Estimates of forest canopy height and aboveground biomass using ICESat

Michael A. Lefsky,[1] David J. Harding,[2] Michael Keller,[3,4] Warren B. Cohen,[5] Claudia C. Carabajal,[6] Fernando Del Bom Espirito-Santo,[7] Maria O. Hunter,[4] and Raimundo de Oliveira Jr.[8]

Received 5 July 2005; revised 7 September 2005; accepted 7 October 2005; published 1 November 2005.

[1] Exchange of carbon between forests and the atmosphere is a vital component of the global carbon cycle. Satellite laser altimetry has a unique capability for estimating forest canopy height, which has a direct and increasingly well understood relationship to aboveground carbon storage. While the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud and land Elevation Satellite (ICESat) has collected an unparalleled dataset of lidar waveforms over terrestrial targets, processing of ICESat data to estimate forest height is complicated by the pulse broadening associated with large-footprint, waveform-sampling lidar. We combined ICESat waveforms and ancillary topography from the Shuttle Radar Topography Mission to estimate maximum forest height in three ecosystems; tropical broadleaf forests in Brazil, temperate broadleaf forests in Tennessee, and temperate needleleaf forests in Oregon. Final models for each site explained between 59% and 68% of variance in field-measured forest canopy height (RMSE between 4.85 and 12.66 m). In addition, ICESat-derived heights for the Brazilian plots were correlated with field-estimates of aboveground biomass (r^2 = 73%, RMSE = 58.3 $Mgha^{-1}$).
Citation: Lefsky, M. A., D. J. Harding, M. Keller, W. B. Cohen, C. C. Carabajal, F. Del Bom Espirito-Santo, M. O. Hunter, and R. de Oliveira Jr. (2005), Estimates of forest canopy height and aboveground biomass using ICESat, *Geophys. Res. Lett.*, *32*, L22S02, doi:10.1029/2005GL023971.

1. Introduction

[2] Accurate estimates of terrestrial carbon storage are required to determine its role in the global carbon cycle, to estimate the degree that anthropogenic disturbance (i.e., land use/land cover change) is altering that cycle, and to monitor mitigation efforts that rely on carbon sequestration through reforestation. Lidar remote sensing, from airborne or satellite platforms, has a unique capability for estimating forest canopy height; this has a direct and increasingly well-understood relationship to aboveground carbon storage [*Lefsky et al.*, 2002]. The Geoscience Laser Altimeter System (GLAS) on the Ice, Cloud and land Elevation Satellite (ICESat) has the potential to provide such information globally; measurement of canopy height is a science objective of the ICESat mission [*Zwally et al.*, 2002]. However, processing of ICESat data to create reliable estimates of forest height is complicated by the pulse broadening [*Harding and Carabajal*, 2005] that occurs in the received echo waveform when the large GLAS laser footprint illuminates vegetation on a sloped surface.

[3] The GLAS laser footprint is elliptical and varies in size as a function of laser operating conditions; the average ellipse size for the laser operations periods used here was 53 × 97 m [*Abshire et al.*, 2005]. For forests on level ground, discrete peaks in the waveform separate the height distribution of reflecting canopy surfaces from that of the underlying ground within this large footprint [*Harding and Carabajal*, 2005]. Over sloped areas, the vertical extent of each waveform increases as a function of the product of the slope and the footprint size, and returns from both canopy and ground surfaces can occur at the same elevation [*Harding and Carabajal*, 2005]. As a result, ancillary topographic information is required to make estimates of canopy height. Given forest stands of uniform height and highly accurate ancillary topography, separation of waveforms into ground and canopy components is straightforward. However, many mature forests are non-uniform in height, and adequate topographic characterization is rare. Therefore, algorithms are needed that are insensitive to the limitations of existing topographic datasets, and can make inferences about stand uniformity from aspects of the waveform itself. In this paper we demonstrate a technique utilizing Shuttle Radar Topography Mission (SRTM) data to correct for the broadening of ICESat waveforms.

2. Methods

2.1. Study Areas

[4] To evaluate the ability of ICESat waveforms to estimate forest canopy height, three field sites were selected. Two sites were selected to represent coniferous (Oregon, USA) and deciduous (Tennessee, USA) forest types located on high slopes (mean slopes > 18%). The third site

[1]Department of Forest Sciences, Colorado State University, Fort Collins, Colorado, USA.
[2]Planetary Geodynamics Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]International Institute of Tropical Forestry, USDA Forest Service, San Juan, Puerto Rico.
[4]Complex Systems Research Center, University of New Hampshire, Durham, New Hampshire, USA.
[5]Forestry Sciences Laboratory, Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon, USA.
[6]NVI, Inc., NASA/GSFC Space Geodesy Branch, Greenbelt, Maryland, USA.
[7]Instituto Nacional de Pesquaisas Espaciais, São Josédos Campos, Brazil.
[8]EMBRAPA Amazônia Oriental, Belém, Brazil.

0094-8276/05/2005GL023971$05.00

(Santarem, Para State, Brazil) was selected as the first of several planned sites within the Amazon basin.

[5] The Oregon sites are in the Willamette National Forest and are predominately associated with temperate coniferous forests of Douglas-fir and western hemlock (*Psuedotsuga menziesii, Tsuga heterophylla*) 45 km south of the H. J. Andrews Experimental Forest. The Tennessee sites are in Great Smoky Mountains National Park, and are associated with both northern hardwoods and mixes of hardwoods and pines. Dominant species include oaks (*Quercus spp.*) and other hardwood species, as well as Virginia and white pine (*Pinus virginiana, P. strobus*). The Santarem sites are within and in the vicinity of the Tapajos National Forests (TNF) near Santarem, in Para State, Brazil. Sampling focused on two areas, km 67 and São Jorge. The km 67 area is relatively undisturbed old-growth forest [*Keller et al.*, 2004], while the São Jorge community contains secondary forests of a range of ages and biomass densities, pasture, and agricultural fields [*Espírito-Santo et al.*, 2005]. The species composition of the Santarem sites is too diverse to describe in this manuscript; readers are directed to the citations above.

2.2. ICESat Data and Geographic Positions

[6] The ICESat data we used here were from cloud-free profiles acquired in October, 2003, February, 2004, and May–June, 2004 during the Laser 2a, 2b, and 2c operations periods, respectively [*Schutz et al.*, 2005]. We obtained geolocated footprint locations from the GLA06 Global Elevation Data Product and, where the GLA06 product did not yield geolocation results for low amplitude returns, from the GLA14 Global Land Surface Altimetry Data Product. To account for systematic pointing errors, footprint geographic positions were refined by comparing the GLA06 elevations with coincident elevations from a DEM. For all ICESat footprint locations along a profile, we calculated a root mean square error (RMSE) between the DEM and GLA06 elevations. The data were then offset in 1 pixel increments within a range of $\mp$200 m along north/south and east/west axes, and the RMSE recalculated. The profile location with the minimum RMSE was then used to define the footprint locations. In North America, we used the USGS 10 m National Elevation Dataset [*Gesch et al.*, 2002] for location refinement. For the Santarem site, we used the 90 m SRTM elevation dataset (resampled to 30 m using bilinear interpolation).

2.3. ICESat Waveform Processing

[7] Waveform extent is defined as the vertical distance between the first and last elevations at which the waveform energy exceeds a threshold level (see *Harding and Carabajal* [2005] for details of waveform processing). In this work, the threshold was determined by fitting a Gaussian distribution to the peak of lowest energy in a histogram of waveform energy, which identifies the mode and standard deviation of background noise in each waveform. The threshold was set to the noise mode plus 4 times the standard deviation. At the leading edge of the waveform, the "signal start" threshold crossing indicates the elevation of the uppermost foliage and/or branches, and the trailing edge threshold crossing indicates the elevation of the lowest illuminated surface, or the "signal end" [*Harding and Carabajal*, 2005]. Where sufficient laser energy is reflected from the ground, this trailing-edge threshold crossing represents the lowest detected ground surface. As a consequence, the waveform includes the height of the canopy as well as the vertical distribution of the ground surface in areas where the ground slope is greater than zero. We excluded from the analysis data not suitable for determining waveform extent due to upper signal truncation, saturation or very low signal-to-noise ratio [*Harding and Carabajal*, 2005]. To determine the effect of upper canopy variability on the height estimate,the extent of the waveform leading edge was manually estimated as the distance between signal start to the first inflection point (peak) in the waveform.

2.4. Terrain Indices

[8] Terrain index was defined as the range of ground surface elevations within one of three sampling windows (3×3, 5×5, and 7×7 DEM pixels) applied to a digital elevation model (DEM) at the GLAS footprint location, without consideration of the resolution of the DEM. For the North American sites (Oregon, Tennessee) we used the 30 m SRTM DEM, in Santarem we used the 90 m SRTM product [*Farr and Kobrick*, 2000]. At each sampling window size, five sampling window patterns were used: the entire ($n \times n$) sampling window, and single pixel wide lines at four azimuths: (0°, 45°, 90°, 135°). For vegetated areas, SRTM DEMs represent the radar phase center elevation, which depends on canopy structure and fractional cover [*Carabajal and Harding*, 2005], and are an approximation for ground elevation.

2.5. Field Sampling

[9] We established field plots located within ICESat footprints where we measured forest canopy properties. In the two U.S. sites, we stratified potential plots by waveform extent and SRTM-derived terrain index, and then randomly selected plots in each class in order to obtain a representative sample. In the Amazon, the difficulty of reaching plots selected at random was too great, and we selected plots on the basis of their proximity to existing roads and open fields.

[10] We modified plot layout and sampling procedures in Tennessee and Oregon from those of the local Forestry Inventory and Analysis programs of the United States Department of Agriculture (USDA) Forest Service [*USDA*, 2004]. For trees with diameter at breast height (DBH) greater then 12.7 cm, we recorded DBH, species, and height on four 7.32 m radius subplots (three subplots located 36.6 meters from a central subplot at azimuths of 360°, 120°, and 240°). To ensure sampling of uncommon large trees at the Oregon locales we also tallied all trees with DBH greater that 61 cm on four annular plots 35.92 m in diameter and centered on each of the subplots, and tallied all trees with DBH greater than 81 cm in a single 112.87 m diameter plot centered on the central subplot.

[11] For old growth forest at Santarem and a remnant logged forest at São Jorge, we established a main plot (20×75 m) along the transect and two perpendicular side plots (40×27.5 m each). In these plots, DBH and maximum height were tallied for all trees with DBH greater than 35 cm. Within the main plot, DBH for all trees with DBH between 10 and 35 cm were recorded on a subplot (10×75 m); for a 30% sample of these smaller trees, we recorded maximum height. Biomass (kg dry mass) for the old growth

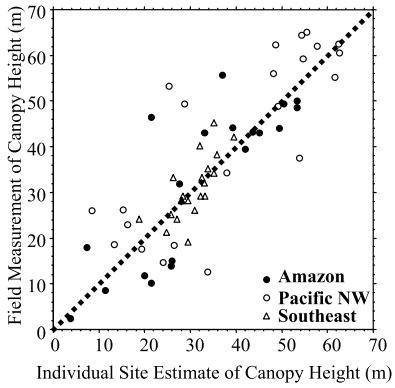

Figure 1. Observed forest maximum canopy height vs. ICESat estimates of same, for the three study areas and overall. See Table 2 for relevant correlation coefficients and RMSE.

forest trees was calculated using a polynomial function of DBH (cm) for tropical moist forest [*Brown*, 1997].

$$\text{Biomass} = 42.69 - 12.80 * \text{DBH} + 1.242 * \text{DBH}^2 \quad (1)$$

[12] For secondary forests, we sampled using various densities of randomly-located subplots along a 75 m long transect; where subplot density varied as a function of stem density (19 2 × 2 m plots in a recently abandoned agricultural fields, eight or nine 4 × 4 m plots in secondary forests). All stems greater than 10 cm DBH were measured in the sub-plots. Maximum height of the tallest tree in a 75 m circle centered on the footprint was also measured. We estimated biomass using allometric relations developed for the Central Amazon by *Nelson et al.* [1999]. Equation (2a) was applied to *Cecropia* spp. and equation (2b) was applied for all other species.

$$\text{Biomass} = 0.081122 * \text{DBH}^{2.4257} \quad (2a)$$

$$\text{Biomass} = 0.13577 * \text{DBH}^{2.4128} \quad (2b)$$

2.6. Relating Remotely Sensed Indices to Canopy Height

[13] The first objective in this work was to determine the relationship between the remotely sensed measures of waveform extent and terrain relief, and field measured maximum tree height, using the following equation:

$$h = b_0(w - b_1 g) \quad (3)$$

Where

h is the measured canopy height
w is the waveform extent in meters
g is the terrain index (i.e. ground extent) in meters
b_1 is the coefficient applied to the terrain index
b_0 is the coefficient applied to the waveform, when corrected for the scaled terrain index.

Equations were fit using the Interactive Data Language implementation of the MPFIT package, a robust least squares curve fitting package based on the Levenberg-Marquardt algorithm (Craig Markwardt, http://cow.physics.wisc.edu/~craigm/idl/idl.html).

[14] To incorporate the extent of the waveform's leading edge, a modified version of the equation was used:

$$h = b_0(w - b_1 g + b_2 l) \quad (4)$$

Where

l is the extent of the leading edge in meters
b_2 is the coefficient applied to the leading edge

3. Results

3.1. Study Area Characteristics

[15] Field measured maximum canopy heights in each study area followed the expected order, with a tallest maximum height of 65 m at the Oregon study area, 55.7 m at Santarem, and 45 m at the Tennessee site. Shortest maximum canopy heights were close to zero in the Oregon and Santarem study areas, but in Tennessee the shortest stands available were 19 m tall. This has the effect of reducing the apparent goodness of fit of our regressions for this site.

3.2. Terrain Index Selection

[16] The correlation between each terrain index and the difference between the extent of the lidar waveform and the field-measured maximum canopy height was used to evaluate each index. At all three sites, the terrain indices derived from a square 3 × 3 matrix of elevations were best able to estimate the height difference.

3.3. Estimation of Plot Maximum Canopy Height

[17] Regression was used to estimate maximum canopy height as a function of waveform extent and the 3 × 3 terrain index. When all three sites were considered in a single regression, the resulting equation explained 48% of variance with an RMSE of 12.14 m, but individual sites had clear biases. A second round of regressions was used to create equations at each of the three sites. Regression equations explained between 59% and 68% of variance at each site (Figure 1 and Table 1), with RMSE in the range of 4.85 to 12.66 m. When the upper canopy slope variable was added to the regression for the Tennessee site, the variance explained increased from 59% to 69% (Table 2).

3.4. Estimation of Plot Aboveground Biomass

[18] The ability of the ICESat-derived maximum canopy heights (Section 3.3) to predict aboveground biomass was tested at the Santarem study area. We do not have estimates of the more complex canopy structure indices used in

Table 1. Results of Regressions Relating Waveform Extent and a Terrain Index to Field Measured Maximum Canopy Height, Using the Equation $h = b_0 (w - b_1 g)$, Where h is Canopy Height, w is Waveform Extent, g is the Terrain Index, and b0 and b1 are Scaling Factors for the Waveform Extent and Terrain Index, Respectively

	R^2	B_0	B_1	Bias, m	RMSE, m	Count
Santarem	68%	1.08249	0.22874	−0.48	9.90	19
Oregon	64%	0.96599	0.05953	−1.71	12.66	24
Tennessee	59%	0.68778	0.14517	0.01	4.85	23
Combined	67%	—	—	−0.76	9.61	66
All	48%	0.88896	0.15427	−0.84	12.14	66

previous studies of canopies [*Lefsky et al.*, 2002], so maximum canopy height squared was used as the independent variable. The resulting equation explains 73% of variance in aboveground biomass (Figure 2),

$$AGBM = 20.7 + 0.098 * H_{est}^2 \quad (5)$$

Where

AGBM is aboveground biomass ($MgHa^{-1}$)

H_{est} is maximum canopy height (m) estimated from ICESat waveforms and SRTM elevation, as in section 3.3

4. Discussion

[19] The results from this study confirm that forest height can be estimated using waveforms from ICESat in combination with a measure of topographic relief. We were able to predict forest heights successfully over a wide range of canopy height and aboveground biomass, for both deciduous and coniferous forests, and over a range of slopes. Prior work showing the strong correlation between lidar-measured canopy height and above ground biomass [*Lefsky et al.*, 2002; *Drake et al.*, 2002] and the biomass result presented here for the Santarem site, provide confidence that ICESat waveforms in combination with SRTM data can substantially contribute to a global inventory of biomass.

[20] Examination of equation coefficients predicting canopy height indicate that additional work is required on the form of the equation itself, as well as the measurement of waveform extent. We expected that B_0 (the scaling parameter for the waveform extent, minus the terrain index, which is itself scaled by B_1) would be close to 1.0, as it is for the Santarem and Oregon study areas. However, at the Tennessee site, B_0 was 0.69. Furthermore, the B_1 parameter ranged from 0.05 to 0.14 for the U.S. sites (the 0.23 for Santarem refers to SRTM 90 data and is not directly comparable). Incorporation of the leading edge extent for all sites and use of a common source of topographic information (i.e., SRTM 90 m data) will be tested for their ability to create a common model for estimating canopy height. In addition, at the Oregon and Tapajos sites, there is a moderate level (r^2 = 0.13) of correlation between canopy height and the terrain index; adoption of a regression technique that considers this colinearity will be assessed in future work.

[21] The results of this work indicate that the combination of ICESat and SRTM data ultimately offers the possibility of a global assessment of forest canopy height, a measurement of fundamental importance heretofore not achievable by other means. The work also provides insights for future work to improve the accuracy of the canopy height estimations.

Table 2. Comparison of Two Regressions the Tennessee Site, the Two Parameter Equation From Table 1 and a Three Parameter Equation Incorporating Extent of the Waveform's Leading Edge[a]

Parameters	R^2	B_0	B_1	B_2	Bias, m	RMSE, m	Count
2	59%	0.68778	0.14517	—	0.01	4.85	23
3	69%	0.62108	0.36924	0.41841	0.01	4.21	23

[a]Three parameter equation is of the form: $h = b_0 (w - b_1 g + b_2 l)$, where l and b_2 are, respectively, the extent and scaling factor for the ICESat waveform's leading edge extent. The other symbols are explained in the caption to Table 1.

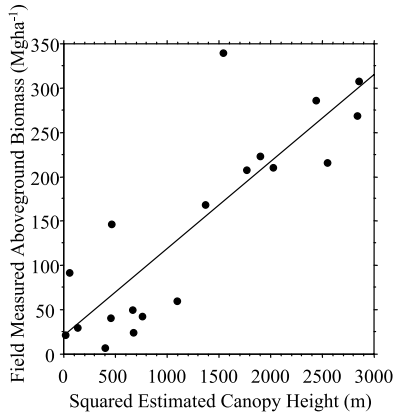

Figure 2. Observed vs. estimated aboveground biomass ($Mgha^{-1}$) for the Tapajos study area (r^2 = 73%, RMSE = 58.3).

References

Abshire, J., et al. (2005), Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance, *Geophys. Res. Lett.*, doi:10.1029/2005GL024028, in press.

Brown, S. (1997), *Estimating Biomass and Biomass Change of Tropical Forests: A Primer*, U. N. Food and Agric. Org., Rome.

Carabajal, C. C., and D. J. Harding (2005), ICESat validation of SRTM C-band digital elevation models, *Geophys. Res. Lett.*, doi:10.1029/2005GL023957, in press.

Drake, J. B., et al. (2002), Relationship between vertical canopy profiles and biomass in a neotropical rainforest, *Remote Sens. Environ.*, *81*, 378–392.

Espírito-Santo, F. D. B., et al. (2005), Mapping forest successional stages following deforestation in Brazilian Amazonia using multitemporal Landsat images, *Int. J. Remote Sens.*, *23*, 635–642.

Farr, T. G., and M. Kobrick (2000), Shuttle Radar Topography Mission produces a wealth of data, *Eos Trans. AGU*, *81*, 583–585.

Harding, D. J., and C. C. Carabajal (2005), ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure, *Geophys. Res. Lett.*, *32*, L21S10, doi:10.1029/2005GL023471.

Gesch, D. M., et al. (2002), The National Elevation Dataset, *Photogramm. Eng. Remote Sens.*, *68*, 5–11.

Keller, M., et al. (2004), Ecological research in the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA): Early results, *Ecol. Appl.*, *14*, S3–S16.

Lefsky, M. A., et al. (2002), Lidar remote sensing of aboveground biomass in three biomes, *Global Ecol. Biogeogr.*, *11*, 393–400.

Nelson, B. W., et al. (1999), Allometric regressions for improved estimate of secondary forest biomass in the central Amazon, *Forest Ecol. Manage.*, *117*, 149–167.

Schutz, B. E., H. J. Zwally, C. A. Shuman, D. Hancock, and J. P. DiMarzio (2005), Overview of the ICESat Mission, *Geophys. Res. Lett.*, doi:10.1029/2005GL024009, in press.

United States Department of Agriculture (USDA) (2004), Forest inventory and analysis national core field guide, internal report, For. Serv., Washington, D. C.

Zwally, H. J., et al. (2002), ICESat's laser measurements of polar ice, atmosphere, ocean, and land, *J. Geodyn.*, *34*, 405–445.

C. C. Carabajal, NVI, Inc., NASA/GSFC Space Geodesy Branch, Code 697, Greenbelt, MD 20771, USA.

W. B. Cohen, Forestry Sciences Laboratory, Pacific Northwest Research Station, USDA Forest Service, Corvallis, OR 97331, USA.

F. Del Bom Espirito-Santo, INPE, Av. dos Astronautas, 1758, São José dos Campos SP, Brasil CEP 12227-010.

R. de Oliveira Jr., EMBRAPA Amazônia Oriental, Belém, PA, Brasil CEP 66095-100.

D. J. Harding, Planetary Geodynamics Branch, NASA Goddard Space Flight Center, Code 698, Greenbelt, MD 20771, USA.

M. O. Hunter and M. Keller, Complex Systems Research Center, University of New Hampshire, Durham, NH 03824, USA.

M. A. Lefsky, Department of Forest Sciences, Colorado State University, Campus Delivery 1472, Fort Collins, CO 80523–1472, USA. (lefsky@fsl.orst.edu)