

Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA

Scott Goetz^{a,*}, Daniel Steinberg^a, Ralph Dubayah^b, Bryan Blair^c

^a Woods Hole Research Center, Falmouth MA 02540, United States

^b Department of Geography, University of Maryland, College Park MD 20742, United States

^c NASA Goddard Space Flight Center, Greenbelt MD 20771, United States

Received 11 August 2006; received in revised form 8 November 2006; accepted 11 November 2006

Abstract

Habitat heterogeneity has long been recognized as a fundamental variable indicative of species diversity, in terms of both richness and abundance. Satellite remote sensing data sets can be useful for quantifying habitat heterogeneity across a range of spatial scales. Past remote sensing analyses of species diversity have largely been limited to correlative studies based on the use of vegetation indices or derived land cover maps. A relatively new form of laser remote sensing (lidar) provides another means to acquire information on habitat heterogeneity. Here we examine the efficacy of lidar metrics of canopy structural diversity as predictors of bird species richness in the temperate forests of Maryland, USA. Canopy height, topography and the vertical distribution of canopy elements were derived from lidar imagery of the Patuxent National Wildlife Refuge and compared to bird survey data collected at referenced grid locations. The canopy vertical distribution information was consistently found to be the strongest predictor of species richness, and this was predicted best when stratified into guilds dominated by forest, scrub, suburban and wetland species. Similar lidar variables were selected as primary predictors across guilds. Generalized linear and additive models, as well as binary hierarchical regression trees produced similar results. The lidar metrics were also consistently better predictors than traditional remotely sensed variables such as canopy cover, indicating that lidar provides a valuable resource for biodiversity research applications. © 2006 Elsevier Inc. All rights reserved.

Keywords: Abundance; Biodiversity; Diversity; General additive models; Habitat; Heterogeneity; Lidar; Regression trees; Remote sensing; Species richness

1. Introduction

The form of the relationship between species richness and environmental variables is known to be dependent on the scale and taxonomic group of interest (Hawkins et al., 2003; Mittelbach et al., 2001). At local scales, the influence of climate is thought to be a lesser influence on richness patterns than competition, predation and habitat variables such as patch area, connectivity, vegetation type, productivity, and land use (Currie et al., 1999; Rosenzweig, 2002; Turner, 2004). Local species richness (within-community *alpha* diversity) has long been known to be influenced by habitat heterogeneity including, in the case of birds, vegetation cover and density

(Cam et al., 2000; Trzcinski et al., 1999), fragmentation and isolation effects (Boulinier et al., 2001; Donovan & Flather, 2002; Villard et al., 1999), land management and anthropogenic disturbance (Allen & O'Connor, 2000; Berg, 1997), and foliage height diversity (MacArthur, 1964) among other factors. Habitat heterogeneity is, however, often complex and difficult to measure *in situ* — a proverbial case of not being able to see the forest for the trees.

Remote sensing has improved our ability to characterize habitat heterogeneity across a range of spatial scales (Avery & Haines-Young, 1990; Kerr & Ostrovsky, 2003; Nagendra & Gadgil, 1999; Turner et al., 2003). There is now a growing body of literature analyzing remote sensing imagery and derived vegetation maps as predictors of species richness patterns (Fairbanks & McGwire, 2004; Gould, 2000; Hurlbert & Haskell, 2003; Johnson et al., 1998). Similar image data products are used in predicting range distributions with multi-

* Corresponding author. Tel.: +1 508 540 9900; fax: +1 508 540 9700.
E-mail address: sgoetz@whrc.org (S. Goetz).

dimensional habitat suitability models (Ferrier, 2002; Guisan & Thuiller, 2005). Radar remote sensing has potential for acquiring information on canopy structure in three dimensions, including some vertical properties of habitat important for many (not just arboreal) life forms (e.g. Bergen et al., in press). Radar imagery has been used, for example, to assess canopy structure in relation to Australian bird species habitat use (Imhoff et al., 1997), but radar is subject to a range of distortions that may limit its practical utility to a wide range of end-users.

A relatively new form of remote sensing provides another means to acquire information on three-dimensional habitat heterogeneity. Light detection and ranging (Lidar) is based on the use of laser light emitted from a source and reflected back to a sensor as it intercepts objects in its path (Dubayah et al., 2000; Lefsky et al., 2002). As the reflected light is detected at the sensor it is digitized, creating a record of returns that are a function of the distance between the sensor and the intercepted object. This entire stream of reflected laser returns is referred to as a waveform. Subcanopy topography, canopy height, basal area, stem diameter, canopy height profiles, canopy cover and biomass have all been successfully derived from large-footprint lidar waveform data in a variety of forest types (Drake et al., 2002; Harding et al., 2001; Hofton et al., 2002; Nelson et al., 2003).

In this paper, we explored the utility of lidar to estimate habitat metrics associated with bird species richness and abun-

dance in the temperate forests of Maryland, where a unique gridded data set of bird observations had been compiled. Our objectives were: (i) to map habitat metrics using lidar, (ii) assess the utility of the metrics for predicting bird species richness, (iii) compare the lidar-derived predictors with other metrics derived from more traditional remote sensing imagery. A secondary objective was to assess the relative utility of different statistical techniques for analyzing the results, including traditional linear regression, general additive models, and binary hierarchical splitting algorithms (decision trees).

2. Study area and data sets

The Patuxent National Wildlife Refuge (PWNR), located in central Maryland (eastern United States), encompasses 5315 ha surrounding the Patuxent and Little Patuxent Rivers (Fig. 1). It was established in 1936 as the only national wildlife refuge for the expressed purpose of supporting wildlife research, most of which is conducted by the Biological Resources Division of the US Geological Survey (www.pwrc.usgs.gov). Onsite USGS-BRD offices house the North American Amphibian Monitoring Program and the widely recognized national Breeding Bird Survey (BBS). The PWNR study site is sufficiently constrained geographically such that factors other than habitat which may potentially influence diversity patterns (e.g., climate and history) were relatively invariant over the period of bird observations (described below).

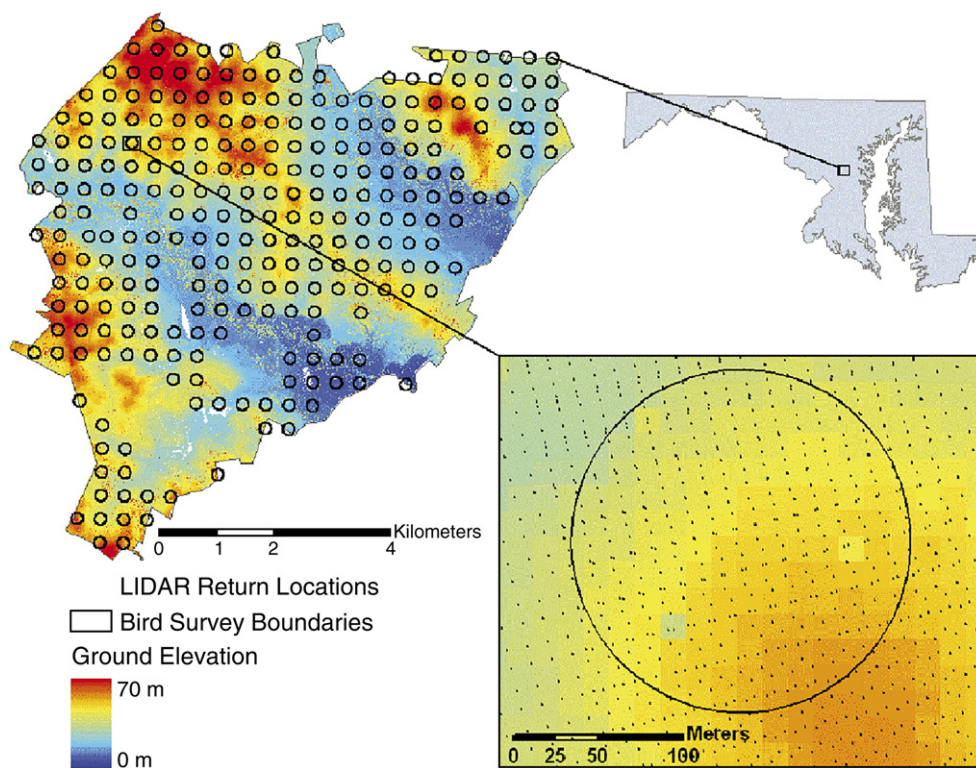


Fig. 1. Patuxent National Wildlife Refuge study area in Maryland, USA. Bird survey locations of 100 m radius are shown as circles overlaid on an image of surface elevations derived from lidar data. The lower elevation blue areas in the image include branches of the Patuxent river. Also shown (lower right) are the individual lidar shot locations (dots) within a bird observation area (circle). Note the variable density of lidar shots in the image, which resulted from multiple flight lines over the study area.

2.1. Bird observations

Avian population data were collected throughout the PNWR in June of 1996 and 1997 by USGS-BRD personnel following the methods described by Ralph et al. (1995). A grid of 266 survey points, spaced at 400 m intervals, was established within the refuge (Fig. 1). At each survey point the abundance of individual bird species encompassed within a 100 m-radius circle about the sampling point was recorded for 5 min using a combination of audio and visual observations. A total of 5042 individuals representing 88 species of birds were observed at the sampling stations within the PNWR and were separated into 6 different guilds based on their respective habitats. These included guild descriptions widely used in the BBS. More specific guild descriptors, such as nesting height (canopy versus ground) and type (cavity versus open cup), were considered too general in terms of our assessment of habitat heterogeneity.

Total species richness and abundance were calculated for each ~30,000 m² survey cell. Forest birds dominated the study area, comprising 31 (35%) of the species and 3128 (62%) of the individuals observed (Table 1). The most common species observed was the red-eyed vireo (*Vireo griseus*), with 479 individuals observed throughout the survey period. Abundance was calculated as the total number of individuals of a species observed within each survey cell, and species richness as the total number of species observed. The abundance and richness of bird species per survey cell ranged from 3 to 113, and from 2 to 27, respectively. Mean abundance and richness were 19.0 and 12.2, respectively. In addition, the observational data were stratified by habitat preference, or avian guild (forest, open-forest, semi-open forest, scrub, suburban and wetland), and the portion of total richness accounted for by each guild. Thus, there were 8 species diversity metrics associated with each survey location: total abundance, total species richness, and 6 per-guild species richness values (Table 1). For the analyses that follow we did not further consider abundance due to the time mismatch between the bird and lidar observations, nor did we analyze the semi-open forest species owing to the generalized habitat preferences of this guild (which included two species of vulture as well as the red-tailed hawk, great horned owl, yellow-shafted flicker, eastern kingbird and brown-headed cowbird). Although two other guilds (open forest and wetland) had only slightly more species, they contained more typical habitat specialists including, in the case of the open forest guild, the

eastern phoebe, eastern meadowlark, eastern bluebird, American kestrel, killdeer and three species of swallow. Wetland species included the red-winged blackbird and belted kingfisher as well as a variety of ducks, gulls, geese and herons.

2.2. Lidar metrics

We used lidar measurements acquired with the Laser Vegetation Imaging Sensor (LVIS) over the PNWR in August of 2003. The LVIS instrument is an imaging laser altimeter, designed and developed at NASA's Goddard Space Flight Center (Blair et al., 1999). It has a 7° field of view within which sampling "footprint" sizes can be varied depending on, among other factors, the altitude at which the instrument is flown. Waveforms are converted to units of distance by accounting for the time elapsed between the initial laser pulse and the return. Geolocation was accomplished by collecting position and directional data at the time of each pulse, to which the ground footprint was referenced post-flight (see Fig. 1).

Full waveform lidar data of the PNWR was acquired at night under clear sky conditions using LVIS from an altitude of 7 km above ground level and a nominal 12 m footprint. A number of different products were derived from the waveform data including ground elevation (Fig. 1), canopy height (CH), the height of median energy (HOME), and the height at which 25 and 75% of the cumulative waveform energy was recorded (Fig. 2). Each of these variables were calculated in reference to the ground return, thus accurate determination of the ground surface is a crucial step in the processing of the data products. Identification of the ground and canopy returns is carried out by an automated algorithm which reduces noise within waveform and then 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. The CH product was derived as the difference in height between the initial canopy return and the ground return. HOME was calculated by locating the median of the entire waveform, including both canopy and ground return energies, and computing the distance between this location and the ground return. A similar technique was used to calculate 25% and 75% energy return heights. We also derived a normalized ratio between the CH and HOME products, which provided an index of the vertical distribution of intercepted canopy elements (biomass) ranging between 0 and 1. We refer to this as the Vertical Distribution Ratio; $VDR = [CH - HOME] / CH$. Areas with a dense canopy and sparse understory tend to exhibit a low VDR due to the relatively short distance between CH and HOME. Areas with a more even vertical distribution of biomass exhibit larger VDR's (closer to 1).

For each of the lidar products described above, the location of each point corresponds to an individual lidar shot and associated waveform. For visualization purposes, we created images of each of the ground elevation, CH, HOME, VDR and other returned energy products using a spatial interpolation scheme. The images were processed for the entire area encompassing the PNWR boundaries (Fig. 2), but all numerical analyses relative to the bird observations that follow were based

Table 1
Bird species richness and abundance stratified by habitat preference guild

| Guild | Species richness | % of total richness | Abundance | % of total abundance |
|------------------|------------------|---------------------|-----------|----------------------|
| Forest | 31 | 35 | 3128 | 62 |
| Scrub | 18 | 20 | 710 | 14 |
| Suburban | 15 | 17 | 875 | 17 |
| Wetland | 9 | 10 | 126 | 2 |
| Open forest | 8 | 9 | 83 | 2 |
| Semi-open forest | 7 | 8 | 120 | 2 |
| Total | 88 | 100 | 5042 | 100 |

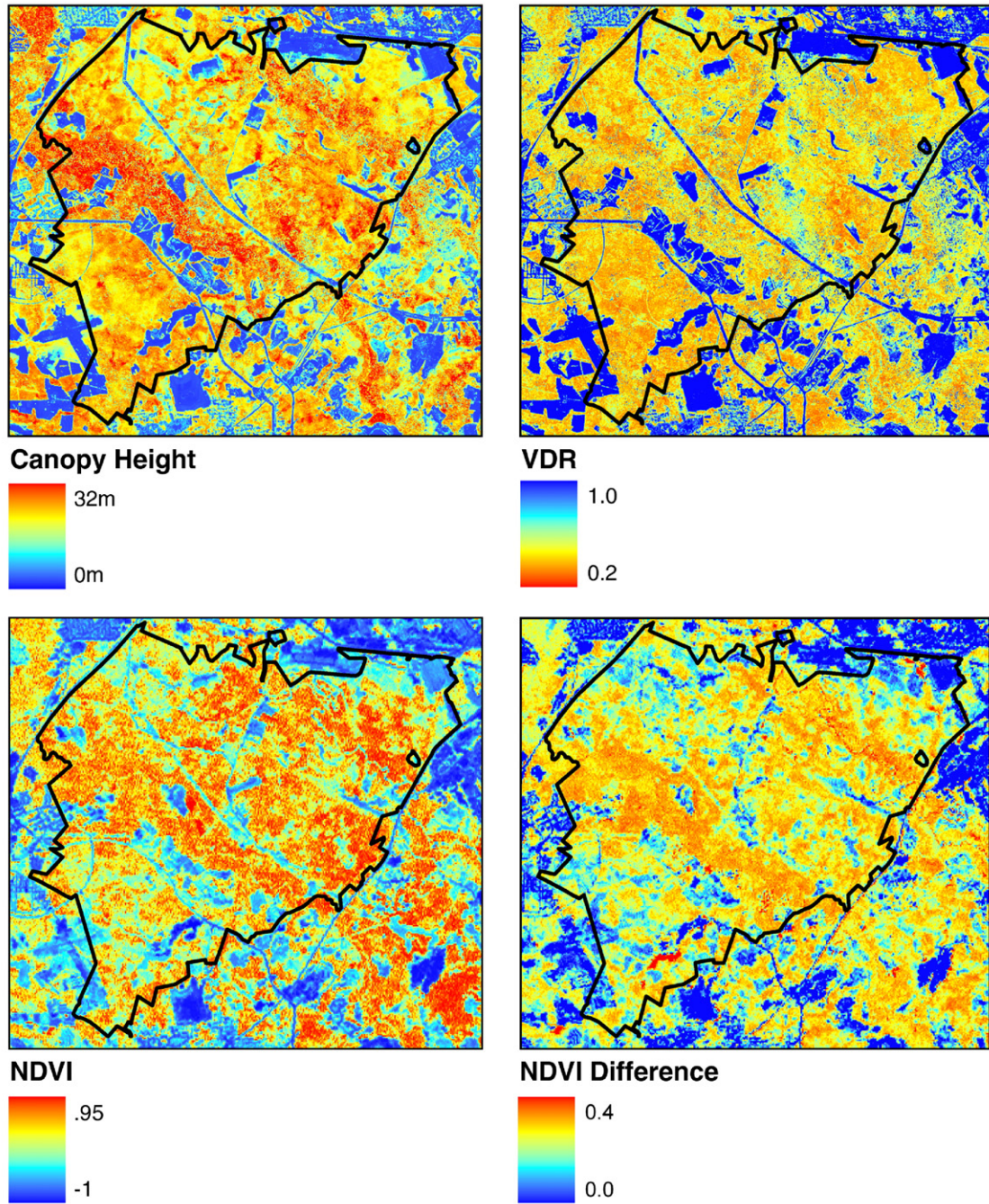


Fig. 2. Lidar image products of the PNWR and surrounding environs depicting vegetation canopy heights and the vertical distribution ratio (VDR, see text). Note the areas occupied by some of the taller forest trees in the canopy height image are within the Patuxent river corridors, visible in Fig. 1. The cross shaped area in the lower right is a light aircraft runway. The lower two images are optical (Landsat ETM) vegetation index products derived for mid-summer (left) and for the difference between summer and leaf-off conditions (right).

on the use of the lidar waveform products (not the interpolated images).

2.3. Optical imagery

In addition to the lidar variables, described above, we processed optical remote sensing imagery from the Landsat series of satellites and analysed those in relation to the bird observation

data set. Rather than considering land cover type, which is relatively invariant across the PNWC, we analyzed spectral vegetation indices derived from two cloud-free Landsat Enhanced Thematic Mapper (ETM) images acquired during both leaf-on and leaf-off conditions. Two derived vegetation index products were examined: leaf-on normalized difference vegetation index ($[\text{infrared} - \text{visible}] / [\text{infrared} + \text{visible}]$) (NDVI), and seasonal NDVI change.

The Landsat ETM scenes (path/row 15/33), acquired on 2 August 2001 and 24 March 2000, were calibrated to spectral radiances and then converted to top-of-atmosphere (TOA) 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 then geographically referenced while also correcting for topographic distortion (i.e., orthorectification). NDVI images were calculated from the referenced TOA reflectances, and the two scenes were differenced in order to calculate the image of seasonal NDVI change (Fig. 2). In this way we were able to consider the seasonality of vegetation cover and density. Most of the PNWR was forested (~80%) but it was not possible to discriminate between forest type classes beyond deciduous versus evergreen habit (Goetz et al., 2004). These distinctions had no utility for analyzing bird species richness patterns, but they did allow us to analyze some of the lidar products by these general forest cover types.

3. Statistical models

Bird survey cell boundaries were intersected with the lidar (LVIS) and optical (ETM) image products (Table 2) using a geographic information system, and statistical summaries of the data falling within the boundaries of each cell were computed for all predictor variables (Fig. 1) using the *R* statistical package (R Development Core Team, 2005). In addition, we calculated horizontal spatial variance within the cells for each of the predictors, as well as minimum and maximum values of each optical variable in order to better account for seasonal changes in vegetation and to increase the potential utility of these data sets.

To visualize the general trends between response and predictor variables, the distribution of each predictor was examined in relation to each response variable, e.g., mean canopy height was plotted with forest species richness, and so on. Four statistical techniques were then used to examine the relationships between avian species diversity and canopy habitat heterogeneity metrics derived from both lidar and optical data sets: Akaike's Information Criteria (AIC), stepwise multiple linear regression (MLR), generalized additive models (GAM), and regression trees. AIC was used to identify the relative importance of predictor variables among all possible sets of predictors for a given response variable (i.e., species richness), where the best performing predictors were identified by the lowest AIC scores. MLR was then used to assess

linear trends between predictor and response variables. MLR is the most common and straightforward statistical method of assessing relationships between variables, but is subject to a range of assumptions such as normal error distributions, as well as limitations due to co-linearity of predictor variables that can result in inflated estimates of explained variance. We assessed co-linearity, and also examined the influence of spatial autocorrelation on variable selection using Moran's-I correlograms.

GAMs and regression trees were used to better account for potential non-linear trends between the response and predictor variables, and to minimize the influence of co-linearity among predictor variables (e.g. Guisan et al., 2002). GAMs require fewer assumptions of data distributions and error structures, assuming only that functions are additive and components can be smoothed by local fitting to subsets of the data. Smoothing parameters were automatically selected based on the effective degrees of freedom and a generalized cross validation criterion in *R*. Regression tree analysis is a technique for partitioning data based on a series of hierarchical binary splits of the predictor variables, forming a tree structure that terminates in nodes associated with discrete ranges in the response variable (Breiman, 2001). Regression trees are non-parametric, where data partitioning at each split minimizes the sum of the squared deviations from the mean in the partitioned groups, thus reducing the residual variance at each successive split. We used a boosting technique in the regression tree models, in which the population was randomly sampled repeatedly to effectively bootstrap the results based on cross-validated explained variance. Pruning the tree results was unnecessary because the sample sizes and number of splits were relatively small for a regression tree approach. The terminal nodes in our regression tree analysis were ranges of bird species richness.

Using these four approaches, the total and the per-guild species richness were modeled using a selected set of significant predictor variables (of the suite listed in Table 2). For each statistical approach, response variables were modeled using: (1) lidar predictors only, (2) optical predictors only, and (3) both lidar and optical variables. This allowed us to identify the most significant predictor variables from lidar and optical derived data sets and to quantify their relative contribution to the total explained variance.

Both the MLR and GAMs were run using a k-fold cross-validated forward stepwise selection procedure. Data selection, model training, and model testing were carried out in an iterative manner proportional to the sample size of the data. For each iteration a unique random selection of 75% of the data was used for model selection and training, and a unique random selection of 25% of the data were used for model testing. Using the training data, each response variable was modeled based on a single predictor (using MLR or GAMs) and these single variable models were tested using the withheld data. After all iterations, the predictor that explained the greatest amount of variation in the test data (based on the mean coefficient of determination) was selected for inclusion in the model. This process was repeated for each remaining response variable, at each step using the withheld data for cross validation, until no remaining predictor explained significantly more (>1%) of the variation in the response variable.

Table 2

Response and predictor variables, with the latter separated into habitat metrics derived from optical and lidar remote sensing data sets

| Response variable (sample size) | Lidar predictors | Optical predictors |
|-----------------------------------|------------------------|------------------------|
| Total richness (266) | Canopy height | NDVI |
| Forest species richness (263) | Canopy height σ | NDVI σ |
| Scrub species richness (191) | HOME | NDVI min, max |
| Suburban species richness (231) | HOME σ | Δ NDVI |
| Wetland species richness (45) | VDR | Δ NDVI σ |
| Open forest species richness (31) | VDR σ | Δ NDVI min, max |
| | Elevation | |

The sample size for each response variable is the total number of survey cells at which the species was observed, of 266 possible locations. σ indicates spatial variability of the variable, and Δ indicates change.

4. Results

The lidar metrics varied between land cover types, as mapped using multi-temporal Landsat imagery. VDR was relatively smaller in evergreen than deciduous forest and higher in forested wetland areas, indicating a more even vertical distribution of biomass in the former and less mid-canopy and understory in the latter. The bird species richness observations showed systematic trends with the various habitat metrics, but also showed substantial variability across the range of habitat properties (Fig. 3). Total richness increased with VDR, for example, but displayed increased between-class variability at higher VDR values. Total richness tended to be somewhat higher at canopy heights below 20 m since species richness of the forest bird guild was smaller than the other guilds combined (Table 1). Conversely, total abundance of forest birds was greater than the other guilds, thus canopy height and total abundance were positively correlated (not shown). Between-class total richness was also more variable for canopy heights below 20 m due to relatively smaller sample sizes in those height classes.

Forest bird species richness increased systematically with canopy height (Fig. 3c), but varied substantially within height classes (e.g., areas with canopy heights between 18–20 m had a median species richness of 7.5 with interquartile ranges of 6 to 10 species). Scrub species showed a marked drop in richness when median canopy height exceeded 9 m (Fig. 3d). Other

associations between individual predictor and response variables are presented in the statistical analyses that follow.

4.1. Linear models

The tests using Moran's-I correlograms indicated that the only significant ($p < 0.05$) spatial autocorrelation was in the forest bird guild, with a lag distance of approximately 1000 m. Spatial autocorrelation does not inflate the explained variance term (R^2) but can lead to inflated (artificially large) sample sizes due to non-independent samples, which may influence significant tests (p -values) and result in the inclusion of non-significant predictor variables. The calculated Akaike Information Criteria (AIC) scores identified the most consistently selected predictors. This assessment indicated that the lidar vertical distribution ratio was the most frequently selected metric among all models with each of the various response variables (Table 3). Co-linearity among the most consistently selected predictors, those with rank correlations greater than ± 0.5 , revealed potentially redundant habitat metrics, such as canopy height and HOME, as well as their derivative spatial variability metrics. The strength of the correlations varied when stratified by areas frequented by the different bird guilds, and correlated predictors have clearly different associations with response variables, thus rather than eliminate these predictors at the outset based on co-linearity we chose to use the iterative stepwise forward selection procedure in the MLR models to incorporate the most relevant predictors (Table 3). Moreover,

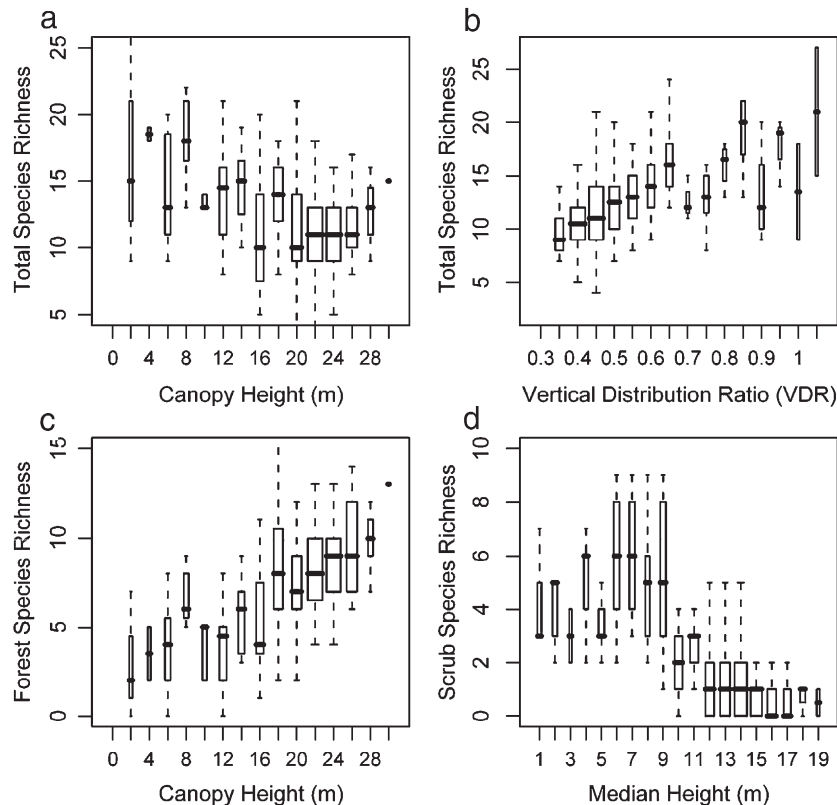


Fig. 3. Boxplots showing the range of response variable (species richness) values relative to key habitat predictor variables for: (a, b) total species richness, (c) forest species richness, (d) scrub species richness. Each box shows the median (horizontal line), quartiles (upper and lower extent of box) and range (dashed vertical lines) for each binned range within the predictor variables. The width of the boxes is proportional to sample size.

Table 3
Significant predictor variables selected using Akaike's Information Criteria, ordered in the rank selected

| Response variable | Predictor variables | | |
|--------------------------------|---------------------|-------------------|---------------|
| <i>All predictors</i> | | | |
| Total richness | VDR | Elevation | HOME σ |
| Forest species | Canopy height | Elevation | HOME σ |
| Open forest species | VDR | HOME σ | – |
| Scrub species | HOME | VDR σ | Elevation |
| Suburban species | VDR | NDVI | – |
| Wetland species | VDR | Elevation | VDR σ |
| <i>Optical predictors only</i> | | | |
| Total richness | NDVI | Δ NDVI | – |
| Forest species | Δ NDVI | – | – |
| Open forest species | NDVI | – | – |
| Scrub species | NDVI min | – | – |
| Suburban species | NDVI | – | – |
| Wetland species | NDVI | Δ NDVI min | – |

Corresponding MLR models of explained variance are provided in Fig. 4. Models based on lidar-only predictors selected the same variables as those using all predictors, except in 1 of 16 cases (see text).

because we compared cross-validated models, using 25% of the data set to test the model developed with 75%, any potential variance inflation was effectively diminished.

The results of the MLR models for total and guild-specific richness are summarized in Fig. 4. Species richness was best predicted when stratified by guild, with different variables selected as the primary predictors across guilds. Scrub — secondary growth species proved to be the most robustly predicted guild, with 45% explained variation in species richness. As in the AIC analysis, the VDR was selected as the primary predictor for total species richness, open forest, suburban and wetland species richness, indicating the importance of vertical habitat distribution for bird diversity. Similar results held for total bird abundance, although we do not include those results here since the bird observations were made in a different year than the lidar acquisitions and abundance can vary substantially on an interannual basis owing to factors other than habitat (e.g. extreme weather events). Canopy height and HOME were selected as primary predictors of forest and scrub species richness, respectively, which also fits with expectations of habitat use for these guilds. No habitat metrics derived from optical remote sensing variables were selected as either primary, secondary or tertiary predictors of bird diversity (or abundance), with the exception of NDVI for suburban guild species richness (secondary). Note the sample sizes (number of sampling areas with guild species present) for the wetland and open forest guilds were small relative to those of the other guilds (Table 2).

Models based solely on optical predictors accounted for significantly less of the variation in species richness than models based on lidar predictors (Fig. 4). As with the lidar based models, species richness was predicted best when stratified by guild. The model of suburban species richness explained the most variability (30%) using only optical habitat predictors. The NDVI was the primary predictor of total species richness, as well as open forest, suburban and wetland species guilds. Forest and scrub species richness were predicted best by maximum NDVI,

seasonal NDVI difference, and minimum NDVI, respectively. The models based on optical predictors rarely selected more than a single significant variable and the metrics of spatial variability were never selected for either NDVI or its seasonality.

Models using both lidar and optical predictors explained no more of the variation in species richness than those based solely on the lidar variables (Fig. 4). Conversely, combining lidar with the optical variables explained an average of 13% more of the variation than models based on optical predictors alone. In all 6 models using the full suite of predictor variables, lidar variables were selected as significant predictors of bird species richness in 15 out of 16 cases (the only exception being the selection of NDVI over the spatial variance of HOME as a secondary predictor of suburban bird richness). Overall predictive capability using all variables was best in forest, open-forest and scrub species richness, and worst in suburban and wetland species richness.

4.2. Additive and non-parametric models

Use of the generalized additive models did not substantially improve upon the MLR models, explaining essentially an equal amount of variation in the response variables. GAMs based solely on optical predictors showed minor improvement (3%) over comparable MLR models, but differences using all variables and lidar-only variables were negligible (<1%). Specific guild richness models that were most improved using GAMs included the scrub and open forest species (5.5% and 7.3%, respectively, with all variables included). Both of these models were significantly improved over MLR ($p < 0.05$). The primary predictor variables selected by the GAMs differed little from those selected by the MLR models, with the exception Open Forest species for which canopy height rather than VDR was selected as the primary predictor.

Results of the regression tree models of species richness were also similar to those of the MLR models. Lidar predictors were

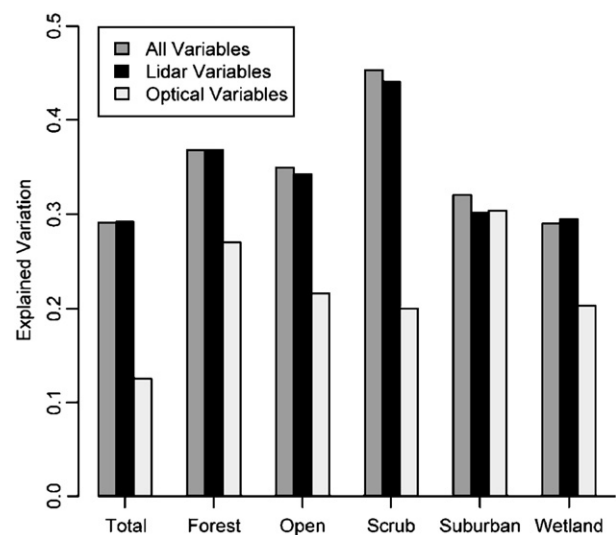


Fig. 4. Total variance in response variables (species richness) explained by habitat metric predictor variables, as derived from lidar versus optical remote sensing data sets. All were significant at $p < 0.01$ or better.

most often selected as the primary variable split, and the species richness models improved when stratified by guild. Models employing only lidar variables predicted species richness and per-guild species richness significantly better than models based on optical variables alone. The explained variation in the response variables accounted for by the regression tree models were comparable to those of the MLR models, and are therefore not further reviewed here. Nonetheless, the regression tree models display the amount of variation explained by the models within the first or second variable split (Fig. 5a). Note from this figure that lower canopy heights and areas with less height

variability contained fewer forest bird species, whereas the greatest species richness occurred in taller canopy forests occupying lower elevation (riparian wetland) areas. Although the apparent R^2 value increased following additional splits, the relative R^2 tended to decrease, indicating that any more complex models would be over-fit. Applying the regression tree model to the lidar data products allowed us to produce maps of species richness, subject to the limits of the statistical fits, for areas both within and outside the PNWC (Fig. 5b). Note the areas outside the PNWR boundaries tend to have lower forest species richness values.

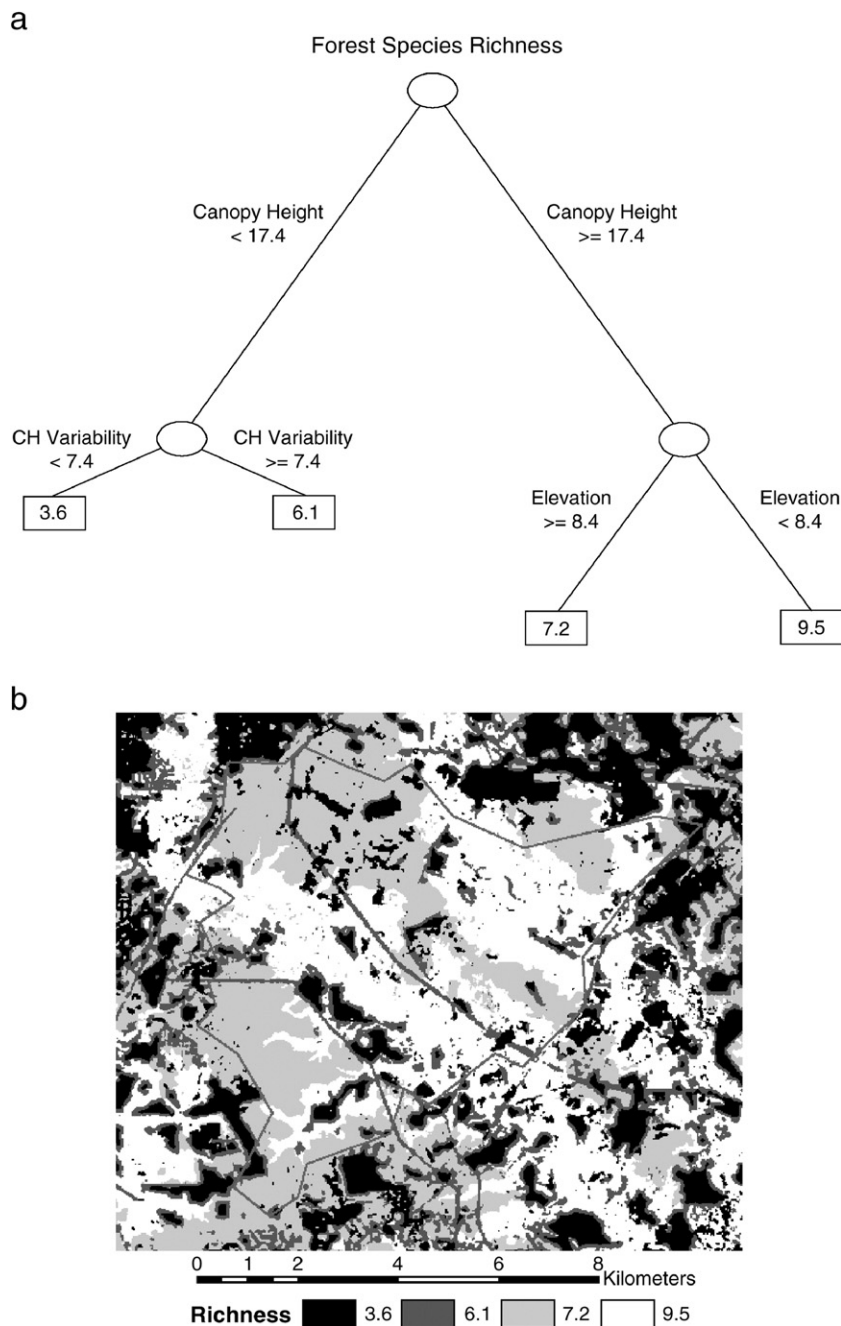


Fig. 5. a. Regression tree output of forest bird species richness. The initial binary split in explained variance (indicated by the vertical spacing between nodes) was on canopy height, followed hierarchically by splits on the spatial variability in canopy height, and ground elevation. Mean species richness is displayed within the terminal nodes. b. Map of forest bird species richness derived from the regression tree model (Fig. 5a) and the lidar data products (Fig. 2).

5. Discussion

Predictions of bird species richness at the PNWR using remotely sensed metrics of habitat heterogeneity, including tree height diversity and vertical mass distribution estimates, were moderately skilled in terms of variance explained (Table 2), and the derived models were robust in that the results were consistent when data were reserved for prediction and then cross-validated. The best models explained 45% of the variation in bird species richness, although more typically 30–40% of species richness was explained using the statistical models stratified by guild. Lidar metrics of habitat heterogeneity performed systematically better than those based on optical remote sensing, and fusion of both optical and lidar improved little upon use of lidar alone. Similar findings have been noted with respect to predicting spotted owl species presence in the Sierra Nevada of California (Hyde et al., 2006).

We note that the best predicted bird guilds were those most easily observed (scrub species) or most frequently observed (forest guilds), whereas those predicted less well were either generalists (suburban or semi-open forest species) or highly restricted in range for areas where vertical habitat dimensionality was minimal (wetland species). Predicting the presence of habitat specialists would be of interest, particularly for threatened species conservation, but the bird observation data set used here were not sufficiently robust to address this topic. Specifically, issues of detectability would need to be considered, including repeated observations to detect rare species and to ensure adequate site occupancy characterization (Boulinier et al., 1998; MacKenzie et al., 2002). Our results indicate that lidar observations are useful in characterizing interior forest habitats, as expressed by the VDR, HOME and other lidar metrics, and these would clearly have utility in identifying habitats suitable to multi-stage canopy specialists. We are currently exploring this topic in another study area with extensive, long-term, well documented bird observations.

Despite the potential benefits of statistical techniques such as GAMs and regression trees over traditional methods like MLR, we found little advantage in these approaches for the current analysis. Part of the limitations we observed may be due to a failure to adequately incorporate interactions among the predictors, although we intentionally limited the predictors to those that could be derived from remote sensing observations (others, such as climate, were relatively invariant across the study area extent). Similarly, the regression tree approach was limited in that the skill of the models did not improve beyond inclusion of the first few variables, which explained most of the variance after two or three binary splits. The overall explained variance was not substantially or systematically improved over the MLR or GAM approaches. Whereas regression trees are a useful non-parametric technique for a variety of applications (Breiman, 2001), they typically function best with larger sample sizes than were available for the current analysis.

These results arise at least partly from the limits of capturing very local scale habitat variability, particularly where that variability is constrained relative to the mobility of the organism of interest. We worked with bird observations because they

were, along with butterflies, the most extensively characterized species data sets available and, in our case, intensively studied in a systematic fashion across a regular grid by widely recognized bird experts (Ralph et al., 1995). Moreover, mapping species diversity is a difficult proposition even with extensive field surveys to guide a suite of predictors derived from various sources, partly due to the detectability of species in different habitat mosaics (Boulinier et al., 1998; MacKenzie et al., 2002). As with the results we present here, other alpha species diversity predictions have performed only moderately well (R^2 values of 50–60%), even over spatial extents broader than those we examined (e.g., Luoto et al., 2004; Seto et al., 2004). This was true at the PNWC because of the dominance of forest vegetation across the site, and the relative horizontal uniformity of the vegetation. Nonetheless, the lidar habitat metrics clearly and consistently performed better than those based on more traditional optical remote sensing because they were able to characterize the vertical habitat structure information relevant to bird diversity. Optical data (Landsat in this case), in contrast, are more sensitive to canopy photosynthetic material and density than vertical structure. Additional variables that can be derived from lidar, including biomass, basal area and stem density, may improve habitat heterogeneity metrics where adequate field data are available to derive statistical descriptors across the study domain. Other lidar data sets, such as those available from the Center for LIDAR Information Coordination and Knowledge (lidar.cr.usgs.gov), may permit extension of this analysis to additional areas where adequate biodiversity observations exist.

Predicting species richness requires consideration of factors other than habitat heterogeneity, even at local scales, including a combination of temporal changes in environmental conditions across a given site, the responses of organisms to those environmental variations, and the interactions of organisms via both intra and interspecific competition. The biotic responses to changes in the availability of resources, often associated with the frequency and type of disturbances, produce range variations that may result in species diversity patterns that vary considerably within a given habitat — despite the best efforts to characterize habitat heterogeneity. Challenges in measuring both current species diversity and associated habitat properties introduce additional uncertainty. Reducing these uncertainties may produce more robust species diversity models and, in this light, we believe the advent of lidar remote sensing has an important role to play in characterizing multi-dimensional habitat heterogeneity. Lidar may ultimately prove most useful when considered in combination with data sets that convey information on other aspects of habitat, such as vegetation type and cover density, as well as temporal variations in habitat properties.

Acknowledgements

We thank Jane Fallon, James Lynch and John Sauer for providing their bird observation data set and supporting documentation, and for fielding queries. This work was supported by NASA Interdisciplinary Science grant NNG04G005G.

References

- Allen, A. P., & O'Connor, R. J. (2000). Interactive effects of land use and other factors on regional bird distributions. *Journal of Biogeography*, 27, 889–900.
- Avery, M. I., & Haines-Young, R. H. (1990). Population estimates for the dunlin *Calidris alpina* derived from remotely sensed satellite imagery of the Flow Country of northern Scotland. *Nature*, 344, 860–862.
- Berg, A. (1997). Diversity and abundance of birds in relation to forest fragmentation, habitat quality and heterogeneity. *Bird Study*, 44, 355–366.
- Bergen, K.M., Gilboy, A.M., Brown, D.G., Gustafson, E.J. (in press). Multi-dimensional vegetation structure in biodiversity informatics: A case study modeling bird species habitat. Ecological Informatics.
- Blair, J. B., Rabine, D. L., & Hofton, M. (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.
- Boulinier, T., Nichols, J. D., Hines, J. E., & Sauer, J. R. (2001). Forest fragmentation and bird community dynamics: Inference at regional scales. *Ecology*, 82, 1159–1169.
- Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E., & Pollock, K. H. (1998). Estimating species richness: The importance of heterogeneity in species detectability. *Ecology*, 79, 1018–1028.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cam, E., Nichols, J. D., & Frather, C. H. (2000). Relative species richness and community completeness: Birds and urbanization in the Mid-Atlantic states. *Ecological Applications*, 10, 1196.
- Currie, D. J., Francis, A. P., & Kerr, J. T. (1999). Some general propositions about the study of spatial patterns of species richness. *Ecoscience*, 6, 392–399.
- Donovan, T. M., & Flather, C. H. (2002). Relationships among North American songbird trends, habitat fragmentation, and landscape occupancy. *Ecological Applications*, 12, 364–374.
- Drake, J., Dubayah, R., Clark, D. A., Knox, R. G., Blair, B., Hofton, M., et al. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79, 305–319.
- Dubayah, R., Knox, J. C., Hofton, M., Blair, J. B., & Drake, J. (2000). Land surface characterization using lidar remote sensing. In M. J. Hill & R. Aspinall (Eds.), *Spatial information for land use management* (pp. 25–38). Singapore: International Publishers Direct.
- Fairbanks, D. H. K., & McGwire, K. C. (2004). Patterns of floristic richness in vegetation communities of California: Regional scale analysis with multi-temporal NDVI. *Global Ecology and Biogeography*, 13, 221–235.
- Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, 51, 331–363.
- 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.
- Goetz, S. J., Jantz, C. A., Prince, S. D., Smith, A. J., Wright, R., & Varlyguin, D. (2004). Integrated analysis of ecosystem interactions with land use change in the Chesapeake Bay watershed. In R. S. DeFries, G. P. Asner, & R. A. Houghton (Eds.), *Ecosystems and land use change* (pp. 263–275). Washington DC: American Geophysical Union.
- Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications*, 10, 1861–1870.
- Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2–3), 89–100.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Harding, D. J., Lefsky, M. A., Parker, G. G., & Blair, J. B. (2001). Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, 76, 283–297.
- Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J. -F., Kaufman, D. M., et al. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105–3117.
- Hofton, M. A., Rocchio, L. E., Blair, J. B., & Dubayah, R. (2002). Validation of vegetation canopy lidar sub-canopy topography measurements for a dense tropical forest. *Journal of Geodynamics*, 34, 491–502.
- Hurlbert, A. H., & Haskell, J. P. (2003). The effect of energy and seasonality on avian species richness and community composition. *American Naturalist*, 161, 83–97.
- Hyde, P., Dubayah, R., Peterson, B., Blair, J. B., Hofton, M., Hunsaker, C., et al. (2006). Mapping habitat suitability using waveform lidar: Validation of montane forest structures. *Remote Sensing of Environment*, 102, 63–73.
- Imhoff, M. L., Sisk, T. D., Milne, A., Morgan, G., & Orr, T. (1997). Remotely sensed indicators of habitat heterogeneity: Use of synthetic aperture radar in mapping vegetation structure and bird habitat. *Remote Sensing of Environment*, 60, 217–227.
- Johnson, D. D. P., Hay, S. I., & Rogers, D. J. (1998). Contemporary environmental correlates of endemic bird areas derived from meteorological satellite sensors. *Proceedings of the Royal Society of London*, 265, 951–960.
- Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology and Evolution*, 18, 299–305.
- Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar remote sensing for ecosystem studies. *BioScience*, 52, 19–30.
- Luoto, M., Virkkala, R., Heikkinen, R. K., & Rainio, K. (2004). Predicting bird species richness using remote sensing in boreal agricultural-forest mosaic. *Ecological Applications*, 14, 1946–1962.
- MacArthur, R. H. (1964). On bird species diversity. *Ecology*, 42(3), 594–598.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- Mittelbach, G. G., Steiner, C. F., Scheiner, S. M., Gross, K. L., Reynolds, H. L., Waide, R. B., et al. (2001). What is the observed relationship between species richness and productivity? *Ecology*, 82, 2381–2396.
- Nagendra, H., & Gadgil, M. (1999). Biodiversity assessment at multiple scales: Linking remotely sensed data with field information. *Proceedings of the National Academy of Sciences*, 96, 9154–9158.
- Nelson, R., Valenti, M., Short, A., & Keller, C. (2003). A multiple resource inventory of Delaware using airborne laser data. *BioScience*, 53, 981–992.
- Ralph, C. J., Sauer, J. R., & Droege, S. (1995). *Monitoring bird populations by point counts*. General technical report PSW-GTR-149 Albany, CA: USDA Forest Service [181 pages www.fs.fed.us/psw/publications/documents/gtr149/gtr_149.html].
- R Core Development Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN3-900051-07-0.
- Rosenzweig, M. L. (2002). *Species diversity in space and time*. Cambridge: Cambridge University Press [436 pp.].
- Seto, K. C., Fleishman, E., Fay, J. P., & Betrus, C. J. (2004). Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing*, 25(20), 4309–4324.
- Trzcinski, M. K., Fahrig, L., & Merriam, G. (1999). Independent effects of forest cover and fragmentation on the distribution of forest breeding birds. *Ecological Applications*, 9, 586–593.
- Turner, J. R. G. (2004). Explaining the global biodiversity gradient: Energy, area, history and natural selection. *Basic and Applied Ecology*, 5, 435–448.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18, 306–314.
- Villard, M. -A., Trzcinski, M. K., & Merriam, G. (1999). Fragmentation effects on forest birds: Relative influence of woodland cover and configuration on landscape occupancy. *Conservation Biology*, 13, 774–783.