Lidar remote sensing of above-ground biomass in three biomes

MICHAEL A. LEFSKY*, WARREN B. COHEN†, DAVID J. HARDING‡, GEOFFREY G. PARKER§, STEVEN A. ACKERY and S. THOMAS GOWER**

*Oregon State University, Forest Sciences Laboratory, 3200 SW Jefferson Way, Corvallis, OR 97331, U.S.A. E-mail: lefsky@fsl.orst.edu; †USDA Forest Service, Forest Sciences Laboratory, 3200 SW Jefferson Way, Corvallis, OR 97331, U.S.A.; ‡NASA Goddard Space Flight Center, Geodynamics Branch and Mail Code 921 Greenbelt, MD 20771, U.S.A.; §Smithsonian Environmental Research Center, PO Box 28, Edgewater, MD 21037, U.S.A.; ¶National Park Service, 909 First Avenue, Seattle, WA 98104, U.S.A.; **Department of Forest Ecology and Management, University of Wisconsin, Madison, WI 53706, U.S.A.

ABSTRACT
Estimation of the amount of carbon stored in forests is a key challenge for understanding the global carbon cycle, one which remote sensing is expected to help address. However, estimation of carbon storage in moderate to high biomass forests is difficult for conventional optical and radar sensors. Lidar (light detection and ranging) instruments measure the vertical structure of forests and thus hold great promise for remotely sensing the quantity and spatial organization of forest biomass. In this study, we compare the relationships between lidar-measured canopy structure and coincident field measurements of above-ground biomass at sites in the temperate deciduous, temperate coniferous, and boreal coniferous biomes. A single regression for all three sites is compared with equations derived for each site individually. The single equation explains 84% of variance in above-ground biomass (P < 0.0001) and shows no statistically significant bias in its predictions for any individual site.

Key words above-ground biomass, biomass measurement, organization of forest biomass. In this study, we compare carbon storage, global carbon cycle, forest biomass, interbiome comparison, lidar remote sensing, SLICER sensor.

INTRODUCTION
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 changing that cycle, and for monitoring mitigation efforts that rely on carbon sequestration through reforestation. Remote sensing has been a key technology in existing efforts to monitor carbon storage and fluxes (Cohen et al., 1996; Running et al., 1999), and has been identified as a probable tool for monitoring compliance with treaties such as the Kyoto protocol (Ahern et al., 1998). However, direct estimation of carbon storage in moderate to high biomass forests remains a major challenge for remote sensing. While remote sensing has had considerable success in measuring the biophysical characteristics of vegetation in areas where plant canopy cover is relatively sparse, quantification of vegetation structure where leaf area index (LAI) exceeds three has been less successful (Waring et al., 1995; Carlson & Ripley, 1997; Turner et al., 1999). High LAI forests, which generally have high above-ground biomass, occur in the boreal, temperate and tropical regions. These forests cover less than 35% of the Earth's terrestrial surface, yet account for 67% of terrestrial net primary productivity (NPP) and 89% of terrestrial biomass (Waring & Schlesinger, 1985). Given their prominent role in global biogeochemistry and the likelihood that these high productivity areas will be prime areas for carbon sequestration efforts, better characterization of high biomass forests using remotely sensed data is desirable. One promising technique is lidar.

Lidar instruments directly measure the vertical structure of forests by determining the distance between the sensor and a target through the precise measurement of the time between the emission of a pulse of laser light from the sensor and the time of detection of light reflected from the target. Waveform-recording lidar systems, such as the SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) device used in this
work (Blair et al., 1994; Harding et al., 1994, 2001) and the Laser Vegetation Imaging System (LVIS, Blair et al., 1999) measure the time-resolved amount of laser energy reflected from the many surfaces of a geometrically complex target. When this distribution of return energy, the lidar waveform, is measured over a vegetation canopy, it records the vertical distribution of light reflected back to the sensor from illuminated vegetation and soil surfaces from the top of the canopy to the ground surface. For forests, relating these waveforms to conventional, primarily non-spatial, measurements of forest structure, such as above-ground biomass and stand basal area, has been a primary research goal (Lefsky et al., 1999a,b; Means et al., 1999; Drake et al., 2002). In this study, we compare the relationships between lidar-measured canopy structure and coincident field measurements of above-ground biomass at sites in the temperate deciduous, temperate coniferous, and boreal coniferous biomes. A single equation, derived from regression analysis using data from all three sites, is compared with equations derived for each site individually. The goal of the work is a simplified method to estimate above-ground biomass at all three sites. Such a method could reduce the effort and expense required to develop global biomass estimates from satellite lidar data such as that to be acquired by the Vegetation Canopy Lidar mission (Dubayah et al., 1997). We focus on the estimation of above-ground biomass because it is related closely to above-ground carbon storage, and allometric equations for its estimation are readily available. While below-ground carbon pools are often as large or larger than above-ground storage, no existing remote sensing system can estimate their magnitude directly.

METHODS

Coincident field plots and lidar data were collected in three distinct sites in the boreal coniferous, temperate coniferous and temperate deciduous biomes. Estimates of above-ground biomass were calculated using established allometric equations and stem data collected in fixed or nested plots. Estimates of canopy height, canopy cover and a variety of canopy-density-weighted heights were calculated from the lidar data. Stepwise multiple regression analysis (IDL, Research Systems Boulder, Colorado, USA) was then performed to determine if general relationships could be established between lidar estimated canopy structure and field estimates of above-ground biomass at all three sites.

Study areas

Field data for the temperate coniferous plots were collected in and near the H.J. Andrews Experimental Forest, located on the west slope of the Cascade Range, in Oregon (~122.25 longitude 44.2 latitude, Van Clev & Martin, 1991). Douglas fir (Pseudotsuga menziesii) is the dominant species in these stands, contributing 90% of all basal area in young stands, and 64% in old-growth stands. Western hemlock (Tsuga heterophylla) is the second most important species and occurs mainly in later successions, contributing 29% of total basal area in old-growth stands (Lefsky et al., 1999a). Data from temperate deciduous plots were collected in and near the Smithsonian Environmental Research Center, located on the western shore of Chesapeake Bay, near Annapolis, MD (~76.55 longitude and 38.87 latitude). The plots are in mixed deciduous forest with an overstorey dominated by Liriodendron tulipifera (Brown & Parker, 1994). Plots for the boreal coniferous type were collected at the Northern Old Black Spruce (NOBS) study area established as part of NASA's BOREAS study (~98.48 longitude and 55.88° latitude); plot data were collected as part of the BigFoot study (Cohen & Justice, 1999). Major cover types at the site include muskeg, black spruce (Picea mariana) forest and wetlands; infrequent patches of jack pine (Pinus banksiana) and aspen (Populus tremuloides) also occur.

Field data collection

Existing publications describe the field data collections for the boreal coniferous (Campbell et al., 1999), temperate deciduous (Brown & Parker, 1994; Lefsky et al., 1999b) and temperate coniferous stands (Lefsky et al., 1999b). Generally, fixed or nested plots were used to tally stems and appropriate allometric equations were used to predict above-ground biomass. At the temperate coniferous site, 21 plots were established coincident with SLICER transects. At the temperate deciduous site, 112 plots were either taken from an existing collection of plots, or subset from an existing 32-ha stem map. At the boreal coniferous site, an existing set of 25 × 25-m field plots were used as a source of field data. At this site, the location of existing SLICER waveforms were compared to the locations of 107 field plots and any plot with more than five waveforms within its boundaries was considered as part of this analysis, for a total of 16 plots.

SLICER data collection and processing

SLICER data were collected from airborne platforms at the temperate coniferous, boreal coniferous and temperate deciduous sites in September 1995, July 1996 and September 1995, respectively. Lidar waveforms were collected over contiguous 10-m footprints for five lidar waveforms across the sampling swath. To estimate canopy height profiles (CHPs, the vertical distribution of foliage and woody surfaces) from the raw SLICER waveforms, Harding et al. (2001) adapted the transformation method developed by MacArthur & Horn (1969). The resulting CHPs serve as a common measurement of forest canopy structure at the three sites. One key factor in the CHP algorithm is a coefficient defined as the ratio of the
average reflectance (at nadir) of the ground and canopy at the laser wavelength (1064 nanometres). For the temperate deciduous and temperate coniferous sites, the ratio of ground and canopy reflectance is assumed to be 2.0. Use of this assumption has been supported by fieldwork comparing lidar estimates and field measurements of canopy cover at these sites (Lefsky, 1997; Means et al., 1999). At the boreal coniferous site, the existence of a high ground-level cover of herbaceous moss and fern species and a small dataset of coincident lidar and field measurements of cover imply that this ratio should be close to 1.0, the value used in calculations for this site.

Canopy structure indices used in this study were calculated for each plot from CHPs following the methods of Lefsky et al. (1999a). A variety of indices were selected from our previous work in the temperate sites and preliminary work with the boreal coniferous dataset. The indices are canopy cover, mean canopy height (MCH, the average height of the waveforms associated with a plot), maximum height (the maximum height of the same set of waveforms), MCH squared, and the mean canopy profile height, defined as:

\[ \text{Mean canopy profile height} = \frac{1}{h} \sum_{i=1}^{h} CHP(i) \]

where: CHP(i) is the normalized canopy height profile at height i, and h is the maximum height of the canopy.

The quadratic canopy profile height is defined as:

\[ \text{Quadratic canopy profile height} = \sqrt{\frac{1}{h} \sum_{i=1}^{h} (CHP(i))^2} \]

In addition to considering the height and canopy cover variables separately, the products of canopy cover and each of the height indices were also considered, as suggested by our initial experience with the boreal site. At the eastern deciduous site, 64 plots with lidar-estimated CHPs were supplemented with 48 plots with field-measured CHPs, to incorporate a greater range of stand conditions. Treating these two datasets as equivalent for the purposes of estimating above-ground biomass is supported by the findings of Lefsky et al. (1999a). Measurements of mean canopy height are not available for the field measured CHPs; a regression between quadratic mean canopy height and mean canopy surface height was developed using those plots with lidar measurements, and then applied to predict mean canopy height. Scatterplots of variable combinations were examined to ensure linearity.

RESULTS

Plot characteristics

Mean canopy height (MCH) at the sites follows the expected order, with boreal coniferous having the shortest maximum and mean MCH, temperate coniferous having the tallest, with the temperate deciduous site in the middle (Table 1). Values for canopy cover for the temperate deciduous site occupy a narrower range than either of the coniferous sites due to the high cover associated with even the youngest of these sites, and the absence of significant disturbance. Mean, minimum and maximum cover are lowest in the boreal coniferous plots, as a consequence of the low productivity of this site, and the juxtaposition of closed forest and open forest/muskeg conditions. Site maxima for above-ground biomass range from 58.5 mg ha\(^{-1}\) for the boreal coniferous plots to 1329.0 mg ha\(^{-1}\) for the temperate coniferous forest; the highest values for the temperate deciduous plots occupy an intermediate position. Figure 1 illustrates characteristic transects of lidar-measured canopy structure at each study site.

Correlation of canopy structure indices and above-ground biomass

Nearly all the canopy structure indices were significantly correlated with above-ground biomass (Table 2), with the exception of canopy cover for temperate deciduous plots. This is due probably to the narrow range of canopy cover conditions observed in those plots. Otherwise, there were few patterns in the correlations that were consistent between all three biomes. For the boreal coniferous site, the product of cover and several of the height indices performed better than the height indices alone. At the temperate deciduous site, the reverse was true, again due probably to the low range of canopy cover, and the resulting non-significant correlation between cover and biomass. At the temperate coniferous site there is no clear difference between the two sets of variables (heights alone and the products of height indices and cover

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canopy cover</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperate deciduous</td>
<td>0.853</td>
<td>0.607</td>
<td>0.938</td>
</tr>
<tr>
<td>Temperate coniferous</td>
<td>0.696</td>
<td>0.285</td>
<td>0.876</td>
</tr>
<tr>
<td>Boreal coniferous</td>
<td>0.312</td>
<td>0.168</td>
<td>0.472</td>
</tr>
<tr>
<td>Mean canopy height (m)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperate deciduous</td>
<td>28.6</td>
<td>9.7</td>
<td>39.5</td>
</tr>
<tr>
<td>Temperate coniferous</td>
<td>35.6</td>
<td>15.3</td>
<td>53.2</td>
</tr>
<tr>
<td>Boreal coniferous</td>
<td>7.3</td>
<td>2.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Above-ground biomass (mg ha(^{-1}))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperate deciduous</td>
<td>312.5</td>
<td>11.4</td>
<td>716.3</td>
</tr>
<tr>
<td>Temperate coniferous</td>
<td>602.0</td>
<td>135.6</td>
<td>1329.0</td>
</tr>
<tr>
<td>Boreal coniferous</td>
<td>29.9</td>
<td>0.0</td>
<td>58.5</td>
</tr>
</tbody>
</table>
Figure 2. Comparison of observed and predicted above-ground biomass (a) Observed above-ground biomass and predicted above-ground biomass for the different models. The predicted biomass was calculated using the non-linear regression equation:

\[ y = b_0 + b_1 x + b_2 x^2 \]

where \( y \) is the predicted biomass, \( x \) is the observed biomass, and \( b_0, b_1, b_2 \) are the coefficients estimated from the regression analysis.

Figure 3. Measurements of forest canopy structure made with NADIR.

The shaded area in the diagram represents the canopy cover, with darker shades indicating higher cover. The green line indicates the canopy height, while the black line represents the ground surface. The diagram also shows the distribution of dominant species within the forest, with species density increasing from the center to the edges of the forest.
Table 2 Correlation coefficients (Pearson’s r) between canopy indices and above-ground biomass from field measurements. See text for methods

<table>
<thead>
<tr>
<th>Canopy cover (%)</th>
<th>Boreal coniferous</th>
<th>Temperate deciduous</th>
<th>Temperate coniferous</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum height (m)</td>
<td>0.84</td>
<td>0.77</td>
<td>0.63†</td>
<td>0.37</td>
</tr>
<tr>
<td>Mean canopy height (m)</td>
<td>0.67†</td>
<td>0.79</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>Mean canopy height squared (m)</td>
<td>0.74††</td>
<td>0.79</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td>Mean canopy profile height (m)</td>
<td>0.70†</td>
<td>0.79</td>
<td>0.93</td>
<td>0.91</td>
</tr>
<tr>
<td>Quadratic mean canopy profile height (m)</td>
<td>0.78†</td>
<td>0.75</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>Cover x maximum height (m)</td>
<td>0.74††</td>
<td>0.80</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>Cover x mean canopy height (m)</td>
<td>0.83</td>
<td>0.74</td>
<td>0.92</td>
<td>0.84</td>
</tr>
<tr>
<td>Cover x mean canopy height squared (m)</td>
<td>0.87</td>
<td>0.51</td>
<td>0.92</td>
<td>0.66</td>
</tr>
<tr>
<td>Cover x mean canopy profile height (m)</td>
<td>0.84</td>
<td>0.78</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td>Cover x quadratic canopy profile height (m)</td>
<td>0.88</td>
<td>0.72</td>
<td>0.81</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Unless otherwise noted, all relationships are significant at $P < 0.0001$. † $P < 0.01$; † † $P < 0.001$; NS: not significant.

Table 3 Slope and intercepts of general biomass equation applied to each site individually, Observed = $B_0 + (B_1 \times$ Predicted). $P(b_0 \neq 0)$ indicates the probability that the intercept of an equation is not equal to zero, and $P(b_1 \neq 1)$ indicates the probability that the slope is not equal to one

<table>
<thead>
<tr>
<th>Intercept $(b_0)$</th>
<th>Slope $(b_1)$</th>
<th>$P(b_0 \neq 0)$</th>
<th>$P(b_1 \neq 1)$</th>
<th>Single equation $R^2$</th>
<th>Individual site equation $R^2$</th>
<th>Mean prediction (Mg ha$^{-1}$)</th>
<th>Standard error (Mg ha$^{-1}$)</th>
<th>% Standard error</th>
<th>Mean residual (Mg ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boreal coniferous</td>
<td>10.11</td>
<td>0.75</td>
<td>0.09</td>
<td>0.19</td>
<td>56%</td>
<td>76%</td>
<td>23.2</td>
<td>2.9</td>
<td>12.6</td>
</tr>
<tr>
<td>Temperate deciduous</td>
<td>11.10</td>
<td>0.93</td>
<td>0.62</td>
<td>0.29</td>
<td>65%</td>
<td>65%</td>
<td>322.7</td>
<td>8.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Temperate coniferous</td>
<td>61.55</td>
<td>0.98</td>
<td>0.30</td>
<td>0.81</td>
<td>87%</td>
<td>87%</td>
<td>552.9</td>
<td>29.4</td>
<td>5.3</td>
</tr>
<tr>
<td>All</td>
<td>-3.34</td>
<td>1.01</td>
<td>0.81</td>
<td>0.84</td>
<td>84%</td>
<td>N/A</td>
<td>323.0</td>
<td>7.5</td>
<td>2.3</td>
</tr>
</tbody>
</table>

$^1$ Standard error as a percentage of mean prediction.

Regression analysis

The correlation analysis identified the mean canopy height squared as the variable with the highest correlation with above-ground biomass for all sites considered together, resulting in the following equation:

$$AB = 0.378 \times MCH^2 \ (R^2 = 84\%, \ P < 0.0001)$$

where $AB$ is above-ground biomass (Mg ha$^{-1}$), and $MCH^2$ is mean canopy height (m) squared.

Analysis of the residuals indicated that the product of mean canopy height and cover had the highest correlation ($r = 0.18$) with those residuals, and was added to the equation, resulting in:

$$AB = 0.342 \times MCH^2 + 2.086 \times COVCHPX \ (R^2 = 84\%, \ P < 0.0001)$$

where $COVCHPX$ is the product of mean cover and mean canopy height.

Although the addition of the $COVCHPX$ variable does not improve the model’s $R^2$, its addition is statistically significant and does reduce the residuals associated with the boreal sites. The predicted values from this equation were regressed against the observed above-ground biomass values for each site separately, and the resulting regression lines were tested to see if they were significantly different from an identity line (Fig. 2). In all three cases, slopes were not significantly different from 1 and intercepts were not significantly different from zero (Table 3). Stepwise multiple regressions were also performed for each site individually, and the resulting $R^2$ values are presented in Table 3. Only in the case of the boreal coniferous site did the single, general equation predict considerably less of the overall variance than did the individual site equation. Standard errors, as a percentage of the mean predicted value, range from 2.5% at the temperate deciduous site to 12.6% at the boreal coniferous sites, although the small number of observations at the boreal site influences the higher value. The mean residuals of the
individual sites ranges from $-10.2$ mg ha$^{-1}$ for the temperate deciduous sites to $49.4$ at the temperate sites, but these biases are relatively small compared to the mean predicted value (3% and 8%, respectively). Of more importance is the bias associated with the boreal coniferous site, 6.8 mg ha$^{-1}$, which is nearly 30% of the mean predicted value, despite the non-significant results obtained in the test of the regression between predicted and observed values for this site. Potential causes of the discrepancy include the varying number of plots at each site and the use of least-square methods for fitting the regression lines, which may not perform optimally when dealing with dependent variables that vary over three orders of magnitude.

CONCLUSION
A single equation can be used to relate canopy structure, remotely sensed using lidar, to above-ground biomass in three distinctly different forested communities. Any suggestion that this result is applicable to forested systems generally must be considered preliminary. Tropical systems are not discussed at all, and it would be necessary to have replicated sites from opinion, is that it indicates that research into that hypothesis could be considered. The primary value of this work, in our view, is that it indicates that research into that hypothesis is reasonable—a conclusion that was not expected at the outset of this study. Forests of the type described in this paper cover 16% of the global land surface and 50% of the forested land surface (Waring & Schlesinger, 1985). If the relationship between forest canopy structure and above-ground biomass is as consistent as suggested in this study, then the estimation of global forest carbon storage, and the monitoring of its change over time, may be greatly simplified. Adoption of a modelling approach would further improve the confidence associated with a simplified relationship. Simple models, starting with the known allometric properties of plants, and incorporating competition for light and space, have already demonstrated that they can reproduce emergent community-level relationships (Enquist & Niklas, 2001). Such an approach should be adaptable to this problem, and could provide the necessary confidence to interpret the global dataset anticipated from the Vegetation Canopy Lidar (VCL) mission, with a minimum of additional fieldwork.

REFERENCES

### BIOSKETCHES

**M.A. Lefsky** is an assistant professor in the Forest Science department at Colorado State University, and emeritus co-director of the Laboratory for Applications of Remote Sensing of Ecology.

**W.B. Cohen** is a research forester with the USDA Forest Service and director of the Laboratory for Applications of Remote Sensing in Ecology.

**D.J. Harding** is a geoscientist with the Geodynamics branch of the Laboratory for Terrestrial Physics at NASA’s Goddard Space Flight Center.

**G.G. Parker** is a forest ecologist with the Smithsonian Environmental Research Center.

**S.A. Acker** is a Supervisory Botanist at Olympic National Park.

**S.T. Gower** is a professor in the Department of Forest Ecology and Management at the University of Wisconsin.