Lidar remote sensing of bird canopy habitat use in the Northeastern United States

Scott J. Goetz[1], Daniel Steinberg[2], Matthew Betts[3], Richard T. Holmes[4], Patrick Doran[5],
Ralph Dubayah[6] & Michelle Hofton[6]

[1]The Woods Hole Research Center, sgoetz@whrc.org
[2]Yale School of Forestry & Environmental Science
[3]Oregon State University, matthew.betts@oregonstate.edu
[4]Dartmouth College, richard.t.holmes@dartmouth.edu
[5]The Nature Conservancy, pdoran@tnc.org
[6]University of Maryland, mhofton@umd.edu

Abstract

A current challenge in biodiversity research is understanding the effect of vegetation structure on the potential of an ecosystem to support species richness and habitat use. We take advantage of the utility of satellite remote sensing, specifically lidar, for improving characterization of habitat structure and apply those advances to an exploration of bird habitat use in New England USA. In this study, we find that lidar metrics of canopy vertical structure and complexity provide unique and significant information for models of habitat use of a neotropical migrant bird species, the black-throated blue warbler, in the Hubbard Brook Experimental Forest, NH. Lidar metrics describing the vertical distribution of canopy elements and the complexity of canopy elements are thus both useful and important for biodiversity research, although we find that other aspects of habitat are equally important, including the type and seasonality of vegetation. Together these variables provide complementary information that advance biodiversity research and emphasize the relevance of remote sensing observations.

Keywords: bird diversity, canopy structure, habitat heterogeneity, habitat use, lidar

1. Introduction

Ecologists have long sought to explain patterns of biodiversity based on latitude, area, evolutionary rates, and other factors. Prior to the availability of satellite data, field-based studies at local spatial scales revealed the strong role of vegetation structure in driving biodiversity. Vegetation structure refers generally to the horizontal and vertical distribution of vegetation. MacArthur and MacArthur (1961) refined the broad concept of vegetation structure by defining foliage height diversity (FHD) as a measure of canopy layering, and suggesting its use as an indicator of biodiversity. Variations on the FHD concept have led to the development of several indices of forest structural complexity incorporating vertical and horizontal variation in tree size, canopy cover, shrub size, shrub cover, coarse woody debris and snags (McElhinny et al. 2005). Vertical and horizontal structural complexity drives biodiversity by creating a greater variety of microclimates and microhabitats, which in turn produce more diverse food and cover for a more diverse range of species (MacArthur and MacArthur 1961, Hunter 1999, Hill et al. 2004). Across landscapes, the distribution of seral stages, patch sizes, and connectivity of patches also influences habitat suitability (Turner et al. 2001). Thus biodiversity managers focus on maintaining variation in tree size, multiple canopy layers, presence of coarse woody debris, and other elements of forest structural complexity within forest stands and creation of a variety of landscape scale seral stages (Hunter 1999, Rapp 2004).

While satellite data have greatly enhanced understanding of the effects of ecosystem energy on biodiversity, no continental-scale quantification of vegetation structure has been available. One goal of our work is to test the utility of current airborne and space-borne data on vegetation structure for studies of biodiversity, and to develop guidelines for the next generation of satellite sensors for quantifying vegetation structure for understanding and managing biodiversity. Towards this end, the field-based understanding of vegetation structure and biodiversity has been greatly advanced by application of airborne lidar. Heterogeneity can be calculated directly from lidar-derived forest structure, using metrics such as vertical distribution ratio (Drake et al. 2002, Goetz et al. 2007, Vierling et al. 2008) and integrated measures of the complexity of the waveform that takes into account the roughness, slope, number of gaussian peaks, and amplitude of peaks in the waveform data (Dubayah et al. 2000, Hofton et al. 2004). Using these and related metrics, including canopy height, Goetz et al. (2007) were able to predict species richness of different bird guilds in the forests of the Patuxent National Wildlife Refuge (Maryland). This was true even in a relatively homogenous forest environment with little variability in traditional optical vegetation indices (e.g. NDVI). Use of lidar data provided an ability to detect variability in vegetation structure and density, which were critical variables describing the habitat use of bird species. Here we explore the habitat suitability (preferences) of a single bird species with specific habitat preferences, the Black-Throated Blue Warbler, *Dendroica caerulescens*, a well studied neotropical migrant breeding in northern hardwoods forests.

2. Methodology

In the Northeast United States, black-throated blue warblers tend to occupy mature deciduous forests with a well-developed and high-density understory (Holmes 1994, Doran and Holmes 2005). Our working hypothesis was that deciduous cover and understory structure and density are both vegetative characteristics that can be identified using optical and lidar remote sensing.

2.1 Study area

We analyzed a long-term data set of bird observations collected at the Hubbard Brook Experimental Forest (HBEF), located in the southern region of the White Mountain National Forest in central New Hampshire. HBEF was established in 1955 as a long-term research site used for the study of forest and watershed dynamics. The HBEF encompasses approximately 3037 hectares of hilly terrain, ranging in elevation from 222m at the lowest point of the brook to 1015m atop Mount Kineo on the southwest rim of the watershed. The region is dominated by northern hardwoods (Sugar Maple, Beech, Yellow Birch, and White Ash) at low to middle elevations, and is dominated by spruce and fir species at higher elevations and along the ridgelines (Schwartz et al. 2001). The forests within and surrounding the HBEF were logged selectively for spruce in the late nineteenth century and were logged intensively for both conifer and hardwood species in the early part of the twentieth century before being established as an experimental preserve in the 1960s.

2.1 Bird data sets

Data on the distribution and abundance of the Black-Throated Blue Warbler were collected in the summers of 1999, 2001, 2002, and 2006. A survey grid consisting of 371 points along 15 north-south transects was established throughout the HBEF (Figure 1) (Schwartz et al, 2001). Points spaced at 100 meter intervals along each transect were visited 3 times during the peak breeding seasons (late May through June) of 1999, 2002, and 2006 and twice in 2001. During each visit the abundance of Black-throated Blue Warblers was surveyed for 10 minutes using fixed radius (50-m radius) point counts (Ralph et al. 1995, Doran and Holmes 2005). Surveys were performed between 0530-1000 by multiple trained observers in order to limit error in observer accuracy.

Using the abundance data collected over the 4 years, annual presence/absence was determined at each of the 371 survey sites. The black-throated blue warbler was considered present at a site if an individual was observed within 50 meters of the survey location, i.e. abundance was greater than 1. Similarly to Doran and Holmes 2005, presence/absence data were used to classify each survey site based on the number of years that the species was present (0, 1, 2, 3, or 4) over the duration of the study. It has been previously shown that high quality sites are consistently occupied year after year, regardless of interannual variability in abundance of a species, whereas low quality sites are only occupied during periods of high population density (Doran and Holmes 2005, Newton 1998, Sergio and Newton 2003). We used this index of multi-year presence as a surrogate for habitat quality or suitability. The greater the number of years occupied by the species, the better the quality of the habitat. The habitat quality index was constructed using all 4 years of data, as well as a separate index using only 3 years of observations (1999, 2001, & 2002), reserving 2006 data for testing.

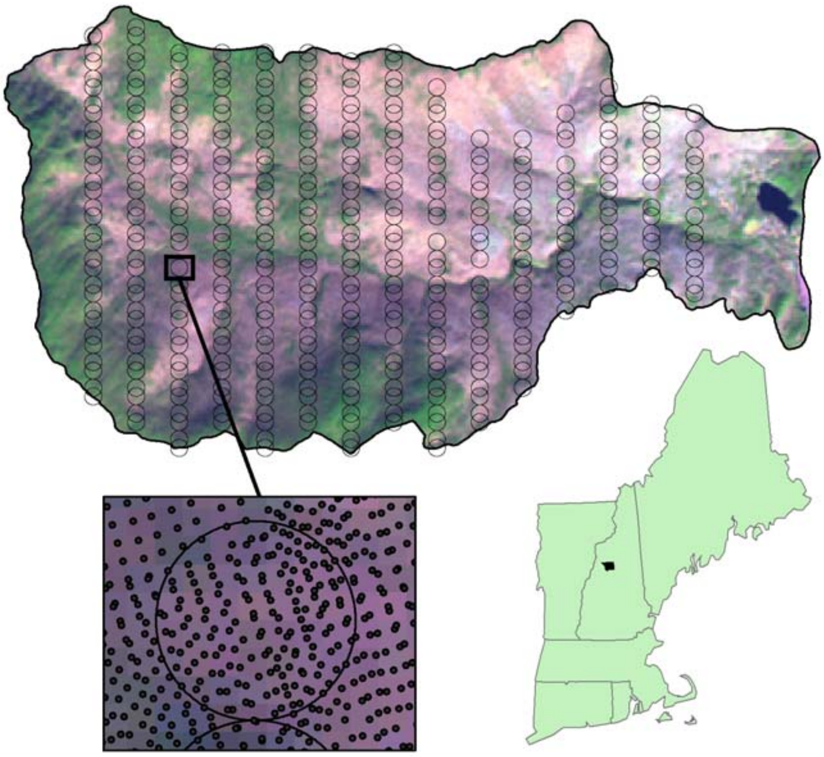

Figure 1. Hubbard Brook study area in central New Hampshire, showing a Landsat NDVI image and gridded bird observation areas. The lower left image shows the density of lidar shots within a given location.

2.2 Lidar data sets

Full waveform lidar data were acquired over the Hubbard Brook Experimental Forest with the Laser Vegetation Imaging Sensor (LVIS) in July of 2003. LVIS is a fully imaging, medium altitude, scanning laser altimeter. It has a 7º field of view within which footprint sizes can be varied from 1 to 80m depending on, among other factors, the altitude at which the instrument is flown (Blair et al. 1999). LVIS digitizes the return signal and converts the waveforms to units of distance by accounting for the time elapsed between the initial laser pulse and the return.

Three products were derived directly from the LVIS waveform data including, ground elevation, canopy height (CH), and the height of median return (HOME). Each of these products is calculated in reference to the ground return; therefore accurate determination of the ground surface is an essential step in producing these metrics. The ground return is identified using an automated algorithm which initially applies a smoothing function to reduce noise in the waveform, and subsequently locates the first increase above a mean noise level, designated as the initial canopy return, and the center of the last Gaussian pulse, designated as the ground return. CH was calculated as the difference in height between the initial canopy return and the ground return. HOME was derived as the difference in height between the median of the entire waveform, including ground and canopy energies, and the ground return.

In addition to canopy height, ground elevation, and median height, we derived two higher-level products which provide information on the vertical distribution of vegetation biomass as well as the structural complexity of the canopy. The Vertical Distribution Ratio, or VDR, is an index of the vertical distribution of intercepted canopy elements (biomass) and ranges between 0 and 1 (Goetz et al. 2007). The VDR is a ratio of the distance between the canopy return and the height of median return to the total canopy height; VDR = [CH-HOME]/CH. In general, forested regions characterized by a dense canopy and sparse understory will exhibit lower VDR values due to the relatively short distance between CH and HOME. Areas characterized by a more even distribution of biomass throughout the vertical profile will exhibit larger VDRs (closer to 1).

Canopy complexity (COMP), by comparison to VDR, is an integrated measurement of the complexity of the waveform and takes into account the roughness, slope, number of gaussian peaks, and amplitude of peaks in a waveform and, like VDR, ranges from 0 to 1. Although we refer to it as canopy complexity, it is not a biophysical measurement as such, but rather a measure of the vertical complexity of the waveform. In forested regions, however, the vertical complexity of a waveform (COMP) is determined by the complexity of vegetation structure.

2.3 Landsat data sets

In addition to the lidar metrics, we examined metrics derived from optical imagery in relation to the bird richness data. Two Landsat ETM scenes (path/row 013/029), acquired in late October of 2000 and August of 1999, were converted to top-of-atmosphere reflectances using in-band spectral irradiances and a solar geometry model to correct for Earth-Sun distances and solar zenith angle variations (Goetz 1997). The images were subsequently georeferenced. The Normalized Difference Vegetation Index was calculated for both the leaf-on (August) and the leaf-off (October) scenes and the two scenes were differenced resulting in an image of seasonal NDVI change. This allowed us to evaluate and consider seasonality in vegetation cover and density.

Vegetation type for the study region was also examined. Using a vegetation type map of the HBEF which delineated regions of deciduous, coniferous, mixed predominantly deciduous, and mixed predominantly coniferous, we produced a continuous grid of percent deciduousness. This was analyzed in addition to the optical and lidar products.

2.4 Spatial and statistical analyses

Using a geographical information system (GIS), bird survey polygons were intersected with the lidar (LVIS) and optical (Landsat ETM) data products (Figure 1). The minimum, maximum, mean, and standard deviation of ground elevation, canopy height, median height, VDR, COMP, leaf-on NDVI, and NDVI difference were computed for all lidar shots or Landsat cells falling within the boundaries of each 50m radius survey cell. These summaries were subsequently examined in relation to the habitat quality index derived from the BTBW occupancy data using an advanced regression tree technique known as "Random Forest."

A "Random Forest" (RF) model builds upon the standard methods of constructing classification and regression trees as a technique for partitioning data based on a series of hierarchical binary splits of the predictor variables, resulting in a tree structure that terminates in nodes associated with discrete ranges in the response variable (Breiman 2001). With RF many trees are iteratively aggregated with cross calibration, reducing error in the overall model via boosting and bagging techniques. In addition to constructing each tree using a different bootstrapped sample of the data, the random forest algorithm incorporates a unique approach to splitting. Typically, each node is split using the optimal split among all predictor variables; in the RF algorithm, each node is split using the best predictor among a subset of predictors chosen at random at that node. This additional layer of randomness significantly increases the accuracy of the model and makes RF robust to overfitting.

Using the RF package in the R programming environment, habitat quality was modeled based on the suite of lidar and optical predictor variables described above. The model was run using both 3 and 4 years of bird population and distribution data. Because the random forest algorithm builds trees based on a bootstrap sample of data (reserving approximately 1/3 of the data for testing), it is not necessary to withhold data for testing after model creation. In spite of this, and in addition to running the model using all 4 years of data, we ran the model based on the first 3 years of data and examined the relationship between predicted habitat quality and occupancy in the fourth year.

3. Results

Lidar and optical data products varied throughout the study region; however, some patterns reflected spatial variation in elevation. General trends between lidar and optical predictor variables and habitat quality are shown in Figure 2. Good quality habitat. i.e. that with greater frequency of occupany, was associated with a dominance of deciduous vegetation, a relatively high canopy height, increased vertical complexity, low VDR, and high seasonal change in NDVI (NDVI difference). Although clear trends exist between habitat quality and the selection of variables displayed, there was a large amount of variability in habitat metrics (both lidar and optical) within a single habitat quality class.

The random forest model based on 4 years of occupancy data explained 47% of the variation in habitat quality. Seasonal difference in NDVI, canopy height, elevation, and canopy complexity were selected as the most important predictors of BTBW habitat quality. This model was subsequently applied to the lidar and optical data summarized for each survey site and we examined the frequency of agreement and disagreement between the predicted and observed habitat quality (Table 1). Although the model produced habitat quality values in the range of 0 to 4, we grouped the range of predicted values into 3 habitat quality groups: best (quality of 3 or 4), average (1 or 2), and poor (0). Of the 251 sites predicted to have the best quality habitat, 199 or 79% of them were occupied for 3 or 4 years over the study period, while 17% were occupied for 1 or 2 years and only 4% were not occupied at all. About 58% of the sites identified as average quality habitat were occupied for 1 or 2 years, while 21% of these sites were occupied for 3 or 4 years, and the remaining 21% were not occupied at all. Of the 75 sites predicted as poor habitat, 73% were not occupied over the 4 years of study.

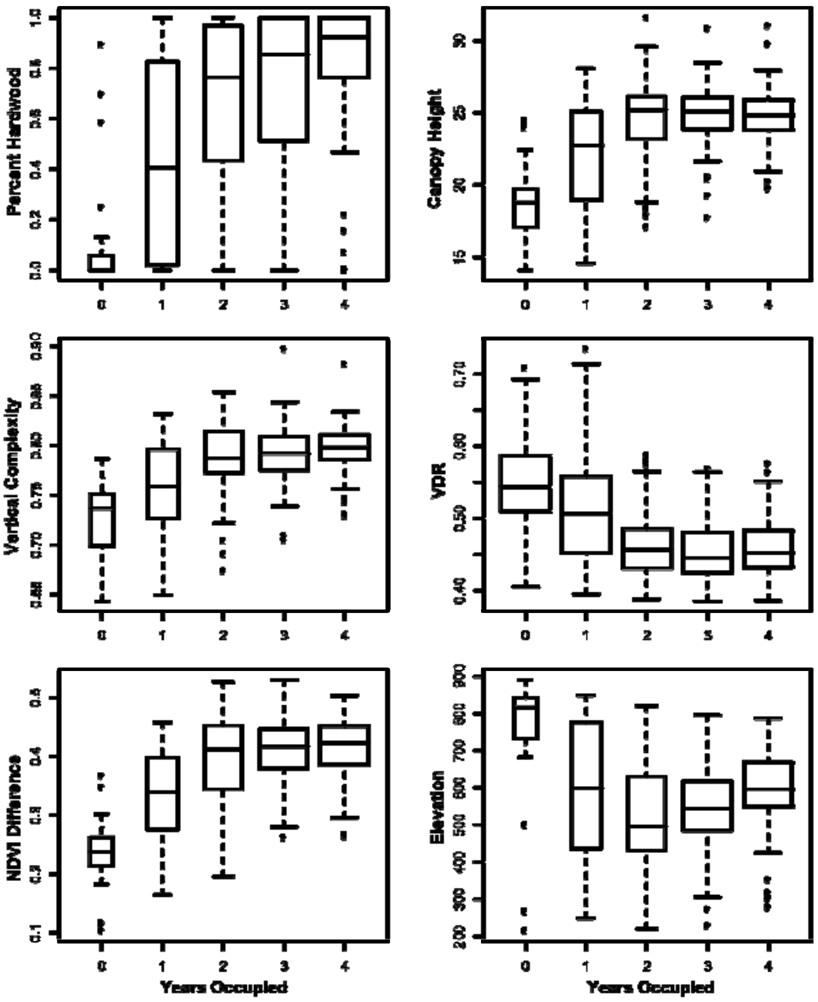

Figure 2. BTBW habitat occupancy relative to the predictor variables used in model development.

As with the model run based on 4 years of data, the random forest model of habitat quality derived from 3 years of occupancy data selected seasonal NDVI difference, elevation, canopy height, and vertical complexity as the strongest predictors of habitat quality. The 3-year model displayed an ability to predict presence/absence in the 4th year of the study period, with 73 sites identified as best quality (quality of 3), 198 as average quality (2) and 100 as poor quality (1 or 0). Thus about 90% of sites identified as best quality data were occupied in the 4th year of the study, while 81% of the average quality sites and only 46% of the poor quality sites were occupied (Table 2).

Table 1: Comparison of predicted occupancy among habitat quality groups.

	Predicted		
	Best (3 or 4 years)	Average (1 or 2 years)	Poor (0)
	251 sites	105 sites	75 sites
Best	199 (79%)	22 (21%)	9 (12%)
Average	42 (17%)	61 (58%)	11 (15%)
Poor	10 (4%)	22 (21%)	55 (73%)

Table 2: Predicted presence or absence among habitat quality groups in year 4 based on model developed using year 1-3 observations.

		Predicted		
		Best (3)	Good (2)	Poor (1 or 0)
		73 sites	198 sites	100 sites
Observed	Present	66 (90%)	162 (81%)	46 (46%)
	Absent	7 (10%)	36 (19%)	54 (54%)

4. Discussion

General trends between remotely sensed metrics of habitat and habitat suitability derived from occupancy data were as expected (Figure 2). Black-throated blue warblers are known to prefer mature forests with a dominance of deciduous vegetation (Doran and Holmes 2005) and we observed a strong positive trend between both habitat quality and percent deciduousness, as well as between habitat quality and canopy height (Figure 2a & b). Clear trends also existed with canopy complexity and the vertical distribution ratio (Figure 2c & d). Habitat quality increased with increasing vertical complexity, demonstrating that a more complex vegetation structure improves habitat for this species. Similarly, lower VDR values were associated with higher quality habitat. Low values of VDR indicate a more uniform distribution of vegetation biomass throughout the canopy profile. Both of these trends may be related to the preference by the black-throated blue warbler for locations with a well-developed understory (Holmes and Doran 2005, Holmes et al. 1996, Steele 1992), particularly the density of hobblebush (*Viburnum alnifolium*) shrubs, i.e., increased occupancy of sites with higher densities of hobblebush. A well-developed understory would have the effect of both lowering VDR, by effectively shifting the median height (HOME) down, and increasing COMP (a similar forest with no understory would have a lower COMP, due to the lack of the understory return in the waveform).

Seasonal change in NDVI was also positively correlated with habitat quality (Figure 2e). This trend is most likely associated with the relationship between vegetation type and seasonal NDVI difference, although it may also indicate a relationship between primary productivity and habitat quality. Deciduous vegetation shows greater phenological changes throughout the growing season than coniferous species, and this trend most likely reflects the preference of the black-throated blue warbler for deciduous forest. Greater seasonal changes in NDVI, however, also indicate greater rates of photosynthesis and primary productivity, which could sustain larger populations of *Lepidoptera*, the black-throated blue warbler's primary food source. The boxplot of elevation as a function of habitat quality (Figure 2f) demonstrates the preference of black-throated blue warblers for low to mid-elevation (400-700m) regions. Areas of higher elevation are typically dominated by coniferous species or sparse vegetation along the rocky ridgelines. Again, the trend between habitat quality and elevation is not directly causal, but is more likely a result of vegetation cover as influenced by elevation.

The random forest model of habitat quality based on lidar and optical predictors and 4 years of occupancy data was skilled in terms of variance explained (Figure 2). When observed and predicted values of habitat quality were grouped into best, average, and poor quality categories, comparison of the observed and predicted values demonstrated strong overall agreement (Table 1). The random forest model based on 3 years of occupancy data was comparably good. Although the percent of variance explained based on the reserved data was relatively low (39%), when the model was applied to the lidar and optical data, and examined in relation to occupancy data from 2006 (year 4), it demonstrated strong predictive power (Table 2). Over 90% of the sites identified as best quality habitat were occupied in 2006. This is particularly interesting because it discounts the effect of individuals showing site fidelity because the 4-year gap between 2002 and 2006 makes it unlikely that the same individuals are returning to the same locations.

Results from sites identified as average or poor habitat were not as compelling. Just over 80% of habitats identified as average were occupied and 46% of the sites identified as poor habitat were occupied in the year 2006. This result, however, could be influenced by the relatively high abundance of black-throated blue warblers in that particular year. Total abundance of the BTBW over all sites in 2006 was higher than in previous years, thus good quality habitat was more limited and greater abundance and occupancy at average and poor habitat sites would be expected.

These results indicate that remotely sensed data can be used to predict habitat quality or suitability of the BTBW, even throughout a relatively homogenous environment. Utilization of lidar data in addition to optical data, provides the ability to sense changes in vegetation structure and density, which adds an integral layer of information to the characterization of habitat, particularly across a region which is relatively homogeneous in terms of spectral reflectance in optical imagery.

Our future work with these data will incorporate additional years of bird observations, focus on a range of different modelling variations, including different methods of sampling the bird observations (e.g. selecting every other grid location and using the remainder for testing), quantifying the influence of spatial autocorrelation, and exploring the utility of an information theoretic modelling approach for assessing habitat quality.

Acknowledgements

We acknowledge support from the NASA Applied Sciences Program, Ecological Forecasting theme, managed by Woody Turner and the NASA Interdisciplinary Science program managed by Diane Wickland. The field data were collected with the logistic support of the Northern Research Station, U.S. Forest Service, Newtown Square, PA with financial support from the National Science Foundation.

References

Blair, J. B., D. L. Rabine and M. Hofton, 1999, The Laser Vegetation Imaging Sensor (LVIS): a medium altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing* 54: 115-122.

Breiman, L. 2001. *Machine Learning* 45, 5–32.

Doran, P. J. and R. T. Holmes, 2005, Habitat occupancy patterns of a forest dwelling songbird: causes and consequences. *Canadian Journal Of Zoology* 83(10): 1297-1305.

Drake, J. B., R. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, R. L. Weishampel, and S. Prince. 2002. Estimation of tropical forest structural characteristics using large-footprint lidar, *Remote Sensing of Environment*, 79: 305-319.

Dubayah , R.O. and J.B. Drake. 2000. Lidar remote sensing for forestry applications. *Journal of Forestry* 98:44-46.

Goetz, S. J., D. Steinberg, R. Dubayah, and B. Blair. 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment* 108:254-263.

Goetz, S. J. 1997. Multi-sensor analysis of NDVI, surface temperature, and biophysical variables at a mixed grassland site. *International Journal of Remote Sensing*, 18:71-94.

Hill, R.A., Hinsley, S.A., Gaveau, D.L.A., & Bellamy, P.E., 2004. Predicting habitat quality for great tits. *International Journal of Remote Sensing,* 25, 4851-4855

Hofton, M. A., L. E. Rocchio, J. B. Blair, and R. Dubayah. 2002. Validation of Vegetation Canopy Lidar sub-canopy topography measurements for a dense tropical forest. *Journal of Geodynamics* 34:491-502.

Holmes, R.T., 1994, Black-throated Blue Warbler (*Dendroica caerulescens*). *In* The birds of North America. No 87. *Edited by* A. Poole and F. Gill. The Academy of Natural Sciences, Philidelphia, and The American Ornithologists' Union, Washington, D.C.

Holmes, R. T., P. P. Marra and T. W. Sherry, 1996, Habitat-specific demography of breeding black-throated blue warblers (*Dendroica caerulescens*): Implications for population dynamics. *Journal Of Animal Ecology* 65(2): 183-195.

Holmes R.T., T.W. Sherry, P.P. Marra, K.E. Petit, 1992, Multiple brooding and productivity of a neotropical migrant, the black-throated blue warbler (*Dendroica caerulescens*), in an unfragmented temperate forest. *The Auk* 109:321-333.

Hunter, M. L. Jr. 1999. Maintaining Biodiversity in Forest Ecosystems. Cambridge University Press, Cambridge, UK.

MacArthur, R. H. and J. W. MacArthur. 1961. On bird species diversity. *Ecology* 42:594-598.

McElhinny, C., P. Gibbons, and C. Brack. 2005. Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management* 218:1-24.

Newton, I, 1998, *Population limitation in birds*. Academic Press, San Diego.

Rapp, V. 2004. Ecosystems and people: managing forests for mutual gains. Science Update 8. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station Portland, OR.

Ralph, C. J., J. R. Sauer, and S. Droege. 1995. Monitoring Bird Populations by Point Counts. General Technical Report PSW-GTR-149, USDA Forest Service, Albany, CA.

Schwartz, P. M., T. J. Fahey, C. W. Martin, T. G. Siccama, and A. Bailey. 2001. Structure and composition of three northern hardwood-conifer forests with differing disturbance histories. *Forest Ecology and Management* 144:201-212.

Sergio, F. and I. Newton, 2003, Occupancy as a measure of territory quality. *Journal Of Animal Ecology* 72(5): 857-865.

Steele, B. B., 1992, Habitat selection by breeding Black-throated Blue Warblers at two spatial scales. *Ornis Scandinavica* 23(1): 33-42.

Turner, M.G., R.H. Gardner, R.V. O'Neill. 2001. Landscape Ecology in Theory and Practice: Pattern and Process. Springer Verlag, New York.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., & Clawges, R.M., 2001. Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment,* 6, 90-9.