Article

Modeling and Mapping Agroforestry Aboveground Biomass in the Brazilian Amazon Using Airborne Lidar Data

Qi Chen 1,2, Dengsheng Lu 1,3,*, Michael Keller 4,5, Maiza Nara dos-Santos 4, Edson Luis Bolfe 4, Yunyun Feng 1 and Changwei Wang 2

Received: 30 September 2015; Accepted: 20 December 2015; Published: 30 December 2015

Academic Editors: Guomo Zhou, Conghe Song, Guangxing Wang, Nicolas Baghdadi and Prasad S. Thenkabail

1 Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, School of Environmental & Resource Sciences, Zhejiang A&F University, Lin An 311300, China; qichen@hawaii.edu (Q.C.); fengyyfm@163.com (Y.F.)
2 Department of Geography, University of Hawaii at Manoa, Honolulu, HI 96822, USA; changwei_wang@scau.edu.cn
3 Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48823, USA
4 Brazilian Agricultural Research Corporation—Embrapa, Campinas, SP 13070-115, Brazil; mkeller.co2@gmail.com (M.K.); maizanara@gmail.com (M.N.-S.); edson.bolfe@embrapa.br (E.L.B.)
5 USDA Forest Service, International Institute of Tropical Forestry, San Juan, PR 00926, USA

* Correspondence: luds@zafu.edu.cn; Tel./Fax: +86-571-6374-6366

Abstract: Agroforestry has large potential for carbon (C) sequestration while providing many economical, social, and ecological benefits via its diversified products. Airborne lidar is considered as the most accurate technology for mapping aboveground biomass (AGB) over landscape levels. However, little research in the past has been done to study AGB of agroforestry systems using airborne lidar data. Focusing on an agroforestry system in the Brazilian Amazon, this study first predicted plot-level AGB using fixed-effects regression models that assumed the regression coefficients to be constants. The model prediction errors were then analyzed from the perspectives of tree DBH (diameter at breast height)—height relationships and plot-level wood density, which suggested the need for stratifying agroforestry fields to improve plot-level AGB modeling. We separated teak plantations from other agroforestry types and predicted AGB using mixed-effects models that can incorporate the variation of AGB-height relationship across agroforestry types. We found that, at the plot scale, mixed-effects models led to better model prediction performance (based on leave-one-out cross-validation) than the fixed-effects models, with the coefficient of determination ($R^2$) increasing from 0.38 to 0.64. At the landscape level, the difference between AGB densities from the two types of models was ~10% on average and up to ~30% at the pixel level. This study suggested the importance of stratification based on tree AGB allometry and the utility of mixed-effects models in modeling and mapping AGB of agroforestry systems.

Keywords: agroforestry; aboveground biomass; lidar; mixed-effects models; allometry; wood density

1. Introduction

Agroforestry denotes land use systems where woody perennials (trees, shrubs, palms, etc.) are cultivated on the same land units as agricultural crops and/or animals [1,2]. In this study, we use agroforestry as a general term to refer to the land use system of cultivating woody perennials, either polyculture or monoculture (i.e., plantation) on agricultural land, regardless the current existence of crops or animals. Over 10 million km$^2$ of agricultural lands have greater than 10% tree cover [3]. Through the provision of diversified products, agroforestry has been advocated and practiced by
many countries to offer a wide range of economic, social, and ecological benefits: (1) increasing the per capita farm income by planning high-value tree products [4]; (2) improving soil fertility and land productivity [5]; (3) increasing household resilience [4]; (4) mitigating the impacts of climate variability and change [6,7]; (5) conserving biodiversity [8–10]; and (6) improving air and water quality [11–13].

Agroforestry offers high potential for carbon (C) sequestration [14] not only because the carbon density of agroforestry is usually higher than annual crops or pasture [15–17] but also because the trees produce fuelwood and timbers that otherwise would be harvested from natural forests [1]. The fine and coarse wood debris of plants are also stored in the soil C pool for long periods [18]. Thus, the role of agroforestry for C sequestration is increasingly recognized by the IPCC (Intergovernmental Panel on Climate Change) [19]. However, for agroforestry to be successful as a strategy for C sequestration, it is necessary to have sound detection and monitoring systems [20].

Remote sensing provides an effective way to estimate and monitor the biomass and C stock of different vegetation types [21] including agroforestry systems [22–25]. In particular, airborne lidar (Light Detection and Ranging) has emerged in the 21 century as the most accurate technology to quantify forest aboveground biomass (AGB) at the landscape level [21,26]. Lidar is especially powerful over forests of high biomass where passive optical imaging or radar sensors have saturation problems. One of the main advantages of lidar is that it can penetrate through the small canopy gaps for detecting vertical structure and extracting ground elevation. The height information derived from lidar is strongly related to biomass for most tree species. A large number of studies have reported the use of airborne lidar for mapping AGB in boreal (e.g., [27–29]), temperate (e.g., [30,31]), and tropical (e.g., [32–36]) forests.

Compared to the large body of literature of lidar remote sensing of carbon and biomass for forests, the use of this frontier technology for quantifying agroforestry AGB is very limited. One of the best examples of agroforestry of relative economic success of colonization in the Brazilian Amazon occurred in the Tomé-Açu [37]. Focusing on an agroforestry system in Tomé-Açu, this study aims to (1) predict plot-level agroforestry AGB using regular fixed-effects regression models that treat the regression coefficients as constants; (2) identify the causes of AGB prediction errors when such models are used; (3) propose new ways to stratify vegetation types and apply mixed-effects models to reduce the AGB prediction errors; and (4) discuss the challenges and future directions for modeling and mapping agroforestry AGB.

2. Study Area

This study was conducted in the Quatro Bocas district in the municipality of Tomé-Açu (2°28′S and 48°20′W), located in the northeastern region of Pará state in Brazil (Figure 1). The study area is 2 km by 5 km. Topography in the region is characterized by low flat plateaus, terraces, and lowlands with altitudes varying from 14 to 96 m. The soils are classified as Ferralsols, Plintosols, and Fluvisols. The average annual rainfall is 2300 mm [38]. The Tomé-Açu region has a humid mesothermal climate—Ami according to the Köppen classification—with high average annual temperatures (26 °C) and relative air humidity rate of about 85%. The original vegetation is lowland dense ombrophilous forest, which has been intensely altered. The landscape mosaic is dominated by pasture, agricultural fields, and secondary forests. Forest remnants are observed especially at the margins of streams.

Tomé-Açu started its agricultural development in the 1920’s, with the beginning of the Japanese immigration to the region. The immigrants implanted horticulture and, later, black pepper (Piper nigrum L.). They were provided with lands by the Brazilian government, which made technological development possible and turned Pará into the greatest black pepper producer in Brazil. With the decay of the black-pepper cycle from the 1970’s on, caused by fusarium blight, the farmers looked for new production alternatives. According to Homma [39], the way out of this ecological crisis for the immigrants was to diversify their activities, with emphasis in fruit crops, especially papaya, melon, acerola, orange, dende (oil palm), cupuacu, passion fruit, and other native and exotic fruit trees and vines that initiated a new economic cycle for the region. The current agroforestry systems
(1 to 34 years old) have a great variety of fruit and timber tree species. Homma [40] pointed out that the success of the region’s agricultural development resulted from the Japanese-Brazilian farmers’ innovative thinking, their holistic view of future markets, and their social-minded spirit, which made possible the creation of the Cooperativa Agrícola Mista de Tomé-Açu (Camta) in 1931, whose intention was to sell vegetables and nowadays commercializes the agroforestry products (fruit, pulp, juice, and oil) in various countries.

Figure 1. The study area—Tomé-Açu at Pará state in Brazil (Data source: Lidar data were acquired in 2013 and 13 plots were measured in 2014) (Note: colors from red to blue represent the elevation of the laser points).

3. Methods

3.1. Field Data Collection and AGB Estimation

Field inventory followed methods consistent with standardized protocols for tropical forests (e.g., [41]). Thirteen field plots of 30 m × 30 m were established in agroforestry regions using a differential GNSS (Trimble GeoXH-6000) during August and October 2014. The plot locations were selected to cover agroforestry fields with different ages and floristic compositions [22]. Diameter at breast height (DBH) was measured with metric tapes. Tree height was measured with tape and clinometer with the estimated accuracy of 10% of the tree height [42,43].

Over the 13 plots, the field crew collected measurements for a total of 1173 live trees to a minimum DBH of 5 cm from 22 species (Table 1). Four out of 13 plots grew monocultures including teak (plots 9 and 10), dende (plot 11), and cupuacu (plot 15), while the rest of the plots were mixed with two to eight species (Table 2).
Table 1. Tree species of the field plots and their wood densities.

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Scientific Name</th>
<th>Tree Count</th>
<th>$\rho$ (g/cm$^3$)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acai</td>
<td>Euterpe oleracea</td>
<td>287</td>
<td>0.41</td>
<td>Palm</td>
</tr>
<tr>
<td>Ameixa (Tallow plum)</td>
<td>Ximenia americana</td>
<td>108</td>
<td>0.64</td>
<td>Other</td>
</tr>
<tr>
<td>Andiroba</td>
<td>Carapa guianensis</td>
<td>7</td>
<td>0.57</td>
<td>Other</td>
</tr>
<tr>
<td>Cacau (Cocoa)</td>
<td>Theobroma glaucum</td>
<td>408</td>
<td>0.53</td>
<td>Other</td>
</tr>
<tr>
<td>Castanha dopara</td>
<td>Bertholletia excelsa</td>
<td>4</td>
<td>0.64</td>
<td>Other</td>
</tr>
<tr>
<td>Cupuacu</td>
<td>Theobroma grandiflorum</td>
<td>72</td>
<td>0.53</td>
<td>Other</td>
</tr>
<tr>
<td>Dende (American oil palm)</td>
<td>Elaeis oleifera</td>
<td>13</td>
<td>0.41</td>
<td>Palm</td>
</tr>
<tr>
<td>Embauba</td>
<td>Cecropia ficifolia</td>
<td>5</td>
<td>0.27</td>
<td>Other</td>
</tr>
<tr>
<td>Freijo cinza</td>
<td>Cordia goeldiana</td>
<td>1</td>
<td>0.50</td>
<td>Other</td>
</tr>
<tr>
<td>Goiaba (Guava)</td>
<td>Myrciaria floribunda</td>
<td>6</td>
<td>0.77</td>
<td>Other</td>
</tr>
<tr>
<td>Guaruba</td>
<td>Clarisia racemosa</td>
<td>5</td>
<td>0.59</td>
<td>Other</td>
</tr>
<tr>
<td>Ipe</td>
<td>Tabebuia chryotricha</td>
<td>34</td>
<td>0.64</td>
<td>Other</td>
</tr>
<tr>
<td>Ipe rosa</td>
<td>Tabebuia roseo-alba</td>
<td>2</td>
<td>0.52</td>
<td>Other</td>
</tr>
<tr>
<td>Mogno (Mahogany)</td>
<td>Svietenia macrophylla</td>
<td>11</td>
<td>0.51</td>
<td>Other</td>
</tr>
<tr>
<td>Molongo</td>
<td>Ambelania acida</td>
<td>2</td>
<td>0.52</td>
<td>Other</td>
</tr>
<tr>
<td>Murta</td>
<td>Styrchnos subcordata</td>
<td>1</td>
<td>0.54</td>
<td>Other</td>
</tr>
<tr>
<td>Paliteira</td>
<td>Clitoria fairchildana</td>
<td>2</td>
<td>0.64</td>
<td>Other</td>
</tr>
<tr>
<td>Parica</td>
<td>Schizolobium amazzonicum</td>
<td>2</td>
<td>0.49</td>
<td>Other</td>
</tr>
<tr>
<td>Pelo de Cutia</td>
<td>Banana guianensis</td>
<td>24</td>
<td>0.61</td>
<td>Other</td>
</tr>
<tr>
<td>Seringa (Rubber)</td>
<td>Hevea brasiliensis</td>
<td>35</td>
<td>0.49</td>
<td>Other</td>
</tr>
<tr>
<td>Tamanqueira</td>
<td>Zanthoxylum rhoifolium</td>
<td>1</td>
<td>0.49</td>
<td>Other</td>
</tr>
<tr>
<td>Teca (Teak)</td>
<td>Tectona grandis</td>
<td>143</td>
<td>0.64</td>
<td>Other</td>
</tr>
</tbody>
</table>

Note: $\rho$ is species-specific wood density (g/cm$^3$).

Table 2. Tree species of individual field plots and plot attributes including AGB, height (H), basal area (BA), and wood density ($\rho_{plot}$).

<table>
<thead>
<tr>
<th>Plot ID</th>
<th>Species (Tree Count)</th>
<th>Type</th>
<th>AGB (Mg/ha)</th>
<th>H (m)</th>
<th>BA (m$^2$/ha)</th>
<th>$\rho_{plot}$ (g/cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acai (90), Cacau (95)</td>
<td>Polyculture</td>
<td>46.0</td>
<td>6.8</td>
<td>18.9</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>Cacau (59), Ipe (29), Parica (2)</td>
<td>Polyculture</td>
<td>78.0</td>
<td>8.6</td>
<td>17.8</td>
<td>0.58</td>
</tr>
<tr>
<td>5</td>
<td>Cacau (50), Seringa (35), other (2)</td>
<td>Polyculture</td>
<td>107.7</td>
<td>9.6</td>
<td>21.1</td>
<td>0.49</td>
</tr>
<tr>
<td>7</td>
<td>Cacau (51), Andiroba (7), Ipe Rosa (2), Molongo (2), Paliteira (2), Cupuacu (1)</td>
<td>Polyculture</td>
<td>159.4</td>
<td>7.4</td>
<td>23.3</td>
<td>0.56</td>
</tr>
<tr>
<td>8</td>
<td>Acai (69), Cacau (63), Cupuacu (25)</td>
<td>Polyculture</td>
<td>124.3</td>
<td>8.0</td>
<td>25.6</td>
<td>0.57</td>
</tr>
<tr>
<td>9</td>
<td>Teca (30)</td>
<td>Monoculture</td>
<td>178.4</td>
<td>18.9</td>
<td>21.1</td>
<td>0.64</td>
</tr>
<tr>
<td>10</td>
<td>Teca (113)</td>
<td>Monoculture</td>
<td>255.5</td>
<td>17.7</td>
<td>31.1</td>
<td>0.64</td>
</tr>
<tr>
<td>11</td>
<td>Dende (13)</td>
<td>Monoculture</td>
<td>219.8</td>
<td>11.2</td>
<td>103.3</td>
<td>0.41</td>
</tr>
<tr>
<td>12</td>
<td>Cacau (45), IPE (3)</td>
<td>Polyculture</td>
<td>13.1</td>
<td>3.9</td>
<td>6.7</td>
<td>0.52</td>
</tr>
<tr>
<td>13</td>
<td>Acai (70), Pelo de Cutia (23), Ameixa (14), Guaruba (5), Goiaba (2), other (2)</td>
<td>Polyculture</td>
<td>41.7</td>
<td>7.5</td>
<td>11.1</td>
<td>0.56</td>
</tr>
<tr>
<td>14</td>
<td>Ameixa (94), Acai (54), Embauba (4), Goiaba (4), other(1)</td>
<td>Polyculture</td>
<td>105.1</td>
<td>9.2</td>
<td>22.2</td>
<td>0.61</td>
</tr>
<tr>
<td>15</td>
<td>Cupuacu (46)</td>
<td>Monoculture</td>
<td>28.1</td>
<td>4.5</td>
<td>15.6</td>
<td>0.53</td>
</tr>
<tr>
<td>16</td>
<td>Cacau (45), Mogno (5), Ipe (2)</td>
<td>Polyculture</td>
<td>10.6</td>
<td>3.5</td>
<td>5.6</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Note: Plot-level wood density ($\rho_{plot}$) was calculated by weighting individual tree wood density with their size (basal area * tree height). H is the mean of individual tree heights.

Based on wood density, DBH, and tree height, the oven-dry aboveground biomass ($\hat{B}_{tree}$) of individual trees was estimated using the following allometric models:
For palm trees (Acai and American oil palm),

$$\hat{B}_{\text{tree}} \text{ (kg)} = 0.001 \cdot \left( \exp\left(0.9285 \cdot \ln\left(DBH^2\right) + 5.7236\right) \cdot 1.05001 \right)$$

(1)

For other trees:

$$\hat{B}_{\text{tree}} \text{ (kg)} = 0.0704 \cdot \left(\rho \cdot DBH^2 \cdot H\right)^{0.9701}$$

(2)

where DBH is in cm, $\rho$ is wood density in g/cm$^3$, H is total plant height in meter. Model 1 was developed from stemmed palms in Amazon [44]. Model 2 was a model derived from the pan-tropical destructive tree AGB database [45]. The model was fitted using a generalized nonlinear least squares method [31] instead of log-transformation-based approaches [45]. Species-specific wood density was based on the Global Wood Density Database [46,47] for all species except American oil palm, for which we found more recent estimate from Fathi [48]. The plot-level AGB density $\hat{B}_{\text{plot}}$ was calculated by summing the individual tree AGB at each plot and then divided by the plot area (900 m$^2$). The AGB density of the plots is within the range of 10.6 to 255.5 Mg/ha (Table 2).

3.2. Airborne Lidar Data Acquisition and Processing

Airborne lidar was collected by GEOID Laser Mapping on 2 September 2013 using an Optech ALTM Orion M200 sensor with an integrated GPS, IMU system at an average height of 853 m above ground and a scan angle of 11°. Global positioning system data were collected at a ground station simultaneously with the flight at a ground survey location to permit post-processing for estimated horizontal and vertical accuracies (1 $\sigma$) of $\pm 0.3$ m and 0.15 m, respectively. The average point density is 24 pt/m$^2$.

The airborne lidar data were processed using the Toolbox for Lidar Data Filtering and Forest Studies (Tiffs) [49]. The lidar point cloud was classified into ground returns and non-ground returns by the data provider. These ground returns were interpolated into a 1-m Digital Terrain Model (DTM). The relative height of every laser point was calculated as the difference between its Z coordinate and the underlying DTM elevation. Based on the relative height of every laser point, statistics such as mean, standard deviation, skewness, kurtosis, quadratic mean height, percentile height (10th, 20th, . . . , 90th), and the proportion of points within different bins (0–5 m, 5–10 m, . . . , 45–50 m) for each plot were extracted. Note that we calculated the lidar metrics based on all returns, including those from ground, to incorporate both horizontal and vertical canopy structure information [20,21]. To map AGB over the landscape, we also generated lidar metric rasters at 30 m resolution, the same dimension as the field plots.

3.3. Lidar-Based AGB Modeling and Mapping

We developed two types of models for biomass estimation. First, we used a fixed-effects model to estimate biomass for all plots. This was developed with multiplicative power models [36,50] at the plot level for biomass estimation:

$$\hat{B}_{\text{plot}} = f_{\text{plot}}(\varphi, z) = \varphi_0 z_1^{\varphi_1} z_2^{\varphi_2} \ldots z_m^{\varphi_m}$$

(3)

where $z_1$, $z_2$, and $z_m$ are different lidar metrics, $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_m = \pi r^2$ are model parameters. To select the most relevant lidar metrics for biomass prediction, we used the same approach as [36] with feature selection using log-scale forward stepwise regression and then iteratively fitted nonlinear multiplicative power models by keeping the statistically significant variables. We then examined the residual errors of the above model from the perspectives of both tree DBH-H relationship and wood
density at the plot scale. We calculated the plot-level average wood density by weighting each tree’s wood density with a proxy of tree stem volume:

\[
\rho_{\text{plot}} = \frac{\sum_i (DBH_i^2 \times H_i) \rho_i}{\sum_i (DBH_i^2 \times H_i)}
\]  

(4)

where \(i\) is the subscript for a tree in a plot and \(DBH_i^2 \times H_i\) is the proxy of tree stem volume.

To test whether AGB prediction can be improved by stratification, we divided the plots into \(n\) different vegetation groups and estimated AGB using nonlinear mixed-effects models. The mixed-effects models assume that the parameters of lidar-based AGB models vary by vegetation groups. Each model parameter is considered as a combination of fixed and random effects:

\[
\phi_k = \beta + b_k
\]  

(5)

where \(\phi_k\) is a parameter for group \(k\) \((k \in [1,n])\), \(\beta\) is the fixed effect or population-level constant, \(b_k\) is the random effect that varies across different vegetation groups. The random effects can be modeled as Gaussian variables with mean of zeros and covariance matrix \(\psi\).

We used the nonlinear power model as the starting point for developing the mixed-effects models. In other words, we assumed the mixed-effects models have the same predictor(s) and model form as the fixed-effects models, but their parameters were modeled as random variables. We began our fitting of the mixed-effects models by assuming the random effects covariance matrix \(\psi\) to be fully unconstructed. If the estimate of any random effects was statistically insignificant from zero across all different vegetation types, the random effect was dropped from the model.

The RMSE (Root Mean Squared Error) and AIC (Akaike Information Criterion) were calculated to compare the performance of different models [51]. Pesudo-\(R^2\) was denoted and calculated as:

\[
R^2 = 1 - \frac{\sum_i (B_{\text{plot},i} - \hat{B}_{\text{plot},i})^2}{\sum_i (B_{\text{plot},i} - \bar{B}_{\text{plot}})^2}
\]  

(6)

where \(i\) is the subscript of individual field plot, \(B_{\text{plot},i}\) is the plot-level AGB estimated from field data, \(\bar{B}_{\text{plot}}\) is the mean of all plots’ AGB field estimates, \(\hat{B}_{\text{plot},i}\) is the predicted plot AGB based on fixed-effects or mixed-effects models. We calculated \(R^2\) and RMSE using both re-substitution and cross-validation methods. We mapped the AGB density by applying the plot-level models using the 30-m airborne lidar metrics at the landscape level.

3.4. Mapping Agroforest Distribution with Visual Interpretation

The results of the mixed-effects models suggested the need to classify the study area into forests, American oil palm, teak, and other agroforests for predicting AGB over the landscape. We classified the landscape via visual interpretation of airborne lidar data in 2D and 3D GIS environment with the aid of Google Earth imagery acquired in 2010. The visual interpretation was mainly based on information such as tree shape, plantation pattern, and color derived from airborne lidar point cloud and high spatial resolution Google Earth imagery. A Canopy Height Model (CHM) of 1-m spatial resolution was generated and used as a backdrop image for delineating the boundary of vegetation patches and roads. Visual interpretation instead of automatic classification was chosen to ensure the highest classification accuracy (without the salt-and-pepper noise in pixel-based classification or boundary inaccuracy in object-based classification), which can help us to conduct a focused analysis of the importance of stratification on landscape-level AGB mapping.

Some remnant forests also exist in the study area. Because the focus of this study is agroforestry, the areas of forests were identified and masked out via visual interpretation. Often, the differences between agroforests and forests were sharp in the lidar image (see Figure 2a,b). However, for some fields that have large trees in the overstory and were only marginally managed with few trees planted
underneath, the distinction between agroforests and forests was not so obvious, both conceptually and visually (see Figure 2c,d). Our strategy was to classify a field into agroforestry only if we saw clear evidence of planting activities (e.g., rows of trees). This might lead to large omission errors but will have minimal commission errors in the identification of agroforests. Such a strategy can be justified because our focus in this study is the AGB of agroforestry. Within the agroforestry areas, the American oil palm and teak can be easily identified with visual interpretation due to their unique characteristics (Figure 3) and the small size of our study area.

\[ \text{AGB} = 13.94 \times h_{0.90}^{0.70} \] – 7

The model $R^2$ is 0.47 and the RMSE is 60.4 Mg/ha (see Figure 4a). Note that the mean AGB of all plots is 105.2 Mg/ha, which means that the coefficient of variation (CV) or relative prediction error is

![Figure 2. Examples (a–d) of agroforestry (left) and forests (right) fields shown in airborne lidar. Color (from red to blue) represents the elevation of the laser points.](image)

![Figure 3. Teak plantations shown in the field (a); aerial photo (b); and lidar data (c); American oil palm sample tree (d); and plantation shown in aerial photos (e); and lidar data (f).](image)
4. Results

With the multiplicative power model and our feature selection procedure, the final plot-level AGB prediction model is a simple power model based on the 90th percentile height:

\[
\hat{B}_{\text{plot}} = 13.94 \times H_{90th}^{0.86}
\]  

(7)

The model \( R^2 \) is 0.47 and the RMSE is 60.4 Mg/ha (see Figure 4a). Note that the mean AGB of all plots is 105.2 Mg/ha, which means that the coefficient of variation (CV) or relative prediction error is almost 60%. In particular, we found that plots 10 and 11 have relatively large residuals. Plot 11 is in an American oil palm plantation. The large residual of plot 11 is mainly caused by the unique DBH-H relationship of American oil palm trees, which implies that the AGB of American oil palm needs to be modeled separately from other vegetation types (see Section 5.1). However, we had only one plot of American oil palm, which eliminates the possibility of independently assessing the model prediction errors. Thus, plot 11 was excluded in our analysis, and the corresponding AGB model (see Figure 4b) developed after feature selection was:

\[
\hat{B}_{\text{plot}} = 8.28 \times H_{80th}^{1.06}
\]  

(8)

where \( H_{80th} \) is the 80th percentile height of lidar points.

Figure 4. Models for predicting plot-level AGB with the American oil palm plot included (a) and excluded (b); and the relationship between residual and wood density at the plot level (c).
Note that when the allometric model for non-palm trees (Equation (2)) was used to estimate AGB, wood density is an important variable. As shown in Table 2, the teak plantation plots (plots 9 and 10) have the highest plot-level wood density. Plot 14 also has high average wood density because some of the species within it, such as ameixa and goiaba have high wood density (see Tables 1 and 2). The large plot-level wood density of these three plots can explain why they all have positive residuals in AGB prediction models (Figure 4b).

The patterns revealed in the residual errors of plot-level AGB model prediction (i.e., plots with large wood density led to large positive residuals) (Figure 4c) suggest that the residuals are not statistically independent, i.e., an obvious violation of the basic assumption of the ordinary least squares (OLS) regression models. A statistical approach that can naturally consider the correlation within individual observations is mixed-effects models. The idea of mixed-effects models is to stratify the observations (i.e., plots in this study) into different groups and allow the model parameters vary as realizations of random variables across the groups. By doing so, the residuals of model prediction within individual groups will become more statistically independent. Mixed-effects models provide a trade-off between fitting all data points with one model and fitting models for each group independently. Hence it is well suited for handling data with few observations within groups [50].

The use of mixed-effects models requires the grouping of individual plots. The relatively large wood density of plots 9, 10, 14, and their positive residuals in the fixed-effects model imply that these three plots can be put into one group. The teak plantations (plots 9 and 10) not only have high wood density but also have the potential of being mapped over large area from remotely sensed data. Thus, we also tried to put only plots 9 and 10 into one group. As a result, we consider two different grouping scenarios as shown in Table 3. As shown in Equation (8), the developed fixed-effects model was a simple power model (in the form of $y = ax^b$, where $a$ and $b$ are model parameters) with the 80th percentile height of lidar points as the predictor. In all of our mixed-effects modeling experiments, we found that only the parameter $b$ had statistically significant random effects. Therefore, in the mixed-effects models, only the parameter $b$ varies with vegetation groups. For grouping scheme A, the model parameters $a = 8.63$, $b_1 = 1.15$, and $b_2 = 0.98$, where $b_1$ and $b_2$ are the parameters for groups 1 and 2, respectively. For grouping scheme B, the model parameters $a = 7.23$, $b_1 = 1.22$, and $b_2 = 1.03$. Note that here the mixed-effects model for a given grouping scheme is equivalent to two simple power models for the two vegetation groups within the scheme (Figure 5).

Table 3. Different schemes of grouping plots for developing mixed-effects models.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Group#1</th>
<th>Group#2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Teak plantation (plot 9, 10)</td>
<td>Non-teak (other plots)</td>
</tr>
<tr>
<td>B</td>
<td>High wood density plots (plot 9, 10, 14)</td>
<td>Other (other plots)</td>
</tr>
</tbody>
</table>

When we used the same plots for calibration and validation (i.e., re-substitution), the $R^2$ was much higher and RMSE was much smaller when we used mixed-effects models in comparison to fixed-effects models (Table 4). The increase of goodness-of-fit in these statistics was expected because mixed-effects models are more complex (i.e., more parameters) than fixed-effects models. AIC is a statistic that penalizes the increase of model fitness due to the use of more complex models [51]. When we fitted mixed-effects models based on two groups, the AIC values decreased slightly for scheme A and by about four for scheme B (note that lower AIC values are preferred for modeling). Figure 5 shows the lidar-based AGB models developed from the field plots as well as their AGB prediction in comparison to the field-based estimates.

We also used leave-one-out cross-validation for calculating $R^2$ and RMSE to further assess the models’ prediction performance (Table 4). We found that the fixed-effects model had $R^2$ as low as 0.38, which was increased to 0.64 and 0.75 when mixed-effects models with scheme A and scheme B were used. This confirms the importance of stratification and mixed-effects modeling for predicting AGB.
Table 4. Comparison of fixed-effects and mixed-effects models for plot-level AGB estimation.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Re-Substitution</th>
<th>Cross-Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE (Mg/ha)</td>
</tr>
<tr>
<td>Fixed-effects model</td>
<td>0.74</td>
<td>40.4</td>
</tr>
<tr>
<td>Mixed-effects model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheme A</td>
<td>0.91</td>
<td>25.9</td>
</tr>
<tr>
<td>Scheme B</td>
<td>0.94</td>
<td>21.6</td>
</tr>
</tbody>
</table>

Figure 5. Comparison of fixed-effects and mixed-effects models for estimating plot AGB. The solid lines in the first-row figures are model curves. The dashed lines in the second-row figures are 1:1 line.

Figure 6 shows vegetation distribution 6a and the estimated AGB 6b and 6c over the whole agroforestry area using mixed- and fixed-effects models, respectively. The mean AGB densities estimated from fixed- and mixed-effects models were 57.0 Mg/ha and 51.7 Mg/ha, respectively. So, the average AGB density estimate from fixed-effects models was about 10% higher than the one from mixed-effects model. Moreover, by checking the spatial pattern of their difference at the pixel level (Figure 6d), we can see that the teak plantation AGB was underestimated by up to ~70 Mg/ha (or 33%) while the other agroforestry AGB were overestimated by up to ~60 Mg/ha (or 25%) using fixed-effects models in comparison to mixed-effects models.
5. Discussion

5.1. Tree DBH-H Relationship

Note that, according to the allometric models used in this study (i.e., Equations (1) and (2)), DBH is the only predictor for estimating palm trees’ AGB and the major factor for predicting other trees’ AGB. However, lidar is more suited for estimating vegetation height than for DBH because the latter usually has to be indirectly estimated from airborne lidar data via its statistical relationship with lidar height metrics. When such allometric models are used, the capability of predicting AGB from lidar is governed by the strength of tree-level DBH-H relationship.

Figure 7 shows the relationship between DBH and tree height for all species based on the field measurements. It is clear that the tree-level data points can be separated into two groups. The group at the right includes 13 American oil palm trees, all from plot 11. Although the other tree species in this study exhibits positive relationship between DBH and height, the American oil palms do not show such a pattern. This might be due to the fact that the plants grow by increasing its diameter rather than height after reaching a certain size. As a result, the American oil palms have unique DBH-H relationship compared to most other tree species in this study. Note that the plot-level AGB model (i.e., Equation (7)) was calibrated to fit AGB from all plots while American oil palms exists in only plot 11. Hence, it is not surprising that the model (see Figure 4a) has large prediction error for the American oil palm plot. The unique DBH-H relationship of American oil palm suggested that it should be treated separately in our modeling. However, we were concerned that the model was not reliable with just one such plot. As a result, we decided to exclude the American oil palm plot in our further analysis.

Figure 6. Maps of vegetation type (a); and AGB predicted with mixed-effects model (b); and fixed-effects model (c); and the difference between AGB predicted with fixed- and mixed-effects models (d). Black color indicates the area masked for analysis.
separately in our modeling. However, we were concerned that the model was not reliable with just one such plot. As a result, we decided to exclude the American oil palm plot in our further analysis.

![DBH-H relationship](image)

**Figure 7.** DBH and tree height (H) relationships for American oil palm and other tree species.

### 5.2. Allometry

One of the biggest uncertainties in our analysis is the tree-level AGB allometric models, especially for palms. For example, Goodman *et al.* [52] developed an allometric model for *Euterpe precatoria*, which is the allometric model that we could find and is the closest to our acai palm species (*Euterpe oleracea*). However, when the model \( \hat{B}_{\text{tree}} (\text{kg}) = 13.589 \times H_s - 108.81 \) was applied to our acai palms, 235 out of 287 trees had negative AGB estimates. That allometric model was developed with 8 trees with a limited DBH range of 12–19 cm. Nevertheless, 242 Acai trees in this study have DBH < 12 cm. The small number and the limited range of the size for trees that were used to calibrate the model [52] can explain the failure of applying such a model into our study area. Unfortunately, no other allometric models specifically for acai palms were found in the literature. We do not have quantitative information about the errors of applying general allometric models (i.e., Equation (1)) to predict the AGB of the palm species in our study. More than 2000 species of palm trees occur in the world. Palm is the most abundant arborescent plant family in western Amazonian forest [52]. Thus, improving our knowledge about the AGB allometry for palms trees, in addition to dicotyledonous trees, is critical for accurate estimation of the AGB and carbon stock in Amazon.

Tree AGB allometric models are key to quantify the errors of remotely sensed AGB. It has been increasingly recognized that it is crucial to characterize the uncertainty of AGB estimates at the pixel and larger spatial scales [36,53–55]. The RMSE based on leave-one-out cross validation only characterized the errors related to the lidar-based AGB models, without considering errors related to tree AGB allometry, model parameters, field measurements, and lidar data. Chen *et al.* [36] did a comprehensive study of different error sources of AGB prediction when airborne lidar data and field data were combined to model and map AGB in an African tropical forest. It was found that errors related to tree AGB allometry and lidar-based regression models were the two major error sources in pixel-level AGB estimates. One of the cornerstones of that study is the availability of the destructively measured tree AGB for characterizing the error of tree AGB allometric models. Although we can apply that methodology to characterize AGB prediction errors for dicotyledonous trees in this study, palms need separate allometric models [52]. Unfortunately, the lack of access to the destructive AGB measurements for palm trees prevents us from fully assessing the errors of agroforestry AGB in this study.
5.3. Mixed-Effects Modeling

A limitation of this study is the small sample size \((n = 12)\) for plot-level AGB modeling. We used mixed-effects models to address the dependence of lidar-AGB relationship on vegetation type and that left us with small numbers of plots per vegetation type. We should use the developed models for AGB prediction for other areas with caution, especially when there are out of the range of the input variables. In the future, more field plots ought to be collected, especially for American oil palm and teak plantations, to develop AGB models for larger areas. The one-year time difference between the times of lidar and field data acquisitions could also contribute to the errors in AGB prediction.

The use of the mixed-effects models for spatially-explicit AGB mapping requires the classification of the agroforestry fields over the whole study area based on their corresponding grouping schemes. Although we found that the grouping scheme B had the best prediction accuracy (Table 4), this model is more challenging to use in our study area because it requires the identification of high wood density areas. The calculation of the average wood density over an area depends on mapping of individual tree species, which is difficult for polyculture agroforestry fields that have multilayer canopy structure. Instead, we chose grouping scheme A for vegetation type classification because teak plantations can be more easily recognized from remotely sensed data owing to the relatively uniform and simple canopy structure (Figure 3a–c). Preliminary results from a supervised classification using 5-m RapidEye satellite imagery (unpublished data) indicated that teak plantations could be classified with a user accuracy and producer accuracy of 97.7% and 96.4%, respectively. Therefore, our methodology, when combined with digital classification of remotely sensed data, has potential to be applied and extended over large areas. One future research direction is to apply object-based image analysis (OBIA) approach and/or hyperspectral imagery to improve the classification of agroforestry fields [56–58].

We excluded forests to stay focused in the study of agroforestry. As shown earlier, it was difficult to separate agroforestry field and forests when an agroforestry field is dense and mixed with many tree species (see Figure 2c,d). A remaining question is whether it is necessary to separate agroforestry and forests in such cases. In other words, it is unclear whether AGB for forests and such agroforestry fields can be fit using just one model. One cold answer this question with comparable field and lidar data over forest and agroforestry plots.

6. Conclusions

Agroforestry has great potential for carbon sequestration while providing multiple benefits to individual farmers and communities. This study examined the potentials and challenges of using airborne lidar data for estimating AGB in an agroforestry system in the Brazilian Amazon. Several significant lessons were learned.

(1) We found strong evidence to support the stratification of vegetation types in agroforestry fields for AGB modeling and mapping. This is in contrast to the widespread use of statistical models with no awareness of different vegetation types in most studies of tropical forest AGB using airborne lidar. Different from forests where trees are selected or adapted via natural processes, the trees in agroforestry are selected and planted by people to maximize economic and other benefits on purpose. Thus, the trees in agroforestry systems usually show more spatially-regularized patterns with a few co-occurring species. The species and structural diversities within agroforestry fields are usually smaller than the ones within forests while the diversities across different agroforestry fields may be larger. Therefore, the need for stratifying vegetation types in agroforestry is stronger than in forests for AGB studies.

(2) This study analyzed the residual errors resulted from regular fixed-effects models and found that the errors have a pattern related to the variations in plot-level wood density. Based on this pattern, agroforestry fields were classified for lidar-based AGB modeling. This is an improvement over previous studies (e.g., [50]) that used existing vegetation classification schemes not specifically developed for lidar-based AGB modeling and mapping.
This study reinforced the utility of mixed-effects models for biomass modeling and mapping. Mixed-effects models can naturally incorporate how different species or groups have different wood densities and thus distinct lidar height—tree AGB relationships. Mixed-effects models also can elegantly cope with the issue of small sample size via adjustment of model parameters as a combination of fixed and random effects. With the classification of agroforestry into teak plantations and other types, we found the mixed-effects models improved the $R^2$ of AGB prediction from 0.38 to 0.64 and reduced the RMSE from 56.4 Mg/ha to 42.9 Mg/ha in comparison to fixed-effects models. We expect this study will encourage the further use of this under-investigated tool in the community of remote sensing of biomass and carbon. This study highlighted the lack of reliable AGB allometry models, especially for palms, to comprehensively quantify the errors of lidar-based AGB estimates. This study also suggested the need to collect more field data over larger areas for automated mapping of plantation types and improved AGB modeling and mapping.

Acknowledgments: This study was financially supported by the China’s National Natural Science Foundation (No# 41571411) and the Zhejiang A&F University’s Research and Development Fund for the talent startup project (No# 2013FR052). Keller, dos-Santos, and Bolfe acknowledge the support from CNPq LBA. Data were acquired by the Sustainable Landscapes Brazil project supported by the Brazilian Agricultural Research Corporation (EMBRAPA), the US Forest Service, and USAID, and the US Department of State.

Author Contributions: Qi Chen developed the analytical framework, did the analysis, and wrote the manuscript. Dengsheng Lu contributed to the analytical framework and data analysis. Michael Keller contributed to the analytical framework and together with Edson Bolfe and Maiza Nara dos-Santos designed the field data collection plan. Edson Bolfe and Maiza Nara dos-Santos organized the field data. Yunyun Feng and Changwei Wang mapped vegetation types. All authors contributed to the editing of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References


43. Hunter, M.O.; Keller, M.; Victoria, D.; Morton, D.C. Tree height and tropical forest biomass estimation. *Biogeosciences* 2013, 10, 8385–8399. [CrossRef]


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).