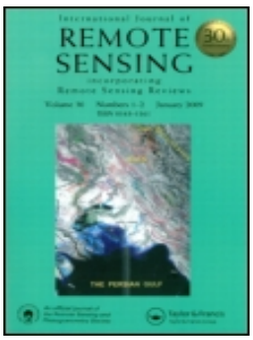

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

Estimation of above-ground forest biomass using metrics based on Gaussian decomposition of waveform lidar data

Wei Zhuang, Giorgos Mountrakis, John J. Wiley Jr. & Colin M. Beier

To cite this article: Wei Zhuang, Giorgos Mountrakis, John J. Wiley Jr. & Colin M. Beier (2015) Estimation of above-ground forest biomass using metrics based on Gaussian decomposition of waveform lidar data, International Journal of Remote Sensing, 36:7, 1871-1889, DOI: 10.1080/01431161.2015.1029095

To link to this article: http://dx.doi.org/10.1080/01431161.2015.1029095

International Journal of Remote Sensing, 2015
Vol. 36, No. 7, 1871–1889, http://dx.doi.org/10.1080/01431161.2015.1029095

Estimation of above-ground forest biomass using metrics based on Gaussian decomposition of waveform lidar data

Wei Zhuang[a], Giorgos Mountrakis[a]*, John J. Wiley Jr.[b], and Colin M. Beier[c]

[a]*Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA;* [b]*Department of Environmental and Forest Biology, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA;* [c]*Department of Forest and Natural Resources Management, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA*

(*Received 1 April 2014; accepted 10 January 2015*)

Large-footprint waveform light detection and ranging (lidar) data have been widely used in above-ground forest biomass estimation. Waveform metrics derived from basic statistics (e.g. percentile of energy) of the lidar waveform, such as canopy height and height of median energy, have been applied to biomass estimation in numerous studies. In this study, a set of metrics based on Gaussian decomposition (GD) results were developed and evaluated for forest above-ground biomass estimation using NASA's laser vegetation imaging sensor (LVIS) data. The GD metrics were designed to explicitly incorporate lidar intensity and vertical structures of canopy layers for biomass estimation. The proposed GD metrics used information related to the above-ground height of each Gaussian centroid and the Gaussian area index (GAI), where GAI is the area covered by a Gaussian function. Two types of novel GD metrics were developed: (1) percentile-height GAI metrics expressing the GAI summation of a subset of Gaussian centroids located within a certain percentile height range; and (2) height-weighted GAI metrics, a summation of GAIs of a waveform weighted by the corresponding heights of their Gaussian centroids. A biomass regression model was built by eight newly developed GD metrics using GAI information and five pre-existing GD-derived metrics that have not previously been used for biomass estimation. The performance of the regression model was then compared to another regression model using 12 previously published metrics (non-GD metrics). The Random Forests (RF) regression algorithm was employed for predicting biomass. The RF out-of-bag results indicated that above-ground biomass estimations using GD metrics achieved significantly better results than those derived from non-GD metrics for deciduous plots (19% lower root mean square error (RMSE), 25% higher coefficient of determination (R^2), and marginally better results in coniferous plots (4% lower RSME, 6% higher R^2). The combination of GD and non-GD metrics achieved comparable biomass estimation results to the model using exclusively GD metrics. GD metrics also showed strong correlation with forest attributes such as mean diameter at breast height (DBH) and stem density. This study contributes to the usage of GD results for accurate estimation of forest above-ground biomass in large-footprint lidar waveform data in temperate deciduous forests, because temperate deciduous forests have been proved challenging in regard to lidar-derived biomass estimations.

1. Introduction

Forests are major global components of terrestrial carbon pools and fluxes (Dixon et al. 1994). Quantifying the capacity of forest landscapes to sequester greenhouse gases and mitigate climate change is of urgent importance (Solomon 2007). To this end, rapid and

*Corresponding author. Email: gmountrakis@esf.edu

repeatable landscape-scale assessment of forest carbon stocks and fluxes is needed for both baseline estimation and change detection (Brown 2002). Remote-sensing data, such as multispectral (Liu et al. 2010), hyperspectral (Anderson et al. 2008), and radar (Zianis and Mencuccini 2004) data, provide excellent opportunities to efficiently estimate biomass across large spatial extents. However, the accuracy of conventional remote-sensing biomass estimation can be limited, in part, as the majority of remote sensors only capture the surface reflection of forest canopy. Since biomass is closely related to the size and spatial distribution of trees in a forest (Clark and Clark 2000), information about vertical forest structure is needed to enhance the accuracy of biomass estimation.

Compared with conventional multispectral remote sensors, large-footprint light detection and ranging (lidar) sensors offer higher potential towards in detecting the vertical structural complexity of forests and estimating forest biomass (Anderson et al. 2006; Dubayah et al. 2010). Currently, several large-footprint waveform datasets are available, including the scanning lidar imager of canopies by echo recovery (SLICER) (Blair et al. 1994), the Geoscience Laser Altimeter System (GLAS) (Zwally et al. 2002), and the Laser Vegetation Imaging Sensor (LVIS) (Blair, Rabine, and Hofton 1999). Zolkos, Goetz, and Dubayah (2013) compared the performance of biomass estimation using lidar, other satellite sensors, such as multispectral and radar, and their combination in a meta-analysis based on published articles. The comparison showed that lidar data could significantly improve the estimation results. Moreover, it was found that the relative standard error (RSE) decreased as the footprint size became larger. Various height metrics and those developed from waveform statistics have frequently been used to characterize forest vertical structure and have been incorporated into linear regression models for biomass estimation. For example, mean height, maximum height, quadratic mean canopy height (QMCH), and Larey's height (Means et al. 1999; Lefsky, Harding, et al. 2005; Lefsky, Hudak, et al. 2005; Boudreau et al. 2008; Asner and Mascaro 2014) were used in biomass estimation. In addition, metrics calculated from lidar waveform energy statistics have also performed well in forest above-ground biomass estimation (Drake, Dubayah, Clark, et al. 2002; Harding et al. 2001). Although not originally designed for biomass estimation, energy statistical metrics, such as relative heights (RHs), are defined as the height at which a certain percentage of total intensity is reached (Blair, Hofton, and Rabine 2006). For example, RH50 is the above-ground height where the intensity summation reaches 50% of the total intensity aggregation. A series of RH metrics have been widely used in biomass estimation studies (Anderson et al. 2006; Hyde et al. 2006) and have shown better performance than that from spectral metrics generated from other sensors, such as Landsat and radar.

Although a variety of metrics have been applied to biomass estimation using linear regression models, their ecological context is limited (Drake, Dubayah, Knox, et al. 2002). Existing metrics are based on statistics expressing the waveform energy distribution. Improved metrics capturing the reflecting energy from canopy clump and gaps may better serve vertical structure characterization as well as biomass estimation (Ni-Meister, Yang, and Kiang 2010; Lee et al. 2011; Widlowski et al. 2011). Peaks in a waveform normally represent strong reflection from canopy clump due to the near-infrared pulse, especially in large-footprint lidar waveforms. Lidar metrics describing peaks have been developed in estimating biomass in small-footprint waveform data but not yet for large-footprint waveforms. For example, Muss, Mladenoff, and Townsend (2011) developed a set of metrics that incorporated area of each peak and the height of peak intensity in a pseudo-waveform from small-footprint waveform data to retrieve a series of biophysical parameters including stand biomass. Lindberg et al. (2012) found strong correlation between vegetation

ratio, which was calculated as the ratio between the intensity sum of vegetation and total intensity, and the vegetation volume measured in 12 m field plots using small-footprint waveform lidar data.

In this article we investigate novel biomass metrics that show strong correlation to vegetation structural attributes focusing on large (rather than small) waveform footprints. More specifically, in this study a novel set of biomass metrics is developed based on parameters derived following a Gaussian decomposition (GD) of the waveform signal. GD derives a number of Gaussian functions representing the vegetation clump within a large waveform footprint (Hofton, Minster, and Blair 2000; Wagner et al. 2006). We hypothesize that the developed metrics perform better than the existing metrics in biomass estimation due to the integration of GD results, such as centre locations, Gaussian widths, and amplitudes of the Gaussian components. Parameters of the Gaussian function in GD for large-footprint lidar data have not previously been used in biomass estimation. We evaluate the predictive capability of these novel metrics in biomass estimation using a Random Forests (RF) regression model. We also compare the performance of our GD-derived metrics to other existing metrics, such as RH metrics in biomass estimation for both deciduous and coniferous forest stands in central New York State, USA.

2. Methods

2.1. Study area, lidar data, and fieldwork

The study area is located in the central region of New York State around and between Syracuse and Ithaca in Cayuga, Onondaga, and Tompkins counties (Figure 1(*b*)). Forests in this region are predominately second-growth since a peak of agricultural activity in the mid- to late nineteenth century (Stanton and Bills 1996), although some remnant forests are present (Smith, Marks, and Gardescu 1993). Successional processes and widespread

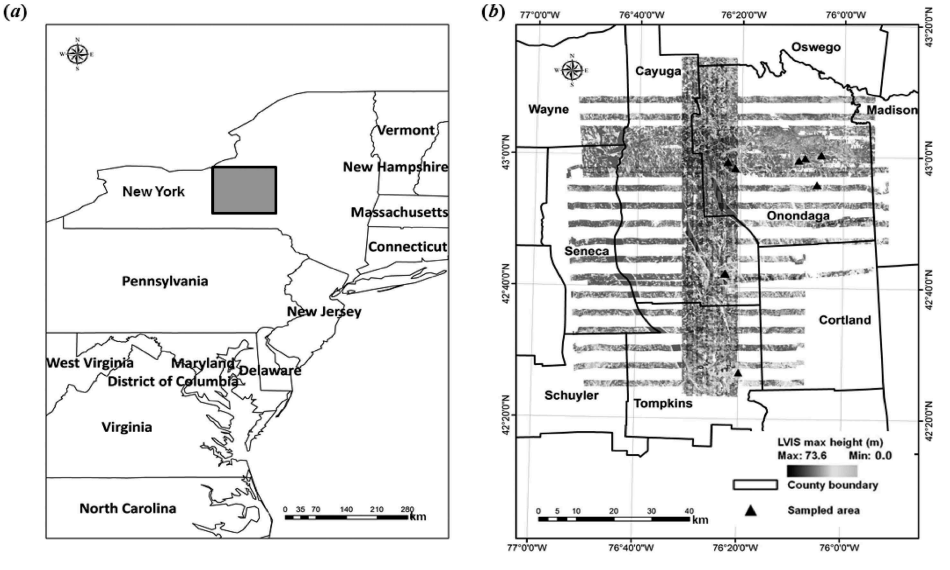

Figure 1. (*a*) Study area located in central New York State (grey box); (*b*) enlarged portion of the study area depicting the lidar data and field sampling location (background images are the height from the LVIS dataset).

tree planting since 1900 have created a matrix of deciduous, coniferous, and mixed stands of various ages across the study area (Nyland, Zipperer, and Hill 1986). In our study, forest ages range from 20 to 100 or more years and various deciduous and coniferous species are present. The major species of coniferous trees include Norway spruce (*Piceaabies*), red pine (*Pinus resinosa*), eastern white pine (*Pinus strobus*), eastern hemlock (*Tsuga canadensis*), and European larch (*Larix decidua*). The deciduous trees include sugar maple (*Acer saccharum*), red maple (*Acer rubrum*), eastern hop-hornbeam (*Ostrya virginiana*), bitternut hickory (*Carya cordiformis*), beech (*Fagus grandifolia*), white ash (*Fraxinus americana*), black walnut (*Juglans nigra*), American basswood (*Tilia americana*), northern red oak (*Quercus rubra*), and black locust (*Robinia pseduoacacia*).

Lidar waveform data were acquired by the LVIS over central New York on 24–26 August 2009 with the leaves fully on. LVIS is a large-footprint waveform lidar sensor developed at NASA's Goddard Space Flight Center (Blair, Rabine, and Hofton 1999). As a simulator of a future satellite mission, the LVIS laser produces a pulse at a wavelength of 1064 nm and illuminates a circular footprint with a diameter of 1–80 m on the ground. It has a large swath of 1 km when the flight height is about 8 km above ground level and a vertical resolution of approximately 0.3 m. The 27 horizontal and 11 vertical LVIS flight lines covering the study area are shown in Figure 1(*b*). The nominal footprint size of the LVIS data used in this study was 20 m.

Fieldwork was carried out July–September 2011 to collect reference data from eight different sites across the study area (Figure 1(*b*)). A total of 82 circular plots, whose centre was geolocated as the waveform footprint centre location using GPS with a mean root mean square error (RMSE) less than 2 m, were sampled to represent typical conditions in the field. A total of 1981 trees were measured. The plot size was 10 m in radius to be consistent with the LVIS nominal footprint size. Based on the spatial energy distribution of the LVIS pulse signal (Blair, Rabine, and Hofton 1999), we developed a sampling protocol involving nested subplots of variable radius. Within the central plot of 7.5 m radius, all stems greater than 5 cm diameter at breast height (DBH; 1.2 m) were measured. In the next concentric plot between 7.5 and 10 m radius, only stems greater than 10 cm DBH were included. Field measurements taken in each plot included species, DBH, and crown height. Plots were categorized into 44 deciduous and 38 coniferous plots based on the dominant species, whose number of stems was more than 80% of the total number of stems, in the canopy layer.

Basal area and stem density by species were calculated for each plot. Species-specific allometric equations (Jenkins 2004) were used to calculate the estimated biomass for each tree within a plot from the fieldwork DBH data. The summation of biomass of all the trees in a plot would later be used as reference data. The minimum biomass for deciduous plots was 15.4 Mg ha^{-1} and the maximum biomass 363.4 Mg ha^{-1}. The range of the coniferous biomass is smaller (i.e. minimum 109.0 Mg ha^{-1} and maximum 339.4 Mg ha^{-1}). The deciduous plots showed a much larger variety in terms of age and biomass than the coniferous plots. The DBH was the key field measurement in the species-specific allometric equations (Table 1).

2.2. *Lidar metrics extraction*

The LVIS waveform data for all the sampled plots were preprocessed for signal smoothing and background noise removal. In the signal-smoothing step, a Gaussian filter was applied to the raw waveform (Chen 2010; Sun et al. 2008). The Gaussian filter was designed according to the Gaussian function that fitted the transmitted pulse of LVIS. The Gaussian

Table 1. Field-derived reference data by forest type for LVIS biomass estimation.

	Deciduous plots	Coniferous plots
Number of plots	44	38
Mean DBH (cm)	26.2	16.1
SD DBH (cm)	13.7	11.6
Mean height (m)	19.8	14.7
SD height (m)	8.1	6.9
Mean basal area (m^2 ha^{-1})	21.2	38.7
SD basal area (m^2 ha^{-1})	15.4	9.9
Mean stem density (number ha^{-1})	835.6	558.7
SD stem density (number ha^{-1})	474.5	313.1
Mean estimated biomass (Mg ha^{-1})	156.5	197.3
SD estimated biomass (Mg ha^{-1})	118.8	50.6

filter had a width twice that of the Gaussian function, and a filter length six times that of the Gaussian function (Chen 2010; Sun et al. 2008). By checking the waveform data for all the sampled plots, the background noise was determined by the maximum intensity of the first 150 bins in a waveform. After deducting the background noise from the raw waveform, the intensity of a bin less than zero was set to zero.

Usually, ground location is assigned to the centroid of the last Gaussian function in LVIS waveform data. However, the location of the last Gaussian centroid often does not reflect accurately the ground location due to overlapping signals from the lower vegetation layer and the ground (Chen 2010). Jutzi and Stilla (2006) showed that one reflecting object (e.g. ground) cannot be distinguished from the other (e.g. lower vegetation layer) if the distance between two reflecting responses in a waveform signal is smaller than 0.85 σ_t (σ_t was the length of the transmitted pulse). Moreover, ground identification by the last Gaussian centroid can be even less accurate in forested areas with slope (Lefsky, Harding, et al. 2005). In this article, because we implemented manual ground identification, metrics that require ground identification express the best-case scenario. To avoid any bias in our results, manual ground identification was applied to both our proposed metrics and the existing ones used as benchmarks. Accurate ground identification is an important topic and an essential step in order to take full advantage of our work, and we have developed algorithms to solve this problem (Zhuang and Mountrakis 2014a, 2014b).

All sampled waveforms were preprocessed and then analysed using GD (Hofton, Minster, and Blair 2000). A novel set of waveform lidar metrics were then developed based on the above-ground height of each Gaussian centroid (H) and Gaussian area index (GAI), which was defined as the area under a Gaussian function (Equation (1)). Three parameters (i.e. width, amplitude, and centroid) that controlled the shape of a Gaussian function were involved in the GAI calculation:

$$\mathrm{GAI} = \int_b^a \mathrm{A}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x, \tag{1}$$

where A, σ, and μ are the amplitude, width, and centroid of a Gaussian function, respectively, and x is the time bin of a waveform. Signal start (parameter a) and signal end (parameter b) were identified as the first and the last signal larger than background noise in a waveform, respectively.

Table 2. Metrics calculated based on GD (H_i represents the height of the centroid of the ith Gaussian function above ground), including two types of proposed metrics derived from GAIs (percentile height metrics and height-weighted metrics) and five other GD metrics, two of which were pre-existing metrics not previously used for biomass estimation, where i indicates the specific number of a Gaussian component and n is the maximum number of the Gaussian components.

Name	Equation	Description
Percentile height metrics		
Percentile height GAI metrics: GAIPer25, GAIPer50, GAIPer75 and GAIPer100	$\sum (\mathrm{GAI})_i$, where $H_{n-1} < H_i < H_n$, H_n is a certain percentile height (25th, 50th, 75th, or 100th), i is a waveform bin, and n is the number of a specific bin	Summation of GAIs with centroids located within a certain percentile height range (e.g. 25%, 50%, 75%, and 100%)
Height-weighted metrics		
GAIHSum	$\sum_{i=1}^{N} (\mathrm{GAI})_i \times H_i$, where $H_i > 0$ and i is a waveform bin	Summation of canopy GAIs, each weighted by the height of its centroid
GAIEHSum	$\sum_{i=1}^{N} (\mathrm{GAI})_i * \exp(H_i * 1.1)$, where $Hi > 0$ and i is a waveform bin	Summation of canopy GAI above-ground height-weighted by an exponentially transformed mean height. The parameter 1.1 was identified through empirical testing
GAIRatio	(GAIHSum)/(GAISum)	Normalized GAIHSum
GAIERatio	(GAIEHSum)/(GAISum)	Normalized GAIEHSum
Other GD-derived metrics		
GAISum	$\sum_{i=1}^{N} (\mathrm{GAI})_i$, where i is a waveform bin	Summation of all GAIs
$(\mathrm{GAI})_{\mathrm{GC}}$	$\frac{\sum(\mathrm{GAI}_{\mathrm{ground}})}{\sum(\mathrm{GAI}_{\mathrm{canopy}})}$	Ratio between ground GAI and total canopy GAI
Mean height	$\sum_{i=1}^{N} \frac{H_i}{N}$, where i is a waveform bin	Mean height above the ground of all the Gaussian centroids; N is the number of Gaussian function in a waveform
First centroid[a]	Max(H_i), where i is a waveform bin	Above-ground height of the highest Gaussian centroid (Hofton, Minster, and Blair 2000)
N[a]	Number of Gaussian functions	Number of Gaussian functions in the decomposition results (Hofton, Minster, and Blair 2000)

Two types of metrics that aggregate GAIs were developed for biomass estimation: percentile height GAI metrics and height-weighted GAI metrics (Table 2). A percentile height GAI metric is the summation of a subset of GAIs, the Gaussian centroids of which are located within a certain percentile height range. The metrics hypothesize that the summation of intensity covered by Gaussian functions at a certain height range can contribute to biomass estimation. To calculate a percentile height GAI metric, a GAI was included in the summation if its associated Gaussian centroid was smaller (closer to the ground) than a predetermined percentile-height threshold, which was fixed and

calculated from Gaussian functions of all waveforms (Table 2). Four percentile height GAI metrics were calculated within the 25th (GAIPer25), 50th (GAIPer50), 75th (GAIPer75), and 100th (GAIPer100) percentile height range, respectively. For example, GAIPer50 would include the summation of all GAIs whose centroid heights were below the 50th percentile and above the 25th percentile of above-ground height. In the case where a portion of a GAI was above the given height percentile (assuming the Gaussian function centre was below the given height percentile), the entire GAI area was included in the calculation. It was assumed that the lidar system settings are similar in all sampling plots and the reflecting patterns of similar ground objects remained consistent (Lindberg et al. 2012; Table 2). Metrics calculated based on GD (H_i represents the height of the centroid of the ith Gaussian function above ground), including two types of proposed metrics derived from GAIs (percentile height metrics and height-weighted metrics) and five other GD metrics, two of which were existing metrics, were not used for biomass estimation.

A height-weighted GAI metric is a weighted summation of all GAIs, where the weight corresponds to the height of the corresponding Gaussian centroid for each GAI. Height-weighted GAI metrics assumed that a taller canopy layer has larger biomass than a lower canopy layer, given the same amount of canopy reflection from each layer. Two height-weighted GAI metrics (GAIHSum and GAIEHSum) were examined, along with their normalized ratios (GAIRatio and GAIERatio) and the summation of all GAIs (GAISum) (Table 2).

- GAIHSum was calculated as the summation of GAIs that represent canopy reflections, with each GAI weighted by the height of its associated Gaussian centroid.
- GAIEHSum summarized all the GAIs using an exponential weight based on the above-ground height of each Gaussian centroid. The calculation of GAIEHSum was based on the observation that tree height can be exponentially correlated with biomass (Zianis and Mencuccini 2004). A similar observation was obtained with our dataset, in which tree height displayed a non-linear relationship with the reference biomass. As such, an exponential relationship was tested in this study by creating the GAIEHSum metric. The exponential parameter was set to 1.1 after a series of experimental tests.
- To calculate the normalized ratio metrics, GAIRatio and GAIERatio, GAIHSum and GAIEHSum were divided by the total of all GAIs (GAISum), respectively (Table 2). Such normalization aimed to compensate for the variability in total received energy across plots due to factors such as atmospheric effects and reflectivity of ground objects.

Additionally, five other GD-based metrics were also examined: (1) GAISum, the summation of all GAIs; (2) mean height, the average above-ground height of all Gaussian centroids; (3) GAIGC, the ratio between ground GAI and the summation of all canopy GAIs; (4) first centroid, the height of the centroid of the highest Gaussian function detected in a waveform; and (5) Num, the number of Gaussian functions detected in a waveform. The last two metrics were pre-existing GD metrics adapted from Hofton, Minster, and Blair (2000), but to date have not been used for biomass estimation.

For comparison purposes, a variety of height-based existing metrics (non-GD metrics) were also calculated. These metrics included RH metrics, QMCH, HMRatio, leading edge, and trailing edge; they also incorporated waveform statistics, such as variance, wavelength, and skewness. These metrics were selected because they had previously

Table 3. Non-GD metrics calculated from LVIS data based on height and statistics.

Name	Equation	Description
Relative height metrics: RH0, RH25, RH50, RH75, and RH100	H_i, where $i < E_{\text{1st, 2nd, 3rd, 4th quantile}}$ and i is a waveform bin	Above-ground height at which 0%, 25%, 50%, 75%, or 100% of the waveform energy occurs (Blair, Hofton, and Rabine 2006)
HMRatio	(RH50)/(RH100)	Ratio between maximum height and height of median energy (HOME) (Drake, Dubayah, Clark, et al. 2002)
QMCH	$\sqrt{\sum_{i=0}^{n} \text{CHP}(h) * h^2}$, where i is a waveform bin, n is maximum waveform bin, and h is the above-ground height for bin i	Canopy height profile was calculated according to the canopy reflection in the waveform; ground–canopy ratio was 2 (Harding et al. 2001; Lefsky et al. 1999)
Leading edge	$H_{\text{maxIntensity/2}}–H_{\text{start}}$, where H is height	Height difference between the first signal larger than half of the maximum intensity of a waveform and the signal start (Lefsky et al. 2007)
Trailing edge	$H_{\text{end}}–H_{\text{maxIntensity/2,}}$ where H is height	Height difference between the first signal larger than half of the maximum intensity of a waveform in the lowest elevation and the signal end (Lefsky et al. 2007)
V	Variance of intensity	Variability in intensity for the canopy portion of the waveform (Garcia et al. 2010)
S	Skewness, $\text{E}\left(\left(\frac{\text{I}-\mu}{\sigma}\right)^3\right)$, where μ is the mean amplitude and σ is the standard deviation of the amplitude	Shape index of the canopy portion of the waveform (Neuenschwander et al. 2008)
L	Wavelength, $N_{\text{First Signal}}–N_{\text{Last Signal}}$, where N is the bin number	Length of signal larger than background noise (Sun et al. 2008)

demonstrated successful estimation of biomass in previous studies or because they were related to the characterization of forest structure (Anderson et al. 2006; Drake, Dubayah, Clark, et al. 2002; Lefsky, Harding, et al. 2005; Meyer et al. 2013). For consistency, the three metrics that required ground identification (RH metrics, QMCH, and HMRatio) also used the manual ground data. Table 3 offers a detailed description and bibliographical information. These metrics were evaluated separately for biomass estimation and compared to the 13 GD metrics.

2.3. Experimental set-up and assessment

Three scenarios were designed to examine the performance of different sets of metrics in biomass estimation. Scenarios 1 and 2 were designed to compare biomass estimation accuracy between the models using GD and non-GD metrics. Scenario 3 was developed to

Table 4. Experiment set-up.

Scenario	Metric	Purpose
Scenario 1	GD metrics	Performance comparison between GD and non-GD metrics from scenarios 1 and 2
Scenario 2	Non-GD metrics	
Scenario 3	Combined metrics*	Use of all information in attempt to obtain highest accuracy

Note: *combined metrics includes both GD and non-GD metrics

investigate the highest potential biomass estimation accuracy using both GD and non-GD metrics. The experimental set-up is shown in Table 4.

RF regression models (Breiman 2001) were established to estimate biomass; these models comprise a collection of binary regression trees. The major advantages of RF are: (1) report generation for the importance of both accuracy and metrics; (2) overfitting reduction through a bootstrap sampling strategy (Cutler et al. 2007; Falkowski et al. 2009); and (3) tolerance to outlier presence in the dataset because a large number of trees are built into the model. RF has been successfully applied in various remote-sensing forest applications, including biomass estimation (Gleason and Im 2012; Hudak et al. 2008; Powell et al. 2010).

In each scenario, the best RF model was identified using the following steps.

(1) An RF model was fitted with all candidate metrics – for example, all GD metrics in scenario 1. The importance of all metrics was then extracted.
(2) Using the importance ranking of the metrics, separate candidate RF regression models were built to sequentially assess the added value of each metric. The first candidate RF model used the first metric in the ranking, the second candidate RF model used the first and second metrics … and the final candidate used all the metrics. Each model was repeated 40 times in order to decrease the variability. For each candidate model the mean RMSE was calculated from 40 repetitions.
(3) The candidate RF model with the lowest mean RMSE was selected as the best RF model (RF_{best}). The metrics associated with RF_{best} were identified as the best metrics from all available metrics. Both RMSE and R^2 of RF_{best} were reported.

In this study, *R* statistical software with an add-on RF package (Liaw and Wiener 2002) was used for building of RF models. For each RF the number of trees was set to 1000, which was determined by the trial-and-error method recommended by Liaw and Wiener (2002). For each tree, by default, approximately two-thirds of data were randomly selected for training while the remaining one-third was used as testing data.

3. Results and discussion

In this section results are reported separately for each of the three metric scenarios – GD, non-GD, and combined metrics. The evaluation for each scenario starts with a sensitivity assessment of the metric importance followed by detailed biomass estimation results for the most successful random forest (RF_{best}) in each scenario. A separate assessment takes place for deciduous and coniferous plots. Correlation between metrics selected in RF_{best} using the combined metrics and other forest attributes, such as mean DBH and stem density, was further investigated.

3.1. Biomass estimation results using GD-derived metrics

3.1.1. Metric importance

Metric importance was evaluated for each GD metric in Table 2. The decrease of normalized mean square error (DNMSE) was used as the importance criterion. For 40 RF regressions and for each metric, the average and standard deviation of DNMSE were calculated as shown in Figure 2(*a*) for deciduous plots and in Figure 2(*b*) for coniferous plots. First centroid and GAIHSum were selected as the most important variables for biomass estimation in deciduous and coniferous plots, respectively.

3.1.2. Regression results

Following the process explained in Section 2.3, RF_{best} was identified. The deciduous RF_{best} model was built using two metrics, first centroid and GAIRatio, while the coniferous RF_{best} model was built using the following metrics: GAIHSum, GAIGC, GAISum, and GAIPer100. Biomass estimation accuracy *versus* reference biomass data was assessed separately for deciduous plots (Figure 3(*a*)) and coniferous plots (Figure 3(*b*)). The RMSE of biomass estimation for deciduous and coniferous plots of the RF_{best} model was 65.6 and 35.2 Mg ha^{-1}, respectively. The R^2 of the RF_{best} regression models was 0.69 and 0.50

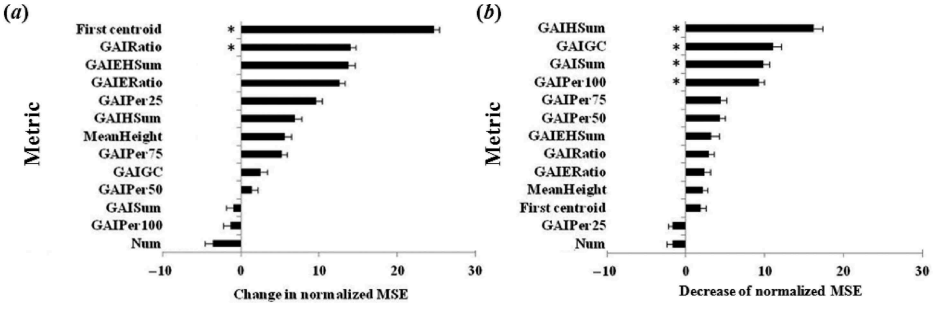

Figure 2. Metrics importance of the GD-derived metrics through 40 RF regressions for (*a*) deciduous and (*b*) coniferous plots. Metrics with star are the selected metrics for biomass regression.

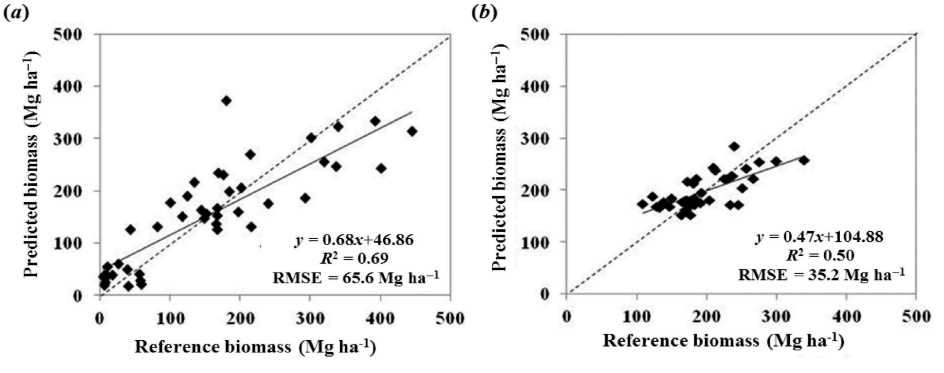

Figure 3. Scatterplots of reference biomass against predicted biomass using GD metrics for the (*a*) deciduous and (*b*) coniferous RF_{best} models. The dashed line is the 1:1 line and the solid line is the regression model.

for the deciduous and coniferous plots, respectively. The coniferous RF_{best} model obtained better RMSE but lower R^2 values because the biomass variance for coniferous plots was much lower than that for deciduous plots (Table 1).

3.2. *Biomass estimation results using pre-existing metrics*

3.2.1. *Metric importance*

Similarly, metric importance was evaluated for each non-GD metric listed in Table 3. The average and standard deviation of the DNMSE distribution for each metric are depicted in Figure 4(*a*) for deciduous plots and in Figure 4(*b*) for coniferous plots. The RH100 and RH75 metrics showed the highest importance in deciduous and coniferous plots, respectively. Also, HMRatio appeared among the top three important metrics for both deciduous and coniferous plots.

3.2.2. *Regression results*

The deciduous RF_{best} model used RH100, QMCH, HMRatio, variance, and waveform length, while the coniferous RF_{best} model was built with HMRatio and RH75. RF biomass estimations using the pre-existing metrics are given in Figure 5 in comparison to the reference biomass data for both deciduous plots (Figure 5(*a*)) and coniferous plots (Figure 5(*b*)). The RMSE of the RF_{best} regression models for deciduous and coniferous plots was 81.0 and 36.3 Mg ha^{-1}, respectively, while R^2 was 0.52 and 0.47 for deciduous and coniferous plots, respectively.

3.3. *Biomass estimation results using combined metrics and all-metric comparison*

3.3.1. *Metric importance and regression results*

In order to explore the possibility of further improving biomass esitimation accuracy, the GD and non-GD metrics were combined. Average metric importance was calculated for each metric along with its standard deviations, as shown in Figure 6(*a*) for deciduous plots and Figure 6(*b*) for coniferous plots. In general, GD-based metrics were ranked higher than non-GD. The best average prediction results were obtained by successively adding

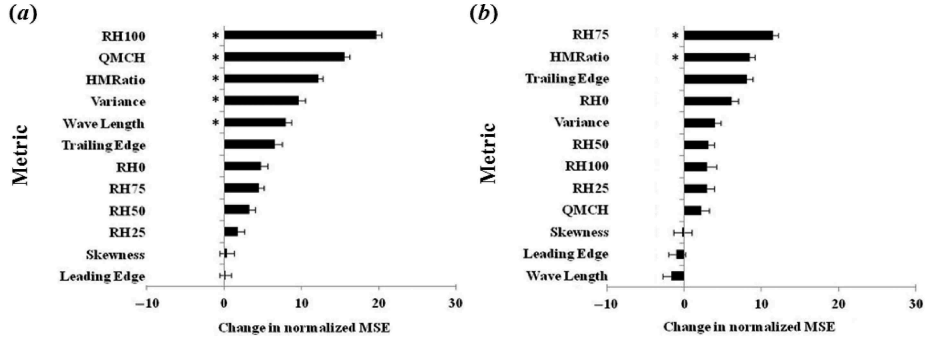

Figure 4. Variable importance of non-GD metrics through 40 RF regressions for (*a*) deciduous plots and (*b*) coniferous plots. Metrics with star are the selected metrics for biomass regression.

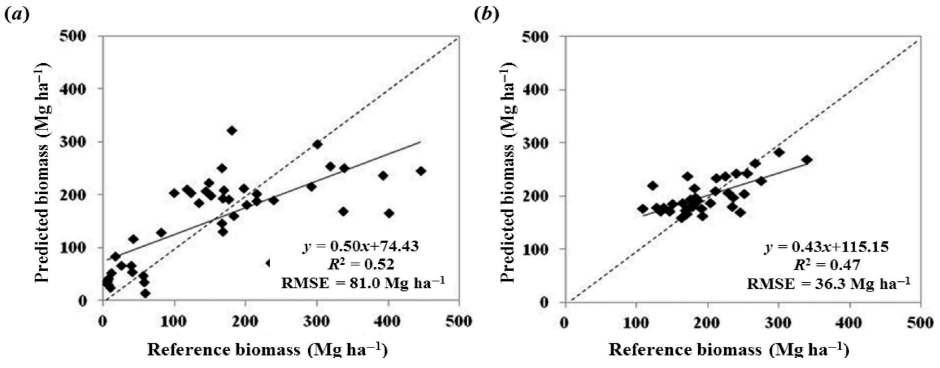

Figure 5. Scatterplots of reference biomass against predicted biomass using non-GD metrics for the (*a*) deciduous and (*b*) coniferous RF_{best} models. The dashed line is the 1:1 line and the solid line is the regression model.

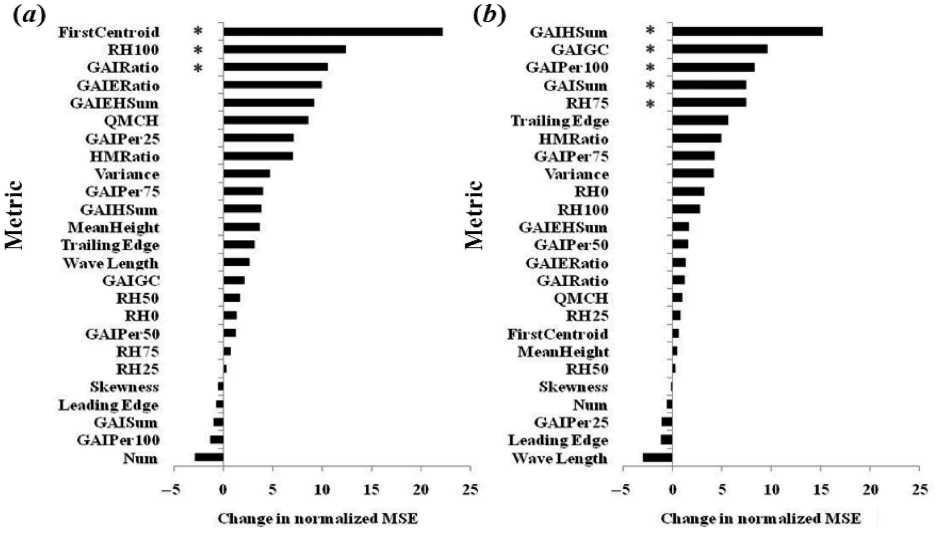

Figure 6. Variable importance of the combined metrics for both (*a*) deciduous and (*b*) coniferous biomass estimation by RF regression. Metrics with star are the selected metrics for biomass regression.

the most important variable according to the RF variable importance results, as mentioned in Section 2.3.

In regard to deciduous plots, the RF_{best} model selected three metrics, two GD (first centroid and GAIRatio derived from GAI) and one non-GD, RH100 (maximum height). In the deciduous RF model, the first centroid metric used previously to characterize mean canopy height (Hofton, Minster, and Blair 2000) showed the highest importance. Although mean height has also been shown to be a good estimator of biomass in field data (Lefsky et al. 1999; Means et al. 1999), first centroid had never previously been used for biomass estimation. In regard to coniferous plots, four GD metrics built by GAI (GAIHSum, GAIGC, GAIPer100, and GAISum) and one non-GD metric (RH75) were selected. The preference for our proposed GD metrics is clear.

In the best prediction results using combined metrics, the RMSE of RF_{best} biomass estimation was 65.8 and 34.5 Mg ha^{-1} for deciduous and coniferous plots, respectively, while R^2 of the RF_{best} regression models was 0.69 and 0.52, respectively.

3.3.2. *Comparison among GD metrics, non-GD metrics, and combined metrics*

The comparison of biomass estimation performance among different metrics is shown in Figure 7. Rather than the single RMSE and R^2 values from RF_{best}, we compared the mean RMSE and mean R^2 values of the 40 repeated RF regression models – all with the same optimized inputs (that with the lowest RMSE was RF_{best}). As can be seen from Figure 7, the GD metrics, when contrasted with the non-GD metrics, achieved significantly better prediction accuracy for biomass estimation (19% lower RMSE, 25% higher R^2) in deciduous plots and marginally better (4% lower RSME, 6% higher R^2) in coniferous plots. In other words, the GD metrics contained more useful information in relation to biomass estimation than the non-GD, and thus the addition of non-GD to GD metrics in the RF models did not significantly improve prediction accuracy. In the 40 RF repetitions

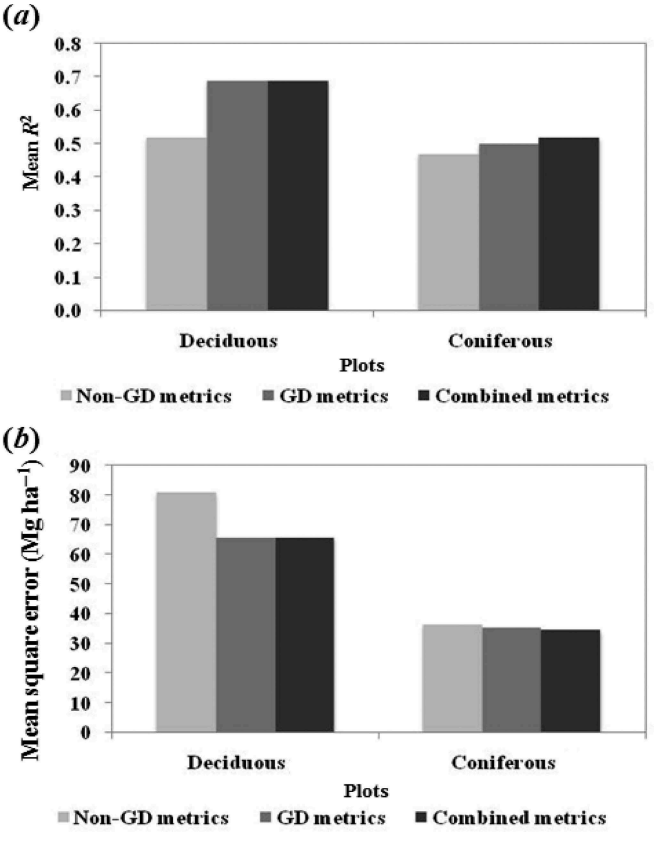

Figure 7. Comparison of biomass estimation results among different sets of metrics: (*a*) mean RMSE and (*b*) mean R^2. The standard deviation of each scenario is indicated by a small bar at the top of each column.

the standard deviation of both RMSE and R^2 values was very low (<0.6 Mg ha^{-1} and <0.001 for both deciduous and coniferous plots), indicating convergence of the models.

3.4. *Correlation of selected metrics to stem density and mean DBH*

In order to investigate further the performance of GD metrics, the relationship between the metrics selected in the RF_{best} of scenario 3 and the mean DBH/stem density for both deciduous and coniferous plots was examined. Starting with the deciduous plots shown in Figure 8, these were dominated by white ash in young (<30 years) plots and by sugar maple in old (>80 years) plots. First centroid and RH100, which were statistically the most important metrics in deciduous plots, were positively correlated with mean DBH in both young and old plots (Figure 8(*a*) and Figure 8(*c*)). As higher trees usually have larger DBH, this strong correlation appears reasonable. Opposite correlations were observed for young and old plots in the relationship between height metrics and stem density (Figure 8(*d*) and Figure 8(*f*)). At the early growth stage of young forest, tree heights and stem density were positively correlated. On the other hand, according to the self-thinning rule (Carleton and Wannamaker 1987; Fahey, Battles, and Wilson 1998; Lonsdale 1990), the negative correlation between height metrics and stem density of old plots was caused by the density-dependent mortality. In addition, GAIRatio as a product of reflection and height also showed strong correlation with forest characteristics (Figure 8(*b*) and Figure 8(*e*)).

These relationships were also examined for coniferous plots. As shown in Figure 9, metrics showed an opposite correlation with mean DBH and stem density, also due to the self-thinning phenomenon in old coniferous plots. Specifically, GAIGC (Figure 9(*b*)) was highly correlated with mean DBH. The positive correlation between GAIGC (i.e. ratio between ground GAI and canopy GAI) and mean DBH, and negative correlation between GAIGC and stem density, meant that the canopy gap between trees became larger when fewer small understorey trees existed. In addition, GAISum (Figure 9(*c*)), which

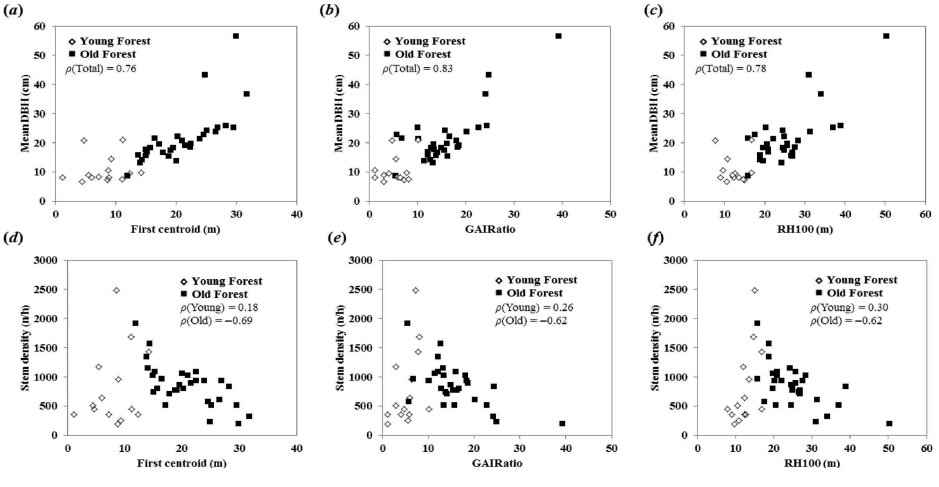

Figure 8. Relationship between metrics used in the best RF model and forest characteristics for deciduous plots. (*a*), (*b*), and (*c*) stem density; (*d*), (*e*), and (*f*) mean DBH within each plot.

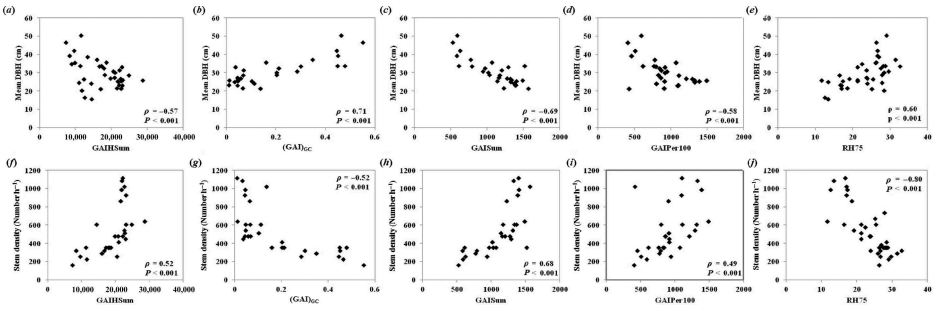

Figure 9. Scatterplots between GD metrics used in the best RF model and forest characteristics for coniferous plots. (*a*), (*b*), (*c*), (*d*) and (*e*), stem density; (*f*), (*g*), (*h*), (*i*), and (*j*) mean DBH within each plot.

represents total canopy reflection, showed a strong negative correlation with mean DBH. Due to the management system used in coniferous plots, older examples contained only a few large trees while younger ones had a greater number of young trees. Thus, strong reflection was recorded from younger coniferous plots with smaller mean DBH, and *vice versa*. Moreover, RH75 (Figure 9(*j*)) and GAISum (Figure 9(*h*)) are highly correlated to stem density. Strong correlation of lidar metrics with mean DBH and stem density has been found to be associated with accurate above-ground biomass estimation (Dubayah and Drake 2000; Lim et al. 2003; Slik et al. 2010; Van Leeuwen and Nieuwenhuis 2010). In our case the selected GD metrics and RH75 exhibit this strong correlation, partially explaining the accurate biomass estimations.

4. Conclusions

This study investigated the use of large-footprint lidar waveform metrics based on Gaussian decomposition (GD) results for forest above-ground biomass estimation. Our study used, for the first time, GD metrics in biomass estimation. A variety of successional stages were included with biomass ranging from 15.4 to 364.3 Mg ha^{-1}. By comparing the GD-derived metrics to results from non-GD metrics, GD-based regression achieved significantly better biomass prediction results for deciduous plots and marginally better results for coniferous plots. This represents a significant contribution, because temperate deciduous forests have proved challenging in regard to lidar-derived biomass estimations.

The improved performance of GD metrics in biomass estimation can be attributed to the incorporation of both waveform amplitude and height information from the parameters of Gaussian functions in GD. Although not explicitly tested in this study, we can infer that the context-specific vertical layer information captured by Gaussian functions has strong linkages to forest characteristics. GD metrics showed strong correlation with forest characteristics (i.e. mean DBH and stem density), these being highly correlated with biomass, thereby increasing our confidence in proposing them as a method suitable for biomass estimation using large-footprint waveform lidar data. Further testing sites would be necessary to assess the potential of GD metrics in a variety of ecological environments.

Acknowledgement

We thank Drs James Bryan Blair and Michelle Hofton for providing the LVIS dataset and the associated RH metrics.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported through NASA's Biodiversity Program [grant number NNX09AK16G].

References

Anderson, J., M. E. Martin, M.-L. Smith, R. O. Dubayah, M. A. Hofton, P. Hyde, B. E. Peterson, J. B. Blair, and R. G. Knox. 2006. "The Use of Waveform Lidar to Measure Northern Temperate Mixed Conifer and Deciduous Forest Structure in New Hampshire." *Remote Sensing of Environment* 105: 248–261. doi:10.1016/j.rse.2006.07.001.

Anderson, J. E., L. C. Plourde, M. E. Martin, B. H. Braswell, M. L. Smith, R. O. Dubayah, M. A. Hofton, and J. B. Blair. 2008. "Integrating Waveform Lidar with Hyperspectral Imagery for Inventory of a Northern Temperate Forest." *Remote Sensing of Environment* 112: 1856–1870. doi:10.1016/j.rse.2007.09.009.

Asner, G. P., and J. Mascaro. 2014. "Mapping Tropical Forest Carbon: Calibrating Plot Estimates to a Simple Lidar Metric." *Remote Sensing of Environment* 140: 614–624. doi:10.1016/j.rse.2013.09.023.

Blair, J. B., D. Coyle, J. Bufton, and D. Harding. 1994. "Optimization of an Airborne Laser Altimeter for Remote Sensing of Vegetation and Tree Canopies." In *Geoscience and Remote Sensing Symposium*, 1994. *IGARSS'94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, International*, 939–941. New York: IEEE.

Blair, J. B., M. A. Hofton, and D. L. Rabine. 2006. "Processing of NASA LVIS Elevation and Canopy (LGE, LCE and LGW) Data Products, Version 1.01." http://lvis.gsfc.nasa.gov

Blair, J. B., D. L. Rabine, and M. A. Hofton. 1999. "The Laser Vegetation Imaging Sensor: A Medium-Altitude, Digitisation-Only, Airborne Laser Altimeter for Mapping Vegetation and Topography." *ISPRS Journal of Photogrammetry and Remote Sensing* 54: 115–122. doi:10.1016/S0924-2716(99)00002-7.

Boudreau, J., R. F. Nelson, H. A. Margolis, A. Beaudoin, L. Guindon, and D. S. Kimes. 2008. "Regional Aboveground Forest Biomass Using Airborne and Spaceborne Lidar in Québec." *Remote Sensing of Environment* 112: 3876–3890. doi:10.1016/j.rse.2008.06.003.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Brown, S. 2002. "Measuring Carbon in Forests: Current Status and Future Challenges." *Environmental Pollution* 116: 363–372. doi:10.1016/S0269-7491(01)00212-3.

Carleton, T. J., and B. A. Wannamaker. 1987. "Mortality and Self-Thinning in Postfire Black Spruce." *Annals of Botany* 59: 621–628.

Chen, Q. 2010. "Retrieving Vegetation Height of Forests and Woodlands over Mountainous Areas in the Pacific Coast Region Using Satellite Laser Altimetry." *Remote Sensing of Environment* 114: 1610–1627. doi:10.1016/j.rse.2010.02.016.

Clark, D. B., and D. A. Clark. 2000. "Landscape-Scale Variation in Forest Structure and Biomass in a Tropical Rain Forest." *Forest Ecology and Management* 137: 185–198. doi:10.1016/S0378-1127(99)00327-8.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. "Random Forests for Classification in Ecology." *Ecology* 88: 2783–2792. doi:10.1890/07-0539.1.

Dixon, R. K., A. M. Solomon, S. Brown, R. A. Houghton, M. C. Trexier, and J. Wisniewski. 1994. "Carbon Pools and Flux of Global Forest Ecosystems." *Science* 263 (5144): 185–190.

Drake, J. B., R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, R. L. Chazdon, J. F. Weishampel, and S. Prince. 2002. "Estimation of Tropical Forest Structural Characteristics,

Using Large-Footprint Lidar." *Remote Sensing of Environment* 79: 305–319. doi:10.1016/S0034-4257(01)00281-4.

Drake, J. B., R. O. Dubayah, R. G. Knox, D. B. Clark, and J. B. Blair. 2002. "Sensitivity of Large-Footprint Lidar to Canopy Structure and Biomass in a Neotropical Rainforest." *Remote Sensing of Environment* 81: 378–392. doi:10.1016/S0034-4257(02)00013-5.

Dubayah, R. O., and J. B. Drake. 2000. "Lidar Remote Sensing for Forestry." *Journal of Forestry* 98: 44–46.

Dubayah, R. O., S. L. Sheldon, D. B. Clark, M. A. Hofton, J. B. Blair, G. C. Hurtt, and R. L. Chazdon. 2010. "Estimation of Tropical Forest Height and Biomass Dynamics Using Lidar Remote Sensing at La Selva, Costa Rica." *Journal of Geophysical Research* 115 (2). doi:10.1029/2009JG000933.

Fahey, T. J., J. J. Battles, and G. F. Wilson. 1998. "Responses of Early Successional Northern Hardwood Forests to Changes in Nutrient Availability." *Ecological Monographs* 68: 183–212. doi:10.1890/0012-9615(1998)068[0183:ROESNH]2.0.CO;2.

Falkowski, M. J., J. S. Evans, S. Martinuzzi, P. E. Gessler, and A. T. Hudak. 2009. "Characterizing Forest Succession with Lidar Data: An Evaluation for the Inland Northwest, USA." *Remote Sensing of Environment* 113: 946–956. doi:10.1016/j.rse.2009.01.003.

García, M., D. Riaño, E. Chuvieco, and F. M. Danson. 2010. "Estimating Biomass Carbon Stocks for a Mediterranean Forest in Central Spain Using Lidar Height and Intensity Data." *Remote Sensing of Environment* 114: 816–830. doi:10.1016/j.rse.2009.11.021.

Gleason, C. J., and J. Im. 2012. "Forest Biomass Estimation from Airborne Lidar Data Using Machine Learning Approaches." *Remote Sensing of Environment* 125: 80–91. doi:10.1016/j.rse.2012.07.006.

Harding, D. J., M. A. Lefsky, G. G. Parker, and J. B. Blair. 2001. "Laser Altimeter Canopy Height Profiles: Methods and Validation for Closed-Canopy, Broadleaf Forests." *Remote Sensing of Environment* 76: 283–297. doi:10.1016/S0034-4257(00)00210-8.

Hofton, M. A., J. B. Minster, and J. B. Blair. 2000. "Decomposition of Laser Altimeter Waveforms." *IEEE Transactions on Geoscience and Remote Sensing* 38: 1989–1996. doi:10.1109/36.851780.

Hudak, A. T., N. L. Crookston, J. S. Evans, D. E. Hall, and M. J. Falkowski. 2008. "Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from Lidar Data." *Remote Sensing of Environment* 112: 2232–2245. doi:10.1016/j.rse.2007.10.009.

Hyde, P., R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker. 2006. "Mapping Forest Structure for Wildlife Habitat Analysis Using Multi-Sensor (Lidar, Sar/Insar, ETM+, Quickbird) Synergy." *Remote Sensing of Environment* 102: 63–73. doi:10.1016/j.rse.2006.01.021.

Jenkins, J. C. 2004. *Comprehensive Database of Diameter-Based Biomass Regressions for North American Tree Species*. Newton Square, PA: United States Department of Agriculture, Forest Service, Northeastern Research Station.

Jutzi, B., and U. Stilla. 2006. "Range Determination with Waveform Recording Laser Systems Using a Wiener Filter." *ISPRS Journal of Photogrammetry and Remote Sensing* 61: 95–107. doi:10.1016/j.isprsjprs.2006.09.001.

Lee, S., W. Ni-Meister, W. Yang, and Q. Chen. 2011. "Physically Based Vertical Vegetation Structure Retrieval from Icesat Data: Validation Using Lvis in White Mountain National Forest, New Hampshire, USA." *Remote Sensing of Environment* 115 (11): 2776–2785. doi:10.1016/j.rse.2010.08.026.

Lefsky, M. A., D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart. 1999. "Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA." *Remote Sensing of Environment* 67: 83–98. doi:10.1016/S0034-4257(98)00071-6.

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. 2005. "Estimates of Forest Canopy Height and Aboveground Biomass Using Icesat." *Geophysical Research Letters* 32: 1–4. doi:10.1029/2005GL023971.

Lefsky, M. A., A. T. Hudak, W. B. Cohen, and S. A. Acker. 2005. "Patterns of Covariance between Forest Stand and Canopy Structure in the Pacific Northwest." *Remote Sensing of Environment* 95: 517–531. doi:10.1016/j.rse.2005.01.004.

Lefsky, M. A., M. Keller, Y. Pang, P. B. De Camargo, and M. O. Hunter. 2007. "Revised Method for Forest Canopy Height Estimation from Geoscience Laser Altimeter System Waveforms." *Journal of Applied Remote Sensing* 1 (1): 013537–013537. doi:10.1117/1.2795724.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by Randomforest." *R News* 2: 18–22.

Lim, K., P. Treitz, M. Wulder, B. St-OngeÃÅ, and M. Flood. 2003. "Lidar Remote Sensing of Forest Structure." *Progress in Physical Geography* 27: 88–106. doi:10.1191/0309133303pp360ra.

Lindberg, E., K. Olofsson, J. Holmgren, and H. Olsson. 2012. "Estimation of 3D Vegetation Structure from Waveform and Discrete Return Airborne Laser Scanning Data." *Remote Sensing of Environment* 118: 151–161. doi:10.1016/j.rse.2011.11.015.

Liu, J., E. Pattey, J. R. Miller, H. McNairn, A. Smith, and B. Hu. 2010. "Estimating Crop Stresses, Aboveground Dry Biomass and Yield of Corn Using Multi-Temporal Optical Data Combined with a Radiation Use Efficiency Model." *Remote Sensing of Environment* 114: 1167–1177. doi:10.1016/j.rse.2010.01.004.

Lonsdale, W. M. 1990. "The Self-Thinning Rule: Dead or Alive?" *Ecology* 71: 1373–1388. doi:10.2307/1938275.

Means, J. E., S. A. Acker, D. J. Harding, J. B. Blair, M. A. Lefsky, W. B. Cohen, M. E. Harmon, and W. A. McKee. 1999. "Use of Large-Footprint Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregon." *Remote Sensing of Environment* 67: 298–308. doi:10.1016/S0034-4257(98)00091-1.

Meyer, V., S. Saatchi, J. Chave, J. Dalling, S. Bohlman, G. Fricker, C. Robinson, M. Neumann, and S. Hubbell. 2013. "Detecting Tropical Forest Biomass Dynamics from Repeated Airborne Lidar Measurements." *Biogeosciences* 10 (8): 5421–5438. doi:10.5194/bg-10-5421-2013.

Muss, J. D., D. J. Mladenoff, and P. A. Townsend. 2011. "A Pseudo-Waveform Technique to Assess Forest Structure Using Discrete Lidar Data." *Remote Sensing of Environment* 115: 824–835. doi:10.1016/j.rse.2010.11.008.

Neuenschwander, A. L., T. J. Urban, R. Gutierrez, and B. E. Schutz. 2008. "Characterization of Icesat/Glas Waveforms over Terrestrial Ecosystems: Implications for Vegetation Mapping." *Journal of Geophysical Research: Biogeosciences* 113: 2005–2012. doi:10.1029/2007JG000557.

Ni-Meister, W., W. Yang, and N. Y. Kiang. 2010. "A Clumped-Foliage Canopy Radiative Transfer Model for a Global Dynamic Terrestrial Ecosystem Model. I: Theory." *Agricultural and Forest Meteorology* 150 (7–8): 881–894. doi:10.1016/j.agrformet.2010.02.009.

Nyland, R., W. Zipperer, and D. Hill. 1986. "The Development of Forest Islands in Exurban Central New York State." *Landscape and Urban Planning* 13: 111–123. doi:10.1016/0169-2046(86)90016-2.

Powell, S. L., W. B. Cohen, S. P. Healey, R. E. Kennedy, G. G. Moisen, K. B. Pierce, and J. L. Ohmann. 2010. "Quantification of Live Aboveground Forest Biomass Dynamics with Landsat Time-Series and Field Inventory Data: A Comparison of Empirical Modeling Approaches." *Remote Sensing of Environment* 114: 1053–1068. doi:10.1016/j.rse.2009.12.018.

Slik, J., S.-I. Aiba, F. Q. Brearley, C. H. Cannon, O. Forshed, K. Kitayama, H. Nagamasu, R. Nilus, J. Payne, G. Paoli, A. D. Poulsen, N. Raes, D. Sheil, K. Sidiyasa, E. Suzuki, and J. L. C. H. Van Valkenburg. 2010. "Environmental Correlates of Tree Biomass, Basal Area, Wood Specific Gravity and Stem Density Gradients in Borneo's Tropical Forests." *Global Ecology and Biogeography* 19: 50–60. doi:10.1111/j.1466-8238.2009.00489.x.

Smith, B. E., P. Marks, and S. Gardescu. 1993. "Two Hundred Years of Forest Cover Changes in Tompkins County, New York." *Bulletin of the Torrey Botanical Club* 120: 229–247. doi:10.2307/2996988.

Solomon, S., ed. 2007. *Climate Change 2007 – The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC*, Vol. 4. Cambridge: Cambridge University Press.

Stanton, B. F., and N. L. Bills. 1996. *The Return of Agricultural Lands to Forest: Changing Land Use in the Twentieth Century. EB(USA)*. Ithaca, NY: Cornell University, Department of Agricultural, Resource, and Managerial Economics.

Sun, G., K. J. Ranson, D. S. Kimes, J. B. Blair, and K. Kovacs. 2008. "Forest Vertical Structure from GLAS: An Evaluation Using LVIS and SRTM Data." *Remote Sensing of Environment* 112: 107–117. doi:10.1016/j.rse.2006.09.036.

Van Leeuwen, M., and M. Nieuwenhuis. 2010. "Retrieval of Forest Structural Parameters Using Lidar Remote Sensing." *European Journal of Forest Research* 129: 749–770. doi:10.1007/s10342-010-0381-4.

Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka. 2006. "Gaussian Decomposition and Calibration of a Novel Small-Footprint Full-Waveform Digitising Airborne Laser Scanner."

ISPRS Journal of Photogrammetry and Remote Sensing 60: 100–112. doi:10.1016/j.isprsjprs.2005.12.001.

Widlowski, J. L., B. Pinty, M. Clerici, Y. Dai, M. De Kauwe, K. De Ridder, A. Kallel, H. Kobayashi, T. Lavergne, W. Ni-Meister, A. Olchev, T. Quaife, S. Wang, W. Yang, Y. Yang, and H. Yuan. 2011. "Rami4pilps: An Intercomparison of Formulations for the Partitioning of Solar Radiation in Land Surface Models." *Journal of Geophysical Research G: Biogeosciences* 116 (2): G02019.

Zhuang, W., and G. Mountrakis. 2014a. "An Accurate and Computationally Efficient Algorithm for Ground Peak Identification in Large Footprint Waveform Lidar Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 95: 81–92. doi:10.1016/j.isprsjprs.2014.06.004.

Zhuang, W., and G. Mountrakis. 2014b. "Ground Peak Identification in Dense Shrub Areas Using Large Footprint Waveform Lidar and Landsat Images." *International Journal of Digital Earth*, ahead-of-print, 1–20. doi:10.1080/17538947.2014.942716.

Zianis, D., and M. Mencuccini. 2004. "On Simplifying Allometric Analyses of Forest Biomass." *Forest Ecology and Management* 187: 311–332. doi:10.1016/j.foreco.2003.07.007.

Zolkos, S., S. Goetz, and R. Dubayah. 2013. "A Meta-Analysis of Terrestrial Aboveground Biomass Estimation Using Lidar Remote Sensing." *Remote Sensing of Environment* 128: 289–298. doi:10.1016/j.rse.2012.10.017.

Zwally, H., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhirne, and R. Thomas. 2002. "Icesat's Laser Measurements of Polar Ice, Atmosphere, Ocean, and Land." *Journal of Geodynamics* 34: 405–445. doi:10.1016/S0264-3707(02)00042-X.