



Article

Estimating Forest Inventory Information for the Talladega National Forest Using Airborne Laser Scanning Systems

Taeyoon Lee ^{1,*} , Can Vatandaslar ^{1,2} , Krista Merry ¹ , Pete Bettinger ¹ , Alicia Peduzzi ¹ and Jonathan Stober ³

¹ Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA; can.vatandaslar@uga.edu (C.V.); kmerry@warnell.uga.edu (K.M.); pbettinger@warnell.uga.edu (P.B.); apeduzzi@uga.edu (A.P.)

² Faculty of Forestry, Artvin Coruh University, 08000 Artvin, Türkiye

³ U.S. Forest Service, Talladega National Forest, Heflin, AL 36264, USA; jonathan.stober@usda.gov

* Correspondence: taeyoon.lee@uga.edu

Abstract: Accurately assessing forest structure and maintaining up-to-date information about forest structure is crucial for various forest planning efforts, including the development of reliable forest plans and assessments of the sustainable management of natural resources. Field measurements traditionally applied to acquire forest inventory information (e.g., basal area, tree volume, and aboveground biomass) are labor intensive and time consuming. To address this limitation, remote sensing technology has been widely applied in modeling efforts to help estimate forest inventory information. Among various remotely sensed data, LiDAR can potentially help describe forest structure. This study was conducted to estimate and map forest inventory information across the Shoal Creek and Talladega Ranger Districts of the Talladega National Forest by employing ALS-derived data and aerial photography. The quality of the predictive models was evaluated to determine whether additional remotely sensed data can help improve forest structure estimates. Additionally, the quality of general predictive models was compared to that of species group models. This study confirms that quality level 2 LiDAR data were sufficient for developing adequate predictive models (R^2_{adj} , ranging between 0.71 and 0.82), when compared to the predictive models based on LiDAR and aerial imagery. Additionally, this study suggests that species group predictive models were of higher quality than general predictive models. Lastly, landscape level maps were created from the predictive models and these may be helpful to planners, forest managers, and landowners in their management efforts.

Keywords: aerial imagery; airborne laser scanning; ALASSO, forest inventory; LiDAR; mixed pine–hardwood forest



Citation: Lee, T.; Vatandaslar, C.; Merry, K.; Bettinger, P.; Peduzzi, A.; Stober, J. Estimating Forest Inventory Information for the Talladega National Forest Using Airborne Laser Scanning Systems. *Remote Sens.* **2024**, *16*, 2933. <https://doi.org/10.3390/rs16162933>

Academic Editor: Nikolay S. Strigul

Received: 4 June 2024

Revised: 6 August 2024

Accepted: 8 August 2024

Published: 10 August 2024



Copyright: © 2024 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/>).

1. Introduction

The ability to adequately characterize and assess forest structure with a high level of accuracy is not only important for the development of a reliable forest plan, but it is also informative for assessments that demonstrate the sustainable management of natural resources [1,2]. Within this context, an understanding of forest conditions, particularly the growing stock (tree and stand volume), aboveground biomass, and basal area, is crucial for planners, forest managers, and landowners. These attributes directly influence the potential revenue and the potential habitat a forest can provide and facilitate opportunities for addressing other management objectives [2,3]. Furthermore, a regularly updated forest inventory is essential for monitoring the spatiotemporal dynamics of forest ecosystems over the length of a planning horizon. For example, an estimate of aboveground biomass can provide insights into the capacity of a forest to sequester carbon, which is considered a critical factor in addressing climate change. This issue may become more important in the future as the forestry sector faces increasing pressure to assess the ability and rate of forests to sequester carbon [4–7]. Additionally, estimates of biomass can provide valuable

information for assessing forest health related to the outbreak of southern pine beetle (*Dendroctonus frontalis*) [8] and for assessing fire risk associated with fuel management [9].

Traditionally, an estimate of a forest inventory has heavily relied on labor-intensive and time-consuming field measurements. Timely field measurements may be limited in terms of their spatiotemporal coverage, may include sampling error, and may not be representative of large-forested areas [10]. While measuring the diameter at breast height (dbh) of trees, identifying the tree species, and counting trees within a sample unit (e.g., plot or prism point) are relatively straightforward methods, estimating the height of each tree may require more effort and include more uncertainty [11], especially in natural, mountainous forest landscapes [7]. Endeavors meant to obtain a sufficient number of well-distributed sampling plots that properly represent an entire forest area remain challenging, due to limited resources and accessibility. To account for these limitations, traditional field measurements have been complemented by products derived from remote sensing systems, which may help address spatial and temporal challenges in developing a forest inventory [3,12]. This approach is sometimes referred to as an enhanced forest inventory [13,14]. Within this framework, various remote sensing systems have been demonstrated to provide relatively accurate and cost-effective forest information, including the development of forest metrics [2,3,10,15,16] and greenhouse gas inventories [17].

Data acquired from Light Detection and Ranging (LiDAR) devices can assist in the development of forestry information. Unlike other remote sensing processes that only provide two-dimensional information, LiDAR can characterize a three-dimensional forest structure, to a certain extent, based on the point density of the LiDAR data [5,16,18]. Indeed, LiDAR can help derive height estimates of an object through the time interval between the emission of a pulse of energy by the LiDAR sensor and the moment that the reflected signal is returned to the device [19]. This type of active remote sensing technology has advanced significantly in the last ten years, resulting in a diverse range of LiDAR systems that can be installed in satellites, mounted on airplanes and unmanned aerial vehicles (UAVs), or used as hand-held devices. Each type of system has advantages and disadvantages. One advantage of using satellite-based laser scanning platforms, such as NASA's Global Ecosystem Dynamics Investigation (GEDI) and Ice, Cloud and Land Elevation Satellite 2 (ICESat-2), is that they can provide multi-temporal data for the entire Earth. For instance, Potapov et al. [20] were able to create a global canopy height model using GEDI data, along with Landsat imagery. Additionally, Dubayah et al. [21] presented an estimate of the mean biomass density for every country covered by the GEDI, with a 1 km resolution.

Terrestrial-based LiDAR platforms can be categorized as either static or mobile. Terrestrial laser scanners (TLSs) capture point clouds by looking upward from ground level, so they are advantageous in capturing the details of forest structure from under the canopy. Especially, static TLS platforms (consisting of a sensor, a tripod, and a GNSS receiver) produce the highest-quality point clouds among LiDAR systems [5,22]. Appropriately employed TLSs can facilitate visualization of tree branches and leaves; therefore, these systems have enormous potential for assisting in the development of new allometric biomass and wood quality relationships [23]. Arseniou et al. [24], for example, were able to estimate woody, aboveground biomass for urban and rural settings, using a TLS platform to identify various tree parts other than the main stem. Nevertheless, TLS platforms have several limitations when employed for forest inventory and operational forest management purposes. One limitation, an occlusion effect, is caused by features hidden or obstructed behind larger diameter trees within the point cloud data, and it poses a challenge when a fixed-position, single-scan approach is used [25]. In addition, the weight and size of a TLS platform can make the effort of moving between field measurement plots challenging. These limitations, along with the cost of the platform, may influence opinions on whether a TLS is a practical alternative for collecting forest inventory information [26,27].

Mobile laser scanning (MLS) platforms, or mobile TLS platforms, capture point clouds by looking horizontally from a height near ground level, where they are held, making them advantageous for capturing details on forest structure from a perspective very similar

to human-collected field measurements. MLS platforms can collect information while a person traverses the sampling plots with the instrument held in their hand or carried in a backpack. Moreover, the development of point clouds that are georeferenced to a local coordinate reference system, using simultaneous localization and mapping (SLAM) algorithms, reduces the need for GNSS. While numerous researchers have illustrated the applicability of MLS for diverse forest inventory tasks [7,10,28,29], these studies often focus on MLS data collected at the plot level. Among them, Vatandaşlar et al. [7] estimated several forest attributes (tree counts, dominant height, basal area, dbh, stand volume, and relative density) within plots located in a near-natural forest landscape. Employing a handheld MLS platform, Vatandaşlar et al. [7] mapped every stem within each field measurement plot and estimated these attributes with RMSEs ranging between 4.5% and 16.4%. However, as with TLS platforms, the number of sampling plots employed will likely influence the accuracy of forest or stand estimates [23]. As with any measurement system, the accuracy of forest attribute estimations can be positively correlated with the number of plots measured [30]. For instance, the diverse forests in the Shoal Creek and Talladega Ranger Districts in the southeastern US cover almost 93,694 ha [31] and, thus, the number of TLS or MLS plots needed to describe the forests' character relatively accurately may be substantial. Therefore, an alternative solution needs to be sought to effectively and efficiently characterize the forest inventory of large, diverse areas, such as this.

Airborne laser scanning (ALS) systems have been considered a suitable choice for helping to describe the forest character of broad areas. Notably, the ALS data acquisition process is not constrained by the accessibility restrictions related to TLS, including static and mobile platforms [5]. The versatility of an ALS system allows the collection of information across diverse temporal and spatial scales [32]. Recent advancements in sensor technologies allow the development of regularly updated and increasingly dense point clouds. For instance, a LiDAR-based forest inventory effort conducted in Ontario two decades ago, resulted in a point cloud dataset with about 0.5 points per m², yet today the development of a point cloud dataset above 40 points per m² can be obtained [33].

While forest attributes, such as tree heights and canopy coverage, can be directly estimated from ALS LiDAR data, other forest characteristics, such as aboveground biomass, growing stock (tree volume), and basal area, can be inferred from LiDAR-derived metrics [5]. The development of these estimates relies on modeling methods, which can range in complexity from regression to random forest models and other machine learning techniques [34]. Distinct models for estimating the characteristics of different forest types (conifer, broadleaved, and mixed) might also be developed, rather than a general model that is applicable to an entire forested area. Further research in this domain is necessary, with a specific emphasis on investigating the nature by which additional spectral data can enhance the predictive capability of LiDAR point clouds. Additionally, as mathematical techniques and remotely sensed data evolve, the most effective combination of methods and data sources needs to be assessed.

A map of forest characteristics for an extensive forest area is an ideal outcome of remote sensing methods. There are two methods for mapping forest attributes using ALS LiDAR data: area based and tree based. Although the tree-based method provides detailed information on individual trees [3,5], the area-based method has been applied for mapping forest attributes over wide areas, including diverse biomes [35]. However, achieving this outcome is complicated by two underlying factors: the multitude of potential independent variables that can be derived from remotely sensed data and the potential correlation amongst these, which can induce a multicollinearity problem [36–38]. Furthermore, a large number of independent variables within predictive models can challenge the application of these models for developing broad-scale GIS databases. Consequently, the selection of independent variables during the model development process is important. Tibshirani [39] suggested a method for developing linear models that estimate forest conditions, while enhancing prediction accuracy, by reducing the number of independent variables. Adhikari et al. [36] recommended the use of adaptive Least Absolute Shrinkage and Selection

Operator (ALASSO), as it effectively eliminated highly correlated independent variables from prediction models.

The main goal of this study is to develop models to estimate forest conditions across a broad area, from information provided by ALS and aerial imagery. This research effort seeks to: (i) evaluate the quality of predictive models developed using the ALASSO method, (ii) evaluate whether additional remotely sensed data (multispectral aerial imagery) can enhance the quality of predictive models that rely on LiDAR point cloud data, and (iii) determine the suitability of general versus species group-specific models for characterizing mixed coniferous and deciduous forests located in the southern United States. We use the Talladega Division (Talladega and Shoal Creek Ranger Districts) of the Talladega National Forest as our case study area, because it represents typical characteristics of natural, pine-dominated forests in southeastern USA. National forest managers do not have comprehensive inventories of the extent of the forests under management, limiting the global decision space for at risk resources, be it related to wildland fire, forest health, endangered species habitat conditions, to help assess where management is needed or has achieved the desired future conditions. Therefore models, maps, and other outcomes derived from predictive models that are based on LiDAR (and other) data, may provide forest managers, researchers, and policymakers with valuable insights to monitor and manage forests throughout the southeastern USA.

2. Materials and Methods

2.1. Study Area

This study was conducted within the boundaries of the Talladega and Shoal Creek Ranger Districts of the Talladega National Forest, located in Alabama, USA (Figure 1). This 93,694-ha part of the national forest lies within the Piedmont and Ridge and Valley ecoregions [31]. The climate in this area generally consists of mild winters and hot summers, which are characteristic of a humid, subtropical climate. Elevation of the land in the Talladega Division varies between approximately 160 m to 735 m above sea level, and annual precipitation is around 1260 mm. The forests in the study area are composed of coniferous species, such as longleaf pine (*Pinus palustris* Mill.), shortleaf pine (*P. echinata* Mill.), and loblolly pine (*P. taeda*), particularly in the uplands and on south-facing slopes. Deciduous tree species are often found in the riparian areas and on north-facing slopes. Oak (*Quercus* spp.), hickory (*Carya* spp.), maple (*Acer* spp.), and yellow poplar (*Liriodendron tulipifera*), are among the more prevalent deciduous tree species in the study area.

2.2. Data Collection

2.2.1. Field Data Collection

A set of 254 fixed area, circular plots (402.6 m² with a 11.32 m radius) were measured by U.S. Forest Service crews between February and April 2022. These field measurement plots were pseudo-randomly located within the operable (upland) lands of the study area. An equal number of field measurement plots were located within eight forest structure classes, defined by the canopy density and tree height from 2012 and 2013 quality level 2 LiDAR data acquisition. The field measurement plots were limited to the interior (rather than the edges) of the management units (stands), and only one plot per management unit was allowed. The center of the field measurement plot was mapped using a Trimble R1 GNSS receiver (Trimble, CO, USA). The field data collection procedure followed the methods described by Laes et al. [40]. Within each field measurement plot, several characteristics of each live tree (height, dbh, tree species, crown ratio, tree crown status (dominant and co-dominant), azimuth, and distance from the plot center) were recorded. A dbh larger than 7.6 cm (3 inches) and a height greater than 0.6 m (2 feet) represented the minimum sizes of the measured live trees.

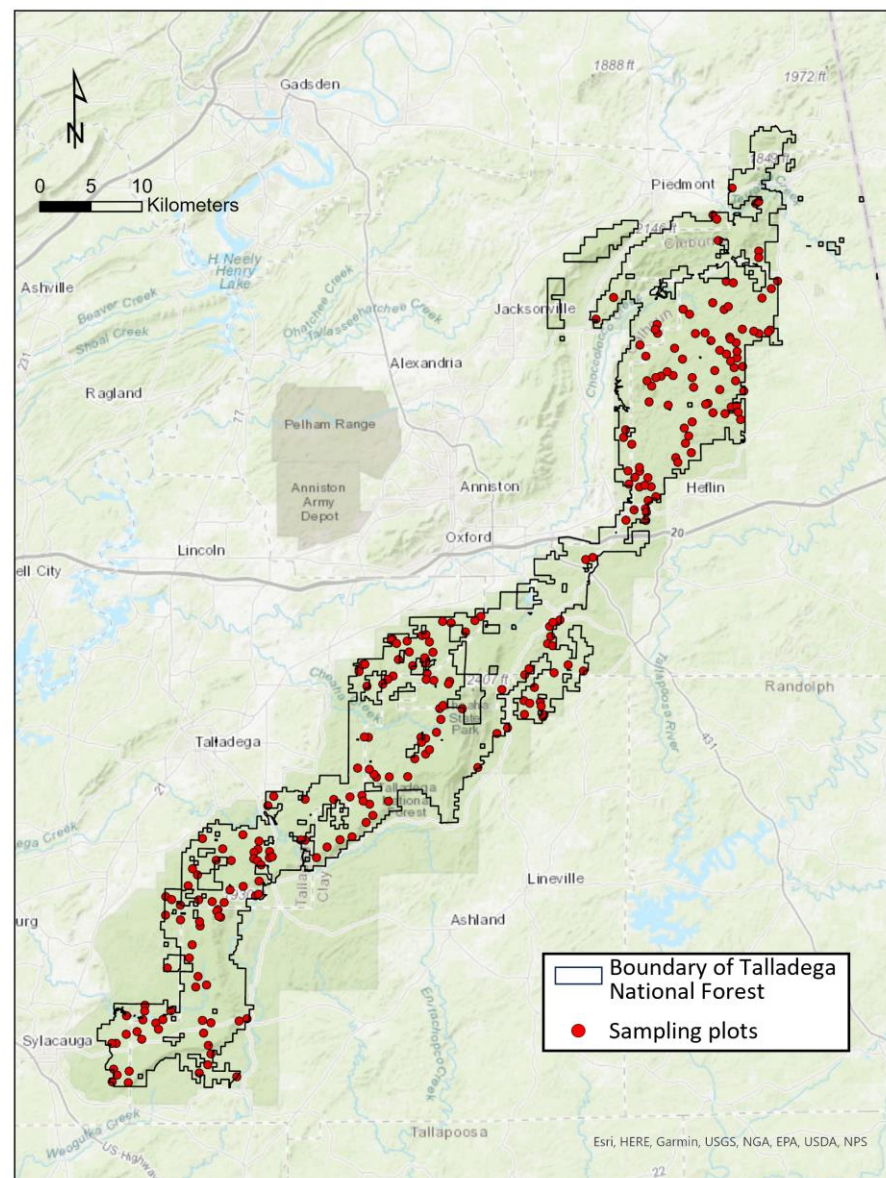


Figure 1. Location of the field measurement plots within the study area (Sources: Esri. “World Topographic Map” [basemap]. 31 January 2024).

2.2.2. Remote Data Collection

True-color and near-infrared images of five counties that cover the study area were captured during the leaf-off season between 2020 (December) and 2021 (February). These images were used to create county mosaic orthophotographs with a spatial resolution of 0.3 m, and they were acquired from the U.S. Department of Agriculture’s National Agricultural Imagery Program (NAIP) [41]. The ALS collection project was conducted by the USGS 3DDEP program, the Alabama Department of Transportation, and the U.S. Forest Service, with ALS data collected using a Leica Terrain Mapper during the leaf-off season between 2020 (December) and 2021 (February). Within the study area, the ALS data contained 13.92 billion data points. The average point density achieved was 9.7 points per m^2 and ranged from 3.5 points per m^2 to 20.3 points per m^2 . The horizontal and vertical accuracy of the LiDAR data (0.71 m and 0.051 m, respectively) was assessed using 441 survey points spread across the State of Alabama. These characteristics satisfy the requirements for, at least, topographic quality level 2 (QL2) [42].

2.3. Data Processing

2.3.1. Field Data Processing

From the field measurement plots, we calculated a variety of forest attributes related to tree heights (minimum, maximum, dominant, and mean merchantable height), tree dbh (minimum, maximum, quadratic mean of dbh, arithmetic mean of dbh, and coefficient of variation of dbh), canopy conditions (mean crown ratio), volume per unit area, and aboveground biomass per unit area. Among these, we focus on the basal area ($\text{m}^2 \text{ ha}^{-1}$), tree volume ($\text{m}^3 \text{ ha}^{-1}$), and aboveground biomass (metric tons per hectare, or Mg ha^{-1}), as dependent variables for the modeling effort. The basal area was estimated based on the dbh, and the tree volume and aboveground biomass were estimated using species-specific allometric equations from the U.S. Forest Service (National Biomass Estimator Library (NBEL) and National Volume Estimator Library (NVEL), which are Excel add-ins developed by the Forest Management Service Center, U.S. Forest Service). Additionally, we classified the field measurement plots by the dominant tree species present, based on the total basal area for a species in a given plot exceeding 70% of the total basal area. As a result, there were only 14 oak-dominant field measurement plots and 149 pine-dominated sampling plots (i.e., pine plots). All other plots indicated no dominance toward a specific tree species (i.e., mixed plots). Accordingly, we separately developed models based on two datasets: (a) using all plots ($n = 254$) and (b) using pine plots ($n = 149$).

2.3.2. Remote Data Processing

From the NAIP imagery, we created vegetation indices (i.e., greenness, Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI)) for the study area using ArcGIS Pro (Esri, Redlands, CA, USA). A total of 24 vegetation indices were developed and utilized as NAIP-derived independent variables (Table 1).

Table 1. Summary of vegetation indices calculated from the NAIP images.

Vegetation Index	Equation *	Calculated Statistics and Its Abbreviation
Greenness	$\frac{G}{R+G+B}$	Minimum of greenness (G_{MIN})
		Maximum of greenness (G_{MAX})
		Range of greenness (G_{RANGE})
		Mean of greenness (G_{MEAN})
		Standard deviation of greenness (G_{STD})
		Sum of greenness (G_{SUM})
		Median of greenness (G_{MEDIAN})
		Greenness at 90 percent (G_{PCT90})
Normalized Difference Vegetation Index, NDVI	$\frac{NIR-R}{NIR+R}$	Minimum of NDVI ($NDVI_{\text{MIN}}$)
		Maximum of NDVI ($NDVI_{\text{MAX}}$)
		Range of NDVI ($NDVI_{\text{RANGE}}$)
		Mean of NDVI ($NDVI_{\text{MEAN}}$)
		Standard deviation of NDVI ($NDVI_{\text{STD}}$)
		Sum of NDVI ($NDVI_{\text{SUM}}$)
		Median of NDVI ($NDVI_{\text{MEDIAN}}$)
		NDVI at 90 percent ($NDVI_{\text{PCT90}}$)
Enhanced Vegetation Index, EVI	$\frac{2.5 \times (NIR-R)}{(NIR+6 \times R-7.5 \times B+1)}$	Minimum of EVI (EVI_{MIN})
		Maximum of EVI (EVI_{MAX})
		Range of EVI (EVI_{RANGE})
		Mean of EVI (EVI_{MEAN})
		Standard deviation of EVI (EVI_{STD})
		Sum of EVI (EVI_{SUM})
		Median of EVI (EVI_{MEDIAN})
		EVI at 90 percent (EVI_{PCT90})

* where R , G , B , and NIR represent the raw pixel values of red, green, blue, and near-infrared bands.

Due to the large file size of the raw LiDAR point cloud data, processing could not be completed with a desktop personal computer. Thus, the LiDAR point cloud data were processed using RStudio (R version 2023.06.1 Build 524) and conducted within the University of Georgia's, Georgia Advanced Computing Resource Center, Sapelo2 Linux (64-bit CentOS 7.9) high-performance computing cluster. To reduce the processing time, parallel computing was used with 21 cores utilized simultaneously, with a total of 900 GB RAM. The raw LiDAR data were first utilized to create a digital terrain model, using only those points in the point cloud classified as bare ground. In creating the digital terrain model, selecting an appropriate algorithm and spatial resolution are important because it directly affects the result of the predicted vegetation metrics [18]. Because of this, the spatial resolution was chosen based on point spacing and the number of points in the point cloud, as suggested by McCullagh [43] and Hengl [44]. The triangular irregular network algorithm, a vector terrain model using Delaunay triangles, was applied to create a digital terrain model with a spatial resolution of 1 m. With the digital terrain model representing the ground, we then generated a normalized LiDAR point cloud, including only the aboveground points. The first returns in terms of the normalized point cloud were then used to create a canopy height model. Noise, resulting from points within the point cloud with values greater than 95% of the height (representing canopy height [2]), was removed to eliminate extreme values (i.e., returns from birds or communication antennas at heights much higher than the forest canopy). LiDAR-derived metrics were then calculated from point clouds that were clipped using the boundary of each field measurement plot using the R package ('lidR', version 4.0.3). We computed 56 LiDAR-derived metrics, based on the normalized LiDAR point clouds, for the 254 field measurement plots. Among the LiDAR-derived metrics, we excluded certain metrics with absolute values (such as the maximum intensity, mean intensity, area, point counts, and total intensity), assuming that the LiDAR sensor was not calibrated for light conditions before data collection, and to avoid developing a model with variables that were return density dependent. Therefore, a total of 74 independent variables were developed, including 50 LiDAR-derived metrics (Table 2) and 24 NAIP-derived metrics, which were used as independent variables in the modeling effort.

Table 2. Summary of the independent variables derived from LiDAR point cloud.

Metrics	Descriptions	Metrics	Descriptions
zmean	Mean height	zpcum x (from 1st to 9th)	Cumulative percentage of return in the ith layer
zsd	Standard deviation of height distribution	isd	Standard deviation of intensity
zskew	Skewness of height distribution	iskew	Skewness of intensity distribution
zkurt	Kurtosis of height distribution	ikurt	Kurtosis of intensity distribution
zentropy	Entropy of height distribution	ipground	Percentage of intensity returned by points classified as "ground"
pzabovemean	Percentage of returns above z mean	ipcumzq x (10th, 30th, 50th, 70th, and 90th)	Percentage of intensity returned below the xth percentile of the height
Pzabove2	Percentage of returns above 2 m	P xth (1, 2, 3, 4, and 5)	Percentage xth returns
zq x (From 5th to 95th)	Xth percentile (quantile) of height distribution	pground	Percentage of returns classified as "ground"

2.4. Modeling

We created 12 models to estimate three forest attributes (basal area, volume, above-ground biomass), with two different sets of field measurement plots (all plots, pine plots) and two types of data sources (LiDAR only, LiDAR + NAIP). Depending on which field measurement plots were employed, two sets of regression models were created: general models and pine models. Due to the high number of independent variables derived from the LiDAR and NAIP data sources, inter-correlation between the independent variables and the complexity of the models was inevitable [3]. To reduce the complexity of the models and to avoid multicollinearity issues, we applied the ALASSO regression. As the ALASSO regression is a regularization technique, it addresses multicollinearity issues by shrinking regression coefficients to zero based on the lambda value, which minimizes the sum of the squared differences between the predicted values and observed values. The best lambda was obtained by applying 10-fold cross-validation. Prior to applying the ALASSO regression, the normality of the data was assessed using the Box–Cox method. Given the results of the Box–Cox method, a natural logarithmic transformation was applied to the dependent variables, improving the linearity of the data, the homogeneity of the residual variances, and the normality of the residuals. Additionally, potential outliers were investigated using Cook’s distance and studentized residuals. The observed outliers without leverage were eliminated. The best models were selected based on various parameters, including the $R^2_{adj.}$, the number of independent variables, the root mean square error (RMSE), Colin Mallow’s C_p , Akaike’s information criterion (AIC), and the Bayesian information criterion (BIC). The models were back transformed by exponentiating the dependent variable and were applied to estimate the forest attributes for the broader study area. Lastly, 10-fold cross-validation was applied to evaluate the quality of the models. While the modeling procedures were conducted using RStudio (R version 2023.06.1 Build 524) on a desktop computer, the broad-scale estimation maps were processed using the Sapelo2 Linux-based high-performance computing cluster.

3. Results

3.1. Field-Based Forest Inventory

The average basal area, tree volume, and aboveground biomass of the 254 field measurement plots was about $23.4 \text{ m}^2 \text{ ha}^{-1}$, $180.9 \text{ m}^3 \text{ ha}^{-1}$, and 40.1 Mg ha^{-1} , respectively (Table 3). For the pine-dominated field measurement plots, the average basal area, tree volume, and aboveground biomass was about $22.3 \text{ m}^2 \text{ ha}^{-1}$, $169.0 \text{ m}^3 \text{ ha}^{-1}$, and 34.6 Mg ha^{-1} , respectively. The range of values for the field measurement plots was large, since the plots were meant to be representative of all the forest conditions in the study area, from early successional to mature forest stages.

Table 3. Descriptive statistics of forest attributes based on field measurements.

	Diameter at Breast Height (cm)	Basal Area ($\text{m}^2 \text{ ha}^{-1}$)	Volume ($\text{m}^3 \text{ ha}^{-1}$)	Aboveground Biomass (Mg ha^{-1})
All plots ($n = 254$)				
Average	22.39	23.43	180.85	40.13
Standard deviation	7.39	10.95	105.48	23.07
Minimum	8.65	0.33	0.77	0.13
Maximum	54.36	53.29	569.93	119.03
Pine plots ($n = 149$)				
Average	22.42	22.28	168.97	34.62
Standard deviation	8.40	16.92	108.26	21.94
Minimum	8.66	0.33	0.77	0.13
Maximum	54.36	51.19	543.27	110.02

3.2. Regression Models

3.2.1. Estimation of Forest Attributes Based on General Models

In nearly every case, the LiDAR + NAIP regression models included a larger number of independent variables than the LiDAR-only models. In the basal area model based on LiDAR + NAIP, five vegetation indices (G_{MIN} , $NDVI_{MIN}$, $NDVI_{MEDIAN}$, EVI_{MAX} , and EVI_{PCT90}) were additionally selected, instead of the $zq30$ variable (30th percentile height distribution from the ground) found in the basal area model based on LiDAR-only (Table 4). The R^2_{adj} values for the basal area models were 0.72 (cross-validation: 0.69) for LiDAR + NAIP and 0.71 (cross-validation: 0.71) for LiDAR only, while the RMSE values were $5.6 \text{ m}^2 \text{ ha}^{-1}$ (cross-validation: $5.90 \text{ m}^2 \text{ ha}^{-1}$) for LiDAR + NAIP and $5.7 \text{ m}^2 \text{ ha}^{-1}$ (cross-validation: $5.91 \text{ m}^2 \text{ ha}^{-1}$) for LiDAR only (Table 5). Other quality metrics, except AIC, also indicated that the model based on LiDAR + NAIP performed slightly better than the model based on LiDAR only. The basal area tended to be underestimated, as its value increased regardless of the data source (Figure 2).

Table 4. The best equations for the general (all plots, $n = 254$) and the pine (pine plots, $n = 149$) models, by forest attribute and data sources.

Forest Variables	Data Sources	Equation
General models		
Basal area	LiDAR + NAIP	$-1.828 + 0.017 \times pzabove2 + 0.015 \times zq25 + 0.017 \times zq95 - 7.83 \times 10^{-4} \times zpcum5 - 0.002 \times zpcum6 + 2 \times 10^{-5} \times isd + 0.24 \times iskw - 0.091 \times ikurt + 0.024 \times ipcumzq90 + 0.043 \times p2th + 0.116 \times G_{MIN} + 0.52 \times NDVI_{MIN} + 0.428 \times NDVI_{MEDIAN} + 3.58 \times 10^{-5} \times EVI_{MAX} - 0.012 \times EVI_{PCT90}$
	LiDAR	$-1.197 + 0.018 \times pzabove2 + 0.017 \times zq25 + 1.13 \times 10^{-4} \times zq30 + 0.015 \times zq95 - 3.79 \times 10^{-4} \times zpcum5 - 0.003 \times zpcum6 + 1.83 \times 10^{-5} \times isd + 0.196 \times iskw - 0.105 \times ikurt + 0.017 \times ipcumzq90 + 0.046 \times p2th$
Volume	LiDAR + NAIP	$-0.148 + 0.018 \times pzabove2 + 0.004 \times zq25 + 0.056 \times zq95 - 7.91 \times 10^{-4} \times zpcum5 - 0.00566 \times zpcum6 + 2.51 \times 10^{-5} \times isd + 0.237 \times iskw - 0.128 \times ikurt - 0.018 \times ipcumzq10 - 0.002 \times ipcumzq30 + 0.029 \times ipcumzq90 - 0.002 \times p1th + 0.029 \times p2th + 0.091 \times G_{MIN} - 0.165 \times G_{RANGE} + 0.522 \times NDVI_{MIN} + 0.229 \times NDVI_{MEAN} + 1.45 \times 10^{-5} \times NDVI_{SUM} + 0.158 \times NDVI_{MEDIAN} + 1.96 \times 10^{-4} \times EVI_{MAX} - 0.010 \times EVI_{PCT90}$
	LiDAR	$0.380 + 0.020 \times pzabove2 + 0.007 \times zq25 + 0.053 \times zq95 - 0.006 \times zpcum6 + 3.088 \times 10^{-5} \times isd + 0.209 \times iskw - 0.121 \times ikurt - 0.020 \times ipcumzq10 - 0.004 \times ipcumzq30 + 0.020 \times ipcumzq90 - 4.642 \times 10^{-5} \times p1th + 0.037 \times p2th$
Aboveground biomass	LiDAR + NAIP	$0.033 + 0.021 \times pzabove2 + 0.062 \times zq95 - 0.004 \times zpcum6 + 3.93 \times 10^{-5} \times isd + 0.132 \times iskw - 0.184 \times ikurt - 0.011 \times ipcumzq10 + 0.083 \times ipcumzq90 - 0.006 \times p1th + 0.031 \times p2th - 1.51 \times 10^{-4} \times pground + 0.385 \times G_{MIN} + 0.505 \times NDVI_{MIN} + 0.218 \times NDVI_{MEAN} + 9.36 \times 10^{-6} \times NDVI_{SUM} - 0.005 \times EVI_{PCT90}$
	LiDAR	$0.516 + 0.022 \times pzabove2 + 0.238 \times zq5 + 0.002 \times zq25 + 0.059 \times zq95 - 0.005 \times zpcum6 + 3.83 \times 10^{-5} \times isd + 0.103 \times iskw - 0.198 \times ikurt - 0.012 \times ipcumzq10 + 0.078 \times ipcumzq90 - 0.006 \times p1th + 0.034 \times p2th$
Pine models		
Basal area	LiDAR + NAIP	$0.657 + 0.009 \times pzabove2 + 3.600 \times zq5 + 0.021 \times zq25 + 0.001 \times zq40 + 0.007 \times zq95 - 0.009 \times zpcum5 + 0.173 \times iskw - 0.067 \times ikurt + 0.057 \times p2th + 0.426 \times NDVI_{MIN} + 0.985 \times NDVI_{MEDIAN} - 8.6 \times 10^{-4} \times EVI_{PCT90}$
	LiDAR	$5.195 + 0.011 \times pzabove2 + 2.873 \times zq5 + 0.021 \times zq25 + 1.681 \times 10^{-4} \times zq40 + 0.003 \times zq95 - 0.008 \times zpcum5 - 1.96 \times 10^{-4} \times zpcum6 + 0.168 \times iskw - 0.061 \times ikurt - 0.004 \times ipcumzq30 - 0.0475 \times ipcumzq90 + 0.057 \times p2th$
Volume	LiDAR + NAIP	$1.593 - 0.002 \times zkurt + 0.021 \times pzabove2 + 8.41 \times zq5 - 1.91 \times zq15 + 0.019 \times zq25 + 0.006 \times zq40 - 0.008 \times zq65 + 0.016 \times zq75 - 0.071 \times zq80 + 0.096 \times zq95 + 0.010 \times zpcum1 - 0.010 \times zpcum5 - 0.005 \times zpcum6 + 1.29 \times 10^{-5} \times zpcum8 + 0.002 \times zpcum9 + 3.04 \times 10^{-5} \times isd + 0.241 \times iskw - 0.16 \times ikurt + 0.016 \times ipcumzq10 - 0.008 \times ipcumzq30 + 0.042 \times p2th - 0.039 \times p5th + 0.486 \times G_{MIN} - 0.132 \times G_{RANGE} + 0.475 \times NDVI_{MIN} + 1.28 \times NDVI_{MEDIAN} + 2.7 \times 10^{-4} \times EVI_{MAX} - 0.004 \times EVI_{STD} - 0.019 \times EVI_{MEDIAN} - 0.015 \times EVI_{PCT90}$
	LiDAR	$3.592 + 0.008 \times pzabove2 + 3.401 \times zq5 + 0.004 \times zq25 + 0.043 \times zq95 - 0.013 \times zpcum5 + 0.222 \times iskw - 0.094 \times ikurt - 2.769 \times 10^{-4} \times ipcumzq10 - 0.033 \times ipcumzq30 + 0.047 \times p2th + 0.013 \times p3th$

Table 4. Cont.

Forest Variables	Data Sources	Equation
General models		
Aboveground biomass	LiDAR + NAIP	$6.466 - 0.005 \times \text{zkurt} + 0.41 \times \text{zentropy} + 0.025 \times \text{pzabove2} + 8.66 \times \text{zq5} - 0.037 \times \text{zq10} - 2.18 \times \text{zq15} + 0.017 \times \text{zq25} - 7.48 \times 10^{-4} \times \text{zq30} + 0.004 \times \text{zq40} - 0.001 \times \text{zq65} + 0.010 \times \text{zq75} - 0.079 \times \text{zq80} + 0.11 \times \text{zq95} + 0.013 \times \text{zpcum1} - 0.009 \times \text{zpcum5} - 0.005 \times \text{zpcum6} + 0.003 \times \text{zpcum9} + 4.75 \times 10^{-5} \times \text{isd} + 0.078 \times \text{iskew} - 0.175 \times \text{ikurt} + 0.020 \times \text{ipcumzq10} - 0.008 \times \text{ipcumzq90} + 0.042 \times \text{p2th} - 0.016 \times \text{p5th} + 0.673 \times G_{\text{MIN}} - 0.094 \times G_{\text{RANGE}} + 0.171 \times G_{\text{STD}} - 0.362 \times G_{\text{MEDIAN}} + 0.364 \times \text{NDVI}_{\text{MIN}} - 0.327 \times \text{NDVI}_{\text{STD}} + 1.51 \times \text{NDVI}_{\text{MEDIAN}} + 1.52 \times 10^{-4} \times \text{EVI}_{\text{MAX}} - 0.002 \times \text{EVI}_{\text{STD}} - 0.033 \times \text{EVI}_{\text{PCT90}}$
	LiDAR	$10.699 + 0.013 \times \text{pzabove2} + 2.062 \times \text{zq5} + 0.052 \times \text{zq95} - 0.011 \times \text{zpcum5} + 0.010 \times \text{iskew} - 0.129 \times \text{ikurt} - 0.010 \times \text{ipcumzq30} - 0.024 \times \text{ipcumzq90} - 0.009 \times \text{p1th} + 0.045 \times \text{p2th}$

Table 5. Summary of statistics for the general models.

Quality Metrics	Basal Area ($\text{m}^2 \text{ ha}^{-1}$)		Total Volume ($\text{m}^3 \text{ ha}^{-1}$)		Total Aboveground Biomass (Mg ha^{-1})	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
$R^2_{\text{adj.}}$	0.72	0.71	0.77	0.77	0.73	0.72
Number of variables	15	11	21	12	16	12
RMSE	5.58	5.73	48.44	49.34	11.68	11.84
R^2	0.74	0.72	0.79	0.78	0.74	0.74
Bias	−0.78	−0.80	−6.27	−6.54	−1.62	−1.64
Bias (%)	−3.33	−3.40	−3.45	−3.61	−4.03	−4.08
AIC	0.08	−76.05	−118.05	−137.76	−137.43	−145.17
BIC	−68.19	−38.28	−47.95	−96.66	−83.13	−104.02
CP	−17.21	0.08	0.11	0.12	0.12	0.13

In the volume models, a higher $R^2_{\text{adj.}}$ value (0.77) was observed (Table 5). Although the overall quality metrics for the LiDAR + NAIP and LiDAR-only models represented a similar level of performance, the increment in the number of independent variables in the LiDAR + NAIP models was noticeable, as it increased from 12 to 21 (Table 4). Specifically, eight independent variables derived from NAIP imagery were selected. Additionally, a LiDAR-derived metric, zpcum 5 (referring to the cumulative percentile of the return from the first layer to the fifth layer, based on LiDAR height measurements divided equally into 10 layers), was also selected as an independent variable in the LiDAR + NAIP model, but was not used in the LiDAR-only model. As observed in the basal area models, the difference between the observed volume and predicted volume tended to increase, with a higher volume regardless of the data sources (Figure 2).

The aboveground biomass models had an $R^2_{\text{adj.}}$ value of 0.73 and 0.72 (cross-validation: 0.64 and 0.65) for the LiDAR + NAIP and LiDAR-only models, respectively (Table 5). Similar to the basal area models, the LiDAR + NAIP model yielded slightly more accurate RMSE values than the LiDAR-only model (LiDAR + NAIP model: 11.7 Mg ha^{-1} (cross-validation: 13.05 Mg ha^{-1}); LiDAR-only model: 11.8 Mg ha^{-1} (cross-validation: 13.09 Mg ha^{-1})). Nevertheless, the difference between the aboveground biomass models depending on the data sources was slight. Five vegetation indices were added into the LiDAR + NAIP model as independent variables (Table 4). Both aboveground biomass models underestimated the biomass for the field measurement plots with a higher aboveground biomass, suggesting that estimation errors increase as the forest stands mature (Figure 2).

The average $R^2_{\text{adj.}}$ and average R^2 values of the 10-fold cross-validation results ranged from 0.64 to 0.73 and from 0.68 to 0.75, respectively (Table 6). The difference between the statistics of cross-validation and the developed models were generally smaller in the LiDAR-based models compared to the LiDAR + NAIP models. Regarding the selection of independent variables, several metrics were commonly selected regardless of the data

source, including: pzabove2, zq95, zpcum6, isd, iskw, ikurt, ipcumzq90, and p2th (all from LiDAR data); and G_{MIN} , $NDVI_{MIN}$, and EVI_{PCT90} (all from NAIP data) (Table 7).

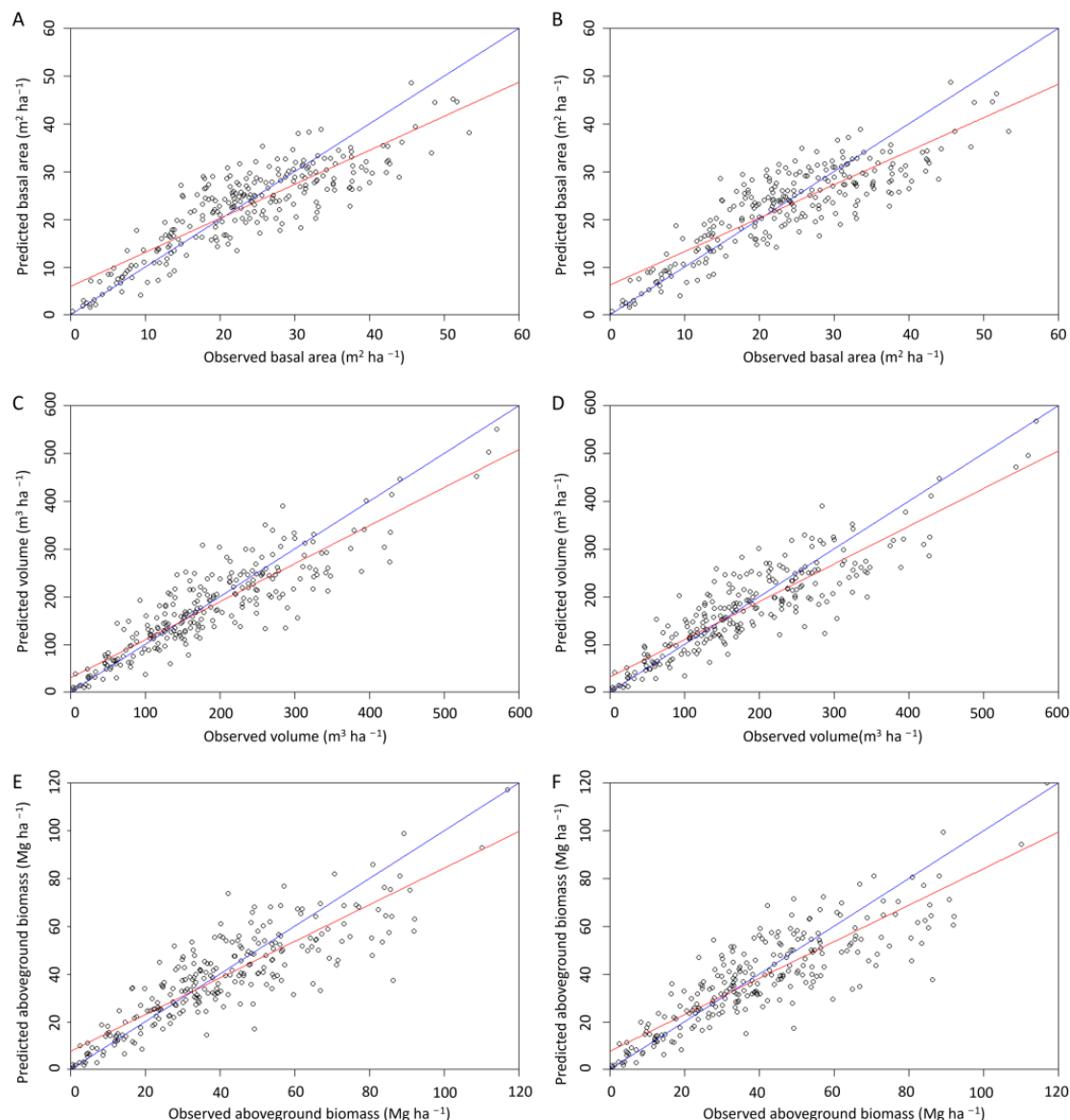


Figure 2. The observed forest attributes versus the predicted forest attributes of the general models ($n = 254$). (LiDAR + NAIP: (A,C,E); LiDAR: (B,D,F)). The blue line represents the 1:1 line (identity line). Red line represents the trend in the model. Dots represent the observed forest attributes.

Table 6. Analysis of 10-fold cross-validation for the general regression models.

Quality Metrics	Basal Area ($m^2 ha^{-1}$)		Total Volume ($m^3 ha^{-1}$)		Total Aboveground Biomass ($Mg ha^{-1}$)	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R^2_{adj}	0.69	0.71	0.67	0.73	0.64	0.65
R^2	0.72	0.69	0.72	0.75	0.69	0.68
RMSE	5.90	5.91	55.87	53.10	13.05	13.09
Bias	−0.76	−0.75	−4.06	−7.56	−1.12	−1.21
Bias (%)	−3.20	−3.20	−2.26	−4.12	−2.63	−2.85

Table 7. The most important independent variables in the best regression models.

	LiDAR Metrics	NAIP Metrics
General and pine model	pzabove2, zq95, iskew, ikurt, p2th	NDVI _{MIN} , EVI _{PCT90}
General model	zpcum6, isd, ipcumzq90	G _{MIN}
Pine model	zq5, zpcum5	NDVI _{MEDIAN}

3.2.2. Estimation of Forest Attributes Based on Pine Models

The R^2_{adj} and RMSE values of the basal area models were 0.81 and 0.80 (R^2_{adj} of cross-validation: 0.79 and 0.75), and $4.8 \text{ m}^2 \text{ ha}^{-1}$ and $5.1 \text{ m}^2 \text{ ha}^{-1}$ (RMSE of cross-validation: $5.21 \text{ m}^2 \text{ ha}^{-1}$ and $5.71 \text{ m}^2 \text{ ha}^{-1}$) for the LiDAR + NAIP and LiDAR-only models, respectively (Table 8). Noticeable improvements in the quality metrics of LiDAR + NAIP models were not observed by supplementing independent variables using NAIP-derived metrics. Interestingly, there was no increase in the number of independent variables, but three NAIP-derived metrics (NDVI_{MIN}, NDVI_{MEDIAN}, and EVI_{PCT90}) were selected instead of LiDAR-derived metrics (zpcum6, ipcumzq30, and ipcumzq90) (Table 4). The visual interpretation of the scatter plots shows that the basal area tended to be underestimated, as the basal area values increased regardless of the data sources (Figure 3). Similar trends in the scatter plots were also observed in the general models, but the distribution of the points was closer to the 1:1 line.

Table 8. Summary of statistics for the pine regression models.

Quality Metrics	Basal Area ($\text{m}^2 \text{ ha}^{-1}$)		Total Volume ($\text{m}^3 \text{ ha}^{-1}$)		Total Aboveground Biomass (Mg ha^{-1})	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R^2_{adj}	0.81	0.80	0.84	0.82	0.83	0.82
Number of variables	12	12	30	11	34	10
RMSE	4.80	5.10	37.86	43.45	7.89	8.93
R^2	0.83	0.81	0.87	0.84	0.87	0.83
Bias	−0.72	−0.75	−3.65	−6.71	−0.68	−1.36
Bias (%)	−3.23	−3.38	−2.12	−3.92	−1.93	−3.89
AIC	−5.84	−54.77	−56.60	−108.83	−49.23	−113.69
BIC	−22.09	−21.01	17.08	−77.80	31.20	−85.33
CP	0.07	0.08	0.06	0.11	0.06	0.10

Regarding the volume models, R^2_{adj} values of 0.84 (cross-validation: 0.78) for the LiDAR + NAIP models and 0.82 (cross-validation: 0.79) for the LiDAR-only models were achieved (Table 8). The RMSE values ranged from $37.9 \text{ m}^3 \text{ ha}^{-1}$ to $43.5 \text{ m}^3 \text{ ha}^{-1}$ (cross-validation: from $47.44 \text{ m}^3 \text{ ha}^{-1}$ to $48.94 \text{ m}^3 \text{ ha}^{-1}$), which were lower than the comparable values from the general regression models. However, the AIC and BIC of the general volume regression models were lower than those of the pine volume regression models. The increment in the number of independent variables in the LiDAR + NAIP models was notable, as it increased from 11 to 30 in the volume models. Specifically, eight independent variables derived from the NAIP images were added to the LiDAR + NAIP regression model (Table 4). Additionally, 12 independent variables derived from the LiDAR point clouds were added to the LiDAR + NAIP regression model.

The aboveground biomass models had R^2_{adj} values of 0.83 (cross-validation: 0.80) for the LiDAR + NAIP models and 0.82 (cross-validation: 0.78) for the LiDAR-only models (Table 8). The RMSE values of the aboveground biomass models were 7.9 Mg ha^{-1} and 8.9 Mg ha^{-1} (cross-validation: 9.4 Mg ha^{-1} and 10.2 Mg ha^{-1}) for the LiDAR + NAIP and LiDAR-only models, respectively. Regarding the AIC and BIC values, our results indicated that the quality of the LiDAR-only models was better than the LiDAR + NAIP models. The

number of independent variables also noticeably increased in the LiDAR + NAIP models (LiDAR + NAIP: 34; LiDAR-only: 10). Ten NAIP-derived metrics were selected, in addition to 16 LiDAR-derived metrics in the LiDAR + NAIP model (Table 4). Similar to the general regression models, both models (LiDAR + NAIP and LiDAR-only) underestimated the aboveground biomass for the field measurement plots, suggesting that estimation errors may increase as the stands mature (Figure 3).

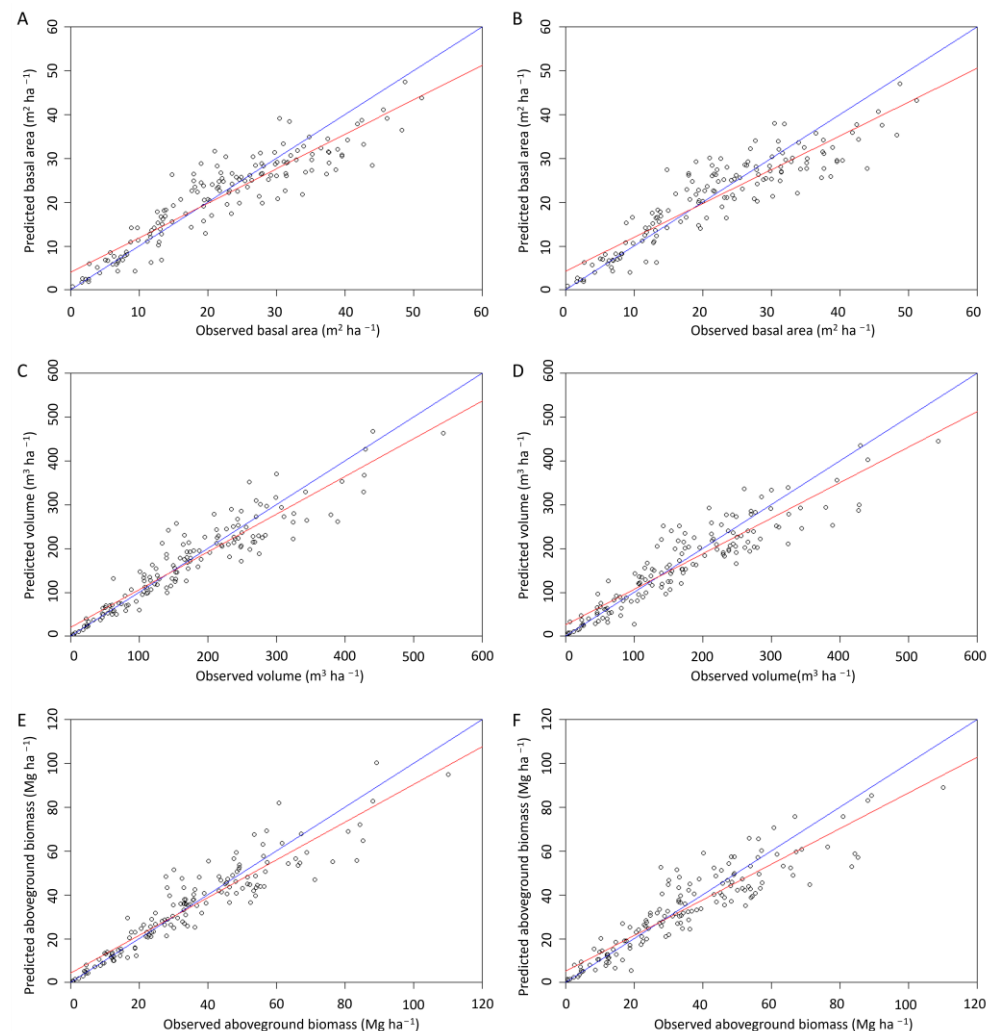


Figure 3. The observed forest attributes versus the predicted forest attributes of the pine models ($n = 149$). (LiDAR + NAIP: (A,C,E); LiDAR: (B,D,F)). The blue line represents the 1:1 line (identity line). Red line represents the trend in the model. Dots represent the observed forest attributes.

The average R^2_{adj} values of the 10-fold cross-validation results of the pine models ranged from 0.75 to 0.80, regardless of the forest attributes and data sources (Table 9). The average R^2 values of the 10-fold cross-validation results for the pine models ranged from 0.78 to 0.84. The developed models based on LiDAR + NAIP for the basal area and aboveground biomass were more robust than the models developed based on LiDAR only. Otherwise, the developed models based on LiDAR + NAIP for the volume were less robust than the models developed based on LiDAR only. Similar to the general models, there were some LiDAR and NAIP-derived metrics that were commonly selected as independent variables in every pine model regardless of the forest attributes: pzabove2, zq5, zq95, zpcum5, is skew, ikurt, and p2th (all from LiDAR data); and NDVI_{MIN}, EVI_{PCT90}, and NDVI_{MEDIAN} (all from NAIP data) (Table 7).

Table 9. Analysis of 10-fold cross-validation for the pine regression models.

Quality Metrics	Basal Area (m ² ha ^{−1})		Total Volume (m ³ ha ^{−1})		Total Aboveground Biomass (Mg ha ^{−1})	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R ² _{adj.}	0.79	0.75	0.78	0.79	0.80	0.78
R ²	0.82	0.78	0.82	0.81	0.84	0.81
RMSE	5.21	5.71	47.44	48.94	9.42	10.24
Bias	−0.72	−0.99	−4.41	−8.43	−0.97	−0.92
Bias (%)	−3.17	−4.33	−2.48	−4.92	−2.21	−2.71

4. Discussion

Estimating forest attributes, such as the basal area, volume, and aboveground biomass, and updating this information regularly, are critical activities for forest management and planning efforts. This study attempted to improve the performance of forest attribute estimations for large areas using LiDAR point clouds and high-resolution, multispectral, remotely sensed data. We investigated the effect of different combinations of remotely sensed data (LiDAR only or LiDAR + NAIP) on the quality of the regression models. Also, we evaluated the quality of regression models depending on the classification of field measurement plots according to the species mixture. To avoid the overfitting issue resulting from multicollinearity between independent variables, the ALASSO method was employed during the modeling process, as this method was suggested by earlier studies [36,45]. A total of 12 models were developed for three forest attributes (basal area, volume, and aboveground biomass), based on the species mixture (all species and pine), as well as the data source (LiDAR only, LiDAR + NAIP). When the general models are applied to the case study landscape, broad-scale maps of the resource conditions can be visualized (Figures 4–6).

The R²_{adj.} values of the developed models ranged from 0.71 to 0.84, which were comparable to other ALS study results [3,34,46]. While many researchers have developed regression models for estimating forest attributes over relatively small-forested areas [10,15,19,34,37,38], there have been few attempts to work in large areas, such as U.S. National Forests. For instance, Leboeuf et al. [46] mapped a merchantable wood volume of a very large area (440,000 km²) using airborne LiDAR. However, this study had a few limitations, including significant temporal differences between the field measurements (2003–2018) and the LiDAR data (2011–2020) collection periods, a limited spatial resolution, and poor representativeness of the field measurement plots. In our study, however, field and remote sensing datasets were collected during a relatively similar period of time (2020–2022). The distribution of field measurement plots is also important in developing robust regression models for forest attribute estimation. To enhance the quality of the regression models, we used a stratified pseudo-random sampling design, by classifying the entire operable study area, while considering recent management activities to improve the balance and representation of the heterogeneity of the forest in the sampling design. Additionally, we surveyed all tree species (both merchantable and non-merchantable >7.62 cm in dbh) within the field measurement plots to enhance the accuracy of the regression models, as suggested by Brown et al. [47].

The R²_{adj.} values and R² values for the pine models were higher than those of the general models, regardless of the forest attributes. The overall quality metrics also indicated that the pine models were higher in terms of quality compared to the general models. Bouvier et al. [3] also confirmed that separate models may result in higher accuracy when compared to general models. Regarding forest attributes, the highest R²_{adj.} values were observed in tree volume models and the lowest R²_{adj.} values were observed in basal area models, regardless of the data sources and field measurement plots (all plots or pine plots). This trend has also been observed in other studies using ALS systems [3,34,48,49]. Sumnall

et al. [50] suggested that basal area models have relatively lower R^2 values than models estimating tree height and biomass. This is likely because, unlike tree height, dbh cannot be directly measured with ALS systems. Dbh is used to estimate the basal area, yet dbh and tree height are often needed to estimate the volume and aboveground biomass [3], therefore it may be logical to presume that the basal area estimation models will underperform when compared to the volume and biomass models.

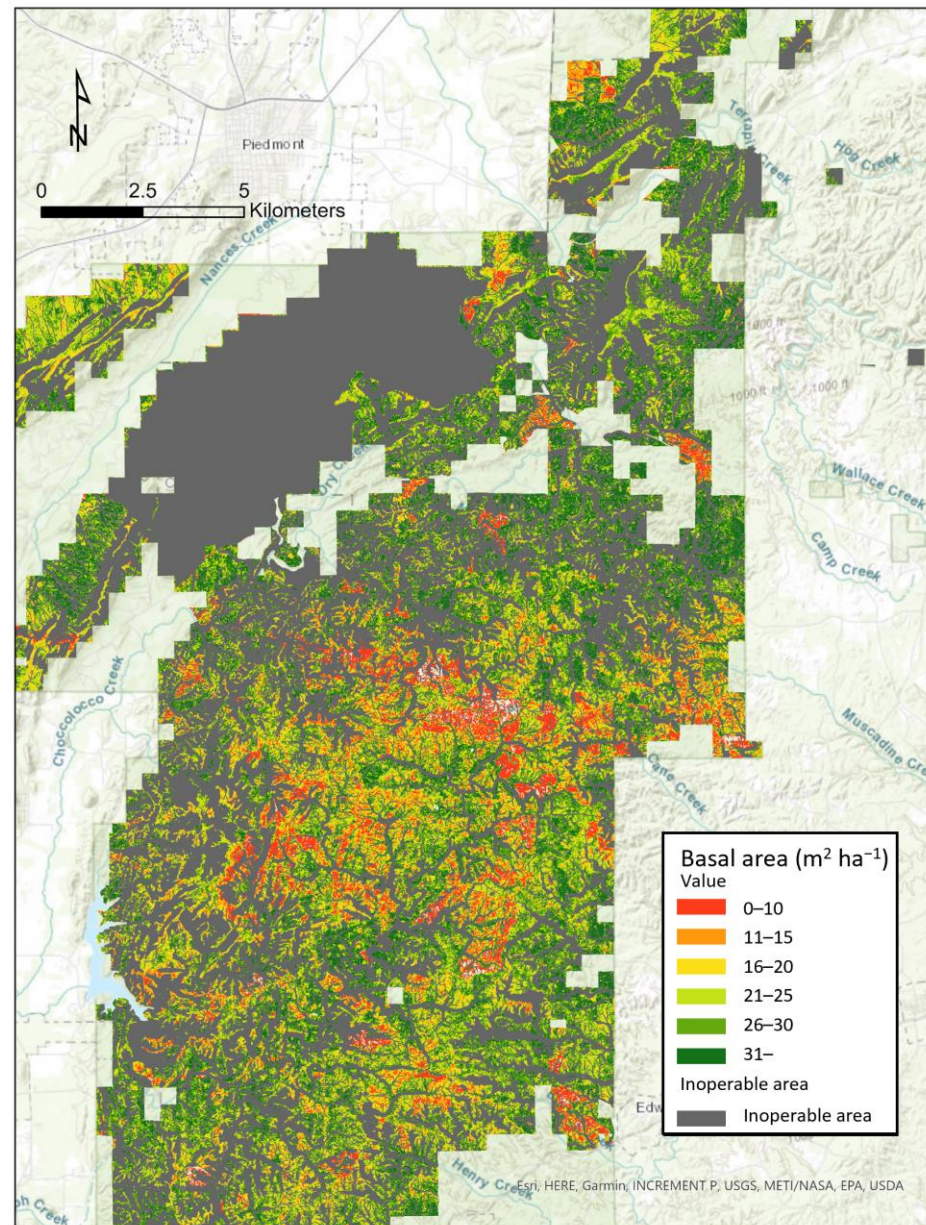


Figure 4. Predicted basal area for a portion of the study area based on the general model (Sources: Esri. “World Topographic Map” [basemap]. 31 January 2024).

There have been many attempts recently to enhance the performance of regression models for estimating forest conditions. They generally involve the development of LiDAR metrics [3,51] and, perhaps, supplementary remotely sensed data [34,47]. In this study, the vegetation indices derived from NAIP imagery were included as Supplemental Data. As the additional NAIP imagery had a higher spatial resolution (0.3 m), we expected the vegetation indices might improve the quality of the regression models to a considerable extent. However, it was observed that the addition of NAIP-derived vegetation metrics was not very influential in improving the quality of the prediction models. Although

LiDAR-only derived models had slightly larger RMSE values and slightly lower R^2 values compared to the LiDAR + NAIP-derived models, the increase in the number of independent variables with the addition of NAIP metrics led us to conclude that the LiDAR-only models were more appropriate for broad-scale mapping efforts.

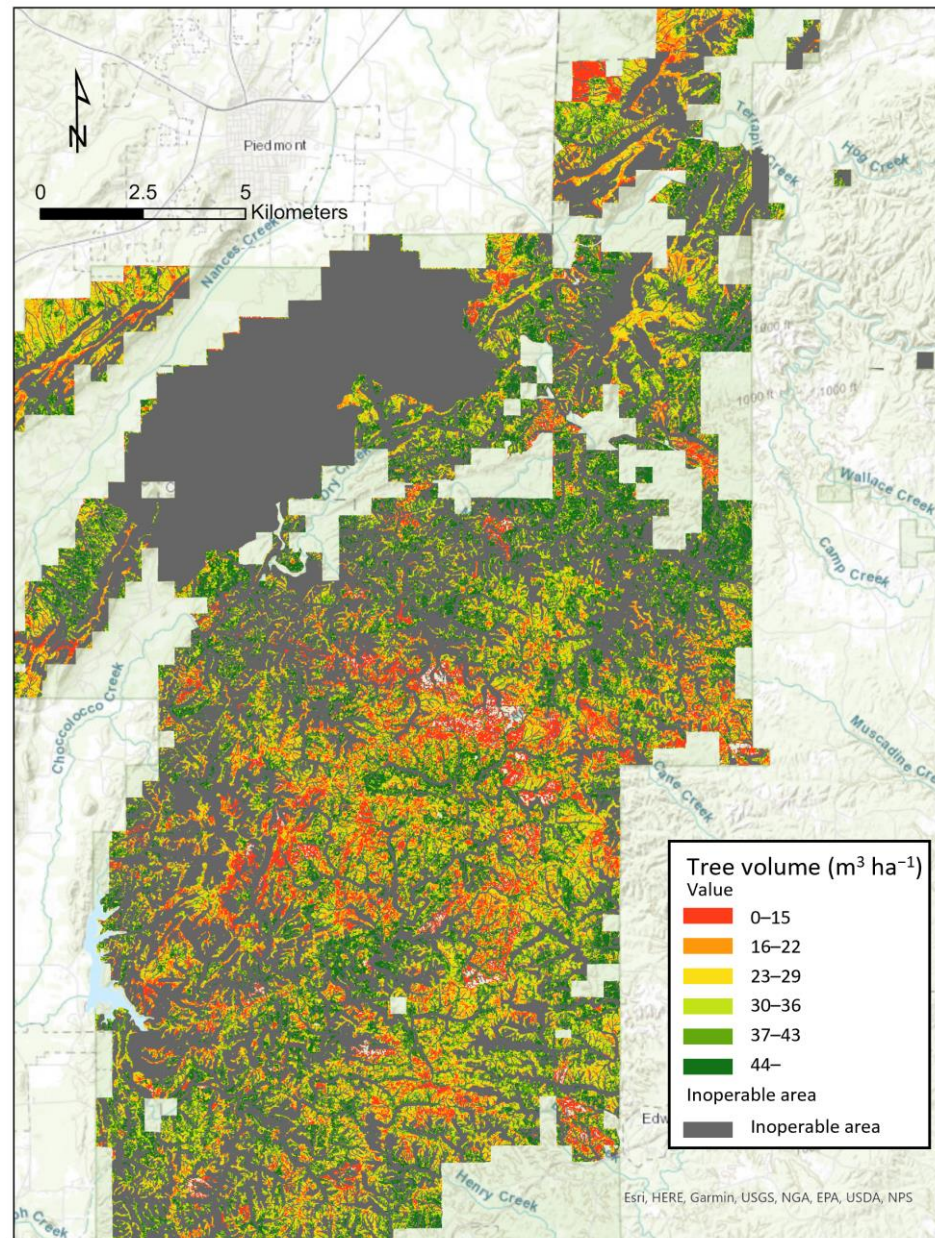


Figure 5. Predicted volume per hectare for a portion of the study area based on the general model (Sources: Esri. “World Topographic Map” [basemap]. 31 January 2024).

Of the 74 metrics derived from the LiDAR and NAIP data sources, 45 were selected as independent variables in at least one regression model. Further, five LiDAR-based metrics and two NAIP-based metrics were selected for every regression model (Table 7). These were, specifically, the LiDAR-based metrics, pzabove2, zq95, iskew, ikurt, and p2th, and the NAIP-based metrics, NDVI_{MIN} and EVI_{PCT90}. LiDAR-based metrics, pzabove2 and p2th, help eliminate the effect of understory vegetation that is not measured in a typical forest inventory survey. The LiDAR-based metric zq95 is widely used to represent forest canopy height. In addition to height-related LiDAR metrics, intensity-related metrics, such as iskew and ikurt, were also crucial in developing the regression models, as they provide

information related to stand density. The inclusion of NAIP-based metrics, $NDVI_{MIN}$ and EVI_{PCT90} , can be explained by the concentration of active chlorophyll in pine tree crowns. Ozkan et al. [34] also confirmed that metrics such as these may be significantly correlated with forest attributes.

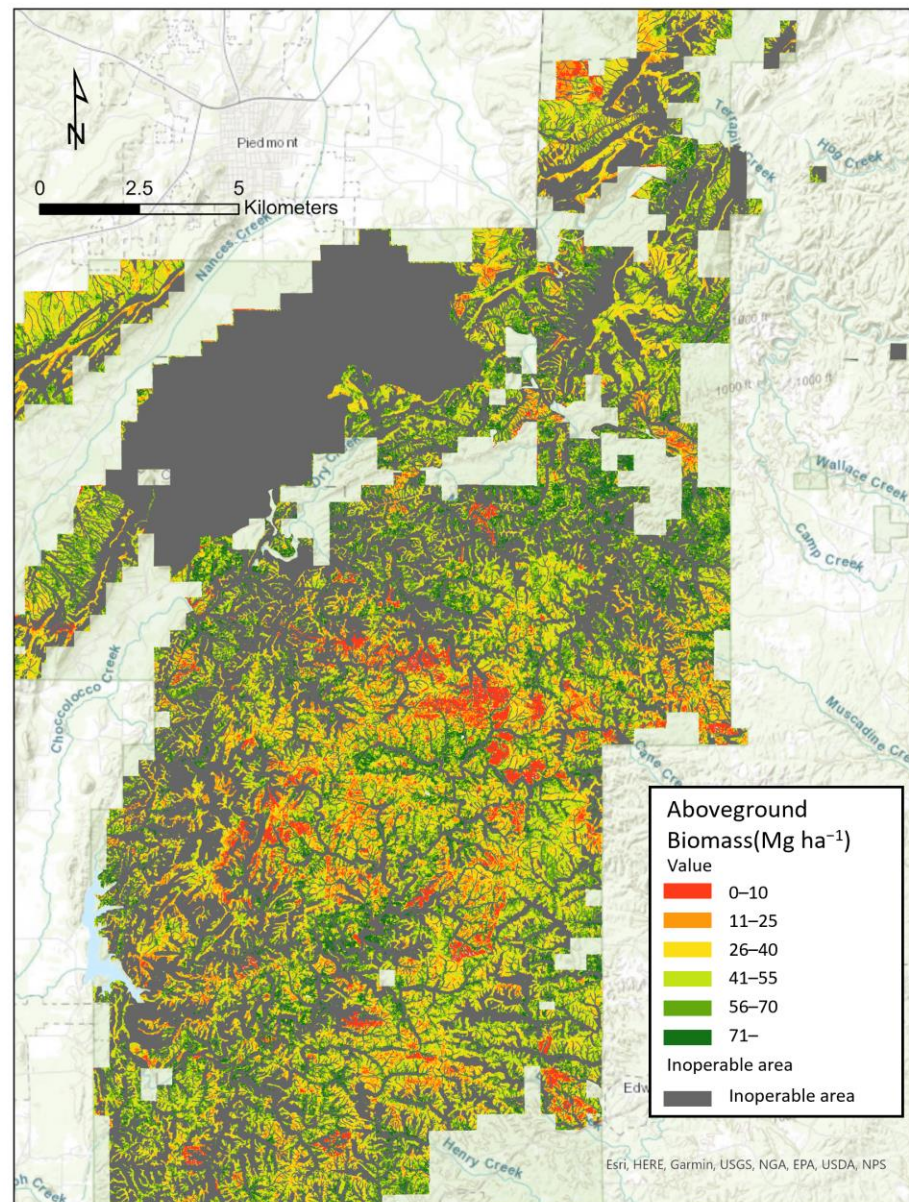


Figure 6. Predicted aboveground biomass for a portion of the study area based on the general model (Sources: Esri. “World Topographic Map” [basemap]. 31 January 2024).

Although the inclusion of NAIP-based metrics improved the accuracy of the regression models we developed, we suggest that LiDAR-derived metrics may be sufficient for developing robust regression models used for operational forest management purposes. Interestingly, the general trend in the spatial pattern of the basal area, volume, and aboveground biomass is due to the high correlation among these forest attributes (Figures 4–6). Since the aboveground biomass of a forest is typically a derivation of growing stock, volume estimates often rely on dbh measurements, and as the basal area is directly related to dbh, these relationships may help in developing simple ratios between forest conditions without the need for additional, elaborate regression models. While this may potentially decrease the reliability of some of the outcomes, the practicality of the overall effort may be worth investigating further. Furthermore, we recognize the limitations inherited in the

present study. For example, the location of the field measurement plots might have caused sampling bias as they are located only in an ‘operable area’. Here, we define operable areas as forestlands where forest management activities might be planned and implemented. Therefore, our models and maps may not be representative of an ‘inoperable area’ where forest management activities are unlikely to be planned due to steep slopes and flatter riparian areas where streams are located, and most of the trees are deciduous. To avoid poor estimates, we eliminated the inoperable areas from our maps and provided estimation results within operable areas only. The relatively low point density of the ALS data used can be considered another limitation of the study. The underestimated results, particularly in dense stands with high aboveground biomass values, suggest that the point cloud density is insufficient to characterize each tree in the canopy height model. This limitation can also be attributed to the few returns from under the canopy, hindering the detection of suppressed trees and undergrowth vegetation as well.

5. Conclusions

The objective of this study was to estimate and map key forest attributes across a wide area of interest, utilizing ALS data and aerial imagery. To achieve this objective, we employed the ALASSO modeling method, relying on 254 field measurement plots from a national forest in Alabama (USA), a mostly natural ecosystem composed of coniferous and deciduous tree species. One of the main conclusions drawn from our findings was that the LiDAR data collected by the ALS system, with a topographic quality level QL2 character, seems to be a sufficient input for the development of regression models that can estimate the basal area, volume, and aboveground biomass at accuracy levels that are acceptable for operational forest inventories. A second conclusion was that the added value of using optical data (aerial imagery) and associated vegetation indices as inputs for the development of regression models for estimating the basal area, volume, and aboveground biomass was negligible, considering the increased model complexity and extra time required to process and analyze the additional model inputs. A third conclusion was that the models exclusively developed for areas dominated by pine species seem to outperform (with respect to the coefficient of determination from cross-validation) general models that were developed for all tree species within the study area. While we affirmed these conclusions for natural, pine-dominated forests in the study area, further research is needed to assess the scalability of the results to other forest ecosystems in the southern USA.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16162933/s1>.

Author Contributions: Conceptualization, T.L., C.V. and P.B.; methodology, A.P. and T.L.; validation, T.L. and C.V.; formal analysis, T.L.; data curation, T.L. and J.S.; writing—original draft preparation, T.L. and C.V.; writing—review and editing, T.L., C.V., P.B., K.M., J.S. and A.P.; visualization, T.L. and K.M.; supervision, P.B.; project administration, P.B.; funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Talladega Division LiDAR Project 2022-Part 2, US Forest Service Agreement No. 22-CS-11080100-232, and in part by Promoting Economic Resilience and Sustainability of the Eastern US Forests (PERSEUS), U.S. Department of Agriculture Award No. 2023-68012-38992.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Bettinger, P.; Boston, K.; Siry, J.P.; Grebner, D.L. *Forest Management and Planning*; Academic Press: London, UK, 2017.
2. Jarron, L.R.; Coops, N.C.; MacKenzie, W.H.; Tompalski, P.; Dykstra, P. Detection of sub-canopy forest structure using airborne LiDAR. *Remote Sens. Environ.* **2020**, *244*, 111770. [[CrossRef](#)]

3. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.P. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sens. Environ.* **2015**, *156*, 322–334. [\[CrossRef\]](#)
4. Chen, Y.; Kershaw, J.A.; Hsu, Y.H.; Yang, T.R. Carbon estimation using sampling to correct LiDAR-assisted enhanced forest inventory estimates. *For. Chron.* **2020**, *96*, 9–19. [\[CrossRef\]](#)
5. Du, L.; Pang, Y.; Wang, Q.; Huang, C.; Bai, Y.; Chen, D.; Lu, W.; Kong, D. A LiDAR biomass index-based approach for tree-and plot-level biomass mapping over forest farms using 3D point clouds. *Remote Sens. Environ.* **2023**, *290*, 113543. [\[CrossRef\]](#)
6. Tian, L.; Wu, X.; Tao, Y.; Li, M.; Qian, C.; Liao, L.; Fu, W. Review of remote sensing-based methods for forest aboveground biomass estimation: Progress, challenges, and prospects. *Forests* **2023**, *14*, 1086. [\[CrossRef\]](#)
7. Vatandaşlar, C.; Seki, M.; Zeybek, M. Assessing the potential of mobile laser scanning for stand-level forest inventories in near-natural forests. *For. Int. J. For. Res.* **2023**, *96*, 448–464. [\[CrossRef\]](#)
8. Nowak, J.T.; Meeker, J.R.; Coyle, D.R.; Steiner, C.A.; Brownie, C. Southern pine beetle infestations in relation to forest stand conditions, previous thinning, and prescribed burning: Evaluation of the southern pine beetle prevention program. *J. For.* **2015**, *113*, 454–462. [\[CrossRef\]](#)
9. Ager, A.A.; Vaillant, N.M.; Finney, M.A. Integrating fire behavior models and geospatial analysis for wildland fire risk assessment and fuel management planning. *J. Combust.* **2011**, *2011*, 572452. [\[CrossRef\]](#)
10. Adhikari, A.; Peduzzi, A.; Montes, C.R.; Osborne, N.; Mishra, D.R. Assessment of understory vegetation in a plantation forest of the southeastern United States using terrestrial laser scanning. *Ecol. Inform.* **2023**, *77*, 102254. [\[CrossRef\]](#)
11. Seki, M.; Sakici, O.E. Ecoregion-based height-diameter models for Crimean pine. *J. For. Res.* **2022**, *27*, 36–44. [\[CrossRef\]](#)
12. García, M.; Riaño, D.; Chuvieco, E.; Danson, F.M. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sens. Environ.* **2010**, *114*, 816–830. [\[CrossRef\]](#)
13. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing technologies for enhancing forest inventories: A review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [\[CrossRef\]](#)
14. White, J.C.; Tompalski, P.; Vastaranta, M.; Wulder, M.A.; Saarinen, N.; Stepper, C.; Coops, N.C. *A Model Development and Application Guide for Generating an Enhanced Forest Inventory Using Airborne Laser Scanning Data and an Area Based Approach*; Canadian Wood Fibre Centre: Victoria, BC, Canada, 2017; Information Report FI-X-018; ISBN 978-0-660-09738-1.
15. Ferraz, A.; Saatchi, S.; Mallet, C.; Meyer, V. Lidar detection of individual tree size in tropical forests. *Remote Sens. Environ.* **2016**, *183*, 318–333. [\[CrossRef\]](#)
16. Lim, K.; Treitz, P.; Wulder, M.A.; St-Onge, B.; Flood, M. LiDAR remote sensing of forest structure. *Prog. Phys. Geog.* **2003**, *27*, 88–106. [\[CrossRef\]](#)
17. McRoberts, R.E.; Næsset, E.; Sannier, C.; Stehman, S.V.; Tomppo, E.O. Remote sensing support for the gain-loss approach for greenhouse gas inventories. *Remote Sens.* **2020**, *12*, 1891. [\[CrossRef\]](#)
18. Bater, C.W.; Coops, N.C. Evaluating error associated with Lidar-derived DEM interpolation. *Comput. Geosci.* **2009**, *35*, 289–300. [\[CrossRef\]](#)
19. García-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing* **2015**, *167*, 24–31. [\[CrossRef\]](#)
20. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **2021**, *253*, 112165. [\[CrossRef\]](#)
21. Dubayah, R.; Armston, J.; Healey, S.P.; Bruening, J.M.; Patterson, P.L.; Kellner, J.R.; Duncanson, L.; Saarela, S.; Ståhl, G.; Yang, Z.; et al. GEDI launches a new era of biomass inference from space. *Environ. Res. Lett.* **2022**, *17*, 095001. [\[CrossRef\]](#)
22. Liang, X.; Kankare, V.; Hyypä, J.; Wang, Y.; Kukko, A.; Haggrén, H.; Yu, X.; Kaartinen, H.; Jaakkola, A.; Guan, F.; et al. Terrestrial laser scanning in forest inventories. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 63–77. [\[CrossRef\]](#)
23. Liang, X.; Kukko, A.; Balenović, I.; Saarinen, N.; Junttila, S.; Kankare, V.; Holopainen, M.; Mokroš, M.; Surový, P.; Kaartinen, H.; et al. Close-range remote sensing of forests: The state of the art, challenges, and opportunities for systems and data acquisitions. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 32–71. [\[CrossRef\]](#)
24. Arseniou, G.; MacFarlane, D.W.; Calders, K.; Baker, M. Accuracy differences in aboveground woody biomass estimation with terrestrial laser scanning for trees in urban and rural forests and different leaf conditions. *Trees* **2023**, *37*, 761–779. [\[CrossRef\]](#)
25. Mathes, T.; Seidel, D.; Häberle, K.H.; Pretzsch, H.; Annighöfer, P. What are we missing? Occlusion in laser scanning point clouds and its impact on the detection of single-tree morphologies and stand structural variables. *Remote Sens.* **2023**, *15*, 450. [\[CrossRef\]](#)
26. Chen, S.; Liu, H.; Feng, Z.; Shen, C.; Chen, P. Applicability of personal laser scanning in forestry inventory. *PLoS ONE* **2019**, *14*, e0211392. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Ryding, J.; Williams, E.; Smith, M.; Eichhorn, M. Assessing handheld mobile laser scanners for forest surveys. *Remote Sens.* **2015**, *7*, 1095–1111. [\[CrossRef\]](#)
28. Gollob, C.; Ritter, T.; Nothdurft, A. Forest inventory with long range and high-speed personal laser scanning (PLS) and simultaneous localization and mapping (SLAM) technology. *Remote Sens.* **2020**, *12*, 1509. [\[CrossRef\]](#)
29. Mokroš, M.; Mikita, T.; Singh, A.; Tomastík, J.; Chudá, J.; Wezyk, P.; Kuželka, K.; Surový, P.; Klimánek, M.; Zięba-Kulawik, K.; et al. Novel low-cost mobile mapping systems for forest inventories as terrestrial laser scanning alternatives. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102512. [\[CrossRef\]](#)

30. Li, C.; Yu, Z.; Dai, H.; Zhou, X.; Zhou, M. Effect of sample size on the estimation of forest inventory attributes using airborne LiDAR data in large-scale subtropical areas. *Ann. For. Sci.* **2023**, *80*, 40. [CrossRef]
31. Stober, J.; Merry, K.; Bettinger, P. Analysis of fire frequency on the Talladega National Forest, USA, 1998–2018. *Int. J. Wildland Fire*. **2020**, *29*, 919–925. [CrossRef]
32. Yao, H.; Qin, R.; Chen, X. Unmanned aerial vehicle for remote sensing applications—A review. *Remote Sens.* **2019**, *11*, 1443. [CrossRef]
33. Penner, M.; Woods, M.; Bilyk, A. Assessing site productivity via remote sensing—Age-independent site index estimation in even-aged forests. *Forests* **2023**, *14*, 1541. [CrossRef]
34. Ozkan, U.Y.; Demirel, T.; Ozdemir, I.; Saglam, S.; Mert, A. Predicting forest stand attributes using the integration of airborne laser scanning and Worldview-3 data in a mixed forest in Turkey. *Adv. Space Res.* **2022**, *69*, 1146–1158. [CrossRef]
35. 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]
36. Adhikari, A.; Montes, C.R.; Peduzzi, A. A comparison of modeling methods for predicting forest attributes using Lidar metrics. *Remote Sens.* **2023**, *15*, 1284. [CrossRef]
37. Chen, X.; Xie, D.; Zhang, Z.; Sharma, R.P.; Chen, Q.; Liu, Q.; Fu, L. Compatible biomass model with measurement error using airborne LiDAR data. *Remote Sens.* **2023**, *15*, 3546. [CrossRef]
38. Næsset, E. Predicting Forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [CrossRef]
39. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *73*, 267–288. [CrossRef]
40. Laes, D.; Reutebuch, S.E.; McGaughey, R.J.; Mitchell, B. Guidelines to Estimate Forest Inventory Parameters from Lidar and Field Plot Data. 2011. 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 (accessed on 3 June 2024).
41. U.S. Department of Agriculture. *National Agriculture Imagery Program (NAIP)*; U.S. Department of Agriculture: Washington, DC, USA, 2023. Available online: <https://naip-usdaonline.hub.arcgis.com/> (accessed on 2 May 2024).
42. U.S. Geologic Survey. *Topographic Data Quality Levels (QLs)*; U.S. Geologic Survey: Reston, VA, USA, 2024. Available online: <https://www.usgs.gov/3d-elevation-program/topographic-data-quality-levels-qls> (accessed on 22 April 2024).
43. McCullagh, M.J. Terrain and surface modelling systems: Theory and practice. *Photogramm. Rec.* **1988**, *12*, 747–779. [CrossRef]
44. Hengl, T. Finding the right pixel size. *Comput. Geosci.* **2006**, *32*, 1283–1298. [CrossRef]
45. Henn, K.A.; Peduzzi, A. Biomass estimation of urban forests using LiDAR and high-resolution aerial imagery in Athens–Clarke County, GA. *Forests* **2023**, *14*, 1064. [CrossRef]
46. Leboeuf, A.; Riopel, M.; Munger, D.; Fradette, M.S.; Bégin, J. Modeling merchantable wood volume using Airborne LiDAR metrics and historical forest inventory plots at a provincial scale. *Forests* **2022**, *13*, 985. [CrossRef]
47. 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. [CrossRef]
48. Hawbaker, T.J.; Keuler, N.S.; Lesak, A.A.; Gobakken, T.; Contrucci, K.; Radeloff, V.C. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res. Biogeophys.* **2009**, *114*, G00E04. [CrossRef]
49. 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]
50. Sumnall, M.J.; Hill, R.A.; Hinsley, S.A. Towards forest condition assessment: Evaluating small-footprint full-waveform airborne laser scanning data for deriving forest structural and compositional metrics. *Remote Sens.* **2022**, *14*, 5081. [CrossRef]
51. Yan, W.Y.; Van Ewijk, K.; Treitz, P.; Shaker, A. Effects of radiometric correction on cover type and spatial resolution for modeling plot level forest attributes using multispectral airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 152–165. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.