Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing

Wenge Ni-Meister,1 Shihyan Lee,1 Alan H. Strahler,2 Curtis E. Woodcock,2 Crystal Schaaf,2 Tian Yao,2 K. Jon Ranson,3 Guoqing Sun,3,4 and J. Bryan Blair3

Received 15 January 2009; revised 26 November 2009; accepted 1 February 2010; published 23 June 2010.

[1] Lidar-based aboveground biomass is derived based on the empirical relationship between lidar-measured vegetation height and aboveground biomass, often leading to large uncertainties of aboveground biomass estimates at large scales. This study investigates whether the use of any additional lidar-derived vegetation structure parameters besides height improves aboveground biomass estimation. The analysis uses data collected in the field and with the Laser Vegetation Imaging Sensor (LVIS), and the Echidna® validation instrument (EVI), a ground-based hemispherical-scanning lidar data in New England in 2003 and 2007. Our field data analysis shows that using wood volume (approximated by the product of basal area and top 10% tree height) and vegetation type (conifer/softwood or deciduous/hardwood forests, providing wood density) has the potential to improve aboveground biomass estimates at large scales. This result is comparable to previous individual-tree based analyses. Our LVIS data analysis indicates that structure parameters that combine height and gap fraction, such as RH100*cover and RH50*cover, are closely related to wood volume and thus biomass particularly for conifer forests. RH100*cover and RH50*cover perform similarly or even better than RH50, a good biomass predictor found in previous study. This study shows that the use of structure parameters that combine height and gap fraction (rather than height alone) improves the aboveground biomass estimate, and that the fusion of lidar and optical remote sensing (to provide vegetation type) will provide better aboveground biomass estimates than using lidar alone. Our ground lidar analysis shows that EVI provides good estimates of wood volume, and thus accurate estimates of aboveground biomass particularly at the stand level.


1. Introduction

[2] Lidar remote sensing provides measurements of the horizontal and vertical vegetation structure of ecosystems. This information will be critical for estimating global carbon storage and assessing ecosystem response to climate change and natural and anthropogenic disturbances. Unlike many other remote sensing measurements, lidars provide direct and indirect measurements of vegetation structure which can be used to estimate global carbon storage [Dubayah and Drake, 2000]. Recent advances in lidar technology have made lidar data widely available to study vegetation structure characteristics and forest biomass. The spaceborne Geoscience Laser Altimeter System (GLAS), part of the ICESat mission, provides global lidar data with a variable diameter of ~70 m footprint spaced at ~170 m [Zwally et al., 2002; Harding and Carabajal, 2005; Lefsky et al., 2005]. Airborne data have also been collected using a Scanning Lidar Imager of Canopies by Echo Recovery (SLICER) with a 15 m footprint and the Laser Vegetation Imaging Sensor (LVIS) with a 20 m/25 m footprint over several large areas for improved vegetation structure characterization [Blair et al., 1999]. Small-footprint multiple return lidar data have also been collected in many regions of the globe [Jupp et al., 2005] and more recently small footprint scanning waveform systems have become operational [Gutierrez et al., 2005; Neuenschwander et al., 2008]. Many studies have demonstrated the potential use of spaceborne and airborne lidar data to map vegetation height, aboveground biomass characteristics, and other vegetation structure parameters [Simard et al., 2008; Lefsky et al., 1999, 2002, 2005, 2007; Harding et al., 2001; Harding and Carabajal, 2005;
Carbon stocks or aboveground biomass are not directly measured by lidar. But many studies have demonstrated the strong relationship between aboveground biomass and lidar measured tree height, ranging from boreal conifers to equatorial rain forests [Lefsky et al., 2005; Drake et al., 2002a, 2002b]. These relationships are used to derive aboveground biomass from lidar-measured vegetation height [Lefsky et al., 2005] or accumulated vegetation lidar returns [Drake et al., 2002a, 2002b, 2003] at large scales.

However, large uncertainties still exist in large-scale aboveground biomass estimates from lidar. Forest aboveground biomass is actually related to several woody structure parameters, including trunk diameter at breast height, height of canopy, stem density, and branch distribution, but height is the only wooden structure parameter directly measured by lidar. Thus aboveground biomass has been indirectly derived based on empirical relationships with lidar measured vegetation height or accumulated lidar returns from vegetation. Often these relationships are site-dependent and lead to large uncertainties when applied over large regions, verifying that height may not be the only structure parameter relating to biomass. Drake et al. [2002a, 2002b] found that the height of medium energy returns (RH50) is better related to aboveground biomass than the height at full energy returns (RH100) due to RH50 being more sensitive to changes in both the vertical arrangement of canopy elements and the degree of canopy openness (including tree density) than height itself. Thus further investigation is required to have a clear understanding of the links between aboveground biomass, vegetation structure parameters measured from field and from lidar in order to develop a more physically based approach for improved aboveground biomass estimates from lidar. This will also help us to better understand the aboveground biomass retrieval accuracy from lidar for the ICESat-II and DESDynl missions. This study is a first step into this direction, and uses field measurements of vegetation structure and both airborne LVI and hemispherical scanning ground lidar EVI data collected in the New England region in 2003 and 2007.

2. Review of Aboveground Biomass Estimate

Two approaches are commonly used to estimate aboveground biomass. One is an allometric approach to estimate biomass based on a given tree diameter at breast height (DBH) and the other one is using wood volume to calculate biomass (see Brown [2002] for an extensive review).

The most common approach used to estimate aboveground biomass is through allometric equations which estimate the aboveground tree biomass (AGB) in relation to a given tree diameter value (D), usually measured at 1.3 m,

\[ AGB = aD^b \]  

where \( a \) and \( b \) are scaling coefficient and scaling exponent, respectively, which vary with species, sites and age. Usually \( a \) and \( b \) are obtained through empirical regression of log-transformed data of biomass and diameter pairs measured from destructively sampled trees. For example, Jenkins et al. [2004] and Ter-Mikaelian and Korzukhin [1997] list \( a \) and \( b \) values from species in the United States. This approach is laborious and time-consuming.

Many studies have investigated how \( a \) and \( b \) are related to stand structure and age. West et al. [1999] used the fractal properties of tree branching networks and developed a geometric model of tree structure which predicts aboveground biomass from tree diameter and estimates the exponential factor \( b = 8/3 = 2.67 \) independently of species, site and age. Enquist et al. [1999] suggested that \( a \) is related to wood density; however, Chambers et al. [2001] found \( a = 0.1 \).

Many other studies found that \( a \) and \( b \) varies by species, sites, and even stand ages. Ketterings et al. [2001] found that \( a \) and \( b \) vary between sites and suggested that \( b \) can be estimated from the site-specific relationship between height and diameter, \( H = kD^{b-2} \) and \( a = r\rho \), where \( r \) is expected to be relatively stable across sites and \( \rho \) is the wood density. Zianis and Mencuccini [2004] used fractal geometry and found that the scaling exponent \( b \) is between 2 and 3 and should be obtained based on tree height/size (or stand age). Their result is consistent with the finding by Ketterings et al. [2001] that \( b \) is related to tree height. Their study also found a statistically significant difference between theoretical and empirical values of the allometric exponent \( b = 2.3679 \) versus \( b = 2.67 \). Pilli et al. [2006] analyzed 49 data sets of different species and found that \( b \) is also related to tree stand age, species and site and \( a \) is correlated to wood density. All these studies indicate that tree height/age based allometric equations are possible especially if such measures can be derived from lidar.

The allometric approach to estimate aboveground biomass requires measurement of the diameter for each individual trees and the availability of allometric equation that are appropriate for each individual tree species. Developing allometric equations for each individual species can be very difficult particularly in tropical forests. Chave et al. [2005] used a single pan-tropical wood volume-based approach to estimate aboveground biomass at plot level. The basic hypothesis is that aboveground biomass for each tree is proportional to wood volume (product of height and basal area) [Chave et al., 2005]:

\[ AGB = F \cdot \left( \rho \cdot \pi \cdot \frac{D^2}{4} \cdot H \right)^{\beta} \]  

where \( H \) is tree height, \( \rho \) is the wood density (dry weight per unit volume in g/cm\(^3\)) and \( \beta < 1 \). \( F \) is a constant factor depending on tree taper. For broadleaf species, \( F = 0.06 \) [Chave et al., 2005]. Their study demonstrates that the inclusion of wood density and wood volume improves biomass estimates.

However the wood volume-based approach is not often used due to the difficulty to obtain height measurements...
3. Site and Data Description

3.1. Study Sites

[11] This study used data collected in temperate forests in New England. Our study sites include many plots in Harvard Forest (HF), MA, Bartlett Experimental Forest (BEF), NH, and the Forest Ecosystem Research site in Howland, ME (see Figure 1). HF is a 60–70 year old mixed deciduous forest. The stand is in the transition hardwoods-white pine-hemlock zone [Spurr, 1956], and is composed mainly of red oak, red maple, yellow birch, white birch, beech, white pine, and hemlock. BEF includes old-growth northern hardwoods with beech, yellow birch, sugar maple, and eastern hemlock. The natural stands in Howland, ME, is the boreal transitional forest and consists of spruce-hemlock-fir, aspen-birch, and hemlock-hardwood mixtures. Our study sites cover large forested regions with various climate conditions and should be a good representation of the forests in the northeastern United States.

3.2. Data Sets

[12] Data sources used in this study include: vegetation structure field data collected in 2003 and 2007, LVIS data collected in 2003 and ground lidar data collected in 2007. There is a four year difference between these data sets. We observed that growth, for example, at the sparse shelterwood site in Howland, Maine, plays a role in LVIS and modeled waveform comparison (W. Yang et al., unpublished manuscript, 2009). However, a field DBH comparison in Howland, MN from 1989 to 2003 does not show much growth. A LVIS and EVI height comparison also does not show any impact of growth so growth was ignored in this study.

3.2.1. Field Measurements

[13] Two sets of ground tree structure field data were used in this study. The first includes 6 stands (28 plots with a 20 m/25 m radius) of tree data collected in summer, 2007 in three forest sites (Harvard Forest, MA, Bartlett Experimental Forest, NH, and the Forest Ecosystem Research site in Howland, ME). The second is a complete stand map collected in 2003 in the Forest Ecosystem Research site in Howland, ME (see Figure 1 for the locations of our study plots and the stem map area). These two data sets were collected over relatively flat terrains.

3.2.1.1. The 2007 Field Data

[14] The 2007 ground data include vegetation structure data collected in six stands with two stands selected in each forest (Harvard Forest, MA, Bartlett Experimental Forest, NH, and the Forest Ecosystem Research site in Howland, ME) and each stand represents the unique canopy cover composition commonly seen in this region, e.g., deciduous forest dominated, conifer forest dominated, or partially harvested forest. Our sites were purposely selected to encompass most forest cover types available in New England. Within each stand, tree data were collected in 5 plots (except only 3 plots near the tower site in Howland). Each is a circle area with a 20 m radius (except for 25 m in the hardwood site in Harvard Forest due to lower tree density). One plot is at the stand center, while the rest are 30–50 m away from the center at each of the 90 degree angles (see the bottom right square in Figure 1). Diameter at Breast Height (DBH), tree species, tree status, and crown status were sampled for trees with DBH ≥ 10cm, smaller trees (DBH between 3 and 10 cm) were measured within 10 m of plot center. In addition, 10 trees were selected in each plot for additional tree height and crown size measurements.

3.2.1.2. The 2003 Stem Map Data

[15] The 2003 stem map in the Forest Ecosystem Research site in Howland, ME is a 200 m × 150 m rectangular area (except for a 30 m × 30 m plot) near the experimental forest’s flux tower (see Figure 1). Tree species, DBH, tree status (alive, dead), and crown status (dominance of the tree crown) were sampled for a total amount of ~7800 trees in 2003. Similar data collected in 1989 were used to verify our tree height allometric equations. Instead of using the originally designed rectangular plot for 2003 data, a circular plot (20 m in radius) was used to be consistent with the other data sets used in this study (see the top right square in Figure 1).
Relationship between single tree species aboveground biomass with wood volume (product of basal area (BA) and tree height (H)) for major tree species in the northeastern United States. BA and H are calculated with DBH ranging from 10 to 68 cm with a 2 cm increment.

3.2.2. Lidar Data

Two sets of lidar data collected at our study sites were used in our analysis. One is the LVIS, or the “Laser Vegetation Imaging Sensor” and the other is a hemispherical scanning ground lidar, the Echidna® validation instrument (EVI) [Jupp et al., 2005, 2009; Strahler et al., 2008].

3.2.2.1. LVIS Data

In summer 2003, LVIS was flown over several regions in New England [Blair et al., 2006]. The LVIS data used in this study were acquired on 18–20 July 2003 in Bartlett, NH, 26 July 2003 in Howland, ME, and 20 July 2003 in Harvard, MA. The flights altitude are ~10 km, producing footprint size of ~20 m, separated by roughly the same distance both along and across track.

LVIS is an airborne laser altimeter system, designed, developed and operated by the Laser Remote Sensing Laboratory at NASA’s Goddard Space Flight Center. The onboard laser generates Gaussian shaped optical pulses at a wavelength of 1064 nm [Blair et al., 1999]. The vertical sampling resolution of LVIS is 30 cm (1 ns). LVIS footprint sizes (diameter) typically vary between 10 to 25 m depending on the mission flight altitude. To be consistent with other data sets, the vegetation structure parameters derived from LVIS were also binned in each plot where ground structure data were collected. LVIS has a published geolocation accuracy of 1 m or less for this data set [Blair et al., 2006]. Our analysis indicates that LVIS geolocation data match very well with IKONOS data.

3.2.2.2. Ground Lidar Data

Concurrently with the ground structure data collection, a hemispherical scanning ground lidar, the Echidna® validation instrument (EVI) [Jupp et al., 2005, 2009; Strahler et al., 2008] was deployed in the same sites as our ground data collection. One set of EVI scan data were collected for each plot in each stand in 2007. EVI, developed by CSIRO Australia as part of its canopy lidar initiative, is a ground-based, upward hemispherical-scanning, full waveform digitized, terrestrial lidar instrument and allows acquisition of vegetation canopy structure data, including height, basal area, and stem counts, as well as accurate information on standing woody and green biomass for carbon balance inventory and mapping [Jupp et al., 2005; Strahler et al., 2008]. EVI utilizes a horizontally positioned laser that emits pulses of near-infrared light at a wavelength of 1064nm. Pulses are emitted at a rate of 2 kHz. Plot and stand level DBH, basal area, tree height in addition to foliage profiles were derived from each plot in our study sites using the EVI data.

4. Analysis

4.1. Field Data Analysis

Based on the vegetation structure parameters collected in 2003 and 2007, we examined the relationship between aboveground biomass for each individual tree and at the plot level for each location with height, DBH, basal area and wood volume. Aboveground biomass for each individual tree was calculated based on DBH-based allometric equations and plot level biomass was calculated as the sum of individual tree biomass within each plot. Vegetation height was not measured for each single tree in the field; however, height was estimated from DBH using DBH-height allometric equations and was evaluated using height and DBH measurements from the same site.

The allometric equations for biomass were chosen from the literature where the ground data collection best approximates the conditions in our study sites [Tritton and Hornbeck, 1982]:

\[ \ln(AGB) = a + b \ln D \]  

where the coefficients $a$ and $b$ vary with species. The allometric equations for height were from Albani et al. [2006]:

\[ H = 1.3 + b_{th} \cdot (1 - e^{b_{2h}D}) \]

where coefficients $b_{th}$ and $b_{2h}$ vary with plant functional types (PFTs), which are differentiated by their leaf physiology, allometry, mortality and dispersal. We verified that the above allometric equation provides good estimates for tree height for our study area by comparing tree height estimates with the tree height estimates in 1989 ($R = 0.87$, RMSE = 2.3 m, for nearly 7800 trees).

4.1.1. Results for Individual Tree Species Based Analysis

Figure 2 compares the relationship between aboveground biomass with wood volume approximated as the product of basal area and tree height for major tree species in New England. The analysis was separated by two groups: conifer/softwood tree species and deciduous/hardwood tree species. Figure 2 shows that for trees within 10 to 68 cm DBH
range, aboveground biomass is almost linearly related to the wood volume except for birch species. Previous studies on allometric equations have shown that the coefficients are dependent on species, possibly on stand age and structure [Pilli et al., 2006; Zianis and Mencuccini, 2004]. As indicated in equation (2), the coefficients (slopes) of the linear relationship between aboveground biomass and wood volume are related to the wood density. Softwood and hardwood show a large difference in wood density, resulting in two distinguished groups as shown in Figure 2. For all hardwood tree species in our study area, except for birch, the biomass does not vary much between individual species. For all softwood tree species, except for northern white-cedar tree, biomass estimates all group together. From this, we conclude that the aboveground biomass and wood volume relationships are vegetation type dependent. This result is consistent with the conclusion from Chave et al. [2005] that had suggested that knowing wood density improves aboveground biomass estimates.

### 4.1.2. Results for Plot Level Analysis

The plot level analysis was based on the two data sets collected in 2003 and 2007. These two data sets have similar geolocation accuracy, and topography and sampling strategy. Data collected in 2003 were binned into 4 × 6 circular plots with 20 m radius, the same plot size as for the 2007 data. Both data sets have about 5 m geolocation accuracy.

Figure 3 compares the relationships among plot level aboveground biomass with various tree structure parameters including quadratic mean DBH (DBH_qm), basal area, quadratic mean tree height (H_qm), top 10% tree height, and wood volume (the product of basal area and tree height) for data collected in 2003 and 2007. It shows that plot level aboveground biomass is related to height metrics and to mean DBH structure parameters with root mean square error (RMSE) ranging around 1.1–4 kg/m² and coefficient of determination $R^2$ ranging from 0.4 to 0.7. However, there are five outliers in biomass and height/DBH relationships (Figures 3a–3d). These outliers are from the shelterwood stand in Howland, MN, where vegetation is very sparse. Quadratic mean of DBH (DBH_qm), quadratic mean tree height (H_qm) and top 10% tree height were large; however, total aboveground biomass was low in these sites.

Over all structure parameters, wood volume (calculated as the product of basal area and top 10% or quadratic mean of tree height) is the best biomass predictor when the analysis was based on vegetation type with RMSE from 1.1 kg/m² to 1.6 kg/m² and $R^2$ from 0.74 to 0.95. Conifer plots show a better relationship ($R^2 = 0.95$) between biomass and wood volume than deciduous ($R^2 = 0.74$). Overall wood volume is a good predictor of plot level aboveground biomass when the forest type is known. The method used in this study is different from Chave et al. [2005] in that their study developed a regression model to directly relate aboveground biomass with diameter and height for each individual tree at a plot level. Our study used basal area and top 10% tree height at the plot level which is more closely related to lidar measured tree height. Our conclusion is consistent with theirs in that woody density and wood volume a good estimate of biomass and our study also indicates that differentiating at least hardwood and softwood vegetation type and utilizing wood volume are critical for biomass estimate.

Figure 3. Relationships between plot level aboveground biomass and different tree structure parameters for all study plots in New England. Solid circles are for deciduous plots, and open circles are for conifer plots.
Spatial relationship between LVIS footprint (open circles) and plot range (solid circles). Plot level LVIS products were obtained by averaging LVIS data for footprint centers located within a 5 m radius circular centered at each plot center.

4.2. LVIS Data Analysis

The lidar data analysis focused on examining the physical meaning of vegetation height metrics and the link between plot level aboveground biomass with height metrics, vegetation cover and structure parameters combining height and gap fraction. This analysis aids us in understanding the error sources in lidar estimated biomass and helps us investigate how much improvement in biomass estimate we can expect by adding additional vegetation structure parameter compared to using height metrics alone.

4.2.1. LVIS Data Processing

To extract vegetation structure parameters for each plot from LVIS data, LVIS footprint-level structure parameters within each plot were binned to obtain plot level LVIS structure parameters. Figure 4 shows the spatial relationship between LVIS footprint and the plot size. To extract LVIS data for the corresponding plot, we averaged all LVIS data occupying an area slightly larger than plot range to reduce geolocation uncertainty. LVIS data (height and coverage) with footprint centers located within a 5 m radius circle centered at each plot center are averaged (using quadratic mean for height) for all the 2003 and 2007 plots (see Figure 4).

Three structure parameters derived from LVIS were used in this study: The first, RH100, calculated as the distance of the top of vegetation returns and the peak of last Gaussian pulse (the ground returns); the second, RH50 as the height of medium energy occurs relative to the last ground returns [Drake et al., 2002a]. Both height metrics are often used to estimate aboveground biomass [Drake et al., 2002a, 2002b, 2003]; and the third, vegetation cover, was calculated from the full lidar waveforms based on the method described below.

Vegetation cover can be extracted from accumulated lidar returns from canopy and ground. However, it is well known that the lidar returns are affected by the canopy and background reflectivity ratio. Based on the basic lidar equations [Ni-Meister et al., 2001], vegetation cover can be directly estimated from accumulated vegetation and background returns, \( R_v \) and \( R_g \) (see Figure 5 on how to obtain \( R_v \) and \( R_g \) from waveforms), and the ratio of canopy volume backscattering coefficient (\( \rho_v \)) and the background reflectivity (\( \rho_g \)), \( \rho_v/\rho_g \):

\[
C = \frac{1}{1 + \frac{\rho_v}{\rho_g} \frac{R_g}{R_v}}
\]

[30] Accumulated vegetation and background returns, \( R_v \) and \( R_g \) for each footprint can be easily retrieved from original lidar energy returns. In this study we propose an approach to estimate the reflectivity ratio, \( \rho_v/\rho_g \), using accumulated canopy and background laser returns, \( R_{v1} \) and \( R_{v2} \), \( R_{g1} \) and \( R_{g2} \) from two adjacent footprints.

[31] The two basic vegetation lidar equations are [Ni-Meister et al., 2001]:

\[
\begin{align*}
R_g &= J_0 \rho_g (1 - C) \\
R_v &= J_0 \rho_v C
\end{align*}
\]

where \( J_0 \) is the beam irradiance of the lidar; \( C \) is the canopy cover. Set \( R'_{g} = R_g / J_0 \) and \( R'_{v} = R_v / J_0 \), the above equations become:

\[
\begin{align*}
R'_{g} &= \rho_g (1 - C) \\
R'_{v} &= \rho_v C
\end{align*}
\]

So,

\[
\frac{R'_{g}}{R'_{v}} + \frac{R'_{v}}{R'_{g}} = 1 \\
1 - C = \frac{R'_{g}}{R'_{v} \rho_v}
\]

With the assumption of constant \( \rho_g \) and \( \rho_v \) from two neighboring footprints, equation (8) gets \( \frac{\rho_v}{\rho_g} \) as a function of ground and canopy returns, \( R_{g1} \) and \( R_{g2} \); \( R_{v1} \) and \( R_{v2} \) from two neighboring footprints:

\[
\begin{align*}
\rho_v &= \frac{R_{v2} R_{v1} - R_{v1} R_{v2}}{R_{v2} - R_{v1}} \\
\rho_g &= \frac{R_{g1} R_{g2} - R_{g2} R_{g1}}{R_{g2} - R_{g1}} \\
\frac{\rho_v}{\rho_g} &= \frac{(R_{v1} - R_{v2})}{(R_{g2} - R_{g1})}
\end{align*}
\]

Figure 4. Spatial relationship between LVIS footprint (open circles) and plot range (solid circles). Plot level LVIS products were obtained by averaging LVIS data for footprint centers located within a 5 m radius circular centered at each plot center.

Figure 5. Illustration of accumulated vegetation returns (Rv) and ground returns (Rg) from a typical lidar waveform.
This method has been evaluated and used to derive corrected lidar waveforms and canopy gap probability profiles from LVIS data (W. Yang et al., unpublished manuscript, 2009). W. Yang et al. (unpublished manuscript, 2009) found that the corrected VLIS waveforms match better with modeled waveforms than the uncorrected ones. Their study also demonstrated that this ratio is larger in deciduous forests sites than in coniferous forests.

4.2.2. LVIS Data Results

Figure 6 compares RH100 and in situ measured vegetation heights calculated as mean top 10%, maximum and quadratic mean tree height at LVIS footprint, plot and stand scales. At the LVIS footprint level (20 m), RH100 is described best as the maximum tree height; however, there are a few outliers showing significant underestimation of maximum tree height. They are likely the results that taller tress are located near the edge of the footprint, where laser energy is weak; therefore there is not enough backscattering from the tree tops to record distinguishable returns [Hyde et al., 2005]. Our result is comparable to other results by Lefsky et al. [2001] and Hyde et al. [2005].

RH100 overestimates the quadratic mean of tree height, but compares reasonably well with top 10% tree height at the footprint level. However, both top 10% and the quadratic mean of tree height has much smaller variations than RH100, partly due to averaging often leading to reduced variations. The other reason could be that in situ tree heights were not direct ground measurements, but were calculated based on allometric equations and such calculation loses its nature variation.

At the plot and stand levels, RH100 and top 10% height and quadratic mean height were all averaged values. RH100 slightly overestimates top 10% tree height, particularly for tall trees; however, both variables show similar spatial variations. RH100 underestimates maximum tree height in most cases particularly for conifer sites. The underestimate
Figure 7. Comparison of plot level RH50 with vegetation cover, ground quadratic mean, and top 10% height and relationship between RH50 and the product of RH100 and LVIS vegetation cover for all study plots (solid circles for deciduous plots and open circles for conifer plots) in New England.

is the result that RH100s at both plot and stand levels, are averaged value and maximum tree heights are not. It appears that crown shape influences RH100 estimates and in conifer forests, skinny crown often reflect too little signal to detect the tree tops, leading to underestimation of tree height for conifer forests. RH100 overestimates quadratic mean of tree height. The overestimation comes from that RH100 is an averaged value of taller trees and quadratic mean includes also small trees.

To better understand RH50, Figure 7 compares RH50 with top 10% and quadratic mean canopy height, canopy cover and canopy volume estimated as the product of vegetation height (RH100) and vegetation cover. It shows that RH50 is related well to vegetation cover, quadratic mean and top 10% averaged vegetation height with $R^2$ greater than 0.7 and RMSE ranging from 1.7 m to 2.0 m. However, RH50 has the closest relationship with canopy volume (RH100* cover) with a coefficient of determination $R^2 = 0.92$ and RMSE = 1 m, indicating that RH50 is a better representation of canopy volume. RH50, the height where 50% of waveform energy occurred, is sensitive to changes in canopy foliage structure, for which canopy height, density, cover and vertical foliage profiles are all determining elements. The value of RH50 is lower in forest stands with lower canopy cover, when more energy can reach the ground [Drake et al., 2002a, 2002b]. Conversely, RH50 is increased in high canopy covered forest. Figure 7 indicates that RH50 is about 60–70% of canopy volume value. Additional analysis (not shown) found that the relationship of RH50 with crown volume (calculated based on crown size and stem density) and total foliage area (foliage area volume density*crown volume), appears to be quite complicated and crown shape may play a role. Our understanding of why RH50 is more closely related to height*cover than to crown volume or foliage area is that RH50 is a combination of height and canopy gap fraction. A physical model like the Geometric Optical and Radiative Transfer (GORT) [Ni-Meister et al., 2001, 2010; Yang et al., 2010] is required for further analysis and the results will be presented in a separate study.

To investigate the relationships of plot level aboveground biomass with RH100, RH50, vegetation cover and canopy volume, Figure 8 compares aboveground biomass with RH100, RH50, vegetation cover and the combination of two. RH100, RH50 and vegetation cover are all good predictors of biomass. A previous study has shown RH50 is a good predictor of aboveground biomass [Drake et al., 2002a, 2002b]. Our LVIS data analysis shows that RH100*cover ($R^2 = 0.85$ and RMSE = 2.4 m) and RH50*cover ($R^2 = 0.87$ and RMSE = 2.2 m) perform similarly or even better than RH50 ($R^2 = 0.84$ and RMSE = 2.5 m). However, Figure 8 also indicates that in deciduous plots, lidar was not very successful in identifying biomass variations.

To understand why RH100*cover and RH50*cover are good predictors of aboveground biomass, Figure 9 compares RH100*cover and RH50*cover with wood volume (basal area * top 10% height). It shows that both metrics are closely related to wood volume, particularly for conifer plots. For deciduous plots, RH100*cover and RH50*cover seem saturated and do not change much with wood volume, indicating that lidar might have difficulties in identifying biomass variations in deciduous forests. This result concurs with Nelson et al. [2007].

High correlation of RH50 with RH100*cover suggests that RH50 is directly related to canopy height and gap fraction, thus wood volume. This finding indicates that biomass should be better described by lidar measured height and canopy gap fraction than height alone. Although nadir-looking above canopy lidar does not provide wood volume information, aboveground biomass can still be estimated with good accuracy from lidar data with structure parameters that combine height and gap fraction, such as RH50, RH100*cover and RH50*cover.

4.3. Analysis of Hemispherical Scanning Ground-Based Lidar Data

Our ground data analysis demonstrates that wood volume (estimated as the product of vegetation height and basal area) is an excellent biomass predictor when knowing forest type as hardwood/deciduous or softwood/coniferous. However, nadir pointing lidar does not provide a direct estimate of wood volume except for tree height. One approach to estimating wood volume is to use a full-digitizing hemispherical-scanning below-canopy lidar like the Echidna® validation instrument (EVI). As demonstrated by Jupp et al. [2009] and Strahler et al. [2008], EVI provides accurate plot level mean DBH, basal area, stem density, foliage profile and tree height structure information. Therefore wood volume can be calculated from EVI basal area and height measurements. By using EVI-basal area and height measurements in New England to estimate wood volume and then aboveground biomass at plot and stand levels we can and would avoid using allometric equations to estimate aboveground at plot and stand levels.
Our first analysis is to compare EVI height with different field measured height metrics. A similar analysis was conducted as Figure 6. EVI height was compared to mean top 10%, maximum and quadratic mean tree heights at plot and stand levels (Figure 10). Similar to RH100, EVI height at both plot and stand levels matches reasonably well with top 10% tree height with slight overestimation. EVI height slightly underestimates maximum tree height at both plot and stand levels. However, the difference from LVIS RH100 is that EVI height accuracy is less crown shape dependent as it is upward looking and crown shape may not play a strong role on height estimate. Similar to RH100, EVI height overestimates quadratic means of tree height.

Further analysis includes comparison of EVI and LVIS canopy height, and EVI DBH and basal area with field data (see Figure 11, top). EVI-height agrees reasonably well with LVIS height with slightly larger LVIS height. EVI basal area matches well with the field measurements.

Finally, aboveground biomass was compared to wood volume estimated by EVI or a fusion of EVI and LVIS at the plot level (Figure 11, middle) and the stand level (Figure 11, bottom). Figure 11 demonstrates that plot level aboveground biomass is well related to wood volume particularly for conifer plots. Although with low samplings, aboveground biomass was better estimated at the stand level for both conifer and deciduous forests ($R^2 = 0.97$ and RMSE = 0.3 kg/m$^2$ for conifer stand and with $R^2 = 0.92$ and RMSE = 1.6 kg/m$^2$ for deciduous stand) than at the plot level. Our analysis demonstrates that hemispherical-scanning ground lidar data from EVI provides high-quality estimates of wood volume, and has a potential to provide accurate estimates of aboveground biomass. This analysis demonstrates the potential strength in biomass estimate at large scale using EVI.

5. Conclusions and Discussion

This study analyzed both field and lidar-based vegetation structure measurements collected in New England at different scales to investigate if additional vegetation structure parameters besides height are highly related to aboveground biomass and will improve aboveground biomass estimate. Field data analysis shows that aboveground biomass for each individual tree and at the plot level is closely related to wood volume and wood density, depending on vegetation type (conifer/softwood or deciduous/hardwood forests). Wood volume at plot level is best approximated by...
the product of basal area and top 10% tree height. Our study indicates that using wood volume and vegetation type rather than tree height alone has the potential to improve biomass estimates.

This result concurs with previous allometric equation-based biomass studies by Pilli et al. [2006], Zianis and Mencuccini [2004], and Chave et al. [2005]. Their studies found that aboveground biomass is closely related to wood volume and wood density. However, our analysis method is different from theirs. Previous analysis calculated the statistics based on individual trees. Our analysis calculates the statistical tree structure parameters first and relates the aboveground biomass with tree structure statistical parameters at large scales. For example, our wood volume at plot level was approximated as the product of mean basal area and top 10% tree height. A similar study at plot level [Chave et al., 2005] did regression analysis to relate biomass with DBH and height for each individual tree.

Previous airborne lidar analysis found that RH50 is a good predictor of aboveground biomass [Drake et al., 2002a, 2002b, 2003]. Our LVIS data analysis shows that structure parameters combining canopy height and gap fraction such as RH100*cover and RH50*cover perform similarly or even better than RH50. High correlations between these structure parameters with wood volume, particularly for conifer forests explain why they are good predictors of aboveground biomass. High correlation between RH50 and RH100*cover suggests RH50 is a better predictor of biomass than RH100. Above-canopy lidar does not provide a direct measure of wood volume; however, structure parameters that combine height and gap fraction, such as RH50, RH100*cover and RH50*cover are highly related to wood volume. Aboveground biomass can therefore still be estimated with good accuracy from lidar data. Finally our analysis also indicates that lidar might not be very successful in identifying biomass variations in deciduous forests, which is consistent with discussions by Nelson et al. [2007]. Further research is necessary to confirm this finding.

Our field data analysis indicates that wood volume (estimated as the product of vegetation height and basal area) is an excellent biomass predictor especially with known dominating forest type to be either deciduous or coniferous forests. Forest type information can be easily extracted from the existing land cover maps derived from optical remote sensing data. Our study therefore implies that the fusion of lidar and forest type information from optical remote sensing will provide better aboveground biomass estimates than lidar alone.

Further, allometric equations to derive aboveground biomass in this study were derived in New England. These equations can vary from site to site due to difference in soil, climate and topography. The best approach would be using allometric equations derived specifically for each site. However, these equations are often not available. The error associated with using one set of allometric equations for all

Figure 10. Comparison of EVI height with field measured tree height metrics. (left) Top 10% mean tree height, (middle) maximum tree height, and (right) quadratic mean tree height at plot (0.12 ha) and stand (~1 ha) scales (solid circles for deciduous plots/stands and open circles for conifer plots/stands) in New England.
stands might lead to some bias in our relationship between biomass with structure parameters. Our findings need to be further tested in other regions.

In particular, this study uses ground and lidar data collected over flat terrain. The relationships between aboveground biomass and vegetation structure parameters are complicated by adding topography [Lefsky et al., 2005, 2007; Hyde et al., 2005; Harding and Carabajal, 2005]. Surface topography has been a known factor in aboveground biomass production [Whittaker et al., 1974]. Uncertainties in LVIS height metrics estimates from lidar also increase with terrain slope [Blair et al., 2006]. Correction of slope effect on height metrics is required to derive accurate vegetation height structure parameters [Lefsky et al., 2005, 2007; Harding and Carabajal, 2005; Hyde et al., 2005; Yang et al., 2010]. A recent study in montane ecosystems indicates that aboveground biomass is more closely related to RH75 than RH50. As this conclusion is different from what was found by Drake et al. [2002a, 2002b, 2003], topography might be a factor for such a relationship. In montane regions, the wood volume might be better related to RH75 than RH50 as found in our study. Our future work includes assessing the relationship between biomass and vegetation structure parameters over sloping terrains by adding the surface topography effect in aboveground biomass estimates from lidar.

Finally, this study and previous work all indicate that deriving height based allometric equations for aboveground biomass estimate is possible. These types of allometric equations will also be dependent on stand age [Ketterings et al., 2001; Pilli et al., 2006; Zianis and Mencuccini, 2004] and optical remote sensing information on stand age [Song et al., 2007; Liu et al., 2008] may have to be fused with lidar remote sensing to obtain optimal results.

Acknowledgments. This work was funded by NASA under grant NNG06G19G. We are also grateful to David D. L. Juppa from CSIRO Marine and Atmospheric Research and Glen Newnham from CSIRO Forest Biosciences for providing the EVI instrument and their effort with data collection and processing. To Xiaoyuan Yang and Feng Zhao for their help in ground data collection and processing EVI and ground data, and to Zhouxun Wang, Jingling Zhang, Mitchell Schull, and Miguel Roman-Colon, for ground data collection.

References


J. B. Blair and K. J. Ranson, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

S. Lee and W. Ni-Meister, Department of Geography, Hunter College, City University of New York, 695 Park Ave., New York, NY 10021, USA.

C. Schaaf, A. H. Strahler, C. E. Woodcock, and T. Yao, Department of Geography and Environment, Boston University, Boston, MA 21005, USA.

G. Sun, Department of Geography, University of Maryland, College Park, MD 20742, USA.