

Research Article

Evaluating Lidar Point Densities for Effective Estimation of Aboveground Biomass

Zhuoting Wu¹, Dennis Dye¹, Jason Stoker², John Vogel¹, Miguel Velasco¹, and Barry Middleton¹

¹US Geological Survey, Western Geographic Science Center, United States ²National Geospatial Program, US Geological Survey, United States

Publication Date: 9 February 2016

DOI: https://doi.org/10.23953/cloud.ijarsg.40



Copyright © 2016 Zhuoting Wu, Dennis Dye, Jason Stoker, John Vogel, Miguel Velasco, and Barry Middleton. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract The U.S. Geological Survey (USGS) 3D Elevation Program (3DEP) was recently established to provide airborne lidar data coverage on a national scale. As part of a broader research effort of the USGS to develop an effective remote sensing-based methodology for the creation of an operational biomass Essential Climate Variable (Biomass ECV) data product, we evaluated the performance of airborne lidar data at various pulse densities against Landsat 8 satellite imagery in estimating above ground biomass for forests and woodlands in a study area in east-central Arizona, U.S. High point density airborne lidar data, were randomly sampled to produce five lidar datasets with reduced densities ranging from 0.5 to 8 point(s)/m², corresponding to the point density range of 3DEP to provide national lidar coverage over time. Lidar-derived aboveground biomass estimate errors showed an overall decreasing trend as lidar point density increased from 0.5 to 8 points/m². Landsat 8-based aboveground biomass estimates produced errors larger than the lowest lidar point density of 0.5 point/m², and therefore Landsat 8 observations alone were ineffective relative to airborne lidar for generating a Biomass ECV product, at least for the forest and woodland vegetation types of the Southwestern U.S. While a national Biomass ECV product with optimal accuracy could potentially be achieved with 3DEP data at 8 points/m², our results indicate that even lower density lidar data could be sufficient to provide a national Biomass ECV product with accuracies significantly higher than that from Landsat observations alone.

Keywords 3DEP; Aboveground Biomass; Essential Climate Variable (ECV); Landsat; Lidar; Point Density; Quality Level

1. Introduction

Accurate estimation, mapping and monitoring of the amount of carbon stored in terrestrial vegetation is crucial to reliable analysis, understanding and projection of the global carbon cycle and its interactions with land use and climate change. Aboveground biomass is defined as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS) (Bojinski et al., 2014). An accurate assessment of the amount of vegetation biomass is crucial for quantifying carbon stocks and the potential of vegetation to sequester carbon, which can have a direct influence on local, regional and

global climate. Aboveground biomass is a common input variable in ecosystem process and climate models, but large uncertainties as well as infrequent updating of biomass data are major potential sources of uncertainties in the model outputs.

It is widely recognized that active remote sensing systems, especially light detection and ranging (lidar), represent the future of large-scale estimation and mapping of terrestrial biomass (Chen et al., 2012; Evans et al., 2009; Zhao et al., 2009; Hall et al., 2005; Lefsky et al., 2002; Su et al., 2016; Gregoire et al., 2016), potentially supporting the operational production of a Biomass ECV product at the national scale. Lidar is an active remote sensing system that can measure the three-dimensional (3-D) structural characteristics of trees and other vegetation which are not directly captured by passive optical land imaging systems such as Landsat. Such 3-D structural information is critical for improved aboveground biomass estimation with greater accuracy (Dubayah and Drake, 2000; Gregoire et al., 2016). Past and ongoing research efforts have developed algorithms that use airborne lidar to quantify forest inventory variables (generally on local spatial scales), such as tree height (Hyppa et al., 2008; van Leeuwen and Nieuwenhuis, 2010; Næsset, 1997), stand structure (Lefsky et al., 2005; Næsset et al., 2005; Kane et al., 2010), leaf area index and cover (Morsdorf et al., 2006; Jensen et al., 2006), timber volume (Yu et al., 2004; Maltamo et al., 2006), and biomass (Boudreau et al., 2008; Lim and Treitz, 2004: Næsset and Gobakken, 2008: Lefsky et al., 2002; Zhao et al., 2009). Given lidar data with sufficiently high pulse density, determination of these fundamental biometric variables (tree height, tree density, fractional cover, and stand structure) with a low level of uncertainty is relatively straightforward. Yet, using these variables to estimate aboveground biomass introduces potentially large errors due to additional complexities and uncertainties, such as field sampling design, model selections and the accuracy of the allometric equations that are used in the estimation procedure (Gregoire et al., 2016).

In 2012, USGS initiated the 3D Elevation Program (3DEP), leveraging partnerships with other federal and state agencies, aiming to provide consistent, standardized national lidar coverage in the coming years (Sugarbaker et al., 2014). According to the National Enhanced Elevation Assessment study, the current level of lidar data availability is at a relatively low quality level (QL3) of <1 points/m² on a 25-year repeat cycle, although the quality level that provides the largest user satisfaction is high quality (QL1) at 8 points/m² on an annual cycle. However, high data quality / pulse density comes with high data acquisition cost, and therefore the most efficient cost-benefit recommendation is to acquire quality level 2 (QL2) at 2 points/m² lidar data on an 8-year cycle for the US (Sugarbaker et al., 2014). Typically, the lower the density of laser pulses, the less comprehensive the tree structure characteristics captured by the laser scanning systems, and therefore the higher the uncertainty of aboveground biomass estimates derived from lidar for individual trees. Yet, for the purpose of a Biomass ECV, it is less critical to accurately depict structure characteristics of each individual tree, and therefore lower pulse density at plot and stand levels can be a compromise to capture larger areas for a national scale operational product.

While the effectiveness of using lidar to map vegetation and estimate biomass has been demonstrated, the high cost of airborne lidar data acquisition poses practical limitations on their application to large-scale biomass mapping. Contrary to the high cost and sporadic acquisition of lidar, Landsat satellites provide wall-to-wall data coverage every 8 days with free access, and can be potential candidates for operational products such as a Biomass ECV. Landsat 8, the latest in the series of Landsat satellites, maintains similar spectral bands as the Operational Land Imager, and improvements in its radiometric performance (greater signal-to-noise ratio and 12-bit quantization) make it potentially more effective for mapping vegetation. However, the lack of 3-D structural information from Landsat data leads to relatively low accuracy in estimating aboveground biomass over large geographic areas (Foody et al., 2003; Hall et al., 2006; Lu, 2005; Avitabile et al., 2012). In the meantime, without suitable accommodations or modifications, the long acquisition cycle of the current 3DEP program and low point density may prevent development and implementation of a

feasible Biomass ECV product in the near future. Here, we aim to facilitate informed decision making and planning for a Biomass ECV product by examining the key factors that underlie this issue in a pilot study area. Specifically, we used our established research field sites in east-central Arizona (Wu et al., 2015) to (1) evaluate the impacts of different density/quality levels lidar data on aboveground biomass estimation accuracy, (2) compare the accuracies of aboveground biomass estimates derived from lidar and Landsat 8 data, and (3) consider the implications of our results for the opportunities and feasibility of an operational national Biomass ECV product.

2. Methods

2.1. Study Area

The study area is located near Point of Pines on the San Carlos Apache Reservation in east-central Arizona (33.39^o N, 109.82^o W, 1829 – 2134 m ASL). Vegetated landscapes in the study area are primarily comprised of forests and woodlands dominated by ponderosa pine (*Pinus ponderosa*), two-needle pinyon (*Pinus edulis*), alligator juniper (*Juniperus deppeana*), Arizona white oak (*Quercus arizonica*), and gambel oak (*Quercus gambelii*). 39 square plots of 30 m by 30 m each were established within the study area, including 21 woodland plots and 18 forest plots respectively. Ponderosa pine forests are continuous and extend throughout the study area; whereas pinyon, juniper and oak species are inter-mixed at lower elevation. Mixed woodland species plots usually have higher tree density and lower biomass than forest plots, which is consistent with the landscape characteristics of the study area.

2.2. Field Biometrics

Field biometric data were collected during three intensive field campaigns from June to August in 2013. To determine the corner coordinates of each plot, two survey-grade Global Positioning System (GPS) receivers were used; a base station GPS receiver was placed near each plot location, with a roving GPS receiver used to accurately map the plot corner points, referenced to the base station receiver. Tree species were recorded and individual tree heights were measured using a Nikon Forestry Pro Laser Range Finder 8381 for tall and mature trees, and an extension pole marked with 0.1 m increments for small and young trees. Tree heights were measured twice and the final tree height for each tree was determined as the average of the two measurements. Both methods achieved vertical accuracies of 0.3 m. The diameter at breast height (dbh) was measured at 1.37 m height for ponderosa pines, and no trees smaller than 1.4 m in height were included in the measurements. For woodland species, the root collar diameter of each stem was measured, and the equivalent diameter at root collar (edrc) was computed using the equation (1) below:

$$edrc = \sqrt{a^2 + b^2 + \dots + x^2},$$
 (1)

Where a, b, ..., x are diameter at root collar for each stem that is 3.8 cm or larger (Chojnacky and Ott, 1986).

Biomass of each individual tree measured in the field was calculated using validated species-specific allometric equations derived from field studies (Clary and Tiedemann, 1986; Grier et al., 1992; Kaye et al., 2005; Navar, 2009) in similar environments.

2.3. Remote Sensing Data

2.3.1. Airborne Lidar

Airborne lidar data were acquired through the U.S. Geological Survey Geospatial Product and Service Contract (GPSC), and the data collection was completed by Woolpert, Inc. (Dayton, OH). The lidar data were acquired using a Leica ALS70 Multiple Pulses in Air (MPiA) lidar sensor, onboard a Cessna 402 on August 8 – 9th, 2013 (Table 1). The Leica ALS70 system used in this study uses a laser that operates in the near infrared spectral region (1064 nm). A total of 52 flight lines were flown to provide comprehensive coverage of the study area at a specified minimum pulse density that equated to an average of 12 points/m² in open areas. When the sensor calibration, data acquisition, and GPS processing phases were complete, Woolpert then processed individual flight lines to derive a raw point cloud file in LASer (LAS) 1.2 format. Ground and non-ground classes were created using the calibrated point cloud files, and survey ground control data were imported and incorporated into an accuracy assessment. Woolpert calculated the vertical accuracy by comparing the lidar bare earth points to the ground surveyed QA/QC points, and reported a vertical accuracy of 0.072 m in flat, open areas, along with a bare-Earth DEM accuracy within 0.068 m, both at a 95 percent confidence level.

Parameter	Performance
Post spacing (minimum)	0.3 m
Aboveground level average flying height	1,829 m
Mean sea level average flying height	3,627 m
Average ground speed	249 kph
Field of view (full)	10 degrees
Pulse rate	292.0 kHz
Scan rate	72.6 Hz
Side lap (minimum)	25%

2.3.2. Landsat 8 Satellite Imaging

Cloud-free Landsat 8 surface reflectance data for September 24th, 2013 were acquired from USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). Normalized Difference Vegetation Index (NDVI) is the most widely used vegetation index for mapping vegetation, taking advantage of strong absorption at the visible spectral region and strong reflectance at the near-infrared spectral region of chlorophyll in the leaves (Rouse et al., 1974; Sellers, 1985; Tucker, 1979). NDVI has been used to estimate biomass (Labrecque et al., 2006; Zheng et al., 2004) as it is a sensitive indicator of canopy structure and leaf chemical content and leaf area, and is especially useful in low-biomass dryland forest and woodland ecosystem with low leaf area index (Gamon et al., 1995) where saturation of the NDVI is not typically observed. NDVI was computed using the red (R) and near-infrared (NIR) bands as (NIR - R)/(NIR + R) using Landsat 8 surface reflectance data from ESPA. Surface reflectance of all 8 visible, near-infrared and shortwave infrared bands as well as computed NDVI were used to estimate aboveground biomass for forests and woodlands in our study area.

2.4. Aboveground Biomass Estimation

2.4.1. Lidar Derived Metrics

Because of the flight-lines overlap, superfluous lidar points were ignored, ensuring a more homogenous sampling density throughout the study area. To isolate the impacts of pulse density on various forest height metrics, a random thinning approach was applied to the original dataset at 12

points/m² to create a gradient of lower point densities at 8, 4, 2, 1 and 0.5 point(s)/m², respectively (Figure 1). For each reduced density point cloud dataset, ground points were classified and tree heights were then calculated relative to the ground surface. A common approach to generate aboveground biomass estimates from remote sensing is to establish statistical regression between laser height measurements and ground sampling plots (Gregoire et al., 2016). Various descriptive statistical metrics commonly used for estimating aboveground biomass (Lefsky et al., 1999; Means et al., 1999; Næsset, 1997; Zhao et al., 2009) were calculated from the entire vertical profile of the vegetation lidar returns to best capture the relationship between the lidar pulse data and field-based estimates of biomass at the plot level. All lidar data processing was conducted using LAStools (http://lastools.org) developed by Martin Isenburg (Isenburg and Schewchuck, 2007; Hug et al., 2004). Basic height statistics metrics, including maximum, minimum, mean, standard deviation, skewness, kurtosis, and quartic average, were computed from the canopy height point cloud data for each field plot. Canopy cover and tree density were computed from the number of vegetation returns versus total returns within each plot. Height percentiles can also be effective in capturing the different levels of laser beam penetration through the tree canopies and therefore reflect the structure characteristics of the vegetation, and have been demonstrated to be very useful in estimating aboveground biomass (Lim and Treitz, 2004). Using LAStools, 10th, 25th, 50th, 75th, and 90th percentiles of height were also calculated from the canopy height point cloud.



Figure 1: Lidar Point Clouds at (A) 8 Points/m² And (B) 2 Points/m² From Side View (Top) and Top View (Bottom), Corresponding to 3DEP Quality Level 1 (QL1) And QL2, Respectively

2.4.2. Regression Models

Multiplicative models were estimated as linear regressions using logarithmic transformations for predictor metrics and field-based aboveground biomass estimates, which have been found to be suitable to estimate vegetation biomass (García et al., 2010; Næsset and Bjerknes, 2001; Næsset and Økland, 2002). The multiplicative model was formulated as equation (2):

International Journal of Advanced Remote Sensing and GIS

$$AGB = a_0 p_1^{a1} p_2^{a2} p_3^{a3},$$
 (2)

Whereas the linear form of equation (2) after logarithmic transformations is shown as equation (3):

$$\ln AGB = \ln a_0 + a_1 \ln p_1 + a_2 \ln p_2 + a_3 \ln p_3,$$
(3)

Where AGB is the field-based aboveground biomass in Mg ha⁻¹, and p_i is the predictor variable selected from the predictor metrics. Such linear regression models were created separately using lidar height derived metrics and Landsat 8 derived predictors.

Stepwise selection was used to select predictor variables to be included in the linear regression models. This method includes the predictor variables showing the highest coefficient of determination (R^2) with the estimator variable, and additional variables were incorporated into the model based on an F-test under the normality assumption of the variables, which was met using logarithmic transformations. To avoid overfitting the models and to identify the most critical predictors for aboveground biomass, a maximum of three predictors can be selected in a single regression model. The variance inflation factor (VIF) was used to identify the existence of collinearity in the selected model. It is commonly accepted that VIF values above 10 indicate multicollinearity which causes standard error inflation (Belsley et al., 2005). We applied a VIF threshold of 10 and the Pearson correlation coefficients to select optimal predictor variables for the aboveground biomass regression that minimized data dimensionality and the presence of multicollinearity within the predictor variables, and to overcome issues of over fitting. Upon model selections, the estimated aboveground biomass was converted back to the original scale, and the root-mean-square error (RMSE, in Mg ha⁻¹) was calculated. This regression modeling process was carried out for three different groups: (1) all forest and woodland plots (n = 39), (2) forest plots only (n = 18), and (3) woodland plots only (n = 21). Upon completion of the regression analysis, a 5 fold cross validation analysis was performed. The differences in the cross validation results and regression model results were evaluated to ensure consistency and accuracy of the regression models. Model performances of aboveground biomass estimates derived from lidar and Landsat 8 were evaluated using R², RMSE, and cross validated RMSE values. All statistical analyses were performed in RStudio (version 0.97.237).

3. Results

3.1. Lidar-Derived Aboveground Biomass Models

We developed regression models to evaluate the statistical relationship between field-based biomass and lidar-derived predictor variables from multiple levels of point densities (Table 2). First, we developed biomass models for all forest and woodland plots, and found 8 points/m² was the most optimal lidar point density level, producing the lowest error of 38 Mg/ha, which is equal to about 36% of average vegetation biomass in our plots (Table 2). Errors associated with the aboveground biomass estimates decreased as the lidar point density increased (p = 0.03, Figure 2). 5-fold cross validated RMSEs were slightly larger than the model RMSEs as we expected, yet they followed the similar trend along the point density gradient (Table 2). Out of all the metrics derived from the airborne lidar dataset, skewness and tree density were the two predictors that were consistently selected for high point densities including 4 and 8 points/m²; whereas standard deviation and kurtosis were the two metrics that were consistently used in the models derived from low density lidar data including 0.5, 1, and 2 points/m² (Table 2).

Second, to account for biomass variation due to broad vegetation types within our study area and to test for differences among vegetation types, we developed separate regression models for forest and woodland plots (Table 2). Among all the models derived from a gradient of point densities, the smallest errors of 32.9 Mg/ha (33%) for the woodland plots and 31.8 Mg/ha (29%) for the forest plots were

produced at 8 points/m² density, respectively (Table 2). Both vegetation type-specific models demonstrated similar descending trends of estimated errors as lidar point density increases (Figure 2). From both model RMSEs and 5-fold cross validated RMSEs, forest plots exhibited a much sharper decline of errors of aboveground biomass estimates with increasing lidar point density than the woodland plots (Table 2 and Figure 2), indicating that the uncertainty associated with estimating forest biomass was more sensitive to source lidar data point density than woodland biomass estimation. Different predictor variables were selected for aboveground biomass models for forest and woodland plots. Forest aboveground biomass models consistently produced smaller errors than the woodland models except for at 1 point/m² (Figure 2). For woodland aboveground biomass models, maximum height and tree density were selected in the highest point density derived models at 8 points/m²; whereas standard deviation was selected in the low point density derived models at 0.5, 1 and 2 points/m² (Table 2). For forest aboveground biomass models, skewness and kurtosis were selected in the higher point density derived models at 4 and 8 points/m²; whereas standard deviation and kurtosis were selected in the lower point density derived models at 0.5, 1, and 2 points/m² (Table 2). For forest plots, kurtosis was proven to be crucial for estimating forest aboveground biomass. Overall, skewness and tree density were the most used metrics to estimate aboveground biomass when using high point density lidar data, while standard deviation and kurtosis were crucial for estimating aboveground biomass at low point density (Figure 3).

3.2. Landsat and Lidar Derived Aboveground Biomass Models Comparison

Using surface reflectance derived from Landsat 8 to estimate aboveground biomass in our study area, red and near infrared bands were selected when all plots were combined (Table 2). The Landsat 8 based model performed at a similar level to those derived from $0.5 - 1 \text{ point/m}^2$ lidar data when all forest and woodland plots were included (Table 2 and Figure 2). For separate forest and woodland models, Landsat 8 performed better in estimating woodland biomass than forest biomass (Table 2 and Figure 2). Green band and NDVI were used in the woodland aboveground biomass models, while coastal and green bands were selected in the forest models (Table 2). Landsat 8 produced larger 5-fold validated RMSEs than all lidar-derived models, including the lowest point densities of $0.5 - 1 \text{ point/m}^2$. Thus, Landsat 8 cannot be interchangeably used with lidar data to estimate forest or woodland aboveground biomass in our study area.

Point Density (points/m ²)	Predictable Variables [*]	R ²	RMSE (Mg/ha)	Percentage of Mean Biomass (%)	5-fold Cross Validated RMSE (Mg/ha)
	All (n = 39)				
8	Skew, Density	0.59	38.1	36.4	38.8
4	Skew, Density, StdDev	0.53	39.6	37.8	41.2
2	StdDev, Kurtosis	0.47	43.0	41.1	46.6
1	StdDev, Kurtosis	0.45	44.7	42.7	48.9
0.5	StdDev, Kurtosis	0.46	42.7	40.8	46.0
Landsat 8	Red, Near Infrared	0.52	43.4	41.5	48.7
	Woodland (n = 21)				
8	Max, Density	0.68	32.9	32.9	37.7
4	Skew, QuarAvg	0.67	34.4	34.4	43.8
2	StdDev	0.61	34.2	34.2	38.9
1	StdDev	0.61	34.9	34.9	40.3
0.5	StdDev	0.56	37.5	37.5	39.5
Landsat 8	Green, NDVI	0.68	37.7	37.7	44.0
	Forest (n = 18)				
8	Skew, Kurtosis	0.52	31.8	28.9	36.1
4	Skew, Kurtosis, StdDev	0.35	33.5	30.4	40.1
2	StdDev, Kurtosis, P50	0.26	34.8	31.6	41.3

Table 2: Aboveground Biomass Models Derived from Multiple Lidar Point Densities and Landsat 8

International Journal of Advanced Remote Sensing and GIS

1	StdDev, Kurtosis	0.10	43.1	39.2	49.6
0.5	StdDev, Kurtosis	0.19	38.9	35.3	47.5
Landsat 8	Coastal, Green	0.15	41.2	37.4	52.7

Skew = Skewness, Density = Tree density, StdDev = Standard Deviation, Max = Maximum height, QuarAvg = Quartic average, P50 = 50th percentile of height



Figure 2: Errors in Aboveground Biomass Estimates at Various Lidar Point Densities, In Comparison to Landsat 8. The Errors were presented as the Percentage of the Root-Mean-Square Error (RMSE) of the Mean Biomass. Aboveground Biomass were Estimated for All Vegetation (Blue), Woodland (Red), And Forest (Green) from Lidar (Circle) and Landsat (Triangle). Linear Regressions between Percentage Error and Point Densities were applied to Lidar Derived Aboveground Biomass Models Only. When Landsat Derived Errors were Larger than Those Derived from All Lidar Data, Landsat Derived Errors Were Presented at 0 Point Density in the Graph for Reference



Figure 3: Usage of Lidar Derived Metrics in the Aboveground Biomass Models. High Point Density Includes 4 and 8 Points/m²; Low Point Density Includes 0.5, 1, and 2 Points/m²

3.3. Cost Considerations for Lidar-Derived Aboveground Biomass Estimates

Since airborne lidar acquisition over large areas is costly, lidar applications necessarily involve cost considerations and tradeoffs. The 3DEP is aiming to provide US national lidar coverage in the next decade, and the current target quality level (QL) is QL2 at 2 points/m², with an option to upgrade to QL1 at 8 points/m² (Sugarbaker et al., 2014). Obviously, the cost of lidar acquisition decreases as point density decreases, and uncertainty of aboveground biomass estimates generally increases as point density decreases, although not necessarily in a linear fashion or uniformly across vegetation types. We used the estimated cost information from 3DEP and the errors of aboveground biomass estimates from our plots to evaluate the cost efficiency of lidar acquisitions to estimate aboveground biomass at various quality levels (Figure 4). The aboveground biomass estimates error at per dollar investment increases linearly as the quality level increases (point density decreases), and QL1 at 8 points/m² provides the best efficiency for the lidar acquisition investment for aboveground biomass applications based on the results from our study area (Figure 4).



Figure 4: Estimate Errors – Acquisition Cost Ratios for Lidar Derived Aboveground Biomass. The Quality Levels (QL) Are: QL1 – 8 Points/m², QL2 – 2 Points/m², and QL3 – 1 Point/m²

4. Discussions

4.1. Lidar Point Density

An important question to ask when producing a large scale product such as a Biomass ECV is whether high point density lidar data are necessary to obtain sufficiently accurate results. From an operational perspective, broad scale (e.g., large regions, statewide, or national) lidar acquisition is usually obtained at low point density at <1 point/m² (Cunningham et al., 2004; Veneziano et al., 2012), which is much lower than the target forest research lidar acquisitions (1-8 points/m²). The next relevant question is whether these low point density lidar data that were intended to create elevation datasets are adequate to produce biomass products with acceptable accuracies.

From our results, the overall accuracies of biomass estimates increased as the source lidar data point density increased, but such trends were not statistically significant for the vegetation type specific models, although these trends may become significant given greater sample size. Some other

research found that high point density did not add much value to extracting forest canopy profiles at the stand level and estimating tree height or volume (Thomas et al., 2006; Goodwin et al., 2006; Lim et al., 2008; Jakubowski et al., 2013), unless the point density was simulated at very low levels (< 0.5 point/m²) where the errors grew significantly (Magnusson et al., 2007; Jakubowski et al., 2013). Results from analysis of our study area indicated that the optimal point density to achieve the best model performance and highest cost efficiency was 8 points/m². These studies are informative for evaluating the effects of point density on biomass estimation accuracy, however our random thinning approach may approximate but not perfectly replicate lower density data as they would be collected using the actual data sampling configurations of a multiple-return lidar sensor. In addition to pulse density, other factors of lidar acquisition such as sensors, flying altitudes, pulse repetition frequencies, and laser pulse power can also impact forest canopy metrics, although such effects are minor (Næsset, 2009; Chasmer et al., 2006).

Because of the rich information of 3-D structures of vegetation from lidar datasets, predictor metrics can be readily extracted from lidar datasets, while additional image analysis, personnel and software costs can occur while using other data sources. Especially for large areas where image processing and analysis workloads increase significantly, lidar can already be less expensive or comparable to image data analysis (Johansen et al., 2010). Another advantage of reduced density lidar data is the reduction of file size for data processing, especially for a large area such as the national Biomass ECV. It has been shown that a 40% reduction in lidar point density with a corresponding significant reduction in file size produced biomass estimates comparable to the full lidar dataset (Singh et al., 2015).

4.2. Lidar Metrics

Evaluation of multiple linear regression models in our analysis illustrated that descriptive statistics metrics such as skewness, kurtosis, and standard deviation explained more variance than direct canopy height related metrics, indicating that the overall summary of the canopy characteristics (kurtosis) and shape of height distribution (skewness) are crucial for aboveground biomass estimates. Only one height percentiles metric was selected in one final model for the forest aboveground biomass estimates (2 points/m²), and maximum height metrics were only used in the woodland aboveground biomass models derived from high point density data. Skewness and tree density were the most commonly used metrics in aboveground biomass models from the high point density lidar data, suggesting that effective estimates of tree density, and therefore accurate biomass estimates, requires higher point density lidar data. In contrast, standard deviation and kurtosis were the most commonly used metrics in the models derived from low point density lidar data, indicating that the variability of height distribution among neighboring points is important to ensure biomass estimate accuracy derived from low density lidar data. For woodland plots, it was challenging to select more than one metric (standard deviation) to predict woodland aboveground biomass using lower density lidar data at 0.5 -2 points/m². Yet, the model performance was not substantially different with one-predictor in the low density models, when compared to two-predictor models from high density lidar data (Figure 2). Thus, low density lidar data can be used to estimate woodland-dominated areas with reasonably sufficient accuracy.

4.3. Aboveground Biomass Model Uncertainty

Quantifying the uncertainty of biomass estimates is crucial for identifying the appropriate data sources and algorithms for a potential national Biomass ECV. Errors can be introduced in multiple steps including tree measurements, allometric equations development, tree-level aboveground biomass prediction, plot-level aboveground biomass estimation, plot-level remote sensing metrics extraction, and remote sensing based model development. Although national scale allometric equations exist (Chojnacky et al., 2014), localized equations are preferred to reduce errors introduced into the fieldbased biomass estimation (Chen et al., 2015). While the lack of field based biometrics and local diameter-based allometric equations limit the operational feasibility of a national Biomass ECV, data routinely collected by the U.S. Forest Service's Forest Inventory and Analysis (FIA) program have been successfully used in lidar-based forest biomass and carbon estimates (Johnson et al., 2014). Because of the comprehensive national sampling strategy of the FIA data, both training and validation sets can be generated from FIA datasets as long as they are independent. For example, national forest inventory data, in combination with spaceborne lidar and optical imagery, have been successfully used in a recent study to estimate biomass of China at the national scale (Su et al., 2016). However, mismatches in the timing of observations can be a source of error, as the years in which FIA plot surveys are performed can vary widely over large areas, and may rarely be coincident with the lidar acquisitions. An additional, potential shortcoming of FIA data for national-scale biomass estimation is that information on grassland or shrubland is not provided, and therefore models have not been developed to estimate grassland or shrubland biomass at a larger scale. Although grassland and shrubland typically have lower aboveground biomass compared to forests and woodlands, large geographic coverage and implications for land management decisions (e.g., grazing and fire management) makes the accurate estimation of their biomass an important issue. Differences between the structure of grasslands and forests/woodlands require different metrics derived from lidar to estimate aboveground biomass, and therefore additional research that extends beyond the scope of this study is needed.

4.4. Operational Feasibility

Although the 3DEP is aimed at accumulating, over time, airborne lidar data that approaches, if not fully achieves, comprehensive national coverage, the current spatial coverage remains far from that goal. In the interim, Landsat 8 surface reflectance data with national coverage and free access could be considered as an alternative candidate to support the creation of a national Biomass ECV. However, in the case of our pilot study area in Arizona, Landsat 8 generally produced higher errors than all lidar-based models, suggesting it cannot simply be used as a substitute of lidar data sources, although additional more sophisticated metrics can be extracted from Landsat data which required further processing and computing. Lidar-based vegetation type specific (forest and woodland) models produced lower errors than all-vegetation combined. Similar to our results, vegetation type specific models improved the aboveground biomass estimates in a mixed Mediterranean forest (García et al., 2010), and land cover maps have been used with other ancillary data to provide regional biomass estimates (Boudreau et al., 2008).

Until national-coverage lidar data are available, a practical approach to a national Biomass ECV product is to apply vegetation type specific models using existing land cover data (Stoker et al., 2014; Avitabile et al., 2012). In our pilot study area, for woodlands, lower point density lidar (e.g., QL3/QL2) were adequate to generate good estimates of woodland aboveground biomass given the small difference across point densities. For forested areas in the pilot plots, high point density lidar (e.g., QL1) for selected dominate forest species areas will be ideal for ensuring the accuracy of aboveground biomass estimates, but lower point density lidar data can also support the production of a Biomass ECV. Further, ponderosa pine forests in our pilot study area may require higher point density for accurate aboveground biomass estimate because of the "tall and thin" shape of conifer trees, and point density requirement may be lower for the broad leaf forests areas elsewhere. In addition, the average aboveground biomass in the semi-arid Southwest US is lower than forests in the Northeast and Northwest US, and therefore a lower point density may be adequate for other higher biomass areas to reach the same overall accuracy target.

Weighing cost and model performance, a high point density operational lidar dataset (e.g., 3DEP QL1) provides the most promising avenue to generating sufficiently accurate Biomass ECV products. Yet, lower point density lidar data still produced better estimates of aboveground biomass than optical data,

and therefore the current QL2 acquisition plan of 3DEP could adequately support the generation of an operational Biomass ECV product, with space for improvement in areas where higher density lidar datasets at 8 points/m² are available. Even though our study area only consists a number of species that are typical of the Southwest US, other research efforts that studied a range of different species (evergreen, deciduous, mixed, plantation, etc.) and stand conditions found similar results that point density decimation effect was minimal, and therefore low point density lidar data can be sufficient to estimate aboveground biomass (Singh et al., 2015; Gonzalez-Ferreiro et al., 2013; Gobakken and Næsset, 2008; Thomas et al., 2006). These findings reinforce our view that it is feasible to use the operational 3DEP QL2 lidar data to develop vegetation type specific aboveground biomass algorithm to create a national scale Biomass ECV given the limitation of current lidar data availability.

5. Conclusion

We compared the effectiveness of lidar data at multiple point densities to estimate aboveground biomass for forest and woodland plots on the San Carlos Apache Reservation in the Southwestern U.S. In general, estimated errors decrease as the point density increases, reaching a turning point at 8 points/m² where the best model performance and cost efficiency was achieved. Forest biomass estimates were more affected by point density levels than woodlands. Therefore, a high density lidar dataset can greatly improve the forest biomass estimates, whereas low density lidar datasets can be sufficient to ensure acceptable accuracy levels for estimating woodland biomass.

Although limited in geographic scope, results from our local scale plot-level research nevertheless provide insights that can be useful in planning effective, potential strategies for developing and implementing a national-scale Biomass ECV product from airborne lidar. Ideally, an operational national lidar dataset at 8 points/m² (QL1) would produce quality aboveground biomass estimates for forest and woodlands as an operational Biomass ECV. However, such high point density lidar data are not yet available at the national scale and the acquisition costs are high. For the near future, we recommend using low point density 3DEP QL2 lidar data, where it is available, to develop vegetation type specific aboveground biomass models, and leveraging higher point density lidar data for forested areas.

This study was conducted in the context of the feasibility of a national Biomass ECV, but it is also relevant to land managers who are potentially making decisions about lidar data acquisition parameters with a set budget. When considering lidar acquisition over large areas, land managers need to make decisions among cost, coverage, density, and accuracy standard. Our plots are located on the San Carlos Apache Reservation, where land managers are particularly interested in using lidar data to estimate vegetation biomass and carbon stocks to inform land management decisions. Our results will enable tribal land managers to make informed decisions on lidar acquisition specifications to maximize the data collection area and ensure reliable results.

Acknowledgements

This study is supported by the USGS Mendenhall, Land Remote Sensing and Land Change Science programs. Thanks to Robert Hetzler, Kelly Hetzler, Charles Truettner, Jacob Higgins, Beverly Maxwell and Terri Victor for providing help in field data collection. Thanks to Birgit Peterson for helpful comments to improve earlier drafts of this manuscript.

References

Avitabile, V., Baccini, A., Friedl, M.A., and Schmullius, C. *Capabilities and Limitations of Landsat and Land Cover Data for Aboveground Woody Biomass Estimation of Uganda.* Remote Sensing of Environment. 2012. 117; 366-380.

Belsley, D.A., Kuh, E., and Welsch, R.E., 2005: *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., and Zemp, M. *The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy.* Bulletin of the American Meteorological Society. 2014. 95; 1431-1443.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., and Kimes, D.S. *Regional Aboveground Forest Biomass Using Airborne and Spaceborne LiDAR in Québec.* Remote Sensing of Environment. 2008. 112; 3876-3890.

Chasmer, L., Hopkinson, C., Smith, B., and Treitz, P. *Examining the Influence of Changing Laser Pulse Repetition Frequencies on Conifer Forest Canopy Returns.* Photogrammetric Engineering and Remote Sensing. 2006. 72; 1359-1367.

Chen, Q., Vaglio Laurin, G., and Valentini, R. *Uncertainty of Remotely Sensed Aboveground Biomass over an African Tropical Forest: Propagating Errors from Trees to Plots to Pixels.* Remote Sensing of Environment. 2015. 160; 134-143.

Chen, Q., Vaglio Laurin, G., Battles, J.J., and Saah, D. *Integration of Airborne Lidar and Vegetation Types Derived from Aerial Photography for Mapping Aboveground Live Biomass.* Remote Sensing of Environment. 2012. 121; 108-117.

Chojnacky, D.C., and Ott, J.S. *Pinyon-Juniper Volume Equations for Arizona Hualapai and Havasupai Indian Reservations.* Research Note INT-363. 1986. 1-4.

Chojnacky, D.C., Heath, L.S., and Jenkins, J.C. *Updated Generalized Biomass Equations for North American Tree Species*. Forestry. 2014. 87; 129-151.

Clary, W.P., and Tiedemann, A.R. *Distribution of Biomass within Small Tree and Shrub form Quercus Gambelii Stands.* Forest Science. 1986. 32; 234-242.

Cunningham, R., Gisclair, D., and Craig, J., 2004: *The Louisiana Statewide Lidar Project*. Louisiana State University, Baton Rouge, Louisiana.

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

Evans, J., Hudak, A., Faux, R., and Smith, A.M. *Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables.* Remote Sensing. 2009. 1; 776-794.

Foody, G.M., Boyd, D.S., and Cutler, M.E.J. *Predictive Relations of Tropical Forest Biomass from Landsat TM Data and Their Transferability between Regions.* Remote Sensing of Environment. 2003. 85; 463-474.

Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J., and Valetini, R. *Relationships between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types*. Ecological Applications. 1995. 28-41.

International Journal of Advanced Remote Sensing and GIS

García, M., Riano, D., Chuvieco, E., and Danson, F.M. *Estimating Biomass Carbon Stocks for a Mediterranean Forest in Central Spain Using Lidar Height and Intensity Data.* Remote Sensing of Environment. 2010. 114; 816-830.

Gobakken, T., and Naesset, E. Assessing Effects of Laser Point Density, Ground Sampling Intensity, and Field Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data. Canadian Journal of Forest Research. 2008. 38; 1095-1109.

Gonzalez-Ferreiro, E., Miranda, D., Barreiro-Fernandez, L., Bujan, S., Garcia-Gutierrez, J., and Dieguez-Aranda, U. *Modelling Stand Biomass Fractions in Galician Eucalyptus Globulus Plantations by use of Different Lidar Pulse Densities*. Forest Systems. 2013. 22; 510-525.

Goodwin, N.R., Coops, N.C., and Culvenor, D.S. Assessment of Forest Structure with Airborne Lidar and the Effects of Platform Altitude. Remote Sensing of Environment. 2006. 103; 140-152.

Gregoire, T.G., Naesset, E., McRoberts, R.E., Stahl, G., Andersen, H., Gobakken, T., Ene, L., and Nelson, R. *Statistical Rigor in Lidar-Assisted Estimation of Aboveground Forest Biomass*. Remote Sensing of Environment. 2016. 173; 98-108.

Grier, C.C., Elliott, K.J., and McCullough, D.G. *Biomass Distribution and Productivity of Pinus Edulis-Juniperus Monosperma Woodlands of North-Central Arizona.* Forest Ecology and Management. 1992. 50; 331-350.

Hall, R.J., Skakun, R.S., Arsenault, E.J., and Case, B.S. *Modeling Forest Stand Structure Attributes Using Landsat ETM+ Data: Application to Mapping of Aboveground Biomass and Stand Volume.* Forest Ecology and Management. 2006. 225; 378-390.

Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., and Stoker, J.M. *Estimating Stand Structure using Discrete-Return Lidar: An Example from Low Density, Fire Prone Ponderosa Pine Forests.* Forest Ecology and Management. 2005. 208; 189-209.

Hug, C., Krzystek, P., and Fuchs, W. Advanced Lidar Data Processing with LAStools. 2004. 12-23.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., and Maltamo, M. *Review of Methods of Small Footprint Airborne Laser Scanning for Extracting Forest Inventory Data in Boreal Forests*. International Journal of Remote Sensing. 2008. 29; 1339-1366.

Isenburg, M., and Schewchuck, J. *LAStools: Converting, Viewing, and Compressing LIDAR Data in LAS Format.* 2007. Available at: http://www.cs.unc.edu/~isenburg/lastools/

Jakubowski, M.K., Guo, Q., and Kelly, M. *Tradeoffs between Lidar Pulse Density and Forest Measurement Accuracy.* Remote Sensing of Environment. 2013. 130; 245-253.

Jensen, J.L.R., Humes, K.S., Conner, T., Williams, C.J., and DeGroot, J. *Estimation of Biophysical Characteristics for Highly Variable Mixed-Conifer Stands using Small-Footprint Lidar*. Canadian Journal of Forest Research. 2006. 36; 1129-1138.

Johansen, K., Phinn, S., and Witte, C. *Mapping of Riparian Zone Attributes using Discrete Return Lidar, Quickbird and SPOT-5 Imagery: Assessing Accuracy and Costs.* Remote Sensing of Environment. 2010. 114; 2679-2691.

Johnson, K., Birdsey, R., Finley, A., Swantaran, A., Dubayah, R., Wayson, C., and Riemann, R. *Integrating Forest Inventory and Analysis Data into a LIDAR-Based Carbon Monitoring System*. Carbon Balance and Management. 2014. 9; 3.

Kane, V.R., McGaughey, R.J., Bakker, J.D., Gersonde, R.F., Lutz, J.A., and Franklin, J.F. *Comparisons between field- and Lidar-Based Measures of Stand Structural Complexity*. Canadian Journal of Forest Research. 2010. 40; 761-773.

Kaye, J.P., Hart, S.C., Fule, P.Z., Covington, W.W., Moore, M.M., and Kaye, M.W. *Initial Carbon, Nitrogen, and Phosphorus Fluxes Following Ponderosa Pine Restoration Treatments.* Ecological Applications. 2005. 15; 1581-1593.

Labrecque, S., Fournier, R.A., Luther, J.E., and Piercey, D. A Comparison of Four Methods to Map Biomass from Landsat-TM and Inventory Data in Western Newfoundland. Forest Ecology and Management. 2006. 226; 129-144.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., and Acker, S.A. *Geographic Variability in Lidar Predictions of Forest Stand Structure in the Pacific Northwest*. Remote Sensing of Environment. 2005. 95; 532-548.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., and Shugart, H.H. Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. Remote Sensing of Environment. 1999. 67; 83-98.

Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., and Gower, S.T. *Lidar Remote Sensing of Above-Ground Biomass in Three Biomes.* Global Ecology and Biogeography. 2002. 11; 393-399.

Lim, K., Hopkinson, C., and Treitz, P. *Examining the Effects of Sampling Point Densities on Laser Canopy Height and Density Metrics.* The Forestry Chronicle. 2008. 84; 876-885.

Lim, K.S., and Treitz, P.M. *Estimation of above Ground Forest Biomass from Airborne Discrete Return Laser Scanner Data Using Canopy-Based Quantile Estimators.* Scandinavian Journal of Forest Research. 2004. 19; 558-570.

Lu, D. Aboveground Biomass Estimation using Landsat TM data in the Brazilian Amazon. International Journal of Remote Sensing. 2005. 26; 2509-2525.

Magnusson, M., Fransson, J.E.S., and Holmgren, J. *Effects on Estimation Accuracy of Forest Variables Using Different Pulse Density of Laser Data*. Forest Science. 2007. 53; 619-626.

Maltamo, M., Eerikainen, K., Packalen, P., and Hyyppa, J. *Estimation of Stem Volume using Laser Scanning-Based Canopy Height Metrics.* Forestry. 2006. 79; 217-229.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E., and McKee, W.A. Use of Large-Footprint Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregon. Remote Sensing of Environment. 1999. 67; 298-308.

Morsdorf, F., Kotz, B., Meier, E., Itten, K.I., and Allgower, B. *Estimation of LAI and Fractional Cover from Small Footprint Airborne Laser Scanning Data Based on Gap Fraction*. Remote Sensing of Environment. 2006. 104; 50-61.

Næsset, E., and Økland, T. *Estimating Tree Height and Tree Crown Properties Using Airborne Scanning Laser in a Boreal Nature Reserve*. Remote Sensing of Environment. 2002. 79; 105-115.

Næsset, E. Effects of Different Sensors, Flying Altitudes, and Pulse Repetition Frequencies on Forest Canopy Metrics and Biophysical Stand Properties Derived from Small-Footprint Airborne Laser Data. Remote Sensing of Environment. 2009. 113; 148-159.

Næsset, E. Determination of Mean Tree Height of Forest Stands using Airborne Laser Scanner Data. ISPRS Journal of Photogrammetry and Remote Sensing. 1997. 52; 49-56.

Næsset, E., and Gobakken, T. *Estimation of Above-And Below-Ground Biomass Across Regions of the Boreal Forest Zone Using Airborne Laser.* Remote Sensing of Environment. 2008. 112; 3079-3090.

Næsset, E., and Bjerknes, K. *Estimating Tree Heights and Number of Stems in Young Forest Stands Using Airborne Laser Scanner Data*. Remote Sensing of Environment. 2001. 78; 328-340.

Næsset, E., Bollandsas, O.M., and Gobakken, T. *Comparing Regression Methods in Estimation of Biophysical Properties of Forest Stands from Two Different Inventories Using Laser Scanner Data.* Remote Sensing of Environment. 2005. 94; 541-553.

Navar, J. Allometric Equations for Tree Species and Carbon Stocks for Forests of Northwestern *Mexico*. Forest Ecology and Management. 2009. 257; 427-434.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W., 1974: *Monitoring Vegetation Systems in the Great Plains with ERTS*. NASA Goddard Space Flight Center 3d ERTS-1 Symposium. 309-317.

Sellers, P.J. *Canopy Reflectance, Photosynthesis and Transpiration*. International Journal of Remote Sensing. 1985. 6; 1335-1372.

Singh, K.K., Chen, G., McCarter, J.B., and Meentemeyer, R.K. *Effects of LiDAR Point Density and Landscape Context on Estimates of Urban Forest Biomass*. ISPRS Journal of Photogrammetry and Remote Sensing. 2015. 101; 310-322.

Stoker, J.M., Cochrane, M.A., and Roy, D.P. Integrating Disparate Lidar Data at the National Scale to Assess the Relationships between Height above Ground, Land Cover and Ecoregions. Photogrammetric Engineering and Remote Sensing. 2014. 80; 59-70.

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., and Fang J. Spatial Distribution of Forest Aboveground Biomass in China: Estimation through Combination of Spaceborne Lidar, Optical Imagery, and Forest Inventory Data. Remote Sensing of Environment. 2016. 173; 187-199.

Sugarbaker, L.J., Constance, E.W., Heidemann, H.K., Jason, A.L., Lukas, V., Saghy, D.L., and Stoker, J.M. *The 3D Elevation Program initiative—A Call For Action*. No. 1399. 2014. 1-35.

Thomas, V., Treitz, P., McCaughey, J.H., and Morrison, I. *Mapping Stand-Level Forest Biophysical Variables for a Mixed Wood Boreal Forest Using Lidar: An Examination of Scanning Density.* Canadian Journal of Forest Research. 2006. 36; 34-47.

Tucker, C.J. *Red and Photographic Infrared Linear Combinations for Monitoring Vegetation.* Remote Sensing of Environment. 1979. 8; 127-150.

Van Leeuwen, M., and Nieuwenhuis, M. *Retrieval of Forest Structural Parameters using LiDAR Remote Sensing*. European Journal of Forest Research. 2010. 129; 749-770.

Veneziano, D., Hallmark, S., and Souleyrette, R., 2012: *Comparison of Lidar and Conventional Mapping Methods for Highway Corridor Studies.* Center for Transportation Research and Education, Iowa State University. Final Report.

Wu, Z., Middleton, B., Hetzler, R., Vogel, J., and Dye, D. *Vegetation Burn Severity Mapping Using Landsat-8 and WorldView-2.* Photogrammetric Engineering and Remote Sensing. 2015. 81; 143-154.

Yu, X., Hyyppä, J., Kaartinen, H., and Maltamo, M. *Automatic Detection of Harvested Trees and Determination of Forest Growth Using Airborne Laser Scanning*. Remote Sensing of Environment. 2004. 90; 451-462.

Zhao, K., Popescu, S., and Nelson, R. *Lidar Remote Sensing of Forest Biomass: A Scale-Invariant Estimation Approach using Airborne Lasers.* Remote Sensing of Environment. 2009. 113; 182-196.

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., and Ryu, S.R. *Estimating Aboveground Biomass Using Landsat 7 ETM+ Data across a Managed Landscape in Northern Wisconsin USA.* Remote Sensing of Environment. 2004. 93; 402-411.