	-0.4	2.3	2.42
Kentucky warbler	long	Y	-3.1	<0.01	209	-4.3	-2.0	1.70
Ovenbird	long	N	1.7	<0.01	311	1.0	2.4	7.36
Rose-breasted grosbeak	long	N	0.0	0.95	190	-1.2	1.2	1.54
Worm eating warbler	long	Y	-1.6	0.040	185	-3.1	0.0	0.73
Yellow-billed cuckoo	long	N	-1.1	<0.01	312	-1.9	-0.4	2.80

Table 3. Metrics developed from the 2000 FIA inventory cycle raw data used to relate variations in bird abundance data to variations in forest structure across space.

Metric	Units	Description
STDAGE	years	Stand age
STDSZCD	none	Stand size code (1 (nonstocked) - 4 (large diam.)): based on the predominant diameter class of live trees in the stand
SITECLCD	none	Site productivity class code. Classification of forest land in terms of its capacity to grow crops of industrial wood. Values ranged from 1 (0-19 cuft/ac/yr) to 7 (255+ cuft/ac/yr)
TPA	stems/ha	Trees per hectare (includes all tallied trees)
DOMTPA	stems/ha	Dominant trees per hectare (>8 cm (3.1496") DBH and dominant or codominant crown class based on FIA crown class code)
SSTPA	stems/ha	Seedling/sapling trees per hectare (<8 cm (3.1496") DBH with overtopped or intermediate crown class based on FIA crown class code)
UTPA	stems/ha	Understory trees per hectare (all trees having an intermediate or overtopped crown class regardless of DBH)
DEADTPA	stems/ha	Trees per hectare with an FIA status code of dead
RCTPA	stems/ha	Trees per hectare having >66% of merchantable volume as cull
DRCTPA	stems/ha	Sum of DEADTPA and DRCTPA
ALLDIAMAVG	cm	Average diameter of all trees
ALLHTAVG	m	Average height of all trees
DDIAMAVG	cm	Average DBH of dominant trees
DOMHTAVG	m	Average height of dominant trees
SSDIAMAVG	cm	Average DBH of seedling/sapling trees
UDIAMAVG	cm	Average DBH of understory trees
RCDIAMAVG	cm	Average DBH of rotten/cull trees
DEADDIAMAVG	cm	Average DBH of dead trees
DRCDIAMAVG	cm	Average DBH of both dead and rotten/cull trees
ALLDIAMCV	none	Coefficient of variation in DBH across all trees
ALLHTCV	none	Coefficient of variation in tree height across all trees
DDIAMCV	none	Coefficient of variation in dominant tree DBH
DHTCV	none	Coefficient of variation in dominant tree height
SSDIAMCV	none	Coefficient of variation in seedling/sapling DBH
UDIAMCV	none	Coefficient of variation in understory DBH
ALLSW_N	none	Shannon-Weiner N calculated for all trees
DSW_N	none	Shannon-Weiner N calculated for all dominant trees
SSSW_N	none	Shannon-Weiner N calculated for seedling/sapling trees
USW_N	none	Shannon-Weiner N calculated for understory trees
14 physiographic class codes (e.g. V19, V21)[a]	none	Physiographic class code: describes land form, topographic position, and soil on moisture available to trees on the plot. Value was the proportion of the plot in a given class. Code examples: Dry tops, moist slope and coves, small drains

[a]See Appendix R for descriptions of each physiographic class code.

Table 4. Metrics developed from the FIA statistical report data developed for each FIA inventory cycle used to relate variations in bird abundance data to variations in forest structure from 1966-2000.

Core table of origin	Metric	Description
1: Area by geographic unit and land class	%ForCover	% of land area in forestland[a]
3: Area of timberland by geographic unit and forest type	%SoftWood	% of timberland[b] in softwood stands
	%HardWood	% of timberland in hardwood stands
4: Area of timberland by geographic unit and stand size class	%Saw	% of timberland in sawtimber[c] stands
	%Pole	% of timberland in poletimber[d] stands
	%NonSawPole	% of timberland in nonstocked[e] or seedling/sapling[f] stands
12: Merchantable volume of growing stock trees on timberland by species and diameter class	SoftWDPoleVol/ha	m^3/ha of softwood trees in pole diameter classes (< 26.55cm (9”) DBH)
	SoftWDSawVol/ha	m^3/ha of softwood trees in sawtimber diameter classes (≥26.55 cm (9”) DBH)
	HDWDPoleVol/ha	m^3/ha of hardwood trees in pole diameter classes (<27.94 cm (11”) DBH)
	HDWDSawVol/ha	m^3/ha of hardwood trees in sawtimber diameter classes (≥27.94 cm (11”) DBH)
	TotalSoftWDVol/ha	m^3/ha of all softwood trees
	TotalHDWDVol/ha	m^3/ha of all hardwood trees
	TotalVol/ha	m^3/ha of all trees

[a] Land with ≥10% stocking of trees of any size over ≥ 0.4049 ha (1ac), or that formerly had such tree cover and is not currently developed for a nonforest use (Griffith and Widmann 2003).
[b] timberland is all forestland that can be managed for the production of wood products and capable of producing ≥1.399 m^3/ha/yr (20cuft/ac/yr) of industrial wood (Miles et al. 2000).
[c] Stocked with ≥10% of minimum full stocking with all live trees with half or more of such stocking in poletimber or sawtimber trees or both, and in which the stocking of sawtimber is at least equal to that of poletimber (Griffith and Widmann 2003).
[d] Stocked with ≥10% of minimum full stocking with all live trees with half or more of such stocking in poletimber or sawtimber trees or both, and in which the stocking of poletimber exceeds that of sawtimber (Griffith and Widmann 2003).
[e] Stocked with <10% of minimum full stocking with all live trees (Griffith and Widmann 2003).
[f] Stocked with ≥10% of minimum full stocking with all live trees with half or more of such stocking in seedlings, saplings, or both (Griffith and Widmann 2003).

Table 5. Landscape metrics[a] developed from 1992 NLCD imagery for use in the analyses examining spatial variations in bird abundance within the 2000 FIA inventory cycle. Acronyms with a C superscript were calculated only within individual cover classes (e.g. deciduous, coniferous), those with an L only at landscape level (across all cover classes). All others were calculated at both scales.

Acronym	Units	Description
PLAND[C]	%	% of the landscape composed of patches in the corresponding cover class
PD	#/100 ha	Patch density: number of patches per 100 hectares
LPI	%	Largest patch index: the percentage of the landscape composed by the largest patch
LSI	none	Landscape shape index: measure of aggregation or clustering of patches in a given cover class (e.g. deciduous patches)
SHAPE_MN	none	Mean shape index: quantifies average patch shape within a given cover class, increases with irregularity in patch shape
SHAPE_AM	none	Area weighted mean shape index: same as mean shape index but weighted by the proportional abundance of that patch in the landscape
SHAPE_CV	%	Mean shape index CV: coefficient of variation of the mean shape index
CPLAND[C]	%	Core percent of the landscape: percentage the landscape in core area of a corresponding cover class
CAI_MN	%	Mean core area index: percent of a patch that is core area, averaged across all patches
CAI_AM	%	Area weighted mean core area index: same as mean core area index but weighted by the proportional of that patch in the landscape
CAI_CV	%	Core area index CV: coefficient of variation of the core area index
CWED	m/ha	Contrast weighted edge density: length (m) of all edge in the landscape, scaled to a per hectare basis, and weighted by each edge weight value
TECI	%	Total edge contrast index: measure of edge contrast in the landscape, approaches 100% when all edge is maximum contrast (based on preset edge weightings)
IJI	%	Interspersion and juxtaposition index: measure of the interspersion or intermixing of patch types; approaches 1 as all patch types are equally adjacent to all other patch types
CONTAG[L]	%	Contagion index: index of the diversity of patch types in the landscape and their relative distribution; approaches 1 when the landscape is a single patch
SHDI[L]	none	Shannon's diversity index
SIDI[L]	none	Simpson's diversity index

[a] See McGarigal et al. (2002) for more detailed descriptions and computational formulae

Table 6. Summary of the differing spatial resolutions, spatial scales, and modelling methodologies employed in the spatial analyses using BBS and FIA data from the 2000 FIA inventory cycle.

Analysis resolution	Avg area (ha) of sample unit	Modelling methodology: Logistic regr.	Class. tree	Regr. tree	GLM
BBS route (n = 227)					
Buffer radius					
100 m	796	x	x	x	
1 km	7693	x	x	x	
10 km	92852	x	x	x	
Multiscale[a]	N/A	x	x	x	
FIA unit (n = 31)	2153798			x	
Physiographic section (n = 16)	3815669				x

[a]The multiscale analyses incorporated variables selected within each of the individual buffer sizes.

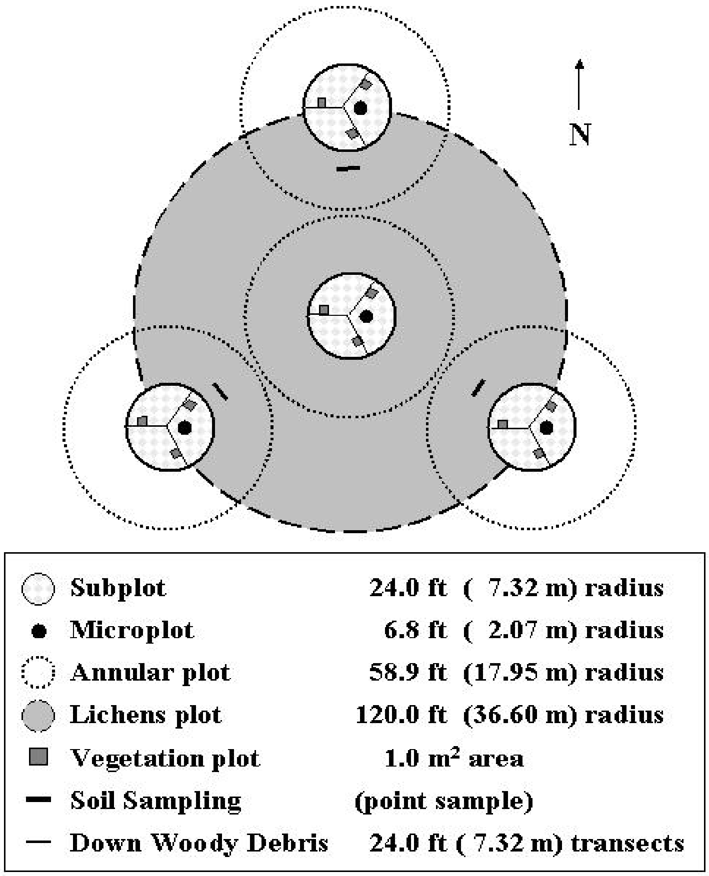

Figure 1. Standardized plot design adopted by all FIA regions and used in all inventories initiated after 1998 for collecting Phase 2 and Phase 3 forest attribute data (Burkman 2002*b*).

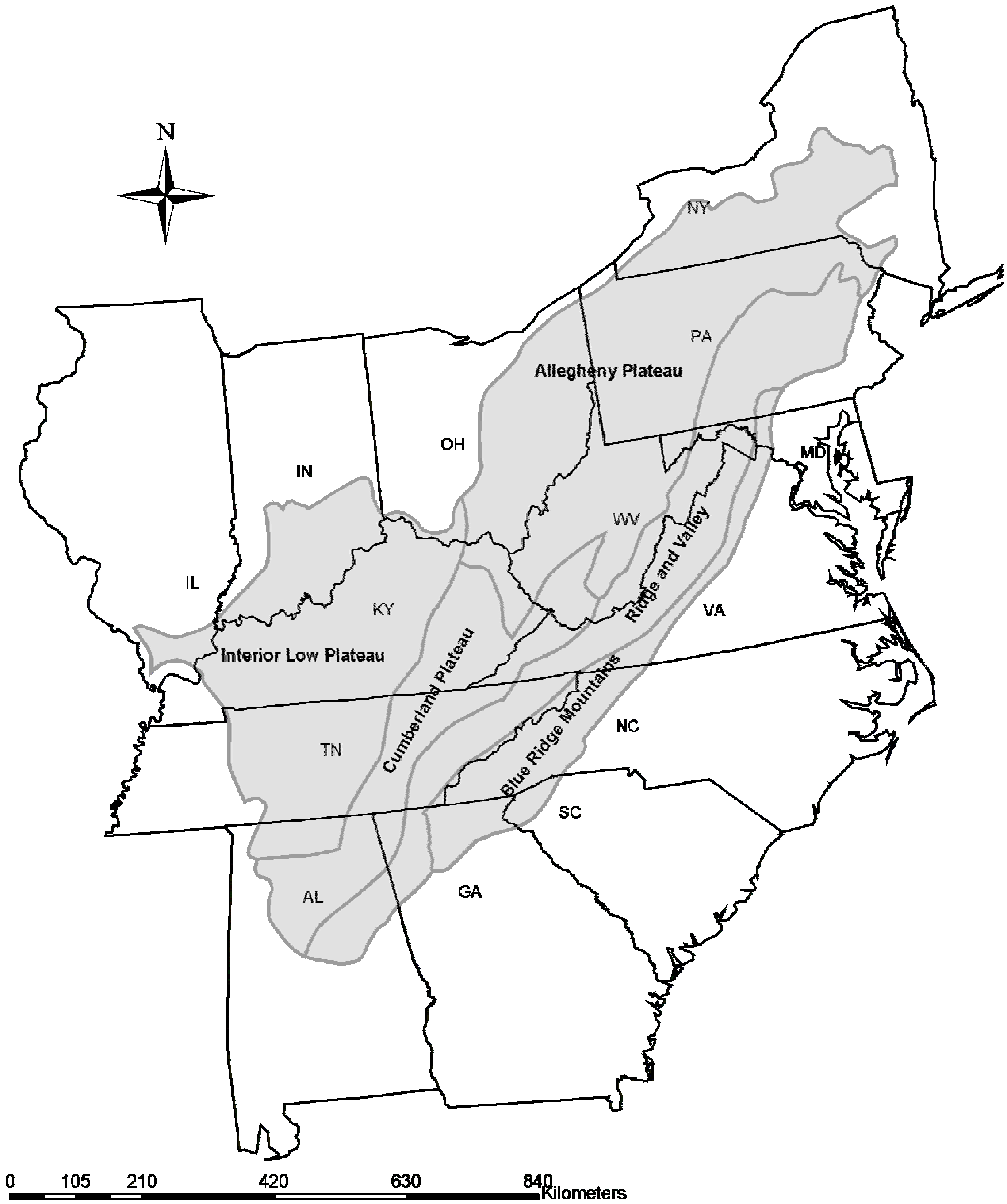

Figure 2a. The 5 physiographic provinces (Bailey 1980) that delineated the study area for this project.

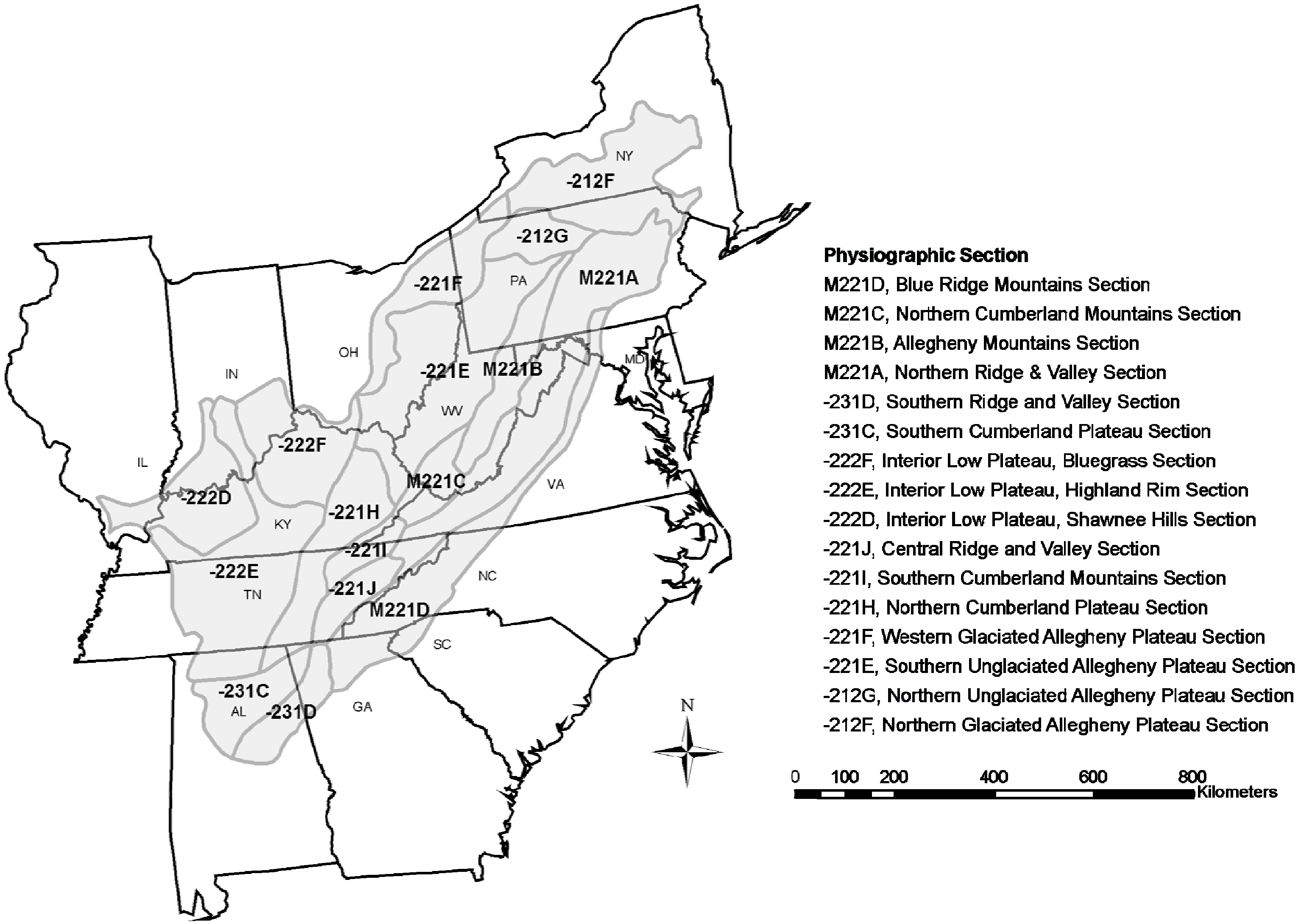

Figure 2b. The physiographic sections (Bailey 1980) included in the study region of this project.

Figure 3. The FIA survey units included within the study region of this project. Survey unit IDs are combinations of the state's number and individual unit's number as defined in the FIA database.

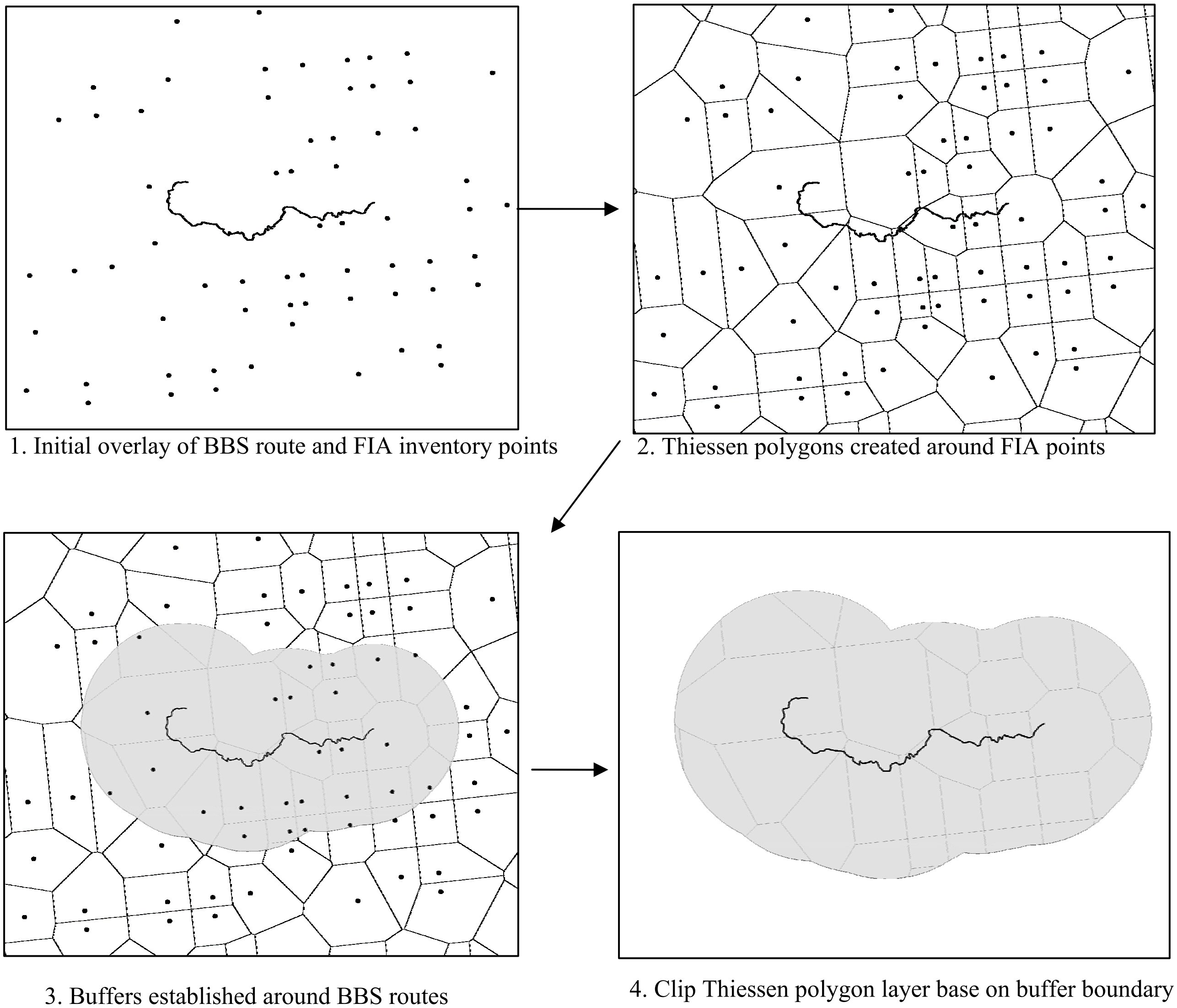

Figure 4. Illustration of steps used to intersect the Thiessen polygon layer developed from the FIA inventory points to the buffers created around each BBS route and thereby FIA metric values to the BBS routes. I followed the same process for the county-level analyses, except that the FIA metrics were summarized by county and then the county boundaries served as the polygons that were intersected with the BBS buffers.

CHAPTER 2

CHALLENGES OF INTEGRATING THE BBS AND FIA DATABASES

Introduction

This chapter describes the issues and obstacles I encountered when integrating the BBS and FIA databases and relates to the fourth objective of this project. However, I feel the information provided in this chapter is necessary for interpreting the results and understanding the overall context and limitations of the study and therefore have placed it here. I have organized the list of obstacles into categories according to the nature of the problem, the data affected, and/or how it impacted my analyses. For each entry, I described the nature of the problem, if and how I addressed the problem, if and how it impacted the analyses and results, and when possible, discuss modifications to the current BBS or FIA structure and protocol that would potentially alleviate the problem.

Challenges Specific to the BBS

- **Observer effects present in the BBS data.**

  Observer effects in the BBS data are well documented (Sauer et al. 1994, Link and Sauer 1998) and take on 2 general forms: a) differences in individual observer ability and b) general improvement in ability among all observers over time (Sauer et al. 1994). Observer effects should be incorporated as covariates into any analyses using BBS data (Sauer et al. 1994), and several methods have been proposed to statistically account for observer effects (e.g. Sauer et al. 1994, James et al. 1996, Link and Sauer 2002). However, most of these methods pertain to the assessment of long-term trends on individual routes and then aggregating those trend estimates to larger scales and are not readily adapted to analyses that focus on abundance variations in space or time relative to variations in habitat (Thogmartin et al. 2004). Those studies that have examined spatial variations in abundance relative to habitat patterns typically included data averaged over 20 years or longer (Flather and Sauer 1996, Thogmartin et al. 2004). In the context of my spatial analyses, where I used the 4-year window to develop route-level abundance estimates specific for the 2000 FIA inventory cycle, 2 observer effects were present. Each route was surveyed by a different individual (in most cases), creating a per route observer effect. Also, 17 routes had >1 observer during the 4-year window, adding temporal observer effect for those routes. While present in the data, the spatial context of these analyses and short timeframe I used to estimate abundance precludes any assessment of these observer effects in the analyses (Geissler and Sauer 1990, John R. Sauer, USGS Patuxent Wildlife Research Center, personal communication). Therefore, they have to be treated as nuisance parameters that were incorporated into model error terms.

A similar problem existed in the temporal analyses even though I consided a 34-year timespan (1966-2000). My response variable was annual abundance of a given bird species averaged across routes, not the overall trend through time or abundance averaged over a long period of time; this precluded application of the methods currently used in the route regression procedures to account for observer bias in trend estimates (Link and Sauer 1994).

Obstacles Specific to the FIA

- **Error introduced into the point-level analyses from the perturbed and swapped FIA plot locations.**

The coordinates I used to locate the FIA inventory plots in the GIS database and develop the Thiessen polygon coverage for my point-level analyses were the uncorrected, or perturbed and swapped, coordinates provided in the FIA database. The FIA introduces this error to protect proprietary information as required by law (Food Security Act of 1985) as well as maintain the ecological integrity of sample plots (McRoberts et al. 2005). While some criteria used in developing this error vary regionally, the overall methodology was supposed to be consistent across all plots (McRoberts et al. 2005, USDA Forest Service 2006). However, considerable variation existed among states in the nature of the uncorrected coordinates provided in the FIA database (FIADB) files (Fig. 5) and introduced a variable and unknown amount of error into my plot-level analyses. The perturbation and swapping does not shift plot locations among counties, so this introduced error was eliminated when summarizing the FIA variables at the county level (McRoberts et al. 2005).

I discuss the effects this introduced error had on my point-level analyses in chapter 3. To summarize, the county level variables had a greater number of significant ($P < 0.05$) Pearson's correlations between bird abundance and the forest metrics and larger average correlation coefficients (though the differences in the magnitude of the coefficients were minor) than the plot level variables (Table 7). The largest discrepancies in the correlations between the plot and county levels occurred at the 100 m buffer scale and decreased to near zero as the buffers become larger (Table 7). This was not surprising given the 10 km radius of the largest buffers was almost 10 times that of the circular radius (1.6 km) used to perturb the FIA plots (McRoberts et al. 2005), and the 10 km buffers often included data from more than 1 county. At the 100 m and 1 km buffer sizes, however, the perturbing and swapping appeared to have a confounding effect on the relationships between the FIA variables and bird abundances, resulting in a poorer fit of the point-level bird-habitat models developed with these data. Similar reductions in model fit have been documented in other model-based studies conducted at fine spatial scales using the erroneous coordinates (McRobert et al.

2005). Also, the unknown nature of this error calls into question the reliability of the bird-habitat relationships that were identified from these data. Therefore, I chose to limit model-selection procedures and detailed discussion of the bird-habitat relationships identified by these analyses to the county-level models only. I also discussed the point-level analyses in the context of the effects perturbing and swapping had on modeling bird-habitat relationships using FIA variables developed at the point-level relative to the county-level.

The obvious solution to this problem is obtaining the true coordinates of the FIA plots for the point-level analyses. Historically, the FIA has provided them on a case-by-case basis if the users can demonstrate convincing need for the exact locations as well confirmation that such information will not be abused and remain confidential. However, I was unable to obtain this information in spite of multiple requests. When the effects of the perturbing and swapping on analyses are unacceptable, the FIA provides users with assistance via their Spatial Data Services (SDS; http://www.fs.fed.us/ne/fia/spatial/index.html). However, this requires the user to submit data for their region(s) of interest to the SDS, where it is integrated with the FIA data and returned, or visiting an SDS center to conduct the analyses. For large-scale, exploratory studies such as this that included over 71,000 plots in multiple FIA regions, with large data requirements and analyses that are fluid and iterative in nature, assistance of this type is both inefficient and impractical. The FIA recognizes the negative impacts this introduced error does have on fine-scale analyses, and efforts are ongoing to better quantify this error and develop methods for circumventing it while still adhering to current confidentiality policies (McRoberts et al. 2005).

- **Among state inconsistencies in the FIA data collected in the field.**

Variability was present among states (and sometimes even within a state) in the raw data collected and available in the FIADB for the 2000 cycle. For example, the seedling table of the FIADB was populated (i.e., had data entered) for 7 of the 14 states in my study area, implying these data were not collected in the remaining states. Tree height (ACTUALHT field in the FIADB tree table) was blank for all trees in some states, blank for trees < 26.55cm DBH (5”, poletimber size) in some, and collected for all DBH classes in others (Table 8).

For those states missing these data, this created null data points for the habitat variables developed at the point and county-level from these variables (e.g. the tree height variables within the understory and seedling-sapling classes; Table 3) that I dealt with on a case-by-case basis and resulted in the complete elimination of some variables from further analyses (Table 8). Therefore, certain structural habitat components that influence habitat selection among forest breeding birds were likely underrepresented in the variables I developed, especially within the seedling-sapling size classes. When intersecting these data with the BBS route buffers and calculating the route level averages for the habitat variables, route-level averages were calculated only from those plots or counties having available data. However, as the frequency of null data points for a given forest variable increased within a buffer, the degree to which the estimated value for the buffer as a whole were representative of the true characteristics on the ground decreased.

The among-state variation in data consistency was a function of the timing of the 2000 inventory in each state and their status in adapting the current standardized survey methodology at the time they completed the inventory (USDA Forest Service 2006, John Scrivani, Virginia Dept. of Forestry, personal communication). Many of the missing variables were from those states that had not yet adopted the current methodology. As more states adopt the current national survey methodology, data consistency among states will improve, decreasing such omissions and variability within the FIADB.

- **Incorporating expansion factors into habitat variable calculations**. (Note: Because I make frequent references to variables in the FIADB that are in English units, I refer primarily to English units of measure in this bullet for clarity.)

The FIA sampling methodology expands individual tree data measured on a plot to per acre and larger area estimates using 2 weighting factors. The first is the current trees per acre (TPACURR), which is the number of trees per acre that an individual tree represents for calculating current estimates of numbers of trees over a given area (USDA Forest Service 2006). In other words, this value represents the number of trees per acre a given sample tree represents at the plot level. The second value is the current expansion (EXPCURR), which equals the number of acres the sample plot represents for making current estimates of area, and the sum of EXPCURR values over sampled plot-level records for a particular state approximates the total area of the state (USDA Forest Service 2006). Multiplying a TPACURR value by its respective plot EXPCURR estimates the number of trees a given tally tree represents within the area the sample plot represents.

While this population-level expansion (where population could be described as any area of interest, such as a county) was relatively straight forward for tree densities, all other tree related variables I developed had to be calculated relative to these expansion factors. For example, consider calculating the average DBH for the area represented by one plot containing 2 tally trees. The first tally tree had a DBH of 10” (25.4cm) and represented 6 trees/ac (TPACURR = 6; 14.8 trees/ha) and the second had a DBH of 3” (7.62cm) and represented 75 trees/ac (185 trees/ha). If the plot's expansion factor was 15,953 ac (6,456 ha), then the 10" tree represented 95,718 trees, the 3" tree represents 1,196,475 trees, and the estimated average DBH = 3.5" (8.9 cm) over the 15,953 acres. A calculation of average DBH (or height) over any area of interest had to be weighted relative to the total number of trees each tally tree represented:

$$\frac{\sum_{j=1}^{n}\sum_{i=1}^{m} DBH_i TPACURR_i EXPCURR_j}{\sum_{j=1}^{n}\sum_{i=1}^{m} TPACURR_i EXPCURR_j}$$

where i = the i^{th} tally tree on the j^{th} plot within the area of interest. The estimated standard deviation for DBH was

$$\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{m} DBH_i^2 TPACURR_i EXPCURR_j - \frac{(\sum_{j=1}^{n}\sum_{i=1}^{m} DBH_i TPACURR_i EXPCURR_j)^2}{\sum_{j=1}^{n}\sum_{i=1}^{m} TPACURR_i EXPCURR_j}}{\sum_{j=1}^{n}\sum_{i=1}^{m} TPACURR_i EXPCURR_j - 1}}.$$

The standard deviation values for DBH and height were of particular importance because I used them to estimate the coefficients of variation for these variables, which served as indices to variability in habitat structure.

Although these formulas are nothing more than algebra, they illustrate the complexity introduced by the inclusion of the TPACURR and EXPCURR variables in calculating what were typically simple descriptive statistics. While this was not a problem *per se*, as such calculations were completed by developing routines or macros within different software packages, they did add extra steps, took time to complete, and increased the risk of computational errors.

- **Accommodating differing land cover conditions occurring within a plot in the calculation of certain variables.**

It was not uncommon for a given inventory plot to encompass different cover types or conditions (e.g. forest and nonforest, very young and old stands). These differences were recorded in the FIADB condition table for a given plot and incorporated in varying degrees into subsequent estimates of stand characteristics developed from the plot. I used 3 of these fields to develop forest variables describing general stand characteristics at the plot level: stand age (STDAGE in the FIADB), stand-size class code (STDSZCD), and site productivity class code (SITECLCD; Table 3). If a plot contained multiple values for any one or combination of these variables, each was indexed with a condition class number (CONDID) and the proportion of the plot in a given CONDID was recorded in the condition proportion (CONDPROP) field in the condition table. I estimated plot-level averages of these 3 variables by calculating their weighted means as a function of the CONDID and CONDPROP fields. Another field in the condition table that recorded similar data was the land class code (LANDCLCD or COND_STATUS_CD in the FIADB), which was a categorical variable ranging in value from 1 to 6 and indexing basic land cover. Most values greater than 1 typically describe nonforest conditions, but with some exceptions. As with the other condition variables, multiple values within a plot were indexed in the CONDID and relative proportions recorded in the CONDPROP field (USDA Forest Service 2006).

If any portion of a plot had a LANDCLCD greater than 1, values for the other condition variables listed above were not recorded in that portion of the plot. This created nonsensical values for those variables when estimating plot-level averages based on their weighted means. For example, assume 50% of a plot was forested (LANDCLCD = 1) and the remaining 50% had a LANDCLCD greater than 1. If the STDSZCD was 4 (large diameter trees; Table 3) for that forested portion, the plot-level weighted average STDSZCD equaled 4 x 0.50 or 2, which corresponded to a stand dominated by small diameter trees and obviously was not an accurate representation of the actual conditions within the plot. To minimize these inaccurate plot-level estimates of the STDAGE, STDSZCD, and SITECLCD variables, I calculated their values on plots with a LANDCLCD equal to 1 across the entire plot and gave them null values otherwise. I felt this approach would produce the most realistic representation of the stand characteristics described by these variables, especially as they were averaged to the county and larger scales.

- **Lack of variables in the FIA relevant to wildlife habitat requirements.**

The historical focus of the FIA has been forestry and timber products; the data present in the FIADB reflect this focus. While many of these variables also provided relevant descriptions of wildlife habitat requirements, variables describing non-timber related components, such as herbaceous cover, dead and down woody debris, and shrub and seedling (<2.54 cm DBH) cover, were largely absent. The recently adapted national survey methodology and incorporation of the Forest Health Monitoring (FHM) program into the FIA increases the diversity of data collected by the FIA and includes variables describing all of these characteristics (Burkmann 2002*a*, USDA Forest Service 2006). Seedling data on stems less than 2.54 cm are collected on the standard FIA plots, known as Phase 2 or P2 plots in the current methodology (USDA Forest Service 2006). The remaining data were part of the FHM and are collected on a small subset of the FIA plots, known as Phase 3 or P3 plots, with 1 plot sampled for every 38,851.2 ha (96,000 ac; Burkmann 2002*b*). These variables typically describe cover attributes important to wildlife at very small scales, but this sampling intensity will likely be too coarse and limit the applicability of these data in wildlife-habitat studies.

Increasing the sampling intensity of the P3 plots is the apparent solution to this problem, but several other limitations need to be addressed. Increasing the sampling intensity of the P3 plots would increase the time to complete the inventory as a whole and may become logistically impractical. Also, data collection on the P3 plots is limited to June through August due to the seasonality of some of the variables, such as herbaceous vegetation (Burkmann 2002*a*), and the number of plots that can be sampled is a function of this limited window. The ephemeral nature of herbaceous vegetation makes it a difficult variable to quantify. For studies such as mine, where bird breeding and nesting habitat was the focus, herbaceous vegetation characteristics measured in August may have little relevance to those present during the breeding season. It may be feasible to include measurements of dead and down woody debris on P2 plots that provide data appropriate for wildlife-habitat studies. These measurements could be estimates of percent cover or other basic indices that are quicker to measure than those collected for the P3 plots and would add minimal time to the inventory. Also, Chojnacky and Heath (2002) developed a method for modeling dead and down woody biomass from existing FIA data that could be incorporated into the FIADB as an additional data field in the plot table.

- **Poor comparability in the FIA data among the 4 inventory cycles for developing forest variables to relate to annual variations in BBS data.**

The lack of consistent data available through the 4 inventory cycles across my study area precluded effective analyses of temporal changes in forest breeding bird relative to changes in forest structure. The data collected and reported by the FIA has changed considerably over the past 40 years and varied greatly among regions (Luppold and McWilliams 2000, Miles et al. 2001, USDA Forest Service 2006). Data from the 1965 and 1975 cycles were limited to the statistical summary reports developed for each cycle. The content, scope, and availability of these reports varied among states and inventory cycles, making it difficult to develop a suite of variables that were consistent through the 35-year duration of my temporal analyses. Many reports contained data averaged across an entire state and therefore contained data not relevant to my study region. Also, the approximate 10-year gaps between the 1965, ’75, ’89, and 2000 inventory cycles necessitated interpolating the values of the variables I was able to develop for the years between these gaps, further reducing their effectiveness in modeling temporal changes in abundance. The existence of FIADB, availability of the raw inventory data, and adoption of the current annual inventory methodology addresses these problems.

Obstacles in Integrating the Databases

- **Spatially relating the FIA variables to the bird abundance data at the BBS route level.**

Choosing a method to intersect the FIA data with the bird abundance data across space was one of the most difficult aspects of this study. Other studies using FIA data to describe wildlife-habitat relationships collected their relevant wildlife data on the FIA plots (Dunk et al. 2002, Welsh et al. 2006) or used FIA plot design and sampling protocol to collected their data (Zielinski et al. 2006), thus avoiding this problem. The FIA collected its data at a finer scale than the BBS, and the degree of overlap between the datasets varied spatially with FIA plot location and BBS route density. Not having the exact FIA plot coordinates added another level of complexity to this problem. When choosing a method for combining the FIA and BBS data, I had to consider several factors including 1) the spatial scale(s) at which birds were potentially responding to variations in cover (Hagan and Meehan 2000), 2) how my choice of scale(s) would affect the nature of and relationships among my dependent and independent variables (Dungan et al. 2002, Bissonette 2003), and 3) the appropriateness of the method given the characteristics of the data. Also, this was a multi-step process that involved interpolating the FIA data over space and then intersecting the 2 datasets for the analyses.

I considered several methods for interpolating the FIA point-level data across space, including kriging, inverse distance weighting (IDW), and Thiessen polygons. Kriging and IDW both create continuous surfaces, which would have been advantageous given that many of variables I developed were continuous variables. However, studies have found that the inherent local variability of forest data created a major source of uncertainty in the predicted values in such surfaces (Hershey 1999). Other studies have shown that some forest variables have no spatial autocorrelation structure, a key assumption of both IDW and kriging. Rather, the structure of the data was better represented by a set of discrete polygons, like those developed from Thiessen polygons (Lowell 1999). Given these considerations, I felt Thiessen polygons would provide the most precise interpolations of the point-level data across space. Also, the spatial characteristics of the Thiessen polygons closely mimicked that of counties, facilitating more meaningful comparisons of analyses between the FIA point and county scales.

By examining the relationships between bird abundance and the weighted averages of each forest variable within 3 overlapping buffer scales at each BBS route, I attempted to simultaneously address the question of selecting the scale(s) at which birds were responding to variations in cover and the effect of scale on the nature and outcome of my analyses.

Similar methods have been used in other studies and have been effective in illustrating multiscale influences on bird habitat selection (Thogmartin et al. 2004, Thogmartin and Knutson 2006).

The disadvantage in using any technique that interpolates a variable over a larger area is the loss of fine scale information about the feature that variable describes. I attempted to choose techniques that minimized this loss of information but remained computationally straightforward. At the FIA plot level (ignoring the perturbed-swapped effect of the plot coordinates), the variables I developed reflected the conditions and cover variability on the ground at the plot. By using weighted averages based on the proportion of Thiessen polygon area within a buffer around a BBS route to estimate values for the FIA within that buffer, the values of each variable were most influenced by those points closest to the route. However, this methodology also reduced important point-level differences that likely influenced bird abundance. For example, one plot may have occurred within a 15 year-old stand and a neighboring plot in a 125 year-old stand. Such spatial heterogeneity in cover was important information, but was not accurately represented as point data were averaged across larger areas. Obviously, such losses of information were compounded as the scale of my analyses became coarser (plot → county → FIA unit → physiographic section), and my results must be interpreted within the context and limitations of their spatial scales and influences they had on the nature of my data and analyses (Dungan et al. 2002).

- **Variability in the timing and duration of 2000 FIA inventory cycle among states.**

The timing and duration of the 2000 FIA inventory cycle varied among the states included in this study, with the cycle's start year ranging 1996-2002 and the duration of the cycle ranging from 1 to 7 years. This high variability was in part caused by the transition of states to the standardized inventory protocol (used for all inventories initiated in 1998) that changed the cycle intervals from approximately 10 years to a pseudo-annual cycle in which states are required to collect data on 20% of the plots annually (USDA Forest Service 2006). This variability in timing and duration created difficulties in relating the BBS data to the FIA data across the entire study region. I wanted the timeframe of the BBS data to match that of the FIA, which varied at the state level, as closely as possible, but I also wanted to maintain consistency in the structure of the BBS data across the entire region to reduce extraneous noise that may confound my analyses. Couched within this issue was the need to account for potential atypical years (e.g. effects of a harsh winter or other widespread phenomenon) when estimating bird abundance from the BBS. Using only 1 or 2 years of BBS data to estimate route-level abundances increased the probability of a spurious year effect. My solution was to use the average time for states to complete the 2000 inventory (4 years), make this interval a moving window with its end year corresponding to the completion year of the inventory in a given state, and only include routes if they had acceptable data for all 4 years within this window. This provided the best temporal match of the BBS to the 2000 cycle while also accounting for potential spurious years in the BBS counts. Because the timeframe of the BBS data varied between states, there was still the potential for an overall year effect, but I was easily able to test for this using an ANOVA. As states complete their transition into the new FIA sampling protocol, all states will be sampling plots annually, thus removing the need for a moving window or similar method.

One drawback of this solution, and of considering a limited timeframe of data in general, was the presence of a detection rate or sampling effort effect in the estimated bird abundances. Average abundance per route for any given species tended to increase with an increase in the time interval (number of years) used to estimate that abundance, indicating that I may have underestimated abundances for some species. Holding other factors constant, estimated abundances for a given species should become asymptotic given enough years of data, but the presence of long-term trends, changes in habitat, and observer effects confounds such estimates. Currently there are no reliable methods of estimating detection rates from BBS data (John R. Sauer, USGS Patuxent Wildlife Research Center, personal communication).

- **Spatial distribution of BBS routes at the FIA unit and physiographic section scales**

BBS Routes are established using a stratified random methodology in an effort to sample a range of conditions and obtain a more accurate estimate of population changes (Sauer et al. 2003). The start point and direction of each route is randomly located within a degree block of latitude and longitude (Sauer et al. 2003), with route placement limited to secondary roads to minimize the impacts of traffic on the counts (Peterjohn et al. 1995). Nonetheless, the spatial distribution and density of routes varies, with the highest density occurring in the northeastern U.S. (Peterjohn et al. 1995), and spatial autocorrelation has been documented in analyses of BBS data (Flather and Sauer 1996, Thogmartin et al. 2004). Also, this methodology limits placement of routes to areas containing secondary roadways, omitting areas with relatively few or limited access road ways. At the point and county scales of my analyses, these spatial variations in coverage were of little concern as I was considering each route individually. At the FIA unit and physiographic section scales, the FIA variables were unit or section averages from points dispersed throughout each region. The spatial arrangement of BBS routes and number of routes per unit and section varied considerably (unit range between 1 and 20 routes; section range between 3 and 55 routes) and was not proportional to area. This problem was compounded for species with geographically restricted distributions and therefore limited number of routes on which they may be detected (e.g. Canada warbler, *Wilsonia canadensis*; Table 2). Therefore, the abundance estimates at these scales reflect inconsistent proportions of a given section or unit, and variations in bird abundance among units or sections may not be reflective of what was really occurring at these scales.

Sauer et al. (2003) addressed similar difficulties when using the BBS to estimate population changes for the North American Bird Conservation Initiative (NABCI) Bird Conservation Regions (BCR) and at smaller scales within those regions. Though the number and distribution of routes varied among BCRs, they found that trend estimates for most BCRs differed little from previous studies (Sauer et al. 2003). However, BCRs encompass several states and even multiple physiographic provinces. For example, the Appalachian BCR (BCR 28) encompasses all physiographic provinces of this study except for the Interior Low Plateau (Fig. 2a; U.S. NABCI Committee 2000). At smaller scales within BCRs (e.g. states and portions therein) that more closely match the FIA unit or physiographic sections in size, they encountered similar problems with variable sample sizes (number of routes) and decreased precision in trend estimates (Sauer et al. 2003). Establishment of additional routes within deficient regions is a potential solution, but the methodology used to establishing them would likely be nonrandom and inconsistent with current methods. Even with additional routes, many bird species would likely remain poorly sampled as they occur in habitats not typically sampled in roadside surveys (Sauer et al. 2003). In general, these shortcomings are a function of the BBS and difficult to improve given the nature of its design (Sauer et al. 2003).

- **Overlap of BBS routes across FIA units and physiographic sections and partitioning the BBS abundance data accordingly.**

For the spatial analyses at the FIA unit and physiographic section scales, several BBS routes overlapped one or more boundaries between units or sections, creating a problem in allocating bird abundance data between them. Sauer et al. (2003) had similar problems in allocating BBS routes to BCRs for developing trend estimates. Their solution was to assign a route to the BCR that contained the greater portion of its length, even though such routes may not represent bird population patterns specific to a single BCR (Sauer et al. 2003). However, this was partially dictated by the route regression methodology used to estimate the population trends, which requires data from the entire route (Geissler and Sauer 1990). Because I was not constrained by such methodological requirements, I divided the average route abundance for each species proportionally, based on the proportion of the route in each unit or section. I felt allocating data proportionally across sample units was more appropriate for my analyses, and it addressed the concern expressed by Sauer et al. (2003) regarding abundance patterns being a function of all units or sections containing the route.

Raw count data summarized at every 10 stops per route were available from the BBS website (Sauer et al. 2005), and use of such data would provide a more accurate allocation of abundance data to a given area. However, the scope of this project made the incorporation of finer scale data impractical. Use of those data also would have been inconsistent with the analyses at the point and county level where I used route-level averages of bird abundance.

Table 7: The average Pearson's *r* value and number of correlations having a *P*<0.05 between the FIA variables (*n* = 43) and route-level bird abundance measures for each species across the 3 buffer sizes created around the BBS routes. Variables were calculated at the FIA plot and county scales with data from the 2000 FIA inventory cycle.

	County level						FIA plot level					
	100 m		1 km		10 km		100 m		1 km		10 km	
Bird species	$\bar{r}$	# of *P* < 0.05	$\bar{r}$	# of *P* < 0.05	$\bar{r}$	# of *P* < 0.05	$\bar{r}$	# of *P* < 0.05	$\bar{r}$	# of *P* < 0.05	$\bar{r}$	# of *P* < 0.05
Cavity												
DOWO	0.163	4	0.161	3	0.169	3	0.175	3	0.163	5	0.172	9
GCFL	0.177	11	0.176	11	0.187	11	0.177	8	0.194	5	0.183	10
PIWO	0.239	26	0.239	24	0.248	26	0.197	20	0.197	20	0.227	23
RBWO	0.283	25	0.281	25	0.287	27	0.214	18	0.226	20	0.261	28
ETTI	0.220	19	0.218	19	0.239	19	0.210	21	0.217	19	0.246	27
Average	0.216	17.0	0.215	16.4	0.226	17.2	0.194	14.0	0.199	13.8	0.218	19.4
Early successional												
CSWA	0.235	18	0.234	18	0.234	21	0.219	6	0.217	9	0.232	12
COYE	0.298	23	0.293	21	0.303	24	0.255	22	0.258	21	0.280	29
RSTO	0.222	15	0.216	15	0.230	18	0.191	7	0.195	11	0.217	13
PRAW	0.203	12	0.195	14	0.203	15	0.196	18	0.198	20	0.222	25
YBCH	0.269	24	0.271	23	0.266	28	0.256	24	0.259	26	0.291	31
Average	0.246	18.4	0.242	18.2	0.247	21.2	0.224	15.4	0.226	17.4	0.248	22.0
Mature forest canopy												
CERW	0.235	15	0.232	15	0.227	19	0.209	7	0.206	7	0.219	17
EAWP	0.193	15	0.196	14	0.195	17	0.206	5	0.199	6	0.200	13
PIWA	0.259	18	0.263	18	0.262	21	0.239	16	0.234	17	0.252	21
REVI	0.317	25	0.323	24	0.340	23	0.200	18	0.217	20	0.247	27
WOTH	0.196	18	0.200	17	0.210	17	0.188	3	0.194	4	0.204	16
Average	0.240	18.2	0.243	17.6	0.247	19.4	0.209	9.8	0.210	10.8	0.224	18.8

Table 7 continued

	County level						FIA plot level					
	100 m		1 km		10 km		100 m		1 km		10 km	
Bird species	$\bar{r}$	# of $P <$ 0.05	$\bar{r}$	# of $P <$ 0.05	$\bar{r}$	# of $P <$ 0.05	$\bar{r}$	# of $P <$ 0.05	$\bar{r}$	# of $P <$ 0.05	$\bar{r}$	# of $P <$ 0.05
Mature forest ground-shrub												
BBCU	0.210	12	0.211	12	0.219	14	0.201	11	0.193	14	0.197	16
BTBW	0.287	16	0.271	15	0.301	17	0.264	9	0.257	12	0.286	15
BAWW	0.315	21	0.314	20	0.319	22	0.229	19	0.228	22	0.254	21
CAWA	0.209	7	0.221	5	0.215	8	0.262	8	0.264	9	0.245	6
HOWO	0.260	22	0.256	20	0.261	22	0.233	20	0.224	17	0.267	19
KEWA	0.210	20	0.200	21	0.211	24	0.199	17	0.205	19	0.241	18
OVEN	0.288	25	0.283	26	0.288	27	0.274	18	0.269	22	0.276	27
RBGR	0.263	28	0.258	28	0.268	29	0.286	19	0.275	20	0.311	26
AMRE	0.255	21	0.260	21	0.257	24	0.244	18	0.226	22	0.264	27
WEWA	0.258	17	0.262	16	0.255	19	0.225	20	0.224	21	0.249	22
YBCU	0.236	22	0.237	21	0.241	24	0.217	19	0.231	17	0.248	20
Average	0.241	19.2	0.240	18.6	0.247	20.9	0.220	16.2	0.221	17.7	0.241	19.7
Overall	0.239	18.2	0.238	17.7	0.244	19.7	0.221	14.5	0.220	15.8	0.240	19.3

Table 8: Descriptions of data inconsistencies in the FIADB for the 2000 FIA inventory cycle among states included in this study and the actions taken to account for them in the analyses.

FIADB table name	FIADB variable name and acronym	Problem encountered	Action taken
Seedling	All variables	Data available for 7 of 14 states	Data too incomplete; omitted from analyses
Tree	Actual height (ACTUALHT)	Not available in AL, IN, IL, NC, TN	Height variables for these states given null values. Values for route buffers originating from states with value into states with null values were estimated from known data points only, reducing precision of estimate relative to actual conditions. When model building, a row containing ≥1 missing data point for any variable was completely omitted from the analyses, thus inclusion of these data effectively decreased my sample size if bird abundance was correlated with a height variable.
		Collected for trees <26.55cm DBH (poletimber size) in GA, KY, VA only	Data too incomplete; omitted seedling-sapling and understory height variables from all analyses
	Crown class code (CCLCD)	Not collected for trees <26.55cm DBH in VA and WV	Assumed the majority of these stems would fall into the seedling-sapling stems; included them in all understory and seedling-sapling variables
	Status code (STATUSCD) and Trees/ac (TPACURR)[a]	For dead trees (status code = 2), some states only estimated trees/ac on a subset of plots; value is zero otherwise	These trees were used for the dead and dead-rotten-cull variables. For plots having dead tree entries but zero trees/ac value for those entries, changed these to null vales and based estimates only on those plots with nonzero trees/ac values.

[a] TPACURR represents the number of trees per acre that the tree represents for calculating current estimates of numbers of trees for a given area.

Figure 5. Variation in the nature of the perturbed and swapped FIA plot locations from the 2000 inventory cycle used in the spatial analyses. Note the clustering of points near the county centroids in Kentucky and the linear pattern of points in Tennessee.