Agricultural and Forest Meteorology 165 (2012) 64–72

Contents lists available at SciVerse ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA

Feng Zhao[a], Qinghua Guo[b], Maggi Kelly[a,c,*]

[a] Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA
[b] School of Engineering, University of California, Merced, CA, USA
[c] Geospatial Innovation Facility, University of California, Berkeley, USA

ARTICLE INFO

Article history:
Received 21 November 2011
Received in revised form 21 May 2012
Accepted 26 May 2012

Keywords:
Lidar
Allometric equations
Biomass
Sierra Nevada
Forest

ABSTRACT

Regression modeling of biomass estimates from an airborne, multiple return lidar system using regional biomass allometric equations differs significantly from those using national scale Jenkins allometric equations with respect to the amount of variation explained, variables selected and variables of importance. Our discrete return lidar data were collected in September 2007 at 121 plots in a conifer dominated forest site in the Sierra Nevada mountains that include a full range of forest density. We regressed field plot-level estimates of biomass derived from field data and two different allometric equations with a range of lidar metrics. We compared regression performance across eight models: (1) point clouds alone, (2) point clouds with an empirical relationship between DBH and height (i.e., volume), (3) individual tree-level metrics, and (4) all data combined, and across two allometric equations – (A) Forest Inventory Analysis (FIA), and (B) Jenkins. In lower biomass plots, the reference above ground biomass (AGB) estimates from regional allometric equations and Jenkins equations were closely related; in plots with large biomass they were different. This finding suggests that published equations from large biomass plots are either not readily available or less represented in national scale allometric equation compiling. Models using reference AGBs calculated from regional allometric equations performed much better than those using reference AGBs calculated from Jenkins allometric equations. In these cases adjusted R^2 improvement ranged from 0.07 to 0.11. The regression model that used regional allometric equations with lidar metrics and individual tree data provided the best overall R^2 (0.79) with lowest RMSE suggesting that in most conditions regional biomass equations should be preferred over national equations. The inclusion of volumetric metrics shows that lidar variables are more sensitive to the reference AGBs calculated from regional allometric equations, and care should be taken when substituting regional equations using national scale compiled allometric equations in regional biomass studies. In addition, consistent with previous studies, the mean height of individual trees identified was chosen by both models with both reference AGBs calculated from regional allometric equations and those calculated from Jenkins equations, supporting the need to identify individual trees for biomass prediction. Based on these results, we conclude that the selection of allometric equations can influence the capacity of lidar data to estimate biomass significantly, and a careful selection of allometric equations is required for regional lidar biomass studies.

1. Introduction

Measurement of above ground biomass (AGB) at local, regional and global scales is critical for estimating global carbon storage and assessing ecosystem response to climate change and anthropogenic disturbances (Hese et al., 2005; Ni-Meister et al., 2010). While optical and radar sensors show potential to provide biomass estimates, lidar remote sensing is generally regarded as a more accurate method because lidar sensors provide more detailed information on canopy structure with the added dimension of vertical height (Chen et al., 2006; Zhao et al., 2011; Yao et al., 2011). Many previous studies have demonstrated the success of lidar estimates of AGB based on a relationship between lidar canopy height metrics (e.g., mean canopy height, canopy height percentiles, Height of Median Energy (HOME) and field-measured reference AGB calculated from regional or national scale allometric equations (Anderson et al., 2006; Lefsky et al., 1999, 2005; Lefsky, 2010; Hyde et al., 2007).

Based on the footprint size and the manner in which the lidar signal is recorded from a target, lidar systems used for biomass

* Corresponding author at: 130 Mulford Hall, #3114, University of California, Berkeley, CA 94720-3114, USA. Tel.: +1 510 642 7272; fax: +1 510 642 1477.
E-mail address: maggi@berkeley.edu (M. Kelly).

0168-1923/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.agrformet.2012.05.019

studies generally fall into two categories: large footprint full waveform systems (e.g., spaceborne Geoscience Laser Altimeter Systems (GLAS); Airborne Laser Vegetation Imaging Sensor (LVIS)) and small footprint multiple discrete-return lidar systems. Large footprint full waveform lidar systems have a footprint diameter of more than 5 m and thus are more suitable for investigations at larger scale, but lack the detailed characterization at individual tree level (Kwak et al., 2010). In contrast, small footprint, discrete return lidar instruments have a much narrower beam (i.e., less than 50 cm), and are generally believed to be able to provide more accurate biomass estimates because information at individual tree level can be characterized. Yet the application of small footprint lidar to large spatial scale studies has been relatively limited due to the necessarily small spatial coverage. Therefore, small footprint, multiple return lidar was generally selected to study regional biomass distribution.

The vast majority of research describing the use of lidar data to estimate biomass uses statistical analysis to infer a relationship between ground-based AGB, developed from field data and allometric equations, and lidar data. In the USA, reference AGB is typically calculated in either of two ways: using a national model approach (e.g., those provided by Jenkins et al., 2003, 2004) or using a regional model approach (e.g., those provided by the USFS Forest Inventory Analysis (FIA) program). The Jenkins allometric equations were originally developed for national level biomass estimation by refitting the data predicted from various allometric equations found in the literature for different tree species. In contrast, regional allometric equations were selected from published papers based on local tree studies, and thus are more suited in detailed regional AGB estimates. While regional allometric equations are widely used for regional biomass estimates, Jenkins allometric equations have also frequently been adopted when a specific allometric equation that matched closely to the study site is not readily available (Hyde et al., 2007; Kim et al., 2009; Zhao et al., 2009; Popescu, 2007; Popescu et al., 2011). In such studies, reference AGB calculated from Jenkins allometric equations is assumed to be accurate enough for regional biomass studies. However, these two sets of allometric equations differ significantly in two aspects. First, the Jenkins allometric equations have only one explanatory variable, DBH, while regional allometric equations commonly require two parameters: DBH and height. Second, the Jenkins allometric equations represent a generalization of available regional allometric equations, and their application to regional scale studies depends on whether the regional allometric equations for some species are representative of an averaged condition of those species across the country. As a result, the influence of the selection of allometric equations on regional AGB studies needs to be carefully examined, rather than replacing regional allometric equations with national scale allometric equations directly or vice versa.

These issues have been noted previously. For example, Popescu et al. (2011) discussed the errors associated with developing national scale equations by compiling species- and site-specific equation that may be biased in favor of species for which published equations exist. Zhou and Hemstrom (2010) reported that there are two main concerns about the Jenkins allometric equations when they are used for finer scale biomass analysis. First, broadleaf tree species groups may lead to biased biomass estimates; Second, the Jenkins allometric equations have only one explanatory variable: DBH and may not be sufficiently accurate for finer scale biomass analysis (Zhou and Hemstrom, 2010). Melson et al. (2011) quantified the model-selection uncertainty for five most numerous tree species in six counties of northwest Oregon, USA, and found that model-selection uncertainty is potentially large enough that it could limit the ability to track forest carbon with the precision and accuracy required by carbon accounting protocols. However, we know of no studies that investigate how the selection of allometric equations will affect the lidar metric selection and biomass models developed for biomass mapping at regional scales. Without enough knowledge of the influence of mixed use of allometric equations on lidar metric selection and AGB regression modeling, a biased understanding of lidar systems' utility on AGB estimates will be inevitable.

To better understand the role allometric equations selection can play on the regional biomass estimates with a small footprint, multiple returns lidar system, this study aims to: (1) quantify the difference in the regional biomass estimates calculated from these two sets of allometric equations; and (2) investigate how the selection of allometric equations influence lidar regression modeling (e.g., variable selection, variation explained) of AGB in a sample site in Sierra National Forest, CA. While this is only a case study for a single region, findings based on this sample study could also be relevant to forests in other regions.

2. Methodology

2.1. Study area

The Sugar Pine study area is located northeast of Oakhurst, CA, and covers approximately 36.1 km^2 (Fig. 1). This area is topographically complex with elevations ranging from 758 m to 2652 m. Primary tree species include (in order of abundance): *Calocedrus decurrens* (Incense Cedar), *Abies concolor* (White Fir), *Pinus ponderosa* (Ponderosa Pine), *Pinus lambertiana* (Sugar Pine), *Sequoiadendro giganteum* (Giant sequoia), *Quercus* (Black Oak), *Quercus ssp* (Live oak), *Cornus nuttallii* (Mountain Dogwood) and *Ainus rhombifolia* (White Alder). Of all the species, conifers comprise 89.57% and the remaining 10% are deciduous species, primarily black oak and live oak. The composition of these primary species shows a general pattern for all plots as a whole, and within each single plot, species composition among plots may vary significantly. For example, some plots can be deciduous dominant.

2.2. Field inventory

A total of 121 field inventory 0.05-acre plots were used in this study. Each plot covers 12.62 m radius area around an accurately located plot center. The first plot was randomly chosen and the following plots are placed on a 500 m grid. When a plot was less than 12.62 m from landing or road surface, it was moved 25 m in a randomly chosen cardinal direction. Plots located in inaccessible areas or on private land were not surveyed. Within each plot forest structural attributes, including species, diameter at breast height (DBH), tree height and height to live crown base (HTLCB) were recorded. The plot was also photographed from four vantage points.

2.3. Plot-level AGB estimation

Plot level AGB (including leaves branches and stems) was estimated using two sets of allometric equations: Jenkins allometric equations and regional allometric equations used by the FIA (Forest Inventory Agency) for the state of California in the pacific northwest region (Waddell and Hiserote, 2005). For each plot, AGB was calculated for each single tree and then were aggregated for each plot.

2.3.1. Jenkins allometric equations

The Jenkins allometric equations model (Jenkins et al., 2003, 2004) estimates total AGB based on tree diameter, and take the general form as follows:

$$B = \exp(b0 + b1\ \ln(\mathrm{DBH})) \quad (1)$$

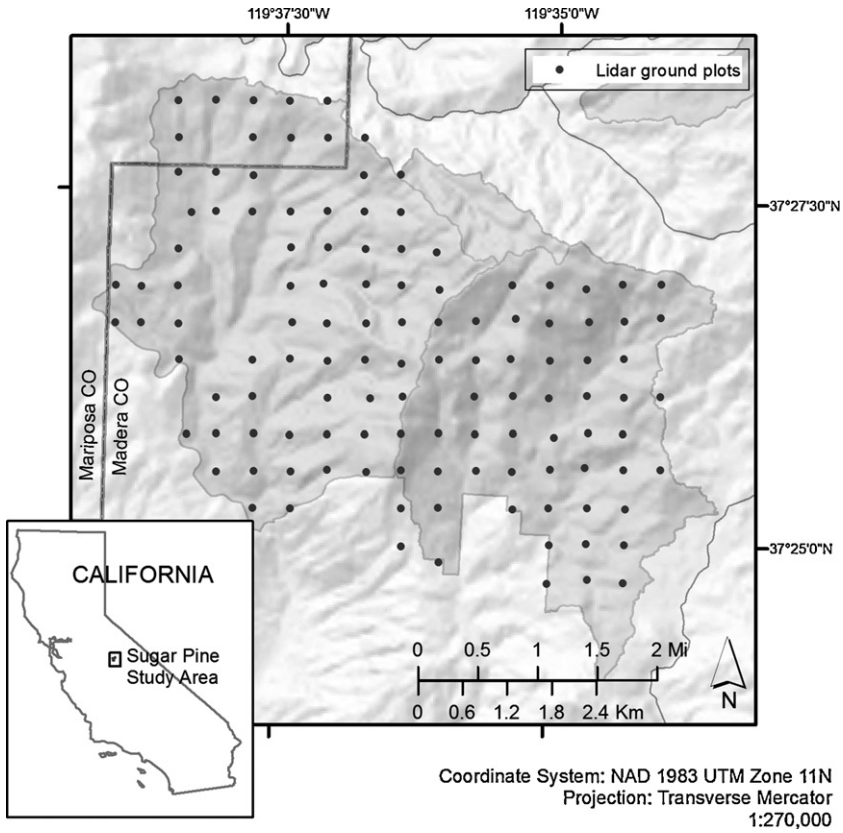

Fig. 1. The Sugar Pine study area.

where B: total aboveground biomass (kg), DBH: diameter in centimeter (cm) at breast height, ln: natural log base "e" (2.71828), $b0$, $b1$: coefficients.

There are 10 AGB equations available associated with 4 hardwood species groups and 5 softwood species group, and 1 woodland group.

2.3.2. Regional (FIA) allometric equations

The FIA program uses separate sets of equations for major AGB components: bole, branch and bark. Tree AGB is calculated and scaled from volume estimates for incense cedar as an example in CA region.

$$B = \mathrm{BOLE} + \mathrm{BRK} + \mathrm{BCH}$$

$$\mathrm{BA} = \mathrm{BDH}^2 \times 0.005454154$$

$$\mathrm{TERM} = ((1.0333(1.0 + 1.382937 \times \exp\left(-4.015292\left(\frac{\mathrm{DBH}}{10.0}\right)\right)))(\mathrm{BA} + 0.87266) - 0.174533)$$

$$\mathrm{CF} = 0.225786 + 4.44236\left(\frac{1}{\mathrm{HT}}\right)$$

$$\mathrm{CV4} = \mathrm{CF} \times \mathrm{BA} \times \mathrm{HT}$$

$$\mathrm{CVTS} = \frac{\mathrm{CV4} \times \mathrm{TERM}}{\mathrm{BA} - 0.087266}$$

$$\mathrm{BOLE} = \mathrm{CVTS} \times W_d$$

For biomass from bark:

$$\mathrm{BRK} = \exp(-13.3146 + 2.8594\ \ln(\mathrm{DBH})) \times 1000$$

For biomass from branch:

$$\mathrm{BCH} = 0.199 + 0.00381 \times \mathrm{DBH} \times \mathrm{HT} \quad (2)$$

where B: biomass of the tree, including tree stem (bole), branches and barks (kg), BOLE: biomass of the stem (kg), BRK: biomass of the bark (kg), BCH: biomass of the branches (kg), BA: basal area (ft^2), CF: cubic from factor, TERM: intermediate parameter, CV4: volume from a 1-ft stump to a 4-in. top (cubic feet), CVTS: total stem volume from ground to tip (cubic feet), W_d: wood density (kg/ft^3), HT: height of trees (ft).

Regional allometric equations are derived mainly from field studies and are thus more suitable for local scale studies. In cases that there are no allometric equations for some species, users normally substitute the equations for species that have similar growth forms. Note that regional allometric equations are selected by users, and thus the consistency of biomass estimates among regions and even trees of a single species may not be stable.

2.4. Airborne lidar data processing

Lidar data was collected in September 2007 using an Optech GEMINI Airborne Terrain Mapper (ALTM) mounted in a twin-engine Cessna Skymaster. The ALTM emits pulses of near-infrared light (1047 nm) at a rate of 100 kmhz. A maximum of four returns were recorded for each pulse. The mean point density within our study area averaged 6 points per m^2.

The point clouds were classified into two classes: ground returns and aboveground returns by the National Center for Airborne Lidar Mapping (NCALM). The Digital Terrain Model (DTM) was generated using the classified ground point and has a spatial resolution of 1 m

Table 1
Regression model cases tested.

Equations	Data inputs			
	Point clouds	Point clouds + volume	Individual trees	Combined
Biomass (FIA)	A1	A2	A3	A4
Biomass (Jenkins)	B1	B2	B3	B4

(Guo et al., 2010). The height of aboveground returns is computed by subtracting the DTM to remove the slope effects on the vertical distribution of aboveground returns.

There are two primary approaches for estimating biomass using lidar metrics: using plot-scale metrics to estimate biomass, and using individual trees (Goerndt et al., 2010). For the first approach, lidar metrics (e.g., mean canopy height, standard deviation of returns, canopy percentiles, etc.) are calculated from either an interpolated canopy height model (CHM) or from the raw point clouds. These metrics are then used to build up a regression relationship with ground reference biomass estimates. The second approach begins by identifying each individual tree using the CHM or the raw point cloud, then the biomass for each individual tree are estimated, and these are aggregated at the plot scale. Similar to biomass estimation at the area-level, lidar metrics from the individual tree level such as the sum of heights of each individual tress can also be used in a regression relationship. The second approach has resulted in higher accuracy as reported in the literature, likely because lidar metrics are derived at the individual tree level (Popescu et al., 2003; Hyyppä et al., 2001; Persson et al., 2002; Kwak et al., 2007). Although our primary focus was not to compare these two approaches, we combined these two approaches in separate models to investigate how the regression modeling performs in both cases, and when combined.

2.4.1. Lidar metrics from point clouds

All four returns corresponding to each plot were clipped from the lidar dataset and several groups of variables were derived from the height distribution of canopy returns. To avoid the inclusion of ground returns, a 1 m threshold was used to remove returns close to the ground from canopy returns. Height metrics were divided into four groups: statistical metrics, canopy height percentiles metrics, canopy transmittance metrics, and foliage profile metrics (Table 2). Statistical metrics include maximum, mean, standard deviation, kurtosis, skewness, and coefficient of variation (CV). Canopy percentile metrics are calculated from 10th, 20th up to 90th. Canopy transmittance metrics are the ratio of returns at each canopy percentile height over the total number of returns.

Lovell et al. (2003) described the methodology to derive an apparent foliage profile from a small footprint lidar. The gap probability from the top to a given height, z, is estimated by the following equations.

$$P_{\text{gap}} = 1 - \frac{\{\#z_j | z_j > z\}}{N} \tag{3}$$

$$L(z) = -\log(P_{\text{gap}}(z)) \tag{4}$$

where $\#z$ is the number of hits down to a height z above the ground and N is the total number of independent lidar shots. The apparent foliage profile is then given by Eq. (4). To test the utility of apparent foliage profile in biomass, three metrics were selected: maximum foliage area volume density, height of maximum foliage volume density and distance between top of the apparent foliage profile and height at maximum foliage volume density. Maximum foliage area volume density reflects the foliage density condition for trees with a specified region, and height of maximum foliage volume density is the height where trees have the maximum foliage density. The distance between top of the apparent foliage profile and height of maximum FAVD could be indicative of the pattern of foliage in the upper crown.

The literature has shown that a volume related measure can help in predicting biomass. For example, Ni-Meister et al. (2010) reported that wood volume is a good predictor of plot level biomass, particularly for conifer forests. Since lidar only provides height related metrics, and there are no available lidar metrics that directly relate to basal area, in this study we integrated the simplified relationship between DBH and height into the canopy percentile height metrics, and produce volumetric metrics in hope of testing the utility of this empirical relationship in predicting the biomass estimates. Specifically, we built an empirical relationship from the collected field data using the DBH and tree height for all measured trees (a total of 1248 trees) in our study area. The equation is as follows:

$$\text{DBH} = 4.37 + 1.81 \times \text{HT} \tag{5}$$

Fig. 2 shows that DBH and height of trees in our study area correlate well, suggesting that this simplified relationship may help improve the regression modeling between lidar metrics and reference biomass. For example, for height at P_10 (Table 2), the volumetric metric is calculated as follows.

$$V\_10 = P\_10 \times (P\_10 \times \text{Eq. (5)})^2 \tag{6}$$

2.4.2. Lidar metrics from individual trees

A variety of methods are available for isolating individual trees from the CHM (Popescu et al., 2003; Hyde et al., 2007; Persson et al., 2002; Kwak et al., 2007). In this study we used a newly developed method developed by Li et al. (2012) to isolate individual trees from the point cloud, which is found to generate superior result with the accuracy of 94% in the mixed conifer forests in the Sierra Nevada region. The algorithm adopts a top-to-bottom region growing approach that segments trees individually and sequentially from the tallest to the shortest from a lidar point cloud. After the height and canopy radius were derived, tree-level metrics (e.g.,

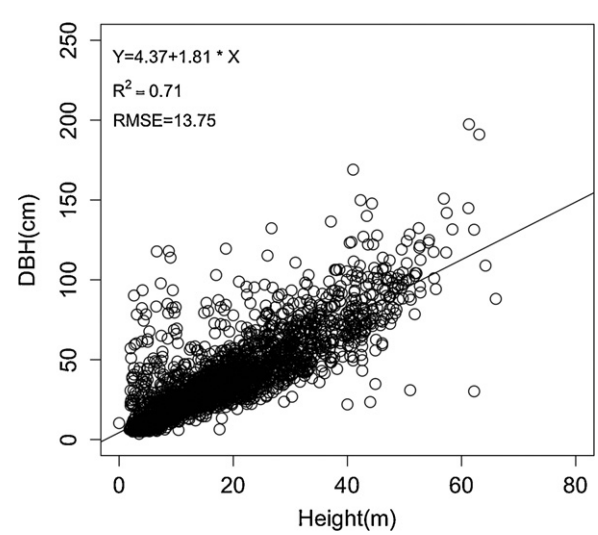

Fig. 2. The general relationship between DBH and height at the individual-tree level.

Table 2
Metrics derived from the height distribution of lidar data.

Lidar metrics from point clouds		Lidar metrics from individual tree detection	
Label	Description	Label	Description
Max_ht	Top height within a plot[a]	Ind_sum_ht	Top height within a plot
Mean_ht	Mean of heights	Ind_mean_ht	Mean of heights
Std_ht	Standard deviation of heights	Ind_sd_ht	Standard deviation of heights
Skew_ht	Skewness of heights	Ind_skew_ht	Skewness of heights
Kur_ht	kurtosis of heights	Ind_kurt_ht	Kurtosis of heights
CV_ht	Coefficient of heights	In_sum_r	Sum of canopy radius
QMCH	Quadratic mean of height	Ind_mean_r	Mean of canopy radius
Canopy_cover	Transmittance at 1 meters above ground	Ind_sd_r	Standard deviation of canopy radius
Max_fp	Maximum foliage profile density[c]	Ind_skew_r	Skewness of canopy radius
Max_fp_height	Height at Max_fp	Ind_kurt_r	Kurtosis of canopy radius
Top_Maxfp_ht	Distance between tree top and the height at Max_fp	Ind_num_tree[e]	Number of trees
P_10. . .P_90	Percentile heights		
D_10. . .D_90	Transmittance at each percentile heights[b]		
V_10. . .V_90	Volumetric metrics at each percentile height[d]		

[a] Height and radius related metrics are in the unit of m.
[b] Transmittance related metrics are unitless.
[c] Foliage profile related metrics are in the unit of m^2/m^3.
[d] Volume related metrics are in the unit of m^3.
[e] The number of trees metric is unitless.

mean height, skewness of heights of all individual trees with a plot) were calculated (Table 2).

2.5. Statistical models

Multiple linear regression method was the primary statistical method used to study the relationship between biomass and predictive variables from lidar data (Hudak et al., 2006; Goerndt et al., 2010). We compared the regression performance across eight models: (1) point clouds alone, (2) point clouds with an empirical relationship between DBH and height (i.e., volume), (3) individual tree-level metrics, and (4) all data combined, and across two allometric equations – (A) FIA, and (B) Jenkins. This resulted in eight model results, which are summarized in Table 1.

2.6. Multiple linear regression

To correct for both non-normality and heteroscedascity (Hudak et al., 2006), the natural log transformation was generally applied to the response variable (i.e., biomass). We also tested the model without natural log transformation on the response variable. The variable selection was conducted using a subset regression technique that using an exhaustive search. Specifically, we used *regsubsets* function in the leaps Package for R (R Development Core Team, 2010). The model statistic used to determine the best subsets was Mallows (1973) Cp statistic which is defined as follows:

$$\mathrm{Cp} = \frac{\mathrm{SSE}}{\mathrm{MSE}_{\mathrm{full}}} - N + 2P \tag{7}$$

where SSE is the error sum of squares of the reduced model with P parameters (including the intercept), MSE is the mean square error of the full model, and N is the number of samples. The models which yield the lowest values of Cp will tend to be similar to those that yield the highest values of adjusted R-squared. The Cp criterion tends to favor models with fewer parameters and was thus selected to determine optimal model for our analysis. To correct for bias in natural log-transformed allometric equations, we adopted a correction factor (Sprugel, 1983).

To compare the performance of regression models, we used two criteria: adjusted R^2 and root mean squared error (RMSE) based on a 10-fold cross validation analysis. To assess the contribution of each selected metrics to the optimal model, we used the *lmg* method in package *relaimpo*, which is a metric that decomposes adjusted R^2 into non-negative contributions that automatically sum to the total R^2.

3. Results

3.1. AGB comparison: Jenkins allometric equations vs. FIA regional allometric equations

There were large differences in biomass density statistics between the two allometric equations. The range of biomass density in our plots is large: 38.6–1132.9 Mg/ha for Jenkins allometric equations compared to 28.8–1442.0 Mg/ha for regional allometric equations. The lower values at 2.5th percentile are 71.4 Mg/ha and 51.3 Mg/ha respectively and the upper values at 97.5th percentile are 645 Mg/ha and 667 Mg/ha for the Jenkins allometric equations and regional biomass equation respectively. The smaller range of biomass density derived from the Jenkins allometric equations may be attributed to the fact that the Jenkins equations reflects the average condition of forest biomass across the country.

The relationship between above ground biomass derived from the Jenkins allometric equations and from the regional allometric equations is shown in Fig. 3. For plots with low biomass, AGB from these two sets of allometric equations are more closely correlated while AGB estimates from regional allometric equations for large biomass plots are significantly less that those derived from the national-scale allometric equation.

3.2. Multiple linear regression models

The adjusted R^2 and RMSE based on 10-fold cross validation test for regression modeling for all models are shown in Table 3, and

Table 3
Results of multiple linear regression models for predicting biomass.

Model	Response variable	AdjR^2	RMSE (Mg/ha)
A1	Biomass (FIA)	0.69	285
A2	Biomass (FIA)	0.72	268
A3	Biomass (FIA)	0.71	355
A4	Biomass (FIA)	0.79	252
B1	Ln (Biomass (Jenkins))	0.62	260
B2	Ln (Biomass (Jenkins))	0.62	266
B3	Ln (Biomass (Jenkins))	0.62	253
B4	Ln (Biomass (Jenkins))	0.68	211

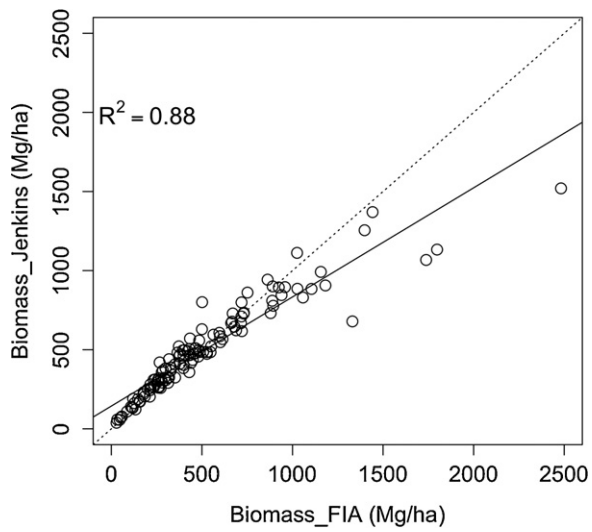

Fig. 3. Biomass comparison derived from two allometric equations. The dashed line is the 1:1 line.

Fig. 4 shows the variables selected and the decomposed adjusted R^2 for each variable. Models with higher adjusted R^2 and low RMSE are indicative of better predictive ability of biomass with lidar-derived variables. Within model type ('A' or 'B'), higher R^2 values correspond to lower RMSE. Models A1–A4 (regional) performed better than B1–B4 (national), improving R^2 by 0.07–0.11, suggesting that in most conditions regional allometric equations are preferred. A4 was the best model with the highest adjusted R^2 (0.79) and the lowest RMSE for all 'A' models (252). The slightly higher cross validation RMSE for A1–A4 as compared to B1–B4 models can possibly be explained by the fact that estimates of above ground biomass estimates calculated from regional allometric equations are generally higher than those calculated from Jenkins allometric equations (Fig. 4).

The results shows that models with all lidar variables (i.e., model A4 and model B4) produced the highest adjust R^2. The variables selected by model A4 based on regional biomass equations include mean of heights derived from individual tree identification, volumetric metric at 90th percentile height, mean of heights derived from point cloud, 50th percentile height, and transmittance at 50th and 20th percentile heights in which the lidar height variables explained the most variance, followed by two transmittance variables. Similarly, four lidar height variables (mean of heights derived from individual tree identification, 70th percentile height, volumetric metric at 70th and 50th percentile height, height at maximum foliage profile density), one transmittance and one canopy radius (transmittance at 1 m above ground and sum of canopy radius derived from individual tree identification, respectively) were selected by model B4 based on the Jenkins allometric equation. In both cases, mean height of individual trees explained a large proportion of variation. This is consistent with previous studies showing that individual tree results have better predictive power for biomass estimation.

Without the addition of volumetric variables, model A1 explained more variance than B1 (0.69 vs. 0.62). The lidar metrics optimized for regression modeling differed significantly. A1 selected six height variables, two foliage profile variables and one transmittance variable, in which height related variables contributed 92% of the explained variation. In comparison, maximum foliage profile and canopy cover variables together explained 21% of the explained variation in B1 model, and height related variables contributed to the rest of explained variation.

Adopting an empirical relationship between DBH and height explained more variation: model A1 (0.69) was improved to model A2 (0.72) when reference AGB estimates from regional allometric equations was used, likely because lidar provided primarily information related to heights, and integrating an empirical relationship into lidar height metrics can lead to a better correlation with reference AGB estimates calculated from volume based regional allometric equations. The Jenkins allometric equations, however, have only one parameter (i.e., DBH) as input. Consequently, the simple empirical relation is not able to improve the correlation between lidar metrics and reference AGB estimates with the Jenkins allometric equation. When volumetric metrics were included, four such metrics contributed evenly to the biomass regression modeling with reference AGB calculated from regional allometric equations, and other two transmittance metrics and a foliage profile related variable, distance between tree top and the height at maximum foliage profile density, explained very few variations. When the Jenkins equations were used to calculate reference AGB, no volumetric metric was adopted and the 50th percentile height was the most significant variable. Both model A2 and model B2 selected transmittance variables. Further, model A2 explained much more variance than B2 (0.72 vs. 0.62). The empirical relationship between DBH and height from field data only improved the relationship between lidar metrics and reference AGB calculated from the regional allometric equation.

Model A3 and model B3 both chose the metric sum of heights derived from individual tree identification and standard deviation of heights derived from individual tree identification. Also the reference AGB calculated from the regional allometric equations showed a better correlation with lidar metrics than reference AGB calculated from the Jenkins allometric equations because reference AGB using regional allometric equations provided more accurate biomass estimates. The variance explained for A3 is 0.71 which is much higher than 0.62 explained by model B3.

4. Discussion

Our results show that published biomass allometric equations from regional and national sources can give substantial variation in plot-level biomass estimates, especially in denser plots. The variation may suggest that large biomass plots are not well represented in the national scale allometric equations derivation due to the fact that it is more time consuming and labor intensive to harvest large biomass plots. In spite of the fact that the biomass density for two sets of allometric equations hold similar patterns statistically, the differences in the biomass density are obvious, indicating the importance of assessing the influence of varied allometric equations in predicting biomass with small footprint lidar systems.

Since both sets of allometric equations employ DBH as input, and the regional allometric equations use one more additional variable (i.e., height), the difference between biomass density at the plot level mainly reflects the aggregated variation of height for all individual trees in a plot. As a result, the comparison of AGB using those two sets of allometric equations shows that tree heights in our study area may generally be higher than the average heights of the same species group across the country.

In models with reference above ground biomass calculated from regional biomass equations, the integration of a simplified, empirical relationship between DBH and height generally improved the performance of the regression models. Its utility can be explained by the fact lidar-derived variables are directly related to height, while reference AGBs calculated from either Jenkins allometric equations or regional allometric equations share a common input, DBH. The simplified transformation from height metrics to volumetric metrics made the association between reference AGBs and

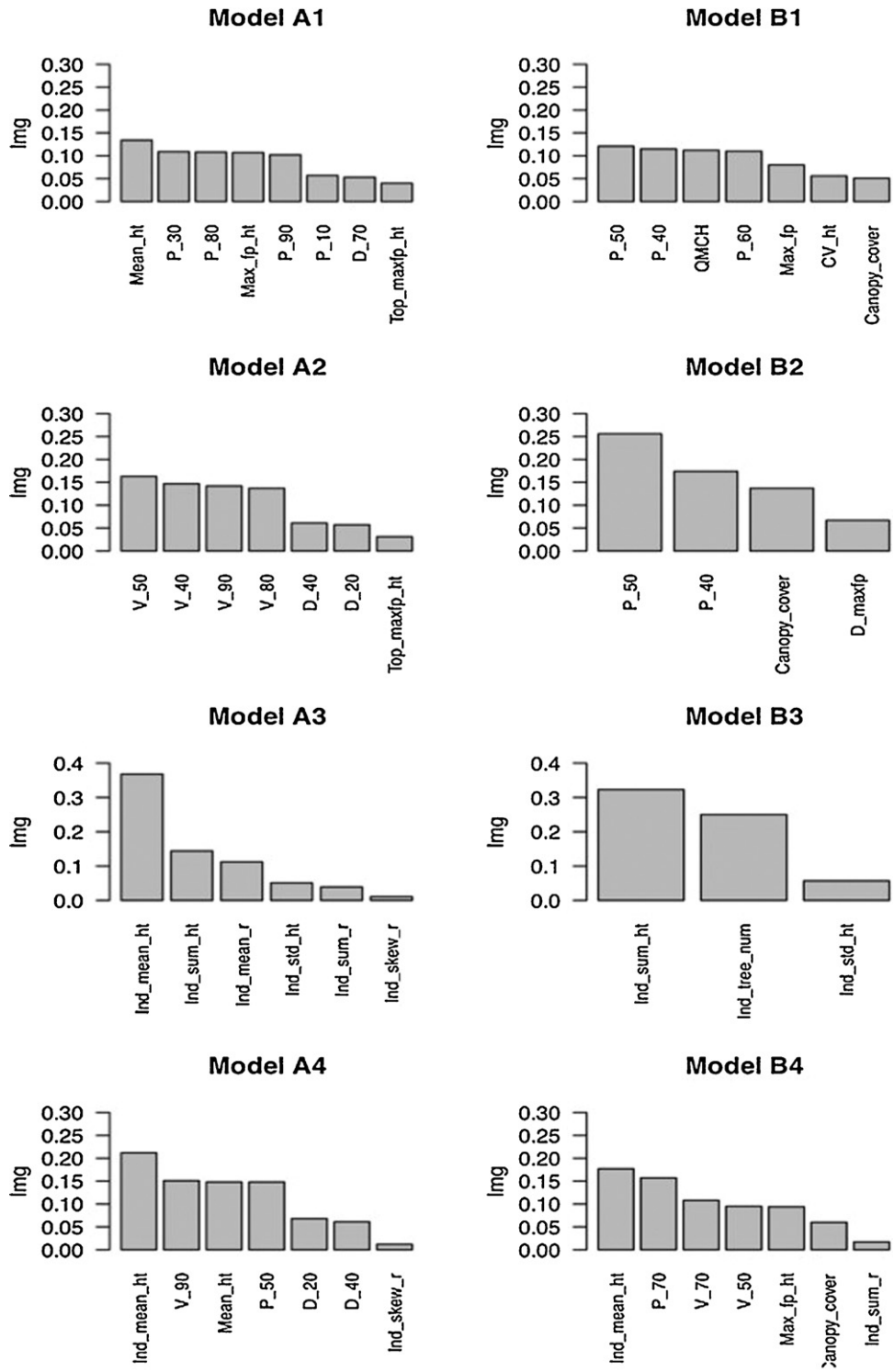

Fig. 4. Variables of importance for selected models.

lidar-derived metrics more direct and close. However, the addition of this empirical relationship did not help in models with AGB calculated from Jenkins equations.

The models using reference AGB calculated from regional biomass equations performed better than those using Jenkins equations. Models with higher adjusted R^2 and low RMSE are indicative of better predictive ability of biomass with lidar-derived variables. For example, in the case where we included volumetric metrics, the explained variance improved from 0.68 to 0.79 for multiple linear models. The superior performance of models using regional biomass allometric equations suggests that reference AGBs using regional biomass equations are not only more accurate to record the biomass estimates, and also are more closely related to the lidar-derived metrics. As a result, one should be cautious when using national scale allometric equations in regional biomass studies. Additionally, the lidar metrics selected varied significantly across the multiple linear models. This was expected given the large difference in the reference AGB in large biomass plots. However, in

case of a study area with only lower biomass plots, our findings may not necessarily hold true.

In the multiple linear models, height-related metrics explained more variation than canopy transmittance and foliage profile related metrics. This implies that reference above ground biomass is primarily related to height and/or DBH directly, and canopy transmittance and profile based metrics contain very limited information to the biomass estimates.

As suggested by Hudak et al. (2006), intensity value, particularly mean intensity values, also prove surprisingly useful in predicting forest basal area and tree density. This suggests that intensity values could potentially improve the predictive power of regression modeling of biomass. Further research is needed to determine how the addition of lidar intensity will improve the regression modeling performance in both reference AGB calculated from regional allometric equations and reference AGB calculated from Jenkins allometric equations. Similarly, integrating optical and Radar data into lidar variable also proved to be useful for biomass prediction (Hyde et al., 2006; Nelson et al., 2007). Thus, assessing influence of varied allometric equations when high spatial resolution imagery is integrated is also worth to be studied.

Although non-parametric approaches are increasingly used to model forest parameters with airborne discrete return lidar data (e.g., Hudak et al., 2006; Yu et al., 2011), further exploration is required for more refined use of non-parametric approaches in biomass estimates using lidar data.

5. Conclusion

In this case study, we tested the sensitivity of lidar metrics to reference above ground biomass estimates calculated from two sets of frequently used allometric equations: the national-scale Jenkins allometric equations and regional allometric equations from the FIA, in a conifer dominated forest site in the Sierra National Forest, CA. We evaluated the performance of these two sets of equations by comparing adjusted R^2 and RMSE across a set of eight models. Our main findings are as follows:

(1) Discrete return lidar-derived variables are more sensitive to reference AGBs calculated from regional allometric equations with respect to the amount of variation explained, variables selected and variables of importance played in multiple linear regression. As a result, care should be taken when substituting national scale allometric equations in regional biomass scale studies and vice versa.
(2) The mean height of individual trees explained a large proportion of variation in the regression modeling of biomass. This is consistent with previous studies that individual tree results have better predictive power for biomass estimation.
(3) This analysis further supports the concerns that national scale allometric equations for regional biomass estimate might favor species with more published allometric equations.
(4) This study illustrates the value of integration of an empirical relationship between DBH and height into multiple linear regression modeling for more accurate biomass estimates.
(5) The availability of, and uncertainly in, allometric equations poses a practical limit to the accuracy with which LiDAR can predict biomass.

The use of airborne lidar system in forest biomass estimation is widely accepted as the most accurate sensors, but rigorous procedures should be taken in selecting appropriate allometric equations to produce reliable and consistent reference biomass estimates. Inconsistencies should be expected if this source of variability is not controlled.

Acknowledgements

This paper is SNAMP Publication Number 14. The Sierra Nevada Adaptive Management Project is funded by USDA Forest Service Region 5, USDA Forest Service Pacific Southwest Research Station, US Fish and Wildlife Service, California Department of Water Resources, California Department of Fish and Game, California Department of Forestry and Fire Protection, and the Sierra Nevada Conservancy.

References

Anderson, J.E., Martin, M.E., Smith, M.L., Dubayah, R.O., Hofton, M., Hyde, P., 2006. The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. Remote Sens. Environ. 105 (3), 248–261.
Chen, Q., Baldocchi, D., Gong, P., Kelly, M., 2006. Isolating individual trees in a savanna woodland using small footprint LIDAR data. Photogram. Eng. Remote Sens. 72 (8), 923–932.
Goerndt, M.E., Monleon, V., Temesgen, H., 2010. Relating forest attributes with area-based and tree-based LiDAR metrics for western Oregon. West. J. Appl. For. 25, 105–111.
Guo, Q., Li, W., Yu, H., Alvarez, O., 2010. Effects of topographic variability and lidar sampling density on several DEM interpolation methods. Photogram. Eng. Remote Sens. 76 (6), 701–712.
Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., Neumann, K., Riedel, T., Schröter, K., 2005. Global biomass mapping for an improved understanding of the CO_2 balance – the Earth observation mission Carbon-3D. Remote Sens. Environ. 94, 94–104.
Hudak, A.T., Evans, J.S., Falkowski, M.J., Crookston, N.L., Smith, A.M.S., Gessler, P.E., Morgan, P., 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral data. Can. J. Remote Sens. 32 (2), 126–138.
Hyde, P., Dubayah, R., Walker, W., Blair, J.B., Hofton, M., Hunsaker, C., 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (lidar, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sens. Environ. 102, 63–73.
Hyde, P., Nelson, R., Kimes, D., Levine, E., 2007. Exploring LiDAR–RaDAR synergy – predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR, and InSAR. Remote Sens. Environ. 106 (1), 28–38.
Hyyppä, J., Kelle, O., Lehikoinen, M., Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height model produced by laser scanners. IEEE Trans. Geosci. Remote Sens. 39, 969–975.
Jenkins, C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49 (1), 12–35.
Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2004. Comprehensive database of biomass equations for North American tree species. General Technical Report NE-319. USDA Forest Service, Newtown Square, Pennsylvania, USA.
Kim, Y., Yang, Z., Cohen, W.B., Pflugmacher, D., Lauver, C.L., Vankat, J.L., 2009. Distinguishing between live and dead standing tree biomass on the north rim of Grand Canyon National Park, USA using small-footprint lidar data. Remote Sens. Environ. 113, 2499–2510.
Kwak, D.A., Lee, W.K., Lee, J.H., Biging, G.S., Gong, P., 2007. Detection of individual trees and estimation of tree height using lidar data. J. For. Res. 12, 425–434.
Kwak, D.A., Lee, W.K., Cho, H.K., Lee, S.H., Son, Y., Kafatos, M., Kim, S.R., 2010. Estimating stem volume and biomass of Pinus koraiensis using LiDAR data. J. Plant Res. 123 (4), 421–432.
Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., Shugart, H.H., 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland. Remote Sens. Environ. 67 (1), 83–98.
Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. Geophys. Res. Lett. 32 (22), L22S02.
Lefsky, M.A., 2010. A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system. Geophys. Res. Lett. 37, L15401.
Li, W., Guo, Q., Jakubowski, M., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogram. Eng. Remote Sens. 78 (1), 75–84.
Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., Coops, N.C., 2003. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Can. J. Remote Sens. 29, 607–622.
Mallows, C.L., 1973. Some comments on Cp. Technometrics 15 (4), 661–675.
Melson, S.L., Harmon, M.E., Fried, J.S., Domingo, J.B., 2011. Estimates of live-tree carbon stores in the Pacific Northwest are sensitive to model selection. Carbon Balance Manag. 6, 2.
Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Ranson, J., Sun, G., Blair, J.B., 2010. Assessing general relationships between above-ground biomass and vegetation structure parameters for improved carbon estimate from vegetation lidar. J. Geophys. Res. 115, G00E11, http://dx.doi.org/10.1029/2009JG000936.
Nelson, R.F., Hyde, P., Johnson, R., Imhoff, M.L., 2007. Investigating RaDAR-LiDAR synergy in a North Carolina pine forest. Remote Sens. Environ. 110 (1), 98–108.

Popescu, S.C., Wynne, R.H., Nelson, R.E., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. 29 (5), 564–577.
Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy 31, 646–655.
Popescu, S.C., Zhao, K., Neuenschwander, A., Lin, C., 2011. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sens. Environ. 115 (11), 2786.
Persson, A., Holmgren, J., Södermann, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogram. Eng. Remote Sens. 68 (9), 925–932.
Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64 (1), 209–210.
R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org.
Waddell, K.L., Hiserote, B., 2005. The PNW-FIA Integrated Database User Guide: A Database of Forest Inventory Information for California, Oregon, and Washington (V2.0). Pacific Northwest Research Station, Portland, Oregon, USA.
Zhou, X., Hemstrom, M.A., 2010. Timber volume and aboveground live tree biomass estimations for landscape analyses in the Pacific Northwest. Gen. Tech. Rep. PNW-GTR-819. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 31.
Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on random forest technique. ISPRS J. Photogram. Remote Sens. 66 (1), 28–37.
Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sens. Environ. 115 (11), 2965–2974.
Zhao, F., Yang, X., Schull, M., Roman-Colon, R., Yao, T., Wang, W., Zhang, Q., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., Newnham, G.J., Richardson, A.D., Ni-Meister, W., Schaaf, C.L., Woodcock, C.E., Strahler, A.H., 2011. Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using Echidna® ground-based lidar. Remote Sens. Environ. 155 (11), 2954–2964.
Zhao, K., Popescu, S.C., Nelson, R.F., 2009. Lidar remote sensing of forest biomass: a scale-invariant approach using airborne lasers. Remote Sens. Environ. 113, 182–196.