



## Article

# Using Airborne Lidar, Multispectral Imagery, and Field Inventory Data to Estimate Basal Area, Volume, and Aboveground Biomass in Heterogeneous Mixed Species Forests: A Case Study in Southern Alabama

Schyler Brown <sup>1,\*</sup>, Lana L. Narine <sup>1</sup> and John Gilbert <sup>2</sup>

<sup>1</sup> College of Forestry and Wildlife Sciences, Auburn University, 602 Duncan Dr., Auburn, AL 36849, USA; lln0005@auburn.edu

<sup>2</sup> Solon Dixon Forestry Education Center, Auburn University, 12130 Dixon Road Center, Andalusia, AL 36420, USA; gilbejo@auburn.edu

\* Correspondence: sbb0056@auburn.edu



**Citation:** Brown, S.; Narine, L.L.; Gilbert, J. Using Airborne Lidar, Multispectral Imagery, and Field Inventory Data to Estimate Basal Area, Volume, and Aboveground Biomass in Heterogeneous Mixed Species Forests: A Case Study in Southern Alabama. *Remote Sens.* **2022**, *14*, 2708. <https://doi.org/10.3390/rs14112708>

Academic Editors: L. Monika Moskal, Francesco Pirotti, H. Jaime Hernández Palma, Gaia Vaglio Laurin and Erico Kutchartt

Received: 24 April 2022

Accepted: 2 June 2022

Published: 4 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Airborne light detection and ranging (lidar) has proven to be a useful data source for estimating forest inventory metrics such as basal area (BA), volume, and aboveground biomass (AGB) and for producing wall-to-wall maps for validation of satellite-derived estimates of forest measures. However, some studies have shown that in mixed forests, estimates of forest inventory derived from lidar can be less accurate due to the high variability of growth patterns in multispecies forests. The goal of this study is to produce more accurate wall-to-wall reference maps in mixed forest stands by introducing variables from multispectral imagery into lidar models. Both parametric (multiple linear regression) and non-parametric (Random Forests) modeling techniques were used to estimate BA, volume, and AGB in mixed-species forests in Southern Alabama. Models from Random Forests and linear regression were competitive with one another; neither approach produced substantially better models. Of the best models produced from linear regression, all included a variable for multispectral imagery, though models with only lidar variables were nearly as sufficient for estimating BA, volume, and AGB. In Random Forests modeling, the most important variables were those derived from lidar. The following accuracy was achieved for linear regression model estimates: BA  $R^2 = 0.36$ , %RMSE = 31.26, volume  $R^2 = 0.45$ , %RMSE = 35.30, and AGB  $R^2 = 0.41$ , %RMSE = 31.31. The results of this study show that the addition of multispectral imagery is not substantially beneficial for improving estimates of BA, volume, and AGB in mixed forests and suggests that the investigation of other variables to explain forest variability is necessary.

**Keywords:** laser; forestry; allometry; parametric; non-parametric; modeling 3DEP; Alabama

## 1. Introduction

Estimates of forest inventory such as basal area (BA), volume, and aboveground biomass (AGB) across large tracts of land provide foresters and ecologists the information necessary to form and implement management strategies at both small and large scales [1,2]. For example, BA (a measure of the cross-sectional area of a tree at breast height) has traditionally been used to manage naturally regenerated forest stands for timber production. BA is also linked to other forest metrics and can be used to estimate volume and biomass [3–6]. BA can also be a useful forest measurement for ecological studies, and a perfect example to demonstrate this point is found in longleaf pine (*Pinus palustris*) forests in the U.S. Southeast. In the U.S. Southeastern, it is not only necessary to know the amount of wooded area of the longleaf pine for conservation of the species, but Red-Cockaded woodpeckers (*Leuconotopicus borealis*) depend on a balance of longleaf pine basal area and stand density that are suitable for cavity nesting [7]. BA is also useful for examining the

amount of woody area infested by various species of invasive insects across large areas [8]. Volume is a measure used by foresters to estimate the cubic amount of wood on an area of land [9,10]. An estimate of volume can give foresters an estimate of the dollar value of the standing timber they are cruising. Volume is also associated with the persistence of invasive species, and having estimates of volume can help foresters better manage both timber product and ecosystem health through stem and growth stocking projections [11]. AGB is the dry weight of carbon stored within forest trees above the ground and is measured in  $\text{Mg ha}^{-1}$  [12]. In ecological studies, AGB is a known driver of the species composition of an ecosystem, and maps of AGB can give ecologists insights into the distribution and composition of organisms across great extents [13]. Because AGB is an estimation of the carbon stored in the trunks of trees, it is critical to understand the role forests play in carbon cycling and climate change. Accurate estimates of AGB can help nations develop strategies to meet goals set by international agreements for climate change, such as those outlined in the Reducing Emissions from Deforestation section of the Paris Climate Agreement [14], and contributing action inventories such as the National Greenhouse Gas Inventories report [15]. Often it can be difficult or impossible to accurately estimate AGB without destructively sampling trees in a forest, and allometric equations developed from field inventory estimating BA, volume, and AGB can fail to account for variables that affect them across a large landscape, such as site index and crowding, among other local factors [2,16]. Inconsistencies in forest inventory methods across the U.S. also cause inaccuracies in regional estimates, as inventory methods are frequently tailored for project-specific goals and thus have differing methods for measuring [17–19]. Lastly, estimates of forest measurements across large areas using traditional mensuration approaches can be time-consuming and costly, and in some instances, variation in landscape features can cause dangers in field work and inaccuracies in estimates [16,18,20–22]. Remote sensing (RS), particularly light detection and ranging (lidar) estimates of BA, volume, and AGB, may overcome these limitations and can produce comparable [23] and more accurate results [24] compared to field-based estimates of large-area forest inventory.

Lidar, an RS method that uses laser scanning to acquire three-dimensional information over the desired area [25], provides an alternative approach to estimating forest inventory metrics [1,21]. By using lidar, forest characteristics such as tree heights and canopy cover can be directly estimated, and other measures such as BA, volume, and AGB can be estimated by means of modeling using a variety of metrics derived from directly measured height and canopy cover [26,27]. The most commonly used approach to modeling BA, volume, and AGB from lidar data is by developing a multiple linear regression from field inventory and lidar variables [28]. Often the best subsets approach, such as forward or backward selection, is used to determine a model with variables that best explain the dependent variable and further validated using a cross-validation or set validation approach [27,29,30]. These models usually contain at least one variable explaining tree height, another explaining canopy cover, and another that accounts for variation in the data, such as a height standard deviation variable [6,31,32]; however, the variables in these models may differ depending on the study site and the foliage type being measured [28,33,34]. Estimates of BA, volume, and AGB can also be acquired from non-parametric machine learning approaches such as random forest (RF), a machine learning algorithm that uses random and iterative samples of the data to produce regression trees and bootstraps data for robust predictive models [35–37].

Estimating forest metrics may be performed by using an area-based approach or a tree-level approach, and previous studies modeling BA, volume, or AGB have seen success in both approaches. For example, in estimating forest attributes from individual trees, researchers found that BA, volume, and tree heights of longleaf pine in Georgia, U.S., could be estimated with accuracies of  $R^2 = 0.18, 0.94,$  and  $0.96$ , respectively [38]. The results for estimating volume and tree heights of individual trees are promising, and the poor results for estimating BA were explained by the loss of height-diameter allometry in southern pines above 25 m. In another study, researchers estimated BA for a plantation of loblolly pines (*Pinus taeda*) (another southern pine species) using an area-based approach

and were able to achieve accuracies of  $R^2 = 0.97$ , and noted that the highly homogeneous environment of the pine plantations that the trees were grown in likely led to such high accuracies [27]. Using an area-based approach across a large area can be beneficial in that it reduces the amount of processing required by the computer. Similarly, variables can be derived from multispectral imagery, such as the normalized difference vegetation index (NDVI) and texture co-occurrence, which have also proven useful for modeling [34]. While predictive modeling using variables from either lidar or multispectral imagery has been successful in many cases, particularly in northern and western forests where pine species dominate [17,30], more difficulty is found in estimating forest inventories in heterogeneously mixed forests [28,34,38]. This, in addition to the perceived cost of remote sensing data, has caused delays in adopting RS technology for use in some forest inventories. However, the combination of lidar data, multispectral imagery, and field data for modeling could improve estimates of BA, volume, and AGB in mixed forests [39–41]. Furthermore, open-source data are becoming more available and can be used for forest inventories in place of otherwise expensive data sources. Finally, the implications of achieving accurate estimates of forest inventory metrics go beyond timber cruising and carbon estimation alone. By achieving wall-to-wall estimates of forest inventory metrics, past and future data can be compared for forest management techniques and landscape ecology analysis. Most importantly, the wall-to-wall maps produced in this study can be used as base maps for validation of satellite lidar data produced by systems such as the Global Ecosystem Dynamics Investigation (GEDI) lidar aboard the International Space Station (ISS) [42]. The goal of this study was to produce wall-to-wall estimates of BA, volume, and AGB in a southern mixed-species forest using open-source RS data. This goal was achieved by meeting the following objectives:

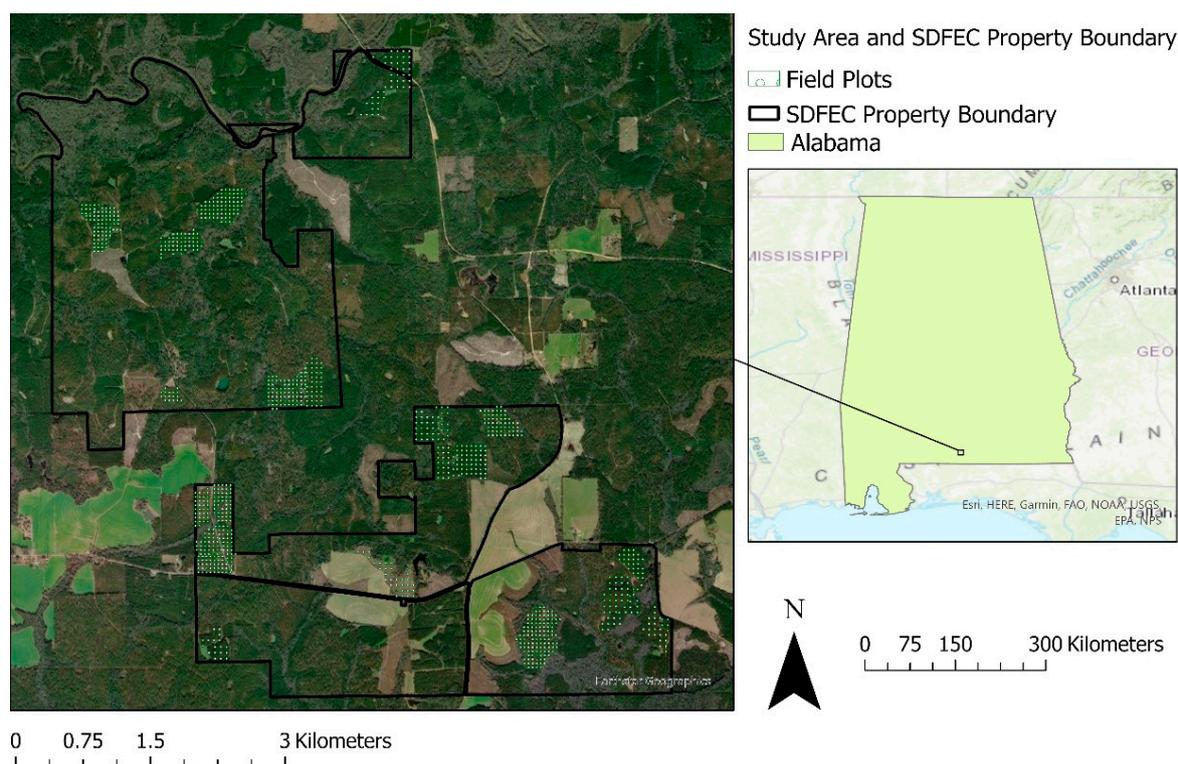
- A. Determine whether the addition of variables derived from multispectral imagery to models previously including only lidar derived variables can improve estimates of BA, volume, and AGB;
- B. Determine what variables are useful for modeling each forest metric (BA, volume, AGB) for Southern mixed forests.

If sufficiently accurate, models predicting forest inventory metrics can be used a number of times in forest inventory [17], and final wall-to-wall map outputs can be used for the validation of satellite data.

## 2. Materials and Methods

### 2.1. Study Area

The study area is a 35.35 km<sup>2</sup> site in Covington and Escambia counties, Southern Alabama (Figure 1). The site is within one of seven-level III ecoregions in Alabama, known as the Southern Plains ecoregion (ecoregion 65), whose natural vegetation is described as southern mixed pine forests and whose geography includes rolling hills and a warmer temperate climate [43]. The region can be further described at the level IV ecoregion (65f) as having loamy acidic soils and dark tea-colored streams with mixed forests interspersed with pine plantations [43–45]. Southern pine species that persist and typically dominate in this region include Longleaf pine (*Pinus palustris*), Shortleaf pine (*Pinus echinate*), Slash pine (*Pinus elliottii*), Loblolly pine (*Pinus taeda*), and Virginia pine (*Pinus virginiana*). The mean elevation of the study site is 59.47 m, with a minimum and maximum elevation ranging from 25 to 99 m, respectively, and it receives approximately 135 cm of rainfall a year. Within the study area is the Solon Dixon Forestry Education Center (SDFEC), a multi-use facility (outreach, extension, teaching, and research) primarily focused on providing hands-on field training to university students. Forest inventory data for this study come from the SDFEC and are described in more detail in the following sub-sections.



**Figure 1.** Study site with SDFEC property boundary and field plots, ESRI “World Imagery” [46].

### 2.1.1. Inventory Data

The SDFEC consists of both naturally regenerated and planted pine, hardwood, and mixed pine-hardwood forest stands intended for timber production. Forest inventory data from the SDFEC were collected from 1048 circular plots of 0.04047 hectares (1/10th acre) size in 304 forest stands (see Figure 1). The number of plots used in this study was reduced from 1048 to 523 to include only plots that were inventoried within two years of the 2017 lidar and imagery data acquisition. Plots are generated using TCruise software [47] at approximately one plot per 0.4047 hectares (one plot per acre) with a distance between plot centers of approximately 44.20 to 51.82 m (145.0 to 170.0 ft). The data included latitude and longitude of plot centers, plot area (0.04047-hectare plots), merchantable timber heights of individual trees (meters), diameter at breast height for individual trees (centimeters), and plot BA ( $\text{m}^2 \text{ha}^{-1}$ ). The plot volume ( $\text{m}^3 \text{ha}^{-1}$ ) and plot sum of AGB ( $\text{Mg ha}^{-1}$ ) were later calculated (see processing approach). The SDFEC uses lumber mill specifications when taking plot data on a forest inventory cruise. Trees surveyed in a plot must meet a minimum diameter at breast height (DBH) of 11.68 cm (4.598 inches) for a 12.70 cm (5.0 inches) class while being at least 4.600 m in height to a 7.620 cm (3.000 inches) top. Trees within plots that do not meet these specifications are not tallied. Furthermore, the SDFEC only tallies trees that local forest product mills are buying, so a variety of species, including Eastern Red Cedar (*Juniperus virginiana*) and Southern Black Cherry (*Prunus serotina*), are not counted.

### 2.1.2. Airborne Lidar

Airborne lidar data used in this study come from the United States Geological Survey (USGS) 3D Elevation Program (3DEP) [48]. The 3DEP is an open-source program originally designed to collect three-dimensional geographic elevation data for a given landscape; however, the data have been used in many fields, including forestry and ecology. The data used in this study were collected by a Leica ALS70-HP RS device mounted in a fixed-wing position aboard Cessna and Partenavia aircrafts. Characteristics of the flight included a scan angle of 45 degrees, a flight height of 2000 AGL meters, and a speed of 130 kts. Lidar scanning characteristics included a 1 m footprint, a pulse density of 2.4 pls/ $\text{m}^2$ , and a scan

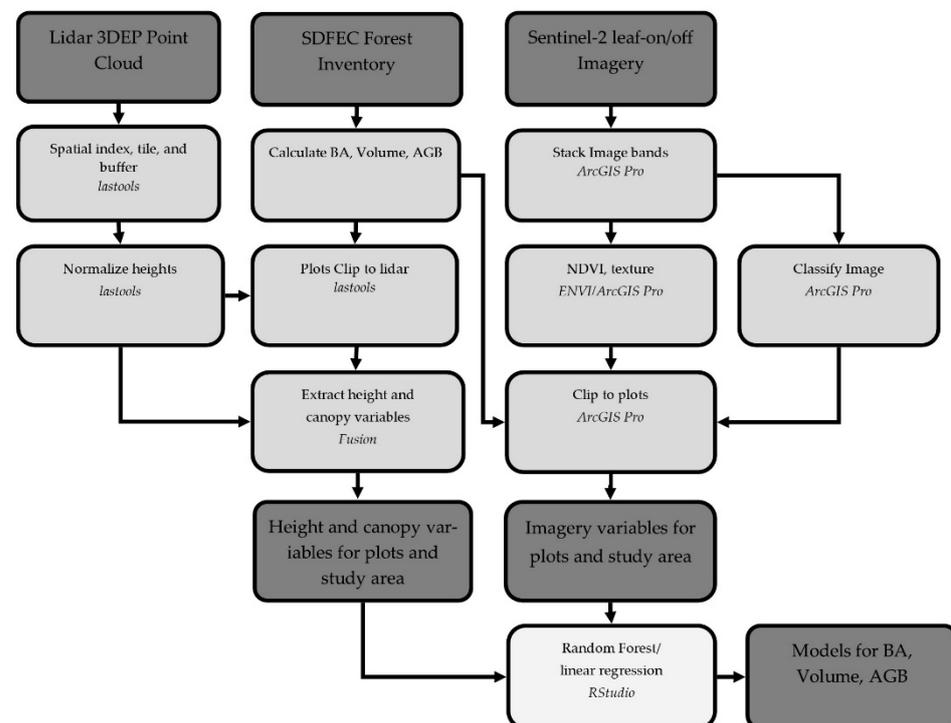
frequency of 35.1 Hz. The scan pattern was triangular. Fifty-four lidar files containing point cloud data collected in 2017 were downloaded from 3DEP using UGet software [49]. Each file contained a las and/or laz tile with associated metadata.

### 2.1.3. Multispectral Imagery

Multispectral imagery data were acquired from Copernicus Sentinel-2. The Copernicus Sentinel-2 mission was developed by the European Space Agency (ESA) with the purpose of providing global and continuous land cover data in addition to map products such as change detection products and land cover maps. The mission aims to provide data for studies focused on climate change, land management, and security [50]. Leaf-on and leaf-off Sentinel-2A Level-1C (L1C) top-of-atmosphere (TOA) reflectance orthoimages were downloaded from USGS EarthExplorer for the dates 19.VII.2016 and 24.II.2017. L1C products are derived from the L1B product and use radiometric and geometric corrections to produce the final product. The downloaded Sentinel-2 imagery consists of a ten-day, ten-meter spatial resolution for four spectral bands, including blue band 2 (490 nm), green band 3 (560 nm), red band 4 (665 nm), and NIR band 8 (842 nm). Imagery for this study consisted of less than twenty percent cloud cover.

### 2.2. Data Processing Approach

Basal area, volume, and AGB were calculated from field inventory data. Variables were derived from lidar and multispectral imagery for the study site. These variables were also extracted from within the boundaries of the field inventory plots for further modeling. A workflow is displayed in Figure 2.



**Figure 2.** Data processing workflow.

#### 2.2.1. Field Inventory

Equations for calculating plot BA, volume, and AGB are shown in Equations (1)–(3), respectively [4,10,12]. Summary statistics of the forest inventory are displayed in Table 1.

**Table 1.** Summary statistics of inventory plots.

	BA (m <sup>2</sup> ha <sup>-1</sup> )	Volume (m <sup>3</sup> ha <sup>-1</sup> )	AGB (Mg ha <sup>-1</sup> )
Min	1.583	4.253	5.498
Max	45.92	412.8	226.4
Mean	18.33	193.7	80.84
Std. dev.	7.121	68.53	33.27

Basal area was calculated using the following equation

$$\text{Sum of BA}_n = \sum((0.00007854 \times \text{DBH}^2)/0.04)_i \quad (1)$$

where DBH is equal to the diameter (centimeters) at breast height (1.30 m), 0.00007854 is a forester's constant (calculated from  $\pi/(40,000)$ ) that converts centimeters to meters squared, and "i" is a tree within plot n [4]. The summed BA per plot was then calculated on a per hectare basis by dividing the value by the plot size (0.04 ha).

In order to match specifications from the DBH and height data from the SDFEC, a merchantable volume equation (Equation (2)) was used to calculate volume on a per hectare basis for hardwoods and pines

$$\text{Sum of Volume (m}^3 \text{ ha}^{-1}\text{)} = \sum(\alpha + \beta(\text{DBH}^2 \times \text{Ht}))_i \quad (2)$$

where  $\alpha$  and  $\beta$  are species-specific parameters estimated from the equation [10]. In this study, species were broadly grouped into pine and hardwood groups. For pines, the longleaf pine parameters were used because they constitute the majority of pine BA, and for hardwood species, the "unknown hardwood" parameters were used because species-specific information for hardwoods was unavailable. The parameters for each group are displayed in Table 2. Lastly, DBH is equal to the diameter at breast height (centimeters), and Ht is the merchantable height of the tree.

**Table 2.** Species-specific parameters for calculating volume.

	Pines	Hardwoods
$\alpha$	−0.4432	0.8235
$\beta$	0.002165	0.001630

Finally, aboveground biomass was calculated on a per hectare basis using the following

$$\text{Sum of AGB (Mg ha}^{-1}\text{)} = \sum(\text{Exp}(\beta_0 + \beta_1 \ln \text{DBH})) \quad (3)$$

where  $\beta_0$  and  $\beta_1$  are species specific parameters (Table 3) [12].

**Table 3.** Species-specific parameters for calculating AGB.

	Pines	Hardwoods
$\beta_0$	−2.536	−2.480
$\beta_1$	2.435	2.483

### 2.2.2. Lidar

Lidar data were processed using LASTools [51] and FUSION [52] software. In LASTools, aboveground tree heights were computed by normalizing points classified as ground returns and measuring the vegetation above ground points. Point cloud data were then clipped to the extent of field inventory plots in ArcGIS Pro. In FUSION, height and canopy

metrics were calculated based on the normalized point cloud and the clipped plots. Plot field inventory data were combined with lidar metrics for modeling. The slope was calculated from a DEM derived from the point cloud and extracted per plot in ArcGIS Pro [53].

### 2.2.3. Multispectral Imagery Processing

Sentinel-2 imagery data were used to derive multispectral variables of interest, including normalized difference vegetation index (NDVI) and second-order co-occurrence texture (Table 4). NDVI is known as a greenness vegetation index ratio with a range of  $-1$  to  $1$  [54]. NDVI was calculated in ENVI [54] using the following equation

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}) \quad (4)$$

where NIR and red are the near-infrared and red band layers derived from the imagery. Second-order co-occurrence texture shows the relationship between one pixel value and its surrounding neighbors based on distance and angularity [55]. These metrics and NDVI were derived in ENVI as 20 m pixel resolution rasters. The means of these values were further extracted from the zone of each field plot in ArcGIS Pro using the zonal statistics as a table tool in ArcGIS Pro.

**Table 4.** Variables used in modeling. Note that repetitive variables such as height percentiles and texture metrics for each of 4 bands were grouped to reduce table size.

Variable	Description
1. Max	Maximum height for a tree in a given plot
2. Mean	Mean height of trees in a given plot
3. Variance	Variance of tree heights in a plot
4. ElevCV	Co-variance of heights in a plot
5. p10, 25, 50, 75, 90, 95, 99	Height percentiles of trees in a plot
6. cancovA	% first returns above 4.6 m
7. cancovB	% all returns above 4.6 m
8. dens_0_10	Count of returns within 0–10 m
9. dens_10_15	Count of returns within 10–15 m
10. dens_15_20	Count of returns within 15–20 m
11. stddev	Standard deviation of tree heights
12. con_1,2,3,4	Texture contrast for bands r,b,g,NIR
13. cor_1,2,3,4	Texture correlation for bands r,b,g,NIR
14. dis_1,2,3,4	Texture dissimilarity for bands r,b,g,NIR
15. ent_1,2,3,4	Texture entropy for bands r,b,g,NIR
16. hom_1,2,3,4	Texture homogeneity for bands r,b,g,NIR
17. mean_1,2,3,4	Mean texture for bands r,b,g,NIR
18. secmom_1,2,3,4	Angular second moment for bands r,b,g,NIR
19. var_1,2,3,4	Texture variance for bands r,b,g,NIR
20. NDVI	NDVI within a plot
21. Land_cover	Landcover classes including Deciduous, Evergreen, and Barren
22. Slope	Slope calculated as a degree of inclination

With a temporal resolution of 10 days, Sentinel-2 data have the advantage of leaf-on and leaf-off acquisitions and made it possible to develop a landcover map distinguishing coniferous and deciduous vegetation in the SDFEC. This was achieved by stacking four bands (blue, green, red, and NIR) from each season (leaf-on and -off) and a canopy height model (CHM) derived from the lidar together, resulting in a nine-band image. Regions of interest were gathered from bands representing leaf-on and leaf-off years, and by using maximum likelihood supervised classification, a landcover map was produced with the following cover types: water, barren land, agriculture, deciduous vegetation, and evergreen vegetation. A confusion matrix was used to assess the accuracy of the supervised classification. The kappa coefficient, an indicator of how the classification of an image compares to ground truth data, was computed from the confusion matrix. This value ranges from 0 to 1, where 0 suggests that no none of the pixels were correctly classified, and 1 indicates that

all pixels were correctly classified. The kappa for the classified image was 0.74, and the user accuracies were highest in the water and evergreen classes (user accuracy > 0.95), with the most confusion occurring in the agricultural and barren land cover classes. The highest producer accuracy occurred in evergreen and deciduous classes (0.73 and 1.00, respectively). Land cover types were extracted from each forest inventory field plot. A plot was then assigned a land cover type based on the dominant land cover type in that plot. Table 4 lists the 52 lidar, spectral, textural, and land cover metrics used in this study.

### 2.3. Data Analysis

Two modeling approaches were applied for predicting BA, volume, and AGB: multiple linear regression and Random Forest (RF). Both modeling approaches were implemented using the R programming language in RStudio [56]. Before modeling via linear regression or RF, the categorical landcover variables needed to be dummy coded to be used in linear regression modeling. Therefore, two data sets were created for each dependent variable (BA, volume, and AGB), one with dummy coded land cover variables and one with categorical land cover variables.

#### 2.3.1. Regression Modeling

For the multiple linear regression approach, the `regsubsets` function from the `leaps` package in R [57] was used to select variables and develop models for BA, volume, and AGB using forward and backward selection. In forward variable selection, a model is produced by starting with one variable and adding variables to the model until the best possible model is produced. Conversely, the backward variable selection includes every variable in the model and removes a variable until the best model is selected. The best five models were developed for up to the best five variable models based on the highest  $R^2$ , lowest Mallows'  $cp$ , and lowest Bayesian Information Criterion (BIC). A Breusch–Pagan test determined that the response variables were not normally distributed, so a log transformation was applied to each [58]. In order to avoid multicollinearity, variance inflation factors (VIF) were calculated for each variable in each model, and models whose predictor variable's VIFs exceeded 10 were removed [59]. A threshold of 0.05 was used to determine variable significance.

#### 2.3.2. Random Forest (RF)

RF is a non-parametric machine learning modeling technique that uses random bagging and a bootstrap sample of the data to create a number of user-specified trees that vote on the best model parameters at each node of each tree [35]. The major benefit of RF in this study is that it can handle a large number of variables and work with data that may otherwise push the boundaries of the assumptions made by multiple linear regression, such as the assumption that there is no multicollinearity. The "RF" function in the `ModelMap`, a package in R [60], was used for modeling BA, volume, and AGB. This package easily allows users to input data and tune hyperparameters to obtain a specified raster output. Important user set hyperparameters for RF include the `mtry` and `ntree` parameters. The `mtry` parameter is the number of randomly selected variables at each node and was automatically optimized. The `ntree` parameter is the number of trees grown in the model and was set to 500 trees. After model building, the model can be used to predict BA, volume, and AGB across the study extent. Raster layers for each lidar and imagery independent variable were created for the extent of the study site with the same number of rows and columns at a 20 m resolution. These rasters were referenced using a look-up table (LUT), and the values for each cell are entered into the RF model to produce an estimation of the desired metric across the study site.

#### 2.3.3. Model Evaluation

Selected linear regression models were evaluated using 10-fold cross-validation. This approach first randomly divides the dataset into 10 groups (folds) of equal sizes and then

tests the data on one group using 1–10 folds to fit the data. The process repeats  $k$  number of times, computing the mean squared error (MSE) for each test group and averages the MSE of all tests to obtain the  $k$ -fold cross-validation estimate [61]. While leaving one out, cross-validation is a common model evaluation approach [27,62,63]; it was shown that  $k$ -fold cross-validation could reduce the test error rate [61]. After  $k$ -fold validation, the predicted values are back-transformed and compared to the observed values. The final model used for mapping was selected based on the  $R^2$ , RMSE, percent root mean squared error (%RMSE), mean absolute error (MAE), and bias. The resulting models were used to create wall-to-wall maps of the independent variables (BA, volume, AGB) using the raster package in R [64]. F-tests were used to determine if models with imagery-derived variables significantly improved the model, and  $R^2$  and %RMSE were compared.

In evaluating the RF model, the out-of-bag error (OOB) was used to determine the prediction error. In the RF model building process, random bootstrap samples were used to build decision trees. Some of the data were left out of each sample; this is the OOB sample. The OOB error is the amount of error produced from wrongly predicted OOB samples. Evaluation of the RF models included calculating the OOB RMSE, %RMSE,  $R^2$ , MAE, and bias. In order to evaluate whether imagery-derived variables were useful in the RF model, variable importance plots were produced. Importance is calculated as the decrease in model accuracy as variables are removed, or the percent increase in mean square error (%IncMSE).

### 3. Results

#### 3.1. Linear Regression Model

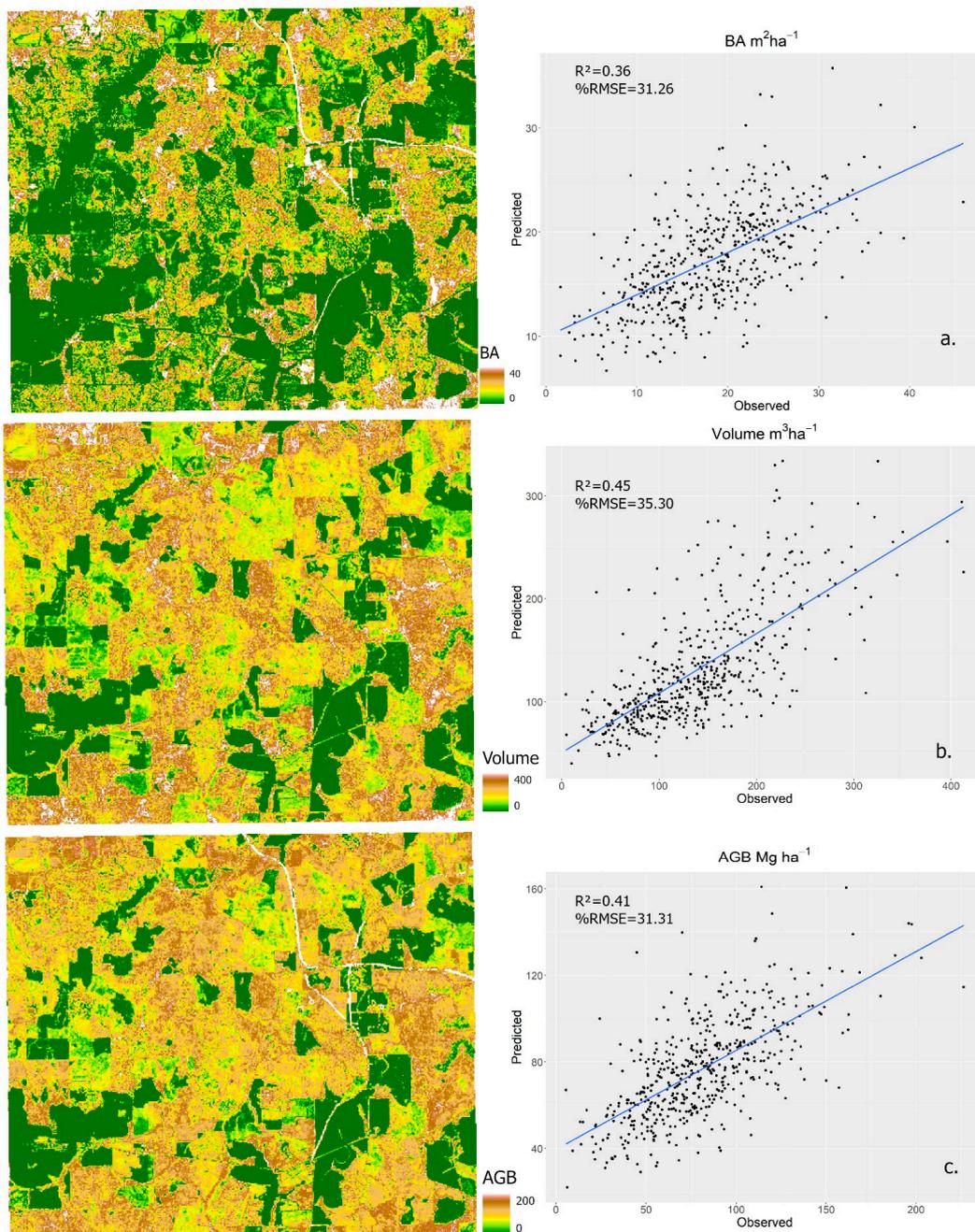
From the stepwise variable selection, the forward selection approach produced the best models, followed by backward selection. A table of the five best models for each response variable is displayed in Appendix A (Table A1). According to the calculated vif's, no variable in a model contributed to multicollinearity ( $VIF < 5$ ).  $p$ -values indicated that each variable was significant to each of the models ( $p < 0.05$ ). Models produced for BA, volume, and AGB in this study generally included a lidar height, canopy, and density metric, though BA did not include any lidar-derived height metric. Beyond three-variable models, the var\_3 variable was included in the four- and five-variable model for BA. In the four- and five-variable model for volume, the hom\_1 variable was included in addition to another lidar-derived density metric. Four and five variable models for AGB included var\_2 and hom\_1 variables, respectively, and the five-variable model included both.

The final model for volume outperformed those of BA and AGB. In model building, the five-variable model for volume had the lowest Mallows' cp and BIC and the highest  $R^2$ . After evaluating models from the 10-fold cross-validation approach, the five-variable model had the lowest error (RMSE = 49.23 m<sup>3</sup> ha<sup>-1</sup>, %RMSE = 35.30) and the highest accuracy between the observed and predicted values ( $R^2 = 0.45$ ). This model included the following variables: p50, cancovA, dens\_15\_20, ElevCV, and hom\_1. An F-test revealed that the hom\_1 variable in the best model significantly improved the model ( $p < 0.05$ ); however, the  $R^2$  and %RMSE were not substantially different; the model without the 1\_hom variable had the following accuracy:  $R^2 = 0.44$ ; %RMSE = 35.96.

For BA, the model with the lowest BIC, Mallows' cp, and highest  $R^2$  was the model with five variables. After validating BA using the 10-fold cross-validation approach, the model with five variables had the least amount of error (RMSE = 5.731 m<sup>2</sup> ha<sup>-1</sup>, %RMSE = 31.26). Furthermore, the predictive accuracy between predicted and observed values using the five-variable model was highest ( $R^2 = 0.36$ ). This model included the following variables: cancovA, dens\_0\_10, dens\_10\_15, dens\_15\_20, and var\_3. An F-test revealed that the var\_3 variable in the best model significantly improved the model ( $p < 0.05$ ); however, the  $R^2$  and %RMSE were not substantially different, and the model without the var\_3 variable had the following accuracy:  $R^2 = 0.35$ ; %RMSE = 31.61.

In modeling AGB, the model with the lowest BIC, Mallows' cp, and highest  $R^2$  was the five-variable model. The five-variable model from the 10-fold cross-validation approach

also had the lowest RMSE ( $25.20 \text{ Mg ha}^{-1}$ ,  $\%RMSE = 31.26\%$ ) and the highest correlation between observed and predicted values ( $R^2 = 0.41$ ). The following variables were included in this model:  $p50$ ,  $cancovA$ ,  $dens_{15\_20}$ ,  $var\_2$ , and  $hom\_1$ . An F-test was used to determine if the addition of one or both of the image-derived variables significantly improved the model. The addition of one or both of the image-derived variables was significant according to the F-test ( $p < 0.05$ ); however, the  $R^2$  and  $\%RMSE$  for all three models were not substantially different, the model without  $hom\_1$  had an accuracy of  $R^2 = 0.41$ ;  $\%RMSE = 31.60$ , the model without both  $hom\_1$  and  $var\_2$  had  $R^2 = 0.38$ ;  $\%RMSE = 32.09$ . Selected models with coefficients are presented in Table 5, and scatterplots showing model prediction and observed values are displayed in Figure 3 with map outputs at a 20 m resolution (approximately the size of the field inventory plots).



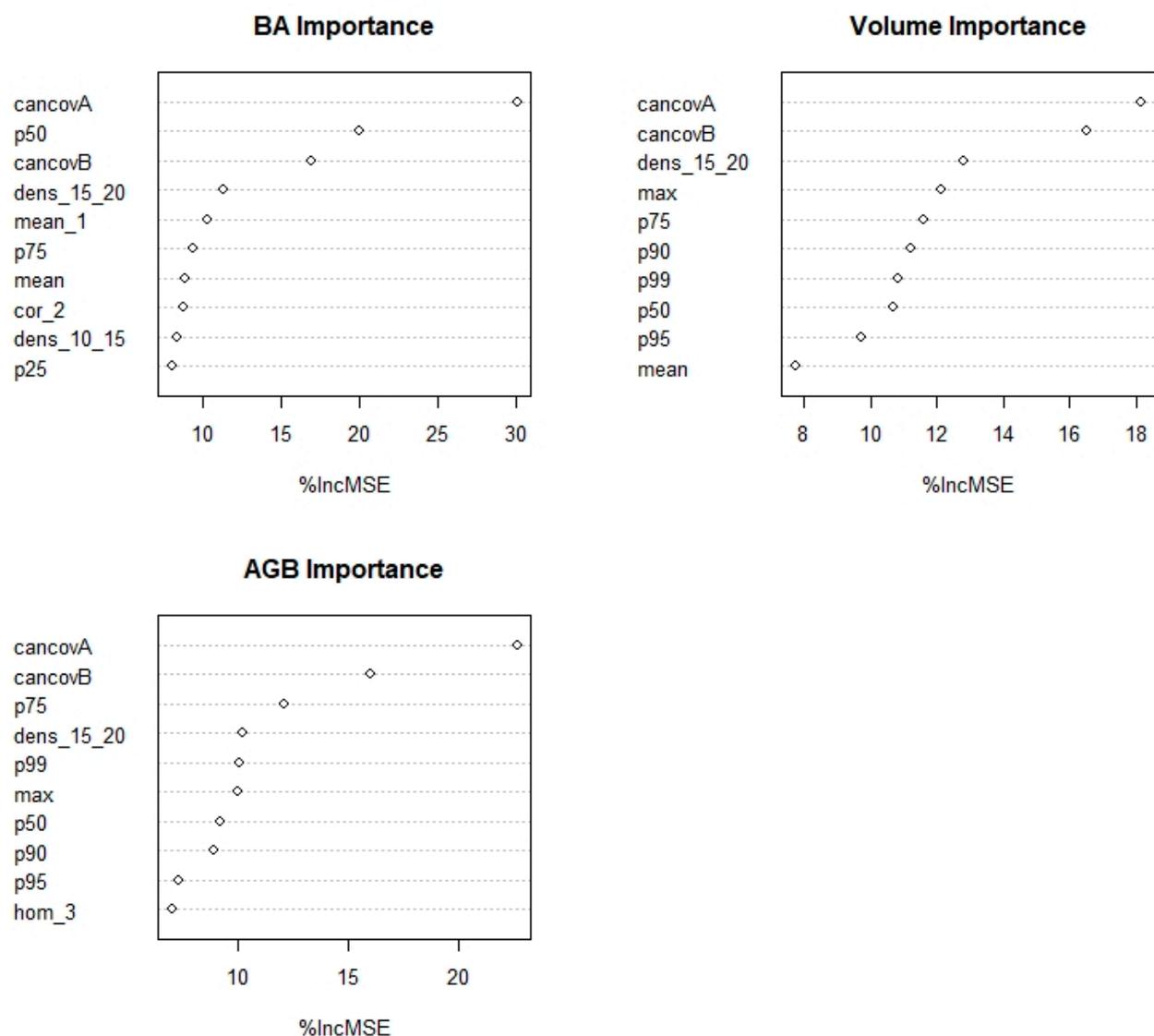
**Figure 3.** Predicted map outputs of BA (a), volume (b), and AGB (c) with scatterplot displaying correlation between observed and predicted values for the selected linear regression model.

**Table 5.** Models predicting BA, Volume, and AGB.

Forest Metric	Model	%RMSE	RMSE	MAE	R <sup>2</sup>	Bias
BA	BA = 1.78 + 0.0173cancovA – 1.73dens_0_10 – 1.06dens_10_15 + 0.672dens_15_20 + 0.0245var_3	31.26	5.731	4.456	0.36	0.9528
Volume	Vol = 3.09 + 0.0769p50 + 0.0145cancovA - 1.03dens_10_15 – 0.590hom_1 + 0.213Elev_CV	35.30	49.23	37.32	0.45	8.287
AGB	AGB = 2.53 + 0.0609p50 + 0.0136cancovA + 0.924dens_15_20 + 0.0406var_2 – 0.432hom_1	31.26	25.20	19.35	0.41	4.064

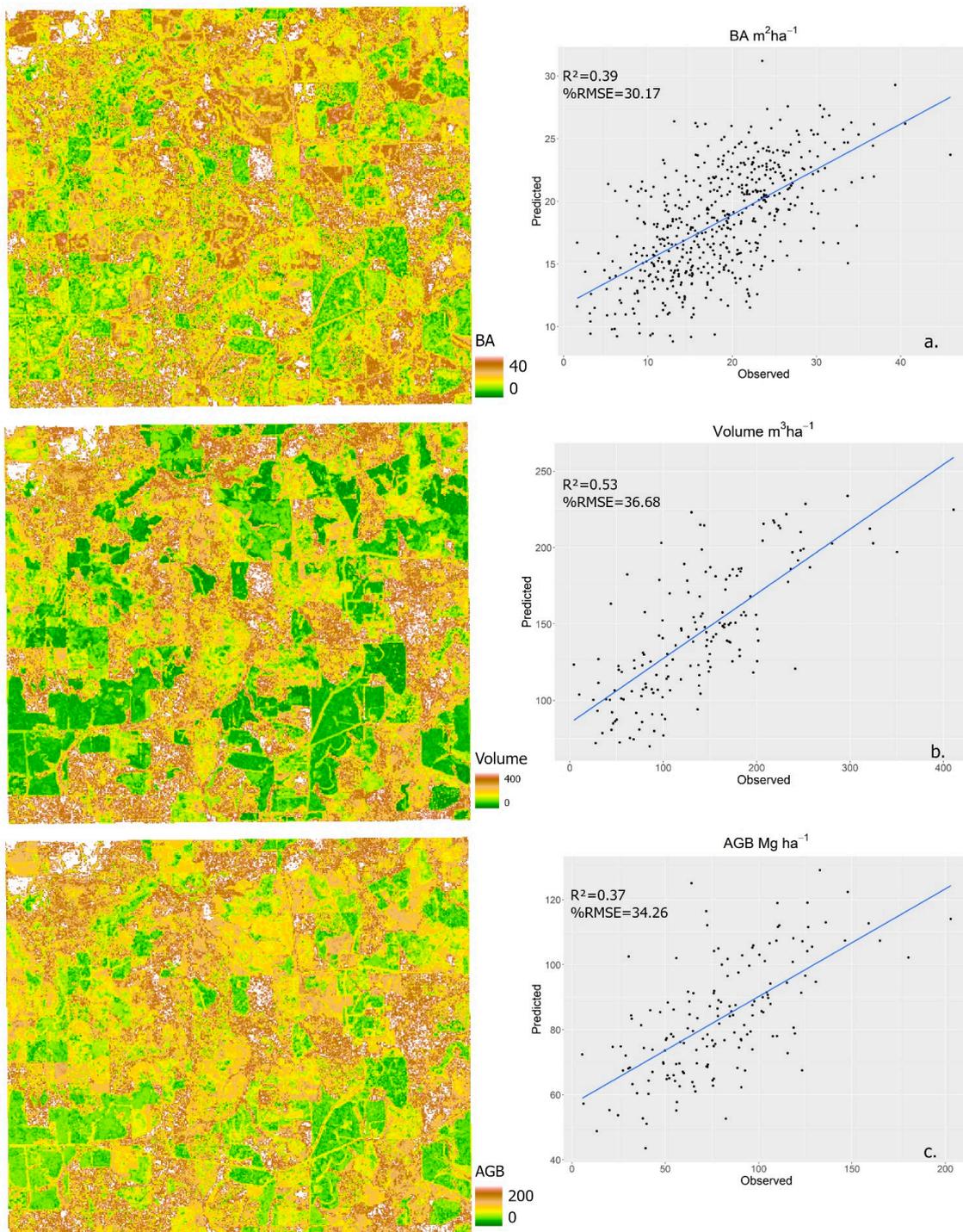
### 3.2. Random Forest Model

The RF models produced accuracies similar to those from the linear regression models. Model diagnostics revealed that volume again had the highest model accuracy ( $R^2 = 0.53$ , RMSE = 50.21 m<sup>3</sup> ha<sup>-1</sup> %RMSE = 36.68), followed by BA ( $R^2 = 0.39$ , RMSE = 5.662 m<sup>2</sup> ha<sup>-1</sup> %RMSE = 30.17) and AGB ( $R^2 = 0.37$ , RMSE = 26.29 Mg ha<sup>-1</sup> %RMSE = 34.26). Table 6 includes model evaluation from the observed vs. predicted values. Variable importance was calculated and plotted (Figure 4). Figure 5 display scatterplots for observed and predicted BA, volume, and AGB from random forest modeling.

**Figure 4.** Importance plots for variables used in RF models.

**Table 6.** Model evaluation of the RF predictions.

RF Model	%RMSE	RMSE	MAE	R <sup>2</sup>	Bias
BA	30.17	5.662	4.449	0.39	0.01053
Volume	36.68	50.21	38.18	0.53	−6.438
AGB	34.26	26.29	20.77	0.37	−4.656

**Figure 5.** Predicted map outputs of BA (a), volume (b), and AGB (c) with scatterplot displaying correlation between observed and predicted values for the RF models.

#### 4. Discussion

Models predicting BA, volume, and AGB in this study ranged in accuracy from  $R^2 = 0.36$  to  $0.53$ . In the RF models, the  $R^2$  values were higher than those produced by the linear regression modeling approach, except for when predicting BA. Both RMSE and %RMSE were lower in linear regression models, except when predicting BA. Other researchers modeling forest metrics reported findings that RF modeling tends to have the smallest RMSE and the least amount of bias [36,65]. Because the predictive accuracy between the two modeling types was similar, the findings of this study suggest that the results of a modeling approach could depend more heavily on the variables used in modeling, as well as the forest type being modeled (mixed forest stands, homogeneous plantations). The models produced for volume had the greatest predictive accuracy ( $R^2 = 0.53$ , RMSE =  $50.21 \text{ Mg ha}^{-1}$ , %RMSE = 36.68), likely because the model used to calculate volume from field inventory included a height variable, while those for BA and AGB only included DBH and associated parameters. By calculating volume from field inventory using an equation that includes a height metric, lidar-derived height variables account for variation in heights when predictive modeling. Less accurate models developed for BA and AGB were also likely due to a number of other factors. For example, the difference in tree heights from when the field inventory was taken and when the lidar tree heights were measured could cause some variation; however, this is likely minimal as a subset of plots was used within two years of the lidar acquisition. The most likely explanation for the unexplained variation in BA and AGB models is due to a combination of the merchantable timber cruising process used to collect this field data and to the high heterogeneity of southeastern forests [34]. In other words, a lot of the foliage measured by lidar and multispectral imagery is not accounted for in the field inventory gathered by inventory cruisers. This is especially seen in the model building process for BA, where variables selected for each model were those whose function either distinguished layers of the canopy (i.e., density metrics trying to explain shrubs and non-inventoried trees beneath the tops of pines) or those that distinguished foliage type (texture bands). Height metrics were excluded from these models because they could not explain BA for all of the foliage within a plot, especially when the data that the models were built from were based only on merchantable timber instead of the entirety of the foliage constituting the plot BA. Furthermore, while there is a relationship between a tree's height and its BA, in many species, this relationship is known to be somewhat weak [28,38], and another reason why there are no height metrics in the model for BA. Similar findings were reported by other researchers working in heavily mixed forest stands, such as those in Canada, where BA was modeled with an  $R^2$  of only 0.093 [28]. In that particular study, the researchers noted that it was difficult to account for most of the vegetation not measured by timber cruisers and recommended using another data source in addition to lidar.

From the models developed in this study, spectral, textural, and landcover variables were considered in addition to lidar to delineate any attributes of the forest not accounted for in the field inventory. As mentioned in the results section, the model for BA included a texture variable (var\_3). Furthermore, the four and five variable models for volume and biomass included texture variables (var\_2, hom\_1). This suggests that in heavily mixed forests, spectral, textural, and landcover class variables have the potential to explain variation missed by lidar variables alone; however, the difference in model accuracy after removing imagery variables was not substantial, and lidar variables alone predicted BA, volume and AGB almost as well. This is not to say that lidar variables are sufficient in estimating lidar in Alabama mixed-species forests, as the accuracy of lidar-derived models were still poor in BA, volume, and AGB. While many papers suggest that lidar alone is sufficient for modeling, most of these studies are developed for study sites whose forests are primarily homogeneous such as those in the western United States [59,66,67]. In estimating BA, for example, a study in western Oregon achieved very high predictive accuracy using a model with two height variables and a canopy cover variable ( $R^2 = 0.96$ ), but the study site was homogeneous in nature, dominated by old-growth coniferous forests [30]. Because of

the variation in allometric dimensions among tree species in mixed forest stands, alternative variables are necessary to accurately predict forest metrics.

## 5. Conclusions

RS data, including airborne and spaceborne lidar, as well as multispectral imagery from satellite-based platforms, are becoming a more available resource to foresters and ecologists. Programs such as the 3D Elevation program from the USGS, the National Agricultural Imagery Program from the USDA, and Sentinel-2 from the ESA provide free data that foresters can use to estimate the amount of BA, volume, and AGB on their stands and across large tracts of forested land. In this study, the potential of estimating BA, volume, and AGB from RS data was demonstrated by means of linear regression analysis and random forest modeling. Predictive accuracies of models were low ( $R^2 = 0.36\text{--}0.53$ ) relative to those in some studies [27,59], and the results presented here suggest that the addition of imagery variables to lidar derived predictive models do not substantially help explain the variation in BA, volume, and AGB. While this study demonstrates the potential for a fast and efficient method of estimating BA, volume, and AGB using freely available data, further investigation of variables is needed to increase the variability of forest structure in mixed forest stands. Lastly, the main limitation of this study is that the field inventory used to build the models included merchantable timber and not the entire vegetation in one forest plot. Therefore, the potential to improve forest metric estimates may be improved if the entirety of the vegetation in a forest plot is measured.

Open-source RS data are increasingly available, and as demonstrated in this study, freely available data can be leveraged for estimating BA, volume, and AGB in a spatially explicit manner. Lastly, improving estimates of BA, volume, and AGB are necessary to produce accurate reference maps that can be used for the validation of forest measures derived from satellite lidar and imagery.

**Author Contributions:** Conceptualization, S.B. and L.L.N.; methodology, S.B. and L.L.N.; formal analysis, S.B.; investigation, S.B.; resources, S.B. and L.L.N.; writing—original draft preparation, S.B.; writing—review and editing, S.B., L.L.N., and J.G.; visualization, S.B.; supervision, L.L.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Forest inventory data is owned by the SDFEC and may be available upon request. All other data are publically available. 3DEP lidar data can be found at <https://prd-tnm.s3.amazonaws.com/LidarExplorer/index.html#/>. Sentinel data can be found at <https://earthexplorer.usgs.gov/>.

**Acknowledgments:** We acknowledge and thank the SDFEC, a part of Auburn University, for collecting and providing the forest inventory data used in this study.

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

## Appendix A

**Table A1.** The best models produced for models of one to five variables for BA, volume, and AGB.

Forest Metric	Model	%RMSE	RMSE	MAE	R <sup>2</sup>	Bias
BA <sub>1</sub>	BA = 2.15 + 0.0101cancovA	37.44	6.865	5.453	0.16	1.333
BA <sub>2</sub>	BA = 1.93 + 0.0164cancovA – 2.62dens_0_10	32.81	6.016	4.741	0.31	1.056
BA <sub>3</sub>	BA = 1.94 + 0.0171cancovA – 1.95dens_0_10 – 0.910dens_10_15	31.91	5.850	4.584	0.34	1.005
BA <sub>4</sub>	BA = 1.86 + 0.0181cancovA – 2.06dens_0_10 – 1.017dens_10_15 + 0.0240var_3	31.59	5.792	4.591	0.36	0.9676
BA <sub>5</sub>	BA = 1.78 + 0.0173cancovA – 1.73dens_0_10 – 1.06dens_10_15 + 0.672dens_15_20 + 0.0245var_3	31.26	5.371	4.456	0.36	0.9528
Volume <sub>1</sub>	Vol = 3.45 + 0.0797p50	39.92	55.75	42.30	0.32	11.85
Volume <sub>2</sub>	Vol = 2.53 + 0.0892p50 + 0.0115cancovA	36.35	50.77	37.72	0.45	8.893
Volume <sub>3</sub>	Vol = 2.82 + 0.0719p50 + 0.0132cancovA – 0.957dens_10_15	35.97	50.24	37.13	0.45	8.698
Volume <sub>4</sub>	Vol = 3.20 + 0.0721p50 + 0.0148cancovA – 1.11dens_10_15 – 0.594hom_1	35.32	49.33	37.30	0.45	8.299
Volume <sub>5</sub>	Vol = 3.09 + 0.0769p50 + 0.0145cancovA – 1.03dens_10_15 – 0.590hom_1 + 0.213Elev_CV	35.30	49.23	37.32	0.45	8.287
AGB <sub>1</sub>	AGB = 3.63 + 0.00989cancovA	40.01	32.20	24.88	0.15	6.368
AGB <sub>2</sub>	AGB = 2.52 + 0.0122cancovA + 0.0568p50	32.65	26.27	19.88	0.38	4.434
AGB <sub>3</sub>	AGB = 2.70 + 0.0132cancovA + 0.0464p50 – dens_10_15	32.46	26.12	19.65	0.38	4.387
AGB <sub>4</sub>	AGB = 2.57 + 0.0141cancovA + 0.0484p50 – 0.626dens_10_15 + 0.0534var_2	31.95	25.72	19.59	0.39	4.160
AGB <sub>5</sub>	AGB = 2.53 + 0.0609p50 + 0.0136cancovA + 0.924dens_15_20 + 0.0406var_2 – 0.432hom_1	31.26	25.20	19.35	0.41	4.064

## References

- Lister, A.J.; Andersen, H.; Frescino, T.; Gatzliolis, D.; Healey, S.; Heath, L.S.; Liknes, G.C.; McRoberts, R.; Moisen, G.G.; Nelson, M.; et al. Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory. *Forests* **2020**, *11*, 1364. [\[CrossRef\]](#)
- Zhang, Y.; Liang, S.; Yang, L. A Review of Regional and Global Gridded Forest Biomass Datasets. *Remote Sens.* **2019**, *11*, 2744. [\[CrossRef\]](#)
- Cannell, M.G.R. Woody Mass of Forest Stands. *For. Ecol. Manag.* **1984**, *8*, 299–312. [\[CrossRef\]](#)
- Elledge, J.; Barlow, B. Basal Area: A Measure Made for Management. 2010, p. 6. Available online: [https://www.aces.edu/wp-content/uploads/2018/10/ANR-1371.REV\\_2.pdf](https://www.aces.edu/wp-content/uploads/2018/10/ANR-1371.REV_2.pdf) (accessed on 22 February 2021).
- Wiant, H.V.W., Jr.; Baumgras, J.E.; Knight, R. *Relation of Biomass to Basal Area and Site Index on an Appalachian Watershed*; Department of Agriculture, Forest Service, Northeastern Forest Experiment Station: Broomall, PA, USA, 1984; pp. 53–56.
- Khan, M.N.I.; Islam, M.R.; Rahman, A.; Azad, M.S.; Mollick, A.S.; Kamruzzaman, M.; Sadath, M.N.; Feroz, S.M.; Rakkibu, M.G.; Knohl, A. Allometric Relationships of Stand Level Carbon Stocks to Basal Area, Tree Height and Wood Density of Nine Tree Species in Bangladesh. *Glob. Ecol. Conserv.* **2020**, *22*, e01025. [\[CrossRef\]](#)
- U.S. Fish and Wildlife Service. *Recovery Plan for the Red-Cockaded Woodpecker (Picoides borealis): Second Revision*; U.S. Fish and Wildlife Service: Atlanta, GA, USA, 2003.
- Vogt, J.; Koch, F.H. Forest Inventory and Analysis Data in Invasive Insect Research. *Am. Entomol.* **2016**, *62*, 46–58. [\[CrossRef\]](#)
- Bond, B. *PB1650 Understanding Log Scales and Log Rules*; The University of Tennessee Institute of Agriculture: Knoxville, TN, USA, 2011; p. 8.
- Oswalt, C.M.; Conner, R.C. *Southern Forest Inventory and Analysis Volume Equation User's Guide*; Department of Agriculture Forest Service, Southern Research Station: Asheville, NC, USA, 2011.
- Tian, N.; Fan, Z.; Matney, T.G.; Schultz, E.B. Growth and Stem Profiles of Invasive. *For. Sci.* **2017**, *63*, 569–576.
- Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsey, R.A. National-Scale Biomass Estimators for United States Tree Species. *For. Sci.* **2003**, *49*, 12–35. [\[CrossRef\]](#)

13. Burns, P.; Clark, M.; Salas, L.; Hancock, S.; Leland, D.; Jantz, P.; Dubayah, R.; Goetz, S.J. Incorporating Canopy Structure from Simulated GEDI Lidar into Bird Species Distribution Models. *Environ. Res. Lett.* **2020**, *15*, 095002. [[CrossRef](#)]
14. United Nations Framework Convention on Climate Change. *UNFCCC Key Decisions Relevant for Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+)*; United Nations Framework Convention on Climate Change: New York, NY, USA, 2016; p. 44.
15. Environmental Protection Agency (E.P.A). *Inventory of U.S. Greenhouse Gas Emissions and Sinks*; EPA: Washington, DC, USA, 2021.
16. Smith, J.E.; Heath, L.S.; Woodbury, P.B. How to Estimate Forest Carbon for Large Areas from Inventory Data. *J. For.* **2004**, *102*, 25–31. [[CrossRef](#)]
17. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T.; Jain, T.B.; Evans, J.S. Transferability of Lidar-Derived Basal Area and Stem Density Models within a Northern Idaho Ecoregion. *Can. J. Remote Sens.* **2018**, *44*, 131–143. [[CrossRef](#)]
18. Wilson, B.T.; Woodall, C.W.; Griffith, D.M. Imputing Forest Carbon Stock Estimates from Inventory Plots to a Nationally Continuous Coverage. *Carbon Balance Manag.* **2013**, *8*, 1. [[CrossRef](#)] [[PubMed](#)]
19. Babst, F.; Bouriaud, O.; Alexander, R.; Trouet, V.; Frank, D. Toward Consistent Measurements of Carbon Accumulation: A Multi-Site Assessment of Biomass and Basal Area Increment across Europe. *Dendrochronologia* **2014**, *32*, 153–161. [[CrossRef](#)]
20. Elzinga, C.; Shearer, R.C.; Elzinga, G. Observer Variation in Tree Diameter Measurements. *West. J. Appl. For.* **2005**, *20*, 134–137. [[CrossRef](#)]
21. Mcroberts, R.E.; Tomppo, E.O. Remote Sensing Support for National Forest Inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [[CrossRef](#)]
22. Westfall, J.A.; Woodall, C.W.; Hatfield, M.A. A Statistical Power Analysis of Woody Carbon Flux from Forest Inventory Data. *Clim. Change* **2013**, *118*, 919–931. [[CrossRef](#)]
23. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273.
24. White, J.C.; Wulder, M.A.; Buckmaster, G. Validating Estimates of Merchantable Volume from Airborne Laser Scanning (ALS) Data Using Weight Scale Data. *For. Chron.* **2014**, *90*, 378–385. [[CrossRef](#)]
25. Chuvieco, C. *Fundamentals of Satellite Remote Sensing an Environmental Approach*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2020; ISBN 978-1138583832.
26. Silva, C.A.; Klauber, C.; Hudak, A.T.; Vierling, L.A.; Jaafar, W.S.W.M.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting Stem Total and Assortment Volumes in an Industrial *Pinus Taeda* L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* **2017**, *8*, 254. [[CrossRef](#)]
27. Silva, C.A.; Klauber, C.; Hudak, A.T.; Vierling, L.A.; Fennema, S.J.; Corte, A.P.D. Modeling and Mapping Basal Area of *Pinus Taeda* L. Plantation Using Airborne LiDAR Data. *An. Acad. Bras. Cienc.* **2017**, *89*, 1895–1905. [[CrossRef](#)]
28. Spriggs, R.A.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A Critique of General Allometry-Inspired Models for Estimating Forest Carbon Density from Airborne LiDAR. *PLoS ONE* **2019**, *14*, e0215238. [[CrossRef](#)]
29. Holmgren, J. Prediction of Tree Height, Basal Area and Stem Volume in Forest Stands Using Airborne Laser Scanning. *Scand. J. For. Res.* **2004**, *19*, 543–553. [[CrossRef](#)]
30. Means, J.E.; Acker, S.A.; Harding, D.J.; Blair, J.B.; Lefsky, M.A.; Cohen, W.B.; Harmon, M.E.; McKee, W.A. Use of Large-Footprint Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sens. Environ.* **1999**, *67*, 298–308. [[CrossRef](#)]
31. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. *A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach*; Canadian Wood Fibre Centre: Victoria, BC, Canada, 2013; Volume 89, ISBN 9a781100223858.
32. Laes, C.D.; Reutebuch, S.E.; Mcgaughey, R.J.; Mitchell, B. Guidelines to Estimate Forest Inventory Parameters from Lidar and Field Plot Data. 2011; pp. 1–22. Available online: [https://fsapps.nwcg.gov/gtac/CourseDownloads/Reimbursables/FY21/Lidar\\_Material/GTAC\\_Guidelines%20to%20estimate%20forest%20inventory%20parameters%20from%20lidar%20and%20field%20plot%20data.pdf](https://fsapps.nwcg.gov/gtac/CourseDownloads/Reimbursables/FY21/Lidar_Material/GTAC_Guidelines%20to%20estimate%20forest%20inventory%20parameters%20from%20lidar%20and%20field%20plot%20data.pdf) (accessed on 3 March 2021).
33. Narine, L.L.; Popescu, S.; Neuenschwander, A.; Zhou, T.; Srinivasan, S.; Harbeck, K. Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sens. Environ.* **2019**, *224*, 1–11. [[CrossRef](#)]
34. Hogland, J.; Anderson, N.; St. Peter, J.; Drake, J.; Medley, P. Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 140. [[CrossRef](#)]
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
36. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Hall, D.E.; Falkowski, M.J. Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245. [[CrossRef](#)]
37. Kim, J.; Popescu, S.C.; Lopez, R.R.; Wu, X.B.; Silvy, N.J. Vegetation Mapping of No Name Key, Florida Using Lidar and Multispectral Remote Sensing. *Int. J. Remote Sens.* **2020**, *41*, 9469–9506. [[CrossRef](#)]
38. Silva, C.A.; Hudak, A.T.; Vierling, L.A.; Loudermilk, E.L.; O'Brien, J.J.; Hiers, J.K.; Jack, S.B.; Gonzalez-Benecke, C.; Lee, H.; Falkowski, M.J.; et al. Imputation of Individual Longleaf Pine (*Pinus Palustris* Mill.) Tree Attributes from Field and LiDAR Data. *Can. J. Remote Sens.* **2016**, *42*, 554–573. [[CrossRef](#)]
39. Hudak, A.T.; Evans, J.S.; Crookston, N.L.; Falkowski, M.J.; Steigers, B.K.; Taylor, R.; Hemingway, H. Aggregating Pixel-Level Basal Area Predictions Derived from LiDAR Data to Industrial Forest Stands in North-Central Idaho. In Proceedings of the Third Forest Vegetation Simulator Conference, Fort Collins, CO, USA, 13–15 February 2007; pp. 133–145.

40. Hudak, A.T.; Evans, J.S.; Falkowski, M.J.; Crookston, N.L.; Gessler, P.E.; Morgan, P.; Smith, A.M.S. Predicting Plot Basal Area and Tree Density in Mixed-Conifer Forest from Lidar and Advanced Land Imager (ALI) Data. In Proceedings of the 26th Canadian Symposium on Remote Sensing, Wolfville, NS, Canada, 14–16 June 2005; pp. 201–208.
41. Popescu, S.C.; Wynne, R.H.; Scrivani, J.A. Fusion of Small-Footprint Lidar and Multispectral Data to Estimate Plot-Level Volume and Biomass in Deciduous and Pine Forests in Virginia, USA. *For. Sci.* **2004**, *50*, 551–565. [[CrossRef](#)]
42. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth’s Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [[CrossRef](#)]
43. Smith, L.; Sidney, L. Ecoregions of Alabama and Georgia. Environment Protection. 1995, p. 75. Available online: [http://ecologicalregions.info/data/ga/alga\\_front.pdf](http://ecologicalregions.info/data/ga/alga_front.pdf) (accessed on 3 March 2021).
44. Omernik, J. Map Supplement Ecoregions of the Conterminous United. *Ann. Assoc. Am. Geogr.* **2014**, *77*, 118–125. [[CrossRef](#)]
45. Range, B.; Lawrence, S. Level III Ecoregions of Alaska. 2001; p. 175. Available online: [https://gaftp.epa.gov/EPADDataCommons/ORD/Ecoregions/ak/ak\\_eco.pdf](https://gaftp.epa.gov/EPADDataCommons/ORD/Ecoregions/ak/ak_eco.pdf) (accessed on 3 March 2021).
46. ESRI “World Imagery” [Basemap]. 2022. Available online: <https://pro.arcgis.com/en/pro-app/2.8/help/mapping/map-authoring/author-a-basemap.htm> (accessed on 4 April 2021).
47. Matney, T. TCruise. 2022. Available online: <https://landmarkspatialolutions.com/forest-inventory-software> (accessed on 24 February 2021).
48. United States Geological Survey (USGS); US Department of Interior 3D Elevation Program. Available online: <https://www.usgs.gov/core-science-systems/ngp/3dep> (accessed on 9 March 2021).
49. UGET. Available online: <https://ugetdm.com/> (accessed on 26 February 2021).
50. Drusch, M.; Gascon, F.; Berger, M. *Sentinel-2 Mission Requirements Document*; ESA: Paris, France, 2010.
51. Isenburg, M. LAStools: Award-Winning Software for Rapid LiDAR Processing. Available online: <http://lastools.org/> (accessed on 20 April 2021).
52. Mcgaughey, R.J. *Agriculture FUSION/LDV: Software for LIDAR Data Analysis and Visualization*; U.S. Department of Agriculture, Forest, Service, Pacific Northwest Research Station, University of Washington: Seattle, WA, USA, 2017.
53. ESRI ArcGIS Pro. Available online: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview> (accessed on 3 February 2021).
54. L3Harris Exelis Visual Information Solutions. ENVI. Available online: <https://www.l3harrisgeospatial.com/Software-Technology/ENVI> (accessed on 22 May 2021).
55. Dinstein, I.; Shanmugam, K.; Haralick, R.M. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *SMC-3*, 610–621.
56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2021.
57. Lumley, T. Package ‘Leaps’. 2020. Available online: <https://cran.r-project.org/web/packages/leaps/leaps.pdf> (accessed on 14 June 2021).
58. Breusch, T.S.; Pagan, A.R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* **1979**, *47*, 1287–1294. [[CrossRef](#)]
59. Sheridan, R.D.; Popescu, S.C.; Gatzliolis, D.; Morgan, C.L.S.; Ku, N.W. Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest. *Remote Sens.* **2015**, *7*, 229–255. [[CrossRef](#)]
60. Freeman, A.E.; Frescino, T.; Freeman, M.E. Package ‘ModelMap’. 2018. Available online: <https://cran.r-project> (accessed on 16 June 2021).
61. Gareth, J.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer Publishing Company, Incorporated: Berlin/Heidelberg, Germany, 2014; ISBN 978-1071614174.
62. Dorado-Roda, I.; Pascual, A.; Godinho, S.; Silva, C.A.; Botequim, B.; Rodríguez-Gonzálvez, P.; González-Ferreiro, E.; Guerra-Hernández, J. Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. *Remote Sens.* **2021**, *13*, 2279. [[CrossRef](#)]
63. Zhao, K.; Popescu, S. Lidar-Based Mapping of Leaf Area Index and Its Use for Validating GLOBCARBON Satellite LAI Product in a Temperate Forest of the Southern USA. *Remote Sens. Environ.* **2009**, *113*, 1628–1645. [[CrossRef](#)]
64. Van Etten, J.; Sumner, M.; Cheng, J.; Baston, D.; Bevan, A.; Bivand, R.; Busetto, L.; Canty, M.; Fasoli, B.; Forrest, D.; et al. Package ‘Raster’ R Topics Documented. 2022. Available online: <https://rspatial.org/raster> (accessed on 15 November 2021).
65. Gagliasso, D.; Hummel, S.; Temesgen, H. A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area. *Open J. For.* **2014**, *4*, 42–48. [[CrossRef](#)]
66. Bright, B.C.; Hudak, A.T.; Mcgaughey, R.; Andersen, H.E.; Negrón, J. Predicting Live and Dead Tree Basal Area of Bark Beetle Affected Forests from Discrete-Return Lidar. *Can. J. Remote Sens.* **2013**, *39*, S99–S111. [[CrossRef](#)]
67. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T. Temporal Transferability of LiDAR-Based Imputation of Forest Inventory Attributes. *Can. J. For. Res.* **2015**, *45*, 422–435. [[CrossRef](#)]