Estimation of tropical forest structural characteristics using large-footprint lidar

Jason B. Drake a,*, Ralph O. Dubayah a, David B. Clark b,c, Robert G. Knox d, J. Bryan Blair d, Michelle A. Hofton a, Robin L. Chazdon e, John F. Weishampel f, Stephen D. Prince a

a Department of Geography, University of Maryland, College Park, MD 20742, USA
b Department of Biology, University of Missouri-St. Louis, St. Louis, MO 63121, USA
c La Selva Biological Station, Costa Rica
d NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
e Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
f Department of Biology, University of Central Florida, Orlando, FL 32816, USA

Received 19 April 2000; received in revised form; accepted 31 August 2000

Abstract

Quantification of forest structure is important for developing a better understanding of how forest ecosystems function. Additionally, estimation of forest structural attributes, such as aboveground biomass (AGBM), is an important step in identifying the amount of carbon in terrestrial vegetation pools and is central to global carbon cycle studies. Although current remote sensing techniques recover such tropical forest structure poorly, new large-footprint lidar instruments show great promise. As part of a prelaunch validation plan for the Vegetation Canopy Lidar (VCL) mission, the Laser Vegetation Imaging Sensor (LVIS), a large-footprint airborne scanning lidar, was flown over the La Selva Biological Station, a tropical wet forest site in Costa Rica. The primary objective of this study was to test the ability of large-footprint lidar instruments to recover forest structural characteristics across a spectrum of land cover types from pasture to secondary and primary tropical forests. LVIS metrics were able to predict field-derived quadratic mean stem diameter (QMSD), basal area, and AGBM with $R^2$ values of up to .93, .72, and .93, respectively. These relationships were significant and nonasymptotic through the entire range of conditions sampled at the La Selva. Our results confirm the ability of large-footprint lidar instruments to estimate important structural attributes, including biomass in dense tropical forests, and when taken along with similar results from studies in temperate forests, strongly validate the VCL mission framework. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Aboveground biomass; Carbon; Forest structure; Tropical forests; Lidar remote sensing; Laser altimeter; Costa Rica

1. Introduction

Tropical forests are among the most structurally complex and carbon-rich ecosystems in the world. This complexity is related both to the size–frequency distribution of woody stems (Clark & Clark, 2000; Denslow & Hartshorn, 1994) and to the three-dimensional arrangement of canopy elements (e.g., leaves, branches, trunks) from the top of the canopy to the ground (Richards, 1996). Variation in tropical forest structure is influenced by underlying environmental conditions (Clark & Clark, 2000; Clark, Clark, & Read, 1998; Laurance et al., 1999), and in turn creates microclimatic (e.g., light, temperature, humidity) gradients (Parker, 1995). These fine-scale gradients modify biological processes such as competition and growth (Clark, Clark, Rich, Weiss, & Oberbauer, 1996; Nicotra, Chazdon, & Iriarte, 1999; Oberbauer, Clark, Clark, Rich, & Vega, 1993; Rich, Clark, Clark, & Oberbauer, 1993), which further modifies organization of forest structural components.

Quantification of forest canopy structure provides information about the primary surfaces of energy and matter exchange between the atmosphere and one of the largest reserves of terrestrial aboveground carbon (Dixon et al., 1994; Perry, 1994). Knowledge of the total carbon content in tropical vegetation provides a critical initial condition for studies at multiple scales that examine carbon flux caused by natural (e.g., landscape-level respiration of woody plants) and anthropogenic (e.g., deforestation and afforestation) pro-
cesses. However, the accurate estimation of structural characteristics (e.g., aboveground biomass—AGBM—or total mass of aboveground living tissues) of tropical forest vegetation remains a major obstacle (Dubayah et al., 1997).

Most studies use forest AGBM, which is approximately 50% carbon, as a surrogate for total aboveground carbon. Because biomass can only be directly measured through destructive sampling, it is usually estimated through a relationship with other measurable properties. Field studies typically use allometric relationships between total biomass and the height or bole diameter of trees. Although field biomass estimation methods are useful for local-scale studies, remote sensing techniques are necessary for the recovery of biomass over broader spatial scales.

Most remote sensing studies illustrate the empirical correlation between forest biomass and the intensity of EM energy (or the ratio of energy at different wavelengths) that is received by the instrument. Unfortunately, many types of remote sensing instruments suffer from the same problem: they are sensitive to changes in biomass in relatively young and/or homogeneous forests, but in older or heterogeneous forests the signal becomes less predictable with respect to changes in biomass.

Several studies have shown that passive optical instruments are insensitive to changes in tropical forest structural characteristics such as AGBM beyond secondary forests of 10–15 years (Curran, Foody, Lucas, Honzak, & Grace, 1997; Foody & Curran, 1994; Moran, Brondizio, Mausel, & Wu, 1994; Sader, Waide, Lawrence, & Joyce, 1989; Steininger, 1996). Metrics from synthetic aperture radar (SAR), such as backscatter, also tend to saturate in dense forest conditions (Imhoff, 1995; Kasischke, Melaart, & Dobson, 1997; Waring et al., 1995; though see also Franson, Walter, & Ulander, 2000; Smith & Ulander, 2000) and have been shown to be insensitive to changes in AGBM for secondary tropical forests with AGBM levels >60 Mg/ha (Luckman, Baker, Kuplich, Yanase, & Frery, 1997).

New large-footprint lidar instruments (Blair, Coyle, Bufton, & Harding, 1994; Blair, Rabine, & Hofton, 1999) may be able to overcome the saturation problems of other remote sensing instruments. These instruments estimate canopy height as well as other parameters related to the vertical arrangement of canopy elements from the top of the canopy to the ground by directly measuring vertical structure (Blair et al., 1999; Lefsky, Harding, Cohen, Parker, & Shugart, 1999b; Weishampel, Ranson, & Harding, 1996). However, this technology has not been applied to tropical forests. The objective of this study was to test the ability of a new airborne large-footprint mapping laser altimeter instrument to accurately estimate tropical forest structural characteristics such as AGBM and basal area in a dense, wet tropical forest. Related large-footprint laser altimeter technology will soon be incorporated in the spaceborne Vegetation Canopy Lidar (VCL) mission (Dubayah et al., 1997; Dubayah, Knox, Hofton, Blair, & Drake, 2000).

In this paper, we first provide a brief background on lidar remote sensing, including previous lidar studies of forest structure, and highlighting some differences between existing systems. Next, we describe our study site and the new lidar instrument used in this study. Finally, we present the results from this lidar instrument and discuss these results in relation to previous remote sensing efforts to estimate tropical forest structure.

2. Lidar remote sensing

Lidar (light detection and ranging) is an active remote sensing technique using laser light. Lidar systems measure the round-trip time for a pulse of laser energy (usually with a near-infrared wavelength for vegetation studies) to travel between the sensor and the target. This incident pulse of energy interacts with canopy (e.g., leaves and branches) and ground surfaces and is reflected back to the instrument. The travel time of the pulse, from initiation until it returns to sensor, is measured and provides a distance or range from the instrument to the object (hence, the common use of the term “laser altimetry,” which is now generally synonymous with lidar).

Current lidar systems for terrestrial applications differ in: (1) whether they record the range to the first return, last return, multiple returns, or fully digitize the return signal; (2) footprint size (from a few centimeters to tens of meters); and (3) sampling rate/scanning pattern. Most commercial airborne lidar systems are low-flying, small-footprint (5–30 cm diameter), high pulse rate systems (1000–10,000 Hz). In addition, most commercial lidar systems record the range to the highest, and/or lowest, reflecting surface within the footprint, and are not fully imaging, using instead many laser returns in close proximity to each other to recreate a surface.

Small-footprint lidar sensors may not be optimal for mapping forest structure for several reasons. First, small-diameter beams frequently miss the tops of trees (see Nelson, 1997). Secondly, because of their small beam size, mapping large areas requires extensive flying. Finally, with systems that only record first and/or last returns, it is difficult to determine if a particular shot has penetrated the canopy all the way to ground. In areas of high canopy, only one in several thousand returns may be from the ground (Blair & Hofton, 1999). If this topography cannot be recovered, accurate height determination is impossible because canopy height is measured relative to the ground.

Large-footprint lidar systems (Blair et al., 1994, 1999) have several advantages that help avoid these problems. First, by increasing the footprint size to the approximate crown diameter of a canopy-forming tree (~10–25 m), laser energy consistently reaches the ground even in dense forests (Weishampel et al., 1996). Second, the larger footprint size also avoids the biases of small-footprint systems that frequently miss the tops of trees. Third, large-footprint
systems fly at higher altitudes and enable a wide image swath, which reduces the expense of mapping large areas on the ground (Blair et al., 1999). Finally, large-footprint lidar systems also digitize the entire return signal (or “waveform”), thus, providing data on the vertical distribution of intercepted surfaces (i.e., canopy and ground elements) from the top of the canopy to the ground.

The conceptual basis for a large-footprint lidar return is illustrated in Fig. 1. The time history of the reflected energy is fully digitized and is converted to units of distance (accounting for the speed of light through the atmosphere). The first energy return above a threshold is used to derive the distance to the canopy top and the midpoint of the last energy return is used to find the range to ground, the subtraction of which yields laser-derived canopy height. The return waveform gives a record of the vertical distribution of nadir-intercepted surfaces (i.e., leaves and branches). At any particular height, the amplitude of the waveform measures the strength of the return. Thus, for surfaces with similar reflectances and geometry within a footprint (and under similar atmospheric conditions), a larger amplitude indicates more canopy material and a smaller amplitude less. The waveform provides only an apparent canopy profile (leaves and branches) because of attenuation of the beam through the canopy and must be adjusted to approximate the true canopy profile (Lefsky et al., 1999b; Ni-Meister, Jupp, & Dubayah, in press).

Fig. 1. Conceptual basis of lidar remote sensing. Incident Gaussian-distributed pulses of laser energy from airborne or spaceborne instruments reflect off various portions of the canopy, resulting in a return waveform where the amplitude of the pulse is a function of the area of reflecting surfaces (leaves and branches) at that height. The entire waveform gives the vertical distribution of surfaces intercepted by the incident beam. Some of the incident light penetrates all the way through the canopy to produce the last large-amplitude Gaussian-shaped spike in the waveform known as the ground return. Lidar systems do not measure canopy height, but rather a target range determined by measuring the travel time of the pulse (accounting for the speed of light through the atmosphere). Canopy height is determined by subtracting the range to the ground from that to the first detectable return or some threshold above that return. The LVIS, a large-footprint airborne scanning lidar instrument that was used in this study, is illustrated in the top portion of the figure.
2.1. Previous lidar studies of forest structure

Measurements from small-footprint laser altimeter instruments have been useful in estimating tree heights (e.g., Magnussen & Boudewyn, 1998; Naesset, 1997; Nelson, Swift, & Krabill, 1988; Nilsson, 1996), percent canopy cover (Weltz, Ritchie, & Fox, 1994), timber volume (Naesset, 1997) and, in some cases, forest AGBM (Nelson et al., 1988). However, these fine-resolution sensors typically yield consistent ground returns only in relatively open forest canopies (Weishampel et al., 1996), thus, making AGBM estimation difficult in dense tropical forests. Previous attempts to estimate tropical forest AGBM using small-footprint laser altimeters have also been complicated by the incompatibility of data sets, e.g., the lack of coincident field- and laser-derived data (Nelson, 1997; Nelson, Gregoire, & Oderwald, 1998; Nelson, Oderwald, & Gregoire, 1997).

Large-footprint lidar measurements incorporating information contained in the laser return waveform have been used to derive canopy height and structure in a variety of canopy closure conditions (e.g., Lefsky et al., 1999b; Means et al., 1999). Because these large-footprint lidar instruments consistently measure subcanopy topography even under conditions of high canopy closure, they have been shown to recover forest canopy structure that is statistically indistinguishable from field measurements (Lefsky, 1997), and are able to accurately capture spatial patterns of canopy heights (Drake & Weishampel, 2000). These instruments have also accurately estimated AGBM in both Douglas fir/western hemlock (Lefsky et al., 1999a; Means et al., 1999) and temperate mixed-deciduous forests (Lefsky et al., 1999b). In both cases, data from the lidar instruments were incorporated into regression models to derive plot-level forest structural (e.g., AGBM) estimates. These relationships were found to be significant even through dense structural conditions. For example, Means et al. (1999) predicted total plot AGBM with $R^2$ values of up to .96 using lidar-based AGBM estimation models through biomass levels of 1300 Mg/ha, far exceeding the normal saturation point of radar ($\approx 150$ Mg/ha from Dobson et al., 1992; Imhoff, 1995; Ranson, Sun, Weishampel, & Knox, 1997; Waring et al., 1995). These results suggest that data from the upcoming (2001) spaceborne VCL (Dubayah et al., 1997) mission will greatly improve global biomass estimates.

3. Methods

3.1. La Selva Biological Station

The La Selva Biological Station (Fig. 2; also see Clark, 1990; Matlock & Hartshorn, 1999; McDade, Bawa, Hespenheide, & Harshorn, 1994) is located near the Sarapiquí River in northeast Costa Rica. Over its 46-year history, La Selva has become one of the most heavily studied tropical field stations in the world (McDade et al., 1994). This 1536-ha area is comprised of a mixture of lowland primary and secondary tropical wet forest (Holdridge, Grenke, Hatheway, Liang, & Tosi, 1971), abandoned pasture, current and...
The primary forest estimated AGBM and basal area values at La Selva are low in comparison with other primary tropical rainforests (Brown, 1997; Brown et al., 1995; Laurance et al., 1999; Saldarriaga, West, Tharp, & Uhl, 1998). This may be due to differences between tropical moist forests (where the majority of studies are conducted) and tropical wet forest such as La Selva (see Clark & Clark, 2000, for a more detailed discussion). Nevertheless, the variety of land cover types and the wealth of ancillary data (e.g., soil, topography, forest structural data) available make La Selva an ideal study site for this study.

### 3.2. Field data

Forest structural data (Table 1) were collected across a successional spectrum ranging from abandoned pasture to primary wet tropical forests (Fig. 2). Primary tropical forest data were collected in eighteen 0.5-ha plots that are part of an ongoing landscape-scale carbon storage and flux study (Clark & Clark, 2000). The plots were stratified over three edaphic conditions: relatively fertile flat inceptisols on old alluvial terraces (A plots), relatively infertile ultisols areas on ridgetops (L plots), and ultisol areas on steep slopes (P plots). All 18 plots were geolocated without knowledge of existing forest structure. This approach eliminates placement biases that can lead to large inaccuracies when AGBM values are extrapolated over a landscape scale (Brown et al., 1995; Clark & Clark, 2000).

Secondary forest data were collected in three different areas that were approximately 14, 22 (Chazdon, 1996; Guariguata, Chazdon, Denslow, Dupuy, & Anderson, 1997; Nicotra et al., 1999), and 31 (Pierce, 1992) years old, respectively, as of March 1998. The 14- and 22-year-old secondary forest plots are each approximately 0.5 ha.

Within the 31-year-old secondary forest area, six circular plots of 12.5-m radius were geolocated so as to approximately coincide with lidar footprints.

In each primary forest plot, all stem diameters greater than 10 cm were measured in a marked location on each tree, either at breast height (1.37 m) or, when necessary, above buttressing (Clark & Clark, 2000). In the 31-year-old secondary forest plots, all stem diameters greater than 10 cm at breast height were measured. In the 14- and 22-year-old secondary forest plots, all stem diameters greater than 5 cm at breast height were recorded. These measurements were taken both as a part of a March 1998 VCL calibration/validation campaign and as a part of the separate long-term studies.

Stem diameters were used to calculate quadratic mean stem diameter (QMSD, Eq. (1)), basal area, and AGBM using an equation for tropical wet forests (Brown, 1997). QMSDs were calculated to compensate for the different (i.e., 5 vs. 10 cm) minimum diameter thresholds that were used in the existing studies. In addition, published AGBM values from agroforestry plots at La Selva (Menalled, Kelty, & Ewel, 1998) and for tropical pastures (Olson, Watts, & Allison, 1983) were used in this study.

\[
\text{QMSD} = \left( \frac{\sum D^2}{n} \right)^{1/2}
\]

where \(D\) is the stem diameter and \(n\) is the number of stem diameters in the area.

### 3.3. Lidar data

The airborne instrument used in this study is the Laser Vegetation Imaging Sensor or LVIS (Blair et al., 1999). LVIS is a medium-altitude imaging laser altimeter designed and developed at NASA’s Goddard Space Flight Center. Variable-sized footprints and a randomly positionable laser beam and 7° telescope field of view allow LVIS to operate in a variety of modes. Footprint diameters can be varied from 1 to 70 m and footprint spacing can be varied both along and across track. The return signal, or waveform, is digitally recorded and converted to units of distance (by accounting for the speed of light through the atmosphere). In this study,
the vertical resolution was approximately 30 cm, as determined by the digitization rate. Ancillary information such as the pointing direction and position of the laser at the time of each pulse (provided using an inertial navigation system and GPS unit) are also recorded. The combination of these data postflight enables the geolocation of the laser footprint on the ground within a global reference frame (usually to better than 1 m horizontal accuracy) (Hofton et al., 2000a).

In March 1998, LVIS was flown in a NASA C-130 airplane to map the La Selva Biological Station and surrounding regions of NE Costa Rica. LVIS was in VCL emulator mode and operated at an altitude of 8 km above the ground to produce eighty 25-m diameter footprints separated by \( \sqrt{25} \text{ m} \) along-track and \( \sqrt{9} \text{ m} \) across-track (see Fig. 1). Unlike previous laser altimeter instruments that record narrow transects of data, LVIS is able to map entire landscapes with relatively few flights. Only LVIS footprints that were coincident with field study plots (see below for details) were selected.

Four metrics were derived from the waveforms (see Fig. 3). Lidar canopy height (LHT) was calculated by identifying: (1) the location within the waveform when the signal initially increases above a mean noise level/threshold (the canopy top), and (2) the center of the last Gaussian pulse (the ground return), and then calculating the distance between these locations. Second, the height of median energy (HOME) was calculated by finding the median of the entire signal (i.e., above the mean noise level) from the waveform, including energy returned from both canopy and ground surfaces. The location of the median energy is then referenced to the center of the last Gaussian pulse to derive a height. The HOME metric is, therefore, predicted to be sensitive to changes in both the vertical arrangement of canopy elements and the degree of canopy openness (including tree density). Third, the height/median ratio (HTRT) is simply the HOME divided by canopy height. The HTRT provides an index of how the location of HOME may change relative to the LHT through succession. Finally, a simple ground return ratio (GRND) was calculated by taking the total intensity (i.e., the number of digitizer counts) contained in all 30-cm vertical bins contained in the last Gaussian peak (Hoffon, Minster, & Blair, 2000b) divided by the sum of the intensity of all other canopy vertical bins of the waveform (see Fig. 3). Thus, GRND provides an approximation of the degree of canopy closure (note that the canopy closure can be directly derived from the waveform with certain canopy assumptions, as found in Means et al., 1999). These four metrics were then incorporated into a stepwise regression procedure to predict field measured basal area and QMSD and field-estimated AGBM.

### 3.4. Analysis

The analysis involving ground and laser altimeter-derived data was performed at both the level of individual LVIS footprints (25-m diameter circle) and at the level of the

**Table 2**

<table>
<thead>
<tr>
<th>Land cover type</th>
<th>Approximate age (year)</th>
<th>LHT (m)</th>
<th>HOME (m)</th>
<th>HTRT</th>
<th>GRND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary forest</td>
<td>old-growth</td>
<td>31.33</td>
<td>20.14</td>
<td>0.64</td>
<td>0.014</td>
</tr>
<tr>
<td>Secondary forest</td>
<td></td>
<td>25.48</td>
<td>16.65</td>
<td>0.65</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>17.68</td>
<td>11.32</td>
<td>0.63</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>23.67</td>
<td>9.21</td>
<td>0.38</td>
<td>0.023</td>
</tr>
<tr>
<td>Agroforestry</td>
<td>7</td>
<td>9.47</td>
<td>1.68</td>
<td>0.17</td>
<td>0.067</td>
</tr>
<tr>
<td>Pasture(^a)</td>
<td>&lt; 5</td>
<td>5.09</td>
<td>0.54</td>
<td>0.11</td>
<td>0.740</td>
</tr>
</tbody>
</table>

\(^a\) Data from two LVIS shots from abandoned pasture sites at La Selva were compared to tropical pasture AGBM data from Olson et al. (1983).
The diameters of all stems within the respective footprints or plots were used to calculate stem basal area. These stem diameters were also used to estimate AGBM using the tropical wet forest allometric equation (Eq. (2)) of Brown (1997):

\[
AGBM = 21.297 - 6.953(D) + 0.740(D^2)
\]

where \(D\) is the stem diameter in centimeters and \(AGBM\) is the estimated oven-dried AGBM in kg for the stem.

Basal area and estimated AGBM were then summed within the footprint or plot and were converted to standard units of area (m²/ha and Mg/ha, respectively). Additionally, the QMSD (Eq. (1)) within each footprint or plot was also calculated.

At the footprint level, the lidar metrics were derived from individual waveforms. Plot-level values were calculated as

### Table 3
Regression equations and values for footprint-level (0.05 ha) forest structural characteristics

<table>
<thead>
<tr>
<th>Forest structural characteristic</th>
<th>Equation</th>
<th>(R^2)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMSD (cm)</td>
<td>(1) (QMSD = 10.39 + 0.72 \times \text{HOME} )</td>
<td>.59</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td>(2) (\log(QMSD) = 1.14 + 0.05 \times \text{LHT} - 0.05 \times \text{HOME} - 0.72 \times \text{GRND} + 2.43 \times \text{HTRT} )</td>
<td>.77</td>
<td>3.74</td>
</tr>
<tr>
<td>Basalarea (m²/ha)</td>
<td>(3) (\text{BArea} = 5.11 - 6.28 \times \text{GRND} )</td>
<td>.27</td>
<td>7.88</td>
</tr>
<tr>
<td></td>
<td>(4) (\text{BArea} = 3.24 + 0.04 \times \text{LHT} - 4.12 \times \text{GRND} + 1.15 \times \text{HTRT} )</td>
<td>.39</td>
<td>7.16</td>
</tr>
<tr>
<td>Estimated AGBM (Mg/ha)</td>
<td>(5) (\log(\text{AGBM}) = 3.58 + 0.07 \times \text{HOME} )</td>
<td>.53</td>
<td>63.17</td>
</tr>
<tr>
<td></td>
<td>(6) (\log(\text{AGBM}) = 2.06 + 0.07 \times \text{LHT} - 0.08 \times \text{HOME} - 1.05 \times \text{GRND} + 3.51 \times \text{HTRT} )</td>
<td>.73</td>
<td>60.02</td>
</tr>
</tbody>
</table>

* All values significant \((P < .01)\).

### Table 4
Regression equations and values for plot-level (0.25–0.5 ha) forest structural characteristics

<table>
<thead>
<tr>
<th>Forest structural characteristic</th>
<th>Equation</th>
<th>(R^2)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMSD (cm)</td>
<td>(1) (QMSD = 8.19 + 0.81 \times \text{HOME} )</td>
<td>.92</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>(2) (QMSD = 5.97 + 0.15 \times \text{LHT} + 1.05 \times \text{HOME} - 0.01[\text{LHT} \times \text{HOME}] )</td>
<td>.93</td>
<td>2.00</td>
</tr>
<tr>
<td>Basal area (m²)</td>
<td>(3) (\text{BArea} = 11.09 + 20.10 \times \text{HTRT} )</td>
<td>.72</td>
<td>3.00</td>
</tr>
<tr>
<td>Estimated AGBM (Mg/ha)</td>
<td>(4) (\text{AGBM} = 26.28 + 6.77 \times \text{HOME} )</td>
<td>.89</td>
<td>22.54</td>
</tr>
<tr>
<td></td>
<td>(5) (\text{AGBM} = 15.64 + 9.54 \times \text{HOME} - 0.01 \times \text{HOME}^3 )</td>
<td>.93</td>
<td>18.39</td>
</tr>
</tbody>
</table>

* Numbers are cross-validated.
* All values significant \((P < .01)\).
the means of metrics for all footprints within the plot. For example, plot-level LHT was calculated for each plot by taking the mean of LHT values for all footprints within the plot (Fig. 4).

Metrics derived from LVIS waveforms (Fig. 3) were then used to estimate the field-derived forest structural summaries at both the footprint and plot levels using stepwise multiple linear regression. During this process, transformations of dependent and independent variables (including square, square root, and logarithmic) were also explored. Those models that were found most predictive were then cross-validated (Cressie, 1991) to define a generalization error (RMSE). Finally, these models were applied to all of the LVIS data over La Selva to produce landscape-level maps of forest structural characteristics.

4. Results

The assortment of plots from different land cover types that are incorporated in this study span the range of forest structural characteristics for the La Selva landscape. Additionally, because the 18 primary forest plots are evenly stratified over the main upland edaphic and topographic conditions at La Selva (Clark & Clark, 2000), the primary forest data are an unbiased, representative sample of the structural conditions found in this primary tropical wet forest.

In general, all of the forest structural characteristics increase through the range of land cover types that were sampled (Table 1). There is a trend of increasing estimated AGBM and QMSD from pasture through secondary forest and into primary tropical forests. The exception to this trend is for the average QMSD value for the six 31-year-old secondary forest sites, which is larger than the average QMSD of the primary forest sites. This exception is likely the result of large remnant trees (i.e., those not cut down when the area was originally deforested), which are found within the 31-year-old forest area (Clark, personal observation) and strongly influence the QMSD values.
Overall, the lidar metrics from these areas (Table 2) are sensitive to changes in forest structure in the different land cover types included in this study. The height metrics (i.e., LHT, HOME, and HTRT) are the most sensitive, and increase with increasing forest age, basal area, and biomass. GRND is generally insensitive to changes in forest structure beyond a secondary forest age of approximately 14 years.

At the footprint level (0.05 ha), metrics from LVIS are able to significantly estimate all three forest structural attributes throughout the range of conditions at La Selva, though in most cases logarithmic transformations of dependent variables was necessary. The HOME metric is the best single predictor of both QMSD (Table 3, Eq. (1)) and AGBM (Table 3, Eq. (5)) and explains 59% and 53% of the variation in these attributes, respectively. The GRND metric was the best single-term predictor of basal area (Table 3, Eq. (3)) and explained 27% of the variation.

For all three structural characteristics, multiple-term equations (Table 3, Eqs. (2), (4), and (6)) explain much higher levels of variation ($R^2$) than the single-term equations. The level of variation of basal area that is explained by both single- and multiple-term equations is over 40% lower than the levels for AGBM and QMSD.

At the plot level (~0.5 ha), all of the relationships between forest structural summaries and lidar metrics are much stronger than at the footprint level. For all single-term equations, the levels of variation in forest structural summaries explained ($R^2$ values) are over 35% higher at the plot level than at the footprint level and range from...
The metric that is the best single-term predictor of both plot-level QMSD (Table 4, Eq. (1)) and AGBM (Table 4, Eq. (2)) values for AGBM and QMSD are higher than for basal area.

Fig. 9. Image of basal area predicted from LVIS data over the La Selva Biological Station using Eq. (3) in Table 4. Note the differences in basal area between younger secondary forest areas and primary forests (see land cover data in Fig. 2).

Fig. 10. Image of AGBM predicted from LVIS data over the La Selva Biological Station using Eq. (4) in Table 4. Note the areas of lower AGBM in western portions of the La Selva that were selectively logged in the late 1970s. In the right inset, AGBM predicted using the footprint-level (~0.05 ha) equation (Eq. (5) of Table 3) clearly reveals the clustering of high AGBM values near streams.
Eq. (4)) is HOME, as is the case at the footprint level (Table 4). The HOME metric explains 89% of the variation in estimated AGBM and 92% of the variation in QMSD with no transformation of the dependent or independent variables. The best single-term predictor of plot-level basal area is HTRT ($R^2 = .72$), instead of GRND, which is the best predictor of basal area at the footprint level.

Multiple-term equations (Table 4, Eqs. (2) and (5)) only marginally improve the relationship for both QMSD ($R^2 = .93$ vs. .92) and estimated AGBM ($R^2 = .93$ vs. .89). For basal area, the single-term equation (Table 4, Eq. (3)) was selected through all combinations and transformations in the stepwise multiple regression procedure.

To further examine the nature of the single-term relationships, they are graphically represented in scatter plots (Figs. 5–7). In all three cases, the metrics significantly estimate forest structural summaries without reaching an asymptote throughout the entire range of conditions at La Selva. In addition, the cross-validated RMSE for all three single-term relationships (Table 4) is low. As a result, these relationships were applied to LVIS data over the entire La Selva landscape to produce images of estimated QMSD (Fig. 8), basal area (Fig. 9), and AGBM (Fig. 10) with a resolution equal to the plot level used in this study (~0.5 ha). In addition, the footprint-level (~0.05 ha) equation for AGBM (Table 3, Eq. (5)) was applied to a subset of the LVIS data within a primary forest area to produce a high-resolution image of estimated AGBM variability (Fig. 10, inset).

5. Discussion

5.1. Field and lidar summaries

The trends in forest structural summaries (Table 1) are primarily in agreement with results from other studies for La Selva. As has been reported in other studies at La Selva (e.g., Guariguata et al., 1997), basal area does not significantly differ between older (i.e., ~22 years) secondary and old-growth forests. The large remnant stems in the 31-year-old secondary forest area result in the exception to the trend of increasing QMSD size through secondary forests and into primary forest plots, as mentioned above. Estimated AGBM is not as sensitive to remnant stems in secondary forest areas, and increases through the range of successional conditions sampled at La Selva.

Lidar-derived metrics are quite sensitive to changes in forest structure, particularly LHT, HOME, and HTRT (Table 2). The LHT from the 14-year-old secondary forest plot is higher than expected, which could be the result of two factors. First, large remnant stems, which can account for as much as 15% of the total basal area in secondary forests at La Selva (Guariguata et al., 1997), coupled with the sensitivity of LVIS canopy height to the maximum detectable canopy surface, may explain why these values are higher than expected. Second, precise plot boundary coordinates were not available for the 14-year-old secondary plot, so the LVIS shots may not be entirely coincident with the actual plot location.

5.2. Footprint-level relationships

The relationships between lidar metrics and field-derived forest structural summaries at the footprint-scale (~0.05 ha), though not as strong as the plot-level relationships (Tables 3 and 4), are significant. The multiple-term equations (Table 3, Eqs. (2) and (6)) selected through a stepwise multiple regression procedure explain 77% of the variation in QMSD and 73% of the variation in estimated AGBM across the range of conditions sampled in this dense tropical landscape. Still, the footprint-scale $R^2$ values are less than 65% of the plot-level single-term equations and less than 80% of the plot-level multiterm regression equations (Tables 3 and 4).

The footprint-level relationships could be negatively affected by two factors. First, the level of variability in forest structure at the scale of an LVIS footprint (0.05 ha) is much higher than at the plot level (~0.5 ha). The level of variation in forest structural characteristics (as determined by calculating the coefficient of variation—CV) is two to three times higher at the footprint scale as the plot level (Table 5). This agrees with other studies that have reported that a plot size of approximately 0.35–0.5 ha is necessary for sampling tropical forest estimated AGBM and other structural characteristics (Brown et al., 1995; Clark & Clark, 2000).

A second factor that contributes to the weaker relationship between forest structural attributes and lidar metrics at the footprint level is geolocation of LVIS observations and stems in each forest area. Although LVIS shots may be geolocated to within 1–2 m (Blair & Hofton, 1999; Hofton et al., 2000a), the location of stems within primary and the 31-year-old secondary forest plots were referenced to a plot corner using a compass and fiberglass measuring tapes. As a result, stem locations may be off by 5–10 m in random (i.e., approximately unbiased) directions. Thus, stems that were included in particular LVIS footprints for this analysis may be as many as 12 m (LVIS shot geolocation uncertainty + stem geolocation uncertainty) outside of the footprint area, and stems that are up to 12 m outside of the

<table>
<thead>
<tr>
<th>Forest structural characteristic</th>
<th>Footprint-level CV</th>
<th>Plot-level CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMSD</td>
<td>22.54</td>
<td>9.60</td>
</tr>
<tr>
<td>Basal area</td>
<td>33.56</td>
<td>11.94</td>
</tr>
<tr>
<td>Estimated AGBM</td>
<td>43.41</td>
<td>14.01</td>
</tr>
</tbody>
</table>

* These values were calculated using only primary and 31-year-old secondary forests.
footprint area and were not included may have truly been within the footprint.

5.3. Plot-level relationships

The relationships between lidar metrics and forest structural characteristics at the plot level are strong (Table 4). Metrics from LVIS are able to explain very high levels of variation ($R^2$ up to .93) in tropical forest AGBM, basal area, and QMSD. These relationships are nonasymptotic (Figs. 5–7) through the entire range of conditions sampled at La Selva and permit estimates of AGBM to approximately the same level of accuracy as related large-footprint lidar studies in temperate coniferous forests (Leffsky et al., 1999a; Means et al., 1999).

The slightly weaker relationship between basal area and lidar metrics is caused by the lack of significant differences in basal area between secondary and primary forest areas that has been reported in field studies at La Selva (Guariguata et al., 1997). Although the frequency distribution of basal area (i.e., number of stems in different size classes) does change, the overall total basal area at La Selva is not significantly different beyond a secondary forest age of approximately 16 years since abandonment (Guariguata et al., 1997). As a result, lidar metrics such as LHT, HOME, and HTRT, which continue to increase with increasing forest age (Table 1), may not be as predictive of changes in total basal area compared to changes in estimated AGBM and QMSD. This may explain why GRND (which is also insensitive to changes beyond a young secondary forest) was the best single-term predictor of basal area at the footprint level (Table 3).

HOME is perhaps the metric with the strongest potential for estimating tropical forest structural characteristics. It is the best single-term predictor of both footprint- and plot-level QMSD and AGBM (Tables 3 and 4). The LHT metric is strongly influenced by the highest detectable canopy surface within a footprint. The HOME metric, however, may be more sensitive to both the vertical arrangement and density of canopy elements. In areas with densely packed canopy materials, less lidar energy will reach the ground, thereby increasing HOME. Conversely, in more open or disturbed areas (e.g., a tree-fall gap), more lidar energy will reach the ground, thus, reducing HOME. Because primary tropical forests represent a spectrum from newly created tree-fall gaps to mature patches with high canopy closure (Lieberman, Lieberman, & Peralta, 1989), the sensitivity of HOME to these changes make it an excellent predictor of forest structural attributes such as biomass.

The RMSE of 22.54 Mg/ha for the AGBM (Table 4, Eq. (4)) is 13.75% of the mean (160.5 Mg/ha) plot-level AGBM for all primary forest areas sampled at La Selva. This value is also approximately 14% of the mean (965 Mg/ha) primary coniferous forest AGBM level (Means et al., 1999). Additionally, this value (the RMSE divided by the mean primary forest AGBM level) is approximately equal to the RMSE divided by the mean primary forest AGBM level (Means et al., 1999). Consequently, the ability of a large-footprint lidar to accurately predict tropical forest structural characteristics across a dense, structurally complex tropical forest landscape is extraordinary.
6. Conclusions

Metrics derived from a large-footprint lidar instrument were significantly correlated with tropical forest structural characteristics at both footprint and plot levels across the entire range of conditions in a structurally complex tropical wet forest. Although the majority of the field-measured forest structural data are from the high-end of the biomass and basal area spectrum at La Selva, the instrument nonetheless successfully measured the structural heterogeneity within this dense primary forest at both footprint (0.05 ha) and plot (~0.5 ha) levels.

At the level of individual LVIS footprints (0.05 ha), the relationship between lidar metrics and forest structural characteristics is weakened by problems of geolocation (stems and LVIS footprints), and the level of variation in forest structure at that scale. Nevertheless, even with these factors the relationships between lidar metrics and tropical forest structural attributes are strong through the entire range of conditions sampled.

At the plot level (0.5 ha), these factors are negligible and the relationship among plot-level forest structural summaries and lidar metrics are extremely strong. The levels of variation explained by metrics from the LVIS instrument at this scale are significantly higher than for any other remote sensing instrument for tropical forests areas to our knowledge. The level of RMSE of the relationships between lidar metrics and QMSS and estimated AGBM is approximately the same as from previous studies in more open temperate forests. As a result, when these relationships are applied to LVIS data over the entire landscape at La Selva, it is possible to examine the relationship between forest structural characteristics and environmental conditions (e.g., topography) and past land use (e.g., selective logging).

Although several of the forest structural characteristics used in this study from La Selva are lower than many moist tropical rainforests (Brown, 1997; Brown et al., 1995; Laurance et al., 1999; Saldarriaga et al., 1988), this is not the case for all forest characteristics at La Selva. For example, the heights of emergent (>50 m) and average canopy-forming trees (~33 m) at La Selva are approximately equivalent to those found in other neotropical rainforests (Richards, 1996).

Perhaps more important from a lidar remote sensing point of view is the light availability at ground level. The degree of canopy closure in primary and secondary forest areas at La Selva is approximately 98–99% at 1 m above ground level (Fetcher, Oberbauer, & Chazdon, 1994; Nota et al., 1999). This is among the highest canopy closure values found in tropical and extratropical forests (Baldocchi & Collineau, 1994). Thus, although large-footprint lidar has proven effective for estimation of forest structure in temperate forests with higher AGBM levels than La Selva, the ability of this technology to recover forest structural characteristics in a dense tropical forest with three to four times higher canopy closure than most temperate forests (Baldocchi & Collineau, 1994) is critical. This ability is particularly important at a global scale because tropical forests are estimated to contain approximately 40% of the carbon in the terrestrial vegetation pool (Dixon et al., 1994). The combination of the previous efforts in temperate forests along with this study in tropical forests strongly confirm that next-generation lidar technology, as found on the VCL mission, will greatly improve global estimates of AGBM and other forest structural characteristics.

Acknowledgments

We are grateful to Nancy Casey-McCabe, Dave Kendig, and Birgit Peterson for their help in processing the LVIS data, and Leonel Campos Ramos, William Conejo Miranda, Marcos Molina, Jeanette Paniagua, Braulio Vilchez, Laura Rochio, and Birgit Peterson for valuable help in collecting field data. We also thank the La Selva Biological Station, the Organization for Tropical Studies, the Carbono Project (funded by DOE and NSF), Wallops Flight Facility Aircraft Programs Office, the National Geographic Institute of Costa Rica, and the Government of Costa Rica. We also thank the anonymous referees for helpful comments on this paper. Our project is funded by a NASA contract to the University of Maryland for the implementation and execution of the VCL Mission under the Earth System Science Pathfinder program.

References


Clark, D. B. (1990). La Selva Biological Station: a blueprint for stimulating


