Available online at www.sciencedirect.com

SCIENCE DIRECT

Remote Sensing of Environment 97 (2005) 249 – 262

Remote Sensing of Environment

www.elsevier.com/locate/rse

Using the spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns

Edward J. Laurent[a,*], Haijin Shi[a], Demetrios Gatziolis[b], Joseph P. LeBouton[c], Michael B. Walters[c], Jianguo Liu[a]

[a]*Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Michigan State University, 13 Natural Resources Building, East Lansing, MI 48824-1222, USA*
[b]*Pacific Northwest Research Station, United States Forest Service, 620 Main Street Suite 400, Portland, OR 97205, USA*
[c]*Department of Forestry, Michigan State University, 126 Natural Resources Building, East Lansing, MI 48824-1222 USA*

Received 27 July 2004; received in revised form 26 April 2005; accepted 29 April 2005

Abstract

We investigated the potential of using unclassified spectral data for predicting the distribution of three bird species over a ~400,000 ha region of Michigan's Upper Peninsula using Landsat ETM+ imagery and 433 locations sampled for birds through point count surveys. These species, Black-throated Green Warbler, Nashville Warbler, and Ovenbird, were known to be associated with forest understory features during breeding. We examined the influences of varying two spatially explicit classification parameters on prediction accuracy: 1) the window size used to average spectral values in signature creation and 2) the threshold distance required for bird detections to be counted as present. Two accuracy measurements, proportion correctly classified (PCC) and Kappa, of maps predicting species' occurrences were calculated with ground data not used during classification. Maps were validated for all three species with Kappa values >0.3 and PCC >0.6. However, PCC provided little information other than a summary of sample plot frequencies used to classify species' presence and absence. Comparisons with rule-based maps created using the approach of Gap Analysis showed that spectral information predicted the occurrence of these species that use forest subcanopy components better than could be done using known land cover associations (Kappa values 0.1 to 0.3 higher than Gap Analysis maps). Accuracy statistics for each species were affected in different ways by the detection distance of point count surveys used to stratify plots into presence and absence classes. Moderate-to-large detection distances (100 m and 180 m) best classified maps of Black-throated Green Warbler and Nashville Warbler occurrences, while moderate detection distances (50 m and 100 m), which ignored remote observations, provided the best source of information for classification of Ovenbird occurrence. Window sizes used in signature creation also influenced accuracy statistics but to a lesser extent. Highest Kappa values of majority maps were typically obtained using moderate window sizes of 9 to 13 pixels (0.8 to 1.2 ha), which are representative of the study species territory sizes. The accuracy of wildlife occurrence maps classified from spectral data will therefore differ given the species of interest, the spatial precision of occurrence records used as ground references and the number of pixels included in spectral signatures. For these reasons, a quantitative examination is warranted to determine how subjective decisions made during image classifications affect prediction accuracies.

Keywords: Landsat; Forest; Birds; Habitat; Prediction; GAP; Wildlife occurrence; Michigan; NDVI; SWIR; Multiple season images; Accuracy assessment; Image classification

1. Introduction

One of the greatest impediments to accurate mapping of the Earth's resources is a paucity of spatially referenced information (Franklin, 2001). Remote sensing technologies could improve this situation for objects detectable at the spectral frequencies and grain, or smallest spatial sampling unit, of the sensor. In the case of wildlife habitat studies, the Landsat Thematic Mapper series of sensors are often used due to their relatively low price, short repeat-time (16 days), spatial resolution (approximately 30 m × 30 m), and spectral

* Corresponding author. Southeast Gap Analysis Project, North Carolina State University, Raleigh, NC 27695-7617, USA. Tel.: +1 517 353 5468.
E-mail address: Ed_Laurent@ncsu.edu (E.J. Laurent).

0034-4257/$ - see front matter
doi:10.1016/j.rse.2005.04.015

resolution capable of detecting differences in vegetation (Lillesand & Kiefer, 1999).

Landsat imagery is often used to map wildlife distribution patterns indirectly. This is accomplished by first classifying images into land cover categories and then reclassifying land cover categories for wildlife occurrence by using known vegetation affinities of each species (e.g., Morrison et al., 1992; Scott et al., 1993). In order to characterize land cover, however, a classification scheme must be used to instruct image processing software how to aggregate pixels of remotely sensed imagery into discrete categories.

Land cover classification schemes divide continuous ecological gradients (Austin, 1985; Whittaker, 1956) into compositionally distinct features to achieve an objective (Foody, 1999; Townsend, 2000; Zube, 1987). While this allows large regions to be divided in ways that humans can understand, wildlife often perceive and respond to landscape heterogeneity in substantially different ways (Johnson et al., 1992, Tang & Gustafson, 1997). For example, rather than choosing habitat based on a particular forest type or age class, some bird species select for understory vegetation or forest structure (Anders et al., 1998; Probst et al., 1992; Stouffer & Bierregaard, 1995). Thus, most land cover classifications have limited utility for predicting patterns of these species' occurrences. Similarly, when classification schemes do not include key habitat features under selection by target species, investigations into the dependence of wildlife species' occurrences on land cover classes are not likely to identify causal mechanisms underlying the observed distributions (Wiens et al., 2002). Thus, the disparity between human and wildlife perceptions of and responses to landscape heterogeneity can add substantial error into statistical analyses and management prescriptions.

In attempts to minimize predictive errors potentially fostered by inappropriate or inaccurate land cover maps, a logical approach to extrapolating patterns of wildlife distributions across large areas is to directly classify imagery for an individual species' occurrence using raw spectral reflectance data. Using this approach, spectral characteristics of locations where species are known to occur are employed when extrapolating their distributions. The underlying assumption is that species' occurrences can be predicted by spectrally detectable components of their habitat. A benefit to this approach is that distribution maps are classified using all occurrence locations separately yet simultaneously, therefore no global model relating spectral variables to occurrence sites need be assumed.

This direct approach to mapping wildlife distributions is growing in popularity (e.g., Conner, 2002; Hepinstall & Sader, 1997; Jenkins et al., 2003a; Laurent et al., 2002). While such studies indicate the potential for bypassing a land cover map, they have not yet provided a strong methodological test of raw satellite imagery's capability to map species' distributions. Some of the choices made by the authors during classification were arbitrary and no framework was established for investigating the influence of these choices, or the range of possible choices on prediction accuracy.

Several decisions must be made during image analyses in order to extrapolate the results of presence/absence surveys across a landscape using spectral information. Some options include the type of imagery, the classification scheme, the method of classification, the choice of pixels used to represent classes, and the parameter values (both spatial and spectral) used to classify maps. While these choices could be described as the "art" of image processing, it is also possible to place them within a hypothesis-testing framework to quantify their effects on prediction accuracy.

To investigate the influence different choices in classification options have on predicting wildlife occurrence maps, we have created a software program (PHASE1) as the first phase in the development of a **H**abitat **A**nalysis **B**y **I**terative **CLASS**ification procedure (HABICLASS). The purpose of HABICLASS is to minimize anthropocentric classification bias when predicting wildlife distribution patterns (first phase) and to determine causes of the predicted patterns (second phases). PHASE1 was created to make the prediction phase methodologically rigorous through the use of statistical models. This is accomplished by creating maps of species' occurrence patterns, iteratively modifying classification methods used to create the maps, and evaluating divergences in prediction accuracies of maps created in different ways. The goal therefore is to identify better means of improving prediction accuracy within a strong inference framework. Here we use PHASE1 of HABICLASS to predict forest bird occurrence with the spatial precision of Landsat imagery. Although these analyses are conducted using spectral data, PHASE1 may be used in combination with any raster dataset.

PHASE1 is superficially similar to other image classification software. It is used to create a spectral signature for each field plot describing grid values at that location. These signatures are then used to identify spectrally similar pixels and classify them into categories to create a thematic map. PHASE1 differs from many other image classification software through its application of cross-validation to stratify and repeatedly sample random subsets of signatures. As a result, a group of maps, rather than a single map, are created from a set of reference plots. The accuracy of the group of maps can then be summarized to describe a statistical distribution of accuracy for comparisons with maps made through other methods. However, reference data for this approach need to be spatially compatible to deny the introduction of scale or aggregation effects (Dutilleul, 1998; Dutilleul & Legendre, 1993) that could influence class descriptions. Reference data used in signature creation should therefore describe the same size area, or grain, for proper use within the PHASE1 classification framework.

To ensure grain equivalency within datasets used in PHASE1, we developed a **G**rain **R**epresentative **A**ssessment and **IN**ventory protocol (GRAIN). GRAIN permits a per-pixel extrapolation of spatially referenced survey data across large areas with the spatial and spectral precision of satellite imagery. For bird point count surveys, GRAIN requires the surveyor's location be georeferenced and the distance between the surveyor and each bird recorded. The georeferenced position is employed to identify the pixel used for signature development in PHASE1. Species presence and absence at those locations define classes used in supervised image classification. However, proximity thresholds (hereafter referred to as detection distances) can be set to exclude species detected beyond specified distances from observers, effectively converting a presence location into an absence. In addition to pixels overlapping georeferenced locations, signatures may also contain values for pixels surrounding the georeferenced position if they meet specified spatial and spectral criteria. Such decisions made during image classification can be iteratively modified, allowing the analyst to repeatedly tune the classification parameters and create a series of maps for investigating the influence of these choices on map accuracies.

The goal of this study was to identify methods of image classification that could predict the regional occurrence of Neotropical migrant warblers with the spectral and spatial precision of Landsat 7 ETM+ imagery and independent from any other system for categorizing the landscape (e.g., land cover maps). Furthermore, we were interested in whether this method could obtain results comparable to Gap Analysis (Donovan et al., 2004; Scott et al., 1993). Gap Analysis is used to classify maps representing species' fundamental niches (Hepinstall et al., 2002, Gap maps), or areas of potential but not necessarily occupied habitat, using known species-habitat associations, land cover data and expert knowledge within a geographic information system. Land cover maps provide the foundation for most predictions within the Gap framework. Thus, comparisons of spectrally derived species' occurrence maps with Gap maps provide a useful and contemporary reference for how well Landsat imagery can be used to directly predict species' occurrences.

Our specific questions were 1) do individual pixels of Landsat 7 ETM+ imagery contain some of the information needed to predict the occurrence of forest bird species, especially those that select for understory conditions during the breeding season? If the answer to question (1) is yes, then 2) can we identify the influence of detection distances for categorizing species occurrence and alternative classification options on prediction accuracy, and 3) how does the accuracy of spectrally derived maps compare to those created from land cover maps using the methods of Gap Analysis?

We used the GRAIN protocol to collect ground reference data of bird occurrence over a large forested region. PHASE1 was employed to automate signature creation describing surveyed plots and classify maps predicting species occurrence. We also examined the influence of varying two spatially explicit classification parameters on prediction accuracy: 1) the window size used to average spectral values in signature creation, and 2) the threshold distance required for bird detections to be counted as present. The accuracy of maps predicting species' occurrences was validated with ground data not used during classification and compared with the accuracy of recent Gap maps. Accuracy was assessed using two common measures, and the information content of these measures was compared.

2. Methods

2.1. Study region

The study region is located in Michigan's Upper Peninsula, USA and includes parts of five counties (Baraga, Dickinson, Iron, Menominee, and Marquette; Fig. 1A). This

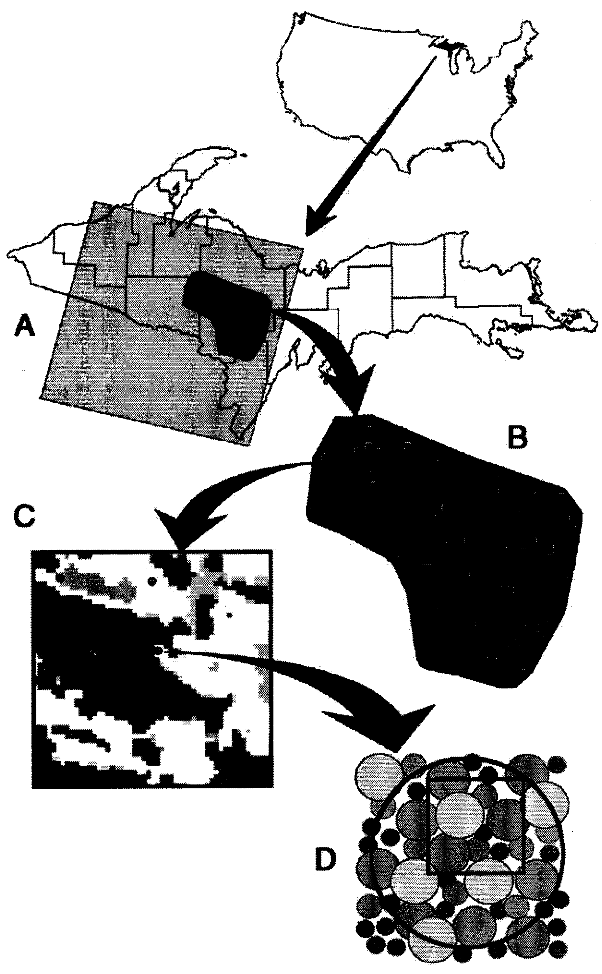

Fig. 1. The study region lies within a single Landsat 7 ETM+ scene and contains parts of five counties in Michigan's Upper Peninsula (A). Within the study region (B), randomly selected landscape units (C; township sections or USGS QQQs with coarse land cover classes within landscape units shown as shades of gray) define areas for the selection of specific plots (D; overview showing different tree species) for bird occurrence surveys. Plots were sampled with a minimum 30-m radius from plot centers (D; large circle), so that a single Landsat 7 ETM+ pixel (D; square) falling within this area could be precisely characterized.

region covers ~400,000 ha or approximately 11% of a single Landsat 7 ETM+ scene (Path 24 Row 28). A majority of the land is owned by state government and industry with management primarily focused on wood production. The study region is an ecologically diverse landscape characterized by a spatial mosaic of forest stands that include upland hardwoods (sugar maple (*Acer saccharum*), quaking aspen (*Populus tremuloides*), yellow birch (*Betula alleghaniensis*), basswood (*Tilia americana*)), lowland hardwoods (black ash (*Fraxinus nigra*), red maple (*Acer rubrum*)), lowland conifers (northern white cedar (*Thuja occidentalis*), black and white spruce (*Picea mariana* and *Picea glauca*), tamarack (*Larix laricina*)) and upland conifers (European larch (*Larix deciduosa*), eastern hemlock (*Tsuga canadensis*), red and jack pine (*Pinus resinosa* and *Pinus banksiana*)). Besides the influence of glacial topography and other edaphic factors, the composition and structure of canopy and understory vegetation has high spatial variability due to differences in forest management practices and gradients of deer browsing pressure (Albert, 1995; Van Deelen et al., 1996).

2.2. Study species

We chose three bird species for analysis (Latin name and American Ornithologists' Union four-letter code): Black-throated Green Warbler (*Dendroica virens*; BTNW), Nashville Warbler (*Vermivora ruficapilla*; NAWA), and Ovenbird (*Seiurus aurocapillus*; OVEN). All three of our study species commonly breed in deciduous and mixed forests of northern North America and select territories on the basis of subcanopy composition and structure in addition to dominant overstory components (Collins, 1981; Collins, et al., 1982; Dunn & Garrett 1997; Morse, 1976). These birds perceive and respond to landscapes at a small spatial grain for many functional reasons including those related to niche partitioning and territory delineation (MacArthur & Pianka 1966; Robinson & Holmes, 1982; Schoener, 1968). Predicting the occurrences of these warblers therefore provides a strong test for assessing the information embedded in Landsat 7 ETM+ imagery.

Although the study species are included in the same family and have all shown a general affinity for breeding in forested areas, they have markedly different habitat associations. BTNW nest in a wide variety of mature coniferous, deciduous and mixed forests (Brewer et al., 1991; Dunn & Garrett, 1997; Morse, 1976). Most foraging and nesting take place in the midlevels of vegetation, therefore requiring a multi-storied layering of vegetation, often with a shrub or coniferous understory component (Collins, 1983; Norton, 1999). NAWA nest within dense ground cover in a variety of wet and dry open woodlands (Brewer et al., 1991; Dunn & Garrett, 1997) and often near ecotones (Williams, 1996). In drier areas, they commonly select early successional forests that arise following fires or deforestation. OVEN nest on the ground and often in large stands of mature deciduous and mixed forests (Brewer et al., 1991; Smith & Shugart, 1987; Zach & Falls, 1979). Dry upland areas are most commonly used but they sometimes are found in lowland forests and swamps. In all OVEN breeding habitat, however, leaf litter is essential for foraging and nest construction (Van Horne & Donovan, 1994). Males of all three species are territorial and vociferous during the breeding period and have distinctive songs that make them easy to distinguish from other species in the study region. In addition, they have characteristic plumages that can be identified by trained observers.

2.3. The GRAIN protocol for field sampling

The GRAIN protocol was used to select and characterize locations for BTNW, NAWA and OVEN occurrence surveys. GRAIN uses the random sampling method described by Lillesand et al. (1998) for the Upper Midwest Gap project but modified to incorporate the field data collection approach of Liu et al. (2001). In 2001, random United States Geological Survey (USGS) quarter quarter quads (QQQs; $n=28$) served as landscape sampling units (LUs). In 2002 and 2003, random township sections ($n=36$ and 32, respectively) served as LUs (Fig. 1B). This switch between roughly equal area QQQs and township sections (1000 ha vs. 1037 ha, respectively) was made in order to allow our data to be more easily integrated with public and private databases such as Michigan's atlas of breeding birds (Brewer et al., 1991).

Within each LU, between 2 and 8 plots (Fig. 1C) were selected for bird surveys for a total of 433 plots surveyed over the course of the study. These plots were a minimum of 90-m apart with the mean minimum distance of 240-m (194-m standard deviation) from the next closest plot. Survey plots encompassed a 30-m radius area so that a single pixel of Landsat 7 ETM+ imagery sensed over plot centers could be precisely characterized by data collected within this area (Fig. 1D). The specific plot selection criterion was that a hypothetical 30-m × 30-m square could be placed anywhere within the plot and perceived by the field crews as having the same vegetation structure and composition as a similar square placed anywhere else within the plot.

Plot selection in general was dependent on two criteria: 1) permission to access the property and 2) maximizing the biotic and abiotic differences among plots to sample the spatial and spectral heterogeneity of the study region. Preference was given to at least one northern hardwood plot in each LU to satisfy collaborative research priorities if northern hardwood plots could be located in the field. Locations of plot centers were georeferenced using a minimum of 80 global positioning system (GPS) point locations collected at 5-s intervals with a Trimble Geo-Explorer 3 (Trimble Navigation Ltd.) and later differentially corrected to a precision of ±5 m using Coast Guard base station data. Site centers were flagged, as were 30-m distances in the 4 cardinal directions (sensu Huff et al.,

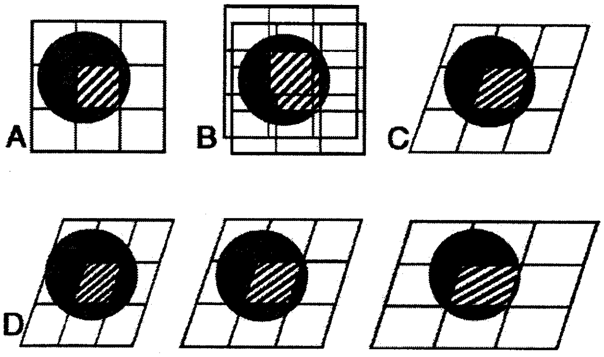

Fig. 2. Factors affecting the ground instantaneous field of view captured by pixels in Landsat imagery include: A) unknown pixel location, B) non-overlapping pixels of multiple images, C) effects of Earth rotation, and D) inclusion of increasing ground coverage within pixels increasingly farther from nadir. Gray circles represent 30-m radius field plots. Each grid represents the ground information contained within a nine-pixel window. The hashed center cell represents the ground area used to describe the sampling plot.

2000). Accuracy of plot center locations was determined to be within the precision of Garmin III+ GPS receivers (15 m; Garmin Corporation) used to navigate to these locations during repeated point count surveys.

There were several reasons why we used a 30-m radius circle for selecting survey locations. First, many of our ground surveys were conducted prior to image acquisition so the exact pixel location was unknown (Fig. 2A). Also, we used multiple season imagery for analysis (see Section 2.4) and pixels from different images do not always overlap (Fig. 2B). Third, the ground information attributed to a pixel is actually a parallelogram that is not square due to the rotation of the Earth as the Landsat 7 satellite travels in a circumpolar orbit (Fig. 2C). Fourth, the Landsat satellites have sensors that sweep back-and-forth past a central point called nadir. The ground instantaneous field of view attributed to pixels farther from nadir is therefore larger than that of pixels closer to nadir (Fig. 2D). Although geometric correction during level 1G processing of Landsat 7 imagery makes adjustments for the last two factors, the resulting image values are still dependent on the geometric limitations of the sensor and the influence of reflective features adjacent to the ground area contained in the pixel (Cracknell, 1998). Finally, circular plots are easier to inventory for birds than parallelograms.

In addition to detections within the 30-m sampling radius, species were also recorded if detected within larger radii circles surrounding plot centers. These radii included 50 m and unlimited distance during all three years. During 2002 and 2003, an additional detection distance of 100 m was used. These larger spatial thresholds are commonly employed during point count surveys for birds (e.g., Ralph et al., 1993; Ralph et al., 1995). Bird detectability was assumed to be consistent among plots and all detections were assumed to be made within approximately 180 m from plot centers (see Wolf et al., 1995).

LUs were divided into 7 (2001), 6 (2002), and 12 (2003) groups for daily bird surveys to minimize travel time (not necessarily minimizing distances) among LUs surveyed during any day. The surveyed order of LU groups, plots among LUs, and survey timing was randomized, as was the selection of observer for each survey. Beginning at sunrise, observers conducted surveys within a 5-h period when weather conditions did not preclude birds from singing (Ralph et al., 1993; Ralph et al. 1995). Observers had been trained to detect the three species in this study and other species in the region by song and sight using tapes, field guides and practice prior to data collection.

Bird species were counted as present if detected by song, call or sight within each detection distance over an 11-min period. Timing of surveys began immediately when observers reached the center of the plot. After 10 min an additional 1 min was spent walking around the plot center to flush elusive individuals within 30 m. The distance of flushed individuals detected <180 m from plot centers by surveyors walking in or out of the plots was also recorded. Each plot was visited three times during the breeding period for these species between June 4 and July 3 in either 2001, 2002, or 2003. At least two observers surveyed each site to account for possible differences in observer's physical abilities to detect the study species (Ramsey & Scott, 1981). A total of 8 observers collected data over the 3 years of this study.

In all, 433 plots ($n=112$, 86 and 235 for years 2001, 2002 and 2003, respectively) were surveyed for bird species' occurrences 3 times during the breeding period (Table 1). Survey plots were placed in a wide variety of land cover classes (see Space Imaging Solutions, 2001 for land cover descriptions) including aspen associations ($n=85$), herbaceous openland ($n=10$), lowland coniferous forest ($n=35$), lowland deciduous forest ($n=14$), lowland shrub ($n=2$), mixed upland conifers ($n=4$), mixed upland deciduous ($n=6$), northern hardwood associations ($n=190$), pines ($n=23$), other upland conifers ($n=15$), upland mixed forest ($n=36$), and upland shrub / low-density trees ($n=10$). Within these land cover classes, surveys were conducted over a wide variety of size classes (non-forest, sapling, pole, mature), management histories (clear cuts, selective cuts, no cuts) and ranges of vertical

Table 1
The number of plots where species were observed within each detection distance

Species	Detection Distance			
	30 m $n=433$	50 m $n=433$	100 m $n=321$	180 m $n=433$
Black-throated Green Warbler	44	175	196	271
Nashville Warbler	88	208	208	271
Ovenbird	150	309	297	409

All 433 plots were surveyed using 30-m, 50-m, and 180-m detection distances and 321 plots were surveyed using the 100-m detection distance.