Available online at www.sciencedirect.com

SCIENCE DIRECT®

Remote Sensing of Environment 96 (2005) 427 – 437

Remote Sensing of Environment

www.elsevier.com/locate/rse

Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems

P. Hyde[a,*], R. Dubayah[a], B. Peterson[a], J.B. Blair[b], M. Hofton[a], C. Hunsaker[c], R. Knox[b], W. Walker[d]

[a]*Department of Geography, University of Maryland, College Park, 20742, MA, United States*
[b]*Laboratory for Terrestrial Physics, NASA's Goddard Space Flight Center, Greenbelt, MA, United States*
[c]*U.S. Forest Service, Department of Agriculture, Fresno, CA, United States*
[d]*Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United Sates*

Received 24 August 2004; received in revised form 24 February 2005; accepted 13 March 2005

Abstract

Consistent and accurate measurements of forest structure at the landscape scale are required by forest ecologists and managers for a variety of applications. Lidar remote sensing has proven to be a valuable tool for measuring these attributes in many ecosystems, including tropical, boreal, and mid-latitude forests. However, there have been few studies in montane forests. Here, we examine the ability of a large footprint lidar system to retrieve forest structural attributes in the highly variable terrain and canopy conditions of the Sierra Nevada mountains in California. Specifically, we examined the impact of slope, elevation, aspect, canopy cover, crown shape, and the spatial arrangement of canopy-forming trees on the accuracy of a large footprint lidar system in retrieving canopy height, canopy cover, and biomass. We found good agreement between field and lidar measurements of height, cover, and biomass at the footprint level, and canopy height and biomass at the stand level. Differences between field and lidar measurements are mainly attributable to the spatial configuration of canopy elements and are less sensitive to topography, crown shape, or canopy cover. The accuracy of canopy cover retrieval was highly sensitive to estimates of ground cover reflectivity and to ground sampling density. The accuracy of biomass retrieval was also good, and comparable to previous efforts in other biomes.

Keywords: Lidar; LIDAR; LVIS; Laser altimetry; Canopy; Structure; Height; Cover; Biomass

1. Introduction

Management of forests for multiple uses, such as timber harvesting and protection of biological diversity, is challenging. Effective management often requires either information about the presence and abundance of organisms – which is not available for many species – or the development of indicators of habitat quality that correlate with species distributions. At the landscape scale, the structure of forests can be quantified and used to predict the occurrence of some species. These structural attributes include the height of the forest canopy, the amount of canopy cover, and biomass.

Field measurements of canopy height and canopy cover are conceptually simple. Direct measurements of biomass are somewhat more problematic because they require destructive sampling, although indirect methods, e.g., allometric equations relating dbh and/or height to biomass, suffice for most applications. However, in situ observations of forest structure, particularly in montane settings, are time-consuming and are often limited by accessibility, resulting in relatively small (ca. <1 ha) field studies. Measurement and mapping of these characteristics from field surveys are generally cost prohibitive at fine spatial scales, across large areas.

Remote sensing methods, primarily multispectral (Hyppa et al., 1998) and more recently hyperspectral (Pu & Gong, 2004), have been explored as cost effective means of measuring forest structural characteristics in a spatially and

* Corresponding author.
E-mail address: phyde@geog.umd.edu (P. Hyde).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2005.03.005

temporally continuous manner. However, these techniques are poorly suited for measuring vertical forest canopy structure (Weishampel et al., 2000). Radar methods, such as interferrometric synthetic aperture radar (insar or ifsar) are better at recovering structure in forests, especially those that are structurally simple or have open canopies (Treuhaft & Cloude, 1999; Treuhaft & Siqueira, 2000), but as is the case with multispectral methods, these are primarily by correlation.

In contrast, lidar remote sensing directly measures important vertical and spatial forest structure. Numerous studies using both small footprint (<0.5 m radius) and large footprint, waveform digitization airborne lidar, have demonstrated its ability to recover structure such as canopy height, canopy cover, canopy height profile, canopy volume, biomass, and basal area at unprecedented accuracies (Dubayah & Drake, 2000; Drake et al., 2002a, 2002b; Lefsky, 1997; Lefsky et al., 1999a, 2001, 1999b; Nelson et al., 1984, 1988; Nilsson, 1996). These studies developed relationships between in situ observations of forest structure and airborne laser data. The field samples in these studies were typically small in number, of limited spatial extent, or were located on relatively flat terrain. Because the application of lidar remote sensing to land surface characterization is relatively new, the accuracies achievable under a variety of environmental conditions is not yet well known, especially for large footprint lidar. Factors such as topographic slope, canopy vertical structure and forest spatial structure (such as clumping of trees) are all hypothesized to affect the accuracies of retrieved structures. This is especially true of montane regions where there is a large range of slopes, elevation, soils, and climate, all of which affect species composition and canopy architecture, such as height and cover. Yet it is in these difficult montane conditions where much forest management takes place. Thus, developing a better understanding of the effects of these factors on lidar retrievals is important.

For example, topographic slope can cause the lidar ground return (the last Gaussian return from the surface) to spread, leading to inaccurate ground determination, and consequently, canopy heights. Slopes can also cause heights retrieved from lidar to be shorter or taller than their actual heights if the stem is located away from the center of the footprint. The reason for this is that heights are determined relative to the mean ground elevation within a footprint, so that a stem which is upslope of the footprint center will appear taller and one which is downslope will appear shorter than the actual height, In addition, the returns from short trees on steep slopes can be convolved or blurred with the surrounding topography. The architecture or shape of individual crowns can impact the accuracy of canopy height retrieval; some amount of canopy penetration can occur with more pointed, typically coniferous, crowns leading to underestimates of canopy height. Aspect and elevation relative to the flight of the aircraft can influence sensor-target geometric relationships and thus the actual size of the lidar footprint, leading to uncertainty about the actual area mapped by a footprint. The spatial distribution of canopy materials within the footprint affects the amount of signal returned to the sensor because of the non-uniform (Gaussian) illumination of the footprint by the lidar pulse; i.e. there is less energy at the edge of the footprint which may cause canopy materials at the edge to go undetected or be underestimated.

Developing a thorough understanding of the effects and interactions of these factors will require many studies. However, as part of the Vegetation Canopy Lidar (VCL) mission (Dubayah et al., 1997), airborne, large footprint lidar data were acquired by the LVIS (Laser Vegetation Imaging Sensor) over Sierra National Forest in California in 1999. The goal of this experiment was to provide calibration and validation data to help define algorithms and subsequent accuracies of forest structure retrievals over montane regions under a wide range of slopes, canopy closures, and environmental conditions.

Along with this aspect of our calibration and validation work, we initiated science studies focused on the application of lidar data to map indicators of habitat suitability for California spotted owls for the USDA Forest Service. In the Sierra Nevada, a collaboration was established between members of the VCL Science Team and scientists with the US Forest Service to assess the efficacy of lidar and other remote sensing data to forest structure. This included an extensive field program to map forest structure coincident with lidar footprints in Sierra National Forest in California. These field data could then be used to establish relationships between forest structure and lidar metrics, which then are applied to the entire set of lidar imagery to produce spatially continuous maps of forest structure. These maps ultimately will be used by the Forest Service to assist in habitat and other forest management.

2. Objectives

The primary objective of this study is to assess the ability of a large footprint lidar to retrieve canopy structure over the diverse montane forests of the Sierra Nevada. Our ultimate goal is to provide spatially continuous maps of forest structure at the landscape scale as a prerequisite for forest management. In the research presented here we compare spatially explicit field measurements of structure to metrics derived from lidar data collected by LVIS. Our field plots cover a large range of slope, aspect, elevation, canopy cover, canopy shape, and arrangement, and thus provide a rigorous assessment of lidar retrieval in montane conditions.

The remainder of this paper is organized as follows. First we describe our collection of field plot data, and give details of the LVIS data acquisition over the Sierra Nevada. Next we present our methods for data processing and analysis of both lidar and field data, including the derivation of canopy height, canopy cover, and biomass. We then give the results of statistical comparisons between field-derived and lidar-derived forest structure. Finally, we discuss the significance of our results relative to the retrieval of forests character-

istics in montane regions, and the implications of these for mapping forest structure.

3. Data collection

3.1. Study area

The study area is located in Sierra National Forest in the Sierra Nevada of California. This site is approximately 60,000 ha, with elevation ranging from 853 to 2743 m (for a complete description, see Hunsaker et al. (2001)). Vegetation types include white fir (*Abies concolor*), red fir (*Abies magnifica*), Sierra mixed-conifer, ponderosa pine (*Pinus ponderosa*), and montane hardwood-conifer.

3.2. Lidar data

Lidar data was collected by the Laser Vegetation Imaging Sensor Blair et al. (1999) during October 1999 while deciduous trees were in leaf-on condition. LVIS is an airborne laser altimeter that records the time and amplitude of a laser pulse reflected off target surfaces. LVIS is a full waveform-digitizing system and records the vertical distribution of nadir-intercepted surfaces at 30 cm vertical resolution. The LVIS instrument flew aboard the NASA C-130H aircraft at about 7 km above ground level. LVIS is an imaging lidar, recording spots or "footprints" illuminated within a 7° potential field of view. For the Sierra Nevada flights these footprints had a nominal radius of 12.5 m, nominally separated by 12.5 m across track and continuous along track. Because of variations in altitude of the plane above the varying topography of the Sierra, actual footprint radii vary between 9 and 11 m. An area of about 175 km^2 was mapped.

The fundamental observation of LVIS is a waveform that gives the vertical distribution of nadir and near nadir-intercepted surfaces. The amplitude of the waveform at any height is proportional to the amount of reflective material intercepted at a particular height, the orientation of that material, and its reflectance (ignoring such effects as multiple scattering within the footprint). Initial processing of the data is required to remove various biases to permit accurate geolocation. The data are then further processed to find ground and canopy returns, to derive various waveform metrics, such as the height of median energy and canopy top height, using automated methods.

3.3. Field plot data

One hundred twenty-four (124) plots centered on laser footprints were distributed throughout the study area using a modified stratified random sampling scheme; the data from 13 plots were excluded because no ground return was discernable in the waveform, leaving 112 suitable for analysis. Although the plots were centered on laser footprints, the actual waveforms were not examined before the stratification (to prevent bias), and therefore there was no attempt to retain only waveforms that showed a strong ground return. The number of plots placed within each land cover type were proportional to their actual distribution within Sierra National Forest, with the exception of the red fir class; this vegetation type was oversampled because of its importance as remnant old-growth. Concentric circular plots were established, with an inner plot of 0.07 ha (15 m radius) and an outer plot of 1 ha (56.4 m radius). The 0.07 ha ("footprint") plots were designed to allow direct comparison of field measurements with individual lidar footprints; the plot size was slightly larger than the nominal footprint to compensate for geolocation errors, if needed. The 1 ha ("stand") plots were designed to be commensurate with existing Forest Service field plots.

Field plot data were collected during the summers of 2000 and 2001 and error-checked in 2002. Within the 0.07 ha plots, all live stems with a diameter at breast height (dbh) $\geq$ 10 cm were inventoried and species type was recorded. The dbh of the stem was measured with fiberglass tapes. The height of the stem, the height of the full crown, and the height of the partial crown (if present) were measured with an Impulse LR laser range finder (Laser Technology, Inc., Englewood, CO). The sweep of the partial crown, if present, was estimated to the nearest 30°. The shape of the crown was characterized as elliptical, umbrella-shaped, conical, or cylindrical. Four crown radii (two each along and across slope) were measured with fiberglass tapes. The bearing of each stem with respect to the plot center was measured with a digital flux gate compass (Laser Technology, Inc., Englewood, CO) and the distance of each stem to the plot center was measured with the laser range finder.

Canopy cover was measured every 5 m along four transects and at the plot center with a moosehorn densiometer (Moosehorn CoverScopes, Medford, OR), following an established Forest Service field protocol. Due to time constraints, on only a subset of plots (n = 40), canopy cover was measured every 3 m on twelve 15 m transects (spaced every 30 degrees for a total of 60 observations) and at the plot center with a moosehorn densiometer and a LAI2000 plant canopy analyzer (Li-Cor, Lincoln, NE). This more intensive sampling was performed to help assess the effects of under/over sampling as it is notoriously difficult to achieve consistent field-sample estimates of canopy cover.

Within the 1 ha plots, all live stems $\geq$ 76 cm dbh were inventoried. The species, the height of the stem, and dbh were also recorded.

4. Data analysis

4.1. Lidar data

The focus of this study is the derivation of three structural measurements, canopy height, canopy cover, and above-ground biomass. Canopy height and canopy cover are directly retrieved from waveform data using algorithms described

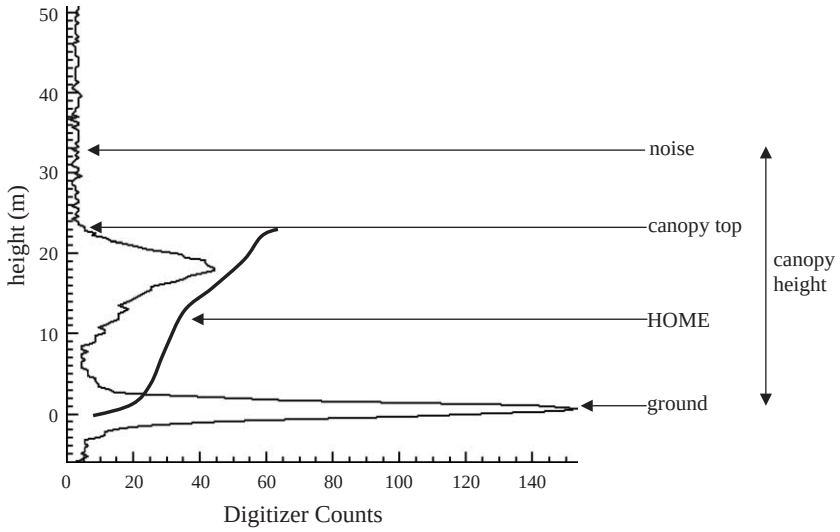

Fig. 1. Diagram of a lidar waveform from LVIS. The top of the canopy is the first return above a noise threshold. The ground is the midpoint of the last return. The bottom of the canopy is the second to last return above a noise threshold. The darker black line is a cumulative energy curve; the height of the median energy return, or HOME, is the height of the point at which half of the total energy in the waveform is above and half is below.

below. Both of these require identification of a ground return in the waveform, and associated with this, the identification of the canopy portion of the waveform. Biomass is not directly measured by LVIS; rather, metrics derived from lidar waveforms, such as canopy height, and height of median energy (HOME), are correlated with canopy structure.

4.1.1. Ground and canopy height

Past studies, e.g. Lefsky (1997) and Lefsky (2000b), have relied on mainly manual methods for finding ground returns, especially where the returns are weak relative to the background noise level. While appropriate for validation studies with small numbers of waveforms, manual methods are impossible for the large numbers of LVIS waveforms used here (about 1 million). Because canopy height is determined relative to the ground, accurately retrieving ground elevation is critical. Thus, an automated algorithm for finding both ground and canopy height was employed.

The algorithm involves: 1) removing noise and resampling the waveform to improve sensitivity and resolution, 2) re-estimating noise statistics, 3) finding the center or mode of the last pulse, i.e., the ground, and 4) detection of the highest surface return above the noise level, i.e., the top of the canopy. The difference between the height of the ground and the top of the canopy is equal to the maximum height of the canopy, often referred to simply as canopy height (Fig. 1). In cases where no ground return could be identified by the algorithm, the data were not used (thus leading to the elimination of some of the field validation waveforms above).

4.1.2. Canopy cover

An algorithm was used to extract canopy cover from LVIS waveforms that uses canopy height as derived from the previous step to define the top of the canopy. The bottom of the canopy is defined as the second to last return above a noise threshold (Fig. 1). Once the ground and canopy portions of the waveform are thus separated, canopy cover is calculated by dividing the canopy portion of the waveform by the total energy in the waveform (canopy return plus ground return).

Differences in reflectivity in the near infra-red portion of the spectrum between canopy and ground must be accounted for. For example, if the canopy is very bright relative to the ground, the amount of energy reflected by the canopy portion will be incorrectly attributed to high closures in the canopy (i.e. the weak reflectance from the ground will be inferred to be because of weak incoming energy rather than weak reflectance). Thus, the ratio of canopy to ground reflectance must be known. Others, e.g., Lefsky et al. (1999a) and Drake et al. (2002b), have assumed a ratio of 2 for a variety of areas. We empirically estimated this ratio through examination of high spatial resolution passive optical (Quickbird, Satellite Imaging Corp., Houston, TX) imagery. Representative samples of forest and non-forested areas were delineated; NIR reflectance was averaged across samples and then ratioed. The resulting correction factor of 1.6 differs significantly from these previous studies.

Table 1
Field plots by vegetation class (described in Mayer, K. E., and W. F. Laudenslayer, Jr. 1988. A Guide to Wildlife Habitats of California. California Department of Forestry and Fire Protection, Sacramento, 166 pp.)

Vegetation class	# Plots
Red fir	36
White fir	19
Ponderosa pine	7
Other pines	4
Sierra mixed-conifer	22
Montane hardwood	2
Montane hardwood-conifer	6
Wet meadow	10
Barren	6

Table 2
Accuracy of predictive models derived through regression analysis

Field measurement	Scale	RMSD (m)	Coefficient of determination (r^2)	Model[a]	n	p
Height (m)	footprint	8.9	0.75	$Y=0.83*LHT+7.85$	112	<0.00
Height (m)	stand	6.4	0.75	$Y=0.59*LHT-0.07$	112	<0.00
Cover (%)	footprint	13.4	0.81	$Y=0.82(CE/(CE+GE))+1.4$	40	<0.00
Biomass (Mg ha^{-1})	footprint	73.5	0.83	$Y=54.1*HOME+5.6*LHT^2-15.1$	112	<0.00
Biomass (Mg ha^{-1})	stand	54.8	0.86	$Y=4.6*$mean LHT $-6.7*$min LHT $+39.2*$mean HOME $-16.5*$median HOME $-41.7*$min HOME $-1.5*$max HOME -45.3	112	<0.00

HOME=height of the median energy return; GE=ground energy; CE=canopy energy.
[a] LHT=LVIS canopy height.

4.1.3. Biomass

While lidar does not measure biomass directly, metrics derived from lidar have proven effective in estimating forest biomass (Drake, 2002b; Lefsky, 1997; Lefsky et al., 1999a, 2001, 1999b; Nelson et al., 1984, 1988; Nilsson, 1996). Canopy height by itself is sufficient in some, more structurally simple, biomes (Lefsky et al., 1999b). In more structurally complex biomes, such as tropical and old-growth Western coniferous forests, some indication of the depth of the canopy is also useful for predicting biomass (Lefsky et al., 1999a; Drake et al., 2001). The metrics used in this study include canopy height, canopy height squared, canopy reflectance, and height of the median energy return (HOME; Fig. 1). These metrics were calculated from waveforms at both the plot and stand levels.

4.2. Field data

Field measurements at the footprint and stand levels include maximum canopy height, i.e., the height of the tallest stem within the plot. Field height data were also pooled according to vegetation type (Table 1), which was used as a proxy for crown shape, e.g., stands of pure red fir tended to be conical or pointed, while deciduous crowns tended to be more rounded. Elevation for each plot was acquired from a 7.5 in. USGS Digital Elevation Model, which was also used to calculate slope. Allometric equations relating stem biomass to height and dbh were obtained from the USDA Forest Service (Waddell & Hiserote, 2003). These equations were applied to the field data to calculate total standing (aboveground) biomass for each stem; the biomass of all stems within the plot was added to provide totals at the footprint and stand levels.

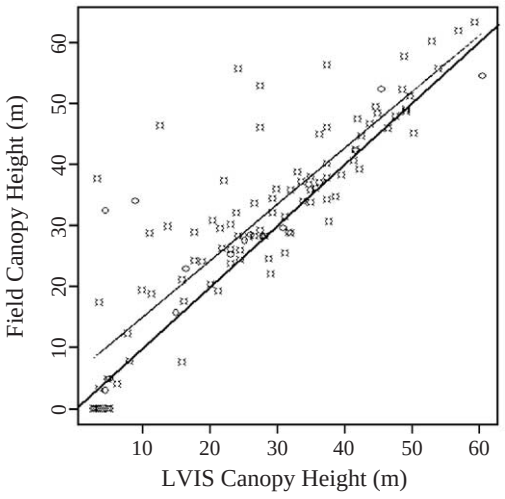

Fig. 2. Scatterplot of LVIS lidar height vs. field canopy heights. There are large underestimates of tree heights in about 14 plots. We hypothesize that these are caused either by lidar footprint sizes that are smaller than expected relative to field plots, or by a lack of sensitivity to trees that occur at the very edges of the lidar footprint (caused by the gaussian drop in incident energy from center to edge).

5. Results

In this section we compare the results of our lidar retrievals with field data. Lidar and field canopy height and

Table 3
Potential sources of error in lidar vs. field canopy height measurements at the footprint level

Potential error source	Coefficient of determination (r^2)	S.E.	n	p
Slope	0.19	11.8 m	112	0.25
Aspect	0.01	13.1 m	112	0.08
Elevation	0.00	13.1 m	112	0.75
Cover	0.00	5.9 %	40	0.08

Residuals from the height regression were regressed against each of the variables given in this table. Note that none of the variables are statistically significant, and thus do not explain any of the residual error in the linear relationship (see Table 2).

Table 4
Accuracy of canopy height measurement (at the footprint scale) as a function of land cover class

Vegetation class	RMSD (m)	Coefficient of determination	n	p
Red fir	8.8	0.72	36	0.00
White fir	8.9	0.60	19	0.00
Ponderosa and Jeffrey pine	11.1	0.43	11	0.03
Sierra mixed-conifer	10.0	0.35	22	0.00
Montane hardwood and Montane hardwood-conifer	4.0	0.88	8	0.00
Wet meadow	9.5	0.54	10	0.02
Barren	3.6	0.85	6	0.00

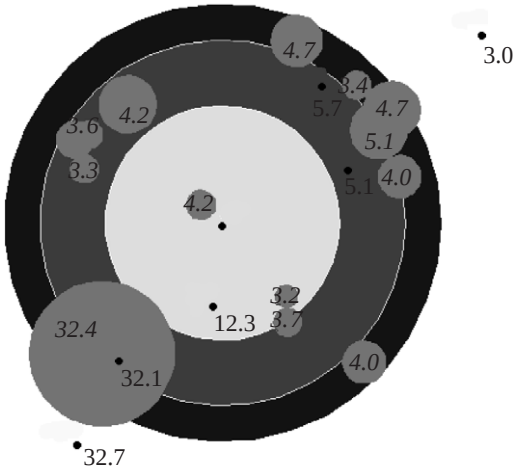

Fig. 3. Example of a large underestimate of tree height caused by edge location of the tallest tree. The radii of the concentric circles are 9, 12.5, and 15 m in length. The very small black circles are the centers of lidar footprints; the text beneath them is the canopy height value at that location. The small gray circles are stems; the radius of the circle is the average of four crown radii measurements taken in the field. The italicized text inside the stem locations is the field measurement of tree height. In this plot, LVIS records the height of the canopy at 4.5 m. The calculated footprint size for this plot is 10.5 m. The large (32.4 m) stem in the lower left sector of the plot is not being recorded by LVIS; the effective footprint size for this observation is closer to 9 m. Note that a second LVIS observation was located directly on this large stem and the measurements are in close agreement.

canopy cover were compared directly using root mean square difference (RMSD) statistics, and through regression analysis at the footprint and stand levels using linear regression techniques (with results summarized in Table 2). The residuals of the regression between field and lidar height measurements were related to slope, aspect, and cover to quantify the effects of these factors on height retrieval. In addition, the effects of crown shape on height estimates were also explored. Lastly, field estimates of biomass were compared to lidar metrics using stepwise multiple linear regression techniques.

5.1. Canopy height

Field and lidar canopy heights showed good agreement (r^2=0.75, RMSD=8.2 m, n=112, p<0.00; Fig. 2) at the footprint level, comparable to other results using large footprint lidar e.g., Lefsky et al. (1999b). Bare plots were included in the comparison. Note that for some plots, field estimates are much larger than lidar estimates, by tens of meters, which weakens the relationship (analyzed further below).

Contrary to expectations, examination of the residuals of the regression showed a weak, and statistically insignificant correlation with slope (r^2=0.19, S.E.=11.8 m, n=112, p=0.25; Table 3). There was no systematic relationship between differences in field vs. lidar measurements and aspect (r^2=0.01, S.E.=13.1 m, n=112, p=0.08), elevation (r^2=0.00, S.E.=13.1 m, n=112, p=0.75), or canopy cover (r^2=0.01, RMSD=5.9 m, n=40, p=0.08) (Table 3).

Plots were grouped together by land cover type (vegetation class) to serve as a rough proxy and distinguisher for crown shape. Field measurements of canopy height from the two montane hardwood plots were pooled with measurements from montane hardwood-conifers, the only other land cover class that contained some deciduous trees. Separate regression models were then fit for each class to assess the effects of crown shape. There were some differences among vegetation classes (Table 4). Field and lidar canopy heights were in closest agreement in the montane hardwood-conifer (r^2=0.88, RMSD=5.4 m, n=6, p<0.00) and combined montane hardwood and montane hardwood-conifer plots (r^2=0.88, RMSD=4.5 m, n=8, p<0.00) and lowest in the Sierra mixed-conifer class (r^2=0.35, RMSD=10.0 m, n=22, p=0.0).

The above analyses were performed assuming a nominal footprint size of 12.5 m. Although efforts were made by the aircraft to follow the terrain, variations in footprint size because of changes in elevation were inevitable. We modeled the expected footprint size for each footprint based on the sensor/target geometry which yielded a range of

Table 5
Canopy height retrieval accuracy as a function of distance

i	n	Root mean-square error	RMSD	Coefficient of determination	S.D.	Bias (field−lidar)	Slope coefficient of determination[a]
3	5	3.74	4.01	0.97	4.33	−0.26	0.18
4	9	3.99	3.87	0.94	4.01	0.58	0.32
5	15	3.44	3.45	0.96	3.52	0.51	0.19
6	24	4.14	4.21	0.95	4.1	1.27	0
7	32	4	4.07	0.94	3.95	1.18	0.01
8	45	4.88	5.26	0.9	4.85	2.16	0.01
9	55	6.38	6.75	0.82	6.38	2.36	0
10	64	7.09	7.69	0.77	7.09	3.1	0
11	77	7.52	8.45	0.74	7.61	3.78	0.02
12	91	7.37	8.43	0.73	7.5	3.93	0
13	101	7.87	9.33	0.69	8.19	4.55	0.01

[a] None of the slope r^2 values are significant above 0.10 level.

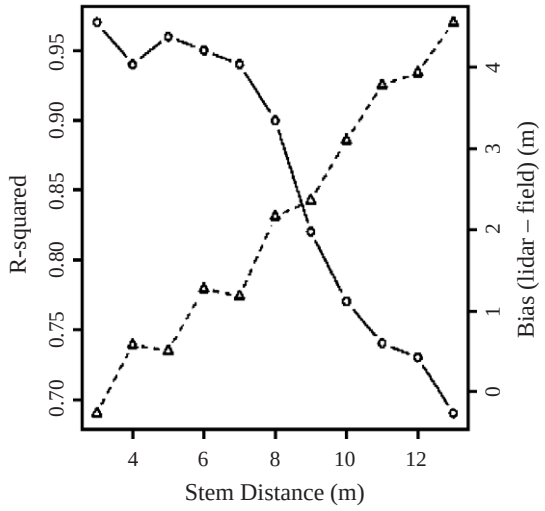

Fig. 4. Canopy height retrieval accuracy and bias as a function of distance. The open circles represent the r^2 values, while the open triangles represent the bias. Accuracy remains high when the tallest stem is within about 9 m, then drops significantly. Bias shows a nearly linear trend with stem distance from the center of the plot.

footprint radii from 9 to 11 m. If the largest stem in field data was beyond 11 m, this would lead to some of the large errors noted above (because the stem would not be in the imaged footprint). For example, a stem map of Plot 18 (Fig. 3) shows the tallest tree occurs at a distance of 12.1 m, whereas the calculated footprint size is 10.5 m. We then compared lidar height with the tallest stem within the calculated footprint radius. The result was an r^2 value of 0.72, RMSD of 7.6 and did not significantly improve the results. However, when separate regressions were performed as a function of footprint size, r^2 values increased from 0.76 for 11 m radius footprint up to 0.88 for 9 m radius. One reason for this increase is that the larger footprint sizes coincidentally include those plots where the largest tree is near the edge of the footprint. This then suggests an issue with sensitivity.

As discussed earlier, the incident energy pattern within an LVIS footprint is not uniform, but rather drops off as a Gaussian, so that errors may increase as the tallest stem moves from the center of footprint to the edges, especially for pointed crowns. We recalculated our height comparisons, this time by grouping plots together as a function tallest stem distance, from 3 to 12.5 m, and assuming a nominal footprint size of 12.5 m. We also excluded barren plots to isolate the effect stem distance. Our results are shown in Table 5 and Fig. 4. The location of the tallest stem relative to the center of the footprint makes a dramatic difference in the results. If the stem is within 8 m of the center, our r^2 values are above 0.90, with RMSD below 5 m. After about 9 m, our r^2 values drop quickly to a low of 0.69. The bias is always positive (field−height) and increases near linearly with distance of the stem. This suggests that LVIS underestimates tree heights as the crown moves to the edge of the footprint (perhaps because of energy drop off).

Given this large effect on the height accuracy, we reassessed the effects of slope by looking for correlation of the height residuals with slope within distance classes (shown in Table 5) (and thus isolating the effect of stem distance). As before, there were no significant correlations of residual error with slope, i.e. none of the remaining variability could be explained by slope, regardless of distance class.

Field and lidar canopy heights also were in good agreement (r^2=0.75, RMSD=6.0 m, n=122, p<0.00) at the stand level. In all cases, there are across-track gaps between footprints due to pitch and roll of the aircraft. Therefore none of the 1 ha plots were completely mapped, although some plots were more thoroughly mapped than others. Differences in field versus lidar canopy height were minimized when >40% of the plot area was mapped (about 35 lidar observations per ha; Fig. 5).

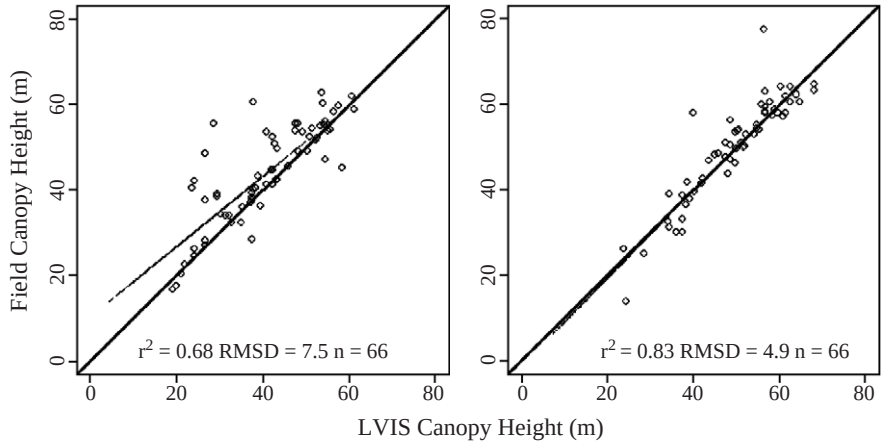

Fig. 5. Field vs. lidar canopy height measurements at the stand level. The panel on the left shows the relationship between field and lidar measured canopy heights where the number of footprints per hectare is less than or equal to 35 (approximately 40% of the 1 ha plot area), while the panel on the right is where the number of shots per hectare is above 35. Solid lines are the 1:1 lines, while dashed line is regression line. In both cases, p<0.00. As can be seen, a more complete imaging of the stand leads to much better height accuracies.

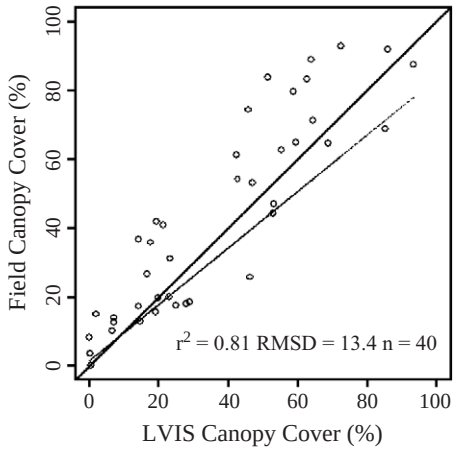

Fig. 6. Field vs. lidar measurements of canopy cover at the footprint level. ($p<0.00$) for the 40 intensively measured plots. Solid and dashed lines are the 1:1 and regression lines, respectively.

5.2. Canopy cover

The agreement between field and lidar measurements of canopy cover was only fair ($r^2=0.54$, RMSD=19.6%, $p<0.00$) for those plots ($n=112$) where the limited sampling protocol was used. In contrast, at the 40 plots that were more intensively field sampled, field and lidar estimates were in good agreement ($r^2=0.81$, RMSD=9.4%, $n=40$, $p<0.00$; Fig. 6). The best results occurred when the moose horn observations were limited to those within 9 m of the plot center. To highlight the importance of using the correct ground to canopy reflectance, we compared results using a ratio of 2 for these 40 plots. In this case, the strength of the relationship decreased significantly ($r^2=0.49$, RMSD=22.0%, $n=40$, $p<0.00$). We do not expect that height or distance would influence canopy cover retrieval accuracy and the residuals of the regression were not related to either of these factors.

5.3. Biomass

Lidar derived biomass agreed well with field-estimated biomass at the footprint ($r^2=0.83$, RMSD=73.5 Mg ha^{-1}, $p<0.00$) and stand levels ($r^2=0.86$, RMSD=54.8 Mg ha^{-1}; $p<0.00$; Fig. 7). Laser height squared and HOME were significant predictors of biomass at the footprint level, while mean and minimum laser height, mean, minimum, median, and maximum HOME were significant predictors of biomass at the stand level.

6. Discussion

Our results suggest that for our study site, the accuracy of canopy height retrieval at the footprint level was not clearly related to topography (slope, aspect, and elevation) or canopy cover, but rather was strongly influenced by the spatial arrangement (more precisely the distance from footprint center) of the largest trees. We hypothesize that this could be the result of two factors: footprint size errors (i.e. assuming the footprint to be 12.5 m in radius when it is actually smaller), or; the Gaussian drop off in power across the footprint, resulting in a lack of sensitivity to canopy material progressively towards the edges of the footprint.

Our results seem to suggest that it is the latter which is more important, because there is little improvement in variance explained or reduction in RMSD by using the exact footprint size vs. a nominal footprint size of 12.5 m. However, our results are somewhat confounded by the fact that the plots with the largest errors are the ones that have their tallest stem at the edge of the footprint, and these coincidentally are the largest radius footprints. Nonetheless, the dramatic increase in accuracy as the tallest stem moves towards the center of the plot strongly suggests that there is a drop in sensitivity as a result of the drop in incident power near the edge of the footprint. This effect should be most

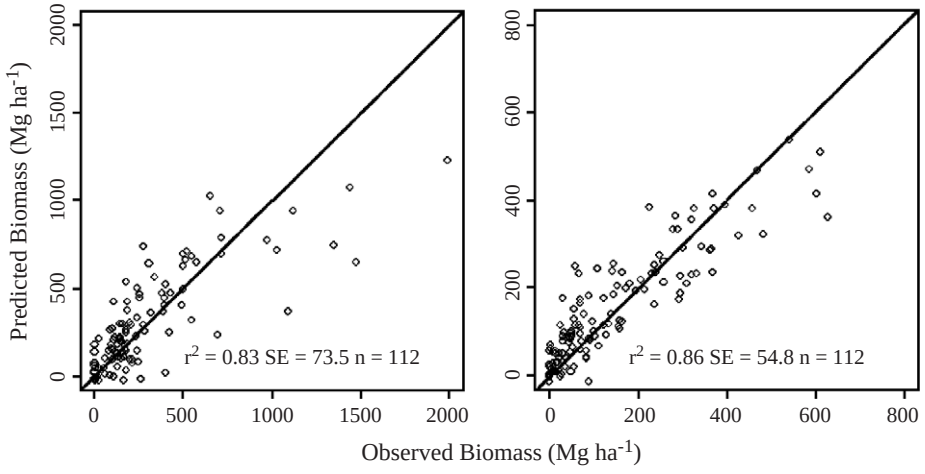

Fig. 7. Results of step-wise multiple linear regression models between lidar metrics and field-measured biomass. The panel on the left is the result of analysis at the footprint level, while the panel on the right s for the stand level. In both cases, $p<0.00$.

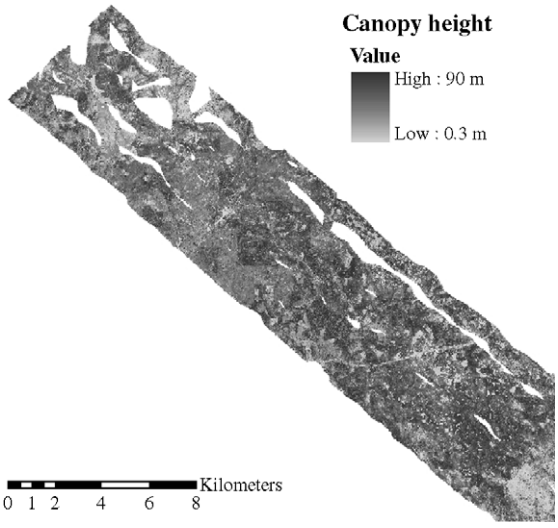

Fig. 8. Lidar-derived canopy height in Sierra National Forest. Abrupt discontinuities in canopy height are visible along some boundaries, which are privately owned "inholdings" within Sierra National Forest. This suggests that different forest management practices produce visible differences in forest structure, some of which are likely important to wildlife. Gaps in coverage are caused by variations in flight lines. North is to the top.

pronounced with crowns that are pointed, where the increase in intercepted leaf area is slow as a function of distance from the top of the stem. The linear increase in positive bias with stem distance also supports this contention. Finding more general conclusions about the interaction of footprint size, footprint energy pattern, crown shape, and stem location will require further studies. For example, we achieved very good results when the actual footprint radius was 9 m, suggesting perhaps that a smaller footprint size is appropriate for coniferous forests that show large variations in stem height over short distances.

Based on previous research (Rocchio, 2000; Hofton et al., 2002), we expected slope to be the major source of error. The maximum expected height error for a typical Sierra location is estimated with the following equation

$$\tan S^{*}D/2, \text{ where } S \text{ is the slope and } D$$

is the footprint diameter.

This gives a predicted error of about 3–4 m, assuming the average slope of our study site, 18%, and footprint radius range of 9–12.5 m. Slope effects should show up not as biases, as it is equally like the tallest stem would be above the center of the footprint or below, but rather as an added scatter to the relationship (i.e. increased RMSD). We observed no relationship of our residual errors with slope (or the tangent of slope). We are unsure why this is so. The lowest RMSD observed was 3.45 m (considering only stems less than 5 m from center in Table 5). The magnitude of this error is about right for slope induced error, but again residuals from our linear regression are uncorrelated with slope. One possible explanation is that there are a set of factors such as uneven energy distribution across the slope, spatial arrangement effects, etc that are masking slope effects or interacting with slope in a non-linear fashion.

Crown shape, using land cover as a proxy, would seem to have an influence on height retrieval accuracy: the more rounded crowns of deciduous trees appear as a discernable feature in the waveform more quickly than more pointed coniferous crowns. However, the small number of plots containing deciduous trees limits our confidence in making generalizations about crown shape. Differences in canopy height measurements are likely to be a function of several convolved factors: crown shape, canopy cover or crown density, and slope, as well as the spatial arrangement of the canopy with respect to the laser footprint.

Our retrievals of canopy cover were encouraging; however our experience in the Sierra Nevada suggests a few cautionary notes. First, when following a standard USFS field sampling protocol, results were poor. It was only at the intensively sampled plots where there was strong agreement between lidar and field estimates. Secondly, lidar underestimated canopy cover relative to field estimates. We would have expected the opposite given the slightly off-nadir collection angle of many LVIS footprints (caused by scanning). This is probably an artifact of the collection process on the ground; human observers are not as likely to observe the very small gaps within canopies, as they are the larger gaps that occur between canopies. These two together highlight the difficulties in ground-truthing canopy cover and the great care that must be taken in designing and implementing a field protocol. Lastly, note the importance of knowing the correct ratio of ground to canopy reflectance. It is not clear to us how stationary (in a spatial

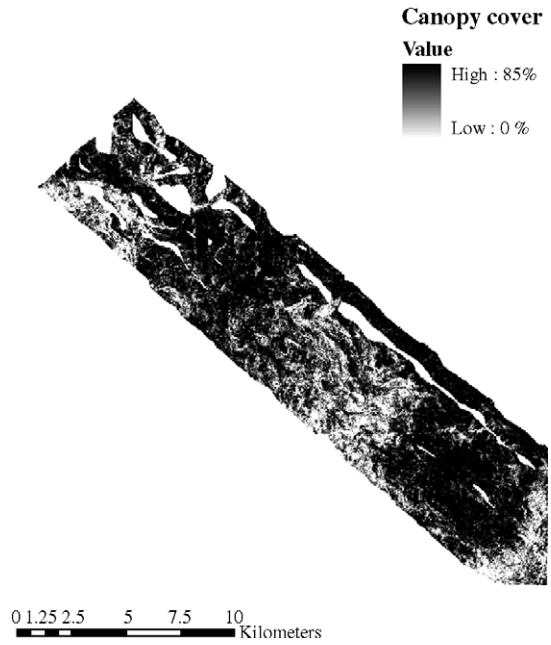

Fig. 9. Lidar-derived canopy cover in Sierra National Forest. Some footprints within very dense stands are identified as "no data" (shown here as 0%) because there was not enough energy in the waveform to penetrate the canopy and provide a reliable ground return.

statistics sense) this ratio is, and therefore we cannot say how often it needs to be re-estimated. The requirement of needing a priori or ancillary data to correctly retrieve canopy cover pushes its estimation somewhat farther away from a "direct" retrieval. It may be possible to estimate this ratio directly from the lidar data if there is some confidence that the outgoing pulse energies leaving the system are about the same, or if different, that they are recorded, as should also be the case with systems that employ a variable gain (to avoid saturation in the recorded return pulse). In theory then, the total NIR energy in vegetated and unvegetated waveforms could be used to estimate the ratio.

Biomass again proved to be reliably estimated by lidar. Forest structural characteristics, such as canopy height, canopy cover, and biomass are correlated with age and site conditions, but the relationship is curvilinear, not linear (Aber, 1979). There are tall stands with high biomass and tall stands with moderate biomass. Distinguishing between the two requires other information besides height. In the tropics, height of the median energy (mean HOME at the stand level) return does just this (Drake et al., 2001). However, in the less dense, coniferous canopies of the Sierras, stand level mean HOME was not as useful because of strong ground returns that skewed HOME towards the ground. Minimum height and minimum HOME helped distinguish stands that contained some shorter-stature trees or bare patches from stands with uniformly tall trees.

Recall that the ultimate goal of our research is to provide spatially continuous maps of forest structure at the landscape scale as a prerequisite for habitat suitability studies.

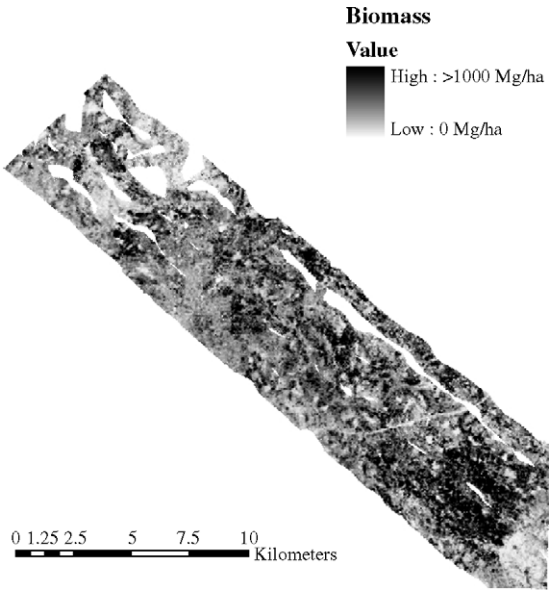

Fig. 10. Biomass in Sierra National Forest. A predictive model of biomass was created through regression analysis, relating field measured stand characteristics to lidar metrics. The patches of high biomass in the southeastern portion of the image are old-growth stands of red fir in and around the Teakettle Experimental Area. Other patches of very high biomass in the image are giant sequoia groves. Most of the patches of very low biomass are rock outcrops.

Using the direct retrieval of canopy height and cover, and the statistical relationship developed between field and lidar data for biomass, we created maps of each of these over the domain of our lidar data (Figs. 8–10). The accuracies observed using our extensive set of field plots gives us some confidence in the accuracy of the final derived map products. It is this mapping of forest structural characteristics at the landscape scale which we believe will be so useful for future habitat studies.

7. Conclusions

Our results have shown the ability of lidar to characterize montane forest canopy structure – canopy height, canopy cover, and biomass – over the wide range of environmental conditions that occur over Sierra National forest, as a prerequisite for large area habitat mapping. From a lidar remote sensing perspective, the Sierra Nevada is difficult terrain: it contains steep slopes, highly variable cover, highly variable backgrounds and a mixture of natural (fire) and anthropogenic (selective logging) disturbance regimes. Thus, our success here is encouraging.

Lidar cannot provide all the indicators required for habitat mapping and indeed, canopy height, canopy cover, and biomass may not necessarily even be the best ones. However, they are nonetheless important metrics that have been exceptionally difficult to estimate over large areas using other field or remote sensing methods. Our ability to now map these provides an opportunity to test and assess their ecological importance in new ways.

There is additional structural information present in the lidar waveform that can be exploited for habitat characterization. For example, ecological theory suggests that the vertical distribution of canopy structural elements is the most important structural attribute for many forest-dwelling species, particularly avifauna. Such vertical distributions, though not studied here, are straightforward to derive from lidar (e.g. canopy height (leaf and branch) profiles), and should prove useful.

Lastly, developing strategies for fusion of lidar with other remotely sensed data will be important for habitat mapping. Multispectral, hyperspectral, multi-angle, and radar methods all provide data that should be highly complementary with the structural information derivable from lidar. For example species-level information might be derived from hyperspectral observations. The creation of such fusion methods will require further research, but could lead to even more powerful approaches to habitat mapping.

Acknowledgements

The authors thank Michelle West, Steve Wilcox, Brian Boroski, Meghan Salmon, Ryan Wilson, Aviva Pearlman, Sharon Pronchik, John Williams, Brian Emmett, and Josh

Rhoads for assistance with field data collection. We also thank Craig Dobson, Leland Pierce, Malcolm North, and Jo Ann Fites-Kaufman for assistance developing the field protocol.

References

Aber, J. D. (1979). Foliage–height profiles and succession in Northern hardwood forests. *Ecology*, *60*, 18–23.

Blair, J. B., Rabine, D. L., & Hofton, M. A. (1999). The laser vegetation imaging sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, *54*, 115–122.

Drake, J., Dubayah, R., Clark, D., Knox, R., Blair, J. B., Hofton, M., et al. (2002a). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, *79*(2–3), 305–319.

Drake, J., Dubayah, R., Knox, R., Clark, D., & Blair, J. B. (2002b). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, *81*(2–3), 378–392.

Dubayah, R., Blair, J. B., Bufton, J. L., Clark, D. B., JaJa, J., Knox, R. G., et al. (1997). Land satellite information for the next decade II. *American Society for Photogrammetry and Remote Sensing*, 100–112.

Dubayah, R., & Drake, J. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, *98*(6), 44–46.

Hofton, M., Rocchio, L., Blair, J. B., & Dubayah, R. (2002). Validation of vegetation canopy lidar sub-canopy topography measurements for a dense tropical forest. *Journal of Geodynamics*, *34*(3–4), 491–502.

Hunsaker, C., Boroski, B., & Steger, G. (2001). Relations between canopy cover and occurrence and productivity of California spotted owls. *Predicting Species Occurences, Issues of Accuracy and Scale*. Covelo, CA: Island Press.

Hyppa, J., Huppa, H., Inkinen, M., & Engdahl, M. (1998). Verification of the potential of various remote sensing devices for forest inventory. *Proceedings of IEEE Geosciences and Remote Sensing Society* (pp. 1812–1814). Pasadena, CA: California Institute of Electrical and Electronics Engineers.

Lefsky, M. A. (1997). *Application of lidar remote sensing to the estimation of forest canopy and stand structure*. Dissertation Thesis, University of Virginia.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, *70*, 339–361.

Lefsky, M. A., Cohen, W. B., & Spies, T. A. (2001). An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in Western Oregon. *Canadian Journal of Forest Research*, *31*, 78–87.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83–98.

Nelson, R., Krabill, W., & Maclean, G. (1984). Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment*, *15*, 201–212.

Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, *24*, 247–267.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, *56*, 1–7.

Pu, R. L. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, *91*, 212–224.

Rocchio, L. (2000). *Lidar remote sensing of sub-canopy topography*. Thesis, University of Maryland, College Park, 101 pp.

Treuhaft, R. N., & Cloude, S. R. (1999). The structure of oriented vegetation from polarimetric interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(5), 2620–2624.

Treuhaft, R. N., & Siqueira, P. R. (2000). Vertical structure of vegetated land surfaces from interferrometric and polarimetric radar. *Radio Science*, *35*(1), 141–177.

Waddell, K., & Hiserote, B. (2003). Technical documentation for the integrated database, version 1.0. *USDA Forest Service Pacific Northwest Research Station*.

Weishampel, J. F., Blair, J. B., Knox, R. G., Dubayah, R., & Clark, D. B. (2000). Volumetric lidar return patterns from an old-growth tropical rainforest canopy. *International Journal of Remote Sensing*, *21*(2), 409–415.