

Article

Not peer-reviewed version

---

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

---

[Taeyoon Lee](#)\*, Can Vatandaslar, [Krista Merry](#), [Pete Bettinger](#), [Alicia Peduzzi](#), [Jonathan Stober](#)

Posted Date: 7 June 2024

doi: 10.20944/preprints202406.0354.v1

Keywords: Aerial imagery; airborne laser scanning; forest inventory; LiDAR; Mixed pine hardwood forest



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

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

Taeyoon Lee <sup>1,\*</sup>, Can Vatandaslar <sup>1,2</sup>, Krista Merry <sup>1</sup>, Pete Bettinger <sup>1</sup>, Alicia Peduzzi <sup>1</sup> and Jonathan Stober <sup>3</sup>

<sup>1</sup> Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA

<sup>2</sup> Faculty of Forestry, Artvin Coruh University, 08000 Artvin, Türkiye

<sup>3</sup> U.S. Forest Service, Talladega National Forest, Heflin AL 36264, U.S.A.

\* 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 Talladega National Forest by employing ALS-derived data and aerial photography. The quality of 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 was 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; forest inventory; LiDAR; mixed pine hardwood forest

## 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 is also informative for assessments that demonstrate the sustainable management of natural resources [1,2]. Within this context, an understanding of forest conditions, particularly 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 [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 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 three-dimensional forest structure to a certain extent based on the point density of the LiDAR data [5,16,18]. LiDAR can help with deriving 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 has been 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. For instance, 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] was able to create a global canopy height model using GEDI data along with Landsat imagery. Additionally, Dubayah et al. [21] presented the estimate of mean biomass densities for every country covered by GEDI with 1 km resolution.

Terrestrial-based LiDAR platforms can be categorized as either static or mobile. Terrestrial laser scanners (TLS) capture point clouds by looking upward from ground level, so they are advantageous in capturing 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 TLS 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, was 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, 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 of whether TLS is a practical alternative for collecting forest inventory [26,27].

Mobile laser scanning (MLS) platforms, which is 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 of forest structure from a perspective very similar to human-collected field measurements. MLS platforms can collect information while a person traverses sampling plots with the instrument held in the person's 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 of the Talladega Division of the Talladega National Forest in the southeastern US cover almost 93,694 ha [31], and thus the number of TLS or MLS plots needed to describe forest 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.

In this context, 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 have further encouraged the adoption of ALS systems, allowing 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 m<sup>2</sup>, yet today the development of a point cloud dataset above 40 points per m<sup>2</sup> 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 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, yet 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 [35,36]. 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 [37] 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. [35] recommended the use of 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 of the Talladega National Forest as our case study area because it represents typical characteristics of natural, pine-dominated forests of southeastern USA. National forest managers do not have comprehensive inventories of the extent under management limiting global decision space for at risk resources be it wildland fire, forest health, endangered species habitat condition to help assess where management is needed or has achieved 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 insight 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 Division of the Talladega National Forest, located in northwestern Alabama, USA. 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 lands in the Talladega Division varies between approximately 160 m to 735 m above sea level, and annual precipitation is around 1260 mm. Historically, the forests in the study area have been 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 m<sup>2</sup> with a 11.32 m radius) were measured by U.S. Forest Service crews between February and April 2022. These plots were pseudo-randomly located within the operable (upland) lands of the study area. An equal number of plots were measured within eight forest structure classes, defined by canopy density and tree height. Plot locations were limited to the interior (rather than edges) of management units (stands), and only one plot per management unit was allowed. Plot centers were mapped using a Trimble R1 GNSS receiver (Trimble, Colorado, USA). The field data collection procedure followed methods described in Laes et al. [38]. Within each plot, several characteristics of each live tree (height, dbh, tree species, 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 measured live trees.

#### 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. The county mosaic images were acquired from the U.S. Department of Agriculture National Agricultural Imagery Program (NAIP) [39]. The ALS data was collected using a Leica Terrain Mapper. Within the study area, the ALS data contained 13.92 billion points. The average point density achieved was 9.7 points per m<sup>2</sup> and ranged from 3.5 points per m<sup>2</sup> to 20.3 points per m<sup>2</sup>. The horizontal and vertical accuracy of the LiDAR data (0.71 m and 0.051 m, respectively) was estimated using 441 survey points spread across the State of Alabama. These characteristics satisfy the requirements for at least topographic quality level 2 (QL2) [40].

2.3. Data Processing

2.3.1. Field Data Processing

From the field measurement plots we estimated 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 basal area (m<sup>2</sup> ha<sup>-1</sup>), tree volume (m<sup>3</sup> ha<sup>-1</sup>), and aboveground biomass (Mg ha<sup>-1</sup>) as dependent variables for the modeling effort. Basal area was estimated based on dbh, and 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 sampling 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 sampling plots and 149 pine dominated sampling plots (i.e., pine plots). All other plots indicated no dominance towards 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) pine plots (n=149).

2.3.2. Remote Data Processing

From 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 <sub>MIN</sub> )
		Maximum of greenness (G <sub>MAX</sub> )
		Range of greenness (G <sub>RANGE</sub> )
		Mean of greenness (G <sub>MEAN</sub> )
		Standard deviation of greenness (G <sub>STD</sub> )
		Sum of greenness (G <sub>SUM</sub> )
		Median of greenness (G <sub>MEDIAN</sub> )
		90 percentage of greenness (G <sub>PCT90</sub> )
Normalized Difference Vegetation Index, NDVI	$\frac{NIR - R}{NIR + R}$	Minimum of NDVI (NDVI <sub>MIN</sub> )
		Maximum of NDVI (NDVI <sub>MAX</sub> )
		Range of NDVI (NDVI <sub>RANGE</sub> )
		Mean of NDVI (NDVI <sub>MEAN</sub> )
		Standard deviation of NDVI (NDVI <sub>STD</sub> )
		Sum of NDVI (NDVI <sub>SUM</sub> )
		Median of NDVI (NDVI <sub>MEDIAN</sub> )
		90 percentage of NDVI (NDVI <sub>PCT90</sub> )
Enhanced Vegetation Index, EVI	$\frac{2.5 \times (NIR - R)}{(NIR + 6 \times R - 7.5 \times B + 1)}$	Minimum of EVI (EVI <sub>MIN</sub> )
		Maximum of EVI (EVI <sub>MAX</sub> )
		Range of EVI (EVI <sub>RANGE</sub> )
		Mean of EVI (EVI <sub>MEAN</sub> )
		Standard deviation of EVI (EVI <sub>STD</sub> )

Sum of EVI ( $EVI_{SUM}$ )

Median of EVI ( $EVI_{MEDIAN}$ )

90 percentage of EVI ( $EVI_{PCT90}$ )

\*Where  $R$ ,  $G$ ,  $B$ ,  $NIR$  represent the raw pixel values of red, green, blue, and the 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 was processed using RStudio (R version 2023.06.1 Build 524) and the University of Georgia’s Georgia Advanced Computing Resource Center’s 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 estimated 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 [41] and Hengl [42]. The triangular irregular network algorithm, a vector terrain model using Delaunay triangles, was applied to create a 1 m spatial resolution digital terrain model. With the digital terrain model representing the ground, we then generated a normalized LiDAR point cloud including only the above ground points. The first returns in the normalized point cloud were then used to create a canopy height model. Noise, resulting from points within the point cloud with negative values or values greater than 95% of the height, were removed to increase the quality of regression models. LiDAR-derived metrics were then calculated from point clouds that were clipped using the boundary of each sampling plot using R package (‘lidR’, version 4.0.3). We extracted 56 LiDAR-derived metrics from the normalized LiDAR point clouds for 254 of sampling plots. Among LiDAR-derived metrics, we excluded certain metrics with absolute values (such as maximum height and 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 are 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 $i^{th}$ 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 $x^{th}$ percentile of height
Pzabove2	Percentage of returns above 2 m	P $x^{th}$ (1, 2, 3, 4, and 5)	Percentage $x^{th}$ returns
zq x (From 5th to 95th)	$x^{th}$ 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, aboveground biomass) with two different sets of field measurement sampling plots (all plots, pine plots) and two types of data sources (LiDAR only, LiDAR + NAIP). Depending on which field measurement sampling 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 LiDAR and NAIP data sources, inter-correlation between independent variables and the complexity of models was inevitable [3]. To reduce the complexity of models and to avoid multicollinearity issues, we applied the ALASSO regression. As 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 predicted values and observed values. The best lambda was obtained by applying 10-fold cross-validation. Prior to applying ALASSO regression, the normality of 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 residual variances, and the normality of 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, root mean square of error (RMSE), Collin Mallow’s  $C_p$ , Akaike’s information criteria (AIC), and 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, 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 sampling plots was about 23.4 m<sup>2</sup> ha<sup>-1</sup>, 180.9 m<sup>3</sup> ha<sup>-1</sup>, and 40.1 Mg ha<sup>-1</sup>, respectively (Table 3). For the pine dominated field measurement plots, the average basal area, tree volume, and aboveground biomass were about 22.3 m<sup>2</sup> ha<sup>-1</sup>, 169.0 m<sup>3</sup> ha<sup>-1</sup>, and 34.6 Mg ha<sup>-1</sup>, respectively. The range of values for the sampling plots was large since the plots were meant to be representative of all forest conditions in the study area, from early successional to mature forest stages.

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

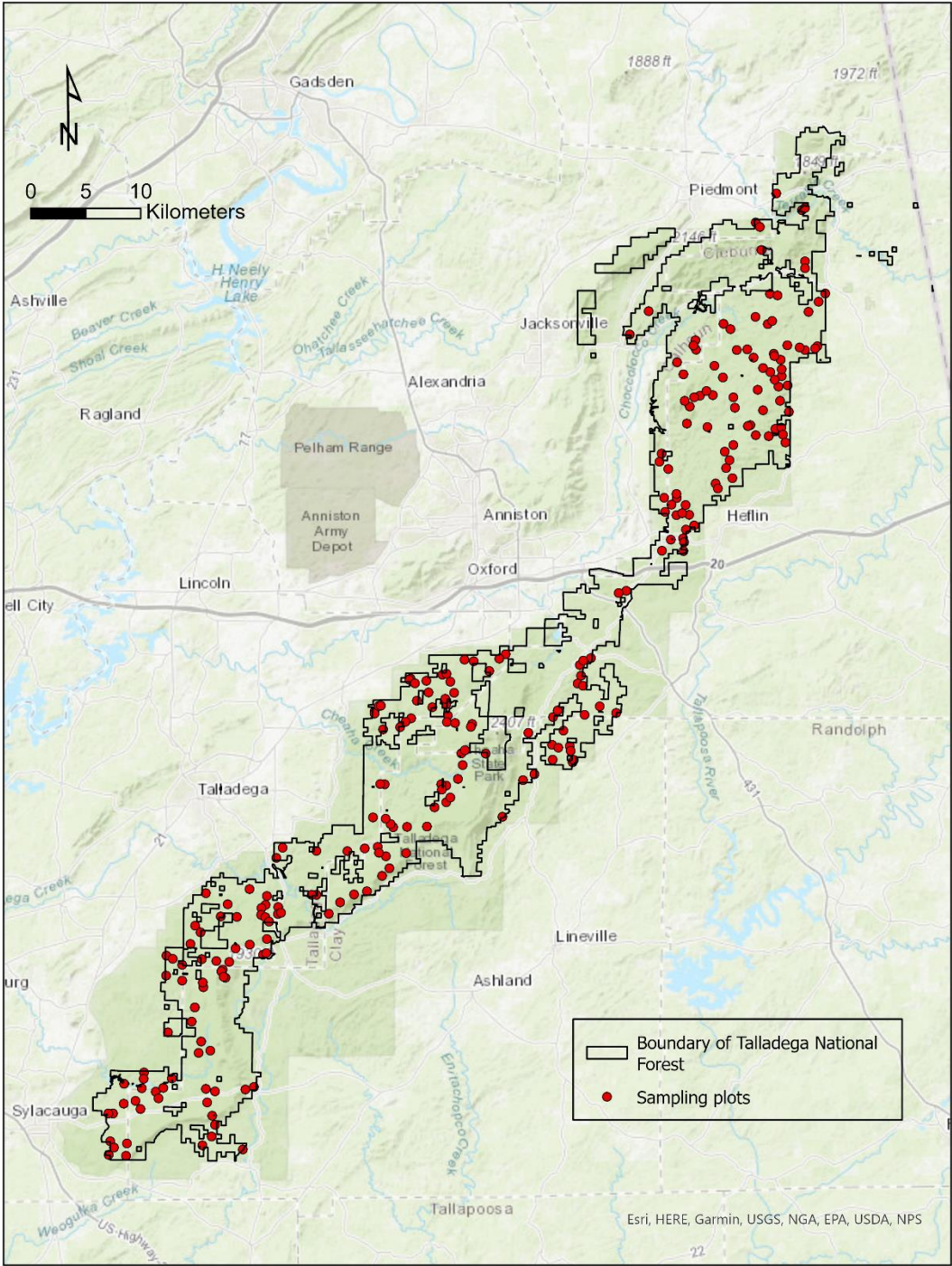
	Diameter at breast height (cm)	Basal area (m <sup>2</sup> ha <sup>-1</sup> )	Volume (m <sup>3</sup> ha <sup>-1</sup> )	Aboveground biomass (Mg ha <sup>-1</sup> )
All plots (n =254)				
Average	22.39	23.43	180.85	40.13
Standard deviation	7.39	10.95	105.48	23.07
Minimum	8.65	0.33	0.77	0.13
Maximum	54.36	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

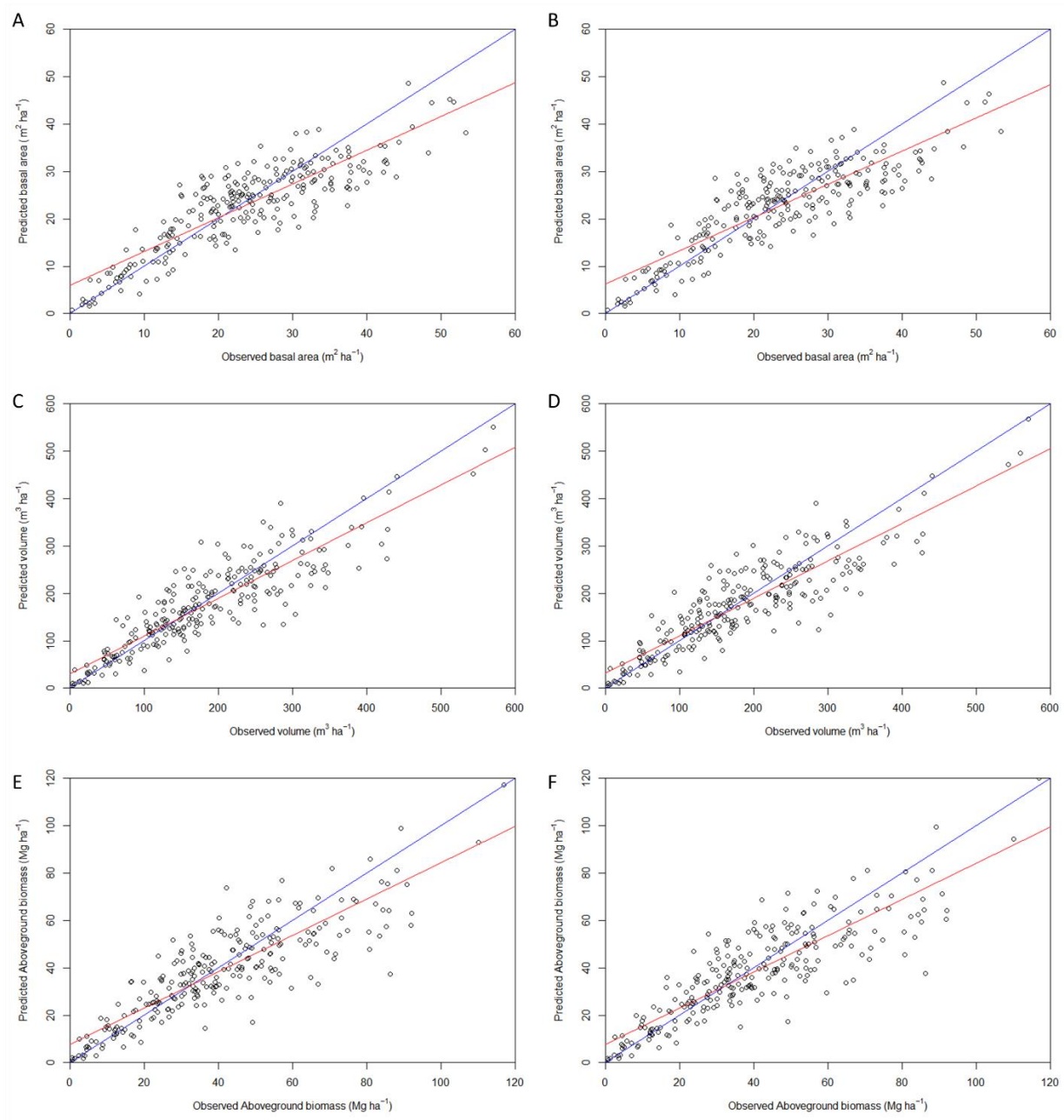


### 3.2.1. Estimation of Forest Attributes Based on General Models

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) LiDAR-only, while 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:  $5.91 \text{ m}^2 \text{ ha}^{-1}$ ) 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 predicted basal area tended to be underestimated as its basal area increased regardless of data sources (Figure 2).



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



**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).

**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	$-1.828 + 0.017 * pzabove2 + 0.015 * zq25 + 0.017 * zq95 - 7.83 * e^{-4} * zpcum5 - 0.002 * zpcum6 + 2 * e^{-5} * isd + 0.24 * iskew - 0.091 * ikurt + 0.024 * ipcumzq90 + 0.043 * p2th + 0.116 * G_{MIN} + 0.52 * NDVI_{MIN} + 0.428 * NDVI_{MEDIAN} + 3.58 * e^{-5} * EVI_{MAX} - 0.012 * EVI_{PCT90}$
	LiDAR	$-1.197 + 0.018 * pzabove2 + 0.017 * zq25 + 1.13 * e^{-4} * zq30 + 0.015 * zq95 - 3.79 * e^{-4} * zpcum5 - 0.003 * zpcum6 + 1.83 * e^{-5} * isd + 0.196 * iskew - 0.105 * ikurt + 0.017 * ipcumzq90 + 0.046 * p2th$
Volume	LiDAR+NAIP	$-0.148 + 0.018 * pzabove2 + 0.004 * zq25 + 0.056 * zq95 - 7.91 * e^{-4} * zpcum5 - 0.00566 * zpcum6 + 2.51 * e^{-5} * isd + 0.237 * iskew - 0.128 * ikurt - 0.018 * ipcumzq10 - 0.002 * ipcumzq30 + 0.029 * ipcumzq90 - 0.002 * p1th + 0.029 * p2th + 0.091 * G_{MIN} - 0.165 * G_{GRANGE} + 0.522 * NDVI_{MIN} + 0.229 * NDVI_{MEAN} + 1.45 * e^{-5} * NDVI_{SUM} + 0.158 * NDVI_{MEDIAN} + 1.96 * e^{-4} * EVI_{MAX} - 0.010 * EVI_{PCT90}$
	LiDAR	$0.380 + 0.020 * pzabove2 + 0.007 * zq25 + 0.053 * zq95 - 0.006 * zpcum6 + 3.088 * e^{-05} * isd + 0.209 * iskew - 0.121 * ikurt - 0.020 * ipcumzq10 - 0.004 * ipcumzq30 + 0.020 * ipcumzq90 - 4.642 * e^{-5} * p1th + 0.037 * p2th$
Aboveground biomass	LiDAR+NAIP	$0.033 + 0.021 * pzabove2 + 0.062 * zq95 - 0.004 * zpcum6 + 3.93 * e^{-5} * isd + 0.132 * iskew - 0.184 * ikurt - 0.011 * ipcumzq10 + 0.083 * ipcumzq90 - 0.006 * p1th + 0.031 * p2th - 1.51 * e^{-4} * pground + 0.385 * G_{MIN} + 0.505 * NDVI_{MIN} + 0.218 * NDVI_{MEAN} + 9.36 * e^{-6} * NDVI_{SUM} - 0.005 * EVI_{PCT90}$
	LiDAR	$0.516 + 0.022 * pzabove2 + 0.238 * zq5 + 0.002 * zq25 + 0.059 * zq95 - 0.005 * zpcum6 + 3.83 * e^{-5} * isd + 0.103 * iskew - 0.198 * ikurt - 0.012 * ipcumzq10 + 0.078 * ipcumzq90 - 0.006 * p1th + 0.034 * p2th$
Pine models		
Basal area	LiDAR+NAIP	$0.657 + 0.009 * pzabove2 + 3.600 * zq5 + 0.021 * zq25 + 0.001 * zq40 + 0.007 * zq95 - 0.009 * zpcum5 + 0.173 * iskew - 0.067 * ikurt + 0.057 * p2th + 0.426 * NDVI_{MIN} + 0.985 * NDVI_{MEDIAN} - 8.6 * e^{-4} * EVI_{PCT90}$
	LiDAR	$5.195 + 0.011 * pzabove2 + 2.873 * zq5 + 0.021 * zq25 + 1.681 * e^{-4} * zq40 + 0.003 * zq95 - 0.008 * zpcum5 - 1.96 * e^{-4} * zpcum6 + 0.168 * iskew - 0.061 * ikurt - 0.004 * ipcumzq30 - 0.0475 * ipcumzq90 + 0.057 * p2th$
Volume	LiDAR+NAIP	$1.593 - 0.002 * zkurt + 0.021 * pzabove2 + 8.41 * zq5 - 1.91 * zq15 + 0.019 * zq25 + 0.006 * zq40 - 0.008 * zq65 + 0.016 * zq75 - 0.071 * zq80 + 0.096 * zq95 + 0.010 * zpcum1 - 0.010 * zpcum5 - 0.005 * zpcum6 + 1.29 * e^{-5} * zpcum8 + 0.002 * zpcum9 + 3.04 * e^{-5} * isd + 0.241 * iskew - 0.16 * ikurt + 0.016 * ipcumzq10 - 0.008 * ipcumzq30 + 0.042 * p2th - 0.039 * p5th + 0.486 * G_{MIN} - 0.132 * G_{GRANGE} + 0.475 * NDVI_{MIN} + 1.28 * NDVI_{MEDIAN} + 2.7 * e^{-4} * EVI_{MAX} - 0.004 * EVI_{STD} - 0.019 * EVI_{MEDIAN} - 0.015 * EVI_{PCT90}$
	LiDAR	$3.592 + 0.008 * pzabove2 + 3.401 * zq5 + 0.004 * zq25 + 0.043 * zq95 - 0.013 * zpcum5 + 0.222 * iskew - 0.094 * ikurt - 2.769 * e^{-4} * ipcumzq10 - 0.033 * ipcumzq30 + 0.047 * p2th + 0.013 * p3th$
Aboveground biomass	LiDAR+NAIP	$6.466 - 0.005 * zkurt + 0.41 * zentropy + 0.025 * pzabove2 + 8.66 * zq5 - 0.037 * zq10 - 2.18 * zq15 + 0.017 * zq25 - 7.48 * e^{-4} * zq30 + 0.004 * zq40 - 0.001 * zq65 + 0.010 * zq75 - 0.079 * zq80 + 0.11 * zq95 + 0.013 * zpcum1 - 0.009 * zpcum5 - 0.005 * zpcum6 + 0.003 * zpcum9 + 4.75 * e^{-05} * isd + 0.078 * iskew - 0.175 * ikurt + 0.020 * ipcumzq10 - 0.008 * ipcumzq90 + 0.042 * p2th - 0.016 * p5th + 0.673 * G_{MIN} - 0.094 * G_{GRANGE} + 0.171 * G_{STD} - 0.362 * G_{MEDIAN} + 0.364 * NDVI_{MIN} - 0.327 * NDVI_{STD} + 1.51 * NDVI_{MEDIAN} + 1.52 * e^{-4} * EVI_{MAX} - 0.002 * EVI_{STD} - 0.033 * EVI_{PCT90}$

LiDAR	$10.699 + 0.013 * pzabove2 + 2.062 * zq5 + 0.052 * zq95 - 0.011 * zpcum5 + 0.010 * iskew - 0.129 * ikurt - 0.010 * ipcumzq30 - 0.024 * ipcumzq90 - 0.009 * p1th + 0.045 * p2th$
-------	---



**Table 5.** Summary of statistics for the general models.

Quality metrics	Basal area (m <sup>2</sup> ha <sup>-1</sup> )		Total volume (m <sup>3</sup> ha <sup>-1</sup> )		Total aboveground biomass (Mg ha <sup>-1</sup> )	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R <sup>2</sup> adj.	0.72	0.71	0.77	0.77	0.73	0.72
# of variables	15	11	21	12	16	12
RMSE	5.58	5.73	48.44	49.34	11.68	11.84
R <sup>2</sup>	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<sup>2</sup>adj. value (0.77) was observed (Table 5). Although the overall quality metrics for LiDAR + NAIP and LiDAR-only models represented a similar level of performance, the increment in the number of independent variables in 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, zpcum5 (cumulative percentile of return in the 5th layer), was also selected as an independent variable in LiDAR + NAIP model yet was not used in the LiDAR-only model. As observed in the basal area models, the difference between observed volume and estimated volume tended to increase with higher volume regardless of data sources (Figure 2).

The aboveground biomass models had R<sup>2</sup>adj. values that were 0.73 and 0.72 (cross-validation: 0.64 and 0.65) for 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<sup>-1</sup> (cross-validation:13.05 Mg ha<sup>-1</sup>); LiDAR-only model: 11.8 Mg ha<sup>-1</sup>(cross-validation:13.09 Mg ha<sup>-1</sup>)). Nevertheless, the difference between aboveground biomass models depending on data sources was slight. Five vegetation indices were added into the LiDAR + NAIP model as independent variables (Table 4). Both aboveground biomass models underestimated biomass for the sample plots having higher aboveground biomass, suggesting estimation errors increase as forest stands mature (Figure 2).

The average R<sup>2</sup>adj. and average R<sup>2</sup> values of 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 statistics of cross-validation and developed models were generally smaller in LiDAR based models compared to the LiDAR + NAIP models. Regarding the selection of independent variables, several metrics were commonly selected regardless of data source, including: pzabove2, zq95, zpcum6, isd, iskew, ikurt, ipcumzq90 and p2th (all from LiDAR data); and G<sub>MIN</sub>, NDVI<sub>MIN</sub> and EVI<sub>PCT90</sub> (all from NAIP data) (Table 7).

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

Quality metrics	Basal area (m <sup>2</sup> ha <sup>-1</sup> )		Total volume (m <sup>3</sup> ha <sup>-1</sup> )		Total aboveground biomass (Mg ha <sup>-1</sup> )	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R <sup>2</sup> adj.	0.69	0.71	0.67	0.73	0.64	0.65
R <sup>2</sup>	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 of the best regression models.

	LiDAR metrics	NAIP metrics
General & pine model	pzabove2, zq95, iskew, ikurt, p2th	NDVI <sub>MIN</sub> , EVI <sub>PCT90</sub>
General model	zpcum6, isd, ipcumzq90	G <sub>MIN</sub>
Pine model	zq5, zpcum5	NDVI <sub>MEDIAN</sub>

3.2.2. Estimation of Forest Attributes Based on Pine Models

The  $R^2_{adj.}$  and RMSE values of basal area models were 0.81 and 0.80 ( $R^2_{adj.}$  of cross-validation: 0.79 and 0.75), and 4.8 m<sup>2</sup> ha<sup>-1</sup> and 5.1 m<sup>2</sup> ha<sup>-1</sup> (RMSE of cross-validation: 5.21 m<sup>2</sup> ha<sup>-1</sup> and 5.71 m<sup>2</sup> ha<sup>-1</sup>) for LiDAR + NAIP and LiDAR, respectively (Table 8). Noticeable improvements in 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<sub>MIN</sub>, NDVI<sub>MEDIAN</sub>, and EVI<sub>PCT90</sub>) were selected instead of LiDAR derived metrics (zpcum6, ipcumzq30, ipcumzq90) (Table 4). The visual interpretation of the scatter plots showed that basal area tended to be underestimated as basal area values increased regardless of the data sources (Figure 3). Similar trends in scatter plots were also observed in general models, but the distribution of points was closer to the 1:1 line.

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

The aboveground biomass models had  $R^2_{adj.}$  values of 0.83 (cross-validation: 0.80) for LiDAR + NAIP and 0.82 (cross-validation: 0.78) for LiDAR-only models (Table 8). The RMSE values of aboveground biomass models were 7.9 Mg ha<sup>-1</sup> and 8.9 Mg ha<sup>-1</sup> (cross-validation: 9.4 Mg ha<sup>-1</sup> and 10.2 Mg ha<sup>-1</sup>) for LiDAR + NAIP and LiDAR-only, respectively. Regarding AIC and BIC values, our results indicated that the quality of LiDAR-only models was better than the LiDAR + NAIP models. The number of independent variables also noticeably increased in LiDAR + NAIP models (LiDAR + NAIP: 34; LiDAR-only: 10). Ten NAIP derived metrics were selected in addition to 16 LiDAR derived metrics in LiDAR + NAIP model (Table 4). Similar to the general regression models, both models (LiDAR + NAIP and LiDAR-only) underestimated aboveground biomass for the sample plots having higher biomass estimates, suggesting estimation errors increased as stands mature (Figure 3).

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

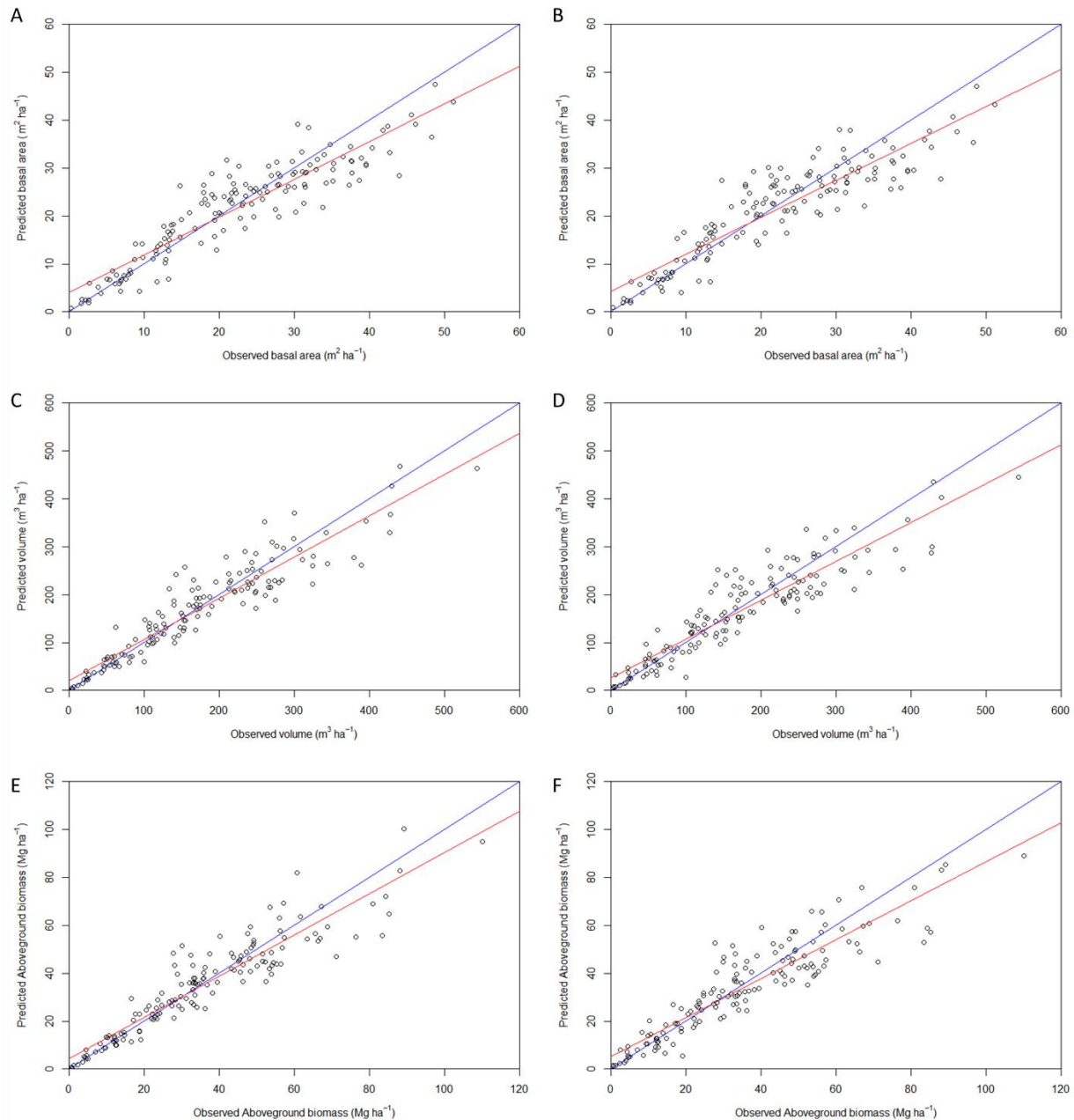
**Table 8.** Summary of statistics for the pine regression models.

Quality metrics	Basal area (m <sup>2</sup> ha <sup>-1</sup> )	Total volume (m <sup>3</sup> ha <sup>-1</sup> )	Total aboveground biomass (Mg ha <sup>-1</sup> )
-----------------	---	---	--

	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R <sup>2</sup> <sub>adj.</sub>	0.81	0.80	0.84	0.82	0.83	0.82
# of variables	12	12	30	11	34	10
RMSE	4.80	5.10	37.86	43.45	7.89	8.93
R <sup>2</sup>	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

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

Quality metrics	Basal area (m <sup>2</sup> ha <sup>-1</sup> )		Total volume (m <sup>3</sup> ha <sup>-1</sup> )		Total aboveground biomass (Mg ha <sup>-1</sup> )	
	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR	LiDAR + NAIP	LiDAR
R <sup>2</sup> <sub>adj.</sub>	0.79	0.75	0.78	0.79	0.80	0.78
R <sup>2</sup>	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



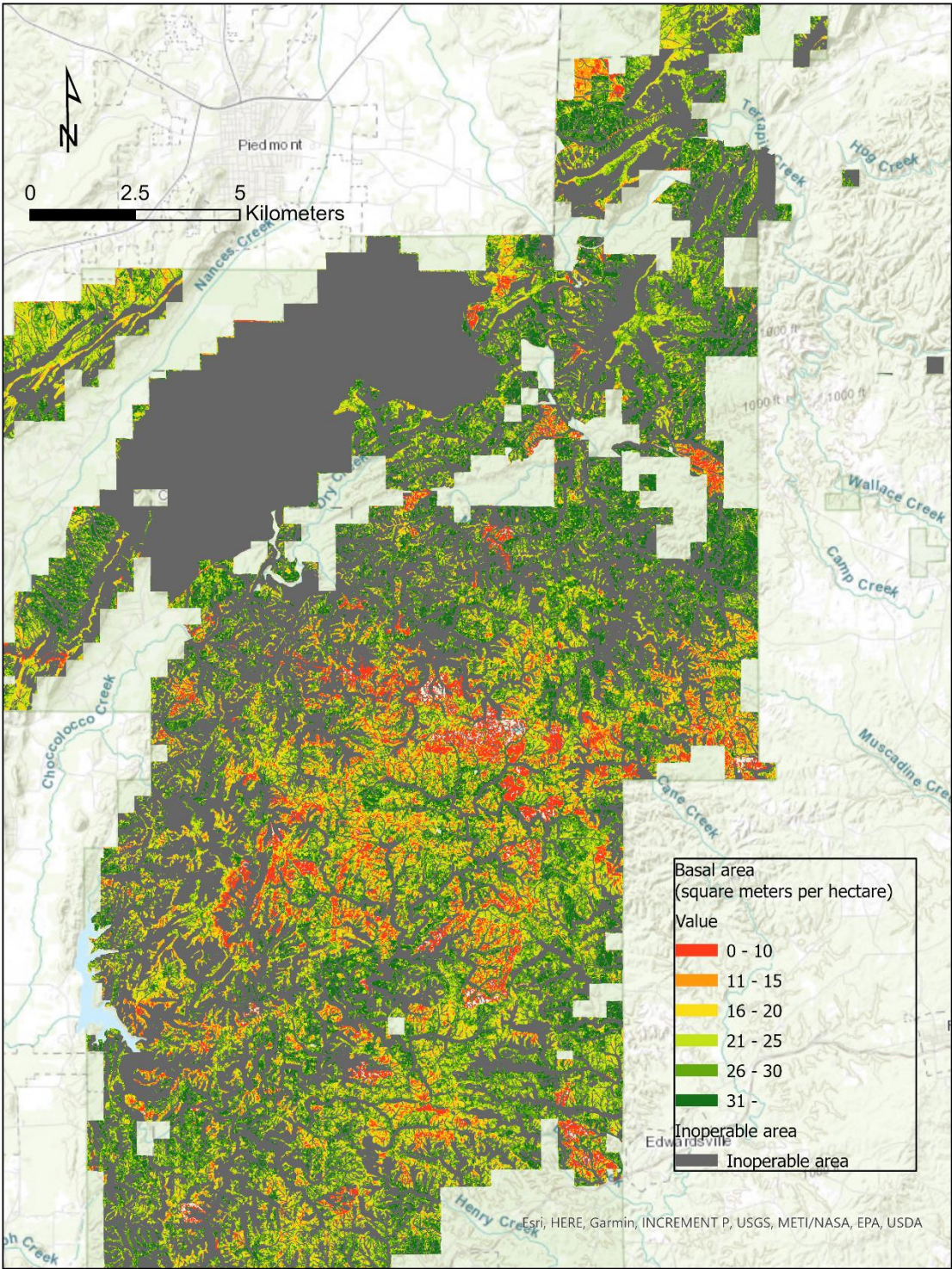
**Figure 3.** The observed forest attributes versus the predicted forest attributes of the pine models ( $n=149$ ). (LiDAR+NAIP: A, C, E; LiDAR: B, D, F).

#### 4. Discussion

Estimating forest attributes such as 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 regression models. Also, we evaluated the quality of regression models depending on the classification of sampling plots according to 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 performance was suggested by earlier studies [35,43]. Eventually, a total of 12 models were developed for three forest attributes (basal area, volume, aboveground biomass) based on the species mixture (all species and pine), as well as data source

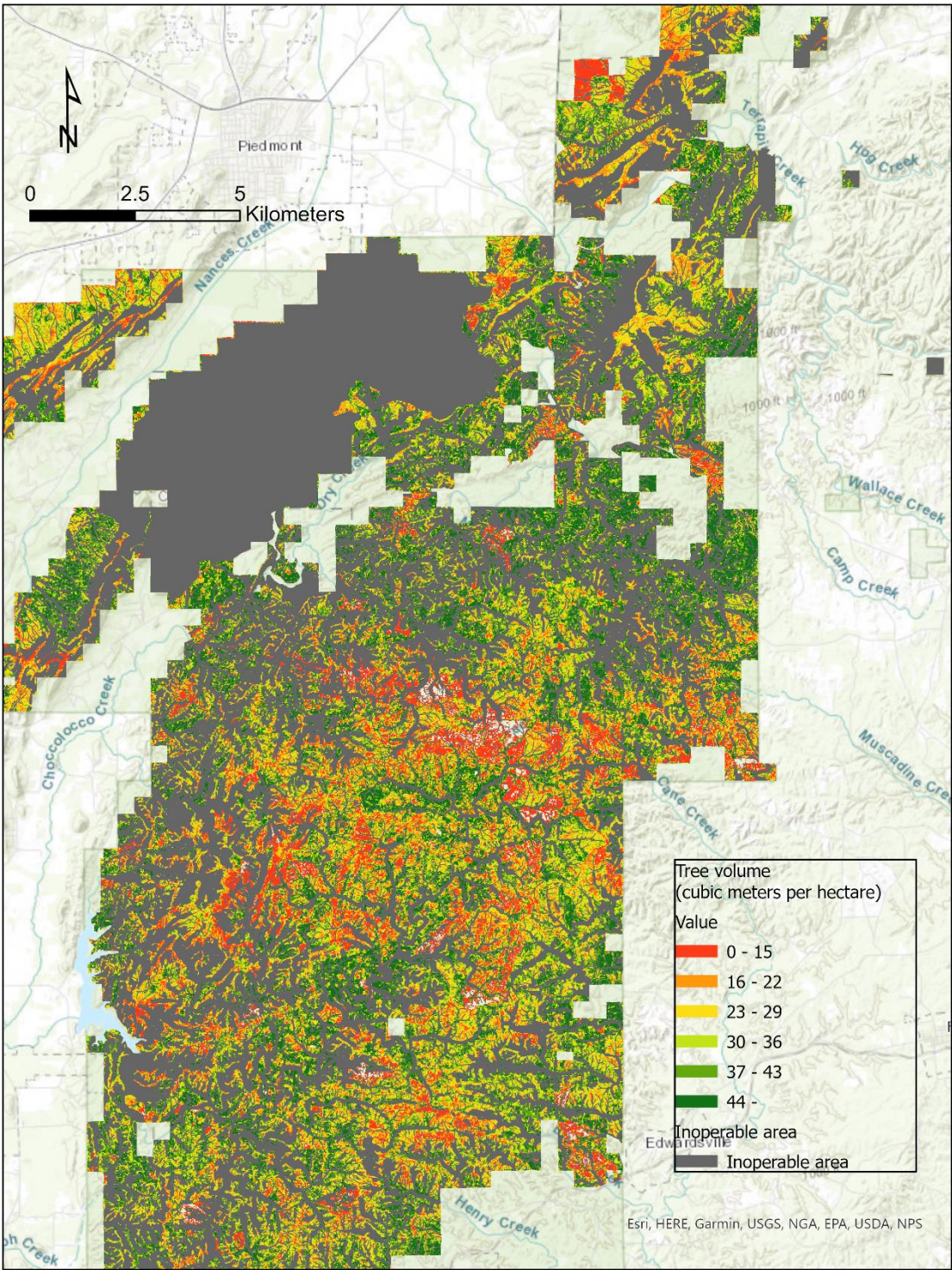


(LiDAR-only, LiDAR + NAIP). When the general models are applied to the case study landscape, broad scale maps of these resource conditions can be visualized (Figures 4–6).



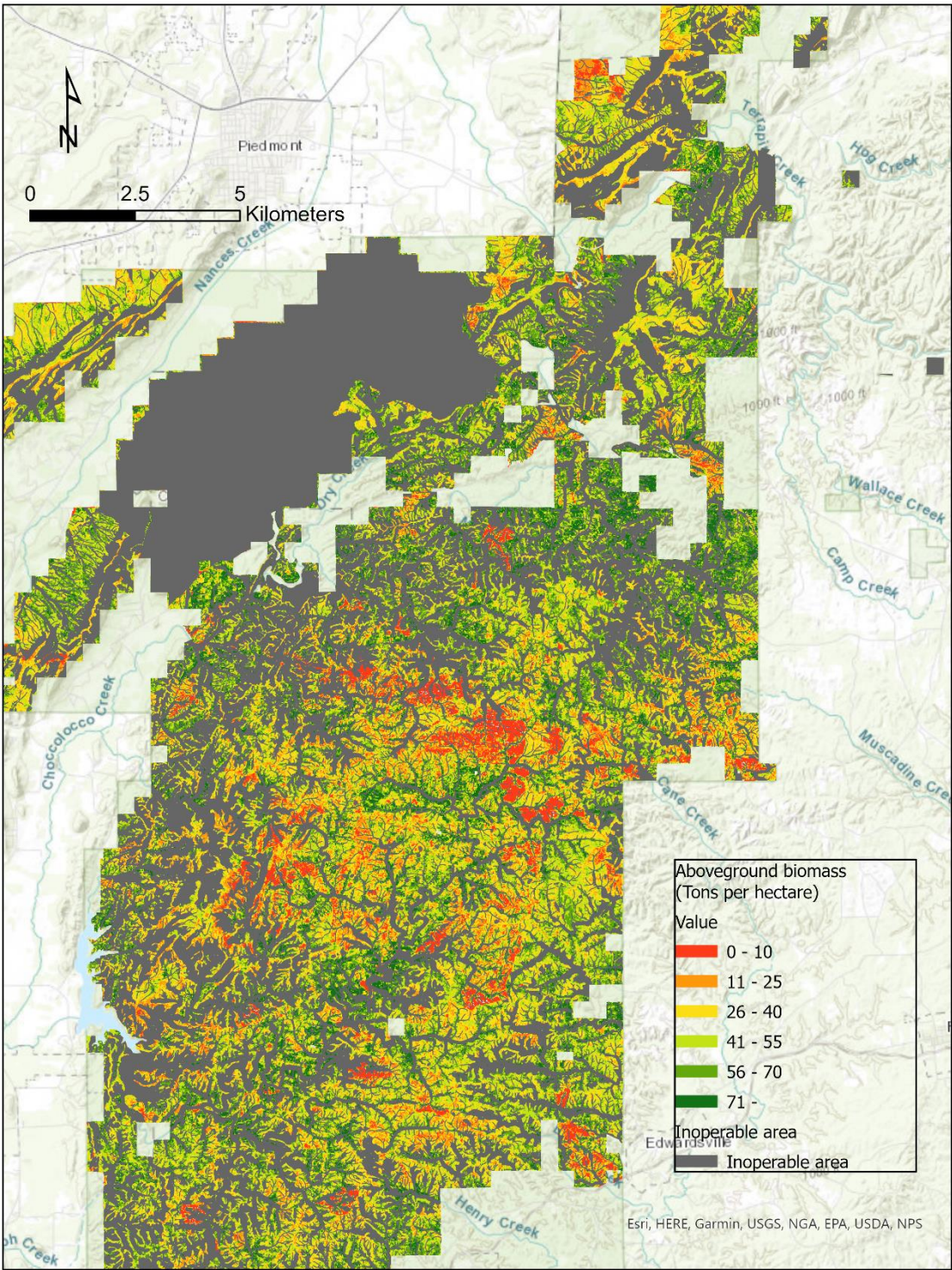
**Figure 4.** Estimated basal area for a portion of the study area based on general model (Sources: Esri. “World Topographic Map” [basemap]. January 31, 2024).





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





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

The  $R^2_{adj}$  values of developed models ranged from 0.71 to 0.84 which were comparable to other ALS study results [3,34,44]. While many researchers have developed regression models for estimating forest attributes over relatively small forested areas [10,15,19,34,36], there have been few

attempts to work in large areas such as U.S. National Forests. For instance, Leboeuf et al. [44] mapped a merchantable wood volume of very large area (440,000 km<sup>2</sup>) using airborne LiDAR. However, this study had a few limitations, including significant temporal differences between the field measurements (2003-2018) and LiDAR data (2011-2020) collection periods, limited spatial resolution, and poor representativeness of sampling plots. In our study, however, field and remote sensing datasets were collected during a relatively similar period of time (2020-2022). The distribution of sample plots is also important in developing robust regression models for forest attribute estimation. To enhance the quality of regression models, we used a stratified pseudo-random sampling design by classifying the entire operable study area with consideration of 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 sampling plots to enhance the accuracy of regression models, as suggested by Brown et al. [45].

The  $R^2_{adj}$  values and  $R^2$  values of pine models were higher than those of the general models regardless of forest attributes. The overall quality metrics also indicated pine models were higher 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  $R^2_{adj}$  values were observed in tree volume models and the lowest  $R^2_{adj}$  values were observed in basal area models regardless of data sources and sampling plots (all plots or pine plots). This trend has also been observed in other studies using ALS systems [3,34,46,47]. Sumnall et al. [48] 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 basal area, yet dbh and tree height are often needed to estimate volume and aboveground biomass [3], therefore it may be logical to observe that the basal area estimation models would underperform the volume and biomass models.

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 [3,49], and perhaps supplementary remotely sensed data [34,45]. In this study, the vegetation indices derived from NAIP imagery were included as supplemental data. As the additional NAIP imagery had higher spatial resolution (0.3 m), we expected the vegetation indices might improve the quality of 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 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 LiDAR-only models were more appropriate for broad-scale mapping efforts.

Of the 74 metrics derived from 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<sub>MIN</sub> and EVI<sub>PCT90</sub>. LiDAR-based metrics pzabove2 and p2th metrics 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 regression models as they provide information related to stand density. The inclusion of NAIP-based NDVI<sub>MIN</sub> and EVI<sub>PCT90</sub> 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.

Although 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 basal area, volume, and aboveground biomass is due to the high correlation among these forest attributes (Figures 4-6). Since aboveground biomass of a forest is typically a



derivation of growing stock, volume estimates often rely on dbh measurements, and as 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.

## 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 topographic quality level QL2 character seems to be a sufficient input for the development of regression models that can estimate 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 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 was needed to assess the scalability of the results to other forest ecosystems of the southern USA.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Estimated basal area for the entire study area based on general model (Sources: Esri. "World Topographic Map" [basemap]. January 31, 2024).; Figure S2: Estimated tree volume for the entire study area based on general model (Sources: Esri. "World Topographic Map" [basemap]. January 31, 2024).; Figure S3: Estimated aboveground biomass for the entire study area based on general model (Sources: Esri. "World Topographic Map" [basemap]. January 31, 2024).

**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., B.P., K.M., J.S., 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:** We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data was created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at <https://www.mdpi.com/ethics>.

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

## References

1. Bettinger, P.; Boston, K.; Siry, J.P.; Grebner, D.L. Forest management and planning. Academic Press, London. 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. <https://doi.org/10.1016/j.rse.2020.111770>
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. <https://doi.org/10.1016/j.rse.2014.10.004>

4. Chen, Y.; Kershaw, J.A.; Hsu, Y.H.; Yang, T.R. Carbon estimation using sampling to correct LiDAR-assisted enhanced forest inventory estimates. *Forestry Chron.* 2020, 96 (1), 9-19. <https://doi.org/10.5558/tfc2020-003>
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. <https://doi.org/10.1016/j.rse.2023.113543>
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 (6), 1086. <https://doi.org/10.3390/f14061086>
7. Vatandaşlar, C.; Seki, M.; Zeybek, M. Assessing the potential of mobile laser scanning for stand-level forest inventories in near-natural forests. *Forestry.* 2023, 96 (4), 448-464. <https://doi.org/10.1093/forestry/cpad016>
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. Forest.* 2015, 113(5), 454-462. <https://doi.org/10.5849/jof.15-002>
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, <https://doi.org/10.1155/2011/572452>
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. <https://doi.org/10.1016/j.ecoinf.2023.102254>
11. Seki, M.; Sakici, O.E. Ecoregion-based height-diameter models for Crimean pine. *J. For. Res.* 2022, 27, 36-44. <https://doi.org/10.1080/13416979.2021.1972511>
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 (4), 816-830. <https://doi.org/10.1016/j.rse.2009.11.021>
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 (5): 619-641. <https://doi.org/10.1080/07038992.2016.1207484>
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. Information Report FI-X-018. 2017, 38. 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. <https://doi.org/10.1016/j.rse.2016.05.028>
16. Lim, K.; Treitz, P.; Wulder, M.A.; St-Onge, B.; Flood, M.; LiDAR remote sensing of forest structure. *Prog. Phys. Geog.* 2003, 27 (1), 88-106. <https://doi.org/10.1191/0309133303pp360>
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 (11), 1891. <https://doi.org/10.3390/rs12111891>
18. Bater, C.W.; Coops, N.C. Evaluating error associated with Lidar-derived DEM interpolation. *Comput. Geosci.* 2009, 35, 289-300. <https://doi.org/10.1016/j.cageo.2008.09.001>
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. <https://doi.org/10.1016/j.neucom.2014.09.091>
20. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; Armston, J.; Dubayah, R.; Blair, J.B.; Hofton, M. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* 2021, 253, 112165. <https://doi.org/10.3390/rs12030505>
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.; Tang, H. GEDI launches a new era of biomass inference from space. *Environ. Res. Lett.* 2022, 17(9), 095001. <https://doi.org/10.1088/1748-9326/ac8694>
22. Liang, X.; Kankare, V.; Hyyppä, J.; Wang, Y.; Kukko, A.; Haggren, H.; Yu, X.; Kaartinen, H.; Jaakkola, A.; Guan, F.; Holopainen, M.; Vastaranta, M. Terrestrial laser scanning in forest inventories. *ISPRS J. Photogramm.* 2016, 115, 63-77. <https://doi.org/10.1016/j.isprsjprs.2016.01.006>
23. Liang, X.; Kukko, A.; Balenović, I.; Saarinen, N.; Junttila, S.; Kankare, V.; Holopainen, M.; Mokroš, M.; Surový, P.; Kaartinen, H.; Jurjević, L.; Honkavaara, E.; Näsi, R.; Liu, J.; Hollaus, M.; Tian, J.; Yu, X.; Pan, J.; Cai, S.; Virtanen, J.P.; Wang, Y.; Hyyppä, J.; 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 (3), 32-71. <https://doi.org/10.1109/MGRS.2022.3168135>
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. <https://doi.org/10.1007/s00468-022-02382-1>



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 (2), 450. <https://doi.org/10.3390/rs15020450>
26. Chen, S.; Liu, H.; Feng, Z.; Shen, C.; Chen, P. Applicability of personal laser scanning in forestry inventory. *PLoS One.* 2019, 14 (2), e0211392. <http://dx.doi.org/10.1371/journal.pone.0211392>.
27. Ryding, J.; Williams, E.; Smith, M.; Eichhorn, M. Assessing handheld mobile laser scanners for forest surveys. *Remote Sens.* 2015, 7 (1), 1095–1111. <http://dx.doi.org/10.3390/rs70101095>
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 (9), 1509. <https://doi.org/10.3390/rs12091509>
29. Mokroš, M.; Mikita T.; Singh A.; Tomastik J.; Chudá J.; Wezyk P.; Kuželka, K.; Surový, P.; Klimánek, M.; Zięba-Kulawik, K.; Bobrowski, R.; Liang, X. Novel low-cost mobile mapping systems for forest inventories as terrestrial laser scanning alternatives. *Int. J. Appl. Earth Obs. Geoinf.* 2021, 104, 102512. <https://doi.org/10.1016/j.jag.2021.102512>.
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. Forest Sci.* 2023, 80, 40. <https://doi.org/10.1186/s13595-023-01209-4>
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 (10), 919–925. <https://doi.org/10.1071/WF20022>
32. Yao, H.; Qin, R.; Chen, X. Unmanned aerial vehicle for remote sensing applications-A review. *Remote Sens.* 2019, 11 (12), 1443. <https://doi.org/10.3390/rs11121443>
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 (8), 1541. <https://doi.org/10.3390/f14081541>
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 (2), 1146–1158. <https://doi.org/10.1016/j.asr.2021.10.049>
35. Adhikari, A.; Montes, C.R.; Peduzzi, A. A comparison of modeling methods for predicting forest attributes using Lidar metrics. *Remote Sens.* 2023, 15 (5), 1284. <https://doi.org/10.3390/rs15051284>
36. 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 (14), 3546. <https://doi.org/10.3390/rs15143546>
37. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)*. 1996, 73 (1), 267–288.
38. Laes, D.; Reutebuch, S.E.; McGaughey, R.J.; Mitchell, B. Guidelines to estimate forest inventory parameters from lidar and field plot data. 2011, Available at: [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)
39. U.S. Department of Agriculture. National Agriculture Imagery Program (NAIP). U.S. Department of Agriculture, Washington, D.C. 2023, <https://naip-usdaonline.hub.arcgis.com/> (Accessed May 2, 2024).
40. U.S. Geologic Survey. Topographic data quality levels (QLs). U.S. Geologic Survey, Reston, VA. 2024, <https://www.usgs.gov/3d-elevation-program/topographic-data-quality-levels-qls> (Accessed April 22, 2024).
41. McCullagh, M.J. Terrain and surface modelling systems: Theory and practice. *Photogramm. Rec.* 1988, 12 (72), 747–779. <https://doi.org/10.1111/j.1477-9730.1988.tb00627.x>
42. Hengl, T. Finding the right pixel size. *Comput. Geosci.* 2006, 32 (9), 1283–1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
43. 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 (5), 1064. <https://doi.org/10.3390/f14051064>
44. 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 (7), 985. <https://doi.org/10.3390/f13070985>
45. 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 (11), 2708. <https://doi.org/10.3390/rs14112708>
46. 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. Biogeo.* 2009, 114, G00E04. <https://doi.org/10.1029/2008JG000870>
47. 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.; Khosravipour, A. Imputation of individual longleaf pine (*Pinus*

- palustris* Mill.) tree attributes from field and LiDAR data. Can. J. Remote Sens. 2016, 42 (5), 554-573. <https://doi.org/10.1080/07038992.2016.1196582>
48. 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 (20), 5081. <https://doi.org/10.3390/rs14205081>
  49. 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. 2020, 169, 152-165. <https://doi.org/10.1016/j.isprsjprs.2020.09.001>

**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.