Using Airborne and Spaceborne lidar for Mapping Forest Attributes in Southern Alabama.

by

Schyler B. Brown

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of Master of Science

Auburn, Alabama
December 10, 2022

Keywords: Forestry, lidar, multispectral, biomass, volume, basal area

Approved by

Lana Narine, Chair, Assistant Professor of Geospatial Analytics
Todd Steury, Associate Dean of Academic Affairs, Associate Professor of Wildlife Biology and Statistics
Becky Barlow, Associate Dean for Extension and Assistant Director for Agriculture, Forestry, and Natural Resource Extension Programs

ABSTRACT

Remote sensing has become a major resource for assessing ecosystems and global carbon cycling. In forest applications, remote sensing data have been used to estimate the amount of carbon stored or lost in forests due to fire, deforestation, and reforestation. In particular, airborne light detection and ranging (lidar) has shown practical effectiveness in estimating forest variables over greater spatial extents. Using models developed from an association of field inventory samples with lidar data, attributes such as basal area, volume, and aboveground biomass can be predicted over larger tracts of forests, allowing for wall-to-wall mapping. Spaceborne lidar can also collect elevation data at broader scales and may be used to efficiently characterize forest structure over an even larger area of land. This work consisted of two main objectives: (1) to develop wall-to-wall maps of basal area, volume and aboveground biomass using airborne lidar for the Solon Dixon Forestry Education Center in southern Alabama, and (2) to investigate the Global Ecosystem Dynamics Investigation (GEDI) lidar for predicting forest structure over a larger spatial extent. The reference maps from chapter 2 had modest accuracy (R^2 = 0.36-0.53), with much of the unexplained model variation due to field data lacking complete vegetation. In the second study, models were developed from GEDI lidar with lower accuracy (R^2 = 0.30-0.36) and the maps produced from GEDI footprints and auxiliary imagery had low accuracy (R^2 = 0.10-0.11). The low accuracy was likely caused by a propagation of errors in the reference maps due to incomplete field inventory and could be further investigated. While lidar data are sufficient for forest mapping, texture variables derived from multispectral imagery are valuable, especially where lidar data is not spatially continuous.

ACKNOWLEDGEMENTS

I want to dedicate this document to a number of people who have made a profound impact on my success at Auburn University. To Catherine, thank you for being by my side through this arduous endeavor. Your constant enthusiasm for what I was studying helped keep me excited and enthusiastic throughout the project. To my parents, thank you for encouraging me to see this through. As a child, I was witness to, and was inspired by the dedication and hard work that put food on our table and made such opportunities as pursuing higher education possible. To my committee members, Dr. Stuery and Dr. Barlow, thank you for your incredible patience. I came to you both with so many questions, and you gave me everything and more with sincere kindness and enthusiasm. Finally, to Dr. Narine, I have so much to be grateful for. By accepting me as a student under your guidance, you have given me an opportunity to achieve something great. More importantly, you have taught me how to think critically, to be self-sufficient, and to be truly hardworking. This short paragraph will never be enough to convey my thanks to each of you here.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 1

1. BACKGROUND

1.1 FOREST MENSURATION

Forest measurements provide the foundation for understanding forest ecosystem dynamics, forest carbon cycling, and healthy commercial forest stewardship. Forest measurements, due to their relationships to conservation and commercial forest stewardship, became an integral part of forest management in the 18th century. The dynamic effects of natural systems on statically managed forests quickly led to the development of ecologically related forest stand measurements such as basal area (BA), volume, and aboveground biomass (AGB) [1,2]. In the following sections, the significance of BA, volume, and AGB will be described, and basic calculations will be presented to give context to the following studies.

A tree's BA is the cross sectional area of the tree stump at breast height and is the estimation of the amount of area a tree takes up in a given space [3]. Typically, BA is initially measured in individual trees in a sample plot and then multiplied by the plot size to get a total BA per acre value. For example, a single tree in a 0.04 hectare (ha) (1/10th acre) plot might have a BA of .05 m^2 (0.54 ft^2). This can then be multiplied by the plot size denominator of 25 (0.04 as a fraction equals 1/25) to get a total BA of 1.25 m^2 ha^{-1} (5.45 ft^2 $acre^{-1}$). Next, this value would be added to all of the other total per ha BA tree

measures, to get total BA in a hectare [3]. The importance of basal area in forestry and ecology cannot be expressed enough. BA can help scientists and foresters get a better understanding of the quality of a habitat, regeneration needs, and when to fell and plant trees. Previous studies have used BA to calculate total wooded area nationally and globally in order to produce management plans for effective forest conservation [4]. For example, researchers in Bangladesh [5], and in Europe [6] found a strong relationship between BA and carbon stores, where BA could be used as an independent variable in models measuring biomass for carbon stock analysis. Another researcher used BA as a factor to determine the growth and recruitment of longleaf pines [7]. Other studies have used BA to help determine extent and preference of habitat needed for various endangered species [8,9]. BA has also been used to measure live and dead wooded area due to invasive species [10,11].

Volume is another important forest metric related to the carbon cycle, habitat quality, and lumber production in forest management. Forest inventory data for productive stands have largely relied on measurements of merchantable volume to help compute costs of standing timber in a forest. Furthermore, volume is key to planning reforestation, estimating fuel load, and collecting data on biomass for carbon storage [12,13]. Timber volume in forest management is commonly measured in board feet. Board feet refers to the number of boards in one tree, and most inventories use the international ¼ rule to measure board feet, as it is known to be one of the most accurate

measures of volume due to its ability to account for taper in a tree [14]. In some forest inventories, particularly those where forest volume is measured for applications beyond merchantable timber, regression-based equations with species specific predictor coefficients might be used to determine a tree's volume [15,16]. These types of equations can be used to estimate the total volume of forest trees or merchantable volume based on timber specifications.

AGB is a highly relevant forest metric that is crucial for determining the exchange of carbon between forests, soils, and the atmosphere. AGB is the amount of dry weight carbon in a tree, and is directly measured by destructive sampling, where a tree is felled and incinerated until what is left is a carbon sample that can be weighed. However, AGB is often estimated with regression-based models with species specific predictor coefficients [17]. Regression-based allometric equations have been developed where carbon is set as a dependent variable, and diameter at breast height (DBH) and tree height are variables whose coefficients are solved for. The result of a model predicting AGB is commonly measured in megagrams per hectare (Mg ha^{-1}) and to convert the result to carbon, the prediction is usually multiplied by a conversion factor of 0.5 [18,19]. AGB estimation has become a dominating topic in many scientific fields, as understanding carbon exchange in forests is vital for mitigating climate change. Many nations now require their scientific and resource departments to estimate AGB in their forest stands, and international committees such as the Paris Climate Accords have

prioritized forest-based carbon estimation in their action plans [20,21]. Because of the urgency for estimating forest structure, many governments and business sectors have investigated the use of remote sensing (RS) data, particularly light detection and ranging (lidar) as it can gather great swaths of 2-dimensional and 3-dimensional forest structure data.

1.2 LIDAR

Estimating forest inventory data such as BA, volume, and AGB for large areas of forests can often be both time consuming and costly [18,22,23]. Furthermore, inaccuracies in forest inventory data can occur depending on the experience of the forester [24], and extrapolation based on forest inventory plots can cause inaccuracies due to site specific geographic differences across a swath of land. In an ideal world, conducting forest inventories would be highly accurate while accruing little to no cost. Lidar data has been shown to minimize the tradeoff between time/cost and accuracy, with minimal bias [25,26], and has shown success in forestry applications, particularly when combining lidar returns measuring height with stand measurements such as BA, volume, and AGB [11,27]. Lidar systems emits laser pulses over the desired area being measured, and records the time and distance the photons traveled from the ground and back to the apparatus to compute heights of various objects [28–30]. The earliest uses of lidar technologies began in the 1960's where the goal was to obtain high accuracy digital terrain models (DTM) for geographic and geological studies; above ground

vegetation returns were considered noise and typically removed [31]. In the 1980's, researchers began using tree return points previously classified as noise, to classify heights of canopies [32]. From here, lidar in forestry applications exploded, and further methods to measure stand metrics were developed throughout 1980's using height and canopy metrics derived from lidar. In 1982, researchers demonstrated that there was a very strong relationship between basal area and volume to lidar derived height metrics, especially maximum canopy height related to biomass [33]. Other researchers showed that total forest biomass and volume could be calculated within 2% of the actual mean biomass or volume [34]. Improvements in acquiring accurate data improved throughout the 90's and early 2000's as researchers began experimenting with variation in flight specifications (height from ground, overlap of flight lines, etc.) [35]. As the technology and resources for acquiring and using lidar become both more advanced and more available, high-resolution mapping large areas of forests can lead to more accurate forest measurements such as volume and BA [11,28].

Lidar remote sensing applications in forestry have used terrestrial, spaceborne and airborne lidar measurement devices [30,36]. However, many projects often use airborne lidar, which uses a lidar system attached to an aircraft to scan forests in tracts [37]. Collecting data in tracts is highly advantageous because it allow researchers to gather large amounts of continuous data in a specific location with varying degrees of accuracy, overlap, and costs [26,30,38]**.** While airborne lidar has a range of accuracy and

costs, high resolution data are becoming more available and more affordable, such as the free datasets from United States Geological Survey (USGS) 3D Elevation Program (3DEP) [39]. Future developments from the USGS include implementing a lidar collection device to NAIP imagery missions to meet both temporal and resolution requirements by programs such as the Forest Inventory Analysis (FIA) [11]. However, until then, there will be downfalls of airborne lidar that are primarily concerned with the fact that the device collecting the data is attached to an aircraft whose limiting factor is fuel [11]. In other words, only so much lidar data can be collected in one flight, and it often takes many flights to acquire large amounts of data, resulting in potential changes in forest characteristics over the time it takes to acquire usable data [40]. Furthermore, airborne lidar systems are limited by their geometry, velocity, and flights must be done in relatively descent weather [41]. This is where the potential of spaceborne lidar becomes valuable. Spaceborne lidar collects elevation measurements over larger swaths of area because the system is mounted to a satellite that orbits the earth, and data can be viewed in many spatial scales [11,41,42]. Spaceborne lidar also typically collects tracts with more than one laser beam and the tract path is continuous and linear, defined by the orbit of the satellite [23,43,44]. The potential for spaceborne lidar is great when large datasets are needed, but falls short when specific areas need to be measured outside of the satellites path, unless models can be derived from reference maps and further interpolated or extrapolated across an extent using auxiliary data [11]. Furthermore, to

the knowledge of the authors, no space-based lidar model has been developed specifically for estimating merchantable volume and AGB across an entire extent. The following literature review will discuss the progress made in mapping BA, volume, and AGB using airborne and space-based lidar, and how the studies produced in this thesis address gaps in the literature.

2. LITERATURE REVIEW

Forest applications of airborne lidar are well established, and particularly useful for estimating forest metrics in homogenous species forest stands and forest plantations. For example, an early study in forests dominated by Western Hemlock (*Tsuga heterophylla*) and Douglas Fir (*Pseudotsuga menziesii*) in the Western U.S., researchers estimated canopy heights, BA, AGB, and leaf biomass with coefficients of determination of 0.95, 0.96, 0.96, and 0.84 respectively [45]. Typical of many studies estimating forest metrics, this study used a linear regression approach with lidar height variables which explained most of the variation in BA and AGB. In estimating timber volume in Spruce (*Picea abies*) stands in Norway, researchers demonstrated promising results using a linear regression model developed from the mean height of lidar returns [13]. The linear regression approach was the standard for much of the early 2000's, and ultimately this led many researchers to experiment with different variables in their models for predicting forest metrics. Typically, a height variable, a canopy cover variable, and a variable describing the density or variation in the canopy and understory are sufficient

for modeling forest metrics such as BA, volume and AGB [46]. However, in some cases, other variables are needed to explain forest structure, such as crown structure imputed from individual tree detection [47] or height percentiles which have become a standard for forest structure estimation [37,48,49]. More recently, researchers have been leveraging non-parametric machine learning algorithms such as Random Forests [50] to map BA, volume and AGB. These types of modeling techniques, particularly Random Forests, have been favored recently for their ability to deal with collinearity and variance inflation, and are also known for having less bias [51] and prediction accuracy comparable or better than linear regression modeling [52,53]. For example, one study compared the accuracy of multiple linear regression, Random Forests, Gradient Nearest Neighbor, and Independent Component Analysis for estimating tree density and BA in North-Central Idaho and found that the Random Forests algorithm outperformed the other algorithms ($R^2 = 0.78$) [51]. In estimating forest structure for tree species most similar to those in this thesis, researchers developed models for BA and volume using the Random Forests algorithm, and airborne lidar in Longleaf Pine (*Pinus palustrus*) and Loblolly Pine (*Pinus taeda*) plantations in Brazil and had very high accuracies ($R^2 = 0.93$ and 0.97 respectively) [54,55]. The most important variables in these random forests models are typically just two to three variables such as canopy height, canopy cover, and a variable that describes the variation in forest density [56,57]. On the flip side, a number of studies have shown that ALS estimation of forest structure is difficult when

forest stands are highly variable in structure and species composition. For example, in deciduous mixed forests of Ontario Canada, researchers estimated BA with much poorer results than those in other forest types ($R^2 = 0.29$), reporting that the correlation between BA and tree heights varies particularly in forests where species composition is very high [58]. Other studies have shown comparable results in mixed species forests. An early study in the Southern Georgia used lidar to estimate AGB and volume with modest results ($R^2 = 0.52$ and 0.55 respectively) noting that high variation in site specific growth influencers had a strong effect on the diameter of trees measured, and that variation in species composition introduced error into the models [34]. Another study estimated biomass over dense tropical forests, and reported that when their model was tested on an independent dataset, their model's predictive capabilities were poor [59]. While authors did not report the statistical results of the model, another study attempted to validate their own estimates with those of the previous study and found poor comparisons ($R^2 = 0.28$) [60]. The difficulty in estimating metrics in these diverse forests necessitates the exploration of additional variables that can help improve forest modeling.

Multispectral imagery has shown varied success in estimating BA, volume and AGB in forest stands and variables from this type of remote sensing application are numerous. A study in Georgia and Alabama used a machine learning approach to estimate BA in pine stands using texture metrics derived from multispectral imagery and found

modest results [61]. Again, however, the authors of this study noted that many of the errors associated with their models were due to the forest stands being composed of a high diversity of species. In another instance, the same researchers compared predictive models of lidar variables to those of multispectral imagery and found that the multispectral imagery models produced comparable results to the lidar based models [26]. These conclusions draw attention to a possibility for modeling forest structure using both lidar and multispectral imagery variables to improve estimates of forest structure in forest stands in southern Alabama. If such a study can produce more accurate predictive models of forest structure, then higher quality reference maps can be developed for upscaling forest structure estimates across greater extents using space-based lidar and imagery.

Although airborne lidar has proven to be highly effective in mapping forest metrics, space-based lidar has the ability to quickly gather data across a greater distance [30]. Using spaceborne lidar in this way is relatively new when compared to methods using airborne and terrestrial lidar, but has shown effectiveness in estimating forest metrics, and a number of approaches have been demonstrated for estimating forest structure using space-based lidar. Estimating forest metrics within the boundary of the laser spot on the Earth's surface (otherwise known as a footprint) along the satellite's track is the most common method, and usually produces highly accurate results. For example, estimates of AGB across the continental U.S. and Mexico using spaceborne lidar and

field data were able to estimate aboveground biomass within 1% accuracy of the field estimated biomass [22], though this method did not produce geographically continuous estimates, but instead predicted AGB within the pulses gathered by the space-based lidar system. Similarly, a study in East Texas compared footprint level estimates of AGB between satellite lidar and airborne lidar and found a very strong correlation between the two ($r = 0.98$). However, effective mapping requires more data than footprint level estimates [62]. Another method for estimating forest structure is to use the footprint-level estimates and extrapolate them to produce a continuous map. This approach usually involves the use of ancillary data such as imagery or radar, whereby space-based lidar estimates of forest structure can be related to imagery or radar values and further upscaled across the extent of the study area. Estimates of AGB and volume in these studies have varied results. For instance, a study estimating AGB in East Texas using airborne lidar, satellite lidar and imagery data found modest to good results in forest mapping, and noted that if field inventory had been available to calibrate the satellite model, higher accuracy could be achieved [63]. A three-phase approach where mapping is achieved via synthesis of field inventory based airborne lidar models, satellite lidar, and imagery could achieve better accuracy. One study using this approach demonstrated good accuracy (R^2 up to 0.76) in estimating AGB in North America, however the resolution of their outputs was somewhat coarse (250 m) [22]. Furthermore, most studies using this three-phase approach do not estimate commercial

forest metrics based on the cruise specifications that most forestry companies do. In a study done in commercial stands, merchantable timber measurements were used to attempt modeling biomass [59], however, another researcher attempting to validate this map pointed out later that the map was highly inaccurate due to measurements of biomass being derived from merchantable heights, and weighing models heavily on BA [60] (though researchers have found that biomass estimations should use BA measures that are closely correlated to height metrics, which can be difficult in mixed species stands [58]). These researchers further point out that highly accurate field data acquired by means of consistent mensuration techniques are needed in order to model forest characteristics from spaceborne lidar, and that caution should be taken when measuring forest structures from merchantable measures due to a variety of biases and errors associated with reference maps [60]. When data for global ecosystem science is desired, or large amounts of forest mensuration data are required for forestry production, spaceborne lidar has the potential to save massive amounts of time in data collection and collaboration efforts [30]**.** Furthermore, new space-based lidar platforms offer products that may be useful for calibrating merchantable forest models (these products will be touched on in the coming sections). Two notable spaceborne lidar missions include the Ice Cloud and land Elevation Satellite (ICESat) and ICESat-2 missions.

Launched in 2003, ICESat was developed with the objective of measuring changes in ice thickness, elevation and distribution. Aboard ICESat was the Geoscience Laser

Altimeter (GLAS) which was a lidar system with three lasers that fired 1064 nm pulses, operating one laser at a time, and a nominal footprint of about 65 meters [64]. Researchers have demonstrated the ability of GLAS for measuring vegetation [64,65]. For example, the ability of ICESat to estimate volume has been demonstrated over large extents of Siberia [66], and in other work ICESat data has been used to measure biomass and canopy cover [22]. While useful for estimating the cryosphere, ICESat was ultimately not optimized for forest measurements. For example, it's large footprint and low pulse density resulted in inaccuracies and the inability to map at higher resolutions. ICESat finished its mission in 2009, after a series of complications with GLAS, leading to the development of ICESat-2 [65].

ICESat-2 was launched in 2018, largely with the objective of continuing the ICESat mission [65]. Unlike ICESat however, the lidar device aboard ICESat-2 were upgraded. Instead of one laser being used at a single time with a 65-meter footprint, ICESat-2 has six beams, producing a < 17-meter footprint. The six beams are produced in pairs, allowing for greater measures in the change of elevation and spatial coverage [65]. Researchers can use this finer resolution to produce more accurate maps of forest attributes. For example, researchers have used ICESat-2 data to estimate aboveground biomass (ABG) and canopy cover in East Texas forests [42], and researchers in Northeastern China have used a machine learning approach to estimate canopy heights [67]. Data collection from spaceborne lidar is new however, with satellites such as

ICESat-2 being launched in just 2018, so there is a need to develop more methods with the ability to properly implement spaceborne lidar datasets into forest ecosystem science and forest management [11]. Furthermore, like the first ICESat mission, ICESat-2 is optimized for estimating properties of the cryosphere.

Another spaceborne lidar mission, Global Ecosystem Dynamics Investigation (GEDI), has also recently launched in 2018 with available data being published in 2020 [43]. GEDI lidar data currently contains the greatest density of footprints of any spaceborne lidar device, allowing users even greater accuracy in their predictive models and mapping at finer resolutions [43,68]. Furthermore, GEDI offers products that may be useful for calibrating models estimating merchantable volume and AGB, such as the L4A footprint level aboveground biomass density product (AGBD) [69]. A potential method for acquiring reference estimates of forest measurements using GEDI lidar could be a three step method where 1: forest measurements are collected on the ground, 2: airborne lidar height and canopy data is gathered, and by means of regression models (created from lidar and ground data), maps estimating forest measures are developed, and 3: spaceborne estimates developed from regression models (created from airborne predictive reference estimates) and extrapolated using imagery data [70]. At the time of this writing, there are few publications that have developed models predicting AGB using a three-phase approach, and to the knowledge of the authors, there are no publications that have mapped merchantable timber volume using data

acquired from GEDI. This study will use plot inventory data combined with airborne lidar data to develop a predictive reference map by means of linear regression modeling that can facilitate the development of prediction models for large-scale forest management AGB and volume using data collected from GEDI. While Sentinel-2 imagery will be used to delineate vegetation and non-vegetation as a processing step, vegetation indices will be considered for modeling.

3. OBJECTIVES

This study occurred in southern Alabama for an extent across the Solon Dixon Forestry Education Center SDFEC in the southern United States. The SDFEC is a research and education station managed by Auburn University used to train students in forestry and wildlife sciences. Located in the Southern Plains ecoregion, the SDFEC is dominated by mixed pine forests and a humid sub-tropical climate. The aim of this study was to develop predictive reference maps of AGB, volume and basal area in the SDFEC. This was achieved using open source airborne lidar, and the maps will be used to further investigate the predictive abilities of GEDI for further estimating merchantable volume, and AGB. The two primary objectives were to:

1. Develop reference maps of BA, volume and AGB within the SDFEC study site.
2. Evaluate the potential of GEDI spaceborne lidar for forest mapping of AGB and volume.

Objectives 1 and 2 are described in further detail in Chapters 2 and 3 respectively. Objective 1, the development of predictive reference maps using airborne lidar, is outlined in Chapter 2. Likewise, Objective 2, the evaluation of GEDI lidar for mapping at greater extents is presented in Chapter 3. These studies are intended to be stand-alone studies, one of which has been published in a peer-reviewed journal (*Remote Sensing*), and the other is under preparation for submission to a peer-reviewed journal publication. Finally, the conclusions of this study are presented in Chapter 4 where the significance of the data from both studies is discussed. A discussion within the conclusion examines pitfalls, limitations, and future work needed to expand and contribute to this field of study.

4.LITERATURE CITED

1. Heilbron, J. Introductory Essay. In *The Quantifying Spirit of the Eighteenth Century*; University of California Press, 1990; pp. 1–24.
2. Nocentini, S.; Ciancio, O.; Portoghesi, L.; Corona, P. Historical Roots and the Evolving Science of Forest Management under a Systemic Perspective. *Can. J. For. Res.* **2020**, *171*, 1–9, doi:10.1139/cjfr-2020-0293.
3. Elledge, J.; Barlow, B. Basal Area: A Measure Made for Management. *Alabama Coop. Ext. Syst.* **2010**, 6.
4. Lugten, G. Food and Agriculture Organization. *Int. J. Mar. Coast. Law* **2020**, *23*, 761–767, doi:10.1163/157180808X353939.
5. Khan, M.N.I.; Islam, M.R.; Rahman, A.; Azad, M.S.; Mollick, A.S.; Kamruzzaman, M.; Sadath, M.N.; Feroz, S.M.; Rakkibu, M.G.; Knohl, A. Allometric Relationships of Stand Level Carbon Stocks to Basal Area, Tree Height and Wood Density of Nine Tree Species in Bangladesh. *Glob. Ecol. Conserv.* **2020**, *22*, doi:10.1016/j.gecco.2020.e01025.
6. Babst, F.; Bouriaud, O.; Alexander, R.; Trouet, V.; Frank, D. Toward Consistent Measurements of Carbon Accumulation: A Multi-Site Assessment of Biomass and Basal Area Increment across Europe. *Dendrochronologia* **2014**, *32*, 153–161, doi:10.1016/j.dendro.2014.01.002.
7. Kara, F.; Loewenstein, E.F.; Brockway, D.G. Effects of Basal Area on Survival and Growth of Longleaf Pine When Practicing Selection Silviculture. *For. Syst.* **2017**, *26*, 1–12, doi:10.5424/fs/2017261-10043.
8. Weiss, S.A.; Toman, E.L.; Corace, R.G. Aligning Endangered Species Management with Fire-Dependent Ecosystem Restoration: Manager Perspectives on Red-Cockaded Woodpecker and Longleaf Pine Management Actions. *Fire Ecol.* **2019**, *15*, doi:10.1186/s42408-019-0026-z.
9. U.S. Fish and Wildlife Service *Recovery Plan for the Red-Cockaded Woodpecker (Picoides Borealis): Second Revision*; 2003;
10. Vogt, J.; Koch, F.H. Forest Inventory and Analysis Data in Invasive Insect Research. *Am. Entomol.* **2016**, *62*, doi:10.1093/ae/tmv072.
11. Lister, A.J.; Andersen, H.; Frescino, T.; Gatziolis, D.; Healey, S.; Heath, L.S.; Liknes, G.C.; McRoberts, R.; Moisen, G.G.; Nelson, M.; et al. Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case

Study from the United States National Forest Inventory. *Forests* **2020**, *11*, 1–41, doi:10.3390/f11121364.

12. Chirici, G.; Giannetti, F.; McRoberts, R.E.; Travaglini, D.; Pecchi, M.; Maselli, F.; Chiesi, M.; Corona, P. Wall-to-Wall Spatial Prediction of Growing Stock Volume Based on Italian National Forest Inventory Plots and Remotely Sensed Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *84*, 101959, doi:10.1016/j.jag.2019.101959.

13. Næsset, E. Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data. *Remote Sens. Environ.* **1997**, *61*, 246–253, doi:10.1016/S0034-4257(97)00041-2.

14. Bond, B. PB1650 Understanding Log Scales and Log Rules. *UT Ext. Publ. For. Trees, Timber* **2011**, 8.

15. Phillips, D.R.; Frederick, D.J. Weight, Volume and Physical Properties of Major Hardwood Species in the Piedmont. *USDA For. Serv.* **1986**.

16. Oswalt, C.M.; Conner, R.C. Southern Forest Inventory and Analysis Volume Equation User ' s Guide. **2011**.

17. Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsey, R.A. National-Scale Biomass Estimators for United States Tree Species. *For. Sci.* **2003**, *49*, 12–35, doi:10.1093/forestscience/49.1.12.

18. Houghton, R.A. Aboveground Forest Biomass and the Global Carbon Balance. *Glob. Chang. Biol.* **2005**, *11*, 945–958, doi:10.1111/j.1365-2486.2005.00955.x.

19. Duncanson, L.; Armston, J.; Disney, M.; Avitabile, V.; Barbier, N.; Calders, K.; Carter, S.; Chave, J.; Herold, M.; MacBean, N.; et al. *Aboveground Woody Biomass Product Validation Good Practices Protocol*; 2021;

20. UNFCCC *The Paris Climate Agreement*; United Nations, 2016; Vol. 2, pp. 78–79;.

21. UNFCCC Key Decisions Relevant for Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+). *Framew. Conv. Clim. Chang.* **2016**, 44.

22. Nelson, R.; Margolis, H.; Montesano, P.; Sun, G.; Cook, B.; Corp, L.; Andersen, H.E.; deJong, B.; Pellat, F.P.; Fickel, T.; et al. Lidar-Based Estimates of Aboveground Biomass in the Continental US and Mexico Using Ground, Airborne, and Satellite Observations. *Remote Sens. Environ.* **2017**, *188*, 127–140, doi:10.1016/j.rse.2016.10.038.

23. Neuenschwander, A.; Pitts, K. Ice, Cloud, and Land Elevation Satellite 2 (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-

Track Products (ATL08). **2020**, *5*, 1–140.

24. Elzinga, C.; Shearer, R.C.; Elzinga, G. Observer Variation in Tree Diameter Measurements. *West. J. Appl. For.* **2005**, *20*, 134–137, doi:10.1093/wjaf/20.2.134.

25. Mcroberts, R.E.; Tomppo, E.O. Remote Sensing Support for National Forest Inventories. **2007**, *110*, 412–419, doi:10.1016/j.rse.2006.09.034.

26. Ahl, R.; Hogland, J.; Brown, S. A Comparison of Standard Modeling Techniques Using Digital Aerial Imagery with National Elevation Datasets and Airborne LiDAR to Predict Size and Density Forest Metrics in the Sapphire Mountains MT, USA. *ISPRS Int. J. Geo-Information* **2019**, *8*, doi:10.3390/ijgi8010024.

27. Means, J.E.; Acker, S.A.; Fitt, B.J.; Renslow, M.; Emerson, L.; Hendrix, C.J. Predicting Forest Stand Characteristics with Airborne Scanning Lidar. *Photogramm. Eng. Remote Sensing* **2000**, *66*, 1367–1371.

28. Okyay, U.; Telling, J.; Glennie, C.L.; Dietrich, W.E. Airborne Lidar Change Detection: An Overview of Earth Sciences Applications. *Earth-Science Rev.* **2019**, *198*, 102929, doi:10.1016/j.earscirev.2019.102929.

29. USGS *Lidar-Base-Specification-2020-Rev-A.Pdf*; 2020;

30. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar Sampling for Large-Area Forest Characterization: A Review. *Remote Sens. Environ.* **2012**, *121*, 196–209, doi:10.1016/j.rse.2012.02.001.

31. Nelson, R.F. How Did We Get Here ? An Early History of Forestry Lidar 1. **2015**, doi:10.5589/m13-011.

32. Burns, J.P. Mapping in Tropical Forests : A New Approach Using the Laser APR The Laser Airborne Profile Recorder Was Used Not Only For. **1982**.

33. Hagel, H.J.; Femppel, J.; Lux, G.; Demling, L. Einsatz Einer Mobilen Video-Anlage Für Die Ausbildung Der Studenten in Der Inneren Medizin. *Fortschr. Med.* **1982**, *100*, 307–308.

34. Nelson, R. Estimating Forest Biomass and Volume Using Airborne Laser Data. **1988**, *267*, 247–267.

35. Reutebuch, S.E.; McGaughey, R.J. LIDAR: An Emerging Tool for Multiple Resource Measurement, Planning and Monitoring. *West. For.* **2008**, *53*, 1–5.

36. Dassot, M.; Constant, T.; Fournier, M. The Use of Terrestrial LiDAR Technology in Forest Science: Application Fields, Benefits and Challenges. *Ann. For. Sci.* **2011**,

68, 959–974, doi:10.1007/s13595-011-0102-2.

37. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar Remote Sensing for Ecosystem Studies. *Bioscience* **2002**, *52*, 19–30, doi:10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2.

38. Nelson, R.; Parker, G.; Hom, M. A Portable Airborne Laser System for Forest Inventory. *Photogramm. Eng. Remote Sensing* **2003**, *69*, 267–273, doi:10.14358/PERS.69.3.267.

39. USGS 3DEP by the Numbers Available online: https://www.usgs.gov/core-science-systems/ngp/3dep/3dep-numbers.

40. Wulder, M.A.; Bater, C.W.; Coops, N.C.; Hilker, T.; White, J.C. The Role of LiDAR in Sustainable Forest Management. **2008**, *84*.

41. Canada, G. of Airborne Versuse Spaceborne Radars Available online: https://www.nrcan.gc.ca/maps-tools-publications/satellite-imagery-air-photos/remote-sensing-tutorials/microwave-remote-sensing/airborne-versus-spaceborne-radars/9397.

42. Narine, L.L.; Popescu, S.; Neuenschwander, A.; Zhou, T.; Srinivasan, S.; Harbeck, K. Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sens. Environ.* **2019**, *224*, 1–11, doi:10.1016/j.rse.2019.01.037.

43. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002, doi:10.1016/j.srs.2020.100002.

44. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Estimation with GEDI LIDAR Waveforms and Bayesian Deep Learning. **2021**.

45. Means, J.E.; Acker, S.A.; Harding, D.J.; Blair, J.B.; Lefsky, M.A.; Cohen, W.B.; Harmon, M.E.; McKee, W.A. Use of Large-Footprint Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sens. Environ.* **1999**, *67*, 298–308, doi:10.1016/S0034-4257(98)00091-1.

46. Laes, D.; Reutebuch, S.E.; Mcgaughey, R.J.; Mitchell, B. *Guidelines to Estimate Forest Inventory Parameters from Lidar and Field Plot Data*; 2011;

47. Popescu, S.C.; Wynne, R.H. Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height. *Photogramm. Eng. Remote Sensing* **2004**, *70*, 589–604,

doi:10.14358/PERS.70.5.589.

48. Næsset, E. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. **2002**, *80*, 88–99.

49. Gobakken, T.; Næsset, E. Estimation of Diameter and Basal Area Distributions in Coniferous Forest by Means of Airborne Laser Scanner Data. *Scand. J. For. Res.* **2004**, *19*, 529–542, doi:10.1080/02827580410019454.

50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:doi.org/10.1023/A:1010933404324.

51. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Hall, D.E.; Falkowski, M.J. Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245, doi:10.1016/j.rse.2007.10.009.

52. Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of Sample Size, Data Type and Prediction Method for Remote Sensing-Based Estimations of Aboveground Forest Biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114, doi:10.1016/j.rse.2014.07.028.

53. Chirici, G.; Giannetti, F.; McRoberts, R.E.; Travaglini, D.; Pecchi, M.; Maselli, F.; Chiesi, M.; Corona, P. Wall-to-Wall Spatial Prediction of Growing Stock Volume Based on Italian National Forest Inventory Plots and Remotely Sensed Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *84*, 101959, doi:10.1016/j.jag.2019.101959.

54. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Jaafar, W.S.W.M.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting Stem Total and Assortment Volumes in an Industrial Pinus Taeda L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* **2017**, *8*, 1–17, doi:10.3390/f8070254.

55. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Fennema, S.J.; Corte, A.P.D. Modeling and Mapping Basal Area of Pinus Taeda L. Plantation Using Airborne LiDAR Data. *An. Acad. Bras. Cienc.* **2017**, *89*, 1895–1905, doi:10.1590/0001-3765201720160324.

56. White, J.C.; Wulder, M.A.; Buckmaster, G. Validating Estimates of Merchantable Volume from Airborne Laser Scanning (ALS) Data Using Weight Scale Data. *For. Chron.* **2014**, *90*, 378–385, doi:10.5558/tfc2014072.

57. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. *A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach*; 2013; Vol. 89;

ISBN 9781100223858.

58. Spriggs, R.A.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A Critique of General Allometry-Inspired Models for Estimating Forest Carbon Density from Airborne LiDAR. *PLoS One* **2019**, *14*, 1–19, doi:10.1371/journal.pone.0215238.

59. Baccini, A.; Laporte, N.; Goetz, S.J.; Sun, M.; Dong, H. A First Map of Tropical Africa's above-Ground Biomass Derived from Satellite Imagery. *Environ. Res. Lett.* **2008**, *3*, doi:10.1088/1748-9326/3/4/045011.

60. Mitchard, E.T.A.; Saatchi, S.S.; Lewis, S.L.; Feldpausch, T.R.; Gerard, F.F.; Woodhouse, I.H.; Meir, P. Comment on "A First Map of Tropical Africa's above-Ground Biomass Derived from Satellite Imagery." *Environ. Res. Lett.* **2011**, *6*, doi:10.1088/1748-9326/6/4/049001.

61. Hogland, J.; Anderson, N.; St. Peter, J.; Drake, J.; Medley, P. Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *ISPRS Int. J. Geo-Information* **2018**, *7*, doi:10.3390/ijgi7040140.

62. Popescu, S.C.; Zhao, K.; Neuenschwander, A.; Lin, C. Satellite Lidar vs. Small Footprint Airborne Lidar: Comparing the Accuracy of Aboveground Biomass Estimates and Forest Structure Metrics at Footprint Level. *Remote Sens. Environ.* **2011**, *115*, 2786–2797, doi:10.1016/j.rse.2011.01.026.

63. Narine, L.L.; Popescu, S.C.; Malambo, L. Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example. *Remote Sens.* **2020**, *12*, 1–16, doi:10.3390/rs12111824.

64. Schutz, B.E.; Zwally, H.J.; Shuman, C.A.; Hancock, D.; DiMarzio, J.P. Overview of the ICESat Mission. *Geophys. Res. Lett.* **2005**, *32*, 1–4, doi:10.1029/2005GL024009.

65. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D.; et al. The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation. *Remote Sens. Environ.* **2017**, *190*, 260–273, doi:10.1016/j.rse.2016.12.029.

66. Nelson, R.; Ranson, K.J.; Sun, G.; Kimes, D.S.; Kharuk, V.; Montesano, P. Estimating Siberian Timber Volume Using MODIS and ICESat/GLAS. *Remote Sens. Environ.* **2009**, *113*, 691–701, doi:10.1016/j.rse.2008.11.010.

67. Li, W.; Niu, Z.; Shang, R.; Qin, Y.; Wang, L.; Chen, H. High-Resolution Mapping of Forest Canopy Height Using Machine Learning by Coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**,

92, 102163, doi:10.1016/j.jag.2020.102163.

68. GEDI Instrument Overview-GEDI Available online: https://gedi.umd.edu/instrument/instrument-overview/.

69. Dubayah, R.O.; Armston, J.; Kellner, J.R.; Duncanson, L.; Healey, S.P.; Patterson, P.L.; Hancock, S.; Tang, H.; Hofton, M.A.; Blair, J.B.; et al. *GEDI L4A Footprint Level Aboveground Biomass Density Version 1*; Oak Ridge, Tennessee, 2021;

70. Margolis, H.A.; Nelson, R.F.; Montesano, P.M.; Beaudoin, A.; Sun, G.; Andersen, H.E.; Wulder, M.A. Combining Satellite Lidar, Airborne Lidar, and Ground Plots to Estimate the Amount and Distribution of Aboveground Biomass in the Boreal Forest of North America. *Can. J. For. Res.* **2015**, *45*, 838–855, doi:10.1139/cjfr-2015-0006.

71. Matney, T.G.; Schultz, E.B. Mesavage and Girard Form Class Taper Functions Derived from Profile Equations. 77–85.

72. L3Harris Exelis Visual Information Solutions 2021.

CHAPTER 2

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

Brown Schyler, Narine L. Lana, Gilbert John, 2022, Remote Sensing, 14, 2708

https://doi.org/10.3390/rs14112708

1.INTRODUCTION

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

[7]. BA is also useful for examining the amount of woody area infested by various species of invasive insects across large areas [8]. Volume is a measure used by foresters to estimate the cubic amount of wood on an area of land [9,10]. An estimate of volume can give foresters an estimate of the dollar value of the standing timber they are cruising. Volume is also associated with the persistence of invasive species, and having estimates of volume can help foresters better manage both timber product and ecosystem health through stem and growth stocking projections [11]. AGB is the dry weight of carbon stored within forest trees above the ground and is measured in Mg ha^{-1} [12]. In ecological studies, AGB is a known driver of the species composition of an ecosystem, and maps of AGB can give ecologists insights into the distribution and composition of organisms across great extents [13]. Because AGB is an estimation of the carbon stored in the trunks of trees, it is critical to understand the role forests play in carbon cycling and climate change. Accurate estimates of AGB can help nations develop strategies to meet goals set by international agreements for climate change, such as those outlined in the Reducing Emissions from Deforestation section of the Paris Climate Agreement [14], and contributing action inventories such as the National Greenhouse Gas Inventories report [15]. Often it can be difficult or impossible to accurately estimate AGB without destructively sampling trees in a forest, and allometric equations developed from field inventory estimating BA, volume, and AGB can fail to account for variables that affect them across a large landscape, such as site index and

crowding, among other local factors [2,16]. Inconsistencies in forest inventory methods across the U.S. also cause inaccuracies in regional estimates, as inventory methods are frequently tailored for project-specific goals and thus have differing methods for measuring [17–19]. Lastly, estimates of forest measurements across large areas using traditional mensuration approaches can be time-consuming and costly, and in some instances, variation in landscape features can cause dangers in field work and inaccuracies in estimates [16,18,20–22]. Remote sensing (RS), particularly light detection and ranging (lidar) estimates of BA, volume, and AGB, may overcome these limitations and can produce comparable [23] and more accurate results [24] compared to field-based estimates of large-area forest inventory.

Lidar, an RS method that uses laser scanning to acquire three-dimensional information over the desired area [25], provides an alternative approach to estimating forest inventory metrics [1,21]. By using lidar, forest characteristics such as tree heights and canopy cover can be directly estimated, and other measures such as BA, volume, and AGB can be estimated by means of modeling using a variety of metrics derived from directly measured height and canopy cover [26,27]. The most commonly used approach to modeling BA, volume, and AGB from lidar data is by developing a multiple linear regression from field inventory and lidar variables [28]. Often the best subsets approach, such as forward or backward selection, is used to determine a model with variables that best explain the dependent variable and further validated using a cross-

validation or set validation approach [27,29,30]. These models usually contain at least one variable explaining tree height, another explaining canopy cover, and another that accounts for variation in the data, such as a height standard deviation variable [6,31,32]**;** however, the variables in these models may differ depending on the study site and the foliage type being measured [28,33,34]. Estimates of BA, volume, and AGB can also be acquired from non-parametric machine learning approaches such as random forest (RF), a machine learning algorithm that uses random and iterative samples of the data to produce regression trees and bootstraps data for robust predictive models [35–37].

Estimating forest metrics may be performed by using an area-based approach or a tree-level approach, and previous studies modeling BA, volume, or AGB have seen success in both approaches. For example, in estimating forest attributes from individual trees, researchers found that BA, volume, and tree heights of longleaf pine in Georgia, U.S., could be estimated with accuracies of $r^2 = 0.18$, 0.94, and 0.96, respectively [38]. The results for estimating volume and tree heights of individual trees are promising, and the poor results for estimating BA were explained by the loss of height-diameter allometry in southern pines above 25 m. In another study, researchers estimated BA for a plantation of loblolly pines (*Pinus taeda*) (another southern pine species) using an area-based approach and were able to achieve accuracies of $r^2 = 0.97$, and noted that the highly homogeneous environment of the pine plantations that the trees were grown in likely led to such high accuracies [27]. Using an area-based approach across a large area

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

in a southern mixed-species forest using open-source RS data. This goal was achieved by meeting the following objectives:

A. Determine whether the addition of variables derived from multispectral imagery to models previously including only lidar derived variables can improve estimates of BA, volume, and AGB;

B. Determine what variables are useful for modeling each forest metric (BA, volume, AGB) for Southern mixed forests.

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

2.MATERIALS AND METHODS

2.1STUDY AREA

The study area is a 35.35 km^2 site in Covington and Escambia counties, Southern Alabama (Figure 1). The site is within one of seven-level III ecoregions in Alabama, known as the Southern Plains ecoregion (ecoregion 65), whose natural vegetation is described as southern mixed pine forests and whose geography includes rolling hills and a warmer temperate climate [43]. The region can be further described at the level IV

ecoregion (65f) as having loamy acidic soils and dark tea-colored streams with mixed forests interspersed with pine plantations [43–45]. Southern pine species that persist and typically dominate in this region include Longleaf pine (*Pinus palustris*), Shortleaf pine (*Pinus echinate*), Slash pine (*Pinus elliottii*), Loblolly pine (*Pinus taeda*), and Virginia pine (*Pinus virginiana*). The mean elevation of the study site is 59.47 m, with a minimum and maximum elevation ranging from 25 to 99 m, respectively, and it receives approximately 135 cm of rainfall a year. Within the study area is the Solon Dixon Forestry Education Center (SDFEC), a multi-use facility (outreach, extension, teaching, and research) primarily focused on providing hands-on field training to university students. Forest inventory data for this study come from the SDFEC and are described in more detail in the following sub-sections.

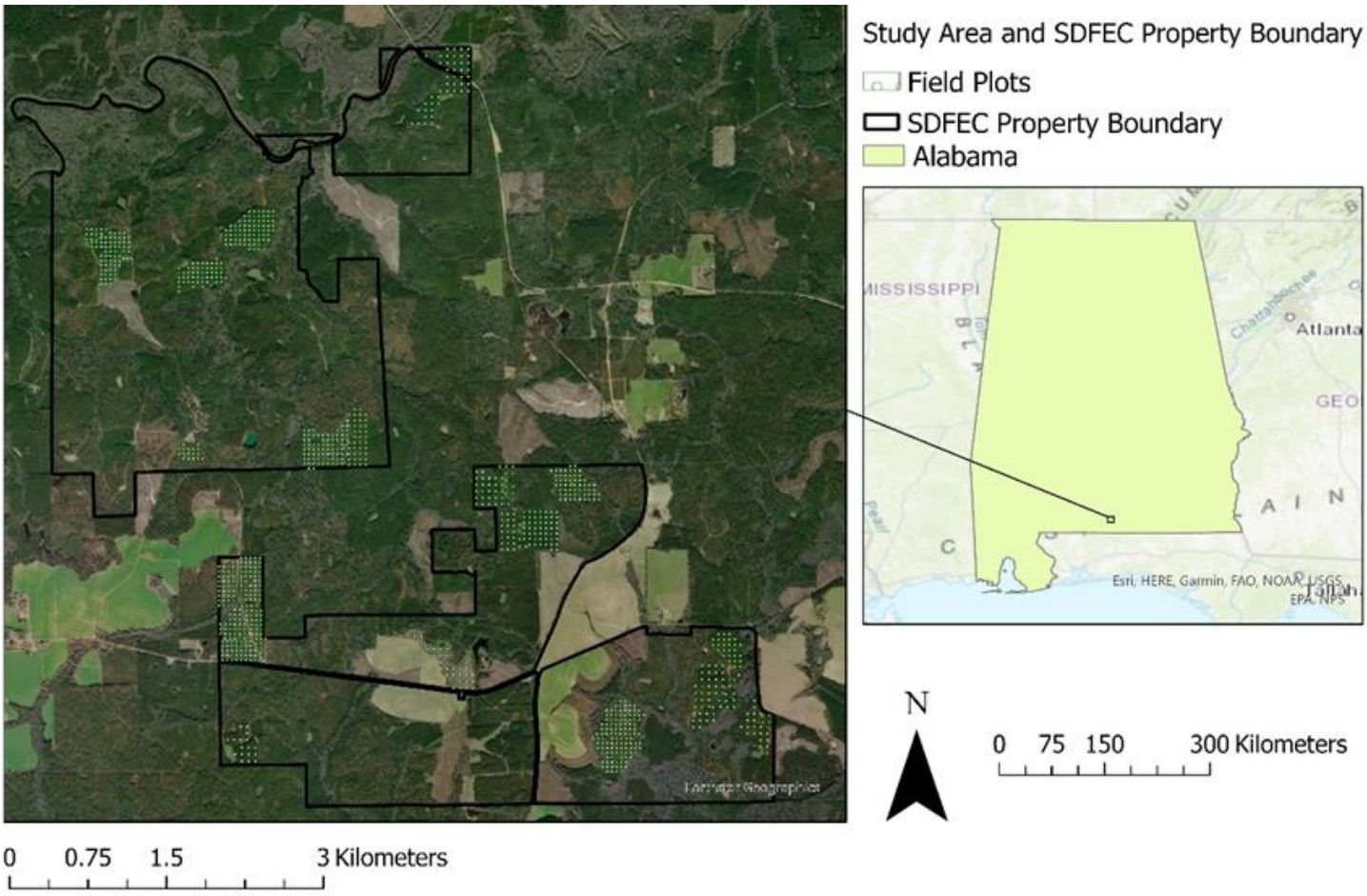

Figure 1 Study site with SDFEC property boundary and field plots represented by neon green circles, ESRI "World Imagery"

2.2INVENTORY DATA

The SDFEC consists of both naturally regenerated and planted pine, hardwood, and mixed pine-hardwood forest stands intended for timber production. Forest inventory data from the SDFEC were collected from 1048 circular plots of 0.04047 hectares (1/10th acre) size in 304 forest stands (see Figure 1). The number of plots used in this study was reduced from 1048 to 523 to include only plots that were inventoried within two years of the 2017 lidar and imagery data acquisition. Plots are generated using TCruise software [46] at approximately one plot per 0.4047 hectares (one plot per acre) with a distance between plot centers of approximately 44.20 to 51.82 m (145.0 to 170.0 ft). The data included latitude and longitude of plot centers, plot area (0.04047-hectare plots), merchantable timber heights of individual trees (meters), diameter at breast height for individual trees (centimeters), and plot BA ($m^2\,ha^{-1}$). The plot volume ($m^3\,ha^{-1}$) and plot sum of AGB (Mg ha^{-1}) were later calculated (see processing approach). The SDFEC uses lumber mill specifications when taking plot data on a forest inventory cruise. Trees surveyed in a plot must meet a minimum diameter at breast height (DBH) of 11.68 cm (4.598 inches) for a 12.70 cm (5.0 inches) class while being at least 4.600 m in height to a 7.620 cm (3.000 inches) top. Trees within plots that do not meet these specifications are not tallied. Furthermore, the SDFEC only tallies trees that local forest product mills are buying, so a variety of species, including Eastern Red Cedar (*Juniperus virginiana*) and Southern Black Cherry (*Prunus serotina*), are not counted.

2.3AIRBORNE LIDAR

Airborne lidar data used in this study come from the United States Geological Survey (USGS) 3D Elevation Program (3DEP) [47]. The 3DEP is an open-source program originally designed to collect three-dimensional geographic elevation data for a given landscape; however, the data have been used in many fields, including forestry and ecology. The data used in this study were collected by a Leica ALS70-HP RS device mounted in a fixed-wing position aboard Cessna and Partenavia aircrafts. Characteristics of the flight included a scan angle of 45 degrees, a flight height of 2000 AGL meters, and a speed of 130 kts. Lidar scanning characteristics included a 1 m footprint, a pulse density of 2.4 pls/m^2, and a scan frequency of 35.1 Hz. The scan pattern was triangular. Fifty-four lidar files containing point cloud data collected in 2017 were downloaded from 3DEP using UGet software [48]. Each file contained a las and/or laz tile with associated metadata.

2.4MULTISPECTRAL IMAGERY

Multispectral imagery data were acquired from Copernicus Sentinel-2. The Copernicus Sentinel-2 mission was developed by the European Space Agency (ESA) with the purpose of providing global and continuous land cover data in addition to map products such as change detection products and land cover maps. The mission aims to provide data for studies focused on climate change, land management, and security

[49]. Leaf-on and leaf-off Sentinel-2A Level-1C (L1C) top-of-atmosphere (TOA) reflectance orthoimages were downloaded from USGS EarthExplorer for the dates 19.VII.2016 and 24.II.2017. L1C products are derived from the L1B product and use radiometric and geometric corrections to produce the final product. The downloaded Sentinel-2 imagery consists of a ten-day, ten-meter spatial resolution for four spectral bands, including blue band 2 (490 nm), green band 3 (560 nm), red band 4 (665 nm), and NIR band 8 (842 nm). Imagery for this study consisted of less than twenty percent cloud cover.

2.5DATA PROCESSING APPROACH

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

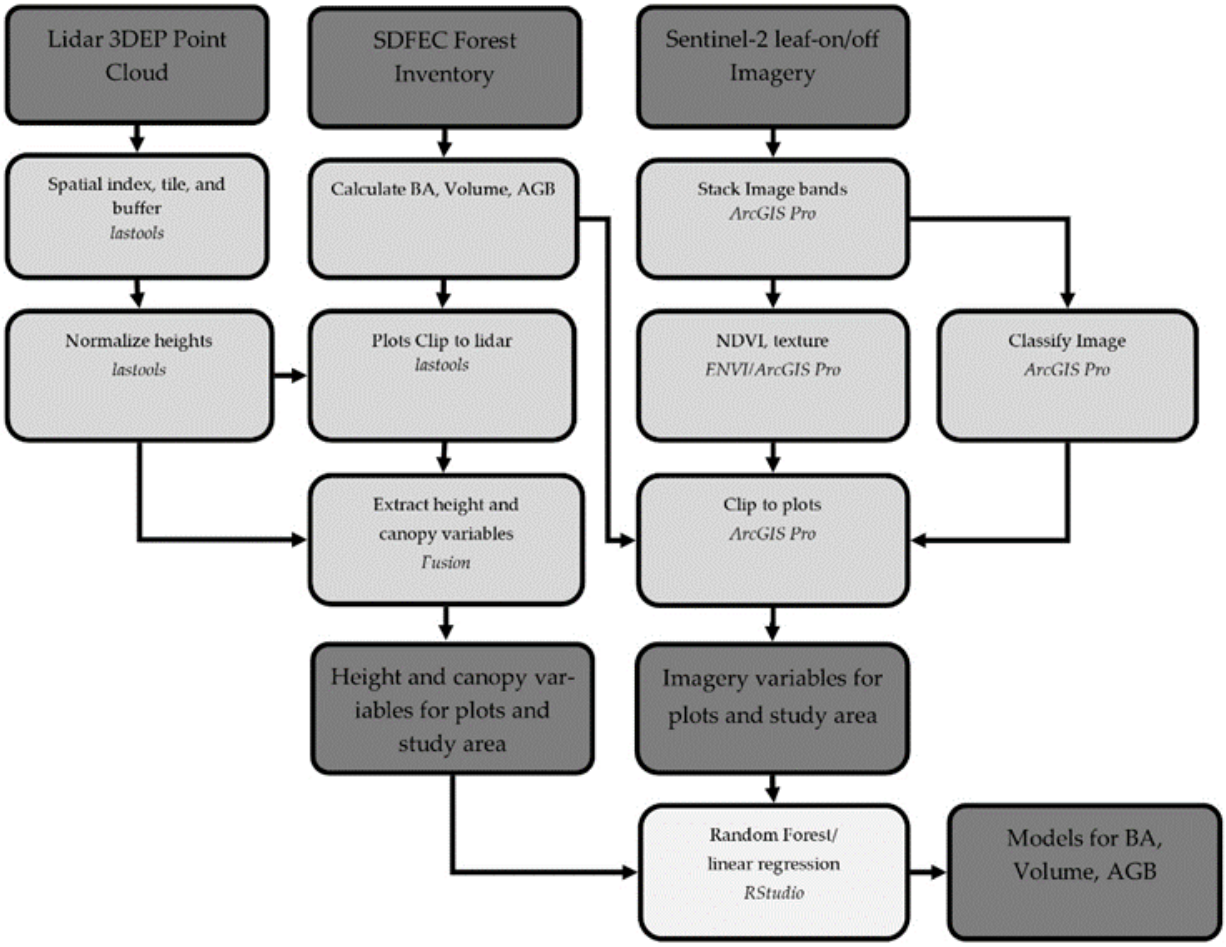

Figure 2 Data Processing and Workflow

2.6FIELD INVENTORY

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

Table 1 Summary statistics of inventory plots for BA, volume, and AGB. Calculated from allometric equations.

	BA ($m^2\ ha^{-1}$)	**Volume ($m^3\ ha^{-1}$)**	**AGB (Mg ha^{-1})**
Min	1.583	4.253	5.498
Max	45.92	412.8	226.4
Mean	18.33	193.7	80.84
Std. dev.	7.121	68.53	33.27

Basal area was calculated using the following equation

$$\text{Sum of } BA_n = \Sigma\left((0.00007854 \times DBH^2)/0.04\right)_i \quad (1)$$

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

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

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

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

Table 2 Species-specific parameters for calculating volume

	Pines	**Hardwoods**
α	−0.4432	0.8235
β	0.002165	0.001630

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

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

where $\beta 0$ and $\beta 1$ are species specific parameters (Table 3) [12].

Table 3 Species-specific parameters for calculating AGB

	Pines	**Hardwoods**
β_0	−2.536	−2.480
β_1	2.435	2.483

2.7LIDAR

Lidar data were processed using LASTools [50] and FUSION [51] software. In LASTools, aboveground tree heights were computed by normalizing points classified as ground returns and measuring the vegetation above ground points. Point cloud data were then clipped to the extent of field inventory plots in ArcGIS Pro. In FUSION, height and canopy metrics were calculated based on the normalized point cloud and the clipped plots. Plot field inventory data were combined with lidar metrics for modeling. The slope was calculated from a DEM derived from the point cloud and extracted per plot in ArcGIS Pro [52].

2.8MULTISPECTRAL IMAGERY PROCESSING

Sentinel-2 imagery data were used to derive multispectral variables of interest, including normalized difference vegetation index (NDVI) and second-order co-occurrence texture (Table 4). NDVI is known as a greenness vegetation index ratio with a range of −1 to 1 [54]. NDVI was calculated in ENVI [53] using the following equation

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

where NIR and red are the near-infrared and red band layers derived from the imagery. Second-order co-occurrence texture shows the relationship between one pixel value and its surrounding neighbors based on distance and angularity [54]. These metrics and NDVI were derived in ENVI as 20 m pixel resolution rasters. The means of these values

were further extracted from the zone of each field plot in ArcGIS Pro using the zonal statistics as a table tool in ArcGIS Pro.

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

Variable	Description
1 Max	Maximum height for a tree in a given plot
2 Mean	Mean height of trees in a given plot
3 Variance	Variance of tree heights in a plot
4 ElevCV	Co-variance of heights in a plot
5 p10, 25, 50, 75, 90, 95, 99	Height percentiles of trees in a plot
6 cancovA	% first returns above 4.6 m
7 cancovB	% all returns above 4.6 m
8 dens_0_10	Count of returns within 0–10 m
9 dens_10_15	Count of returns within 10–15 m
10 dens_15_20	Count of returns within 15–20 m
11 stddev	Standard deviation of tree heights
12 con_1,2,3,4	Texture contrast for bands r,b,g,NIR
13 cor _1,2,3,4	Texture correlation for bands r,b,g,NIR
14 dis _1,2,3,4	Texture dissimilarity for bands r,b,g,NIR
15 ent _1,2,3,4	Texture entropy for bands r,b,g,NIR
16 hom _1,2,3,4	Texture homogeneity for bands r,b,g,NIR
17 mean_1,2,3,4	Mean texture for bands r,b,g,NIR
18 secmom_1,2,3,4	Angular second moment for bands r,b,g,NIR
19 var_1,2,3,4	Texture variance for bands r,b,g,NIR
20 NDVI	NDVI within a plot
21 Land_cover	Landcover classes including Deciduous, Evergreen, and Barren
22 Slope	Slope calculated as a degree of inclination

With a temporal resolution of 10 days, Sentinel-2 data have the advantage of leaf-on and leaf-off acquisitions and made it possible to develop a landcover map distinguishing coniferous and deciduous vegetation in the SDFEC. This was achieved by stacking four bands (blue, green, red, and NIR) from each season (leaf-on and -off) and a canopy height model (CHM) derived from the lidar together, resulting in a nine-band image. Regions of interest were gathered from bands representing leaf-on and leaf-off years, and by using maximum likelihood supervised classification, a landcover map was produced with the following cover types: water, barren land, agriculture, deciduous vegetation, and evergreen vegetation. A confusion matrix was used to assess the accuracy of the supervised classification. The kappa coefficient, an indicator of how the classification of an image compares to ground truth data, was computed from the confusion matrix. This value ranges from 0 to 1, where 0 suggests that no none of the pixels were correctly classified, and 1 indicates that all pixels were correctly classified. The kappa for the classified image was 0.74, and the user accuracies were highest in the water and evergreen classes (user accuracy > 0.95), with the most confusion occurring in the agricultural and barren land cover classes. The highest producer accuracy occurred in evergreen and deciduous classes (0.73 and 1.00, respectively). Land cover types were extracted from each forest inventory field plot. A plot was then assigned a land cover type based on the dominant land cover type in that plot. Table 4 lists the 52 lidar, spectral, textural, and land cover metrics used in this study.

2.9DATA ANALYSIS

Two modeling approaches were applied for predicting BA, volume, and AGB: multiple linear regression and Random Forest (RF). Both modeling approaches were implemented using the R programming language in RStudio [55]. Before modeling via linear regression or RF, the categorical landcover variables needed to be dummy coded to be used in linear regression modeling. Therefore, two data sets were created for each dependent variable (BA, volume, and AGB), one with dummy coded land cover variables and one with categorical land cover variables.

2.10REGRESSION MODELING

For the multiple linear regression approach, the regsubsets function from the leaps package in R [56] was used to select variables and develop models for BA, volume, and AGB using forward and backward selection. In forward variable selection, a model is produced by starting with one variable and adding variables to the model until the best possible model is produced. Conversely, the backward variable selection includes every variable in the model and removes a variable until the best model is selected. The best five models were developed for up to the best five variable models based on the highest r^2, lowest Mallow's cp, and lowest Bayesian Information Criterion (BIC). A Breusch–Pagan test determined that the response variables were not normally distributed, so a log transformation was applied to each [57]. In order to avoid multicollinearity,

variance inflation factors (VIF) were calculated for each variable in each model, and models whose predictor variable's VIFs exceeded 10 were be removed [58]. A threshold of 0.05 was used to determine variable significance.

2.11RANDOM FOREST

RF is a non-parametric machine learning modeling technique that uses random bagging and a bootstrap sample of the data to create a number of user-specified trees that vote on the best model parameters at each node of each tree [35]. The major benefit of RF in this study is that it can handle a large number of variables and work with data that may otherwise push the boundaries of the assumptions made by multiple linear regression, such as the assumption that there is no multicollinearity. The "RF" function in the ModelMap, a package in R [59], was used for modeling BA, volume, and AGB. This package easily allows users to input data and tune hyperparameters to obtain a specified raster output. Important user set hyperparameters for RF include the mtry and ntree parameters. The mtry parameter is the number of randomly selected variables at each node and was automatically optimized. The ntree parameter is the number of trees grown in the model and was set to 500 trees. After model building, the model can be used to predict BA, volume, and AGB across the study extent. Raster layers for each lidar and imagery independent variable were created for the extent of the study site with the same number of rows and columns at a 20 m resolution. These rasters were

referenced using a look-up table (LUT), and the values for each cell are entered into the RF model to produce an estimation of the desired metric across the study site.

2.12MODEL EVALUATION

Selected linear regression models were evaluated using 10-fold cross-validation. This approach first randomly divides the dataset into 10 groups (folds) of equal sizes and then tests the data on one group using 1–10 folds to fit the data. The process repeats k number of times, computing the mean squared error (MSE) for each test group and averages the MSE of all tests to obtain the k-fold cross-validation estimate [60]. While leaving one out, cross-validation is a common model evaluation approach [27,61,62]; it was shown that k-fold cross-validation could reduce the test error rate [60]. After k-fold validation, the predicted values are back-transformed and compared to the observed values. The final model used for mapping was selected based on the r^2, RMSE, percent root mean squared error (%RMSE), mean absolute error (MAE), and bias. The resulting models were used to create wall-to-wall maps of the independent variables (BA, volume, AGB) using the raster package in R [63]. F-tests were used to determine if models with imagery-derived variables significantly improved the model, and r^2 and %RMSE were compared.

In evaluating the RF model, the out-of-bag error (OOB) was used to determine the prediction error. In the RF model building process, random bootstrap samples were

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

3.RESULTS

3.1LINEAR REGRESSION MODEL

From the stepwise variable selection, the forward selection approach produced the best models, followed by backward selection. A table of the five best models for each response variable is displayed in Appendix A (Table A1). According to the calculated vif's, no variable in a model contributed to multicollinearity (VIF < 5). p-values indicated that each variable was significant to each of the models ($p < 0.05$). Models produced for BA, volume, and AGB in this study generally included a lidar height, canopy, and density metric, though BA did not include any lidar-derived height metric. Beyond three-variable models, the var_3 variable was included in the four- and five-variable model for BA. In the four- and five-variable model for volume, the hom_1 variable was included in addition to another lidar-derived density metric. Four and five

variable models for AGB included var_2 and hom_1 variables, respectively, and the five-variable model included both.

The final model for volume outperformed those of BA and AGB. In model building, the five-variable model for volume had the lowest Mallow's cp and BIC and the highest r^2. After evaluating models from the 10-fold cross-validation approach, the five-variable model had the lowest error (RMSE = 49.23 $m^3 ha^{-1}$, %RMSE = 35.30) and the highest accuracy between the observed and predicted values (r^2= 0.45). This model included the following variables: p50, cancovA, dens_15_20, ElevCV, and hom_1. An F-test revealed that the hom_1 variable in the best model significantly improved the model ($p < 0.05$); however, the r^2 and %RMSE were not substantially different; the model without the 1_hom variable had the following accuracy: r^2 = 0.44; %RMSE = 35.96.

For BA, the model with the lowest BIC, Mallow's cp, and highest r^2 was the model with five variables. After validating BA using the 10-fold cross-validation approach, the model with five variables had the least amount of error (RMSE = 5.731 $m^2 ha^{-1}$, %RMSE = 31.26). Furthermore, the predictive accuracy between predicted and observed values using the five-variable model was highest (r^2 = 0.36). This model included the following variables: cancovA, dens_0_10, dens_10_15, dens_15_20, and var_3. An F-test revealed that the var_3 variable in the best model significantly improved the model ($p < 0.05$); however, the r^2 and %RMSE were not substantially different, and the model without the var_3 variable had the following accuracy: r^2 = 0.35; %RMSE = 31.61.

In modeling AGB, the model with the lowest BIC, Mallow's cp, and highest r^2 was the five-variable model. The five-variable model from the 10-fold cross-validation approach also had the lowest RMSE (25.20 Mg ha^{-1}, %RMSE = 31.26%) and the highest correlation between observed and predicted values ($r^2 = 0.41$). The following variables were included in this model: p50, cancovA, dens_15_20, var_2, and hom_1. An F-test was used to determine if the addition of one or both of the image-derived variables significantly improved the model. The addition of one or both of the image-derived variables was significant according to the F-test ($p < 0.05$); however, the r^2 and %RMSE for all three models were not substantially different, the model without hom_1 had an accuracy of $r^2 = 0.41$; %RMSE = 31.60, the model without both hom_1 and var_2 had $r^2 = 0.38$; %RMSE = 32.09. Selected models with coefficients are presented in Table 5, and scatterplots showing model prediction and observed values are displayed in Figure 3 with map outputs at a 20 m resolution (approximately the size of the field inventory plots).

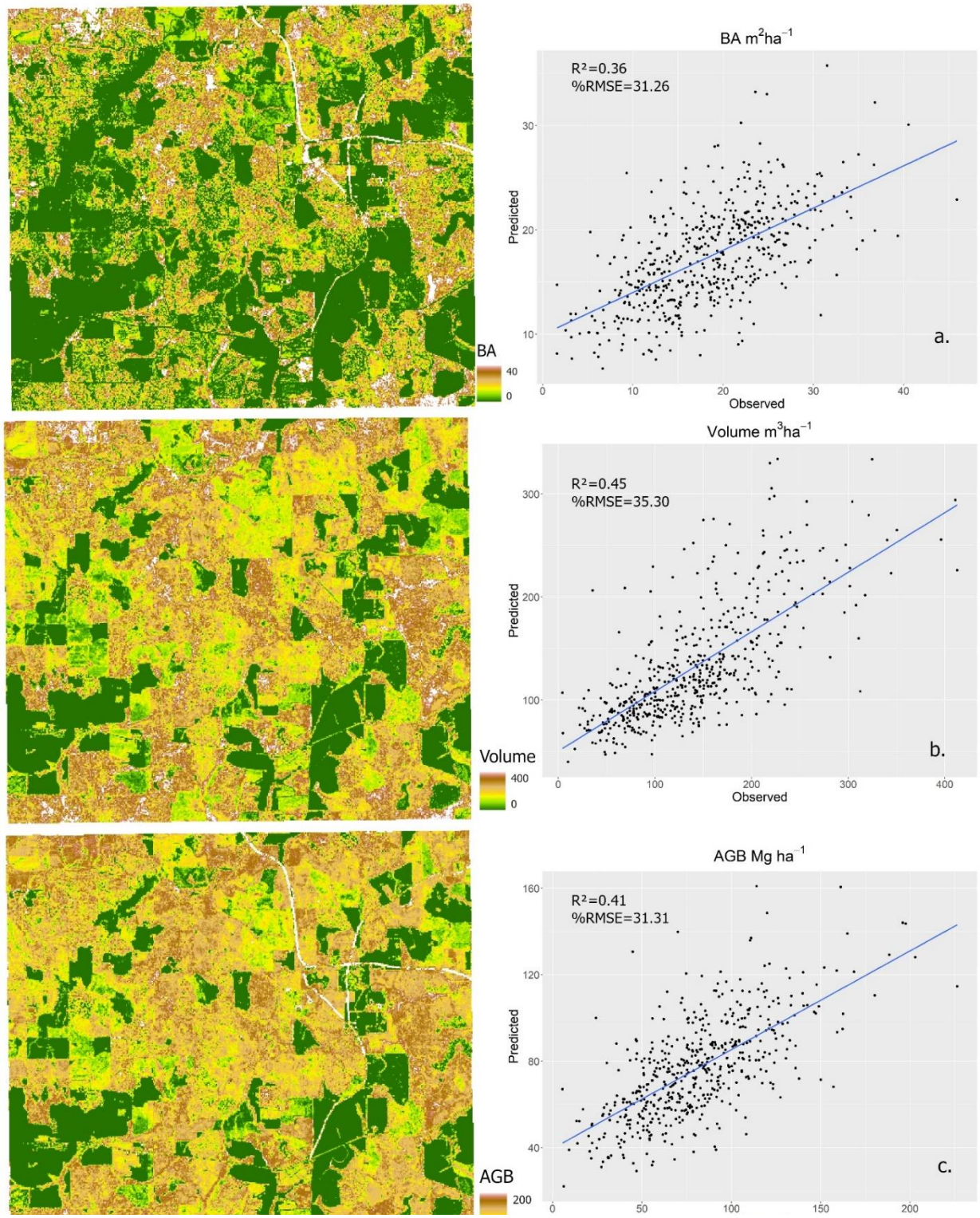

Figure 3 Predicted map outputs of BA (a), Volume (b), and AGB (c) with scatterplots displaying correlation between observed and predicted values for the selected linear regression models.

Table 5 Best models predicting BA, Volume, and AGB developed from linear regression modeling.

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

3.2RANDOM FOREST MODEL

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

Table 6 Model evaluation of RF predictions. Derived from observed and predicted values.

RF Model	%RMSE	RMSE	MAE	r^2	Bias
BA	30.17	5.662	4.449	0.39	0.01053
Volume	36.68	50.21	38.18	0.53	−6.438
AGB	34.26	26.29	20.77	0.37	−4.656

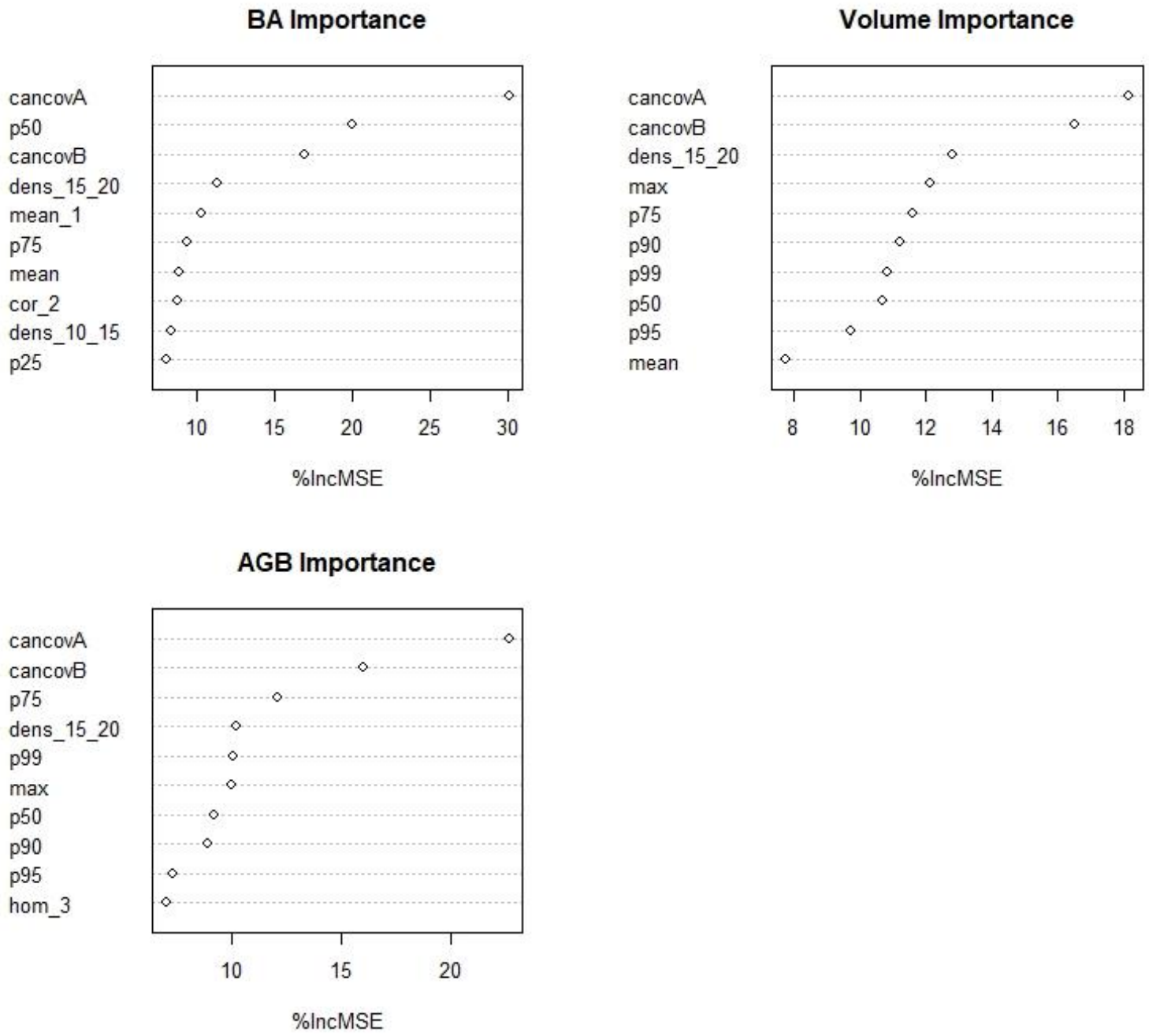

Figure 4 Importance plots for variables used in RF models. Most important variables are those at the top of the plots, and are based on % increase in MSE.

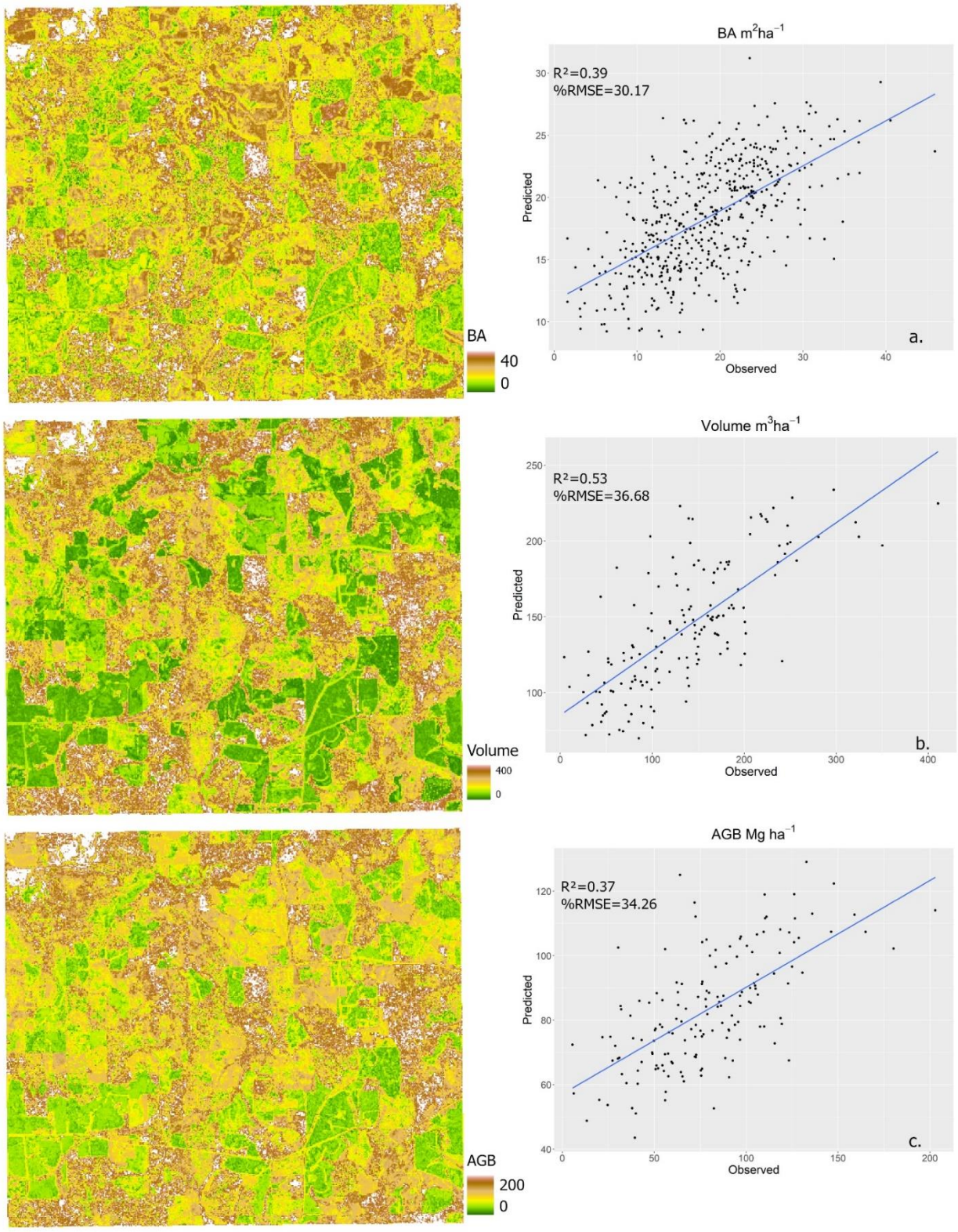

Figure 5 Predicted map outputs of BA (a), volume (b), and AGB (c) with scatterplots displaying correlation between observed and predicted values for the RF models.

4.DISCUSSION

Models predicting BA, volume, and AGB in this study ranged in accuracy from $r^2 = 0.36$ to 0.53. In the RF models, the r^2 values were higher than those produced by the linear regression modeling approach, except for when predicting BA. Both RMSE and %RMSE were lower in linear regression models, except when predicting BA. Other researchers modeling forest metrics reported findings that RF modeling tends to have the smallest RMSE and the least amount of bias [36,64]. Because the predictive accuracy between the two modeling types was similar, the findings of this study suggest that the results of a modeling approach could depend more heavily on the variables used in modeling, as well as the forest type being modeled (mixed forest stands, homogeneous plantations). The models produced for volume had the greatest predictive accuracy ($r^2 = 0.53$, RMSE = 50.21 Mg ha^{-1} %RMSE = 36.68), likely because the model used to calculate volume from field inventory included a height variable, while those for BA and AGB only included DBH and associated parameters. By calculating volume from field inventory using an equation that includes a height metric, lidar-derived height variables account for variation in heights when predictive modeling. Less accurate models developed for BA and AGB were also likely due to a number of other factors. For example, the difference in tree heights from when the field inventory was taken and when the lidar tree heights were measured could cause some variation; however, this is likely minimal as a subset of plots was used within two years of the lidar acquisition. The most likely explanation

for the unexplained variation in BA and AGB models is due to a combination of the merchantable timber cruising process used to collect this field data and to the high heterogeneity of southeastern forests [34]. In other words, a lot of the foliage measured by lidar and multispectral imagery is not accounted for in the field inventory gathered by inventory cruisers. This is especially seen in the model building process for BA, where variables selected for each model were those whose function either distinguished layers of the canopy (i.e., density metrics trying to explain shrubs and non-inventoried trees beneath the tops of pines) or those that distinguished foliage type (texture bands). Height metrics were excluded from these models because they could not explain BA for all of the foliage within a plot, especially when the data that the models were built from were based only on merchantable timber instead of the entirety of the foliage constituting the plot BA. Furthermore, while there is a relationship between a tree's height and its BA, in many species, this relationship is known to be somewhat weak [28,38], and another reason why there are no height metrics in the model for BA. Similar findings were reported by other researchers working in heavily mixed forest stands, such as those in Canada, where BA was modeled with an r^2 of only 0.093 [28]. In that particular study, the researchers noted that it was difficult to account for most of the vegetation not measured by timber cruisers and recommended using another data source in addition to lidar.

From the models developed in this study, spectral, textural, and landcover variables were considered in addition to lidar to delineate any attributes of the forest not accounted for in the field inventory. As mentioned in the results section, the model for BA included a texture variable (var_3). Furthermore, the four and five variable models for volume and biomass included texture variables (var_2, hom_1). This suggests that in heavily mixed forests, spectral, textural, and landcover class variables have the potential to explain variation missed by lidar variables alone; however, the difference in model accuracy after removing imagery variables was not substantial, and lidar variables alone predicted BA, volume and AGB almost as well. This is not to say that lidar variables are sufficient in estimating lidar in Alabama mixed-species forests, as the accuracy of lidar-derived models were still poor in BA, volume, and AGB. While many papers suggest that lidar alone is sufficient for modeling, most of these studies are developed for study sites whose forests are primarily homogeneous such as those in the western United States [58,65,66]. In estimating BA, for example, a study in western Oregon achieved very high predictive accuracy using a model with two height variables and a canopy cover variable ($r^2 = 0.96$), but the study site was homogeneous in nature, dominated by old-growth coniferous forests [30]. Because of the variation in allometric dimensions among tree species in mixed forest stands, alternative variables are necessary to accurately predict forest metrics.

5.CONCLUSIONS

RS data, including airborne and spaceborne lidar, as well as multispectral imagery from satellite-based platforms, are becoming a more available resource to foresters and ecologists. Programs such as the 3D Elevation program from the USGS, the National Agricultural Imagery Program from the USDA, and Sentinel-2 from the ESA provide free data that foresters can use to estimate the amount of BA, volume, and AGB on their stands and across large tracts of forested land. In this study, the potential of estimating BA, volume, and AGB from RS data was demonstrated by means of linear regression analysis and random forest modeling. Predictive accuracies of models were low (r^2 = 0.36–0.53) relative to those in some studies [27,58], and the results presented here suggest that the addition of imagery variables to lidar derived predictive models do not substantially help explain the variation in BA, volume, and AGB. While this study demonstrates the potential for a fast and efficient method of estimating BA, volume, and AGB using freely available data, further investigation of variables is needed to increase the variability of forest structure in mixed forest stands. Lastly, the main limitation of this study is that the field inventory used to build the models included merchantable timber and not the entire vegetation in one forest plot. Therefore, the potential to improve forest metric estimates may be improved if the entirety of the vegetation in a forest plot is measured.

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

6.LITERATURE CITED

1. Lister, A.J.; Andersen, H.; Frescino, T.; Gatziolis, D.; Healey, S.; Heath, L.S.; Liknes, G.C.; McRoberts, R.; Moisen, G.G.; Nelson, M.; et al. Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory. *Forests* **2020**, *11*, 1–41, doi:10.3390/f11121364.

2. Zhang, Y.; Liang, S.; Yang, L. A Review of Regional and Global Gridded Forest Biomass Datasets. *Remote Sens.* **2019**, *11*, doi:10.3390/rs11232744.

3. Cannell, M.G.R. Woody Mass of Forest Stands. *For. Ecol. Manage.* **1984**, *8*, 299–312.

4. Elledge, J.; Barlow, B. Basal Area: A Measure Made for Management. *Alabama Coop. Ext. Syst.* **2010**, 6.

5. Jr, H.V.W.; Baumgras, J.E.; Knight, R. Relation of Biomass to Basal Area and Site Index on an Appalachian Watershed. **1984**, 53–56.

6. Khan, M.N.I.; Islam, M.R.; Rahman, A.; Azad, M.S.; Mollick, A.S.; Kamruzzaman, M.; Sadath, M.N.; Feroz, S.M.; Rakkibu, M.G.; Knohl, A. Allometric Relationships of Stand Level Carbon Stocks to Basal Area, Tree Height and Wood Density of Nine Tree Species in Bangladesh. *Glob. Ecol. Conserv.* **2020**, *22*, doi:10.1016/j.gecco.2020.e01025.

7. U.S. Fish and Wildlife Service *Recovery Plan for the Red-Cockaded Woodpecker (Picoides Borealis): Second Revision*; 2003;

8. Vogt, J.; Koch, F.H. Forest Inventory and Analysis Data in Invasive Insect Research. *Am. Entomol.* **2016**, *62*, doi:10.1093/ae/tmv072.

9. Bond, B. PB1650 Understanding Log Scales and Log Rules. *UT Ext. Publ. For. Trees, Timber* **2011**, 8.

10. Oswalt, C.M.; Conner, R.C. Southern Forest Inventory and Analysis Volume Equation User ' s Guide. **2011**.

11. Tian, N.; Fan, Z.; Matney, T.G.; Schultz, E.B. Growth and Stem Profiles of Invasive. **2017**, *63*, 569–576.

12. Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsey, R.A. National-Scale Biomass Estimators for United States Tree Species. *For. Sci.* **2003**, *49*, 12–35, doi:10.1093/forestscience/49.1.12.

13. Burns, P.; Clark, M.; Salas, L.; Hancock, S.; Leland, D.; Jantz, P.; Dubayah, R.; Goetz, S.J. Incorporating Canopy Structure from Simulated GEDI Lidar into Bird Species Distribution Models. *Environ. Res. Lett.* **2020**, *15*, doi:10.1088/1748-9326/ab80ee.

14. UNFCCC Key Decisions Relevant for Reducing Emissions from Deforestation and

Forest Degradation in Developing Countries (REDD+). *Framew. Conv. Clim. Chang.* **2016**, 44.

15. Environmental Protection Agency (E.P.A) *Inventory of U.S. Greenhouse Gas Emissions and Sinks*;

16. Smith, J.E.; Heath, L.S.; Woodbury, P.B. How to Estimate Forest Carbon for Large Areas from Inventory Data. *J. For.* **2004**, *102*, 25–31, doi:10.1093/jof/102.5.25.

17. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T.; Jain, T.B.; Evans, J.S. Transferability of Lidar-Derived Basal Area and Stem Density Models within a Northern Idaho Ecoregion. *Can. J. Remote Sens.* **2018**, *44*, 131–143, doi:10.1080/07038992.2018.1461557.

18. Wilson, B.T.; Woodall, C.W.; Griffith, D.M. Imputing Forest Carbon Stock Estimates from Inventory Plots to a Nationally Continuous Coverage. *Carbon Balance Manag.* **2013**, *8*, 1–15, doi:10.1186/1750-0680-8-1.

19. Babst, F.; Bouriaud, O.; Alexander, R.; Trouet, V.; Frank, D. Toward Consistent Measurements of Carbon Accumulation: A Multi-Site Assessment of Biomass and Basal Area Increment across Europe. *Dendrochronologia* **2014**, *32*, 153–161, doi:10.1016/j.dendro.2014.01.002.

20. Elzinga, C.; Shearer, R.C.; Elzinga, G. Observer Variation in Tree Diameter Measurements. *West. J. Appl. For.* **2005**, *20*, 134–137, doi:10.1093/wjaf/20.2.134.

21. Mcroberts, R.E.; Tomppo, E.O. Remote Sensing Support for National Forest Inventories. **2007**, *110*, 412–419, doi:10.1016/j.rse.2006.09.034.

22. Westfall, J.A.; Woodall, C.W.; Hatfield, M.A. A Statistical Power Analysis of Woody Carbon Flux from Forest Inventory Data. **2013**, doi:10.1007/s10584-012-0686-z.

23. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273.

24. White, J.C.; Wulder, M.A.; Buckmaster, G. Validating Estimates of Merchantable Volume from Airborne Laser Scanning (ALS) Data Using Weight Scale Data. *For. Chron.* **2014**, *90*, 378–385, doi:10.5558/tfc2014072.

25. Chuvieco, C. *Fundamentals of Satellite Remote Sensing An Environmnetal Approach*; Third.; CRC Press, 2020; ISBN 978-1138583832.

26. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Jaafar, W.S.W.M.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting Stem Total and Assortment Volumes in an Industrial Pinus Taeda L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* **2017**, *8*, 1–17, doi:10.3390/f8070254.

27. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Fennema, S.J.; Corte, A.P.D. Modeling and Mapping Basal Area of Pinus Taeda L. Plantation Using Airborne LiDAR Data. *An. Acad. Bras. Cienc.* **2017**, *89*, 1895–1905, doi:10.1590/0001-3765201720160324.

28. Spriggs, R.A.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A Critique of General Allometry-Inspired Models for Estimating Forest Carbon Density from Airborne LiDAR. *PLoS One* **2019**, *14*, 1–19, doi:10.1371/journal.pone.0215238.

29. Holmgren, J. Prediction of Tree Height, Basal Area and Stem Volume in Forest Stands Using Airborne Laser Scanning. *Scand. J. For. Res.* **2004**, *19*, 543–553, doi:10.1080/02827580410019472.

30. Means, J.E.; Acker, S.A.; Harding, D.J.; Blair, J.B.; Lefsky, M.A.; Cohen, W.B.; Harmon, M.E.; McKee, W.A. Use of Large-Footprint Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sens. Environ.* **1999**, *67*, 298–308, doi:10.1016/S0034-4257(98)00091-1.

31. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. *A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach*; 2013; Vol. 89; ISBN 9781100223858.

32. Laes, D.; Reutebuch, S.E.; Mcgaughey, R.J.; Mitchell, B. *Guidelines to Estimate Forest Inventory Parameters from Lidar and Field Plot Data*; 2011;

33. Narine, L.L.; Popescu, S.; Neuenschwander, A.; Zhou, T.; Srinivasan, S.; Harbeck, K. Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sens. Environ.* **2019**, *224*, 1–11, doi:10.1016/j.rse.2019.01.037.

34. Hogland, J.; Anderson, N.; St. Peter, J.; Drake, J.; Medley, P. Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *ISPRS Int. J. Geo-Information* **2018**, *7*, doi:10.3390/ijgi7040140.

35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:doi.org/10.1023/A:1010933404324.

36. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Hall, D.E.; Falkowski, M.J. Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245, doi:10.1016/j.rse.2007.10.009.

37. Kim, J.; Popescu, S.C.; Lopez, R.R.; Wu, X. Ben; Silvy, N.J. Vegetation Mapping of No Name Key, Florida Using Lidar and Multispectral Remote Sensing. *Int. J. Remote Sens.* **2020**, *41*, 9469–9506, doi:10.1080/01431161.2020.1800125.

38. Silva, C.A.; Hudak, A.T.; Vierling, L.A.; Loudermilk, E.L.; O'Brien, J.J.; Hiers, J.K.;

Jack, S.B.; Gonzalez-Benecke, C.; Lee, H.; Falkowski, M.J.; et al. Imputation of Individual Longleaf Pine (Pinus Palustris Mill.) Tree Attributes from Field and LiDAR Data. *Can. J. Remote Sens.* **2016**, *42*, 554–573, doi:10.1080/07038992.2016.1196582.

39. Hudak, A.T.; Evans, J.S.; Crookston, Nicholas L. Falkowski, M.J.; Steigers, B.K.; Taylor, R.; Hemingway, H. Aggregating Pixel-Level Basal Area Predictions Derived from LiDAR Data to Industrial Forest Stands in North-Central Idaho. In: Havis, Robert N.; Crookston, Nicholas L., Comps. 2008. Third Forest Vegetation Simulator Conference; 2007 February 13–15; Fort C. *Third For. Veg. Simulator Conf.* **2008**, 133–145.

40. Hudak, A.T.; Evans, J.S.; Falkowski, M.J.; Crookston, N.L.; Gessler, P.E.; Morgan, P.; Smith, A.M.S. Predicting Plot Basal Area and Tree Density in Mixed-Conifer Forest from Lidar and Advanced Land Imager (ALI) Data. *Proc. 26th Can. Symp. Remote Sens.* **2005**, 201–208.

41. Popescu, S.C.; Wynne, R.H.; Scrivani, J.A. Fusion of Small-Footprint Lidar and Multispectral Data to Estimate Plot-Level Volume and Biomass in Deciduous and Pine Forests in Virginia, USA. *For. Sci.* **2004**, *50*, 551–565, doi:10.1093/forestscience/50.4.551.

42. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton,

M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002, doi:10.1016/j.srs.2020.100002.

43. Smith, L.; Sidney, L. Ecoregions of Alabama and Georgia. *Environ. Prot.* **1995**, 75–75.

44. Omernik, J. Map Supplement Ecoregions of the Conterminous United. *Ann. Assoc. Am. Geogr.* **2014**, *77*, 118–125.

45. Range, B.; Lawrence, S. Level III Ecoregions of Alaska. **2001**, 175.

46. Matney, T. TCruise 2022.

47. United States Geolotical Survey (USGS); US Department of Interior 3D Elevation Program Available online: https://www.usgs.gov/core-science-systems/ngp/3dep.

48. UGET Available online: https://ugetdm.com/.

49. Drusch, M.; Gascon, F.; Berger, M. *Sentinel-2 Mission Requirements Document*; 2010;

50. Isenburg, M. LAStools: Award-Winning Software for Rapid LiDAR Processing Available online: http://lastools.org/.

51. Mcgaughey, R.J. Agriculture FUSION / LDV : Software for LIDAR Data Analysis and Visualization. **2017**.

52. ESRI ArcGIS Pro.

53. L3Harris Exelis Visual Information Solutions 2021.

54. Dinstein, I.; Shanmugam, K.; Haralick, R.M. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *SMC-3*, 610–621.

55. Team, R.C. R: A Language and Environment for Statistical Computing 2021.

56. Lumley, T. Package ' Leaps .' **2020**.

57. Breusch, T.S.; Pagan, A.R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* **1979**, *47*, 1287–1294, doi:https://doi.org/10.2307/1911963.

58. Sheridan, R.D.; Popescu, S.C.; Gatziolis, D.; Morgan, C.L.S.; Ku, N.W. Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest. *Remote Sens.* **2015**, *7*, 229–255, doi:10.3390/rs70100229.

59. Freeman, A.E.; Frescino, T.; Freeman, M.E. Package ' ModelMap .' **2018**.

60. Gareth, J.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R.*; Springer Publishing Company, Incorporated, 2014; ISBN 978-1071614174.

61. Dorado-Roda, I.; Pascual, A.; Godinho, S.; Silva, C.A.; Botequim, B.; Rodríguez-Gonzálvez, P.; González-Ferreiro, E.; Guerra-Hernández, J. Assessing the

Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. **2021**, doi:10.3390/rs13122279.

62. 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, doi:10.1016/j.rse.2009.03.006.

63. Etten, J. Van; Sumner, M.; Cheng, J.; Baston, D.; Bevan, A.; Bivand, R.; Busetto, L.; Canty, M.; Fasoli, B.; Forrest, D.; et al. Package ' Raster ' R Topics Documented : **2022**.

64. Gagliasso, D.; Hummel, S.; Temesgen, H. A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area. *Open J. For.* **2014**, *04*, 42–48, doi:10.4236/ojf.2014.41008.

65. Bright, B.C.; Hudak, A.T.; McGaughey, R.; Andersen, H.E.; Negrón, J. Predicting Live and Dead Tree Basal Area of Bark Beetle Affected Forests from Discrete-Return Lidar. *Can. J. Remote Sens.* **2013**, *39*, doi:10.5589/m13-027.

66. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T. Temporal Transferability of LiDAR-Based Imputation of Forest Inventory Attributes. *Can. J. For. Res.* **2015**, *45*, 422–435, doi:10.1139/cjfr-2014-0405.

CHAPTER 3

VOLUME AND ABOVEGROUND BIOMASS MAPPING USING GEDI AND SENTINEL-2 DATA IN SOUTHERN ALABAMA

1.INTRODUCTION

Human-induced climate change is a growing concern for both the persistence of ecological systems and human societies [1]. Furthermore, deforestation due to development and increased agriculture is a well-known driver of climate change and the loss of biodiversity [2–4]. It is therefore imperative that governments around the world have accurate and high-resolution estimates of forest structure to support management of forests and their ability to act as carbon sinks rather than anthropogenic sources of climate change. Southern Alabama is a region of critical importance for the export of timber, and therefore plays a major role in both the economy of Alabama and the absorption of atmospheric carbon. However, climate change, deforestation, and the need for forest maps will inevitably change the way foresters in the Southern Alabama do forestry and manage timber stands [5]. Key forest inventory metrics such as timber volume and aboveground biomass (AGB) are commonly used for forest management and therefore useful for understanding the role of forests in the carbon cycle. AGB is a measure of the dry aboveground carbon of a tree, and is directly related to the amount of carbon stored by forests, typically found by multiplying the dry mass by a

conversion factor of 0.5 [6,7]. It is estimated that up to 300 peta-grams of carbon (Pg C) is stored in the world's forests, which is equivalent to approximately one-third of the carbon in the atmosphere [8]. In the southeastern U.S., the forest industry can act as a source of carbon with the removal of pine forests for the production of paper products, however with proper management the change in carbon stock can remain constant [9]. Linked to biomass, volume is the total amount of wood in a tree or forest, and is also a useful measure of the monetary value of a forest stand and habitat suitability [10–12]. Estimating volume and AGB across large extents based only on traditional forest inventory measurements can be problematic due to the costs of labor, inconsistencies in mensuration methodologies, and variation in geographic factors influencing forest growth, such as soil quality and water availability [13–15]. Technologies such as light detection and ranging (lidar) have been used to produce maps of key forest inventory metrics across somewhat large spatial extents, and studies using airborne laser scanning (ALS) (lidar from aircraft) have shown major success in estimating forest metrics across large areas [16–18], however few have used space-based lidar for merchantable forestry operations. Estimation of AGB and volume from ALS have typically been accomplished through modelling approaches in which known AGB and volume from field plots are used to find highly correlated lidar variables and their coefficients. The lidar variables are known to vary based on the type of forest being measured [19–21], and the type of modelling used [21–23], but generally a height, canopy, and variance variable are

sufficient for estimating AGB and volume [24,25]. One major advantage of ALS over field inventory is that it offers full coverage across a study stie due to its ability to directly and continuously measure forest structure. This has also been demonstrated to be more accurate and cost-effective than the method of estimating forest structure from interpolation of forest plots spread across an area [26]. While the extent of forest structure measured by lidar is great relative to field inventory plots, it is nevertheless limited by fuel costs and are therefore used more frequently in local projects. One study demonstrated the ability of ALS to accurately estimate AGB (R^2 = 0.81; RMSE = 37.69 Mg ha^{-1}) but noted that the extent of a lidar acquisition is limited by its high costs, suggesting that regional or global estimation using ALS is difficult without expensive repeat flights [27]. Another study achieved high accuracy estimates of volume from airborne lidar (R^2 = 0.94; RMSE = 12.02 m^3ha^{-1}), but noted that while the cost of lidar acquisition was comparable to field inventory estimates, the cost is still very high resulting in limited coverage [17].

Spaceborne lidar (lidar system mounted on a satellite) may provide opportunities to overcome limited coverage of ALS, as they can collect data almost continuously while in orbit. Missions such as Ice, Cloud, and land Elevation Satellite (ICESat) and ICESat-2 have been useful in measuring canopy cover, AGB, and forest volume across large extents [28–32]. ICESat was launched in 2003 with the objective of measuring changes in polar ice over time, and used it's Geoscience Laser Altimeter System (GLAS), a single-

beam full- waveform lidar to measure properties of ice [32]. A 2009 study used height and canopy variables from GLAS to model and estimate timber volume in Siberian forests [33]. This study demonstrated early on the usefulness of spaceborne lidar for estimating forest metrics, but the authors noted the overestimation of forest metrics due to the instrument's large footprint and single laser system. ICESat-2 was designed to carry on the mission of ICESat, but also includes mission objectives for estimating forest characteristics. Aboard ICESat-2 is the Advanced Topographic Laser Altimeter System (ATLAS), a six-beam photon-counting laser system with a footprint < 17m (an upgrade from GLAS single beam large footprint lidar) that could take more accurate measurements of ice and forest characteristics [34]. Recent studies have used ICESat-2 to estimate AGB [28,35], even though ATLAS is optimized for quantifying properties of the cryosphere including ice thickness and density. However, another space-based lidar mission that was launched the same year as ICESat-2 is the Global Ecosystem Dynamics Investigation (GEDI), which was designed with the intent of measuring ecosystem structure.

GEDI is another spaceborne lidar aboard the International Space Station, launched in December 2018 that was specifically designed for vegetation mapping [36]. The main objective of the mission is to understand how land-use and climate change are impacting ecosystem structure. Embedded in this main objective are three primary goals: (1) get a better understanding of the Earth's carbon balance with tropical and

temperate forests, (2) understand how land surface impacts carbon dioxide in the atmosphere, and (3) understand how ecosystem structure affects biodiversity [36]. These goals are possible due in part to the components of GEDI itself: GEDI is composed of three lasers (two strong power beams and one coverage beam), one of which is split in two (coverage), and the lasers are dithered across a given track to produce 8 total lidar tracks with 25m diameter footprints. The platform was scheduled for a two-year mission and extended through 2023 aimed at estimating land-based vegetation structure in order to better understand climate cycling and ecosystem dynamics [36]. GEDI includes several publicly available products including height and canopy data (L2; 25m footprints), and Level 4 (L4) aboveground biomass density (AGBD) products available in both 25-meter footprints (L4A) and one-kilometer grids (L4B). The L4A product was produced by modeling a relationship between field calculated AGBD and simulated GEDI waveforms from airborne lidar data, and then using the model to predict footprint-level AGBD [37]. The idea behind simulating GEDI data was to develop accurate models for AGBD across a number of different regions, instead of estimating AGBD using GEDI height data and a general parametric model, which would cause inaccuracies throughout different forest types. L4B gridded AGBD products are produced from the L4A product at a 1-kilometer spatial resolution.

Already, several studies have used GEDI data in ways that align with the aforementioned mission objective and goals. In estimating vegetation structure, Lang et

al. [38] developed a Bayesian convolutional neural network model using GEDI and ALS data that could estimate global canopy heights accurately (RMSE 2.7m) with low bias. These canopy height estimates are crucial for estimating AGB, as canopy height is typically the most important variable for modelling and estimating AGB [39,40]. Furthermore, the same study demonstrated that both weak and strong beams are useful for estimating canopy height, which has been limited with previous sensors [41]. Another study analyzed the accuracy of GEDI canopy heights by comparing them with ALS canopy height measurements [42]. The researchers reported good accuracy in GEDI height estimations in all forest types in the United States (RMSE = 2.62 m; %RMSE less than 25%), but noted that geolocation errors and the difference in data acquisition between GEDI and the ALS were likely responsible for any unexplained variability.

In terms of volume, one study demonstrated the use of GEDI to improve estimates across Changbai mountains in northeast China with higher accuracy than previous sensors (R^2=0.88; %RMSE = 15.21%) [43]. Another study used GEDI data to estimate volume in Brazilian Eucalyptus plantations with very high accuracy (R^2 = 0.90; 24.39 m^3 ha^{-1}, noting that the best explanatory variable for predicting volume was the 90^{th} height percentile [44]. Beyond the two previously mentioned studies, little work has been done to estimate volume from GEDI, and to the knowledge of the authors, no work has been done to estimate merchantable volume for forests in the Southeastern U.S. by taking advantage of the high density of GEDI data and the l4A AGBD as a predictor variable to

help calibrate forest metrics models. Understanding the ecological impacts of forest structure is another of GEDI's primary mission goals that researchers have been exploring. For example, researchers have used both simulated GEDI data from an unmanned aerial vehicle (UAV) and available GEDI waveform data to estimate fuel load in tropical forests with good accuracy (R^2=0.88) [45]. Determining fuel load helps researchers understand how disturbance dynamics affect ecosystems as the climate changes [46]. Forest structure also impacts the presence of animals as well [47], and GEDI offers the potential to estimate habitat suitability and species distribution across large extents. For example, when developing species distribution models (SDM) for bird species in California, researchers found that simulated GEDI-derived canopy structure variables were among the most important in predicting avian occurrence [48]. Finally, a major goal of the GEDI mission is to understand the balance of carbon between the atmosphere and forests. Estimating AGB is a common approach to understanding the changes in carbon concentration over time. Though GEDI provides estimated footprint-level AGBD, other researchers may produce their own estimates of AGB for comparison with those of GEDI's, or optimize multiple sensors for improved estimation of AGB. Dorado-Roda et al. [49] used GEDI height variables to estimate footprint level AGB in Mediterranean forests, and found that AGB could be predicted more accurately in homogeneous coniferous forests than in mixed deciduous forests. Authors suggested that, as shown in previous studies, forest estimates of AGB using lidar data is less

accurate in dense heterogenous forests [22,50]. In another study, researchers combined simulated ICESat-2 and simulated GEDI data to estimate AGB and found improved results when estimates were compared with those using only one or the other of the two sensors [51]. These latter two studies are significant because they present the usefulness of GEDI products for optimizing estimates of AGB in specific study areas and the synergistic use of data sources for accurate estimates [52]. As such, it is necessary to further explore methods for estimating AGB by combining data sources to potentially derive more accurate estimates, as well as to develop maps with complete coverage of AGB. Furthermore, current GEDI estimates of AGB are available as 25m footprints or 1-kilometer grids, and its use for estimating volume is not well-understood for forests within the Southeastern U.S., so work needs to be done to achieve full coverage at finer scales.

Sentinel-2 data, in conjunction with GEDI footprints, offer opportunities to achieve full coverage estimates of AGB and volume at potentially finer spatial scales [53]. Sentinel-2 consists of two polar orbiting satellites, each composed of a 13-band multispectral imagery sensor. It's mission is designated for observing global and continuous land cover data in addition to providing map products such as change detection products and land cover maps. Further mission goals include providing data for studies focused on climate change, land management, and security. By estimating volume and AGB with GEDI footprints, Sentinel-2 imagery can be used to upscale these parameters to

produce wall-to-wall maps. While this approach is not necessarily unique [28,54,55], no study has used it to map merchantable AGB and volume in the southeastern U.S. In this study, ground data is included as part of a three-step approach to mapping merchantable volume and AGB; first developing wall-to-wall reference maps of merchantable volume and AGB from field data and airborne lidar [50], second, predicting merchantable volume and AGB in GEDI footprints by developing models from reference maps and GEDI L2 and L4 metrics, and finally, leveraging a Sentinel-2 scene for wall-to-wall mapping. Ultimately, the aim of this study is to evaluate GEDI L2 and L4 data for determining key forest inventory metrics in the southeastern U.S. The specific objectives were to: (1) estimate forest volume and AGB using footprint-level GEDI data, and (2) develop maps of forest volume and AGB using GEDI and Sentinel-2 data.

2.METHODS

2.1STUDY SITE

A 35.35 km^2 extent crossing Covington and Escambia counties in Southern Alabama was used as the primary area of interest in this study (Chapter 2, figure 1), though volume and AGB will be mapped across the extent of a Sentinel-2 scene. The area is located in the southeastern plains type III ecoregion (ecoregion 65) composed of southern mixed pine forests on rolling hills with a warm climate [56]. The area is further

described at the level IV ecoregion (65F) as having loamy acidic soils with dark tea-colored streams and mixed forests interspersed with pine plantations [57,58]. The mean elevation of the study site is 59.47 meters, with a minimum and maximum elevation ranging from 25 to 99 meters respectively. The study area also encompasses the Solon Dixon Forestry Education Center (SDFEC), a multi-use facility (outreach, extension, teaching, and research) primarily focused on providing hands-on field training to university students. Forest inventory data for this study comes from the SDFEC and is described in more detail in the following sub-sections.

2.2AIRBORNE LIDAR-DERIVED REFERENCE MAPS

Airborne lidar reference maps can be highly valuable for comparing space-based estimates of metrics like AGB, volume, canopy cover, and canopy heights. Furthermore, airborne reference maps can be used for calibrating models of AGB and volume with spaceborne lidar predictor variables. In this study, independent airborne lidar derived reference maps were used for developing models that predict merchantable AGB and volume. Specific details on the development of airborne lidar-derived reference maps are described in a previous study [50], but are summarized here.

Airborne lidar data come from the United States Geological Survey (USGS) 3D Elevation Program (3DEP) [59]. 3DEP is an open-source program originally designed to collect three-dimensional geographic elevation data for a given landscape, however the

data have been used in many fields including forestry and ecology. The data used for the production of the reference maps were collected by a Leica ALS70-HP RS device mounted in a fixed wing position aboard Cessna and Partenavia aircrafts. Characteristics of the flight included a scan angle of 45 degrees, a flight height of 2000 AGL meters, and a speed of 130 kts. Lidar scanning characteristics included a 1-meter footprint, a pulse density of 2.4 pls/meter^2, and a scan frequency of 35.1 Hz. The scan pattern was triangular. 54 lidar files containing point cloud data collected in 2017 were downloaded from 3DEP using UGet software [60]. Each file contained a las and/or laz tile with associated metadata.

Forest inventory data for the SDFEC consists of both naturally regenerated and planted pine, hardwood, and mixed pine-hardwood forest stands intended for timber production. Forest inventory data from the SDFEC was collected from 1,048 circular plots of .04 hectare (1/10th acre) size, in 304 forest stands (see Figure 1). The number of plots used to produce the airborne reference maps from a previous study [50] was reduced from 1048 to 523 to include only plots that were inventoried within two years of the 2017 airborne lidar. The data included latitude and longitude of plot centers, plot area (0.04 hectare plots) merchantable timber heights of individual trees (meters), and diameter at breast height for individual trees (centimeters). The plot volume (m^3ha^{-1}) and plot sum of AGB ($Mgha^{-1}$) were later calculated using methods described in a previous study [50]. The SDFEC uses lumber mill specifications when taking plot data

on a forest inventory cruise. Trees surveyed in a plot must meet a minimum diameter at breast height (dbh) of 11.68 cm for a 12.7 cm class, while being at least 4.6 meters in height to a 7.62 cm top. Trees within plots that do not meet these specifications are not tallied. Furthermore, the SDFEC only tallies trees that local forest product mills are buying, so a variety of species including Eastern Red Cedar (*Juniperus virginiana*) and Southern Black Cherry (*Prunus serotina*) are not counted.

To develop wall-to-wall reference maps of AGB and volume, airborne lidar-derived variables from processed lidar point clouds were used as predictors in a linear regression, where AGB and volume were dependent variables in model building. The subsequent models were fit and validated using a 10-fold cross validation approach and used to map AGB and volume across the study site, an approach well documented in other studies [59,61]. The maps have a resolution of 20m and were resampled in ArcGIS Pro to match the approximate spatial resolution of GEDI footprints (25m). The map produced for volume has an accuracy of R^2=0.45, %RMSE=35.30. The map for AGB have an accuracy of R^2=0.41 %RMSE=31.31.

2.3GEDI DATA

GEDI HDF5 files were downloaded from NASA EarthData Search, a data pool that allows a user to download earth observation (EO) data from an area of interest without visiting each DAAC website for a product of interest. Data used in this study include 38

HDF5 files of GEDI L2A relative height metrics (rh), 38 HDF5 files of L2B elevation and canopy cover metrics, and 38 HDF5 files of GEDI L4 data containing footprint level aboveground biomass density (AGBD) from May, 2019 to January, 2022. L2 products are processed and stored at the Land Processing Distributed Active Archive Center (LP DAAC). L4A data are processed and stored at the Oak Ridge National Laboratory (ORNL) DAAC. The download included all available data collected to the date of download containing the area of interest, totaling 463,233 25-meter footprints. L2 rh metrics, elevation and canopy cover metrics were downloaded for modeling both volume and AGB, as height and canopy metrics have been demonstrated to correlate strongly with both volume and AGB [29,62]. As mentioned above, GEDI data are collected aboard the ISS using three lasers, one of which is split to produce four beams dithered across eight tracks. Two of the lasers are full power beams and produce 15 m/J beams. The third laser is considered a coverage beam and produces two 4.5 m/J beams. Each beam footprint has a diameter of 25m, optimized for capturing canopies of large trees and being able to reach the ground beneath dense canopies [36]. Lidar full waveform collected in each footprint is further processed into the previously mentioned L2 products. L2 rh metrics are quantiles of heights of waveform energy relative to the ground elevation (ground elevation is computed as the mode of lowest returns). Similar to canopy height percentiles used in earlier airborne lidar studies [17,27,63], rh metrics can give researchers an idea of the vegetative heights in forests. Canopy cover related

metrics include canopy cover calculated by determining the ratio between the vertical canopy energy distribution (subtracting the ground estimation from the energy waveform) and ground energy for a given footprint [36]. Also included in L2B canopy metrics are plant area volume density (pavd) percentiles, leaf area index (LAI), essentially the amount of foliage in the canopy is calculated from the vertical canopy energy, and foliage height diversity (FHD), the total leaf area in a given space, which is calculated from LAI. Finally, the L4A footprint AGBD are calculated from regional allometric equations derived from the relationship between airborne lidar canopy and height variables, and field inventory AGB [37].

2.4GEDI DATA PROCESSING

From the downloaded HDF5 files for L2A, L2B, and L4A products, the following items were extracted and formatted as a csv file using the package *rhdf5* [64] in RStudio [65]: latitude, longitude, beam number, shot number, quality flag, degrade flag, rh0-100, canopy cover, LAI, FHD, and AGBD. The data were then filtered in ArcGIS Pro [66] using a mask produced from classified Sentinel-2 imagery (See Sentinel-2 processing section), such that GEDI shots only in forested areas remained in the dataset. After filtering shots, the final number of footprints was 22,160 footprints. The *summarize raster within* tool in ArcGIS Pro was used to determine the volume and AGB of the reference maps within the extent of the GEDI shots. The mean values of the cells in a GEDI footprint was used as the parameter to gather reference map values within the zones of

the GEDI shots. These values were joined to the GEDI footprint height and canopy metric data for modelling. The final number of GEDI shots paired with coincident airborne reference AGB and volume was 2,313**.**

2.5GEDI MODELS OF VOLUME AND AGB

Aside from the 1km resolution GEDI AGB product available for download, higher resolution predictive models of volume and AGB from GEDI are currently scarce, particularly for southeastern forests. Here we calibrate volume and AGB models using GEDI L2 and L4 products and high resolution airborne lidar reference maps for the purpose of estimating merchantable timber volume and AGB at a finer scale (25 meters in the case of GEDI). Models predicting footprint-level volume and AGB were produced from the reference map volume and AGB and GEDI L2 products via multiple linear regression. Using the *regsubsets* function in the *leaps* R package [67] in RStudio, L2 variables that correlated with reference volume and AGB were selected by means of forward, backward, and best subsets variable selection. Up to five variables were considered in each model to avoid overfitting the model, and variance inflation factors (VIF's) were examined to avoid collinearity in the models [49]. Models with the highest R^2, and lowest cp, and BIC with VIF's lower than 5 were used on the testing data. A Breusch-Pagan test determined is used to test for heteroscedasticity of the dependent variables. If the residuals were heteroscedastic, then a log-transformation was applied to the variable. Models were subsequently used to predict volume and AGB across all of

the filtered GEDI footprints in the extent of the Sentinel-2 scene (see Sentinel-2 data and processing).

2.6 MODEL VALIDATION

A K-fold cross validation approach was used to test and evaluate the models predictive capabilities, and refit the dependent variables. This is a process by which the data are first split into k number of groups (folds) and then k-1 groups are used to train the data, and the remaining group is used for testing the model. The process is repeated k number of times until each group has been tested, and the mean squared error (MSE) from each test is averaged (the validation error). In this study, a ten-fold validation method was used to train and test the models. Predicted data was back-transformed and the RMSE, %RMSE, MAE, R^2 and bias were calculated from the error between airborne-derived AGB and volume, and predicted GEDI volume and AGB. In this study, validation was carried out by using an error term similar to the one used in a validation study by Fernandez et al. [68]:

$$Error_i = AGB_{ALS\,i} - AGB_{GEDI\,i} \quad (1)$$

where the error of sample "i" of AGB or volume is the difference between the AGB or volume of the ALS and GEDI of "i". From this error, calculations were performed for comparisons including the following (using AGB as an example):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Error_i)^2}{\sum_{i=1}^{n}(AGB_{ALS\,i} - mean(AGB_{ALS}))^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Error_i)^2} \quad (3)$$

$$RMSE\% = \frac{RMSE}{mean(n_{obs})} \times 100 \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Error_i| \quad (5)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n} Error_i \quad (6)$$

2.7 *SENTINEL-2 DATA AND PROCESSING*

GEDI offers dense coverage of lidar data in the form of footprint-level products, or 1 km gridded products, but wall-to-wall estimation of forest characteristics at a finer scale is possible with the use of ancillary imagery data. Previous work has shown that the use of imagery metrics such as NDVI, EVI, texture, and land cover classification correlate well with estimates of forest characteristics [28,69], and can be further leveraged for wall-to-wall mapping when combined with GEDI lidar. In this study, Sentinel-2A, L1C data were processed to produce such metrics (Table 1) for estimating volume and AGB across the entire Sentinel image. Leaf-off Sentinel-2 L1C data for the date of November, 2021 were downloaded for a 290km area containing the extent of the study site, and the entirety of the scene was further processed and evaluated. The Sentinel image was downloaded from the USGS Earth explorer website [70] and contained 13 bands, four of which have a spatial resolution of 10m, and were central to this study: band 2 blue (490nm), band 3 green (560nm), band 4 red (665), and band 8 Near Infra-Red (NIR)

(842nm). After the data were downloaded, the Sentinel Application Platform (SNAP) application was used to atmospherically correct the imagery [71]. To prepare the data for mapping, commonly used imagery metrics (table 1**)** were derived at the resolution of the GEDI footprints (25m). Texture metrics, also known as second order co-occurrence layers were developed in ENVI [72]. Texture metrics are calculated from the distance and angularity of pixel values surrounding a pixel for which texture is being calculated. The purpose of these layers is to distinguish forest characteristics in the image that may relate to volume and AGB. Other imagery-derived layers included normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and land cover, which were computed using ArcGIS Pro software. NDVI is a measure of greenness in vegetation, and is often used to assess forest health, but has been demonstrated as a useful metric for estimating AGB [28,73]. NDVI was calculated from the red and NIR bands:

$$NDVI = (NIR - Red)/(NIR + Red) \quad (7)$$

Similarly, EVI is a metric used for estimating forest health but can be particularly useful in areas with high vegetation density, like those in this study area. EVI leverages the blue band in addition to the red and NIR bands and is calculated as:

$$EVI = 2.5 * ((NIR - Red) / (NIR + 6 * Red - 7.5 * Blue + 1)) \quad (8)$$

Finally, land cover variables were derived from a classification of the leaf-off Sentinel-2 imagery using ENVI software [72]. First, the entire image was classified into seven main landcover types: developed, barren, agriculture, water, deciduous, mixed forest, and evergreen forest. This was done using a maximum likelihood classifier. After the classification, a confusion matrix was used to assess the accuracy of the classification output, compared with randomly selected ground truth data points within the study area. The overall accuracy (the total number of correctly classified pixels divided by the total number of pixels) was 79%. The kappa coefficient or the measure of agreement between truth and classified pixels classes, was also used to assess the classification accuracy. This measure ranges from 0 (not accurate at all) to 1 (all pixels correctly classified). Our classification had a kappa coefficient of 0.73, which we determined was suitable for this study. Once the classification was complete, the three forest land cover types were extracted from the classification.

Table 1: Imagery derived multispectral variables used for mapping.

Variable	Description
1 con_1,2,3,4	Texture contrast for bands Blue, green, red, NIR
2 cor _1,2,3,4	Texture correlation bands blue, green, red, NIR
3 dis _1,2,3,4	Texture dissimilarity bands blue, green, red, NIR
4 ent _1,2,3,4	Texture entropy bands blue, green, red, NIR
5 hom _1,2,3,4	Texture homogeneity bands blue, green, red, NIR
6 mean_1,2,3,4	Mean texture bands blue, green, red, NIR
7 secmom_1,2,3,4	Angular Second Moment bands blue, green, red, NIR
8 var_1,2,3,4	Texture variance bands blue, green, red, NIR
9 NDVI	Normalized Difference Vegetation Index
10 EVI	Enhanced Vegetation Index
11 LC	Forest land cover including evergreen, deciduous, and mixed.

2.8 FOREST MAPPING

Wall-to-wall mapping was achieved by first relating footprint level estimates of volume and AGB to imagery metrics using a machine learning approach and then using the model to extrapolate AGB and volume across the forested areas of the study site. The approach used in this study takes advantage of the machine learning random forest-based regression algorithm in the R statistical package *ModelMap* [74]. Random Forests is a machine learning algorithm developed by Leo Breiman's [75], in which decision trees are developed based on random assortments of the training data. The trees then make predictions on the outcome based on the data within the tree, and each tree votes on which outcome is most accurate. The model cannot predict data that lies outside the numeric range of the training data, and this could lead to over or under estimations of truth in a prediction should the model be predicting outside of the area where the data were collected. Input data in the forest model included the variable to predict (volume or AGB) and the variables derived from the Sentinel-2 imagery. The method is considered a supervised machine learning technique, as the user inputs and modifies a number of hyperparameters to adjust the outcome of the forest model. These hyperparameters include the number of trees in the forest, the leaf size for each of the trees, and maximum tree depth. The number of trees in the forest is the total number of decision trees in the model, and typically the more trees in the model, the more accurate the model will predict. In this study, the number of trees was optimized based on an

initial number of runs testing each model diagnostic. The minimum leaf size is the smallest number of nodes on a tree, or the lowest number of observations used. This hyperparameter is known to cause some overfitting in the model if it is set high relative to the number of trees where the data contains a lot of noise. Therefore, in this study the minimum leaf size was set to 5, which is the default for statistical package. Maximum tree depth is a parameter that determines the maximum number of splits in a tree, and is a data driven parameter. Other parameters include the data available to each tree, and the amount of training data excluded for validation. Although the modelmap package can be used right away to train and predict, model training was done a number of times to find a balance of the hyperparameters that would give the most accurate model before predicting and outputting a raster. Furthermore, the number of data excluded for validation was initially set to 0% to find stable variable importance, and tune hyperparameters. Next, training data was excluded for validation and was set to 30% to test and evaluate the final model prediction. The output maps were generated at the same grid size as the GEDI footprints (25-meters). Model predictions were evaluated using the R^2 and RMSE.

3.RESULTS

3.1GEDI FOOTPRINT LEVEL ESTIMATES OF VOLUME AND AGB

Models were produced using GEDI L2 and L4A footprint metrics, and the reference values of volume and AGB spatially coincident with the GEDI footprints. In model building, the best models were composed of five GEDI variables. None of the variables in the best models had VIF values over 5, and none of the models showed heteroscedasticity according to the Breusch-Pagan test. For volume, the best model included the AGBD, cover, pavd1, fhd, and rh90 variables (Table 2). Using a 10-fold cross validation method to refit and test the model, an R^2 of 0.36 was produced, with an RMSE of 36.45 m^3 ha^{-1}. In this model, the AGBD and rh90 variables had a Pearson's correlation higher than 0.50. The cover and pavd1 variables correlated less well with volume ($R < 0.20$) and the fhd variable had a Pearson's R correlation of 0.40. Figure 1 contains a graph of the predicted and observed footprint level predictions for volume.

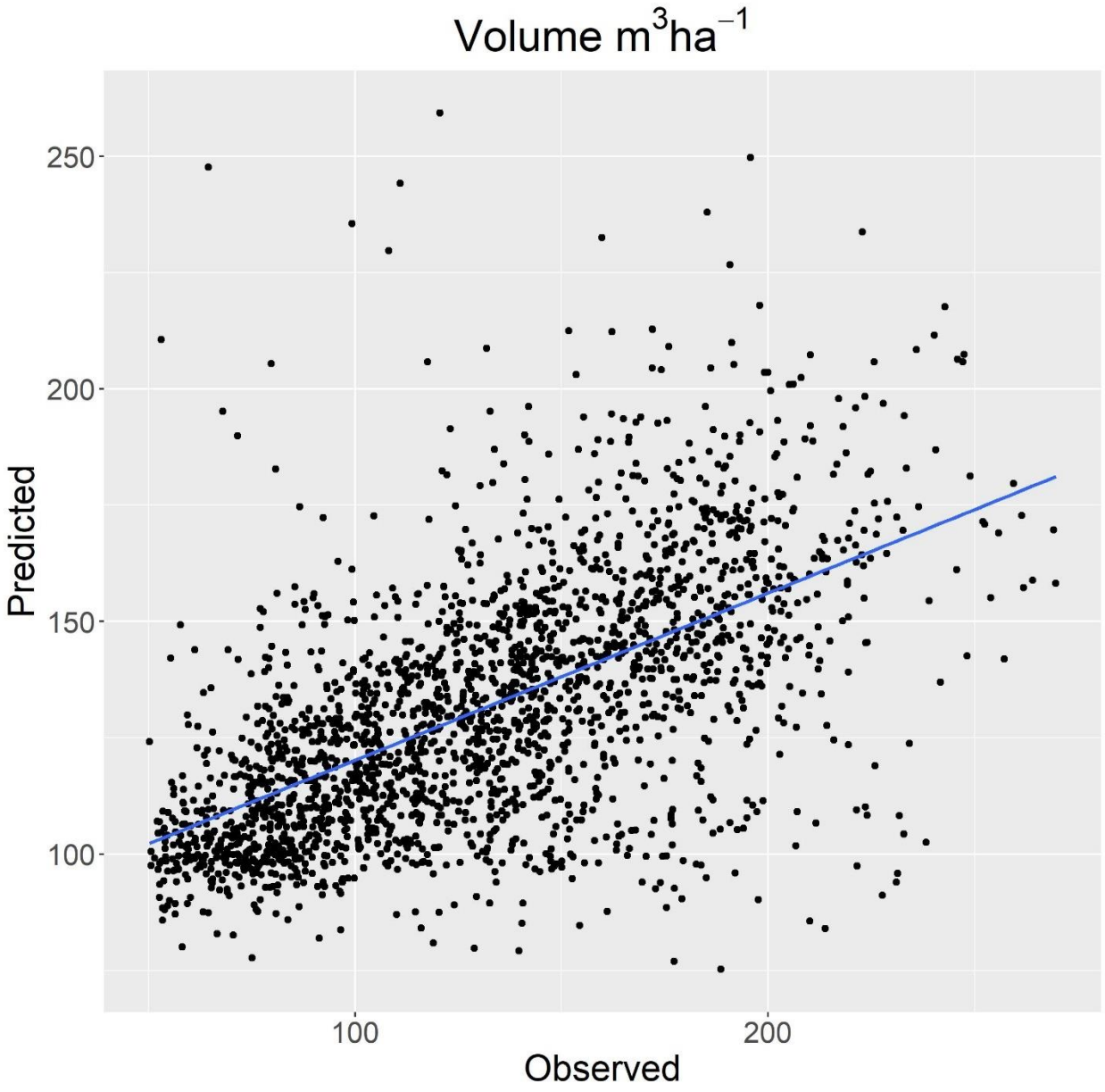

Figure 1: Scatterplot of observed and predicted Volume from GEDI footprint models.

Similarly, the model predicting biomass most accurately was the five-variable model with variables including AGBD, cover, pavd1, fhd, and rh80. This model had an R^2 of 0.30 and an RMSE of 15.60 Mg ha^{-1}. Like the variables in the volume model, those in the AGB that correlated best (Pearson's R < 0.50) with AGB, were the AGBD and rh80 variables. The next best correlated variable was fhd with a Pearson's correlation of 0.39. Finally, the least correlated variables were cover and pavd1 (R < 0.20). Figure 2 includes a graph showing the relationship between footprint level observed and predicted AGB.

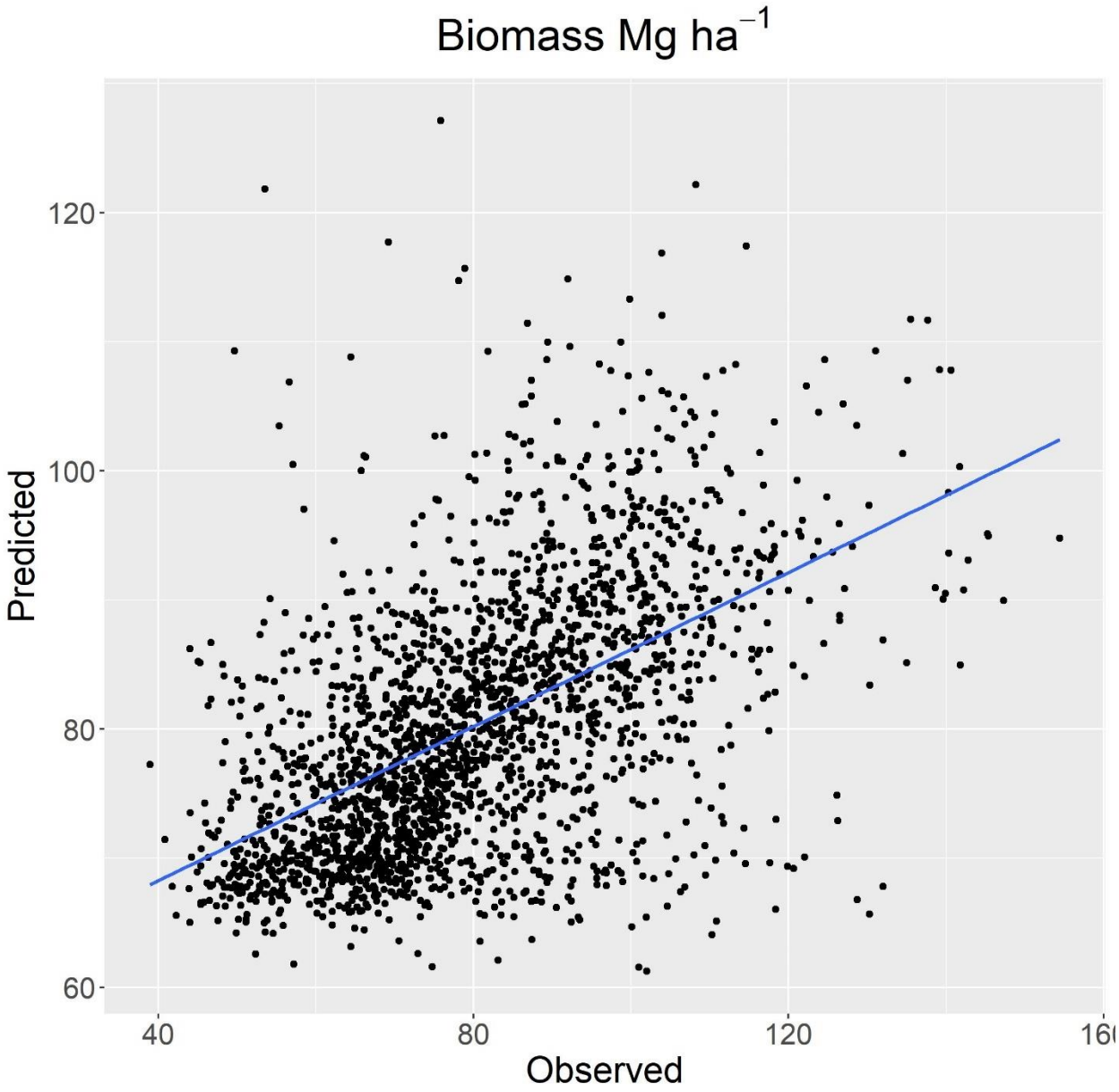

Figure 2: Scatterplot of observed and predicted AGB for GEDI footprint models.

Table 2: GEDI models predicting footprint level volume and AGB with evaluation metrics derived from observed and predicted values.

Forest Metric	Model	R^2	RMSE	%RMSE	MAE	Bias
Volume	Vol = 132 + 0.221agbd – 44.8cover – 43.2pavd1 – 22.5fhd + 3.54rh90	0.36	36.45	27.65	28.68	$-7.134x10^{-16}$
AGB	AGB = 74.7 + 0.0804agbd -12.9cover-15.6pavd1 – 5.10fhd+1.18rh80	0.30	15.60	19.37	11.81	$-7.276x10^{-16}$

3.2WALL-TO-WALL MAPS OF VOLUME AND AGB

Estimates of volume and AGB were upscaled across a 30x30 km extent using the Sentinel-2 image. The decision to use only part of the scene was to restrict variation in land cover that would otherwise occur beyond areas related to lumber plantation. Model diagnostics from the random forest model provided information on the accuracy of the map predictions as well as the importance of variables in the model. Both models were optimized using 500 trees, and both models were set to automatically tune the number of predictors used to select the split in the trees. For volume the model optimized 6 randomly sampled variables at each split. After testing the model on 30% of the data left out for validation, the model R^2 was 0.11 and the RMSE was 30.55 m^3 ha^{-1}. The chart below (figure 3) displays the most important variables for the volume model, and in general the top 5 most important variables were associated with the NIR band. In modeling AGB, the model also optimized 6 randomly sampled variables at each split. After testing the model on 30% of the data left out for validation, the model R^2 was 0.10 and the RMSE was 11.08 Mg ha^{-1}. Variable importance again included those

that were associated with the NIR band (figure 4). Map outputs of the forest volume and AGB at a 25-m resolution are displayed below (figure 5 and figure 6).

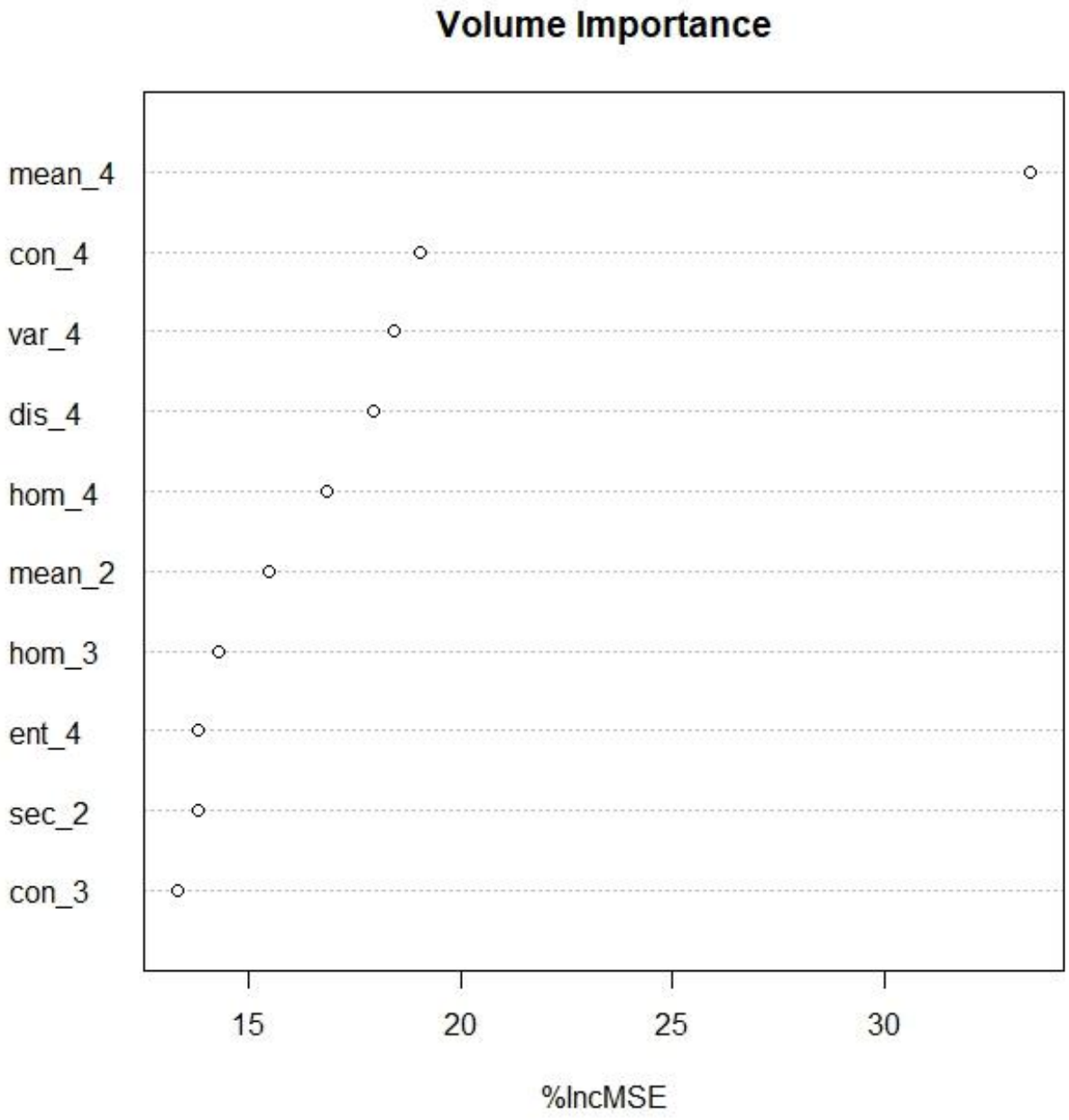

Figure 3: Variable importance plot for RF volume model. Most important variable to model is towards the top of the plot and is based on the % increase in MSE were the variable removed from the model.

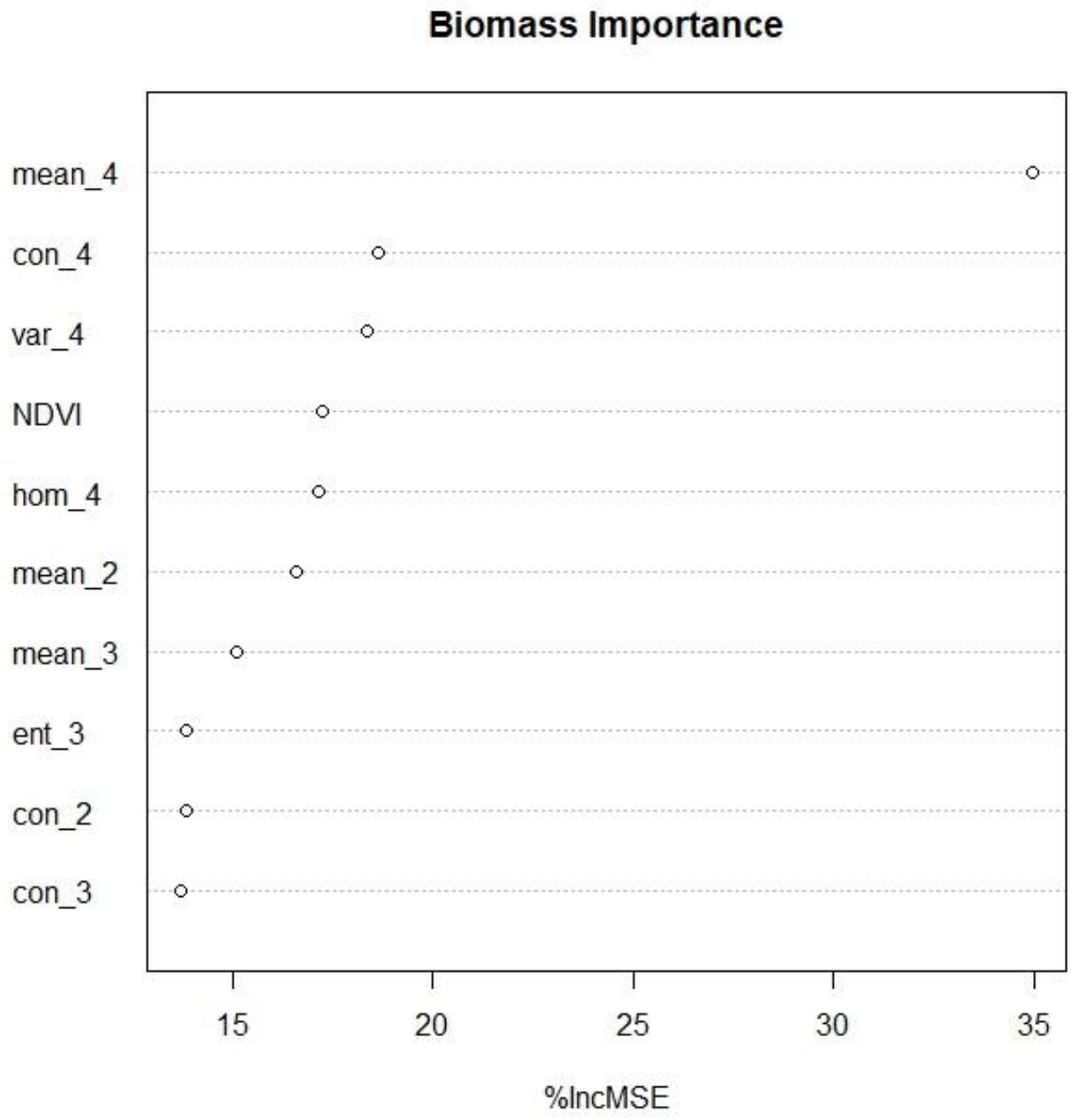

Figure 4: Variable importance plot for RF AGB model. Most important variable to model is towards the top of the plot and is based on the % increase in MSE were the variable removed from the model.

Volume

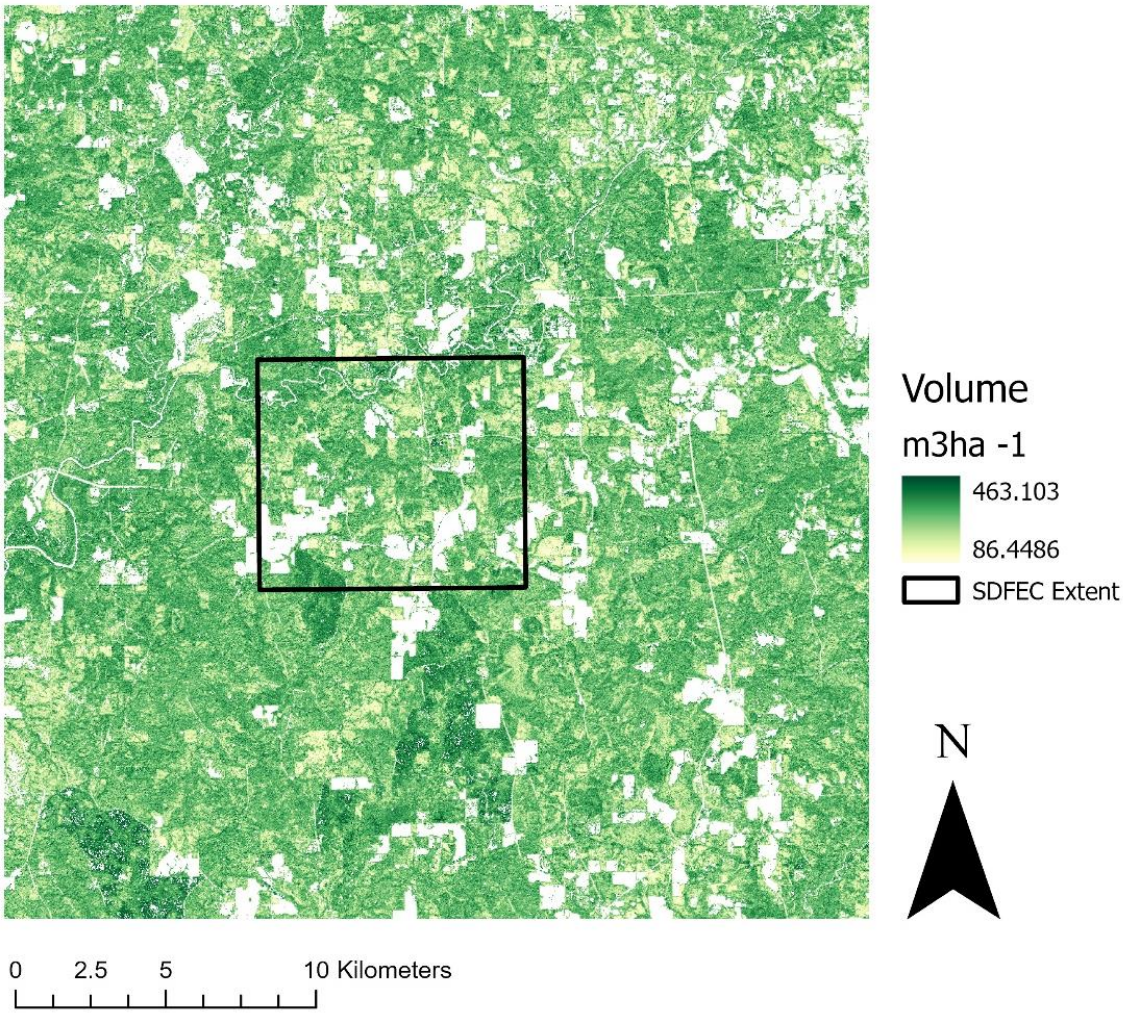

Figure 5: Map output of volume across extent. Maps were developed from the relationship between Senitnel-2 layers (NDVI, EVI, texture, etc.) and predicted GEDI footprint values.

Biomass

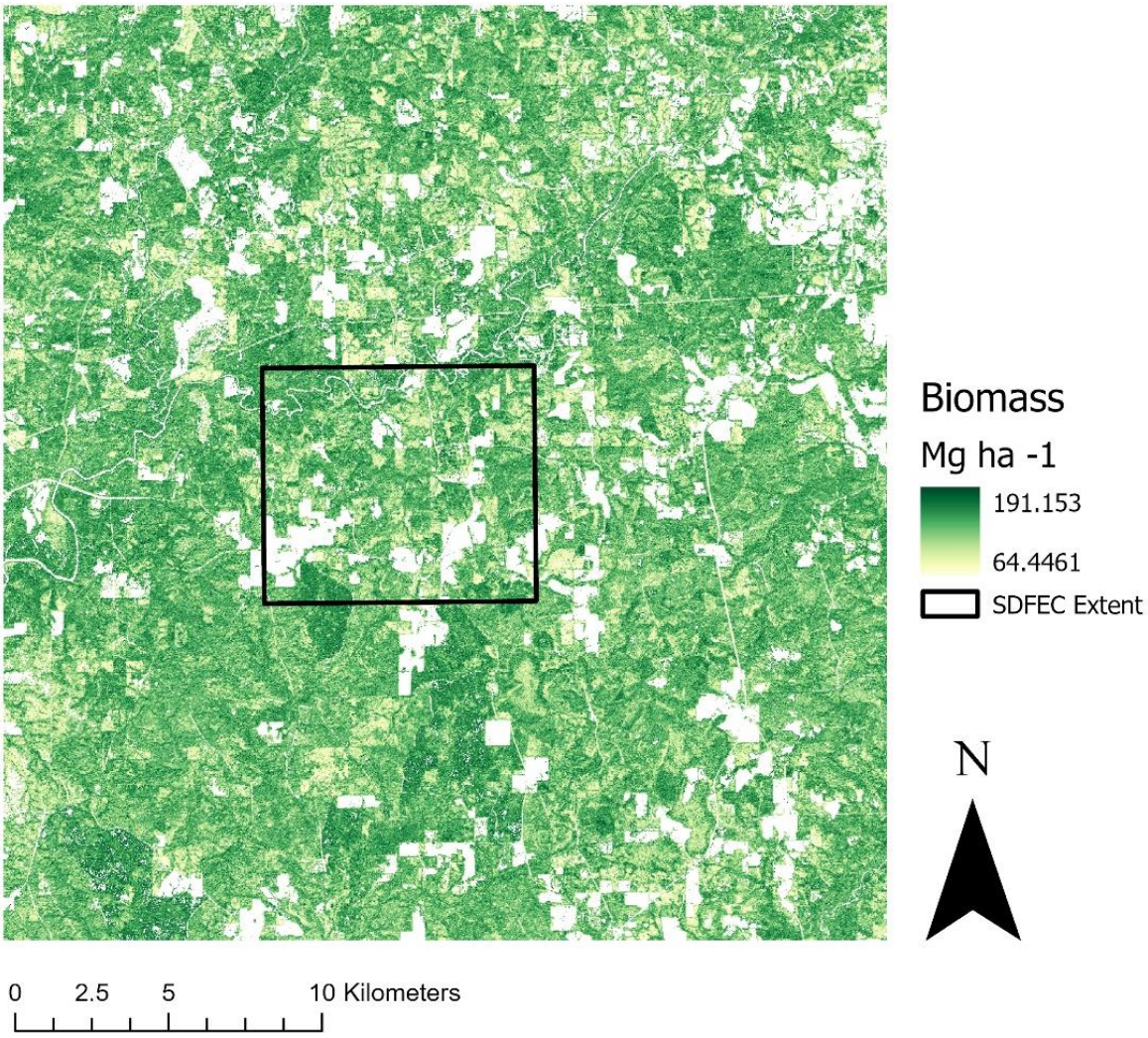

Figure 6: Map output of AGB across extent. Maps were developed from the relationship between Senitnel-2 layers (NDVI, EVI, texture, etc.) and predicted GEDI footprint values.

4.DISCUSION

In this study, we demonstrated a method to upscale AGB and volume predictions as wall-to-wall maps using airborne lidar reference maps, GEDI lidar data, and Sentinel-2 data. The models produced from GEDI variables to predict footprint volume and AGB were not very accurate (R^2 = 0.30 to 0.36), however this is likely due to the high error associated with the reference maps which propagated into the GEDI footprint model estimates. In a previous study [50], we discuss the error of the reference maps being attributed to the highly selective process of the field data used to produce the maps. Furthermore, the highly selective field data used to produce the airborne reference maps, and subsequently the GEDI footprint models differ from the GEDI L4 product whose AGBD is produced from a less selective subset of vegetation in a given footprint. In other words, many species of trees in a field plot from our study area were not tallied, while L4A footprints were far less selective [76]. This likely led to inaccuracies when building the GEDI footprint-based model in this study, as the dependent variable from the reference maps only represent a subset of total vegetation occurring in the coincidental GEDI footprint, compared to the GEDI L4A product.

Unexplained variability in the models may also be attributed to the temporal difference in the reference maps (2017) and the GEDI data (2020-2021), however it has been well documented that temporal difference has little effect on the accuracies of a model built with temporally variable data [7]. For example, one study demonstrated that a 7-year

temporal difference in field data and remote-sensing maps had little effect on the precision of modeling biomass, so long as the forest itself has not been drastically changed [77]. Local factors such as soil type, and water availability could greatly affect the volume and AGB of a forest stand, and could be added as random effects variables in future studies. Consideration of other model types that consider geographic variation such as a geographically weighted regression could be beneficial at greater extents [73]. Another useful method to deal with variability caused by changes in local factors across a geographic range would be the synthesis of spatially and temporally coincident ICESat-2 and GEDI data, which has been demonstrated by simulated GEDI data and ICESat-2 data [51].

To maximize the accuracy of the footprint level predictions, we selected the five variable models, and though these models showed no signs of collinearity, it is important to note that the models with height metrics alone were not considerably less accurate (see appendix A2 for all models). Furthermore, while a number of variables could have been added to the model to increase the explained variability, parsimony is crucial to a successful predictive regression model as it should be able to perform well each time it is used [23]. Other studies have demonstrated that height variables typically dominate the predictive capabilities of regression models. For example, a study estimating longleaf pine volume in Brazilian plantations used just two height variables and managed to produce a model that could account for 97% of the variation

in predicted volume [61]. Many other studies have used height variables in conjunction with a canopy cover and/or canopy density variable to account for volume or AGB variability, a very well demonstrated method reflective of the volume model in this paper [20,78]. While it is expected that height variables would account for the majority of the variability in the model, canopy cover did contribute to the accuracy of the model, and is consistent with previous work demonstrating the importance of canopy cover as a biomass and volume predictor [76]. Another important variable in the footprint level models for AGB and volume was the foliage-height diversity variable. Because the SDFEC includes both naturally regenerated stands, and even age planted stands, the diversity of foliage density between stands would cause a lot of variation in model predictions, however the inclusion of this variable is able to account for some of this variability. Other studies have also demonstrated that this variable is increases the accuracy of models with complex forest structure [42,49].

After developing models and predicting footprint level volume and AGB across the extent of the downloaded Sentinel-2 scene, wall-to-wall maps were developed. Similar to the GEDI models, the forest-based regression predictions of volume and AGB were not accurate (R^2 = 0.11 & 0.10). Potential sources of error include inaccuracies of the airborne lidar reference maps [50] and the subsequent inaccuracies of the GEDI footprint estimates of volume and AGB. One study used airborne lidar reference maps and ICESat-2 data in conjunction with Landsat data to estimate AGB in southeastern

Texas and Alabama with good results (R^2 = 0.58), noting that errors in their estimates occurred due to a lack of ground data and a temporal variability between their reference map and ICESat-2 data [28]. This result supports our findings in that the accuracy of field data (or lack thereof) can lead to inaccuracies in model predictions. Other studies that upscale space-based forest metric predictions using multispectral imagery have found moderate accuracy. For example, researchers estimated canopy heights in northeastern China using ICESat-2, Sentinel 1 and 2, and Landsat, and found accuracies ranging from R^2 = 0.68 to 0.78 [79]. The variable importance measures revealed both interesting and expected results. For example, in both random forests models, EVI was not among the top ten important variables in the model, which contradicts a previous study using Sentinel-2 variables to predict AGB and volume in the same study site [50]. Similarly, the NDVI variable was only among the top ten most important variables in the model predicting AGB, but not for the model predicting volume. We did find that the mean_4 and con_4 variables were the top two most important variables in both models, which has been demonstrated as an important variable in other studies as well [55,79]. Because the mean_4 and con_4 variables are derived from the NIR band from Sentinel-2, they are capable of determining distinguishing vegetation than the other bands [69]. These results are an important insight for distinguishing the variation in dense forests like those in the southeastern U.S., particularly where calibrating a model for estimating merchantable timber volume

is the goal. Furthermore, previous studies have demonstrated that dense and highly variable forests can cause more difficulties when modeling forest parameters [22,80], and the results of the random forests model offer insight into what variables are most important for estimating forest metrics like AGB and volume.

5.CONCLUSIONS

Mapping volume and AGB across great extents using remote sensing data is the standard for modern ecologists and foresters interested in landscape level ecology and carbon estimation. However, despite the wide range of data sources available to researchers in the present era, methods to efficiently map volume and AGB at large extents and fine scales are still difficult, and efforts to map merchantable volume are few. Here we present an approach to developing wall-to-wall maps of merchantable timber volume and AGB by taking advantage of the extent and density of GEDI lidar footprint data in the southern Alabama, and the L4 AGBD product as a variable in modeling. By first estimating footprint level volume and AGB, other ancillary data sources can be leveraged to create wall-to-wall maps of volume and AGB at finer extents (25m) than those available in the L4 gridded product (1km). Limitations of this study include the inaccuracies associated with the reference maps, which caused high unexplained error in the final maps. Furthermore, despite the high spatial density of GEDI footprint data, spatial gaps can still cause unexplained variation in the capabilities of predictive models. We did however find usefulness in the GEDI L4A

AGBD as a predictor variable for modeling merchantable timber volume and AGB. This study also demonstrates the usefulness of Sentinel-2 variables, particularly those that take advantage of the NIR band, for upscaling estimates of volume and AGB across larger extents. Further work should be carried out to estimate the uncertainty propagated from the data sources used to produce the final product [7]. Further investigation of methods for efficiently and accurately mapping ecosystems at fine scales may continue to be the most important aspect of ecological research in the future, as climate change and anthropogenic altercations to ecosystems create new problems for both humankind and organisms in the biosphere. This fact will have a severe influence on timber production, as important regulations will shift the way foresters log in their forests [5]. Therefore, estimating merchantable timber volume and AGB is crucial to economies in the southeastern U.S., who heavily rely on the forest industry. The results of this study demonstrate a potential method for helping investigate these problems, and future studies can benefit from using adaptations of this method with improved techniques and technologies.

6.LITERATURE CITED

1. UNFCCC *The Paris Climate Agreement*; United Nations, 2016; Vol. 2, pp. 78–79;.

2. Brooks, T.M.; Mittermeier, R.A.; Mittermeier, C.G.; Da Fonseca, G.A.B.; Rylands, A.B.; Konstant, W.R.; Flick, P.; Pilgrim, J.; Oldfield, S.; Magin, G.; et al. Habitat Loss and Extinction in the Hotspots of Biodiversity. *Conserv. Biol.* **2002**, *16*, 909–923, doi:10.1046/j.1523-1739.2002.00530.x.

3. Houghton, R.A.; Hall, F.; Goetz, S.J. Importance of Biomass in the Global Carbon Cycle. *J. Geophys. Res. Biogeosciences* **2009**, *114*, 1–13, doi:10.1029/2009JG000935.

4. Houghton, R.A. The Annual Net Flux of Carbon to the Atmosphere from Changes in Land Use 1850-1990. *Tellus, Ser. B Chem. Phys. Meteorol.* **1999**, *51*, 298–313, doi:10.3402/tellusb.v51i2.16288.

5. UNFCCC Key Decisions Relevant for Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+). *Framew. Conv. Clim. Chang.* **2016**, 44.

6. Houghton, R.A. Aboveground Forest Biomass and the Global Carbon Balance. *Glob. Chang. Biol.* **2005**, *11*, 945–958, doi:10.1111/j.1365-2486.2005.00955.x.

7. Duncanson, L.; Armston, J.; Disney, M.; Avitabile, V.; Barbier, N.; Calders, K.; Carter, S.; Chave, J.; Herold, M.; MacBean, N.; et al. *Aboveground Woody Biomass Product Validation Good Practices Protocol*; 2021;

8. Mitchard, E.T.A. The Tropical Forest Carbon Cycle and Climate Change. *Nature* **2018**, *559*, 527–534, doi:10.1038/s41586-018-0300-2.

9. Gu, H.; Williams, C.A.; Hasler, N.; Zhou, Y. The Carbon Balance of the Southeastern U.S. Forest Sector as Driven by Recent Disturbance Trends. *J. Geophys. Res. Biogeosciences* **2019**, *124*, 2786–2803, doi:10.1029/2018JG004841.

10. Bond, B. PB1650 Understanding Log Scales and Log Rules. *UT Ext. Publ. For. Trees, Timber* **2011**, 8.

11. Fang, J.Y.; Wang, G.G.; Liu, G.H.; Xu, S.L. Forest Biomass of China: An Estimate Based on the Biomass-Volume Relationship. *Ecol. Appl.* **1998**, *8*, 1084–1091, doi:10.1890/1051-0761(1998)008[1084:FBOCAE]2.0.CO;2.

12. Shifley, S.R.; Thompson, F.R.; Dijak, W.D.; Larson, M.A.; Millspaugh, J.J. Simulated Effects of Forest Management Alternatives on Landscape Structure and Habitat Suitability in the Midwestern United States. *For. Ecol. Manage.* **2006**, *229*, 361–377, doi:10.1016/j.foreco.2006.04.030.

13. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar Sampling for Large-Area Forest Characterization: A Review. *Remote Sens. Environ.* **2012**, *121*, 196–209, doi:10.1016/j.rse.2012.02.001.

14. Elzinga, C.; Shearer, R.C.; Elzinga, G. Observer Variation in Tree Diameter Measurements. *West. J. Appl. For.* **2005**, *20*, 134–137, doi:10.1093/wjaf/20.2.134.

15. Wulder, M.A.; Bater, C.W.; Coops, N.C.; Hilker, T.; White, J.C. The Role of LiDAR in Sustainable Forest Management. **2008**, *84*.

16. Popescu, S.C.; Wynne, R.H.; Scrivani, J.A. Fusion of Small-Footprint Lidar and Multispectral Data to Estimate Plot-Level Volume and Biomass in Deciduous and Pine Forests in Virginia, USA. *For. Sci.* **2004**, *50*, 551–565, doi:10.1093/forestscience/50.4.551.

17. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Fennema, S.J.; Corte, A.P.D. Modeling and Mapping Basal Area of Pinus Taeda L. Plantation Using Airborne LiDAR Data. *An. Acad. Bras. Cienc.* **2017**, *89*, 1895–1905, doi:10.1590/0001-3765201720160324.

18. Næsset, E. Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data. *Remote Sens. Environ.* **1997**, *61*, 246–253, doi:10.1016/S0034-4257(97)00041-2.

19. Jaskierniak, D.; Lane, P.N.J.; Robinson, A.; Lucieer, A. Extracting LiDAR Indices to Characterise Multilayered Forest Structure Using Mixture Distribution Functions. *Remote Sens. Environ.* **2011**, *115*, 573–585, doi:10.1016/j.rse.2010.10.003.

20. Sheridan, R.D.; Popescu, S.C.; Gatziolis, D.; Morgan, C.L.S.; Ku, N.W. Modeling Forest Aboveground Biomass and Volume Using Airborne LiDAR Metrics and Forest Inventory and Analysis Data in the Pacific Northwest. *Remote Sens.* **2015**, *7*, 229–255, doi:10.3390/rs70100229.

21. Ahl, R.; Hogland, J.; Brown, S. A Comparison of Standard Modeling Techniques Using Digital Aerial Imagery with National Elevation Datasets and Airborne LiDAR to Predict Size and Density Forest Metrics in the Sapphire Mountains MT, USA. *ISPRS Int. J. Geo-Information* **2019**, *8*, doi:10.3390/ijgi8010024.

22. Spriggs, R.A.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A Critique of General Allometry-Inspired Models for Estimating Forest Carbon Density from Airborne LiDAR. *PLoS One* **2019**, *14*, 1–19, doi:10.1371/journal.pone.0215238.

23. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T.; Jain, T.B.; Evans, J.S. Transferability of

Lidar-Derived Basal Area and Stem Density Models within a Northern Idaho Ecoregion. *Can. J. Remote Sens.* **2018**, *44*, 131–143, doi:10.1080/07038992.2018.1461557.

24. Lister, A.J.; Andersen, H.; Frescino, T.; Gatziolis, D.; Healey, S.; Heath, L.S.; Liknes, G.C.; McRoberts, R.; Moisen, G.G.; Nelson, M.; et al. Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory. *Forests* **2020**, *11*, 1–41, doi:10.3390/f11121364.

25. White, J.C.; Wulder, M.A.; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. *A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach*; 2013; Vol. 89; ISBN 9781100223858.

26. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273.

27. Popescu, S.C.; Zhao, K.; Neuenschwander, A.; Lin, C. Satellite Lidar vs. Small Footprint Airborne Lidar: Comparing the Accuracy of Aboveground Biomass Estimates and Forest Structure Metrics at Footprint Level. *Remote Sens. Environ.* **2011**, *115*, 2786–2797, doi:10.1016/j.rse.2011.01.026.

28. Narine, L.L.; Popescu, S.C.; Malambo, L. Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example. *Remote Sens.* **2020**, *12*, 1–16, doi:10.3390/rs12111824.

29. Nelson, R.; Margolis, H.; Montesano, P.; Sun, G.; Cook, B.; Corp, L.; Andersen, H.E.; deJong, B.; Pellat, F.P.; Fickel, T.; et al. Lidar-Based Estimates of Aboveground Biomass in the Continental US and Mexico Using Ground, Airborne, and Satellite Observations. *Remote Sens. Environ.* **2017**, *188*, 127–140, doi:10.1016/j.rse.2016.10.038.

30. Neuenschwander, A.; Pitts, K. Ice, Cloud, and Land Elevation Satellite 2 (ICESat-2) Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-Track Products (ATL08). **2020**, *5*, 1–140.

31. Narine, L.L.; Popescu, S.; Neuenschwander, A.; Zhou, T.; Srinivasan, S.; Harbeck, K. Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data. *Remote Sens. Environ.* **2019**, *224*, 1–11, doi:10.1016/j.rse.2019.01.037.

32. Schutz, B.E.; Zwally, H.J.; Shuman, C.A.; Hancock, D.; DiMarzio, J.P. Overview of

the ICESat Mission. *Geophys. Res. Lett.* **2005**, *32*, 1–4, doi:10.1029/2005GL024009.

33. Nelson, R.; Ranson, K.J.; Sun, G.; Kimes, D.S.; Kharuk, V.; Montesano, P. Estimating Siberian Timber Volume Using MODIS and ICESat/GLAS. *Remote Sens. Environ.* **2009**, *113*, 691–701, doi:10.1016/j.rse.2008.11.010.

34. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D.; et al. The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation. *Remote Sens. Environ.* **2017**, *190*, 260–273, doi:10.1016/j.rse.2016.12.029.

35. Nandy, S.; Srinet, R.; Padalia, H. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophys. Res. Lett.* **2021**, *48*, doi:10.1029/2021GL093799.

36. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002, doi:10.1016/j.srs.2020.100002.

37. Dubayah, R.O.; Armston, J.; Kellner, J.R.; Duncanson, L.; Healey, S.P.; Patterson, P.L.; Hancock, S.; Tang, H.; Hofton, M.A.; Blair, J.B.; et al. *GEDI L4A Footprint Level Aboveground Biomass Density Version 1*; Oak Ridge, Tennessee, 2021;

38. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Estimation with GEDI LIDAR Waveforms and Bayesian Deep Learning. **2021**.

39. Næsset, E.; Gobakken, T. Estimation of Above- and below-Ground Biomass across Regions of the Boreal Forest Zone Using Airborne Laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090, doi:10.1016/j.rse.2008.03.004.

40. Zhao, K.; Popescu, S.; Nelson, R. Lidar Remote Sensing of Forest Biomass: A Scale-Invariant Estimation Approach Using Airborne Lasers. *Remote Sens. Environ.* **2009**, *113*, 182–196, doi:10.1016/j.rse.2008.09.009.

41. Neuenschwander, A.; Guenther, E.; White, J.C.; Duncanson, L.; Montesano, P. Validation of ICESat-2 Terrain and Canopy Heights in Boreal Forests. *Remote Sens. Environ.* **2020**, *251*, 112110, doi:10.1016/j.rse.2020.112110.

42. Wang, C.; Elmore, A.J.; Numata, I.; Cochrane, M.A.; Huang, J.; Zhao, Y.; Li, Y.; Wang, C.; Elmore, A.J.; Numata, I.; et al. Factors Affecting Relative Height and Ground Elevation Estimations of GEDI among Forest Types across the Conterminous USA ABSTRACT. *GIScience Remote Sens.* **2022**, *59*, 975–999,

doi:10.1080/15481603.2022.2085354.

43. Chen, L.; Ren, C.; Zhang, B.; Wang, Z.; Liu, M.; Man, W.; Liu, J. Improved Estimation of Forest Stand Volume by the Integration of GEDI LiDAR Data and Multi-Sensor Imagery in the Changbai Mountains Mixed Forests Ecoregion (CMMFE), Northeast China. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102326, doi:10.1016/j.jag.2021.102326.

44. Fayad, I.; Baghdadi, N.; Alvares, C.A.; Stape, J.L.; Bailly, J.S.; Scolforo, H.F.; Zribi, M.; Le Maire, G. Estimating Canopy Height and Wood Volume of Eucalyptus Plantations in Brazil Using Gedi Lidar Data. *Int. Geosci. Remote Sens. Symp.* **2021**, *2021-July*, 5941–5944, doi:10.1109/IGARSS47720.2021.9553091.

45. Vieira, R.; Alberto, C.; North, E.; Cardil, A.; Hamamura, C.; Lopes, B.; Faria, D.; Brancalion, P.H.S.; Eduardo, G.; Paula, A.; et al. Remote Sensing of Environment Large Scale Multi-Layer Fuel Load Characterization in Tropical Savanna Using GEDI Spaceborne Lidar Data. **2022**, *268*, doi:10.1016/j.rse.2021.112764.

46. Krofcheck, D.J.; Hurteau, M.D.; Scheller, R.M.; Loudermilk, E.L. Prioritizing Forest Fuels Treatments Based on the Probability of High-Severity Fire Restores Adaptive Capacity in Sierran Forests. *Glob. Chang. Biol.* **2018**, *24*, 729–737, doi:10.1111/gcb.13913.

47. MacArthur, R.H.; MacArthur, J.W.; Preer, J. On Bird Species Diversity. II. Prediction of Bird Census from Habitat Measurements. *Am. Nat.* **1962**, *96*, 167–174, doi:10.1086/282219.

48. Burns, P.; Clark, M.; Salas, L.; Hancock, S.; Leland, D.; Jantz, P.; Dubayah, R.; Goetz, S.J. Incorporating Canopy Structure from Simulated GEDI Lidar into Bird Species Distribution Models. *Environ. Res. Lett.* **2020**, *15*, doi:10.1088/1748-9326/ab80ee.

49. Dorado-Roda, I.; Pascual, A.; Godinho, S.; Silva, C.A.; Botequim, B.; Rodríguez-Gonzálvez, P.; González-Ferreiro, E.; Guerra-Hernández, J. Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. **2021**, doi:10.3390/rs13122279.

50. 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*, doi:https://doi.org/10.3390/rs14112708.

51. Silva, C.A.; Duncanson, L.; Hancock, S.; Neuenschwander, A.; Thomas, N.; Hofton, M.; Fatoyinbo, L.; Simard, M.; Marshak, C.Z.; Armston, J.; et al. Fusing

Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping. *Remote Sens. Environ.* **2021**, *253*, 112234, doi:10.1016/j.rse.2020.112234.

52. Duncanson, L.; Armston, J.; Disney, M.; Avitabile, V.; Barbier, N.; Calders, K.; Carter, S.; Chave, J.; Herold, M.; Crowther, T.W.; et al. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. *Surv. Geophys.* **2019**, *40*, 979–999, doi:10.1007/s10712-019-09538-8.

53. Drusch, M.; Gascon, F.; Berger, M. *Sentinel-2 Mission Requirements Document*; 2010;

54. Chen, L.; Ren, C.; Bao, G.; Zhang, B.; Wang, Z.; Liu, M.; Man, W.; Liu, J. Improved Object-Based Estimation of Forest Aboveground Biomass by Integrating LiDAR Data from GEDI and ICESat-2 with Multi-Sensor Images in a Heterogeneous Mountainous Region. *Remote Sens.* **2022**, *14*, doi:10.3390/rs14122743.

55. Verhelst, K.; Gou, Y.; Herold, M.; Reiche, J. Improving Forest Baseline Maps in Tropical Wetlands Using Gedi-Based Forest Height Information and Sentinel-1. *Forests* **2021**, *12*, 1–16, doi:10.3390/f12101374.

56. Range, B.; Lawrence, S. Level III Ecoregions of Alaska. **2001**, 175.

57. Omernik, J. Map Supplement Ecoregions of the Conterminous United. *Ann. Assoc. Am. Geogr.* **2014**, *77*, 118–125.

58. Smith, L.; Sidney, L. Ecoregions of Alabama and Georgia. *Environ. Prot.* **1995**, 75–75.

59. United States Geolotical Survey (USGS); US Department of Interior 3D Elevation Program Available online: https://www.usgs.gov/core-science-systems/ngp/3dep.

60. UGET Available online: https://ugetdm.com/.

61. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Jaafar, W.S.W.M.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting Stem Total and Assortment Volumes in an Industrial Pinus Taeda L. Forest Plantation Using Airborne Laser Scanning Data and Random Forest. *Forests* **2017**, *8*, 1–17, doi:10.3390/f8070254.

62. Means, J.E.; Acker, S.A.; Fitt, B.J.; Renslow, M.; Emerson, L.; Hendrix, C.J. Predicting Forest Stand Characteristics with Airborne Scanning Lidar. *Photogramm. Eng. Remote Sensing* **2000**, *66*, 1367–1371.

63. Watt, M.S.; Meredith, A.; Watt, P.; Gunn, A. Use of LiDAR to Estimate Stand Characteristics for Thinning Operations in Young Douglas-Fir Plantations. *New Zeal. J. For. Sci.* **2013**, *43*, 1–10, doi:10.1186/1179-5395-43-18.

64. Fischer, B.; Smith, M.; Pau, G. Rhdf5: R Interface to HDF5 2022.

65. RStudio Team RStudio: Integrated Development for R. 2020.

66. ESRI ArcGIS Pro.

67. Lumley, T. Package ' Leaps .' **2020**.

68. Fernandez, J.; Velikova, M.; Glennie, C.L. Validation of ICESat-2 ATL08 Terrain and Canopy Height Retrievals in Tropical Mesoamerican Forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1–1, doi:10.1109/jstars.2022.3163208.

69. Hogland, J.; Anderson, N.; Affleck, D.L.R.; Peter, J.S. Using Forest Inventory Data with Landsat 8 Imagery to Map Longleaf Pine Forest Characteristics in Georgia, USA. *Remote Sens.* **2019**, *11*, doi:10.3390/rs11151803.

70. USGS USGS Earth Explorer Available online: https://earthexplorer.usgs.gov/.

71. ESA SNAP - ESA Application Platform 2022.

72. L3Harris Exelis Visual Information Solutions 2021.

73. Gagliasso, D.; Hummel, S.; Temesgen, H. A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area. *Open J. For.* **2014**, *04*, 42–48, doi:10.4236/ojf.2014.41008.

74. Freeman, A.E.; Frescino, T.; Freeman, M.E. Package ' ModelMap .' **2018**.

75. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:doi.org/10.1023/A:1010933404324.

76. Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S.; et al. Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sens. Environ.* **2022**, *270*, doi:10.1016/j.rse.2021.112845.

77. McRoberts, R.E.; Næsset, E.; Gobakken, T. The Effects of Temporal Differences between Map and Ground Data on Map-Assisted Estimates of Forest Area and Biomass. *Ann. For. Sci.* **2016**, *73*, 839–847, doi:10.1007/s13595-015-0485-6.

78. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Falkowski, M.J.; Smith, A.M.S.; Gessler, P.E.; Morgan, P. Regression Modeling and Mapping of Coniferous Forest Basal Area and Tree Density from Discrete-Return Lidar and Multispectral Satellite Data. *Can. J. Remote Sens.* **2006**, *32*, 126–138, doi:10.5589/m06-007.

79. Li, W.; Niu, Z.; Shang, R.; Qin, Y.; Wang, L.; Chen, H. High-Resolution Mapping of Forest Canopy Height Using Machine Learning by Coupling ICESat-2 LiDAR

with Sentinel-1, Sentinel-2 and Landsat-8 Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102163, doi:10.1016/j.jag.2020.102163.

80. Hogland, J.; Anderson, N.; St. Peter, J.; Drake, J.; Medley, P. Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *ISPRS Int. J. Geo-Information* **2018**, *7*, doi:10.3390/ijgi7040140.

CHAPTER 4

1.CONCLUSIONS

Mapping forest structure has become central to many government and corporate organizations who rely on forests for carbon mitigation, conservation, and/or timber product operations. As discussed in the introduction of this document, many methodologies have been used to estimate forest metrics across an extent. However, as technologies in remote sensing improve, more accurate and higher resolution maps can be developed to support forest management. In the case of southern Alabama where forestry is a leading driver of the economy, mapping forest structure is going to become an essential part of forestry organizations as dramatic changes in the atmosphere and biosphere cause changes in the way forests are managed. In other words, foresters in southern Alabama will need maps of merchantable volume, BA, and AGB in order to find a balance between carbon mitigation, ecosystem health, and monetary gain from timber products. While maps of merchantable forest metrics have been produced commonly by airborne lidar [1,2], few have used spaceborne lidar for this explicit purpose [3], and none to the knowledge of the authors have used GEDI for this purpose in southern Alabama. Furthermore, no study has leveraged the L4A footprint level AGBD variable for calibrating merchantable timber volume and AGB models in southern Alabama. In this thesis, we used airborne lidar and field data to produce reference maps of merchantable timber volume, BA, and AGB across an extent

containing the SDFEC. These reference maps facilitated the investigation of GEDI spaceborne lidar for predictive merchantable timber and AGB mapping. A number of interesting findings come from these studies. For example, despite the evidence that separately, lidar and multispectral imagery have the ability to estimate forest structure [4,5], the combination of these variables do not necessarily improve models with just lidar variables. Furthermore, the linear regression models produced with lidar variables were comparable to the RF models which included both lidar and multispectral variables, except in the case of volume, which was improved by the use of the RF algorithm. There are a number of potential modeling techniques demonstrated for mapping forest structure which may suite a given study area better than others. For example, further work might investigate the use of random effects variables for prediction, or a bimodal distribution model that could help account for understory forest characteristics. However, linear regression modeling and RF have in the literature been consistently appropriate for successfully mapping forest structure. Interestingly, the models predicting volume using airborne lidar data were the most accurate (R^2 = 0.53) when compared to BA (R^2 = 0.39) and AGB (R^2 = 0.41). This is likely because the model used to calculate volume from the field data included a height variable, which helps account for the variation in volume due to tree height measured by lidar, while those for BA and AGB did not. Allometric equations used for calculating BA, volume and AGB from field data could benefit from variables that account for more variation

caused by a tree's unique characteristics such as height and taper. Although the models developed from Chapter 2 were modest (R^2 = 0.39 to 0.53), one of the major discussion points was the residual error in the model potentially caused by the field data. Forest inventory data from the SDFEC were highly selective, excluding non-merchantable trees, and measuring only trees that meet cruising specs. Lidar scanning, on the other hand, estimate all returns within a given pulse, and therefore measure all vegetation in a plot leading to mismatch between the field data and lidar data. Therefore, we highly recommend that foresters measure as much vegetation in a plot as possible, particularly in dense, mixed species forests. While this may take more time per plot for cruising, ultimately it will save time in the long run after highly accurate estimates of desired forest metrics are predicted across an extent. Another potential source of error that future researchers should consider is the difference in management in stand types throughout the study area. Natural stands and plantation stands have very different vegetation profiles, and may ultimately require different models or variables in the models. The forest inventory reference data could, for example, reflect this difference as a variable in modeling, however this study did not have that data. Nevertheless, the maps themselves can be useful for comparing areas of low and high BA, volume and AGB throughout the study site. Furthermore, changes in forest structure can be assessed as the lidar data used in the study (USGS 3DEP) [6] are repeated every three to four years, and most importantly they are free data which is useful for forestry

organizations trying to save on operational costs. Similarly, the GEDI spaceborne lidar data used in the second study (Chapter 3) were free, open-source data, however it is not geographically continuous data like airborne lidar data. Instead, if mapping is desired, researchers should consider a multi-phase approach that includes reference data and ancillary data to achieve spatially continuous maps. In our study, we leveraged Sentinel-2 data to extrapolate reference-based GEDI derived footprint level estimates of merchantable volume and AGB across a greater extent. While GEDI L4A footprint level estimates of AGBD are available as an estimate of biomass, we were more interested in estimating merchantable timber volume and AGB and instead used the L4A product as a variable to help calibrate our models. Although our footprint level models had low accuracy, the process of model building revealed useful results. For example, both of the best models predicting timber volume and AGB used the L4A AGBD variable. This is not surprising as the relationship between merchantable timber volume, AGB, and AGBD is quite strong. This and the height variable in each model contributed the most to explaining variation. More surprisingly was the fact that both models depended heavily on the foliage height diversity (fhd) variable instead of another height variable, demonstrating the importance of estimating variation in forest structure due to vegetation density. Therefore, future models of volume and AGB in dense heterogenous forests like those in southern Alabama would do well to consider forest density-based variables in their models. When relating footprint estimates of volume and AGB to

Sentinel-2 data for wall-to-wall mapping, we found that variables derived from the NIR band were most useful to mapping, but surprisingly the NDVI and EVI variables were not among the most important variables for predicting volume. Using these methods, we were able to develop maps of merchantable volume and AGB at a resolution of 25m, across a large extent (30km x 30km), and assess which variables from GEDI and Sentinel-2 were most important to forest mapping. Future work from this study could include estimating the errors propagated through each model into the final map products. Other work might involve the development of high resolution (1-meter) reference maps through methods such as individual tree detection or by relating highly accurate terrestrial or drone lidar data to airborne or spaceborne lidar. However, the most important work to consider in future analysis as suggested by the entirety of this thesis, is to have field data (or other auxiliary reference data) that completely represents the vegetation measured by the lidar system. This means that foresters should measure every species of tree contributing a significant amount of biomass or volume to a forest stand. Other vegetation types in a plot should be considered for measurement as well, because despite not having financial value, there is considerable value for ecosystem dynamics and carbon stock. Furthermore, in stands where understory vegetation is quite dense, the airborne lidar will measure it and introduce it as residual error in models where the field data only includes large merchantable trees. In both chapters of this thesis, stand types were delineated by major species groups (hardwood, pine, etc.)

and introduced as variables in models, but contributed nothing to the RF models, and were not included in linear regression models. Though anecdotal, we suspect that if the field data included more of the deciduous trees in a given plot, the land cover variable would have contributed significantly more to the models.

2.LITERATURE CITED

1. White, J.C.; Wulder, M.A.; Buckmaster, G. Validating Estimates of Merchantable Volume from Airborne Laser Scanning (ALS) Data Using Weight Scale Data. *For. Chron.* **2014**, *90*, 378–385, doi:10.5558/tfc2014072.
2. Woods, M.; Pitt, D.; Penner, M.; Lim, K.; Nesbitt, D.; Etheridge, D.; Treitz, P. Operational Implementation of a LiDAR Inventory in Boreal Ontario. *For. Chron.* **2011**, *87*, 512–528, doi:10.5558/tfc2011-050.
3. Baccini, A.; Laporte, N.; Goetz, S.J.; Sun, M.; Dong, H. A First Map of Tropical Africa's above-Ground Biomass Derived from Satellite Imagery. *Environ. Res. Lett.* **2008**, *3*, doi:10.1088/1748-9326/3/4/045011.
4. Hogland, J.; Anderson, N.; Affleck, D.L.R.; Peter, J.S. Using Forest Inventory Data with Landsat 8 Imagery to Map Longleaf Pine Forest Characteristics in Georgia, USA. *Remote Sens.* **2019**, *11*, doi:10.3390/rs11151803.
5. Buffington, K.J.; Dugger, B.D.; Thorne, K.M.; Takekawa, J.Y. Statistical Correction of Lidar-Derived Digital Elevation Models with Multispectral Airborne Imagery in Tidal Marshes. *Remote Sens. Environ.* **2016**, *186*, 616–625, doi:10.1016/j.rse.2016.09.020.
6. United States Geolotical Survey (USGS); US Department of Interior 3D Elevation Program Available online: https://www.usgs.gov/core-science-systems/ngp/3dep.

APPENDICES

Appendix A

Table A1. *The best 1-5 variable models for models for BA, volume and AGB developed from the relationship between field inventory based estimates of BA, volume, and AGB, and airborne lidar variables in chapter 2.*

Forest Metric	Model	%RMSE	RMSE	MAE	r^2	Bias
BA_1	BA=2.15 + 0.0101cancovA	37.44	6.865	5.453	0.16	1.333
BA_2	BA=1.93 + 0.0164cancovA-2.62dens_0_10	32.81	6.016	4.741	0.31	1.056
BA_3	BA=1.94+0.0171cancovA-1.95dens_0_10-0.910dens_10_15	31.91	5.850	4.584	0.34	1.005
BA_4	BA=1.86+0.0181cancovA-2.06dens_0_10-1.017dens_10_15+0.0240var_3	31.59	5.792	4.591	0.36	0.9676
BA_5	BA=1.78+0.0173cancovA-1.73dens_0_10-1.06dens_10_15+0.672dens_15_20+0.0245var_3	31.26	5.371	4.456	0.36	0.9528
$Volume_1$	Vol=3.45+0.0797p50	39.92	55.75	42.30	0.32	11.85
$Volume_2$	Vol=2.53+0.0892p50+0.0115cancovA	36.35	50.77	37.72	0.45	8.893
$Volume_3$	Vol=2.82+0.0719p50+0.0132cancovA-0.957dens_10_15	35.97	50.24	37.13	0.45	8.698
$Volume_4$	Vol=3.20+0.0721p50+0.0148cancovA-1.11dens_10_15-0.594hom_1	35.32	49.33	37.30	0.45	8.299
$Volume_5$	Vol=3.09+0.0769p50+0.0145cancovA-1.03dens_10_15-0.590hom_1+0.213Elev_CV	35.30	49.23	37.32	0.45	8.287
AGB_1	AGB=3.63+0.00989cancovA	40.01	32.20	24.88	0.15	6.368
AGB_2	AGB=2.52+0.0122cancovA+0.0568p50	32.65	26.27	19.88	0.38	4.434
AGB_3	AGB=2.70+0.0132cancovA+0.0464p50-dens_10_15	32.46	26.12	19.65	0.38	4.387
AGB_4	AGB=2.57+0.0141cancovA+0.0484p50-0.626dens_10_15+0.0534var_2	31.95	25.72	19.59	0.39	4.160
AGB_5	AGB=2.53+0.0609p50+0.0136cancovA+0.924dens_15_20+0.0406var_2-0.432hom_1	31.26	25.20	19.35	0.41	4.064

Appendix A2

Table A2. *The best 1-5 variable models produced for volume and AGB developed from the relationship between airborne lidar reference maps and GEDI footprint level variables.*

Forest Metric	Model	R^2	RMSE	%RMSE	MAE	Bias
$Volume_1$	Vol = 62.2 + 3.88rh90	0.29	38.46	29.24	30.69	$6.429x10^{-16}$
$Volume_2$	Vol = 72.1 – 107pavd1 + 4.04rh90	0.33	37.36	28.40	29.61	$-8.488x10^{-16}$
$Volume_3$	Vol = 145 - 38.1fhd – 111pavd1 + 6.05rh90	0.35	36.80	27.94	29.15	$1.287x10^{-15}$
$Volume_4$	Vol = 145 + 0.139agbd – 32.8fhd – 126pavd1 + 4.22rh90	0.36	36.60	27.76	28.81	$1.506x10^{-15}$
$Volume_5$	Vol = 132 + 0.221agbd – 44.8cover – 43.2pavd1 – 22.5fhd + 3.54rh90	0.36	36.45	27.65	28.68	$-7.134x10^{-16}$
AGB_1	AGB = 56.9 + 1.52rh90	0.26	16.01	19.96	12.31	$-2.861x10^{-16}$
AGB_2	AGB = 59.8 – 33.6pavd1 + 1.59rh80	0.28	15.75	19.64	12.07	$2.158x10^{-16}$
AGB_3	AGB = 62.4 + 0.0686agbd – 38.5pavd1 + 0.816rh80	0.29	15.64	19.47	11.88	$1.600x10^{-16}$
AGB_4	AGB = 76.0 – 6.41fhd + 0.0690agbd – 40.3pavd1 + 1.13rh80	0.30	15.59	19.41	11.83	$-1.102x10^{-16}$
AGB_5	AGB = 74.7 + 0.0804agbd -12.9cover-15.6pavd1 – 5.10fhd+1.18rh80	0.30	15.60	19.37	11.81	$-7.276x10^{-16}$