MODELING FOREST ATTRIBUTES USING HIGH-RESOLUTION REMOTE SENSING FOR DATA-DRIVEN DECISION-MAKING IN FOREST MANAGEMENT

by

ANGEL ADHIKARI

(Under the Direction of Alicia Peduzzi)

ABSTRACT

Industrial plantations are rapidly expanding to counter deforestation and meet urbanization-driven timber demands. This growth requires efficient data-driven management to enhance yields in the face of rising wood consumption, population, illegal logging, and land-use change. Plantations, like the southeastern U.S. loblolly pines, often underperform due to generic management approaches. This underscores the demand for robust data-driven management systems, which rely on the accurate and repetitive measurement of forest attributes. Collecting such data through traditional on-site data collection is not feasible. It necessitates high-resolution remote sensing technologies like LiDAR for effective forest management.

To address these challenges, this dissertation first compares the performance of four different modeling methods for predicting various plot level forest attributes necessary for decision-making using aerial LiDAR data: (1) Least Squares Regression (LSR), (2) Adaptive Least Absolute Shrinkage and Selection Operator (ALASSO), (3) Random Forest (RF), and (4) Generalized Additive Modeling Selection (GAMSEL). The results revealed that no single method exhibits a significant advantage on accuracy over the

others. Nevertheless, the ALASSO method exhibited shighltly higher accuracy. Moreover, it is considered less biased and more accessible compared to other three. The second part of this dissertation focuses on the development of methods for assessing understory vegetation in plantation forests, particularly in the southeastern coastal plain pine plantations, where the high competition from evergreen understory species constrains pine growth. Our study identifies Terrestrial Laser Scanning (TLS) as a promising tool for understory vegetation assessment, as the model based on variables derived from this sensor demonstrates high accuracy in detecting and quantifying understory biomass. However, implementing this sensor for larger areas is not practical. Therefore, the third part of this dissertation addresses the upscaling of the TLS-derived model to Unmanned Aerial Vehicles and Aerial Laser Scanning-derived variables, enabling its application to larger areas. The result shows that both the UAV and ALS-based methods can assess the understory biomass. However, the ALS-based upscaled models were more accurate and robust, demonstrating efficacy and potential for enhanced forest management and decision-making.

INDEX WORDS: Lidar, Forest Attributes, Modeling Methods, Understory Vegetation, Terrestrial Laser Scanning, Unmanned Aerial Vehicles, Aerial Laser Scanning

MODELING FOREST ATTRIBUTES USING HIGH-RESOLUTION REMOTE SENSING FOR DATA-DRIVEN DECISION-MAKING IN FOREST MANAGEMENT

by

ANGEL ADHIKARI

BSc, Tribhuvan University, Nepal, 2013

MSc, The University of Tokyo, Japan, 2020

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

MODELING FOREST ATTRIBUTES USING HIGH RESOLUTION REMOTE SENSING FOR DATA-DRIVEN DECISION MAKING IN FOREST MANAGEMENT

by

ANGEL ADHIKARI

Major Professor: Alicia Peduzzi

Committee: Cristian R. Montes
Deepak R. Mishra
Nathaniel Osborne
Bronson Bullock

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

ACKNOWLEDGEMENTS

This research project would not have been possible without the invaluable support of several individuals. First and foremost, I want to express my most sincere appreciation to my major professor, Dr. Alicia Peduzzi, who provided expert guidance throughout my graduate studies, consistently made time to address any research challenges that arose, and helped me maintain a clear focus on the tasks at hand. Her encouragement during tough times was invaluable, and I truly appreciated the opportunity to learn and grow alongside her. I would also like to extend my gratitude to Dr. Cristian R. Montes, Dr. Deepak Mishra, Dr. Nathaniel Osborne, and Dr. Bronson Bullock for their contributions to the formulation and completion of this study as members of my committee.

I am truly grateful to the Warnell School of Forestry and Natural Resources, the University of Georgia, for funding my assistantship and the Plantation Management Research Cooperative (PMRC) for supporting my field data collection.

I am grateful to the Precision Forestry Lab members Katrina Henn, Suveksha Jha, Sudhir Payare, Jackson Glomb, Daniel Hewitt, and Wood and Fiber Quality lab members Nawaraj Pokhrel and Sameen Raut for their help in my field data collection and processing, and all fellow students, and my family members who were always there when I needed someone to talk.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS viii

LIST OF TABLES xiii

LIST OF FIGURES xiv

CHAPTER

1 INTRODUCTION & LITERATURE REVIEW 1

1.1 Introduction 1

1.2 Goals, Objectives, and Hypotheses 7

1.3 References 9

2 A COMPARISON OF MODELING METHODS FOR PREDICTING FOREST ATTRIBUTES USING LIDAR METRICS 14

2.1 Introduction 16

2.2 Materials and Methods 22

2.3 Results 37

2.4 Discussion 47

2.5 Conclusions 52

2.6 References 53

3 ASSESSMENT OF UNDERSTORY VEGETATION IN A PLANTATION FOREST OF THE SOUTHEASTERN UNITED STATES USING TERRESTRIAL LASER SCANNING....60

3.1 Introduction....63

3.2 Methods....70

3.3 Results....89

3.4 Discussion....97

3.5 Conclusion....102

3.6 Abbreviations....102

3.6 References....104

4 ASSESSING UNDERSTORY VEGETATION BIOMASS IN PINE PLANTATION FOREST WITH MULTI-SCALE LIDAR....109

4.1 Introduction....111

4.2 Methods....116

4.3 Results....136

4.4 Discussion....153

4.5 Conclusions....159

4.6 References....160

5 CONCLUSION....165

APPENDICES

A Chapter 2 Appendix 168

LIST OF TABLES

Page

Table 2.1 Summary statistics of forest attributes from ground measurement 26

Table 2.2 Description of plot-level metrics derived from lidar data..... 27

Table 2.3 Summary statistics of thirty-nine lidar metrics..... 28

Table 2.4 Summary statistics of forest attributes from ground measurement 37

Table 2.5 The collinearity index value for all variables selected in the LS model for volume, basal area, and dominant height..... 39

Table 2.6 List of variables and their coefficients selected for Volume, BA, and Dominant height model by adaptive lasso regression method..... 41

Table 2.7. K-fold cross-validation result summary for all response variables with four different modeling methods 45

Table 3.1 Distribution of sample plots based on height and understory abundance..... 73

Table 3.2: Description of sub-plot level metrics derived from TLS data 82

Table 3.3: Summary statistics of understory biomass and twenty-seven subplot level TLS metrics..... 83

Table 3.4: Least Square regression models summary for the understory biomass prediction with TLS-based variables. Calibration denotes the performance metrics when fitting with the full data set, and CV depicts the mean value after five-fold cross-validation..... 92

Table 4.1 Sample plot distribution within understory abundance and dominant height criteria. Understory abundance level one refers to the presence of very low to no understory, and six represents very high abundance in the sample plot. 119

Table 4.2: The essential parameters of the three LiDAR datasets. 123

Table 4.3: Point density summary at different height levels 124

Table 4.4: Description of UAV and ALS-based metrics derived for upscaling TLS-based model..... 125

Table 4.5: Summary statistics of TLS predicted biomass map and major ALS and UAV-LS variables at 8 m resolution. The difference in TLS predicted biomass summary statistics for ALS and TLS is due to the different sample sizes (no. of pixels). 137

Table 4.6. The estimated understory biomass and its uncertainty after Monte Carlo simulation (Unit: $Mg.ha^{-1}$) 152

LIST OF FIGURES

Page

Figure 2.1 Study area and plots Location in the Bio-Bio region, central Chile..... 23

Figure 2.2 The illustration of the 9-fold cross-validation. The dataset is divided into nine random subsets; from there, eight are used for training and one for validation. This process is repeated nine times with different subsets for validation..... 36

Figure 2.3 Pearson correlation coefficients for each response variable and the four highly correlated predictor variables. (a) Pearson correlation coefficients plot for BA; (b) Pearson correlation coefficients plot for volume; (c) Pearson correlation coefficient 38

Figure 2.4. Variable importance plot for dominant height, BA, and Volume in Random Forest models. 40

Figure 2.5. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for basal area estimation with the functions for the last value of the lambda (Y-axis)..... 43

Figure 2.6. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for volume estimation with the functions for the last value of the lambda (Y-axis)..... 43

Figure 2.7. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for dominant height estimation with the functions for the last value of the lambda (Y-axis). 44

Figure 2.8. Each panel shows the residual vs. fitted plot for volume with ALASSO models. R^2 and RMSE are the model accuracy before cross-validation, and CV_R^2 and CV RMSE show the model accuracy result of 9-fold cross-validation. 46

Figure 3.1 Study area map showing the spatial distribution of sample plots inside Nassau County, Florida, US 71

Figure 3.2: Dominating evergreen understory species in the study area, a) Gallberry (Ilex glabra), b) saw palmetto (Serenoa repens), c) ti-ti (Cyrilla racemiflora), and d) fetterbush (Lyonia lucida)..... 72

Figure 3.3: Graphical representation of a sample plot layout and subplot location for destructive sampling 75

Figure 3.4: The process for collecting understory samples and measuring their dry biomass. First, a $1m^2$ PVC frame was used to establish an understory subplot (a). Then, after conducting destructive sampling, the samples were placed in bags and labeled (b). Next, the collected samples were dried in an oven (c) for 48 hours at 105°C before dry biomass weight measurement (d)..... 76

Figure 3.5: Figure Illustrating the LiDAR point cloud (filtered under 3 meters) before (left) and after (right) the removal of tree stems. 81

Figure 3.6: Mean height and understory cover-based volume extraction method. (a) Destructively sampled understory vegetation at ground level. (b) Blue dots represent the clipped point clouds for the subplot. (c) The mean height of the point clouds in meters and understory cover estimates based on 0.2 meters grid counting. 84

Figure 3.7: Voxel-based volume extraction method. (a) Destructively sampled understory vegetation at ground level (b) Blue dots represent the clipped point clouds for the subplot (c) Fitted voxel of resolution 0.2 meters in the point clouds. 85
Figure 3.8 Three-dimensional α-Hull fitting-based volume extraction method. (a) Destructively sampled understory vegetation at ground level. (b) Blue dots represent the clipped point clouds for the subplot. (c) The structure in red is the fitted α-Hull with an alpha value of 0.2 on the point clouds. 86
Figure 3.9: Boxplot of the understory biomass and six primary LiDAR metrics, displaying the Median, quartiles, and outliers, providing a concise visualization of central tendency, dispersion, and Skewness of the selected variables..... 90
Figure 3.10: Pearson correlation coefficients plot between understory biomass and with all derived volume-based metrics and highly correlating statistical TLS metrics. Darker blue and bigger squares represent higher correlation, and lighter blue and smaller squares represent lower correlation. 91
Figure 3.11: The plot on the top displays the relationship between predicted and observed understory biomass, and the plot on the bottom is the residual vs. predicted plot. 95
Figure 3.12: Understory biomass predicted maps for 40m x 40m sample plots, showing four different abundance of understory biomass amounts at 1m resolution. Plot no is the assigned plot number out of 60 plots. 96
Figure 4.1 Overview of the research area and the distribution of the sample plots in Nassau County, Florida, US..... 117
Figure 4.2 Figure shows the UAV-flight plan to collect lidar data, which was created in UGCS software and projected over google earth pro base map. In the figure, yellow lines,

which are 70 meters above ground, are the UAV flight lines, and the red square in the middle represents the sample plot (40m × 40m) boundary. 122
Figure 4.3: Forest vertical profile showing LiDAR point cloud density and distribution comparison plot for TLS, UAV-LS, and ALS for one of the sample plot..... 127
Figure 4.4: Flow chart of the data processing and modeling..... 135
Figure 4.5: Pearson correlation coefficients between TLS-predicted understory biomass and the ten highest correlated UAV-LS derived metrics (a) and ALS-derived metrics (b). 139
Figure 4.6 Observed vs Predicted plot comparing UAV Lidar-based predicted understory biomass (kg) against TLS-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.61 and RMSE of 11.37 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.6 and RMSE of 11.46 kg). 141
Figure 4.7 Residuals plot for the UAV Lidar-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions. 142
Figure 4.8 Observed vs Predicted plot comparing ALS-based predicted understory biomass (kg) against TLS-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.69 and RMSE of 10.75 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.67 and RMSE of 10.79 kg). 144

Figure 4.9 Residuals plot for the ALS-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions. 145
Figure 4.10 Observed vs Predicted plot showing ALS-based predicted understory biomass (kg) against UAV-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.71 and RMSE of 9.607 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.70 and RMSE of 9.641 kg). 147
Figure 4.11 Residuals plot for the UAV-LS to ALS-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions. 148
Figure 4.12: Variables selected in the UAV-LS-based model and their correlation with the TLS-based biomass. 149
Figure 4.13 Variables selected in the ALS-based model and their correlation with the TLS-based biomass. 150
Figure 4.14 Variables selected in the ALS-based model and their correlation with the UAV-LS-based biomass. 150

CHAPTER 1

1.1 Introduction

The large-scale industrial plantation establishment started in the 1960s, primarily in response to escalating concerns about the rapid deforestation of natural forests (Olorunnisola 2023). The rising demand for timber products due to soaring urbanization on a global scale further prompted the establishment of plantation forests (FAO 2016, 2020). Industrial plantations are typically characterized by efficient management practices, the production of high-quality timber, rapid growth rates, shorter rotation lengths, and the incorporation of tree improvement technology (Campinhos 1999). Consequently, they yield higher timber products in a short period per unit area. These plantation forests serve as the major source of raw materials for the forest industry, satisfying about one-third of the world industrial roundwood demand, thus mitigating the pressure on primary forests and modified natural forests (Barua, et al. 2014).

Over the past few decades, there has been a significant expansion of plantation forests, covering about 3% of the global forest area, however contributing up to 35% of the global roundwood production (FAO, 2020). Plantation forests cover 291 million hectares, experiencing an average annual growth rate of 1.84% since 1990. The most significant annual increments have been observed in North and Central America (2.51%) and South America (2.38%)(FAO 2020). Moreover, these forests are projected to significantly expand their share of global forest coverage, with industrial plantations expected to double in a few decades. Furthermore, by 2050, 71% of timber harvested in South

America is predicted to originate from industrial plantations (Daigneault, et al. 2008). With this rate, the supply of industrial roundwood sourced from forest plantations is expected to increase from its current level of approximately 520 million cubic meters to nearly 1.5 billion cubic meters by 2050 (McEwan, et al. 2020). However, the demand for roundwoods is predicted to reach six billion cubic meters by 2050 (Barua, et al. 2014).

The major driving factors for that increased demand include increased wood consumption by escalating economies, particularly in developing nations, increasing population growth, illegal logging, and converting forested land to agricultural purposes in tropical regions (McEwan, et al. 2020). These trends are expected to shrink naturally forested areas, consequently increasing dependence on plantation forests. While expanding plantation forests is inevitable and necessary, it is also expected to bring about an augmented demand for enhanced management and monitoring systems, as well as data-driven automated decision-making systems (McEwan, et al. 2020).

To meet this growing demand, it is necessary to refine these technologies, with a promising possibility being the use of high-resolution remote sensing techniques. For example, similar to expanding plantations of eucalyptus, particularly in South America, loblolly pine plantations represent some of the most intensively managed forests in North America, specifically in the south (Oswalt, et al. 2019; Phillips 2013). Nevertheless, even with the most intensive management practices, such plantations often produce timber at rates lower than their potential (Borders, et al. 2001; Zhao, et al. 2016).

Efforts to increase productivity have led to the development of integrated stand management approaches that aim to maximize growth suited to the specific site. In this context, accurate and regular assessments of various forest attributes related to timber

production and growth play a vital role. Additionally, inadequate site nutrition, compounded by competition from other vegetation, has been identified as a significant limiting factor for plantation forest growth, specifically, in the case of loblolly pine in the southern region (Allen, et al. 1990; Fox, et al. 2007). Addressing these challenges requires a comprehensive understanding of site-specific stand attributes, including those related to inter and intra-species competition and spatial distribution (Gao, et al. 2016; Young, et al. 2023). This understanding is crucial for enhancing the yield and improving ecological service functions in plantation forests.

To develop site-specific management approaches, it is essential to regularly estimate and monitor various attributes, including diameter at breast height (DBH), tree height, dominant height, the volume of merchantable standing trees, crown volume, plantation survival, and competing vegetation. These attributes have traditionally been determined through on-site measurements at sample plots (Santoro, et al. 2011). Alternatively, a satellite-based approach, assisted by estimating forest attributes, has been widely adopted to estimate such variables across a continuous spatial distribution (Bijalwan, et al. 2010; Gao, et al. 2013). However, passive optical data can only capture the canopy in two dimensions, making them insensitive to sub-canopy features, such as basal area, tree height, attributes from codominant and understory layer, and lacks spectral and spatial details to classify and quantify individual plants at the species level (Morel, et al. 2011; Wing, et al. 2012). LiDAR data have been explored to address this limitation, as they are sensitive to the three-dimensional properties of forests.

Laser Scanning (LS) has emerged as a promising remote sensing technology, offering a solution to the limitations associated with conventional remote sensing methods. LS is an

active remote sensing system equipped with an electromagnetic energy source that has revolutionized remote sensing technology over the last three decades (Ørka, et al. 2022; Salas, et al. 2010). One of the major highlights of this sensor is the ability to directly measure the three-dimensional structure of scanned areas and extract bio-spatial data, such as overstory and understory vegetation attributes, terrain surface structure and surface reflectance (Reutebuch, et al. 2005). Among the LS systems, LiDAR (Light Detecting and Ranging) stands out for its remarkable performance, offering a high-density point cloud, intensity measurements for signal returns, multiple echoes per laser pulse, and horizontal and vertical positioning accuracy down to the centimeter level (Popescu, et al. 2004). Due to these attributes, LiDAR has proven to be a valuable tool for characterizing and monitoring forests across the landscape (Asner, et al. 2012; Coops, et al. 2021; Ioki, et al. 2014; Popescu, et al. 2004; Salas, et al. 2010). Using information from the height and intensities of LiDAR point clouds, various metrics such as, stastistical distribution of height and intensities, density of 3D point clouds in different height percentiles, density ratios, and canopy cover, at the plot, grid, or individual tree level can be derived (Coops, et al. 2021). These metrics, extracted from point cloud data, are then compared with plot-level data to estimate stand-level attributes.

Nonetheless, several challenges exist in estimating attributes using LiDAR metrics, primarily due to high dimensionality and redundancy. High dimensionality refers to the vast number of variables extracted from LiDAR data, which can make analysis complex and computationally demanding. Furthermore, the issue of multicollinearity arises, where these numerous metrics are often interdependent, mainly due to the reliance of all generated metrics either on the height or intensity of the LiDAR returns. This issue is not

specific to any particular reason or species but is a general problem in the application of LiDAR for extracting forest attributes (Kim, et al. 2019). These correlated variables can lead to unreliable statistical inferences and pose significant challenges in distinguishing the unique contribution of each metric. Therefore, given the high-dimensional and often correlated nature of these metrics derived from 3D LiDAR data, robust statistical modeling approaches are required to obtain meaningful information and accurate inventory parameter estimation(García-Gutiérrez, et al. 2011). Such approaches are essential to address the intricacies of the data and ensure the extraction of reliable, useful information from this advanced sensing technology (Marrs, et al. 2019).

Previous studies have proposed various parametric, semiparametric, and nonparametric modeling methods, including Ordinary Least Square (OLS) regression, Nonlinear Least Square (NLS) regression, Random-Forest (RF) based imputation, geographically weighted regression (GWR), Artificial Neural Networks (ANN), Support Vector Regression (SVR), and more to establish reliable models linking ground attributes with LiDAR metrics (Holopainen, et al. 2006; Ioki, et al. 2014; Monnet, et al. 2011; Niska, et al. 2009; Packalén, et al. 2011; Shin, et al. 2016; Strunk, et al. 2008). However, it is important to note that each method has its limitations. There is a potential for exploring other parametric and semiparametric approaches that can effectively handle data with higher dimensions and redundancy for forest attributes estimation using LiDAR data.

In addition to its applicability in improving forest measurement and monitoring, LiDAR technology plays a crucial role in detecting and quantifying understory vegetation, a significant concern in intensively managed plantation forests. LiDAR distinct advantage lies in its ability to penetrate the canopy and create comprehensive 3D profiles of entire

forests (Venier, et al. 2019). Over the past two decades, numerous studies have utilized airborne LiDAR to assess understory vegetation (Campbell, et al. 2018; Jubanski, et al. 2012; Li, et al. 2021). However, it is important to recognize that airborne LiDAR has limitations in assessing understory vegetation, especially in forests with dense and tall overstory vegetation, due to its lower pulse density and occlusion caused by the overstory (Campbell, et al. 2018).

To address this limitation, Terrestrial Laser Scanning (TLS) has emerged as a promising alternative. This ground-based sensor offers a much higher point-cloud density, with the potential to collect up to 1000 points per square meter (Li, et al. 2021). Studies have acknowledged its efficiency in characterizing understory vegetation, surpassing other field-based or remote sensing methods. However, a significant constraint of TLS is its applicability to small areas or plot-level assessments due to ground-level data collection's time and resource-intensive nature.

To overcome this constraint, a feasible approach is to upscale the TLS-based method of understory assessment to aerial LS-based methods, such as Unmanned Aerial Vehicle (UAV-LS) for stand-level assessments and Airborne Laser Scanning (ALS) for landscape-level assessments. UAV-LS involves mounting or attaching LiDAR sensors to drones or UAVs, enabling the collection of three-dimensional (3D) geospatial data at relatively low altitudes. In contrast, airborne Laser Scanning, often referred to as Airborne Laser Scanning, involves the installation of LiDAR sensors onboard aircraft or helicopters to acquire 3D data from an elevated point above the surface of the earth, in the case of forests, above the canopy. ALS can cover larger areas more quickly than TLS

but at the cost of a reduced point density, making it suitable mostly for regional-level mapping and topographic surveys.

Conversely, UAV-LS balances the high point density of TLS and the aerial coverage, making it a flexible and suitable choice for various applications. It is particularly valuable for monitoring small to medium-sized areas with high precision, while airborne LiDAR is commonly used for large-scale assessment and planning. Thus, upscaling the TLS-based approach to aerial platform-based LS methods will aid plantation managers and policymakers in monitoring tree growth and competing vegetation and determining control measures and silvicultural prescriptions to enhance plantation productivity.

In summary, this study explores the application of LiDAR technology from various platforms to address three critical aspects of plantation forests. Firstly, it aims to improve modeling methods for estimating and monitoring overstory attributes. Secondly, it focuses on detecting and quantifying understory vegetation in plantation forests using TLS. Lastly, it deals with upscaling the TLS-based approach to more practical LS platforms, such as UAV and Airborne LiDAR, to aid in monitoring and data-driven decision-making for plantation forest management.

1.2 Goals, Objectives, and Hypotheses

1.2.1 General Goal

To improve methods for forest attribute estimation and understory biomass prediction using ground and aerial laser scanning techniques.

1.2.2 Study objectives

1. To identify the best modeling methods for predicting stand-level forest attributes using lidar metrics.

(a) Hypothesis: advanced modeling methods incorporating a variety of lidar metrics and addressing multicollinearity will generate robust models and yield more accurate forest attribute predictions than traditional modeling approaches.

(b) Methodological objective: To train and validate advanced statistical models utilizing a diverse set of lidar-derived metrics and compare their performance for predicting stand-level forest attributes.

2. To assess understory vegetation biomass utilizing Terrestrial Laser Scanning data.

(a) Hypothesis: Terrestrial lidar data, with its 3-dimensional nature and capability to capture object shapes and structure and its high density of returns due to its proximity to objects or ground-based acquisition, will exhibit a strong correlation with understory vegetation biomass than alternative field-based methods, providing a more efficient and accurate way to estimate understory biomass.

(b) Methodological objective: Develop and validate a predictive model based on Terrestrial Laser Scanning (TLS) derived metrics to accurately estimate understory vegetation biomass, ensuring that the generated predicting variables or metrics are meaningful, and the model is mechanistically sound.

3. To generate a robust understory biomass predictive model utilizing airborne lidar data for large-scale forest management decision-making.

(a) Hypothesis: Having an improved and spatially explicit understory biomass estimate derived from TLS data can be effectively upscaled to Unmanned Aerial Vehicle (UAV) and Airborne Laser Scanning (ALS)-derived variables while improving the relationships between the dependent and independent variables without a significant loss of accuracy and across large spatial extents.

(b) Methodological objective: Upscale the TLS-derived model by calibrating UAV-lidar and ALS data variables against the TLS-derived biomass estimates, ensuring its scalability and accuracy over larger spatial areas.

1.3 References:

1. Allen, H., Dougherty, P.M. and Campbell, R. 1990 Manipulation of water and nutrients—practice and opportunity in southern US pine forests. *Forest Ecology and Management*, **30** (1-4), 437-453.

2. Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M. *et al.* 2012 A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia*, **168** (4), 1147-1160.

3. Barua, S., Lehtonen, P. and Pahkasalo, T. 2014 Plantation vision: potentials, challenges and policy options for global industrial forest plantation development. *International Forestry Review*, **16** (2), 117-127.

4. Bijalwan, A., Swamy, S., Sharma, C.M., Sharma, N.K. and Tiwari, A. 2010 Land-use, biomass and carbon estimation in dry tropical forest of Chhattisgarh region in India using satellite remote sensing and GIS. *Journal of Forestry Research*, **21**, 161-170.

5. Borders, B.E. and Bailey, R.L. 2001 Loblolly pine—pushing the limits of growth. *Southern Journal of Applied Forestry*, **25** (2), 69-74.

6. Campbell, M.J., Dennison, P.E., Hudak, A.T., Parham, L.M. and Butler, B.W. 2018 Quantifying understory vegetation density using small-footprint airborne lidar. *Remote sensing of environment*, **215**, 330-342.

7. Campinhos, E. 1999 Sustainable plantations of high-yield shape Eucalyptus trees for production of fiber: the Aracruz case. *New Forests*, **17** (1), 129-143.

8. Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K. *et al.* 2021 Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, **260**, 112477.

9. Daigneault, A.J., Sohngen, B. and Sedjo, R. 2008 Exchange rates and the competitiveness of the United States timber sector in a global economy. *Forest Policy and Economics*, **10** (3), 108-116.

10. FAO. 2020 Global forest resources assessment. https://www.fao.org/3/CA8753EN/CA8753EN.pdf (10/25/2023, 2023).

11. Fox, T.R., Jokela, E.J. and Allen, H.L. 2007 The Development of Pine Plantation Silviculture in the Southern United States. *Journal of Forestry*, **105** (7), 337-347.

12. Gao, T., Yang, X., Jin, Y., Ma, H., Li, J., Yu, H. *et al.* 2013 Spatio-temporal variation in vegetation biomass and its relationships with climate factors in the Xilingol grasslands, Northern China. *PLoS One*, **8** (12), e83824.

13. Gao, T., Zhu, J., Deng, S., Zheng, X., Zhang, J., Shang, G. *et al.* 2016 Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. *International Journal of Applied Earth Observation and Geoinformation*, **52**, 155-165.

14. García-Gutiérrez, J., González-Ferreiro, E., Mateos-García, D., Riquelme, J. and Miranda, D. 2011 *A Comparative Study between Two Regression Methods on LiDAR Data: A Case Study*, 311-318 p.

15. Holopainen, M. and Kalliovirta, J. 2006 Modern data acquisition for forest inventories. In *Forest Inventory*, Springer, pp. 343-362.

16. Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M.-H., Wong, W.V.C., Ling, Z.-Y. *et al.* 2014 Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR. *Forest ecology and management*, **328**, 335-341.

17. Jubanski, J., Ballhorn, U., Kronseder, K., Franke, J. and Siegert, F. 2012 Detection of large above ground biomass variability in lowland forest ecosystems by airborne LiDAR. *Biogeosciences Discussions*, **9** (8).

18. Kim, Y.J., Nam, B.H. and Youn, H. 2019 Sinkhole detection and characterization using LiDAR-derived DEM with logistic regression. *Remote Sensing*, **11** (13), 1592.

19. Li, S., Wang, T., Hou, Z., Gong, Y., Feng, L. and Ge, J. 2021 Harnessing terrestrial laser scanning to predict understory biomass in temperate mixed forests. *Ecological Indicators*, **121**, 107011.

20. Marrs, J. and Ni-Meister, W. 2019 Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sensing*, **11** (7), 819.

21. McEwan, A., Marchi, E., Spinelli, R. and Brink, M. 2020 Past, present and future of industrial plantation forestry and implication on future timber harvesting technology. *Journal of Forestry Research*, **31**, 339-351.

22. Monnet, J.-M., Chanussot, J. and Berger, F. 2011 Support vector regression for the estimation of forest stand parameters using airborne laser scanning. *IEEE Geoscience and remote sensing letters*, **8** (3), 580-584.

23. Morel, A.C., Saatchi, S.S., Malhi, Y., Berry, N.J., Banin, L., Burslem, D. *et al.* 2011 Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management*, **262** (9), 1786-1798.

24. Niska, H., Skon, J.-P., Packalen, P., Tokola, T., Maltamo, M. and Kolehmainen, M. 2009 Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs. *IEEE Transactions on Geoscience and Remote Sensing*, **48** (3), 1076-1085.

25. Olorunnisola, A.O. 2023 The Past, Present and Future Outlook of the Wood Industry in Nigeria. In *Wood Industry-Past, Present and Future Outlook*, IntechOpen.

26. Ørka, H.O., Jutras-Perreault, M.-C., Candelas-Bielza, J. and Gobakken, T. 2022 Delineation of Geomorphological Woodland Key Habitats Using Airborne Laser Scanning. *Remote Sensing*, **14** (5), 1184.

27. Oswalt, S.N., Smith, W.B., Miles, P.D. and Pugh, S.A. 2019 Forest resources of the United States, 2017. *General Technical Report-US Department of Agriculture, Forest Service. Forest Service*.

28. Packalén, P., Mehtätalo, L. and Maltamo, M. 2011 ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. *Annals of Forest Science*, **68** (6), 1085-1092.

29. Phillips, R.B. 2013 Global dynamics of the pulp and paper industry–2013. *NZ Journal of Forestry*, **58** (3), 35.

30. Popescu, S.C. and Wynne, R.H. 2004 Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, **70** (5), 589-604.

31. Reutebuch, S.E., Andersen, H.-E. and McGaughey, R.J. 2005 Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. *Journal of Forestry*, **103** (6), 286-292.

32. Salas, C., Ene, L., Gregoire, T.G., Næsset, E. and Gobakken, T. 2010 Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, **114** (6), 1277-1285.

33. Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I. *et al.* 2011 Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sensing of Environment*, **115** (2), 490-507.

34. Shin, J., Temesgen, H., Strunk, J.L. and Hilker, T. 2016 Comparing Modeling Methods for Predicting Forest Attributes Using LiDAR Metrics and Ground Measurements. *Canadian Journal of Remote Sensing*, **42** (6), 739-765.

35. Strunk, J.L., Reutebuch, S.E. and Foster, J.R. LiDAR inventory and monitoring of a complex forest.

36. Venier, L.A., Swystun, T., Mazerolle, M.J., Kreutzweiser, D.P., Wainio-Keizer, K.L., McIlwrick, K.A. *et al.* 2019 Modelling vegetation understory cover using LiDAR metrics. *Plos One*, **14** (11), e0220096.

37. Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., Gitelman, A. and Olsen, M.J. 2012 Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. *Remote Sensing of Environment*, **124**, 730-741.

38. Young, J.B., Bullock, B.P. and Montes, C.R. 2023 Assessing Mid-rotation Loblolly Pine and Competing Vegetation Responses to Post-thin Fertilization and Herbicide Application in the Southeastern United States. *Journal of Forestry*, fvad035.

39. Zhao, D. and Kane, M. 2016 Loblolly pine tree equations for volume-to-weight conversion ratios, aboveground biomass and biomass allocation. *Warnell School of Forestry and Natural Resources, Plantation Management Research Cooperative (PMRC) Tech. Rep*, **1**.

CHAPTER 2

A COMPARISON OF MODELING METHODS FOR PREDICTING FOREST ATTRIBUTES USING LIDAR METRICS[1]

[1] This chapter is based in full on the previously published article [Adhikari, A., Montes, C. R., & Peduzzi, A. (2023). A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics. Remote Sensing, 15(5), 1284.] Reprinted here with the permission of publisher.

Abstract

Recent advancements in laser scanning technology have demonstrated great potential for the precise characterization of forests. However, a major challenge in utilizing metrics derived from lidar data for forest attribute prediction is the high degree of correlation between these metrics, leading to multicollinearity issues when developing multivariate linear regression models. To address this challenge, this study compared the performance of four different modeling methods for predicting various forest attributes using aerial lidar data: (1) Least Squares Regression (LSR), (2) Adaptive Least Absolute Shrinkage and Selection Operator (ALASSO), (3) Random Forest (RF), and (4) Generalized Additive Modeling Selection (GAMSEL). The study used three primary plot-level forest attributes (volume, basal area, and dominant height) as response variables and thirty-nine plot-level lidar metrics as explanatory variables. A k-fold cross-validation approach was used, with consistent folds to assess the performance of each method. Our results revealed that no single method demonstrated a significant advantage over the others. Nonetheless, the highest R^2 values of 0.88, 0.83, and 0.87 for volume, basal area, and dominant height, respectively, were achieved using the ALASSO method. This method was also found to be less biased, followed by GAMSEL and LSR.

Keywords: forest attributes; least squares regression; adaptive least absolute shrinkage and se-lection operator; random forest; generalized additive modeling selection; eucalyptus plantation

2.1 Introduction

Forest investment planning requires accurate inventories describing a given stands size and product distribution. To achieve this, attributes such as breast height diameter (DBH), tree heights, merchantability, and volume of standing trees are regularly sampled to describe the value of a given stand. However, traditional methods of field-based inventories are often time-consuming and expensive due to the cost of establishing adequate samples capturing the existing variability (Lovell, et al. 2005; McRoberts, et al. 2007; White, et al. 2016). Moreover, the variables collected through field-based inventories are limited within the instrumental range and accessibility of a field crew, and their credibility highly depends on the quality and quantity of the field samples (Liang, et al. 2016). Furthermore, studies demanding more comprehensive variables, such as biomass or tree taper studies, require destructive sampling of a subsequent number of trees. The feasibility of such sampling is extremely limited by capital, organization, labor, and protected areas (Brown 1997; Picard, et al. 2012). Therefore, over the last two decades, researchers have focused on improving the use of remotely sensed information as auxiliary variables in forest inventories. The advent of inexpensive multispectral satellite data (Holopainen, et al. 2006) or the use of hyperspectral imagining, radar, and laser scanning has opened the door to new developments (Pajeres 2015; Toth, et al. 2016; Zhu, et al. 2018), not only allowing fast and repetitive data collection over a large spatial area but also allowing for a reduction in inventory cost.

Airborne laser scanning (ALS) is a type of active remote sensing system with its own source of electromagnetic energy that has revolutionized remote sensing technology over the last three decades (Ørka, et al. 2022; Salas, et al. 2010). The major highlight of this revolution is its capability to measure the three-dimensional structure of imaged areas

directly and to extract bio-spatial data, such as aboveground vegetation-related data, from geospatial information, such as terrain surface, using laser pulses (Reutebuch, et al. 2005). In ALS, a laser scanner is attached to the aerial platform, such as crewed or uncrewed aircraft, which distributes the transmitted pluses across the flight direction, measuring the 3D position of points at the surface with an accuracy of a few decimeters in vegetation canopies (Næsset, et al. 2004). LiDAR (Light Detecting and Ranging) is one of the ALS systems that presents a remarkable performance by offering a high density of points, intensity measurements for the returning signal, multiple echoes per laser pulse, and centimeter accuracy for horizontal and vertical positioning (Popescu, et al. 2004). Because of such features, a lidar is a valuable tool for forest characterization and monitoring across the landscape (Asner, et al. 2012; Coops, et al. 2021; Ioki, et al. 2014; Popescu, et al. 2004; Salas, et al. 2010).

Two approaches for forest attributes estimation using three-dimensional Lidar point clouds are commonly used for forest attribute estimation: the Individual tree-based approach(Li, et al. 2012; Popescu, et al. 2003) and the area-based approach(Næsset 2002; Naesset 1997). Individual tree-based approaches are applied when single-tree level attributes are required, and high-density lidar data is available. For single-tree level attribute extraction, several automated and semi-automated algorithms, such as the local maximum-based methods (Koch, et al. 2006; Silva, et al. 2016a), local curvature methods (Bian, et al. 2014), the watershed methods (Hirata, et al. 2009; Kwak, et al. 2007), and many other methods have been proposed and applied by several researchers in past years. However, these algorithms are difficult to implement due to the generation of omission and inclusion errors when individualizing trees (Bouvier, et al. 2015; Goerndt, et al.

2010). Besides, joining and overlapping tree crowns can be problematic when identifying individual trees (Gleason, et al. 2011). Therefore, applying an individual tree-based approach in complex stands such as uneven-aged natural forests and plantations with high stem density is difficult. Area-based approaches (ABA) is another method that is more commonly used in diverse forest biomes for forest attribute estimation and mapping (Hall, et al. 2005; Kronseder, et al. 2012; Lim, et al. 2003). In this approach, various plots or grids, level metrics such as mean height, dominant height, the density of 3D point clouds in different height percentiles, skewness, kurtosis, canopy cover, etc., are generated using the echo heights (backscattering of laser beam towards the sensor each time the laser beam is totally or partially intercepted by an object) and intensities (the strength of laser beam that returned to the sensor) of lidar point cloud data (Coops, et al. 2021). Those metrices extracted from point cloud data are then compared with the plot-level data to estimate stand level attributes.

However, some serious challenges exist in estimating stand-level attributes using lidar metrics. Among the major challenges are the high data dimensionality and redundancy in some of them (Marrs, et al. 2019). For instance, Shi et al. (2018) evaluated the correlation among 37 frequently used lidar metrics. Their result indicated that about 60 percent of lidar metrics correlate above 70 percent to each other (Shi, et al. 2018). Thus, such high dimensional and highly correlating metrics obtained from 3D lidar data require a robust statistical modeling approach to get meaningful information and accurate inventory parameter estimation. To produce reliable models relating ground attributes with lidar metrics, several studies have proposed various parametric, semiparametric, and nonparametric modeling methods, such as Ordinary Least Square (OLS) regression

(Holopainen, et al. 2006; Ioki, et al. 2014; Strunk, et al. 2008), nonlinear least square regression (Packalén, et al. 2011), Random-Forest (RF) based imputation (Hudak, et al. 2008; Latifi, et al. 2010; Shi, et al. 2018), Geographically Weighted Regression (GWR) (Chen, et al. 2012; Shin, et al. 2016) Artificial Neural Networks (ANN), Support Vector Regression (SVR) and others.

The linear regression model (LM) is one of the most used regression methods for modeling remote sensing data with field-measured data.The key advantage of using this method is the simplicity and easy interoperability of the resulting model. However, when applying this method to develop a model predicting forest attributes using lidar-derived metrics, multicollinearity between different metrics is one of the major problems limiting its applicability. Multicollinearity, also called collinearity, refers to a phenomenon in which explanatory variables in a multiple regression model are highly linearly related. This implies that two or more variables provide the same information in more than one way (Holcomb 1999). Moreover, traditional linear model-fitting methods, such as Ordinary Least Square (OLS) regression, require a large sample size, which also increases the regression variances in the case of multicollinearity. Such substantial variances of regression coefficients make it challenging to test the hypothesis concerning the effects of the predictors (Osborne, et al. 2002).

To overcome the issue of multicollinearity and the high dimension of predictor variables, nonparametric machine learning techniques, such as Support Vector Regression (SVR), Random Forest (RF), k-nearest neighbor imputation (kNN), etc., have been introduced as alternatives to traditional regression analysis. For example, Shi et al. (2018) used the random forest to classify tree species using lidar metrics and demonstrated its usefulness

for highly correlated variables (Shi, et al. 2018). Furthermore, Pascual et al. (2019) applied kNN imputation based on a random forest approach to estimate the forest attributes (Pascual, et al. 2019). However, the usual pitfall of the random forest method is that the models are complex, not easily interpretable, and tend to overfit, given a large number of predictors used (Shah, et al. 2014). Moreover, some research found higher bias with random forests while predicting out of the range covered by the training data. (Pascual, et al. 2019; Shah, et al. 2014).

Another approach to deal with multicollinearity is using semi-parametric methods like Adaptive Least Absolute Shrinkage and Selection Operator (ALASSO) regularization in regression analysis (Duzan, et al. 2015; Tibshirani 2011) and Generalized Additive Models Selection (GAMSEL) (Chouldechova, et al. 2015). Unlike OLS, these regression methods can accommodate nonlinear relations and have more flexibility in terms of statistical assumptions. Moreover, they are known for their variable selection and regularization ability and for enhancing a models prediction accuracy and interpretability, especially for high-dimension data (Chouldechova, et al. 2015; Sun, et al. 2022). In the case of selecting variables to predict forest attributes with highly correlated lidar metrics, such ability can help select the most suitable predictors to develop a robust model. However, the applicability of these semi-parametric methods has not been explicitly explored for forest attribute estimation utilizing Lidar data.

This research aims at developing equations to estimate forest attributes using two different modeling approaches—parametric and nonparametric regression for forest attribute prediction using lidar metrics. We evaluated four modeling methods: (1) Least Squares Regression (LSR), (2) Adaptive Least Absolute Shrinkage and Selection

Operator (ALASSO), (3) Random Forest (RF), and (4) Generalized Additive Modeling Selection (GAMSEL). The last three methods aim to directly address the multicollinearity problem in ABA methodology while providing flexible model forms that might improve the relationship between lidar metrics and ground information.

2.2 Materials and Methods

2.2.1 Study site

This study uses the data from an intensively managed *Eucalyptus globulus* plantation administrated by Hancock Timber Resource Group in Arauco, Bio Bio region of central Chile (37°21'S; 73°20'W). Plantation ages range from 6 to 12 years. The area lies in a sub-humid and temperate climatic region with a maritime influence and four dry months per year. Temperature varies between 5.0 and 22.8°C, with an average of 12.6°C, and the mean annual rainfall is up to 1376 mm (Ramírez, et al. 2009). The predominant topography is hilly, with an elevation range of 80m to 740m. Soils in the plantation areas are deep marine sediments with loamy red clays on the surface and dark red clays in-depth, classified as Merilupo (CIREN 1999). The soils are well structured, favoring deep root growth development and better soil water holding capacity. Figure 2.1 shows the study area and the distribution of inventory plots.

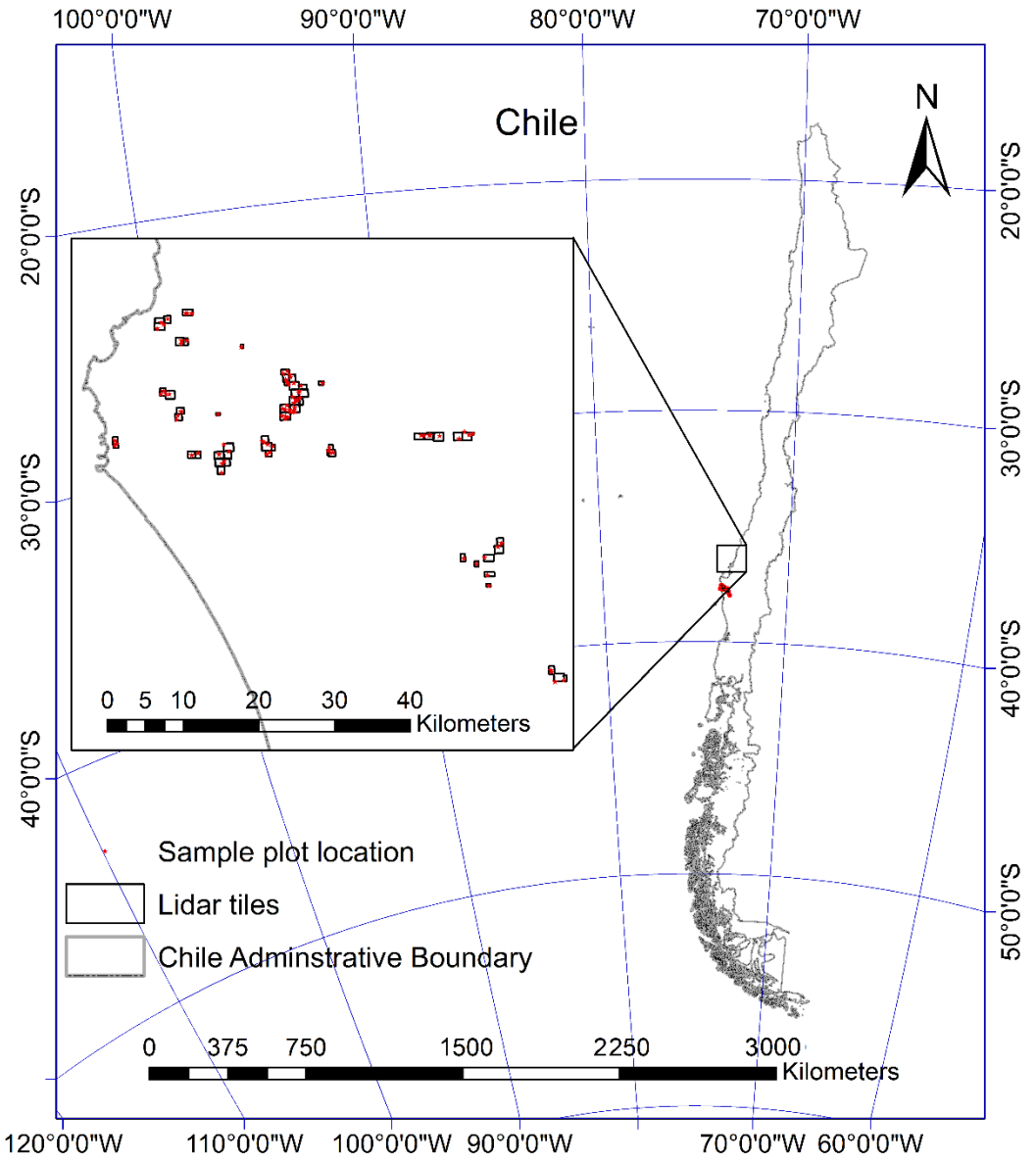

Figure 2.1 Study area and plots Location in the Bio-Bio region, central Chile

2.2.2 Field Data Collection

The field crew collected inventory data from April 12th to the 21st of 2021. Ninety circular plots of 400 m^2 were established inside 69 Lidar data acquisition square sites.

The center of each circular plot was positioned using a global navigation satellite system. In each plot, the diameter of all trees greater than 15 centimeters at the breast height was measured using a diameter caliper, and the heights of ten dominating trees were measured with the Vertex. The latter information was used to develop a regression between height and diameter to estimate all unmeasured heights from the plot using a linearized Schumacher type of equation.

This study is focused on predicting three major variables needed for forest characterization: dominant height, basal area, and volume. Dominant height generally refers to the average height of the tallest trees in a forest. It is frequently used as a metric to assess a forest ecosystems general composition and vitality, with taller trees serving as an indicator of a more advanced and developed forest structure (Wang, et al. 2007). Similarly, the basal area measures the cumulative cross-sectional area of all tree stems at a given height, typically at breast height. It is a widely utilized variable in forestry and ecological management for characterizing the structural attributes of a forest stand. Likewise, tree volume is another important metric in forest ecology and management, which refers to the total volume of a tree stem, including the trunk and branches. It provides information on determining wood quality, amount of biomass stored, harvest estimation, and various others essential for forest management and monitoring.

This study calculated plot level dominant height as the average height from the 100 largest trees per hectare (the ones with the largest diameter) (Wang, et al. 2007). Basal area (BA) and volume were calculated by summing the individual tree data at the plot level and used as the response variable for this study. Equations of calculating plot BA and volume are shown in equations 2.1 and 2.2.

Basal area was calculated using equation 2.1,

$$BA = \sum_{i=0}^{n} \left(\frac{(\pi \times (DBH^{i})^2}{4 \times 10000} \right) \quad (2.1)$$

Where DBH is the breast height diameter (1.37m from the ground) and 10000 is a conversion factor that converts cm to square meters.

The merchantable volume per plot was calculated using the volume equation shown in equation 2.2 up to a utilization index of 6 cm. The equation was derived from a sample of randomly selected eucalyptus trees between the ages of 5 and 12 years by the Hancock Timber Resource Group in Arauco.

$$Volume = \beta_0 \times TPH^{\beta 1} \times BA^{\beta 2} \times HDom^{\beta 3} \times Age^{-\beta 4} \quad (2.2)$$

Where TPH is the number of trees per hectare, BA is a basal area, $HDom$ is the dominant height of the plot, and Age is the Age of the stand. β_0, β_1, β_2, β_3 and β_4 (β_0 = 0.39239906, β_1 = 0.04033281, β_2 = 0.84491105, β_3 = 0.85388192, β_4 = - 0.18976973) are eucalyptus plantation specific parameters.

2.2.3. Lidar Data

This study used laser scanning data covering the study area collected by Hancock Timber Resource Group on April 5th, 2021. The airborne laser scanner used was a discrete lidar sensor with a field of view of 40 degrees and a flight altitude was 900 meters. The average pulse density (the average number of pulses returned from the surface) was 43.6 points/m^2 for the study area. Further specifications of the scanner are summarized in

Table 2.1. For this study, we used 63 tiles distributed between 37° 32' 12.32" S and 37° 58' 39.32" S latitude and between 73° 38' 55.22" W and 73° 06' 23.08" W longitude.

Table 2.1 Summary statistics of forest attributes from ground measurement

Parameters	**Descriptions**
Flight date	April 5th, 2021
Altitude (m) above ground level	900
Scan angle (degrees)	20 degrees
Pulse density per m^2	43.6
Laser wavelength	1550 nm

2.2.4. Individual tree detection

While the primary focus of this study is to determine plot-level attributes, we also briefly examine the estimation of individual tree-level attributes, as it has been challenging in eucalyptus plantation forests in this region due to the intricate crown structure (Corte, et al. 2022). For individual tree detection, the first step of data pre-processing, which involves noise filtering and normalization, was done by The Hancock Timber Resource Group. Then, we used the lidR package [56] in the R environment [57] for further processing. In order to identify individual trees using height-normalized lidar data, we first generated a 0.5 m resolution Canopy Height Model (CHM) using the 'pitfree' algorithm in the lidR package [58]. Subsequently, we utilized the CHM to apply a tree segmentation algorithm for individual tree detection, employing a local maxima-based algorithm (Popescu, et al. 2004). Finally, we used a circular moving window with a fixed tree window size of 3 m × 3 m. Next, we applied this function to all sixty-three lidar tiles

to calculate the number of trees in each plot for window sizes. The estimated number of trees was then added to the explanatory variable list for tree density estimation.

2.2.5. Lidar metrics generation

Many studies have demonstrated that forest attributes such as basal area, stand volume, and dominant height have a significant relationship with lidar height percentiles and density metrics (Asner, et al. 2012; Brown, et al. 2022; Silva, et al. 2016b). Therefore, our explanatory variables in this study were lidar height percentiles and density metrics. To generate the lidar height percentiles and density metrics listed in Table 2.2, we first clipped the normalized point cloud by the shape files defining the perimeter of the plots measured in the field. After that, 39 metrics were calculated at each plot level using elevation and pulse return values. A detailed description of those 39 metrics is presented in Table 2.3. Similarly, the summary statistics of all lidar metrics are described in Table 3.

Table 2.2 Description of plot-level metrics derived from lidar data

Abbreviations	**Description**
Hmax, Hmean, Hmode, Hmed	Maximum, Mean, Mode, and median values of the echo heights within a plot
Hstd, Hvar, Hcv, Hkur, Hske	Standard deviation, variance, coefficient of variation, kurtosis, and skewness of the echo heights within a plot
H01, H05, H10, H15, H20, H25, H30, H35, H40, H45, H50, H55, H60, H65, H70, H75, H80, H90, H95, H99	Echo height distribution percentiles (1^{st}, 5^{th}, 10^{th}, ……, 90^{th}, 95^{th}, 100^{th}) within a plot.
CRR	Canopy relief ratio = (height mean − height minimum)/(height maximum − height minimum)
P1minht, P1mean, P1mode	Percentage of first returns above minimum height, mean and mode
P_all_minht, P_all_mean, P_all_mode	Percentage of all returns above minimum height, mean, and mode

Abbreviations	Description
R_1st	(All returns above 0.5m/total first returns) *100
R_all_mean	(All returns above mean/total first returns) *100
R_all_mode	(All returns above mode/total first returns) *100

Table 2.3 Summary statistics of thirty-nine lidar metrics

Lidar metrics	Minimum	Maximum	Mean	SD
Hmax	12.57	32.51	21.15	3.96
Hmean	4.24	15.93	10.15	2.51
Hmode	0.51	25.55	6.87	6.08
Hmed	2.62	17.99	10.70	3.44
Hstd	3.11	9.14	5.42	1.25
Hvar	9.66	83.61	30.95	14.74
Hcv	35.69	86.81	55.09	12.53
Hkur	1.45	3.27	1.99	0.36
Hske	-1.00	1.20	-0.21	0.41
H01	0.53	1.37	0.77	0.19
H05	0.71	2.82	1.44	0.48
H10	0.94	5.17	2.34	1.04
H15	1.15	7.07	3.32	1.59
H20	1.28	9.62	4.39	2.12
H25	1.45	12.17	5.53	2.60
H30	1.70	13.90	6.66	2.96
H35	2.11	14.82	7.74	3.23
H40	2.35	15.52	8.76	3.42
H45	2.48	16.95	9.73	3.49
H50	2.62	17.99	10.70	3.44
H55	2.80	18.74	11.62	3.36
H60	3.04	19.81	12.46	3.31
H65	3.72	21.35	13.27	3.28
H70	4.84	22.76	14.02	3.25
H75	6.41	24.09	14.73	3.25
H80	8.77	26.05	16.11	3.34
H90	10.07	27.31	16.88	3.43
H95	10.92	28.68	17.87	3.56
H99	11.80	30.33	19.37	3.76
CRR	0.25	0.62	0.47	0.08
P1minht	37.95	96.87	73.01	14.57

Lidar metrics	Minimum	Maximum	Mean	SD
P_all_minht	31.01	92.68	59.21	13.81
R_1st	40.07	162.79	91.61	25.97
P1mean	21.86	74.60	45.94	12.07
P1mode	8.39	92.85	52.58	22.17
P_all_mean	17.21	48.79	31.79	7.36
P_all_mode	0.48	79.14	39.69	19.69
R_all_mean	22.23	96.95	49.31	14.60
P_all_mode	0.68	147.66	60.85	31.51

2.2.6. Modelling Methods

In this study, we tested four modeling approaches to predict ground-based variables. Two parametric methods: Ordinary Least Square (OLS), which provides us with a benchmark with respect to other common studies, and Adaptive Least Absolute shrinkage and Selection Operator (ALASSO), which is known to allow for variable selection and to solve multicollinearity issues, at the expense of a small bias in the final prediction. Also, two non-parametric methods: Random Forest regression (RF) and Generalized Additive Model Selection (GAMSEL). The former deals with collinearity by building decision trees with different sets of predictors and averaging them, and the later combines a series of basis functions with a penalization term (similar to ALASSO) to overcome multicollinearity. Each method was used to predict basal area, dominant height, and merchantable volume using the Lidar metrics described in Table 1. Details of each modeling approach are briefly introduced in the sections below:

2.2.6.1. Ordinary Least Square Regression

Ordinary Least Square (OLS) regression is a traditional data modeling approach, well-established for its simplicity and easy-making inferences with good predictive performance. The OLS fitting was performed using field-measured plot attributes as a response variable and Lidar metrics as explanatory variables. To avoid bias due to the

multicollinearity between linear predictors, we used a subset regression approach to eliminate the high correlation between the explanatory variables to avoid bias due to the multicollinearity between linear predictors. This approach identifies the subset of Lidar-derived variables that best predicts the response variable using an efficient branch-and-bound algorithm implemented in the 'leaps' package in R to run the all-subset regression. To identify the best models from all-subset regression, we considered using Adjusted R-square (adj. R^2), Mallow's C_P (C_P), and Bayesian Information Criterion (BIC) because of their capability to address the overfitting issue. Model selection statistics such as the coefficient of determination (R^2), which in modeling is the measure for goodness-of-fit based on the proportion of explained variance, get higher, and the residual sum of squares (RSS) gets smaller when the model variables increase (Barrett 1974). Thus, the model, including all variables, always results in the best selection statistics. Adjusted R^2 deal with this problem by adjusting the value of R^2 by considering the impact of additional independent variables in the model. Similarly, C_P statistics address the limitation of R^2 by adding a penalty to the training RSS to adjust the number of predictors in the model (James, et al. 2013). For a Least Square Model with a fixed number of predictors, the Cp is computed using equation 2.3.

For a Least Square Model with a fixed number of predictors, the Cp is computed using equation 3.

$$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2) \tag{2.3}$$

where $\hat{\sigma}^2$represents the estimated variance, n is the number of observations, d is the number of predictors, and the term $2d\hat{\sigma}^2$ is a penalty term that increases as the number of predictors in the model increases. Likewise, BIC is derived from a Bayesian point of

view, and similar to C_p, we selected the model with the lowest BIC value. For a Least square model containing d predictors, the BIC is computed using equation 2.4.

$$BIC = \frac{1}{n\hat{\sigma}^2}(RSS + \log(n)d\hat{\sigma}^2) \quad (2.4)$$

Despite this method, variables selected by all subset regression can also have a high correlation. Therefore, once several "best" models were identified based on adj R^2, C_P, and BIC, the variance inflation factor (VIF) and conditional indices (CI) were used to examine if variables were highly correlated. In a model, VIF and CI reflect the magnitude of the correlations between the independent variables. Thus, the larger values for these indices indicate higher multicollinearity. According to Belsley, Kuh, and Welsch (1980), condition indices around 10 suggest weak correlations, while between 30 and 100 indicate moderate to strong correlations. If the condition index exceeds 100, it signs serious multicollinearity issues (Belsley, et al. 1980). Likewise, a VIF of above ten indicates the occurrence of multicollinearity among predictor variables (Belsley, et al. 2005; Holcomb 1999; Kutner 2004). Hence, in this study, models that included explanatory variables with a VIF > 10 and CI > 30 were discarded from the final model.

2.2.6.2. Adaptive Least absolute shrinkage and selection operator regression

The least absolute shrinkage and selection operator (LASSO) is a method of regression that has been gaining popularity in recent years, specifically in the statistical and machine learning field, due to its ability to handle high dimensional data sets (Zhao, et al. 2006). This type of regression adds a penalization term to the loss function that shrinks the regression coefficients toward zero (Tibshirani 2011). This is done by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value (λ). After

shrinkage, variables with a regression coefficient of zero are excluded from the model and, thus, constrain the complexity of the model (Ranstam, et al. 2018).

Penalized regression methods attempt to achieve two fundamental goals of regression, predicting accurately and selecting the relevant variables simultaneously (Koirala, et al. 2021). The major advantage of LASSO over conventional variable selection methods is that it performs a continuous variable selection and is more computationally feasible for larger datasets like the variables used in this study. The lasso regression loos function can be defined as follows:

$$Y = \beta_0 + \lambda_1 \times \beta_1 \times x_1 + \lambda_2 \times \beta_2 \times x_2 \dots\dots\dots\dots\dots \lambda_n \times \beta_n \times x_n \quad (2.5)$$

where Y are dependent variables, and in our case, ground measured values for dominant height, basal area, and volume. Similarly, $x_1, x_2, \dots x_n$ are the independent variable which in our case will be lidar metrics, $\beta_0, \beta_1, \beta_2 \dots \dots \beta_n$ are the regression coefficient for the independent variables. λ is the lasso shrinkage parameter that can shrink coefficients to zero.

This method needs a grid search method to find the best regularization parameters. Moreover, no proper regularization parameter allows this method to exploit the oracle properties defined by Fan and Li(Fan, et al. 2001). In linear regression, an estimator that identifies the correct subset of actual variables and has an optimal estimation rate is known as the oracle properties. Thus, the adaptive lasso, which has oracle properties, was proposed as an alternative to improve the variable selection properties of LASSO (Zou 2006). It has the same advantages as the lasso; additionally, it avoids overfitting and penalizing large coefficients.

In linear regression, lasso seeks to minimize:

Residual sum of square + λ × (Sum of the absolute value of the magnitude of coefficients)

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}|\beta_j| = RSS + \lambda \sum_{j=1}^{p}|\beta_j| \quad (2.6)$$

where λ indicates the shrinkage amount and is chosen through 10-fold cross-validation, β_j are the estimated coefficients. When λ = 0, it implies all features are included and, thus, equivalent to the linear regression.

Likewise, adaptive lasso seeks to minimize the following:

$$RSS + \lambda \sum_{j=1}^{p}\hat{w}_j|\beta_j| \quad (2.7)$$

where $\hat{w}_j$ is an adaptive weights vector parameter in the equation. This weight vector performs different regularizations for each coefficient.

We used the k-fold cross-validation approach for this study to choose the best λ value. For this approach, we first randomly divided the dataset into ten sub-samples of equal size. After that, nine sub-samples were used to develop a prediction, and the remaining sub-sample validated the model. This procedure was performed ten times, with each of the ten sub-samples being used for validation and the other for model development. The result is produced by combining the ten different validation results for a range of λ values and choosing the preferred λ, which is then used to determine the final model. We used *"glmnet"* package in R to perform adaptive lasso regression for this study.

2.2.6.3. Random Forest

The RF algorithm is a nonparametric ensemble learning method for classification or regression based on several decision trees, which was developed by Breiman (2001). This

method has several advantages with respect to linear regression when best Lidar metrics selections are the ultimate goal: e.g. (1) Variable deletion is not required since it can handle a large number of input variables; (2) it estimates the variables that are important in the classification, measured as the mean decrease accuracy; (3) the generated forests can be saved for future use on other data; and (4) it reduces overfitting and is, therefore, more accurate compared to boosted regression-based methods that are trained on the same data (Cootes, et al. 2012). The regression was carried out using the 'random forest' (Liaw, et al. 2002) package in R. Accuracy of the RF-based method also depends on tuning parameters, such as 'Ntree' (i.e., number of trees grown) and 'mtry' (i.e., number of predictors sampled for splitting at each node); thus, their values have to be optimized carefully. Parameter values with complex rules tend to overfit the training data; as a result, a model performs very well for the training data but may yield the worst prediction for the independent data (Probst, et al. 2019). In the case of RF, the Out-of-Bag (OOB) observation method is widely used for selecting suboptimal parameter values to avoid model overfitting. For this study, we used 'tuneRF' function in random forest packages, which search for an optimal value for the parameters based on the OOB error estimation method.

2.2.6.4. Generalized Additive Model Selection (GAMSEL)

GAM (Generalized Additive Model) is a nonparametric extension of linear regression models which splits the regression lines into multiple segments and then applies local smoothing functions to track the non-linearity in the relationships between the dependent and independent variables (Hastie, et al. 1987). In this study, we used an extension of GAMs called the Generalized Additive Model Selection (GAMSEL), a penalized

likelihood method for fitting sparse generalized additive models, specifically for data with high dimensions. By allowing the effect of each variable to be estimated as either a linear, low-complexity curve or a zero as determined by the data, this method interpolates between null, linear, and additive models (Chouldechova, et al. 2015). We used the 'gamsel' package in the R (Chouldechova, et al. 2018), where we first fitted the GAM model with all independent variables. After that, like the ALASSO model, we applied k-fold cross-validation with ten folds to find the best value for the penalization term. Which is then used to fit the final model.

2.2.7. Model Validation

The K- fold cross-validation method was used to assess the accuracy of the four modeling approaches evaluated in this study. This approach involves random division of the observations set into k groups, or folds, of equal size. Then the first fold was used as a validation set, and the model was fitted on the remaining k-1 folds (James, et al. 2013). To do this, we first divided the datasets into nine folds (9 plots in one-fold). Then the first fold was used as a validation set, and the model was fitted on the remaining eight folds. This procedure was repeated nine times with different subsets for validation. Finally, the model was used to predict the data on the first fold. An illustration of k-fold cross-validation is shown in Figure 2.2.

Figure 2.2 The illustration of the 9-fold cross-validation. The dataset is divided into nine random subsets; from there, eight are used for training and one for validation. This process is repeated nine times with different subsets for validation.

2.3. Results

2.3.1. Summary statistics and variable relationship

In the field measured 90 plots, the highest plot dominant height measured was 35.23 meters, and the lowest was 12.63 meters. Likewise, timber volume varied between 26m^3/ha to 363m^3/ha and BA between 6.21m^2/ha and 36.79m^2/ha. A more detailed statistical summary of field-measured attributes is shown in Table 2.4.

Table 2.4 Summary statistics of forest attributes from ground measurement

Response	**Minimum**	**Maximum**	**Median**	**Mean**	**SD**
Basal area (m^2/ha)	6.25	36.75	19.5	19.25	7
Dominant height (m)	12.67	35.23	19.48	20.49	4.11
Volume (m^3/ha)	29.75	363	139.25	130.5	63.25
Tree density (trees/ha)	750	1925	1375	1354.4	10

Figure 2.3 shows Pearson correlation coefficients between each response variable and the highly correlated four corresponding lidar metrics. That shows the 75th and 80th percentile, representing the middle and upper canopy of a stand, strongly correlate with the BA and the Volume. Furthermore, the 90th and 95th percentile had the highest correlation coefficient of 0.93 with the dominant height. Besides these four variables shown in figure 3, all other response variables also demonstrate moderate to high correlations with the response variables.

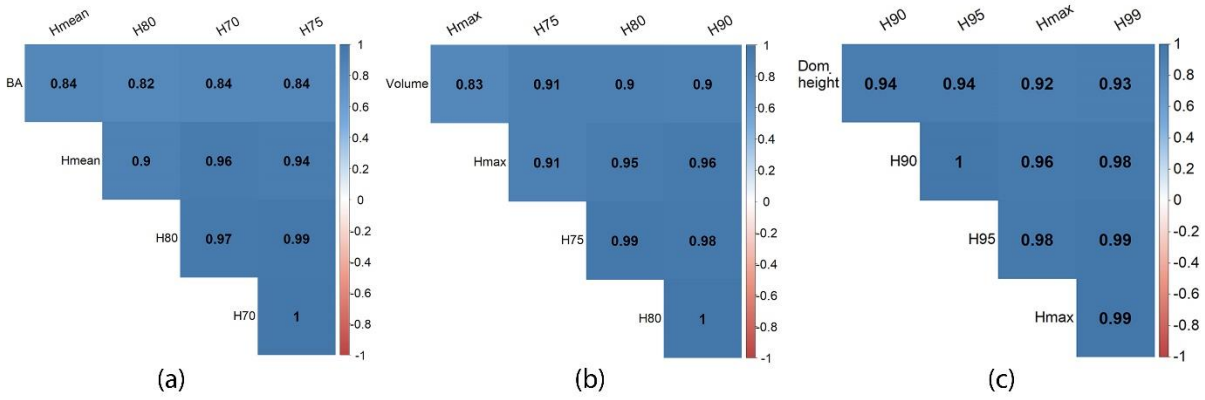

Figure 2.3 Pearson correlation coefficients for each response variable and the four highly correlated predictor variables. (a) Pearson correlation coefficients plot for BA; (b) Pearson correlation coefficients plot for volume; (c) Pearson correlation coefficient

2.3.2. Least squares regression model

After the iterative variable selection process using least squares regression, the final model retained only two predictors that satisfied the multicollinearity test, i.e., VIF and CI. In the final model for volume and basal area, the 75th percentile of echo height and the percentage of first return above mean echo height were the two variables retained (equations 2.8 and 2.9). For both equations, the 75th percentile had the most significant coefficient, indicating its higher relative importance for predicting both volume and basal area.

$$Volume = -5.271 + 0.5176p_{75} + 0.0612rp_1 \quad (2.8)$$

$$BA = -0.414 + 0.0486p_{75} + 0.0103rp_1 \quad (2.9)$$

where p_{75} is the 75th percentile of echo height, and rp_1 is the percentage of first returns above mean

Likewise, for dominant height, as shown in equation 2.10, the 90th and 40th percentiles of echo height are the two variables that are held in the model while keeping the collinearity indices within the acceptable range. Details of VIF and CI values for each selected variable are presented in Table 2.5.

$$Dom.height = 1.466 + 1.196p_{90} + 0.132p_{10} \quad (2.10)$$

where p_{90} is the 90th percentile of echo height and x_2 is the 40th percentile of echo height.

Table 2.5 The collinearity index value for all variables selected in the LS model for volume, basal area, and dominant height

	Volume		**Basal area**		**Dominant height**	
	p_{75}(H75)	rp_1(P1mean)	p_{75} (H75)	rp_1 (P1mean)	p_{90} (H90)	p_{10} (H40)
VIF	1.78	1.78	1.83	1.83	1.6	1.6
CI	9.46	13.26	9.41	13.32	6.36	13.58

3.3. Random Forest

In RF-based models, height percentile variables were the most important for all three response variables. As shown in Figure 2.4, the 90th, 75th, and 80th percentiles of echo height were the most significant variables for estimating volume. Similarly, all echo height metrics that describe the upper canopy level, namely 95th, 80th, 90th, and 99th percentiles, were the most important variables for the dominant height estimation model. Finally, in the basal area model, the mid-canopy echo heights, i.e., 60th and 65th percentiles of height, and the variable "percentage of first return above mean" are the most important metrics. A summary of all important variables is shown in Figure 2.4.

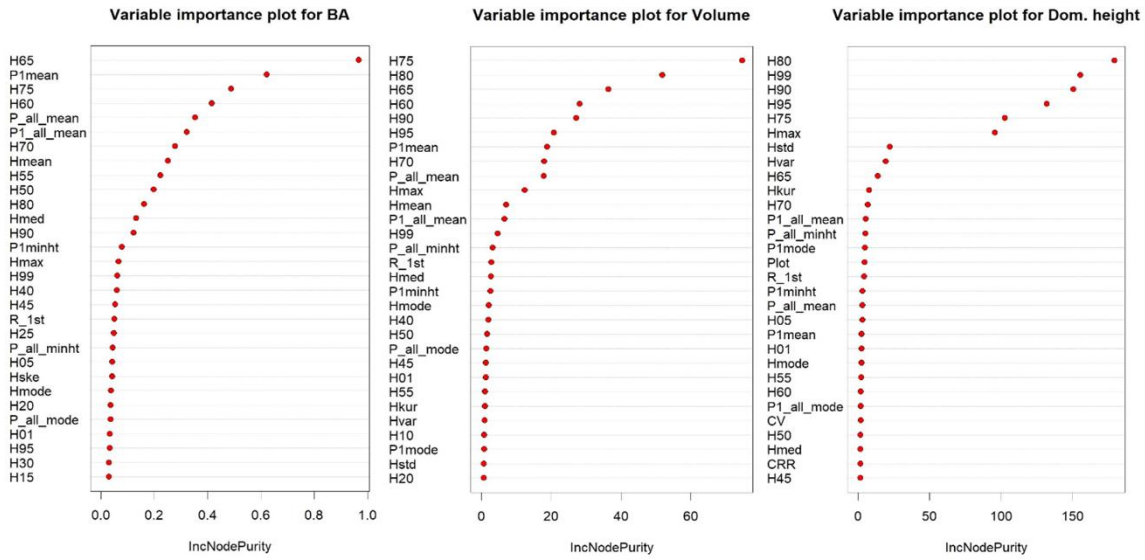

Figure 2.4. Variable importance plot for dominant height, BA, and Volume in Random Forest models.

2.3.4 ALASSO Regression model

The ALASSO method retained nine, eight, and eight variables on the final model for volume, basal area, and dominant height, respectively. The 1st percentile of echo height and the first return ratio over mean appeared in all three models; likewise, the 75th percentile, 85th percentile of echo height, percentage of first return above minimum height, and above mean were other selected variables in the model. For volume prediction, the 1st percentile of echo height was the most significant variable, with a coefficient of 1.122, followed by the 80th percentile height, with a coefficient of 0.25. The canopy relief ratio was the most crucial variable in the basal area model, followed by the 1st percentile of echo height with a coefficient of 0.549 and 0.08, respectively. Likewise, the 1st percentile of echo height with a coefficient of 0.69 has the highest impact, followed by the 90th percentile with a coefficient of 0.58 on the dominant height model.

A list of all variables and their coefficients selected for volume, BA, and Dominant height model with ALASSO is shown in Table 2.6.

Table 2.6 List of variables and their coefficients selected for Volume, BA, and Dominant height model by adaptive lasso regression method

	Coefficient		
	Volume	**BA**	**Dom. Height**
Intercept (β_0)	-5.701	-0.731	1.572
Height Maximum			0.069
Height standard deviation			0.137
Height variation	0.031		
Height skewness	0.385		
Height 01st percentile	1.122	0.075	0.684
Height 10th percentile		0.021	
Height 20th percentile	0.032		
Height 75th percentile	0.087	0.002	
Height 80th percentile	0.251	0.036	
Height 90th percentile			0.578
Height 99th percentile			0.275
Canopy relief ratio		0.549	
Percentage of first returns above the min height		0.003	0.002
(All returns above 0.5m/total first returns) *100			0.002
The percentage of first returns above mean	0.034	0.003	0.014
The percentage of first returns above mean	0.028	0.009	
The percentage of all returns above mean	0.014		
Total no. of variables	9	8	8

3.5 Generalized Additive Model Selection

Using the GAMSEL method, we obtained 11 variables for the Basal Area (BA) model, nine for the volume model, and five for the dominant height model. A common variable in all three models was the 'first return percentage above the mean', which reflects the canopy cover and structure in the stand. It was also one of the highly significant variables for volume and basal area, as shown in Figures 2.5 and 2.6. In a similar manner, for basal

area and volume, height means, height kurtosis, first percentile, 25th percentile, 80th percentile, and all returns above the mean and first return ratio were the repeating variables. Moreover, for volume, besides the first return percentage above the mean, the 80th and 25th percentile were the other key variables in the model representing the upper and lower canopy, respectively. Likewise, for basal area, echo height means, all return percentages above the mean and 75th percentile were the key variables in addition to the first return percentage above the mean, 80th, and 25th percentile.

As described earlier, one significant benefit of the GAM model is that it can incorporate a non-linear relationship between the response and predictor variable. Hence, in our final GAMSEL model, all variables retained had a non-linear relationship with respect to the response variable. The detailed relation between all selected variables (with nonzero coefficients) for basal area, volume, and dominant height models with the function of the last lambda value is shown in Figures 2.5, 2.6, and 2.7, respectively.

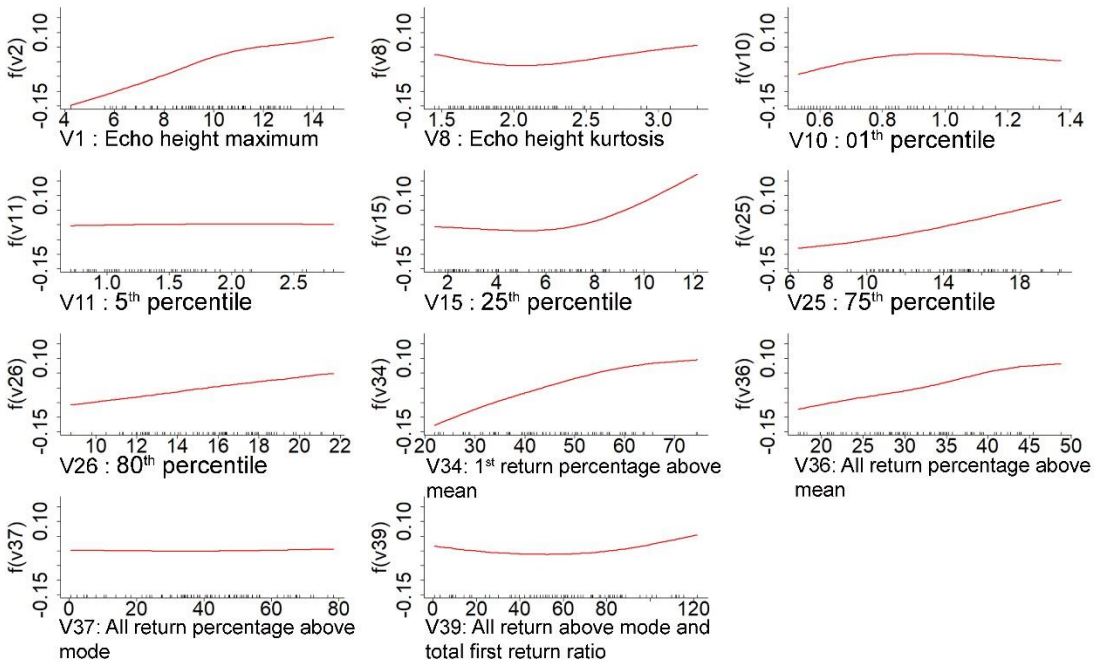

Figure 2.5. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for basal area estimation with the functions for the last value of the lambda (Y-axis).

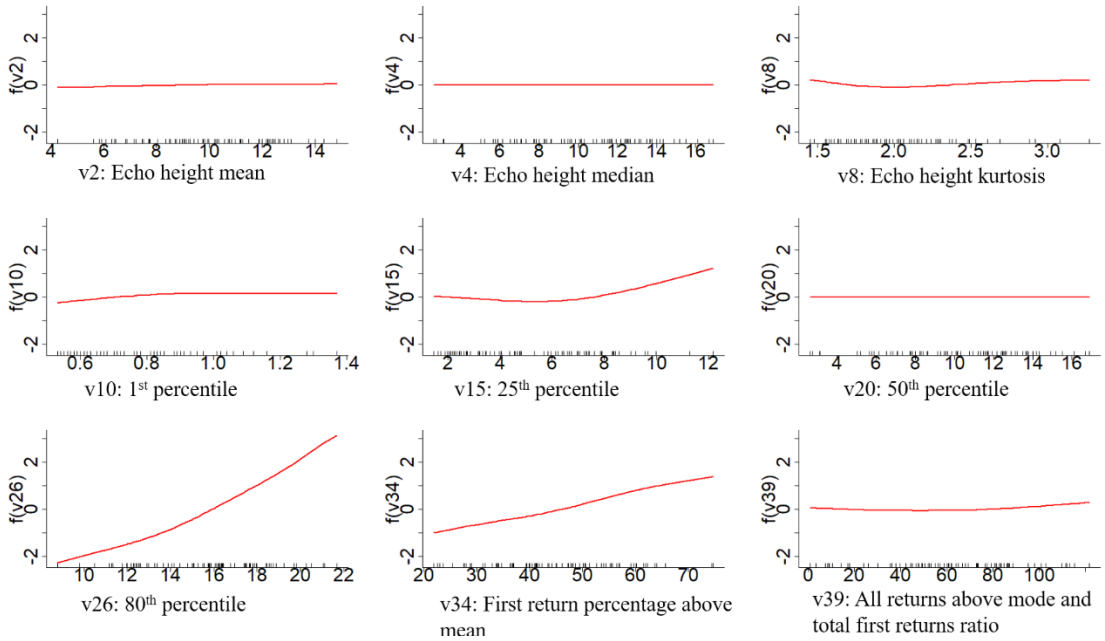

Figure 2.6. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for volume estimation with the functions for the last value of the lambda (Y-axis).

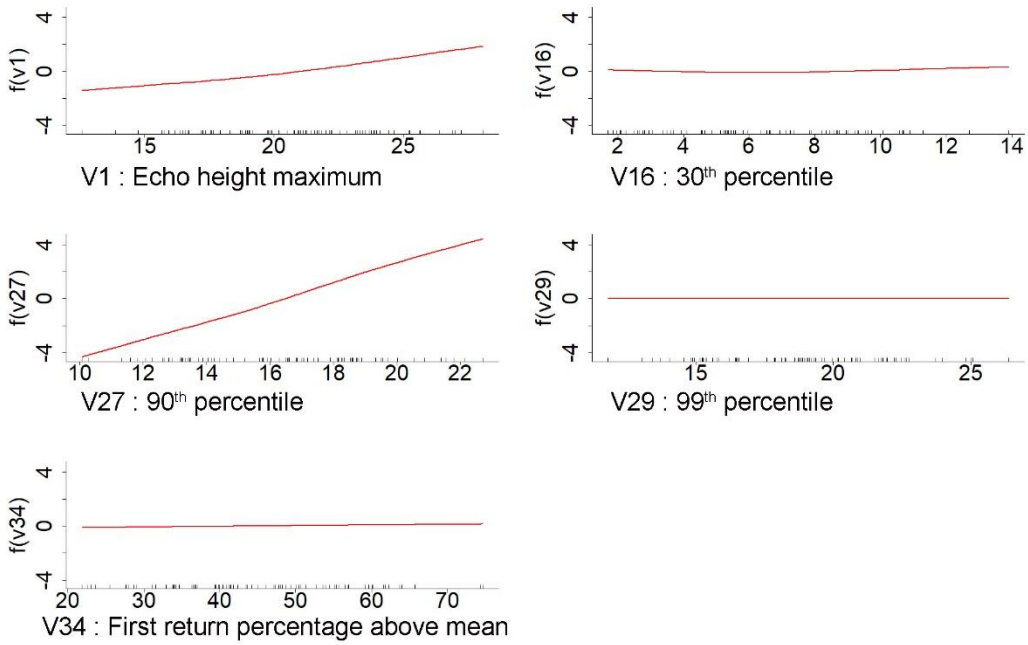

Figure 2.7. Each panel shows the relationship of the variable with nonzero coefficients (X-axis) in the GAMSEL model for dominant height estimation with the functions for the last value of the lambda (Y-axis).

2.3.6 Model comparison summary

In this comparison, for volume, the highest R^2 of 0.88 was achieved with ALASSO and GAMSEL, and ALASSO yielded the lowest RMSE of 0.781 m^3/plot. Likewise, the highest R^2 of 0.83 and the lowest RMSE of 0.109 for the basal area were achieved with the ALASSO method. In the dominant height case, ALASSO and GAMSEL yielded the highest R^2 of 0.87. The least RMSE of 1.353 m was achieved using the Least square modeling method. Consistent with the RMSE, the lowest Mean Square Error (MSE) was attained for the Volume and Basal Area models using the ALASSO method, followed by the GAMSEL method. For the Dominant Height model, the minimum MSE was achieved using the Least Squares (LS) method, followed by the ALASSO method. Table 2.7

summarizes the k-fold cross-validation results for four modeling methods. Figures 2.8 show the residual vs. fitted plot for volume, BA, and dominant height for the ALASSO model. Overall, the ALASSO method demonstrated better consistency and produced residual plots with fewer biases compared to the other three methods. A detailed residual vs. fitted plot for volume, BA, and dominant height for all four modeling methods is shown in Appendix B.

Table 2.7. K-fold cross-validation result summary for all response variables with four different modeling methods

Response Variable	**Model**	**R^2**	**RMSE**	**MSE**
Volume (m^3/plot)	LS	0.87	0.807	0.651
	ALASSO	0.88	0.781	0.634
	RF	0.86	0.846	0.713
	GAMSEL	0.88	0.789	0.622
Basal Area (m^2/plot)	LS	0.8	0.121	0.0145
	ALASSO	0.83	0.109	0.0129
	RF	0.8	0.122	0.0148
	GAMSEL	0.814	0.116	0.0135
Dom. Height (m)	LS	0.86	1.353	1.831
	ALASSO	0.87	1.359	1.849
	RF	0.83	1.487	2.317
	GAMSEL	0.87	1.421	2.021

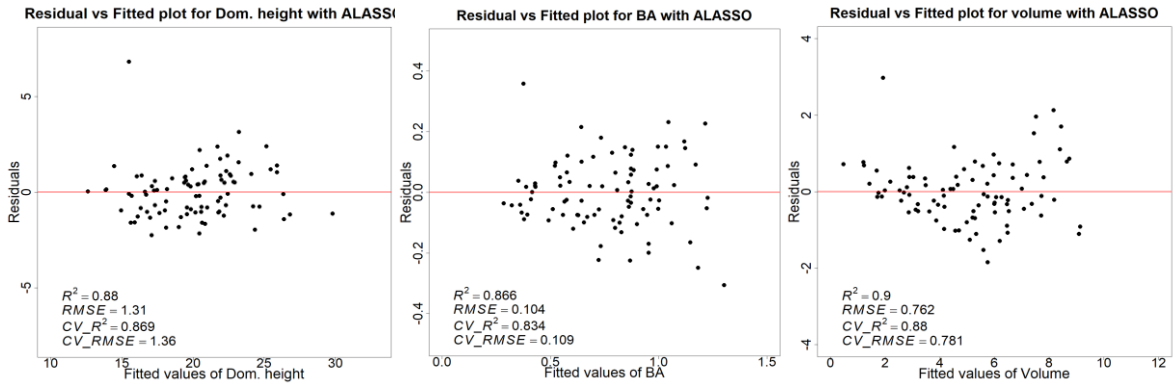

Figure 2.8. Each panel shows the residual vs. fitted plot for volume with ALASSO models. R2 and RMSE are the model accuracy before cross-validation, and CV_R2 and CV RMSE show the model accuracy result of 9-fold cross-validation.

2.3.7 Individual trees detection:

The individual tree detection method performed very poorly, yielding a correlation of 0.018 when comparing the number of trees detected by the model with the tree density measured in the field. However, when combining the number of trees estimated with the segmentation algorithm with the other thirty-nine metrics and applying the least square regression, the model accuracy (adjusted R^2) increased to 0.15. Likewise, for model accuracy, the random forest and GAMSEL regression methods achieved 0.18 and 0.178, respectively. Additionally, when incorporating all 39 lidar metrics and the number of trees estimated by the algorithm for each plot as explanatory variables in the ALASSO regression method, we obtained a comparable Adjusted R-squared value of 0.18 to that of the other methods.

2.4. Discussion

In recent decades, several forest-related industries around the globe have adopted laser scanning as auxiliary information to complement forest inventories. However, the modeling methods in use for forest parameter estimation are still not adequately exploiting the information in the point clouds (da Cunha Neto, et al. 2021). In this study, we compared the performance of four different modeling methods to predict various forest attributes using aerial lidar data. When comparing those four methods for volume estimation, adaptive lasso performed better than the other three. Although the overall coefficient of determination (adj. R^2) of 0.88 and the lowest RMSE of 19.53 m^3/ha achieved in this study with adaptive lasso regression, the difference compared to other methods is very nominal, i.e., 19.73 m^3/ha, 20.18m^3/ha, and 21.15 m^3/ha with respect to GAMSEL, LS (Least Square), and random forest, respectively. The RMSE of 19.53m^3/ha is low compared to other studies in similar forest conditions. A possible reason for this might be lidar data with better point cloud density (44 points/m^2) and a relatively smaller study area than other studies. Leite et al. 2020 compared a few parametric and nonparametric regression methods for area-based volume estimation of a eucalyptus plantation using low-lidar data (5 points/m^2) and concluded the best result of adj. R^2 0.83 and RMSE 40.71 m^3/ha using the artificial neural network (ANN) and lowest of adj. R^2 0.79 and RMSE 46.76 m^3/ha with the linear regression method (Leite, et al. 2020). Likewise, Silva et al. (2016) applied the principal component approach for predicting stem volume in eucalyptus plantations using lidar metric generated from medium-density (10/points/m^2) lidar point clouds. They achieved an adj R^2 of 0.87 and RMSE of 27.60 m^3/ha (Silva, et al. 2016b).

In this study 75th percentile of height was included in all four models as a vital metric for estimating the volume. After the 75th percentile, 90th percentile, canopy relief ratio, and percentage of first return above mean are primary metrics predicting volume. Consistent with the outcome obtained for the Volume model, the optimal model for the Basal Area was determined using the Adaptive Lasso Regression method. Adaptive lasso regression slightly improved the prediction for the basal area with an overall adj. R^2 of 0.83 and RMSE 2.73 m^2/ha compared to the adj. R^2 of 0.81, 0.80, and 0.80 and RMSE 2.90 m^2/ha, 3.03 m^2/ha, and 3.05 m^2/ha with GAMSEL, LS, and random forest, respectively. Likewise, for dominant height, both the adaptive lasso and GAMSEL model achieved the highest adj. R^2 of 0.87 followed by LS and random forest regression methods with adj. R^2 of 0.86 and 0.83, respectively. However, in terms of better RMSE for dominant height, the least square method resulted in a slightly better RMSE of 33.83 m/ha, followed by adaptive lasso, GAMSEL, and random forest with an RMSE of 33.98 m/ha, 33.53 m/ha, and 38.82 m/ha, respectively. Other researchers, such as Brown et al. (2022), compared linear and random forest methods for basal area estimation using variables from low-density lidar and multispectral imageries. In that study, they achieved a slightly better result with the random forest method (R^2 of 0.39 and 0.36, and RMSE of 5.662 m^2/ha and 5.731 m^2/ha with random forest and LS, respectively) but overall low accuracy compared to our study (Brown, et al. 2022). In another study by Li et al. (2022), they proposed a model with good generalization capability for the basal area using lidar-derived metrics for eucalyptus plantation forests achieving R^2 of 0.71 and rRMSE of 18.22% (Li, et al. 2022) using the least square method.

Furthermore, Li et al. (2022) used an exhaustive combination of lidar metrics with a clear meaning in forest mensuration. They formulated tree type-specific and region-generalized 86 models and sorted the most meaningful variables for estimating stand volume and basal area. In that study, they found the 75^{th} and 95^{th} percentile, the percentage of first returns above minimum height, and standard height deviation as important variables for basal area estimation. Similarly, the 95^{th} percentile, first return ratio above mean height, and coefficient of a height variance for volume estimation in eucalyptus forest (Li, et al. 2022). The 50^{th} to 75^{th} percentile represents the middle and upper part of the stand canopy; likewise, the 80^{th} to 99^{th} percentile characterizes the upper part of the canopy (Bouvier, et al. 2015; Næsset 2014). Therefore, including these mid and upper-canopy percentiles and density metrics representing the crown structure and cover makes the model more generalizable and significant. This result corresponds with the variables selected in our models. In our study, the 75^{th} percentile of echo height and the percentage of first returns above the mean are the two significant variables in the LS model for basal area and volume. Comparably, the foremost contributing variables in the random forest model are the height percentiles representing the canopy middle and upper parts, and the minimal contributing variables are below the 40^{th} percentile. Corresponding to LS and random forest, the most significant variables in the GAMSEL model are also from the middle and upper percentiles and density ratio variables representing canopy structure and cover. However, the variables included in adaptive lasso are distinct from the other three since the most contributing variable for the basal area is the canopy relief ratio, and for volume and dominant height is the first percentile of echo height. Although these selected variables are not among the topmost variables in the other three models, they

have similar properties to those selected variables in the other three models. The canopy relief ratio is the ratio of the mean and maximum value of all echo heights in a given area, signifying the canopy structure like other density metrics. Similarly, the first percentile represents the bottom portion of the stem and tends to fluctuate evenly according to the stem density.

We tested the performance of popular parametric, semiparametric, and nonparametric modeling methods for forest attribute estimation. In our test, the semiparametric models, i.e., adaptive lasso and the GAMSEL, slightly improved the estimates; however, none performed significantly differently from the others. Kangas et al. (2016), in their study of aboveground biomass (AGB) estimation, tested parametric, semiparametric, and nonparametric models using real-world and simulated lidar-based variables. In that comparison, although they achieved better accuracy with the nonparametric model, i.e., local constant kernel, they found it leading to underestimation of variance; thus, they recommended using the semiparametric, i.e., GAM, which they found more consistent and suitable for the internal model (Kangas, et al. 2016). On the other hand, even though parametric models such as least square regression are a more straightforward method, applications at a practical scale are scarcer. One primary reason is that when the number of dependent variables is very high, evaluating each variable to get a meaningful model conforming to all assumptions is not practical (Opsomer, et al. 2007). Hence, even though our results show the nominal difference in accuracies between the evaluated models, the semiparametric model seems more fitting to apply in a practical scenario.

In our study, the individual tree detection method failed to estimate the total number of trees in the plot correctly. One probable reason for such low correlations can be the small

sample plot size of just 400m^2 resulting in the insufficient number of trees to adjust the error due to the trees in the plot boundary. Moreover, in the field inventory, a standard GNSS receiver was used to establish the plot center, which might also have contributed to the low accuracy of ITD. Some studies have suggested that using a higher-density lidar point cloud will improve ITD accuracy. However, a study was done in a similar eucalyptus plantation forest by Corte et al. (2022), and applying a similar ITD approach in high-density lidar data (>1400 pts/m^2) did not noticeably contribute to improving accuracy (average R^2 = 0.24). They mentioned complex stand structure and irregular branching forms of the eucalyptus causing omission and commission when outlining trees, in addition to error associated with the GPS positioning as the major limitation (Corte, et al. 2022). Those mentioned limitations are also the case in our study. In addition, half of the sample plots are from regrowth forests with complex branching and multiple stems, which was challenging to delineate the crown of every branch.

2.5. Conclusions

In our study, the assessment of various methods for forest attribute estimation using high-dimensional lidar metrics showed no method to be superior to the others in terms of significant advantage. The ALASSO method, however, demonstrated the highest R2 values of 0.88, 0.83, and 0.87 for volume, basal area, and dominant height, respectively. Furthermore, GAMSEL and LSR followed this method regarding performance and lack of bias.

In conclusion, this study provides insights into the performance of different modeling methods for predicting forest attributes using aerial LiDAR data. In light of this result and the ease of implementation, we recommend using the ALASSO method for forest attribute estimation. However, further research is needed to investigate the generalizability of these results to other forest types and conditions.

2.6 References:

1. Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M. *et al.* 2012 A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia*, **168** (4), 1147-1160.

2. Barrett, J.P. 1974 The coefficient of determination—some limitations. *The American Statistician*, **28** (1), 19-20.

3. Belsley, D.A., Kuh, E. and Welsch, R. 1980 Identifying influential data and sources of collinearity. *Regression Diagnostics*.

4. Belsley, D.A., Kuh, E. and Welsch, R.E. 2005 *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

5. Bian, Y., Zou, P., Shu, Y. and Yu, R. 2014 Individual tree delineation in deciduous forest areas with LiDAR point clouds. *Canadian Journal of Remote Sensing*, **40** (2), 152-163.

6. Bouvier, M., Durrieu, S., Fournier, R.A. and Renaud, J.-P. 2015 Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sensing of Environment*, **156**, 322-334.

7. Brown, S. 1997 *Estimating biomass and biomass change of tropical forests: a primer*. Food & Agriculture Org.

8. Brown, S., Narine, L.L. and Gilbert, J. 2022 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 Sensing*, **14** (11), 2708.

9. Chen, G., Zhao, K., McDermid, G.J. and Hay, G.J. 2012 The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data. *International Journal of Remote Sensing*, **33** (9), 2909-2924.

10. Chouldechova, A., Hastie, T. and Spinu, V. 2018 gamsel: Fit regularization path for generalized additive models. *R package version*, **1** (1).

11. Chouldechova, A. and Hastie, T.J. 2015 Generalized Additive Model Selection. *arXiv: Machine Learning*.

12. CIREN. 1999 Agrological study of VIII region ‘‘Soil description, materials and symbols’’ (original in
Spanish). *Tomo 2, Publicacio´n Ciren No. 121*.

13. Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K. *et al.* 2021 Modelling lidar-derived estimates of forest attributes

over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, **260**, 112477.

14. Cootes, T.F., Ionita, M.C., Lindner, C. and Sauer, P. Robust and Accurate Shape Model Fitting Using Random Forest Regression Voting. Springer Berlin Heidelberg, pp. 278-291.

15. Corte, A.P.D., da Cunha Neto, E.M., Rex, F.E., Souza, D., Behling, A., Mohan, M. *et al.* 2022 High-Density UAV-LiDAR in an Integrated Crop-Livestock-Forest System: Sampling Forest Inventory or Forest Inventory Based on Individual Tree Detection (ITD). *Drones*, **6** (2), 48.

16. da Cunha Neto, E.M., Rex, F.E., Veras, H.F.P., Moura, M.M., Sanquetta, C.R., Käfer, P.S. *et al.* 2021 Using high-density UAV-Lidar for deriving tree height of Araucaria Angustifolia in an Urban Atlantic Rain Forest. *Urban Forestry & Urban Greening*, **63**, 127197.

17. Duzan, H. and Shariff, N.S.B.M. 2015 Ridge regression for solving the multicollinearity problem: review of methods and models. *Journal of Applied Science*.

18. Fan, J. and Li, R. 2001 Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, **96** (456), 1348-1360.

19. Gleason, C.J. and Im, J. 2011 A review of remote sensing of forest biomass and biofuel: options for small-area applications. *GIScience & Remote Sensing*, **48** (2), 141-170.

20. Goerndt, M.E., Monleon, V.J. and Temesgen, H. 2010 Relating forest attributes with area-and tree-based light detection and ranging metrics for western Oregon. *Western Journal of Applied Forestry*, **25** (3), 105-111.

21. Hall, S., Burke, I., Box, D., Kaufmann, M. and Stoker, J.M. 2005 Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management*, **208** (1-3), 189-209.

22. Hastie, T. and Tibshirani, R. 1987 Generalized Additive Models: Some Applications. *Journal of the American Statistical Association*, **82** (398), 371-386.

23. Hirata, Y., Furuya, N., Suzuki, M. and Yamamoto, H. 2009 Airborne laser scanning in forest management: individual tree identification and laser pulse penetration in a stand with different levels of thinning. *Forest Ecology and Management*, **258** (5), 752-760.

24. Holcomb, J.P. 1999 Applied Regression Analysis: A Research Tool. *The American Statistician*, **53** (2), 170.

25. Holopainen, M. and Kalliovirta, J. 2006 Modern data acquisition for forest inventories. In *Forest Inventory*, Springer, pp. 343-362.

26. Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. and Falkowski, M.J. 2008 Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, **112** (5), 2232-2245.

27. Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M.-H., Wong, W.V.C., Ling, Z.-Y. *et al.* 2014 Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR. *Forest ecology and management*, **328**, 335-341.

28. James, G., Witten, D., Hastie, T. and Tibshirani, R. 2013 *An introduction to statistical learning*. Springer.

29. Kangas, A., Myllymäki, M., Gobakken, T. and Næsset, E. 2016 Model-assisted forest inventory with parametric, semiparametric, and nonparametric models. *Canadian Journal of Forest Research*, **46** (6), 855-868.

30. Koch, B., Heyder, U. and Weinacker, H. 2006 Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing*, **72** (4), 357-363.

31. Koirala, A., Montes, C.R. and Bullock, B.P. 2021 Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US. *Forest Ecology and Management*, **479**, 118610.

32. Kronseder, K., Ballhorn, U., Böhm, V. and Siegert, F. 2012 Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, **18**, 37-48.

33. Kutner, M., C. Nachtsheim, and J. Neter. 2004 *Applied Linear Regression Models*. 4 edn: McGraw-Hill/Irwin, New York.

34. Kwak, D.-A., Lee, W.-K., Lee, J.-H., Biging, G.S. and Gong, P. 2007 Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, **12** (6), 425-434.

35. Latifi, H., Nothdurft, A. and Koch, B. 2010 Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, **83** (4), 395-407.

36. Leite, R.V., Amaral, C.H.d., Pires, R.d.P., Silva, C.A., Soares, C.P.B., Macedo, R.P. *et al.* 2020 Estimating stem volume in eucalyptus plantations using airborne LiDAR: A comparison of area-and individual tree-based approaches. *Remote Sensing*, **12** (9), 1513.

37. Li, C., Chen, Z., Zhou, X., Zhou, M. and Li, Z. 2022 Development of Generalized Estimation Models of Forest Inventory Attributes Using an Exhaustive Combination of Airborne Lidar-Derived Metrics. *Available at SSRN 4104346*.

38. Li, W., Guo, Q., Jakubowski, M.K. and Kelly, M. 2012 A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering & Remote Sensing*, **78** (1), 75-84.

39. Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H. *et al.* 2016 Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, **115**, 63-77.

40. Liaw, A. and Wiener, M. 2002 Classification and regression by randomForest. *R news*, **2** (3), 18-22.

41. Lim, K., Treitz, P., Baldwin, K., Morrison, I. and Green, J. 2003 Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, **29** (5), 658-678.

42. Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C. and Culvenor, D.S. 2005 Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, **214** (1), 398-412.

43. Marrs, J. and Ni-Meister, W. 2019 Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sensing*, **11** (7), 819.

44. McRoberts, R.E. and Tomppo, E.O. 2007 Remote sensing support for national forest inventories. *Remote sensing of environment*, **110** (4), 412-419.

45. Næsset, E. 2002 Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote sensing of environment*, **80** (1), 88-99.

46. Næsset, E. 2014 Area-based inventory in Norway–from innovation to an operational reality. In *Forestry applications of airborne laser scanning*, Springer, pp. 215-240.

47. Naesset, E. 1997 Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote sensing*, **52** (2), 49-56.

48. Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M. *et al.* 2004 Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, **19** (6), 482-499.

49. Opsomer, J.D., Breidt, F.J., Moisen, G.G. and Kauermann, G. 2007 Model-assisted estimation of forest resources with generalized additive models. *Journal of the American Statistical Association*, **102** (478), 400-409.

50. Ørka, H.O., Jutras-Perreault, M.-C., Candelas-Bielza, J. and Gobakken, T. 2022 Delineation of Geomorphological Woodland Key Habitats Using Airborne Laser Scanning. *Remote Sensing*, **14** (5), 1184.

51. Osborne, J.W. and Waters, E. 2002 Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, **8** (1), 2.

52. Packalén, P., Mehtätalo, L. and Maltamo, M. 2011 ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. *Annals of Forest Science*, **68** (6), 1085-1092.

53. Pajeres, G. 2015 Overview and current status of remote sensing applications based on unmanned aerial vehicles. *Photogramm. Eng. Remote Sens*, **81** (4), 281-329.

54. Pascual, A., Bravo, F. and Ordoñez, C. 2019 Assessing the robustness of variable selection methods when accounting for co-registration errors in the estimation of forest biophysical and ecological attributes. *Ecological Modelling*, **403**, 11-19.

55. Picard, N., Saint-André, L. and Henry, M. 2012 Manual for building tree volume and biomass allometric equations: from field measurement to prediction. *Manual for building tree volume and biomass allometric equations: from field measurement to prediction, FAO; Food and Agricultural Organization of the United Nations (2012)*.

56. Popescu, S.C. and Wynne, R.H. 2004 Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, **70** (5), 589-604.

57. Popescu, S.C., Wynne, R.H. and Nelson, R.F. 2003 Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian journal of remote sensing*, **29** (5), 564-577.

58. Probst, P., Wright, M.N. and Boulesteix, A.L. 2019 Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, **9** (3), e1301.

59. Ramírez, M., Rodríguez, J., Peredo, M., Valenzuela, S. and Mendonça, R. 2009 Wood anatomy and biometric parameters variation of Eucalyptus globulus clones. *Wood Science and Technology*, **43** (1), 131-141.

60. Ranstam, J. and Cook, J.A. 2018 LASSO regression. *British Journal of Surgery*, **105** (10), 1348-1348.

61. Reutebuch, S.E., Andersen, H.-E. and McGaughey, R.J. 2005 Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. *Journal of Forestry,* **103** (6), 286-292.

62. Salas, C., Ene, L., Gregoire, T.G., Næsset, E. and Gobakken, T. 2010 Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, **114** (6), 1277-1285.

63. Shah, A.D., Bartlett, J.W., Carpenter, J., Nicholas, O. and Hemingway, H. 2014 Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *American Journal of Epidemiology*, **179** (6), 764-774.

64. Shi, Y., Wang, T., Skidmore, A.K. and Heurich, M. 2018 Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS Journal of Photogrammetry and remote sensing*, **137**, 163-174.

65. Shin, J., Temesgen, H., Strunk, J.L. and Hilker, T. 2016 Comparing Modeling Methods for Predicting Forest Attributes Using LiDAR Metrics and Ground Measurements. *Canadian Journal of Remote Sensing*, **42** (6), 739-765.

66. Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K. *et al.* 2016a Imputation of individual longleaf pine (Pinus palustris Mill.) tree attributes from field and LiDAR data. *Canadian journal of remote sensing*, **42** (5), 554-573.

67. Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.e. *et al.* 2016b A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry: An International Journal of Forest Research*, **89** (4), 422-433.

68. Strunk, J.L., Reutebuch, S.E. and Foster, J.R. LiDAR inventory and monitoring of a complex forest.

69. Sun, Y., Jin, X., Pukkala, T. and Li, F. 2022 Predicting Individual Tree Diameter of Larch (Larix olgensis) from UAV-LiDAR Data Using Six Different Algorithms. *Remote Sensing*, **14** (5), 1125.

70. Tibshirani, R. 2011 Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73** (3), 273-282.

71. Toth, C. and Jóźków, G. 2016 Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, **115**, 22-36.

72. Wang, Y., LeMay, V.M. and Baker, T.G. 2007 Modelling and prediction of dominant height and site index of Eucalyptus globulus plantations using a

nonlinear mixed-effects model approach. *Canadian Journal of Forest Research*, **37** (8), 1390-1403.

73. White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T. and Tompalski, P. 2016 Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, **42** (5), 619-641.

74. Zhao, P. and Yu, B. 2006 On model selection consistency of Lasso. *The Journal of Machine Learning Research*, **7**, 2541-2563.

75. Zhu, L., Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H. and Haggren, H. 2018 A review: Remote sensing sensors. *Multi-purposeful application of geospatial data*, 19-42.

76. Zou, H. 2006 The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** (476), 1418-1429.

CHAPTER 3

ASSESSMENT OF UNDERSTORY VEGETATION IN A PLANTATION FOREST OF THE SOUTHEASTERN UNITED STATES USING TERRESTRIAL LASER SCANNING[2]

[2] This chapter is based in full on the previously published article [Adhikari, A., Peduzzi, A., Montes, C. R., Osborne, N., & Mishra, D. R. (2023). Assessment of understory vegetation in a plantation forest of the southeastern United States using terrestrial laser scanning. Ecological Informatics, 77, 102254.] Reprinted here with the permission of publisher.

Abstract

Forests of the southeastern United States are home to large timber industries with substantial contributions to global round wood production and paper products. Despite the success of plantations and the large timber industries in this area, pine growth remains constrained due to the competition between planted pine and the species in the understory. Moreover, effective control of this interspecies competition had shown a significant two- to four-times increase in stand productivity. Thus, this study aims to evaluate the use of laser scanning derived data from Terrestrial Laser Scanner (TLS) to assess understory vegetation biomass, as conventional methods utilizing optical imagery have yet to be effective in quantifying and mapping evergreen understory in the southeastern coastal forests of the United States. For this study, we utilized TLS to scan the entire forest profile within 60 sample plots of an operational loblolly pine plantation in Nassau County, Florida, and collected understory biomass data through destructive sampling. We compared three TLS-based volume estimation methods for predicting understory biomass and applied the Adaptive Least Absolute Shrinkage and Selection Operator (ALASSO) regression method to derive an optimal model by integrating the most efficient volume estimation methods and other TLS-derived standard metrics. Our study identifies the 20^{th} percentile of the echo height and a 3D volume metric based on mean height and understory cover as the most significant explanatory variables for the optimal model. The model exhibits high accuracy, with Adjusted R-squared (Adj. R^2) of 0.80 and Root Mean Square Error (RMSE) of 234.8 grams per square meter (g/m^2). Additionally, the mean height and understory-based volume estimation method outperformed other methods, such as voxel count and three dimensional (3D) alpha hall-

based method, with Adj. R^2 of 0.79, 0.47, and 0.57, and RMSE values of 288, 448.6, and 413 g/m^2, respectively, when used as a single variable in the model. The resulting model successfully predicted and quantified understory vegetation, showcasing TLSs potential to accurately capture biomass variation, particularly in evergreen-dominated pine plantation forests in the southeastern United States coastal regions. As a tool for monitoring understory, TLS can be used to aid plantation forest managers in identifying areas that require control measures for enhanced management practices.

Keywords: Understory biomass assessment, voxelization, alpha-hull fitting, mean-height, understory cover-based volume estimation method

3.1 Introduction

Forests of the southeastern United States are home to large timber industries producing twelve percent of the worlds industrial round woods and nineteen percent of pulp and paper products (Oswalt, et al. 2014). In addition, decades of forest research efforts and successful pine plantation silviculture have established this region into the wood basket of the United States (Fox, et al. 2007; Schultz 1997). However, despite considerable progress in silvicultural practices such as site preparation, nursery management, fertilization, and weed control, the growth rate of existing pine plantations still lags behind those of other temperate and subtropical regions (Allen, et al. 2005). Low soil fertility and interspecies competition between herbaceous and woody plants for resources are widely recognized as one of the major limiting factors for pine growth in the South (Adegbidi, et al. 2002; Allen, et al. 2005; Peduzzi, et al. 2010). In particular, the competition between pine, hardwoods, and woody evergreen shrub species commonly found in the coastal plain represents a substantial limitation on pine growth. Furthermore, studies have shown that controlling competing vegetation can significantly improve stand productivity and survival rates. For example, Miller et al. (2003) conducted a study on loblolly pine plantations across 13 sites in 7 states in the southeastern US, where they implemented a vegetation control treatment. Their findings revealed a remarkable 30-148 percent increase in wood volume yield after 15 years post-treatment (Miller, et al. 2003). Similarly, Michael (1980) conducted a study on longleaf pine in three sites located in Alabama. The experiment involved a similar vegetation control treatment, resulting in a substantial 40 percent increase in wood volume yield after 20 years (Michael 1980). Furthermore, studies conducted in other plantation areas have consistently demonstrated a two- to four-times increase in stand productivity

following the implementation of vegetation control measures (Fortson, et al. 1996; Wagner, et al. 2006; Zhang, et al. 2013).

However, assessing understory vegetation has been challenging. Fixed plots and ocular-based sampling are the commonly used methods for understory vegetation assessment (Bonham 2013). Nevertheless, these field-based methods are often time-consuming, expensive, and usually limited to small areas (Eskelson, et al. 2011; Lovell, et al. 2005). Furthermore, visual estimation of understory vegetation often leads to inconsistency among observers (Willem, et al. 2000). In a few other studies, understory vegetation had been estimated using the destructive sampling method to quantify dry biomass (Sabo, et al. 2008; Stagg, et al. 2006). However, field-based methods have intrinsic error sources, such as bias in species selection, inadequate sample size capturing existing variability, and representativeness of sampling locations (Barker, et al. 2001). These difficulties and associated uncertainties limit the usefulness of field-based methods, particularly for understory vegetation estimation.

Remote sensing (RS) methods appear to be a more rational and feasible approach for the timely detection and quantification of competing understory vegetation as they can collect repetitive measurements across large areas and with difficult accessibility. Over the last three decades, researchers have focused on improving the use of remotely sensed information beyond just auxiliary variables in field-based data collection. However, conventional RS data such as multispectral satellite imagery from Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) have limited applicability for understory vegetation detection due to overstory vegetation interference and a lack of plant-level and site-specific methodologies (Singh, et al. 2018). In addition, multispectral

optical data are not sensitive enough to three-dimensional (3D) vegetation structures (Keane, et al. 2001). While they may allow for broad classification by identifying the presence or absence of certain species, they lack the minute spectral details needed to differentiate understory from overstory and the capacity to estimate the volume of competition and its corresponding impact on planted trees. This lack of spectral and spatial details makes it challenging to classify and quantify individual plants at the species level, thus limiting our ability to assess smaller areas or objects like understory vegetation (Wing, et al. 2012). Additionally, passive sensor data primarily contain information from a forest canopy or overstory portion. Consequently, researchers have found a weak relationship between overstory trees and understory vegetation cover, with coefficient of determination (R^2) values ranging from 0.27 to 0.34 (Eskelson, et al. 2011; Suchar, et al. 2010).

Few studies have used phenological approaches in combination with passive RS data for over- and understory vegetation detection (Clinton, et al. 2010; Singh, et al. 2018). This approach uses time-series RS data to detect and map understory vegetation based on the phenological differences between understory and overstory species for any part of the year. However, in the case of the pine plantation forests in the coastal plain along the southeastern United States, competing understory are mostly woody evergreen species with no significant phenological differences, which makes this region more complicated for understory vegetation assessment. Also, to date, only a few studies have examined methods for predicting and mapping understory vegetation. These studies have found challenges with above canopy conventional RS methods (Peduzzi, et al. 2010; Venier, et al. 2019), highlighting the need for a sensor/dataset capable of capturing the below

canopy structure and providing more accurate and detailed information on understory vegetation.

One alternative method is Laser Scanning, specifically LiDAR (Light Detection And Ranging), an active remote sensor that uses electromagnetic energy capable of constructing 3D maps of the complete vegetation profile (Reutebuch, et al. 2005; Venier, et al. 2019). A laser scanner measures the 3D position of points at the surface, achieving an accuracy of a few decimeters in vegetation canopies and offering high-density point clouds (Popescu, et al. 2004). Terrestrial Laser Scanning (TLS) is a ground-based lidar sensor that collects a very high point-cloud density of up to 1,000 pts/m^2. TLS can collect 3D understory vegetation structures at a high spatial resolution and centimeter-to-millimeter-level accuracy (Lague, et al. 2013; Newnham, et al. 2015). With these merits, TLS had been recognized as an effective nondestructive method for improving the accuracy of understory biomass prediction by enhancing the resolution and range of measurement. However, despite its recognized benefits, only a few studies have utilized TLS for understory vegetation estimation. Existing literature had primarily focused on developing methods and comparisons (Olsoy, et al. 2014; Xu, et al. 2020), and there is still a lack of exploration on the application of TLS for understory quantification and biomass estimation, particularly in the pine plantation in the southeastern United States coastal forests.

Therefore, this study addresses this research gap and explores the feasibility of implementing TLS-based metrics to help quantify evergreen understory vegetation more accurately than traditional methods in pine plantations in the southeastern United States. In addition, we compare model results using TLS metrics extracted from previously

published methods such as voxelized (Cooper, et al. 2017; Venier, et al. 2019) and alpha-hull fitting (Olsoy, et al. 2014). We also introduce a new approach adapted from Li et al. (2021), based on mean height and understory cover, for extracting understory volume from TLS data (Li, et al. 2021).

Furthermore, the study explores the practical significance of employing a TLS-based method to accurately quantify evergreen understory vegetation in pine plantations of the southeastern United States. The study area in the southeastern coastal region is known for its large understory biomass due to the favorable environment created by warm temperatures, high humidity, and abundant rainfall (Johnson, et al. 2017). However, accurately estimating the amount of biomass in the understory had proven to be challenging for forest managers. The primary reasons for these challenges are the high costs associated with measurement, inconsistencies in methodologies employed for measurement, and the complex nature of the understory vegetation in this region, which predominantly consists of evergreen shrubs and hardwood trees. Moreover, while several studies have investigated methods for overstory biomass estimation, little attention had been given to understory vegetation estimation and prediction (Campbell, et al. 2018). Thus, this study aimed to contribute to this understudied area by investigating understory biomass prediction using TLS in the southeastern coastal region. Although some studies have explored related aspects, to our knowledge, no previous research had specifically examined the application of TLS data for understory biomass prediction in this particular region. This study fills this research gap and provides an innovative approach for forest managers to estimate understory biomass in this region.

Several studies have explored the potential of terrestrial and airborne LiDAR data for aboveground biomass assessment. However, only a few of these studies have explored ways to include understory biomass assessments (Ferraz, et al. 2016). One study employing laser scanning for understory assessment by Campbell et al. (2018) his study used small-footprint airborne lidar to assess understory density, achieving an R^2 of 0.44 with multiple linear regression modeling and normalized point density as an explanatory variable. Their study emphasized that higher-density LiDAR pulses are more favorable for estimating understory vegetation. In this context, it is worth noting that TSL stands out for its ability to generate very high-density data. It is particularly advantageous for capturing the subtleties of the understory layers, bringing great potential for improving understory assessments. By utilizing TLS, this research addresses the need for accurate and detailed information on understory vegetation, providing a valuable tool for guiding plantation forest managers in implementing effective control measures.

Overall, this study aims to achieve the following primary objectives:

1. Assessing Evergreen Understory Vegetation: The first objective of this study is to address the research gap by exploring the feasibility of utilizing TLS-based metrics to quantify evergreen understory vegetation more accurately than traditional methods. By employing TLS data, we aim to capture the intricacies of the understory layers and provide detailed information on understory vegetation in the study area known for its complex nature and large understory biomass.

2. Comparative Analysis of Volume Estimation Methods: The second objective involves comparing different volume estimation methods for quantifying understory vegetation

biomass using TLS data. This comparative analysis will help identify the most effective method for accurate biomass estimation in the southeastern coastal region.

3. Fine-Scale Understory Biomass Mapping: The third objective is to apply the optimal model, identified through comparative analysis and modeling, for mapping understory biomass and quantifying variations within the sample plots at a fine scale.

3.2. Methods

3.2.1 Study area

This study was carried out in an intensively managed loblolly pine plantation forest under the administration of Rayioner Inc, located in Nassau County, Florida (30°60'N; 81°80'W). The county is characterized by a warm and humid summer season and mild winter conditions. Average summer temperatures reach 26°C, while the remaining months of the year average around 17°C. A substantial proportion of the annual precipitation, approximately 65% (an average of 132.8 cm), occurs during the summer months, while the remaining months receive an average of approximately 7.62 cm of rainfall (Watts 1991). The soils in this county are predominantly sandy and well-drained, with low nutrient content (Nassau County Extension, 2020). The dominant vegetation in the region consists of loblolly pine (*Pinus taeda*) and slash pine (*Pinus elliottii*) plantation forests, which have been intensively managed to optimize timber production by promoting growth and yield. Despite intensive preparation and establishment of plantation forests, understory vegetation abundance in the study area remains significantly high, with the majority comprising native evergreen understory vegetation. Notably, gallberry (*Ilex glabra* (L.) Gran) and saw palmetto (*Serenoa repens* (Bartr,) small) are recognized as the predominant species, while other species include ti-ti (*Cyrilla recemiflora* L.), fetterbush (*Lyonia Lucida* (Lam) K. Koch), staggerbush (*Lyonia ferruginea* (Walter) Nutall), sweet bay (*Magnolia virginiana* L.), blueberries (*Vaccinium spp.*), saw greenbrier (*Similax bona-nox* L.), cat greenbrier (Similax glauca Walt.), and woody evergreen species such as red maple (*Acer rubrum* L.) and sweetgum (*Liquidamber styraciflua* L.)(Lauer, et al. 1999; Peduzzi, et al. 2010). Figure 3.1

illustrates the study area and the distribution of the inventory plots, and Figure 3.2 shows the dominating understory vegetation.

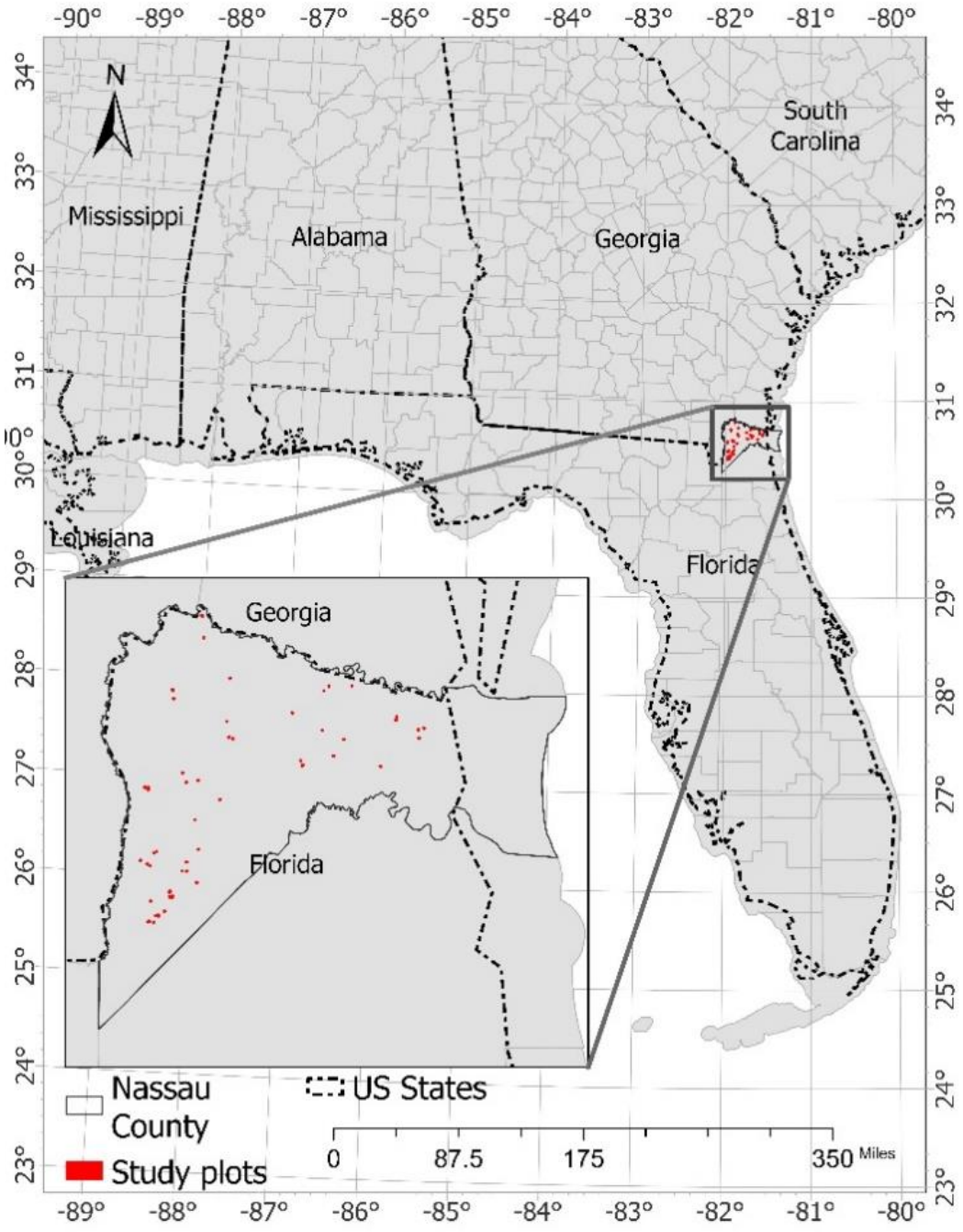

Figure 3.1 Study area map showing the spatial distribution of sample plots inside Nassau County, Florida, US

Figure 3.2: Dominating evergreen understory species in the study area, a) Gallberry (Ilex glabra), b) saw palmetto (Serenoa repens), c) ti-ti (Cyrilla racemiflora), and d) fetterbush (Lyonia lucida)

3.2.2 Field Planning and data collection

For the field data collection, we classified loblolly plantation stands with ages ranging from 10 to 21 years old and without thinning into six strata based on understory vegetation abundance and four strata based on mean stand height. An understory density raster file of 10-meter spatial resolution created by Rayioner Inc. based on LiDAR data was used to classify understory vegetation abundance. We reclassified the density raster values into six broad classes using natural breaks methods in ArcGIS pro 2.9.1. In our classification scheme, we assigned specific classes to represent varying levels of understory vegetation, with class one denoting areas characterized by a very low presence of understory vegetation, while class six represented regions with a significantly high abundance of understory vegetation. This approach enabled us to effectively capture and quantify the degree of understory vegetation cover across the study area. Similarly, we

classified the mean height of all selected stands into four bins (4.6m – 8.3m, 8.3m – 13.3m, 13.3m – 17.1m, and 17.1m – 25m). Based on these two criteria, with even distribution, we sampled sixty plots of 1600 m^2 (40m × 40m) inside all strata, as shown in Table 3.1.

Table 3.1 Distribution of sample plots based on height and understory abundance

	Stand Dominant Height bins (meters)				
Understory bins	**(4.6, 8.3)**	**(8.3, 13.3)**	**(13.3, 17.1)**	**(17.1, 25)**	**Total**
1 (No understory)	2	2	2	4	**10**
2	3	2	2	3	**10**
3	3	2	3	2	**10**
4	2	3	3	2	**10**
5	3	2	3	2	**10**
6 (Very high understory)	2	4	2	2	**10**
Total	**15**	**15**	**15**	**15**	**60**

Although we were sampling operational forest stands and did not always find enough stands to represent all bins, we ensured that each stratum contained at least two plots and no more than four plots while maintaining ten plots in each understory bin and fifteen plots in each height bin. In addition, during plot establishment, to ensure a consistent variability of understory vegetation within the sample plots, we kept a buffer distance of 40 meters between them and other land-use areas during the establishment of the plot. This was necessary because the areas adjacent to the road and other land-use areas exhibit a distinct distribution of understory vegetation influenced by such edges. However, establishing square-shaped 60 plots in the field using conventional methods such as measuring tape and a compass-based approach was not feasible due to the tall and dense understory vegetation in the sampled stands. Therefore, we utilized a Trimble Satellite-Based Augmentation System (SBAS) corrected Global Navigation Satellite System

(GNSS) receiver with an accuracy of ±1.5 meters to locate and establish all four corners of sample plots with high understory abundance.

Following plot establishment (Figure 1), we established two subplots of $1m^2$ (1m×1m) each, as shown in Figure 3.3, at the center of two diagonally opposite quadrants of each plot for destructive sampling of understory vegetation, and their location was recorded using the SBAS corrected GNSS receiver. These subplots served as the site for destructive sampling of understory vegetation, providing essential ground truth data. After establishing plots and subplots, TLS data was collected. Ensuring an appropriate point density and scanning range is crucial for TLS data acquisition, which relies on defining an optimal walking path. To ensure the desired point density, we followed existing literature recommendations by maintaining a spacing below 10 meters (Gollob, et al. 2019). Moreover, the operator walked around each subplot to maintain an adequate number of returns from the subplot, pointing the sensor toward the subplot center during the scanning process. We aimed to optimize data quality and capture sufficient and uniform point clouds for each subplot by employing these measures.

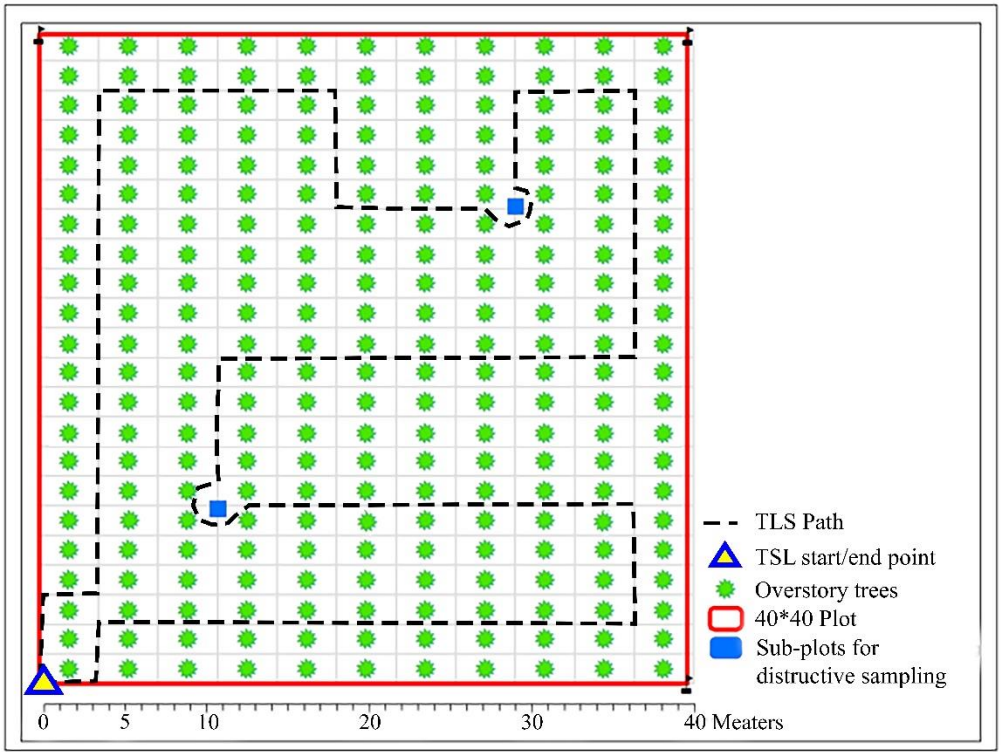

Figure 3.3: Graphical representation of a sample plot layout and subplot location for destructive sampling

After the terrestrial laser scanning, destructive sampling was undertaken, collecting all green plants inside the subplots (Figure 3.4). Out of 120 subplots, we gathered a destructive sample from 100 of them. The remaining 20 were subplots where, although the plot had understory vegetation present, the area within the subplot location or boundary did not. After destructively collecting samples, the green weights of each sample were recorded in the field. Then, the samples were dried at 105°C in an oven for 48 hours or until a constant dry weight was reached and weighed again to record their dry weight (Figure 4). The obtained dry weight value for each subplot was considered the dry understory biomass and used as a dependent variable for the modeling process.

Figure 3.4: The process for collecting understory samples and measuring their dry biomass. First, a 1m^2 PVC frame was used to establish an understory subplot (a). Then, after conducting destructive sampling, the samples were placed in bags and labeled (b). Next, the collected samples were dried in an oven (c) for 48 hours at 105°C before dry biomass weight measurement (d).

3.2.3 LiDAR data acquisition and processing

We collected ground-based LiDAR data using the GeoSLAM ZEB HORIZON personal mobile laser scanner (PLS) developed by GeoSLAM Ltd. (Nottingham, UK) on August 2022. The handheld device (weighing 1.3 kg) is equipped with a VLP-16 sensor (Velodyne LiDAR Inc., Morgan Hill, CA, USA), an optional Firefly 8si camera capable of capturing 4k resolution images, and an Inertial Measurement Unit (IMU). The VLP-16 (weighing 0.83 kg) is a 16 channels Light Detection and Ranging (LiDAR) unit capable of measuring up to 300,000 points per second in single return mode within a maximum range of 100 m. The device uses a continuous wavelength of 903 nm (near infrared) to determine distances with a range accuracy of ±3 cm. The VLP-16 has a field of view of 360° × 30° with a vertical angular resolution of 2° and a horizontal angular resolution of 0.1°–0.4°. The device has an internal rotation rate that ranges from 5 to 20 Hz. A revolving housing that rotates around an axis perpendicular to the internal rotation axis of

the VLP-16 at a rate of 31 min^{-1} is used to connect the VLP-16 to the ZEB HORIZON. The combination of internal and external rotation creates an angular field of view of 360° × 270°.

Data acquisition with ZEB HORIZON starts with IMU initialization to establish the local coordinate reference system. To convert this local reference system into a global coordinate reference system, a survey-graded and real-time SBAS (Satellite-Based Augmented system) corrected Trimble R2 GNSS system (manufactured by Trimble, U.S.A.) was mounted on top of the LiDAR sensor while maintaining the same central axis. The trajectory generated by the GNSS system was utilized to match the ZEB HORIZON-based trajectory facilitates converting the local coordinate system to the global coordinate reference system.

The scanned data was converted into LAS file format and geo-rectified using GeoSLAM Hub and Zeb Locate processor, proprietary software provided by the sensor manufacturer. The lidR package within the R environment was used for further processing, including noise filtering, ground point classification, DEM generation, and normalization. Specifically, the Isolated Voxels Filter (IVF) based noise algorithm, with a voxel resolution 5 and the maximum number of 'other points' set to 2, was applied for noise filtering. In order to generate a Digital Elevation Model (DEM) and normalize the LiDAR data, a prior classification of ground and non-ground points was achieved by using a Cloth Simulation Filter (CSF) (Zhang, et al. 2016). The CSF process inverted the data and sought the boundary plane that included all the points from above by simulating a cloth falling onto the surface. Following the classification of ground points, a DEM raster with 20 cm resolution was generated using the k-Nearest Neighbor Inverse-

Distance Weighting Algorithm (kNN-IDW) with the six nearest neighbors in the lidR package (Roussel, et al. 2020). The resulting DEM was then utilized for height-normalization of the LiDAR data, which involves eliminating the influence of terrain elevation from the measured LiDAR point cloud, thus focusing exclusively on the height variations of the vegetation or objects of interest. We subtracted the DEM derived from the LiDAR data from the corresponding LiDAR point cloud to perform the height normalization. This process effectively removed the contribution of terrain elevation, enabling us to analyze the vegetation heights accurately.

After normalization, LiDAR data was clipped to the plot level using the vector files defining the plot perimeter established in the field. Although handheld LiDAR systems ensure high point density, it is important to highlight that the major drawback of this system is an inconsistency in point density across the acquired point cloud (Chen, et al. 2019). This inconsistency arises due to the movement of the operator while scanning, which can result in variations in the distance between the LiDAR sensor and the scanned object (Yang, et al. 2013). For instance, when the operator moves closer to the object, the point density in that region increases, capturing more points and finer details. Conversely, when the operator moves farther away, the point density decreases, resulting in a sparser representation of the object. This variation in point density can pose challenges for subsequent analyses or applications that rely on consistent and uniform point distributions. For tasks such as individual tree detection, terrain modeling, or feature extraction, a consistent point density is desired to ensure accurate and reliable results. To overcome this limitation, we down-sampled the point density to 1 point/cm^3 using the point cloud decimation function in the lidR package (Roussel, et al. 2020) in the R

environment (Team 2013). First, this algorithm creates a grid with a specified resolution. Next, it filters the point cloud by randomly selecting points in each cell, thus producing point clouds with uniform densities throughout the coverage area. Finally, the algorithm computes the proportion of points or pulses to retain for each cell using the actual local density and the desired density. If the desired density exceeds the actual density, the algorithm returns an unchanged set of points.

Furthermore, detecting and removing the overstory pine tree stems is necessary to ensure that only understory competing vegetation was present in the plot. The most commonly employed method for individual tree detection in TLS data is the 2D-layer searching method (Aschoff, et al. 2004; Lindberg, et al. 2012). This technique involves identifying tree trunks based on point clustering or circle finding in a cross-sectional slice obtained at a specific height from the original point cloud data. However, this method proved ineffective in our case due to the presence of tall and dense shrubs and woody competing vegetation, which exhibited similar point clustering characteristics in the 2D slice within the study site. Another commonly employed approach for individual tree detection is local maxima filtering using a canopy height model, which is primarily designed to detect tree tops (Dalponte, et al. 2016; Silva, et al. 2016). However, this method falls short in accurately identifying stem locations when applied to terrestrial LiDAR data due to the distinct differences between tree top and trunk locations. Hence, in this study, we have developed a novel voxelized-density based approach explicitly tailored for TLS data and stem detection, which enables accurate identification of stem locations.

This method assumes that a tree stem can be represented as a continuous point trajectory extending from the ground to the top, characterized by the longest uninterrupted sequence

of points along the z-axis. The proposed methodology includes several key steps. Firstly, we implemented voxelization to achieve a uniform distribution of points and alleviate the impact of point clustering. This process involved dividing the point cloud data into small cubic regions, or voxels, each measuring 0.5 meters. Subsequently, we utilized the voxelized point cloud to generate a density raster with a resolution of 0.5 meters. This raster quantifies the density of voxels within each grid, thereby providing valuable information regarding the distribution of tree stems. Based on the principle that stem locations exhibit higher voxel density compared to crown and open areas, we applied a maximum local filter with a window size of 1.8 meters to detect tree stem locations. The selection of this specific window size was based on the spacing of the plantations and field observations, ensuring its appropriateness within the study context. Finally, we employed a percentile-based thresholding method based on density values to classify and identify tree stem locations. In our particular conditions, we determined that a density threshold ranging from the 47^{th} percentile to the 99^{th} percentile yielded satisfactory results. However, it is crucial to note that this threshold can be adjusted to accommodate different vegetation types and study settings. By employing this voxelized density-based approach, we successfully detected and located tree stems and removed them from the point clouds, focusing specifically on the understory layers and competing vegetation within the designated plot (Figure 3.5). Additionally, we filtered out all LiDAR point clouds above three meters, which we consider to be the maximum height for understory layers based on the field observation (Figure 3.5).

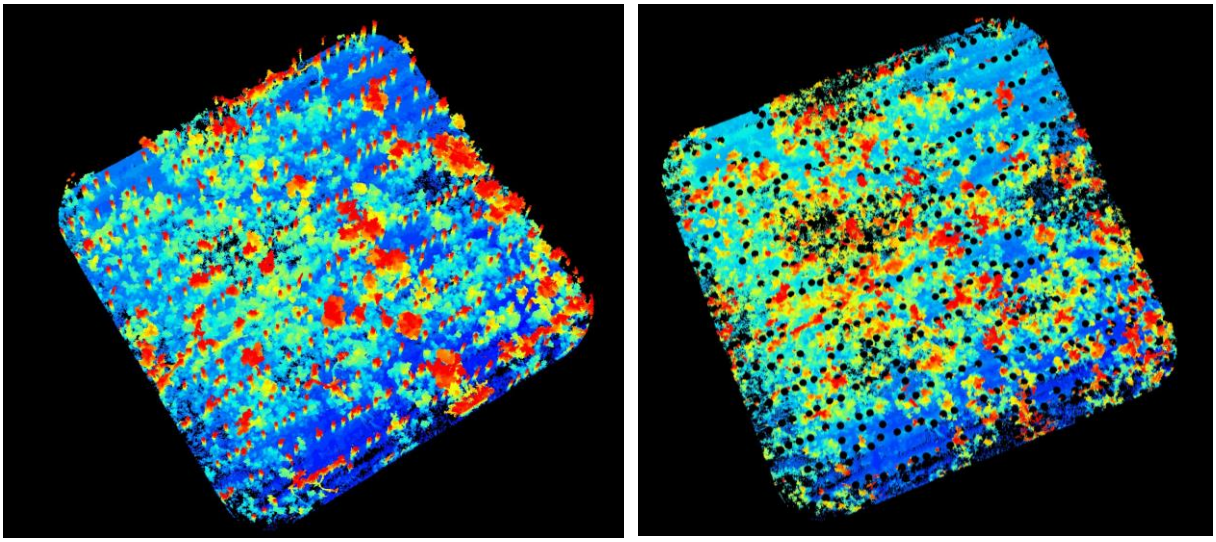

Figure 3.5: Figure Illustrating the LiDAR point cloud (filtered under 3 meters) before (left) and after (right) the removal of tree stems.

This method demonstrates robust performance with high-density LiDAR data, particularly with TLS. However, it exhibits a significant limitation related to its sensitivity to point cloud density. Hence, cautious implementation is advised when using this method with low-density LiDAR data. The accuracy of individual tree detection heavily relies on the continuous representation of tree stem and canopy in the point clouds. To mitigate this limitations associated with low-density point clouds, it is recommended to employ larger voxel sizes, at least ten times greater than the point density of the data sets, and to use higher height percentile thresholds, particularly for broadleaf species. By adhering to these recommendations, the method can be better optimized to address various LiDAR data scenarios, ensuring accurate and reliable individual tree detection results.

3.2.4 Variable extraction from LiDAR data

Following preprocessing, downscaling, and stem removal, we clipped the LiDAR data using the subplot boundary file. Next, we utilized the clipped $1m^2$ LiDAR point clouds to generate various metrics for modeling understory vegetation. The variables generated for this study include seven statistical variables, nine percentiles variables, two variables representing understory cover, and nine variables representing volume. Volume variables are based on three methods: voxelized, alpha-hall fitting, and mean height and understory cover based. A detailed description of the subplot-level metrics derived from TLS is presented in Table 3.2, and summary statistics are in Table 3.3.

Table 3.2: Description of sub-plot level metrics derived from TLS data

Abbreviation	**Description**
Max ht, Mean ht, Median, Median, Standard Dev, Skewness, Kurtosis	Maximum, Mean, Median, Standard deviation, Skewness, and Kurtosis values of the echo heights within a sub-plot
P90, P80, P70, P60, P50, P40, P50, P40, P30, P20, P10	Echo height distribution percentiles (90th, 80th, 70th, ……, 30th, 20th, 10th) within a sub-plot
Cover_10cm, cover_20cm	Understory cover per square meter generated from a $10cm^2$ and $20cm^2$ grid
UcMh_Vol_10cm, UcMh_20cm	Product of subplot level mean height and understory cover at 10 cm^2 and 20 cm^2 grid.
Vol Voxel_5cm, Vol Voxel_10cm, Vol Voxel_20cm	The volume of the sub-plot generated by fitting a voxel of size $5m^3$, $10m^3$, and $20cm^3$
Vol α-Hull_10cm, Vol α-Hull_20cm, Vol α-Hull_30cm, Vol α-Hull_40cm, Vol α-Hull_50cm	The volume of the sub-plot generated by fitting an α-Hull with α values of 0.1, 0.2, 0.3, 0.4, 0.5

Table 3.3: Summary statistics of understory biomass and twenty-seven subplot level TLS metrics

Variables	Mean	Maximum	Minimum	Median	Standard Deviation
Understory Biomass (grams)	856.11	2878.00	8.13	701.50	627.52
Maximum height (m)	1.91	3.00	0.28	1.81	0.85
Mean height (m)	0.97	2.47	0.06	0.85	0.60
Understory $cover_{10cm}$ (m^2)	0.79	1.00	0.06	0.89	0.24
Understory $cover_{20cm}$ (m^2)	0.91	1.04	0.16	1.00	0.17
Volume mean $height_{10cm}$ (m^3)	0.81	2.45	0.00	0.72	0.56
Volume mean $height_{20cm}$ (m^3)	0.91	2.47	0.01	0.82	0.59
Volume $Voxel_{10cm}$ (m^3)	0.54	2.14	0.04	0.44	0.42
Volume $Voxel_{20cm}$ (m^3)	1.03	2.87	0.15	0.86	0.59
90^{th} Percentile	1.51	2.94	0.07	1.41	0.82
80^{th} Percentile	1.38	2.89	0.06	1.20	0.80
70^{th} Percentile	1.25	2.82	0.05	1.04	0.76
60^{th} Percentile	1.13	2.74	0.05	0.95	0.74
50^{th} Percentile	0.98	2.67	0.05	0.82	0.68
40^{th} Percentile	0.85	2.52	0.05	0.72	0.64
30^{th} Percentile	0.71	2.40	0.04	0.59	0.57
20^{th} Percentile	0.55	2.26	0.04	0.45	0.49
10^{th} Percentile	0.39	2.15	0.04	0.27	0.43
Skewness	0.05	4.98	-2.68	-0.11	1.03
Kurtosis	3.58	32.12	1.17	2.49	3.89
Standard Deviation	0.45	1.23	0.03	0.38	0.26
Volume α-$Hull_{10cm}$ (m^3)	0.32	1.64	0.00	0.22	0.32
Volume α-$Hull_{20cm}$ (m^3)	0.58	2.26	0.01	0.42	0.48
Volume α-$Hull_{30cm}$ (m^3)	0.75	2.51	0.02	0.62	0.56
Volume α-$Hull_{40cm}$ (m^3)	0.85	2.60	0.02	0.75	0.59
Volume α-$Hull_{50cm}$ (m^3)	0.92	2.63	0.04	0.85	0.60

Median	0.98	2.67	0.05	0.82	0.68
Minimum height	0.05	0.47	0.04	0.04	0.05

3.2.4.1. Mean height and understory cover extraction method

To calculate the mean height of the subplot, we averaged the height values of all point clouds that fell within one square meter subplot and were less than 3 meters above ground level. To assess the understory cover, we placed a two-dimensional grid of 10 cm resolution over the subplot area and counted the number of grids containing point clouds. We then calculated the ratio between the number of grid cells with point clouds and the total number of grid cells to determine the understory cover (Fig. 3.6). Similarly, to measure the understory cover ratio at a 20 cm resolution, we applied a two-dimensional grid of a 20 cm resolution.

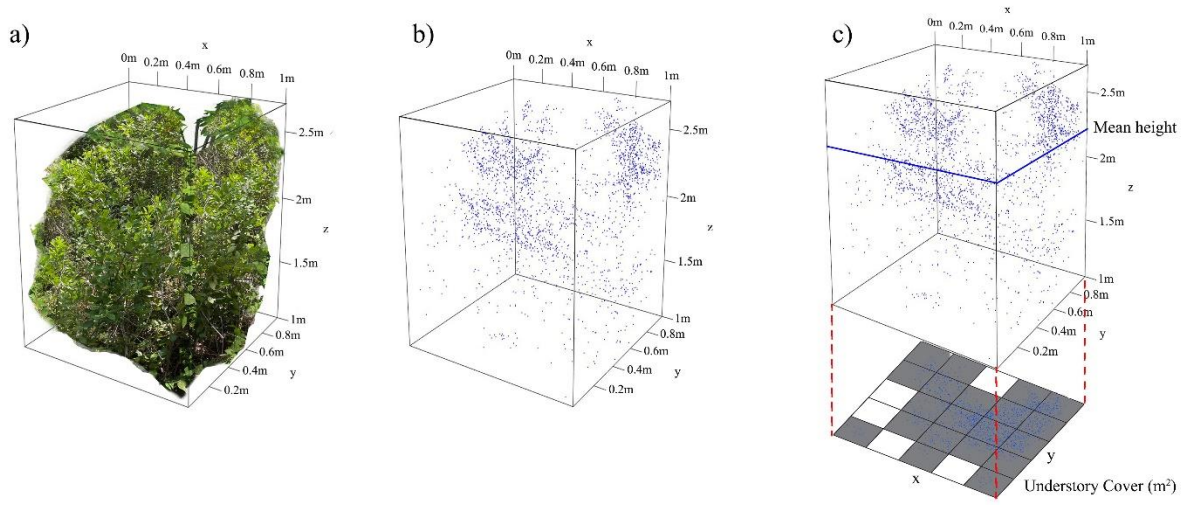

Figure 3.6: Mean height and understory cover-based volume extraction method. (a) Destructively sampled understory vegetation at ground level. (b) Blue dots represent the clipped point clouds for the subplot. (c) The mean height of the point clouds in meters and understory cover estimates based on 0.2 meters grid counting.

3.2.4.2. Volume estimation using the voxel-based method

Voxel fitting is a popular method for estimating the volume of objects using LiDAR data (Cooper, et al. 2017; Hancock, et al. 2017). This approach involves dividing the point cloud data into a three-dimensional grid of equally sized cubic cells or 'voxels' (Fig. 3.7). The accuracy of this method depends on the resolution of the voxel grid and the suitability of the chosen model (Béland, et al. 2014). Considering the accuracy dependency on voxel resolution, we fitted a voxel at three different resolutions (5 cm^3, 10 $cm^{3,}$ and 20 cm^3) and calculated the subplot volume for each resolution. We used the LidR package in the R environment to fit voxels and estimate volume.

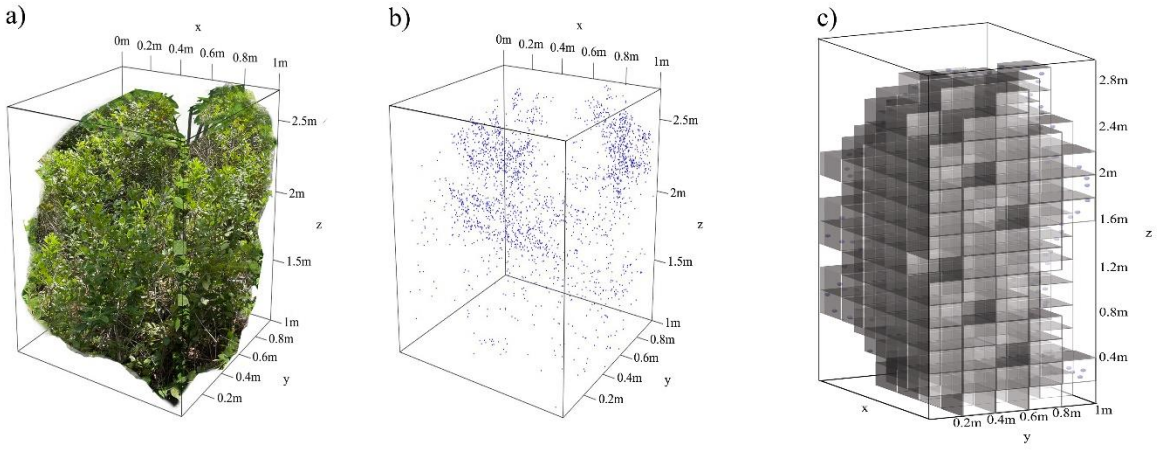

Figure 3.7: Voxel-based volume extraction method. (a) Destructively sampled understory vegetation at ground level (b) Blue dots represent the clipped point clouds for the subplot (c) Fitted voxel of resolution 0.2 meters in the point clouds.

3.2.4.3 Volume estimation using three-dimensional α-Hull fitting

This technique involves creating a three-dimensional hull that encompasses all the LiDAR points and varies in its level of detail according to a chosen alpha value (Fig 3.8). The parameter alpha controls the level of detail of the hull, and finding the best alpha

value is crucial for obtaining accurate volume estimates (Wallace 2013). This method is particularly useful for irregularly shaped objects such as the understory vegetation and can provide accurate volume estimates when the alpha value is chosen carefully. To ensure precision in our study, we fitted the model using five alpha values (0.1, 0.2, 0.3, 0.4, and 0.5), creating five different variables that we evaluated with linear regression. We used the alphashape3d package(Lafarge, et al. 2016) in R to generate the alpha hulls covering the subplots and calculate the volume of it.

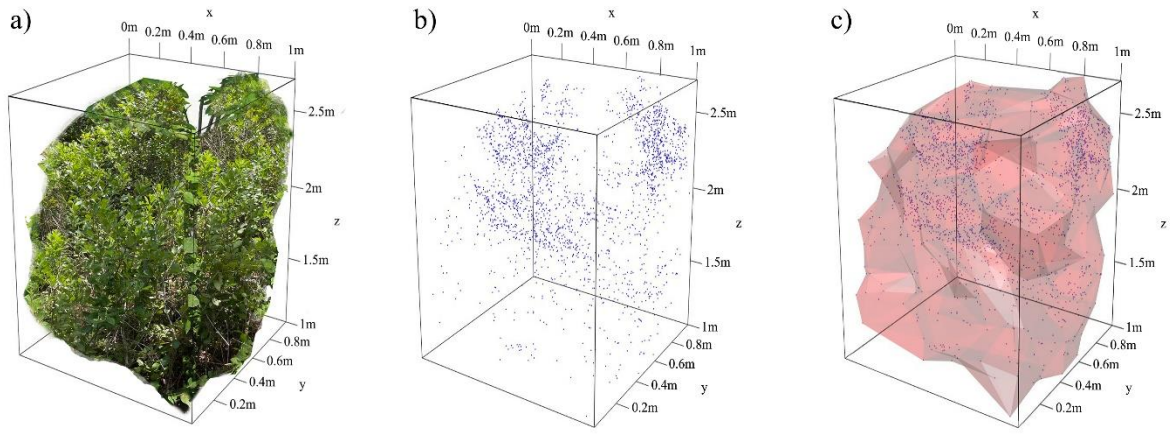

Figure 3.8 Three-dimensional α-Hull fitting-based volume extraction method. (a) Destructively sampled understory vegetation at ground level. (b) Blue dots represent the clipped point clouds for the subplot. (c) The structure in red is the fitted α-Hull with an alpha value of 0.2 on the point clouds.

3.2.5 Statistical analysis

3.2.5.1 Model validation

In this study, we evaluated the accuracy of models with different explanatory variables using the k-fold cross-validation method. This approach involves randomly dividing a set of observations into k subsets, or folds, of equal size. The first fold is then utilized as a validation set, while the model is fitted on the remaining k-1 folds (Kuhn 2015). To

implement this method, we used the caret package (Kuhn 2015) within the R environment, which allowed us to split the datasets into seven folds. Subsequently, the model was fitted on six of the seven folds, while the first fold served as the validation set. This process was repeated seven times with distinct validation subsets. To evaluate the performance of the regression models using this validation method, we employed various metrics, including the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Akaike Information Criterion (AIC). MBE was determined by the following equation.

$$MBE\ (Bias) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) \qquad (3.1)$$

Where Y is the observed biomass value, $\hat{Y}$ is predicted biomass value at the i subplot, and n is the number of sampling locations.

3.2.5.2 Selection of Volume Extraction Method

To determine the most suitable method for estimating volume and to identify the optimal resolution, we utilized a simple least-square regression approach. In this approach, we used the volume variable derived from each method as the single explanatory variable to model understory biomass. We modeled all three methods at various simulated resolutions and evaluated their performance through k-fold cross-validation. The method that exhibited the best performance was selected for further modeling in combination with other statistical TLS metrics.

3.2.5.3 Modeling Biomass with Selected Volume Metrics and TLS Metrics

For the final model, we employed the Adaptive LASSO (Least Absolute Shrinkage and Selection Operator) method (Zou 2006). One major challenge in linear regression with

high-dimensional data is multicollinearity, which refers to strong correlations among predictor variables, leading to unstable and unreliable model estimates (Holcomb 1999). In our study, where we incorporate various TLS metrics as predictors, multicollinearity can arise due to their inherent relationships since they are all LiDAR echo height-dependent variables. Adaptive LASSO solves these challenges by automatically selecting variables while simultaneously addressing multicollinearity. It had outperformed other methods, such as least square regression and random forest for variable selection and modeling, specifically in LiDAR data analysis (Adhikari, et al. 2023). Furthermore, the Adaptive LASSO overcomes the limitations of traditional LASSO by providing oracle properties, which means it can select the true meaningful variables with higher probability, even in the presence of correlated predictors. In our study, the application of Adaptive LASSO allowed us to address multicollinearity challenges associated with incorporating highly correlated TLS metrics as predictors. By automatically selecting the most relevant variables, Adaptive LASSO provided a more robust and accurate model for biomass estimation, which we further validated using a 7-fold cross-validation method.

3.3. Results

3.3.1 Summary statistics

In this study, we used field-measured understory biomass as a dependent variable. We tested twenty-seven lidar metrics as independent variables generated from 100 sample subplots of $1m^2$ from 60 sample plots. Within those 100 subplots, the understory vegetation biomass varied between 81.3 kilograms to 28,780 kilograms, with a mean of 8561.1 kilograms per hectare. Similarly, the maximum mean height of the LiDAR points within a subplot was 2.47 meters, and the minimum was 0.26 meters. The estimated volume of clipped LiDAR point clouds representing the understory vegetation subplot yielded varying results when three different methods were employed at varying resolutions. When using the voxel-based method, the average volume was determined to be 0.54 m^3 with a minimum of 0.04 m^3 and a maximum of 2.14 m^3 while keeping the voxel of 10 cm^3; and the mean volume was 1.03 m^3 with a minimum of 0.15 m^3 and a maximum of 2.87 m^3 when the voxel resolution was increased to 20 cm^3. For mean height and understory cover-based volume, the mean volume increased from 0.79 m^3 (min: 0.00m^3, max: 2.45 m^3) to 0.91 m^3(min: 0.15 m^3, max: 2.87 m^3) when the resolution changed from 10 cm^2 to 20 cm^2.

Similarly, with the α-Hull method of volume estimation, the mean volume increased from 0.32 to 0.58 and 0.92 when fitting the hull using α values of 0.1, 0.2, and 0.5, respectively. Figure 3.9 is a boxplot showing the statistical distribution of the understory biomass and six major variables. A detailed statistical summary of the field-measured understory biomass and LiDAR metrics is provided in Table 3.3.

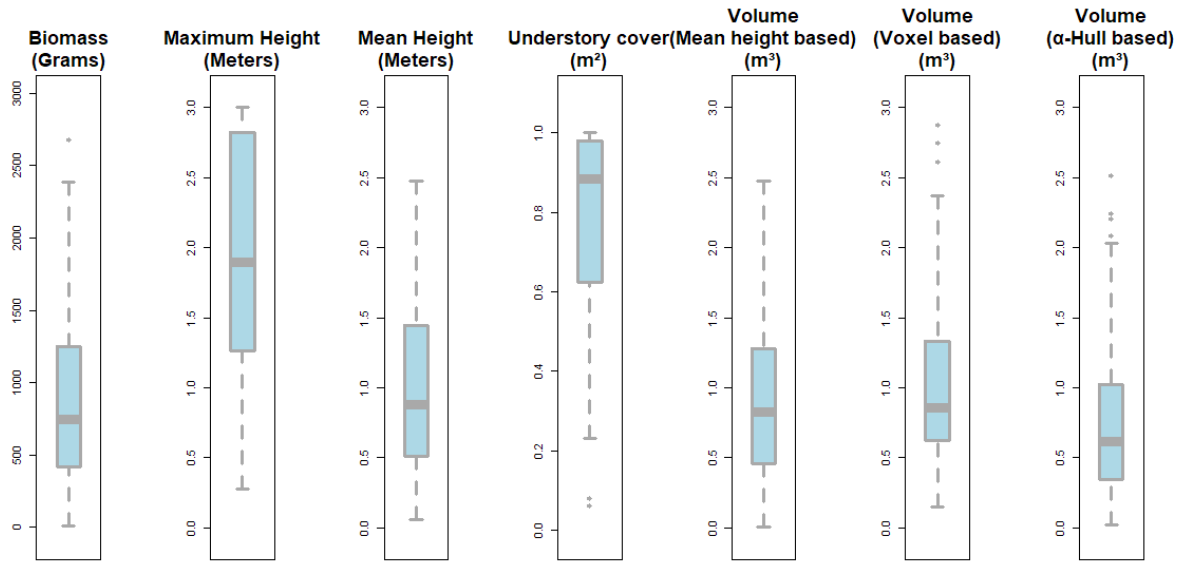

Figure 3.9: Boxplot of the understory biomass and six primary LiDAR metrics, displaying the Median, quartiles, and outliers, providing a concise visualization of central tendency, dispersion, and Skewness of the selected variables.

3.3.2 Variable correlation

The pearson correlation between understory vegetation biomass and the major LiDAR metrics is presented in Figure 3.10. Volume based on mean height and understory cover at a 10 cm grid resolution showed the highest correlation coefficient (0.89) with the understory biomass. The following four variables with high correlation coefficients were mean height-based volume at 20 cm, 40th percentile, mean height, and Median of all returns within the subplot with coefficients of 0.86, 0.82, 0.81, and 0.80, respectively. Moreover, this study evaluated the volume correlation based on different methods at varied resolutions. The voxel-based method achieved the highest correlation coefficient of 0.70 when keeping the voxel resolution of 20 cm^3. Additionally, the study investigated the effect of changing the alpha value in the alpha hall-based volume estimation method. The highest correlation coefficient of 0.74 was achieved when the alpha value was kept at

0.5. Overall, the results suggest that volume based on mean height and understory cover at 10 cm resolution is the most promising variable for estimating understory vegetation biomass using LiDAR metrics for this dataset and this study.

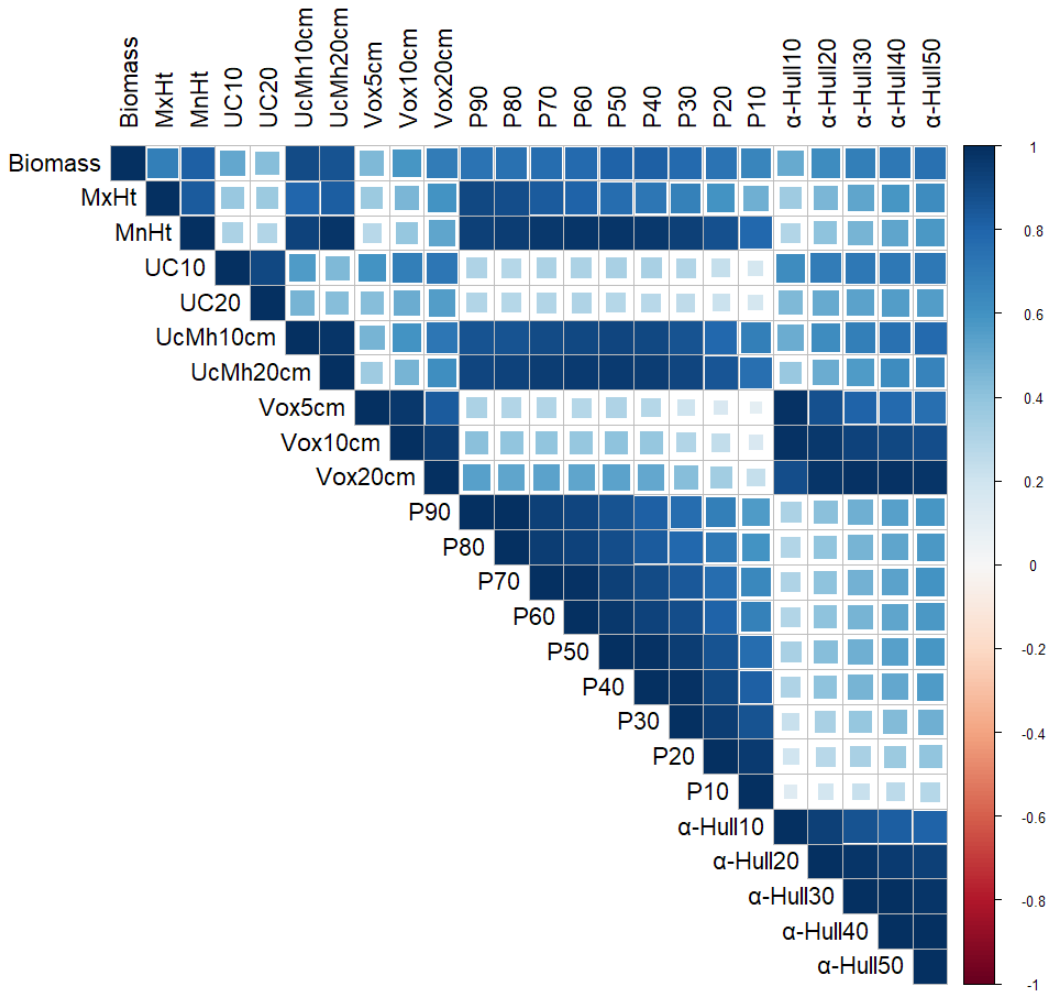

Figure 3.10: Pearson correlation coefficients plot between understory biomass and with all derived volume-based metrics and highly correlating statistical TLS metrics. Darker blue and bigger squares represent higher correlation, and lighter blue and smaller squares represent lower correlation.

3.3.3 Comparing 3D volume estimation Methods for Biomass modeling

The 3D volume estimation method comparision result in Table 3.4 shows that among the three-volume estimation methods we tested at different resolutions, volume based on understory cover and mean height at 10 cm resolution is the most significant contributor or explanatory variable for understory biomass estimation. With this variable alone, the model achieved an Adjusted R^2 of 0.79 and RMSE of 288.0 grams on cross-validation datasets, which the highest compared to any other models with a single explanatory variable. Regarding the best resolution for variable generation using TLS, we found that for the voxel-based method, a voxel size of 20 cm^3 explained biomass better than 10 cm^3 (R^2_{adj} of 0.33 and 0.47, respectively). Similarly, in the α-Hull-based method, a higher α value of 0.5 was suitable for biomass estimation, resulting in an R^2_{adj} of 0.55, compared to R^2_{adj} of 0.23 when α was 0.1.

Table 3.4: Least Square regression models summary for the understory biomass prediction with TLS-based variables. Calibration denotes the performance metrics when fitting with the full data set, and CV depicts the mean value after five-fold cross-validation

y	**x**	**Performance metrics for the linear regression**					
		R^2_{adj}		**RMSE (g)**		**Bias**	**AIC**
		Cal.	**CV**	**Cal.**	**CV**		
Understory Biomass	UcMh_Vol_$_{10cm}$	**0.80**	**0.79**	**276.5**	**288.0**	**0.00**	**1414.2**
	UcMh_Vol_$_{20cm}$	0.74	0.73	318.7	315.6	0.00	**1440.6**
	Voxel-based vol_{10cm}	0.34	0.33	507.1	508.4	0.00	**1533.5**
	Voxel-based vol_{20cm}	0.48	0.47	447.9	448.6	0.00	**1508.7**
	α-Hull-based vol_{10cm}	0.24	0.23	539.3	538.7	0.00	**1545.9**
	α-Hull-based vol_{20cm}	0.42	0.38	479.8	482.7	0.00	**1525.3**
	α-Hull-based vol_{30cm}	0.47	0.45	453.2	456.2	0.00	**1513.1**

y	x	Performance metrics for the linear regression					
		R^2_{adj}		RMSE (g)		Bias	AIC
		Cal.	CV	Cal.	CV		
	α-Hull-based vol_{40cm}	0.52	0.51	434.1	434.6	0.00	**1502.4**
	α-Hull-based vol_{50cm}	0.57	0.55	418	413.1	0.00	**1494.9**

3.3.4. Biomass estimation model

After selecting the mean height-based volume estimation method at a 10 cm resolution as the dependent variable for explaining the volume of the understory, we performed an Adaptive LASSO regression analysis by incorporating this variable along with other TLS percentile and statistical metrics listed in Table 3.3. The Adaptive LASSO model identified two out of the sixteen variables as significant contributors to the model (equation 3.2).

$$Y = 56.57 + 884.51 \times x_1 + 169.25 \times x_2 \quad (3.2)$$

Where Y is understory biomass, x_1 is the product of the mean height and understory cover, and x_2 is the 20th percentile of echo height.

This result indicates that including the 20th percentile in the mean height-based model led to a substantial improvement in model performance compared to single-variable models. The Adaptive LASSO model exhibited favorable performance measures in both calibration and cross-validation datasets. In the calibration dataset, the model achieved an Adjusted R^2 of 0.84 and an RMSE of 233 grams. In the cross-validation dataset with a 7-fold approach, the model yielded an Adjusted R^2 of 0.8 and an RMSE of 235 grams. Notably, the RMSE showed an 18 percent improvement compared to the best single variable model in Table 3.4. These results underscore the enhanced predictive accuracy of the Adaptive LASSO model, suggesting its effectiveness in capturing the relationships

between the selected variables and the biomass of the understory. Furthermore, Figure 3.11 depicts the observed versus predicted plot and residual plot, providing insights into the model performance. The plot reveals a cross-validation positive bias of 0.44, suggesting a slight tendency for the model to overestimate the understory biomass when it falls below 500 grams and underestimate it when it exceeds 1500 grams. However, the model exhibits reasonably accurate predictions within the 500 to 1500 grams range. Additionally, the residual plot showcases a random distribution of residuals, indicating that the model adequately captures the variability in the data with no apparent patterns or systematic deviations. Moreover, using the final model, figure 3.12 shows the understory prediction map for sample plots with different extreme understory biomass conditions.

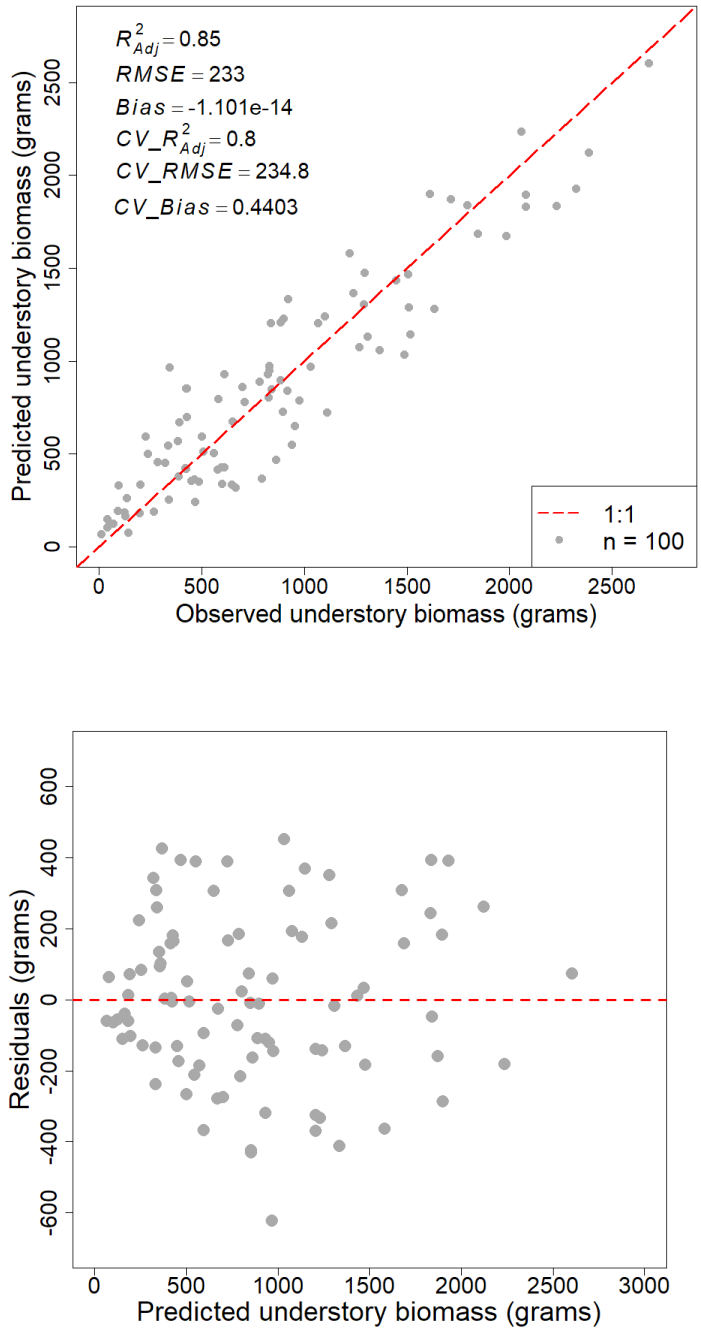

Figure 3.11: The plot on the top displays the relationship between predicted and observed understory biomass, and the plot on the bottom is the residual vs. predicted plot.

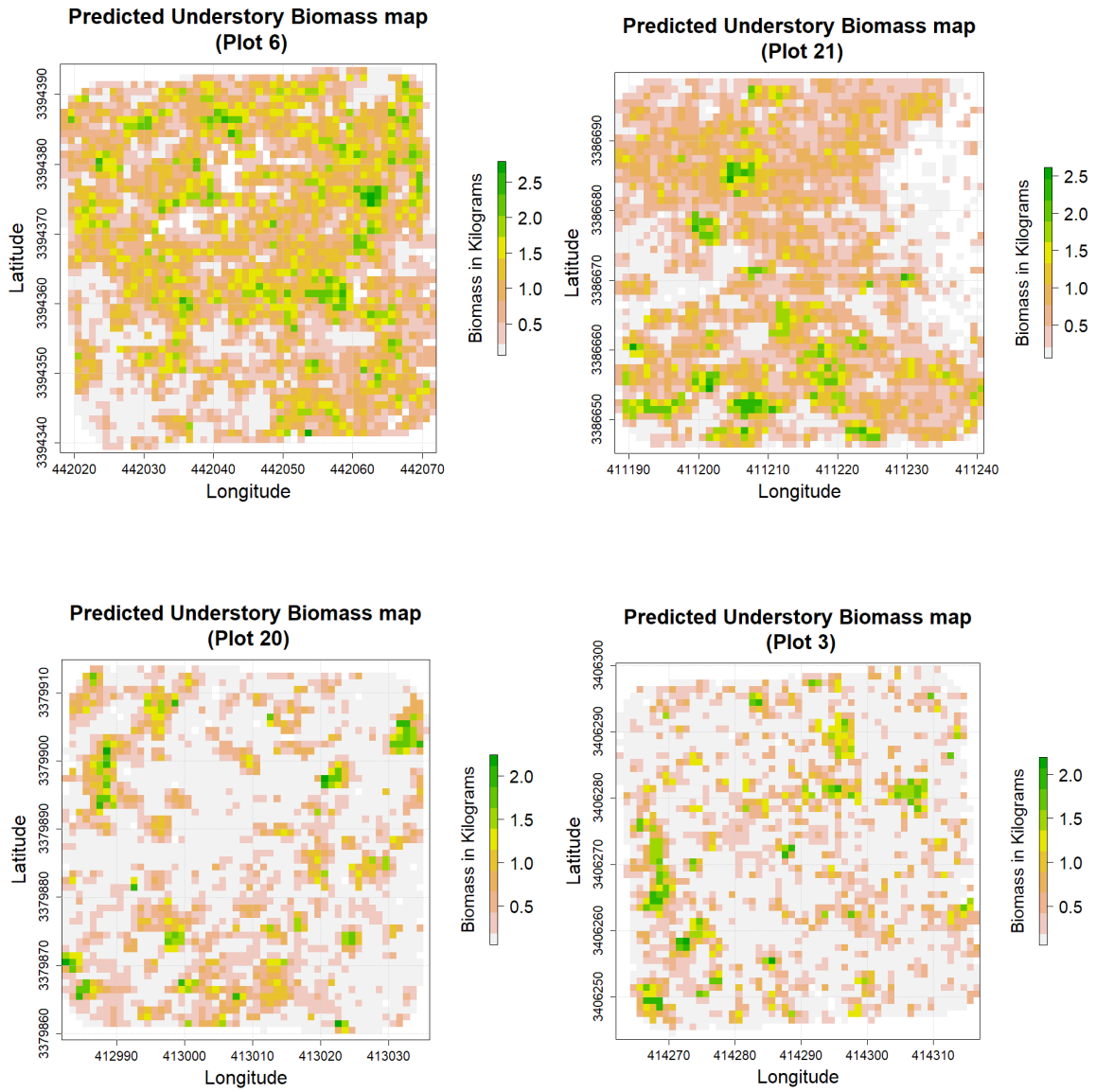

Figure 3.12: Understory biomass predicted maps for 40m x 40m sample plots, showing four different abundance of understory biomass amounts at 1m resolution. Plot no is the assigned plot number out of 60 plots.

3.4. Discussion

This study presents an efficient understory biomass prediction model based on two TLS-derived variables: the 20th percentile of echo height and a 3D volume metric based on mean height and understory cover. Our model enables precise quantification and prediction of understory biomass with a high level of accuracy (Adj. R^2 = 0.80, RMSE=234.8 g/m^2) across plantation stands, even at a resolution of one meter for biomass cover estimation. By adopting this non-destructive approach for assessing understory vegetation, we address the limitations posed by the high cost associated with field-based measurement and the complex nature of the understory vegetation in the coastal plains of the southeastern United States. The understory vegetation in pine plantations in this region is predominantly characterized by evergreen shrubs (with scattered hardwood trees), which have initially proven challenging to assess accurately using traditional field-based methods as well as remote sensing techniques such as Landsat imageries and airborne LiDAR (Campbell, et al. 2018; Peduzzi, et al. 2010). However, the TLS-based approach had been shown to surpass field-based methods in accuracy, speed, and precision, with data collection taking only 10 to 30 minutes, including initial sensor calibration and data logging time after the scan, depending on the abundance of understory vegetation (Li, et al. 2021). This efficiency is attributable to the scan range of up to 100 meters and the high point density of TLS, which enables precise predictions of subtle variations in understory vegetation abundance down to one-meter resolution at the stand level. Such capabilities not only benefit plantation managers for monitoring competition between overstory and understory but also hold the potential for supporting ecological studies, including understory carbon assessments and wildlife habitat studies.

One major issue of implementing laser scanning for vegetation attributes and biomass estimation is generating and selecting the most meaningful variables because hundreds of LiDAR metrics can be generated using LiDAR data (Adhikari, et al. 2023). Most studies have used LiDAR metrics summarizing statistical distributions, percentiles, and density ratios as explanatory variables (Campbell, et al. 2018; Ferraz, et al. 2016). However, few suggested better predictability with volume-based metrics, such as voxel fitting (Cooper, et al. 2017; Venier, et al. 2019), 3D convex-hall fitting (Olsoy, et al. 2014), and mean height-based volume (Li, et al. 2021).

To develop a more robust and generalizable prediction model, we incorporated and carefully evaluated both types of metrics in this study. Our results revealed that mean height-based volume estimation method exhibited a stronger correlation with understory biomass, achieving an R^2 of 0.79 and an RMSE of 288.0 grams, compared to the alpha-hull-based volume estimation methods and the method based on voxel count, which achieved R^2 values of 0.55 and 0.47, and RMSE values of 413.1 and 448.6 grams/m^2, respectively when used as a single variable in the model. Although the mean height-based volume metric alone demonstrated a high correlation, we observed few negative biomass predictions in areas with low biomass and negative echo heights. However, this issue was effectively addressed when incorporating the mean height-based volume metric with other standard metrics. Our findings highlight the significance of incorporating volume-based metrics alongside other standard metrics, resulting in a more stable and accurate estimation model for understory vegetation biomass. This comprehensive approach enhances the model predictive performance and contributes to a more reliable assessment of understory biomass across diverse levels of understory abundance.

This study shows a higher correlation between biomass and TLS-derived 3D volume metrics than other standard metrics. However, the resolution used to generate these volume-based metrics emerges as a crucial parameter that significantly influences the overall predictivity of the variable. In our study, we simulated and tested several resolutions for all three methods and found that each method exhibited sensitivity to the resolution employed. For the mean height-based volume metrics, we evaluated two resolutions,10 cm, and 20 cm, to calculate the understory cover. With that, we observed that the accuracy increased with a higher resolution, indicating that a finer level of detail in the data contributes to a more precise estimation of understory biomass for this particular method.

Conversely, in the case of voxel-based estimation, we tested two voxel resolutions, 10 cm, and 20 cm. However, the accuracy of voxel-based estimation improved when using a coarser resolution of 20 cm. Likewise, the alpha-hull-based method also displayed sensitivity to resolution, but similarly to the voxel-based method. We examined five different alpha values: 0.1, 0.2, 0.3, 0.4, and 0.5, to evaluate their impact on accuracy and found that the higher alpha value (e.g., 0.5) improved accuracy in estimating understory biomass for this method. The detailed results of these resolution tests are presented in Table 3.4, which provides insights into the performance of each method under varying resolutions. These findings underscore the importance of carefully selecting an appropriate resolution for each specific volume-based method to achieve the most accurate understory biomass predictions. Moreover, our results suggest that mean height-based methods are better suited for high-density LiDAR data. In contrast, voxel-based or alpha-hull-based methods may be more suitable for predicting understory biomass for

lower-density data. Nevertheless, further investigations are necessary to validate this observation conclusively.

Li et al. (2021) conducted a comparable study in a temperate mixed broadleaf conifer forest. They used volume-based metrics derived from terrestrial LiDAR to estimate understory herbs and shrub biomass, similar to our study. Their results showed an R^2 value of 0.69 and 31.6% RMSE for shrubs, which aligns with our findings of an R^2 value of 0.8 and 27.15% RMSE. However, their study suggests a logarithmic relationship, while our results indicate a more linear relationship between understory and TSL-based metrics. One possible explanation for this discrepancy could be the high variability in our sample data, which spans a wide range of 8.13 grams per square meter to 2878 grams per square meter, in contrast to the minimum biomass of 57.21 grams per square meter and a maximum of 310.66 grams per square meter in their study. Moreover, including the 20th percentile in height mean and understory cover-based volume model substantially improved model accuracy. This variable plays a crucial role in capturing the density of vegetation near the ground and effectively complements the other variables included in the model.

TLS has a notable limitation in its inconsistent point density, which is influenced by the distance to the scanned object and the movement of the operator. This leads to variations in the distance between the sensor and the object being scanned, resulting in areas with high and low point densities (Chen, et al. 2019; Yang, et al. 2013). Since TLS is a handheld sensor, its speed is dependent on the walking speed of the operator. Ensuring consistent point density would require maintaining a consistent walking speed, which is impractical, particularly in densely vegetated forest areas. This variation in point cloud

density significantly impacts the accuracy of such LiDAR metric-depended models, particularly shape-based volume metrics like voxelization and convex hull fitting. In order to investigate the impact of varying point cloud density, different density levels and resolutions were simulated in our study. The results revealed that the correlation between dry biomass and volumes derived from the mean height and understory cover approach exhibited a relatively stable relationship compared to other shape-based methods (Figure 6). This suggests that the volume estimation method utilizing mean height and understory cover is less susceptible to variations in density. Consequently, the mean height and understory cover approach offers a more robust and reliable means of estimating volumes in comparison to alternative shape-based methods. This research contributes to understanding the limitations and potential solutions for attribute estimation in forests with high occlusion using TLS.

The study also has some limitations. First, the model was developed and evaluated in a single study area. Therefore, testing the model in other study areas is important to ensure that it is generalizable. Second, the model was only trained with data from a loblolly pine plantation; therefore, testing it in other forest types and different tree species compositions will allow model adjustment and increase in robustness. Despite these limitations, the results of this study provide a valuable tool for predicting understory biomass in dense and evergreen understory vegetation in the Southeastern US.

3.5. Conclusion

In conclusion, this study presents an efficient TLS-derived metrics-based model capable of predicting understory biomass up to one-meter resolution, thereby addressing the limitations posed by cost and vegetation complexity in assessing understory vegetation within the coastal forests of the southeastern United States. Volume-based metrics, particularly mean height-based volume in combination with the 20^{th} percentile of echo height, exhibit a robust and superior correlation with understory biomass compared to other metrics, achieving an R^2 of 0.80 and RMSE of 234.8 grams.

However, a key limitation of this method is its applicability to small areas or stand-level assessments due to the requirement for ground-level LiDAR data collection. Future work may involve upscaling the approach to UAV-based or aerial LiDAR, extending its applicability to larger areas and landscape-level studies.

3.6 Funding

The Warnell School of Forestry and Natural Resources at the University of Georgia provided funding support for the author's graduate assistantship and the Plantation Management Research Cooperative for field support.

3.7 Acknowledgments

The authors would like to thank Rayonier Inc for their assistance with field planning and support.

3.8 Abbreviations

AIC = Akaike Information Criterion

ALASSO = Adaptive Least Absolute Shrinkage and Selection Operator

CSF = Cloth Simulation Filter

DEM = Digital Elevation Model

GNSS = Global Navigation Satellite System

IMU = Inertial Measurement Unit

IVF = Inverse Variable Filter

kNN-IDW = k-Nearest Neighbors with Inverse Distance Weighting

LASSO = Least Absolute Shrinkage and Selection Operator

LiDAR = Light Detection and Ranging

MBE = Mean Bias Error

MH = Mean echo height

MODIS = Moderate Resolution Imaging Spectroradiometer

PLS = Personal Laser Scanner

R^2 = Coefficient of determination

Adj. R^2 = Adjusted coefficient of determination

RMSE = Root Mean Square Error

RS = Remote Sensing

SBAS = Satellite-Based Augmentation System

TLS = Terrestrial Laser Scanning

UC = Understory Cover

UAV = Uncrewed Aerial Vehicle

3.9 References

1. Adegbidi, H., Jokela, E., Comerford, N. and Barros, N. 2002 Biomass development for intensively managed loblolly pine plantations growing on Spodosols in the southeastern USA. *Forest Ecology and Management*, **167** (1-3), 91-102.

2. Adhikari, A., Montes, C.R. and Peduzzi, A. 2023 A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics. *Remote Sensing*, **15** (5), 1284.

3. Allen, H.L., Fox, T.R. and Campbell, R.G. 2005 What is ahead for intensive pine plantation silviculture in the South? *Southern Journal of Applied Forestry*, **29** (2), 62-69.

4. Aschoff, T. and Spiecker, H. 2004 Algorithms for the automatic detection of trees in laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **36** (Part 8), W2.

5. Barker, M.G. and Pinard, M.A. 2001 Forest canopy research: sampling problems, and some solutions. In *Tropical forest canopies: ecology and management*, Springer, pp. 23-38.

6. Béland, M., Baldocchi, D.D., Widlowski, J.-L., Fournier, R.A. and Verstraete, M.M. 2014 On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR. *Agricultural and Forest Meteorology*, **184**, 82-97.

7. Bonham, C.D. 2013 *Measurements for terrestrial vegetation*. John Wiley & Sons.

8. Campbell, M.J., Dennison, P.E., Hudak, A.T., Parham, L.M. and Butler, B.W. 2018 Quantifying understory vegetation density using small-footprint airborne lidar. *Remote sensing of environment*, **215**, 330-342.

9. Chen, S., Liu, H., Feng, Z., Shen, C. and Chen, P. 2019 Applicability of personal laser scanning in forestry inventory. *PLOS ONE*, **14** (2), e0211392.

10. Clinton, N.E., Potter, C., Crabtree, B., Genovese, V., Gross, P. and Gong, P. 2010 Remote Sensing–Based Time-Series Analysis of Cheatgrass (Bromus tectorum L.) Phenology. *Journal of Environmental Quality*, **39** (3), 955-963.

11. Cooper, S.D., Roy, D.P., Schaaf, C.B. and Paynter, I. 2017 Examination of the Potential of Terrestrial Laser Scanning and Structure-from-Motion Photogrammetry for Rapid Nondestructive Field Measurement of Grass Biomass. *Remote Sensing*, **9** (6), 531.

12. Dalponte, M. and Coomes, D.A. 2016 Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in ecology and evolution*, **7** (10), 1236-1245.

13. Eskelson, B.N., Madsen, L., Hagar, J.C. and Temesgen, H. 2011 Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, **57** (3), 212-221.

14. Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Gonçalves, G., Silva, C.A. *et al.* 2016 Airborne Lidar Estimation of Aboveground Forest Biomass in the Absence of Field Inventory. *Remote Sensing*, **8** (8), 653.

15. Fortson, J.C., Shiver, B.D. and Shackelford, L. 1996 Removal of competing vegetation from established loblolly pine plantations increases growth on Piedmont and Upper Coastal Plain sites. *Southern Journal of Applied Forestry*, **20** (4), 188-193.

16. Fox, T.R., Jokela, E.J. and Allen, H.L. 2007 The Development of Pine Plantation Silviculture in the Southern United States. *Journal of Forestry*, **105** (7), 337-347.

17. Gollob, C., Ritter, T., Wassermann, C. and Nothdurft, A. 2019 Influence of Scanner Position and Plot Size on the Accuracy of Tree Detection and Diameter Estimation Using Terrestrial Laser Scanning on Forest Inventory Plots. *Remote Sensing*, **11** (13), 1602.

18. Hancock, S., Anderson, K., Disney, M. and Gaston, K.J. 2017 Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. *Remote Sensing of Environment*, **188**, 37-50.

19. Holcomb, J.P. 1999 Applied Regression Analysis: A Research Tool. *The American Statistician*, **53** (2), 170.

20. Johnson, K.D., Domke, G.M., Russell, M.B., Walters, B., Hom, J., Peduzzi, A. *et al.* 2017 Estimating aboveground live understory vegetation carbon in the United States. *Environmental Research Letters*, **12** (12), 125010.

21. Keane, R.E., Burgan, R. and van Wagtendonk, J. 2001 Mapping wildland fuels for fire management across multiple scales: Integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire*, **10** (4), 301-319.

22. Kuhn, M. 2015 A Short Introduction to the caret Package. *R Found Stat Comput*, **1**, 1-10.

23. Lafarge, T., Pateiro-Lopez, B. and Pateiro-Lopez, M.B. 2016 Package 'alphashape3d'. *Recuperado el*, **26**.

24. Lague, D., Brodu, N. and Leroux, J. 2013 Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). *ISPRS Journal of Photogrammetry and Remote Sensing*, **82**, 10-26.

25. Lauer, D.K. and Glover, G.R. 1999 Stand level pine response to occupancy of woody shrub and herbaceous vegetation. *Canadian journal of forest research*, **29** (7), 979-984.

26. Li, S., Wang, T., Hou, Z., Gong, Y., Feng, L. and Ge, J. 2021 Harnessing terrestrial laser scanning to predict understory biomass in temperate mixed forests. *Ecological Indicators*, **121**, 107011.

27. Lindberg, E., Holmgren, J., Olofsson, K. and Olsson, H. 2012 Estimation of stem attributes using a combination of terrestrial and airborne laser scanning. *European Journal of Forest Research*, **131**, 1917-1931.

28. Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C. and Culvenor, D.S. 2005 Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, **214** (1), 398-412.

29. Michael, J. 1980 Long-term impact of aerial application of 2, 4, 5-T to longleaf pine (Pinus palustris). *Weed Science*, **28** (3), 255-257.

30. Miller, J.H., Zutter, B.R., Zedaker, S.M., Edwards, M.B. and Newbold, R.A. 2003 Growth and yield relative to competition for loblolly pine plantations to midrotation—a southeastern United States regional study. *Southern Journal of Applied Forestry*, **27** (4), 237-252.

31. Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B. *et al.* 2015 Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*, **1** (4), 239-251.

32. Olsoy, P.J., Glenn, N.F., Clark, P.E. and Derryberry, D.R. 2014 Aboveground total and green biomass of dryland shrub derived from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **88**, 166-173.

33. Oswalt, S.N., Smith, W.B., Miles, P.D. and Pugh, S.A. 2014 Forest Resources of the United States, 2012: a technical document supporting the Forest Service 2010 update of the RPA Assessment. *Gen. Tech. Rep. WO-91. Washington, DC: US Department of Agriculture, Forest Service, Washington Office. 218 p.*, **91**.

34. Peduzzi, A., Allen, H.L. and Wynne, R.H. 2010 Leaf area of overstory and understory in pine plantations in the flatwoods. *Southern Journal of Applied Forestry*, **34** (4), 154-160.

35. Popescu, S.C. and Wynne, R.H. 2004 Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, **70** (5), 589-604.

36. Reutebuch, S.E., Andersen, H.-E. and McGaughey, R.J. 2005 Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. *Journal of Forestry*, **103** (6), 286-292.

37. Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R., Meador, A.S. *et al.* 2020 lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, **251**, 112061.

38. Sabo, K.E., Hart, S.C., Sieg, C.H. and Bailey, J.D. 2008 Tradeoffs in overstory and understory aboveground net primary productivity in southwestern ponderosa pine stands. *Forest Science*, **54** (4), 408-416.

39. Schultz, R.P. 1997 *Loblolly pine: the ecology and culture of the loblolly pine (Pinus taeda L.)*. US Government Printing Office.

40. Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K. *et al.* 2016 Imputation of individual longleaf pine (Pinus palustris Mill.) tree attributes from field and LiDAR data. *Canadian journal of remote sensing*, **42** (5), 554-573.

41. Singh, K.K., Chen, Y.-H., Smart, L., Gray, J. and Meentemeyer, R.K. 2018 Intra-annual phenology for detecting understory plant invasion in urban forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, **142**, 151-161.

42. Stagg, R.H. and Scott, D.A. Understory growth and composition resulting from soil disturbances on the long-term soil productivity study sites in Mississippi, pp. 52-56.

43. Suchar, V.A. and Crookston, N.L. 2010 Understory cover and biomass indices predictions for forest ecosystems of the Northwestern United States. *Ecological Indicators*, **10** (3), 602-609.

44. Team, R.C. 2013 R: A language and environment for statistical computing.

45. Venier, L.A., Swystun, T., Mazerolle, M.J., Kreutzweiser, D.P., Wainio-Keizer, K.L., McIlwrick, K.A. *et al.* 2019 Modelling vegetation understory cover using LiDAR metrics. *Plos One*, **14** (11), e0220096.

46. Wagner, R.G., Little, K.M., Richardson, B. and Mcnabb, K. 2006 The role of vegetation management for enhancing productivity of the world's forests. *Forestry*, **79** (1), 57-79.

47. Wallace, L. Assessing the stability of canopy maps produced from UAV-LiDAR data. IEEE, pp. 3879-3882.

48. Watts, F.C. 1991 *Soil Survey of Nassau County, Florida*. Soil Conservation Service: Washington, D.C.

49. Willem, W. and Mead, B.R. 2000 Ocular estimates of understory vegetation structure in a closed Picea glauca/Betula papyrifera forest. *Journal of Vegetation Science*, **11** (2), 195-200.

50. Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., Gitelman, A. and Olsen, M.J. 2012 Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. *Remote Sensing of Environment*, **124**, 730-741.

51. Xu, K., Su, Y., Liu, J., Hu, T., Jin, S., Ma, Q. *et al.* 2020 Estimation of degraded grassland aboveground biomass using machine learning methods from terrestrial laser scanning data. *Ecological Indicators*, **108**, 105747.

52. Yang, B. and Dong, Z. 2013 A shape-based segmentation method for mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, **81**, 19-30.

53. Zhang, J., Powers, R.F., Oliver, W.W. and Young, D.H. 2013 Response of ponderosa pine plantations to competing vegetation control in Northern California, USA: a meta-analysis. *Forestry*, **86** (1), 3-11.

54. Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X. *et al.* 2016 An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sensing*, **8** (6), 501.

55. Zou, H. 2006 The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** (476), 1418-1429.

CHAPTER 4

ASSESSING UNDERSTORY VEGETATION BIOMASS IN PINE PLANTATION FOREST WITH MULTI-SCALE LIDAR[3]

[3] Adhikari, A., Peduzzi, A., Montes, C. R., Osborne, N., & Mishra, D. R. To be submitted to Ecological Indicators.

Abstract

This research assesses understory vegetation biomass in southeastern U.S. coastal forests using laser scanning data from ground and aerial sensors, addressing the limitations of optical imagery. We collected lidar data from 60 sample plots with ground-based and Unmanned Aerial Vehicle (UAV)-mounted lasers in intensively managed unthinned loblolly pine plantation stands in Nassau County, Florida. Additionally, we utilized publicly available Airborne Laser Scanning (ALS) data collected by the National Oceanic and Atmospheric Administration (NOAA), covering our study plots. The dry weight of destructively sampled understory vegetation from two 1 m^2 subplots in each plot provided biomass values for model calibration purposes. We applied ALASSO regression to TLS-derived metrics, creating an optimal biomass model, which was then upscaled to UAV and ALS lidar-derived variables. Results indicate that while TLS is precise (Adjusted R^2 =0.85 and RMSE = 0.233 kg/m^2), it is not practical for large areas. Conversely, ALS-based models show high accuracy (Adjusted R^2 = 0.7 and RMSE = 0.167 kg/m^2) and predictions within an acceptable range of uncertainty, demonstrating efficacy and potential for enhanced forest management and decision-making.

Keywords: Understory biomass assessment, upscaling, ALS, UAV, biomass modeling, LiDAR

4.1 Introduction

Forest ecosystems are typically characterized by understory and overstory communities. The overstory layer is the top layer, usually above 3 meters in height, and consists of woody plants. The understory layer falls underneath the overstory layer, usually less than 3 meters, and includes herbaceous vegetation such as plants, shrubs, saplings, seedlings, stumps, and grasses (Adhikari, et al. 2023b; Li, et al. 2020). Both ecological communities play vital roles in the functionality of forest ecosystems. Nevertheless, in an intensively managed plantation pine forest that is common in the southeastern United States, the presence of understory is not favorable for the productivity of overstory trees. Although understory vegetation does not directly affect the access to light in the canopy, it still increases the competition for site resources such as soil nutrition, which is documented as one of the major limiting factors for site productivity in the coastal forest of the southeastern United States (Adegbidi, et al. 2002; Allen, et al. 2005; Peduzzi, et al. 2010).

Furthermore, several studies in similar plantation areas have shown a two to four-times increase in overstory productivity and significant improvement in survival after controlling the competing vegetation (Fortson, et al. 1996; Wagner, et al. 2006; Zhang, et al. 2013). For instance, Michael (1980) conducted a comprehensive study on longleaf pine at three distinct sites in Alabama, applying a similar vegetation control treatment that resulted in a substantial 40% increase in wood volume yield after 20 years. Similarly, Miller, et al. (2003) conducted a study spanning 13 sites across seven states in the southeastern US, where they applied a treatment to control vegetation competition in loblolly pine plantations. Their findings suggest an increase in wood volume yield ranging from 30% to 148% after a 15-year post-treatment period (Miller, et al. 2003).

These studies highlight the apparent correlation between the accelerated growth of planted pine forests and the reduction of early herbaceous and long-term woody competition. Thus, timely monitoring and quantification of competing vegetation are necessary to enhance the productivity of a plantation forest.

However, assessing understory vegetation poses significant challenges. Assessment approaches commonly involve fixed plots and ocular-based sampling techniques (Bonham 2013). Moreover, these field-based methodologies are frequently characterized by inherent drawbacks, including time intensiveness, high-cost associations, and limited to small areas (Eskelson, et al. 2011; Lovell, et al. 2005). In contrast, remote sensing (RS) methods emerge as a more rational and viable alternative for the timely detection and quantification of competing understory vegetation. RS techniques facilitate the acquisition of repetitive measurements across extensive and hard-to-reach areas. However, it is essential to note that conventional RS data, including multispectral satellite imagery from sensors like Landsat and the Moderate Resolution Imaging Spectroradiometer (MODIS), exhibit limited utility in understory vegetation detection due to the low spatial resolution and overstory canopy inferences. This limitation arises from interference by overstory vegetation and the absence of plant-level, site-specific methodologies (Singh, et al. 2018).

Laser Scanning (LS) presents a promising alternative for effectively capturing information about understory vegetation by overcoming the limitations associated with conventional optical remote sensing methods. Laser Scanning, particularly LiDAR (Light Detection And Ranging), is an active remote sensing technology employing electromagnetic energy to generate comprehensive 3D representations of vegetation

structures (Reutebuch, et al. 2005; Venier, et al. 2019). LiDAR operates by precisely measuring the 3D coordinates of surface points, achieving an accuracy level within a few decimeters within vegetative canopies, and yielding dense point clouds (Popescu, et al. 2004). Terrestrial Laser Scanning (TLS) is a ground-based lidar sensor that collects a very high point-cloud density, enabling the acquisition of 3D vegetation structures at high spatial resolution and centimeter-to-millimeter-level accuracy (Lague, et al. 2013; Newnham, et al. 2015). The applicability of TLS for forest monitoring has been demonstrated in scholarly publications since the early 2000s (Hopkinson, et al. 2004; Jupp, et al. 2009; Strahler, et al. 2008). Initially, its primary application was focused on measuring conventional structural parameters commonly used in forestry, such as tree height and diameter at breast height (DBH). However, over time, its utility expanded to include the comprehensive assessment of whole-stand volumes, enhancing the precision of above-ground biomass estimations (Calders, et al. 2015; Momo Takoudjou, et al. 2018).

Due to its 3D nature and enhanced ground-level resolution compared to the canopy, TLS holds great potential to address the challenges associated with understory vegetation assessment, where conventional methods often fall short due to their inefficiency. Furthermore, existing research has demonstrated its effectiveness in accurately predicting and quantifying understory vegetation. This not only aids forest managers in identifying areas with high understory competition but also supports habitat assessments that rely on understory vegetation, such as bushes and shrubs (Ashcroft, et al. 2014; Li, et al. 2021). An exemplary case is a study by Adhikari et al. (2023), TLS data from 60 sample plots of an operational loblolly pine plantation in Nassau County, Florida, was utilized and

modeled with understory biomass data collected through destructive sampling. Their findings suggest a strong correlation (R^2 of 0.80 and RMSE 234.8 grams) between TLS-derived metric and the understory biomass. Leveraging the remarkably high point density of TLS, their model demonstrated the capability to predict understory biomass at resolutions as fine as one meter. However, a major constraint of the TLS-based method is its applicability to small areas or plot-level assessments due to ground-level LiDAR data collection requirements, which are time and resource-intensive.

One way to address the limitation of the TLS approach is the traditional design-based approach, which involves sampling a subset of population units to extrapolate overall population characteristics (Dumelle, et al. 2022). This method is typically used for attribute estimation of the overall population. In the context of plantation forest management, identifying specific areas of high competition is more critical than general biomass estimation (Wing, et al. 2012). Thus, design-based approaches fall short of accurately mapping spatial distribution, necessitating a model-based approach incorporating data from the entire area. This is where an upscaling approach becomes relevant.

Field-based methods for understory estimation are often impractical and due to the extensive spatial coverage needed to calibrate sensors for large areas (Li, et al. 2021). In this context, TLS serves as an effective bridge, accurately linking field-measured data with larger-scale airborne sensors. Upscaling involves transitioning from a TLS-based model to aerial LiDAR approaches, such as Unmanned Aerial Vehicle Laser Scanning (UAV-LS) for stand-level assessments and Airborne LiDAR Scanning (ALS) for landscape-level assessments.

UAV-LS involves attaching LiDAR sensors to drones or UAVs, facilitating the acquisition of three-dimensional (3D) geospatial data at relatively low altitudes. Conversely, airborne LS involves the installation of sensors aboard an aircraft or helicopter to acquire 3D data from an elevated vantage path above the Earths surface. Airborne LS is distinguished by its ability to cover larger areas more quickly than TLS at the cost of reduced point density, making it well-suited for regional mapping, forestry, and topographic surveys. In contrast, UAV-LS offers a compromise between the high-point density of TLS and the aerial coverage, making it more flexible and suitable for a wide range of applications. UAV LiDAR is particularly valuable for monitoring small to medium-sized areas with high precision. Therefore, upscaling the TLS-based approach to an aerial LS-based approach could aid plantation managers in monitoring competing vegetation and deciding control measures to enhance pine growth in this region.

Thus, this study aims to assess and predict understory vegetation biomass by evaluating multi-scale lidar data. Specifically, we seek to 1) Upscale predicted understory biomass from a TLS-derived model to UAV lidar-derived variables at a resolution as fine as 8 meters. 2) Upscale predicted understory biomass from a TLS-derived model to airborne lidar-derived variables with a resolution as fine as 10 meters.

4.2 Methods

4.2.1 Study area

This study is based on an intensively managed loblolly pine plantation forest managed by Rayonier Inc., situated in Nassau County, Florida (30°60′N; 81°80′W) as shown in Figure 4.1. The county has a warm, humid summer climate and mild winter conditions, with an average temperature during summer reaching 26 °C. In contrast, the year remaining months have an average temperature of approximately 17 °C. The summer months receive a substantial portion of the annual precipitation, accounting for around 65% (averaging 132.8 cm). In contrast, the remaining month average rainfall is approximately 7.62 cm (Watts 1991).

The soil in this county is classified as sandy and well-drained, with limited nutrient content (Nassau County Extension, 2020). The dominating overstory vegetation in the area includes loblolly pine (*Pinus taeda*) and slash pine (*Pinus elliottii*) plantation forests, which have been intensively managed to optimize timber production by promoting growth and yield. Likewise, gallberry (*Ilex glabra* (L.) Gran) and saw palmetto (*Serenoa repens* (Bartr,) small) are the major evergreen understory vegetation, and ti-ti (*Cyrilla recemiflora L.*), fetterbush (*Lyonia Lucida* (Lam) K. Koch), staggerbush (*Lyonia ferruginea* (Walter) Nutall), sweet bay (*Magnolia virginiana* L.), blueberries (*Vaccinium spp.*), saw greenbrier (*Similax bona-nox* L.), cat greenbrier (*Similax glauca* Walt.), in addition to woody evergreen varieties like red maple (*Acer rubrum* L.) and sweetgum (*Liquidamber styraciflua L.*) are others species found in understory portion of the forest growing in this area.

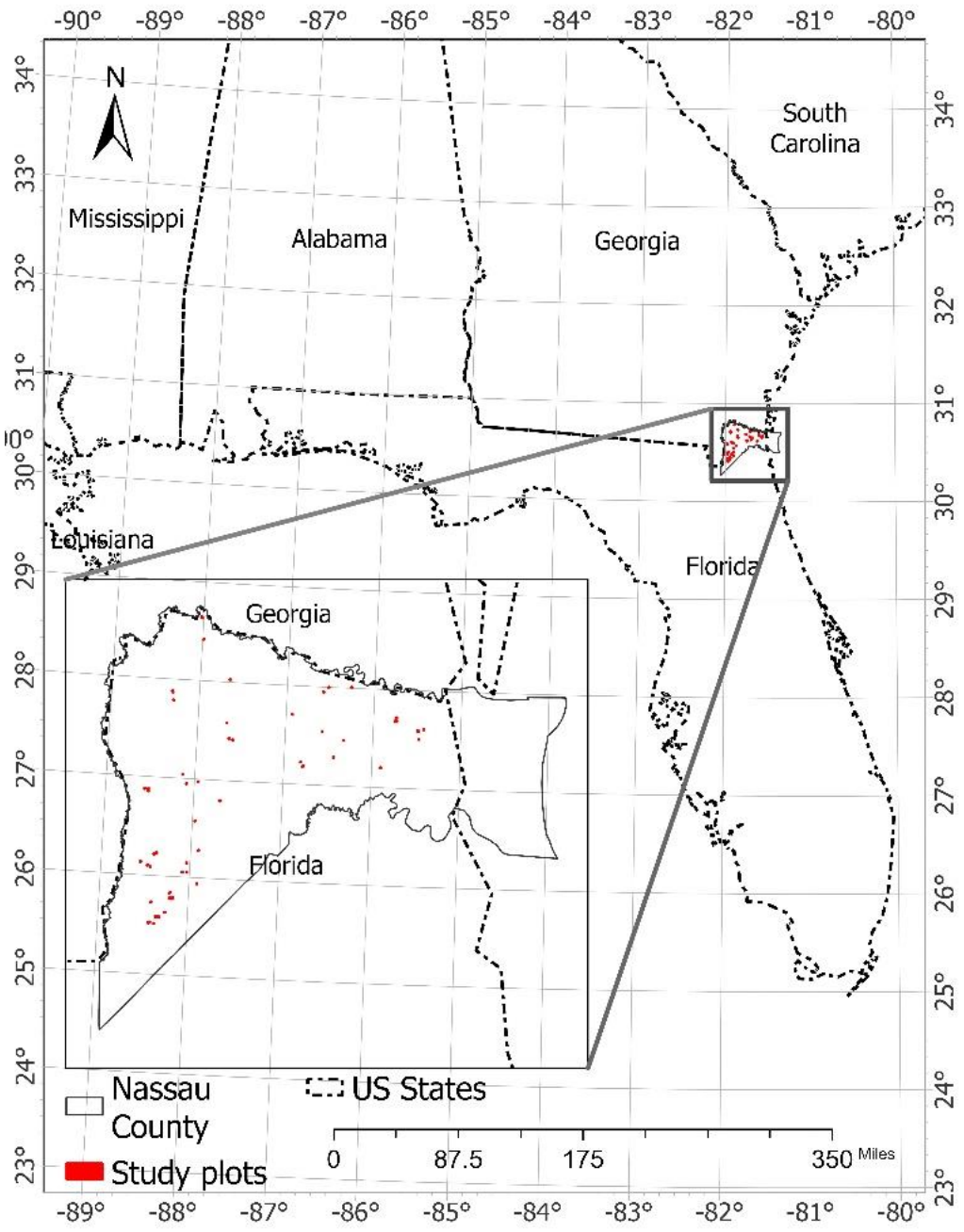

Figure 4.1 Overview of the research area and the distribution of the sample plots in Nassau County, Florida, US.

4.2.2 Field planning and data collection

4.2.2.1 Understory Biomass Data

To collect understory biomass data from the field, we first selected all unthinned loblolly pine plantation stands within the age range of 11 to 21 years across the study area. After selecting the stands, we established 60 sample plots within these stands using a stratified random sampling approach employing two strata: dominant height and abundance of understory vegetation.

We used six classes generated from the understory density raster file with a spatial resolution of 10 meters to establish sample plots based on the abundance of understory vegetation strata. Rayonier Inc. generated this file based on low-resolution airborne LiDAR data. Likewise, the dominant height strata were based on four height bins determined by the mean height of the stand. Employing these two criteria, we established a square-shaped sample plot of 1600 square meters (40m × 40m) distributed evenly within all strata. Refer to Figure 4.1 for a visual representation of the geographic distribution of these sample plots. After establishing the sample plots, we installed two subplots of $1m^2$ (1 m × 1 m) positioned at the center of diagonally opposite quadrants of each plot. From those established subplots, all green plants growing inside were destructively sampled. Out of 120 subplots, we collected understory samples from 100 of them. The remaining 20 subplots did not yield samples due to the absence of vegetation within their designated boundaries, even though understory vegetation was present in the surrounding plot area.

After destructively collecting samples from 100 plots, the samples were dried at 105 °C in an oven for 48 hours or until a constant dry weight was achieved. Upon reaching this

constant dry weight, the samples were weighed using a precise weighing scale with an accuracy of 10 grams to record their dry weight. The obtained dry weight value for each subplot was considered the dry understory biomass and used as a dependent variable for the modeling process. For further plot design and field planning specifics, refer to(Adhikari, et al. 2023b).

Table 4.1 Sample plot distribution within understory abundance and dominant height criteria. Understory abundance level one refers to the presence of very low to no understory, and six represents very high abundance in the sample plot.

Stand Dominant Height	**Understory Abundance levels**						
Class (meters)	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
(4.6 - 8.3)	2	3	3	2	3	2	**15**
(8.3 - 13.3)	2	2	2	3	2	4	**15**
(13.3 - 17.1)	2	2	3	3	3	2	**15**
(17.1 - 25)	4	3	2	2	2	2	**15**
Total Plots	**10**	**10**	**10**	**10**	**10**	**10**	**60 Plots**

4.2.2.2 TLS data acquisition and processing

We used the GeoSLAM ZEB HORIZON personal mobile laser scanner developed by GeoSLAM Ltd. (Nottingham, UK) to collect high-density ground-based LiDAR data on August 2022. This handheld device is equipped with a 16-channel VLP-16 sensor (Velodyne LiDAR Inc., Morgan Hill, CA, USA) that is capable of measuring up to 300,000 points per second in single return mode with a maximum range of 100 meters. The LiDAR sensor is integrated with a survey-grade, real-time SBAS (Satellite-Based Augmented System) corrected Trimble R2 GNSS system precisely aligned on top of the LiDAR sensor along the central axis. This configuration ensures a seamless

transformation of the locally generated coordinate system from the LiDAR sensor to a global coordinate reference system.

In the post-data acquisition phase, a series of preprocessing steps were conducted, that includes; (1) noise filtering, such as isolated points, air points, and low points, are eliminated using the Isolated Voxels Filter (IVF) based noise filtering algorithm. (2) Downsampling point clouds to maintain consistent point density across the plot. (3) The classification of ground and non-ground points using a Cloth Simulation Filter (CSF) (Zhang, et al. 2016). (4) Height normalization using a digital terrain model (DTM) of 0.2m resolution generated using the k-nearest Neighbor Inverse-Distance Weighting Algorithm (kNN-IDW). (5) Clipped normalized LiDAR data to the plot level (40m×40m) using the vector files defining the plot perimeter established in the field. (6) Tree stem identification and removal using density-based individual tree detection methods (Adhikari, et al. 2023b). (7) Filtering out all LiDAR point clouds above three meters, which we consider to be the maximum height for understory layers based on the field observation. (8) Clipping lidar point clouds representing the subplots of $1m^2$ using subplot center coordinates recorded in the field. The preprocessing, individual tree detection, stem removal, and subplot clipping were implemented using lidR package (Roussel, et al. 2020) in the R environment (Team 2013). More detail on the sensor, field design for TLS data acquisition, and data processing is available in Adhikari, et al. (2023b).

4.2.2.3 UAV LiDAR data acquisition and processing

The UAV-lidar data was obtained in the winter 2022 with the Headwall UAV system. The Headwall unit comprises a DJI M600 Pro UAV, a Velodyne VLP-16 Lite (Velodyne

LiDAR Inc., Morgan Hill, CA, USA) lidar sensor, and an Inertial Measurement Unit (IMU). The VLP-16, weighing 0.59 kg, is a Light Detection and Ranging (LiDAR) featuring 16 channels. The unit can measure up to 600,000 points per second in dual return mode within a maximum range of 100 m. The device uses a continuous wavelength of 903 nm (near infrared) to measure distances with a range accuracy of ±3 cm. The VLP-16 has a horizontal field of view of 360° × 30° with a vertical angular resolution of 2° and a horizontal angular resolution of 0.1°–0.4°. The VLP-16 analyzes multiple returns and reports either the strongest, last, or both returns. In this study, we used both return modes, where the sensor records two returns (first and last) when the distance between two objects situated within the same laser beam is greater than 1m. It has an L1/L2 dual-frequency GNSS receiver for generating post-processing kinematic (PPK) trajectories. For this study, considering the sensor range and required flight time, we maintained a flying height of 75 m above the ground and a speed of 5 ms^{-1} with an approximate horizontal distance between the adjacent flight lines of 20 m, producing a high-density lidar point cloud (141 points per meter squared). Furthermore, as shown in Fig. 4.2, the flight lines of the UAV were designed to exhibit a high degree of redundancy, achieving an overlap of 90% within the designated survey area.

Following acquiring LiDAR data from the UAV platform, similar to TLS procedures, the subsequent processing stages include several key steps. These included noise filtering, ground and non-ground points classification, height normalization using a DEM of resolution 0.5 m, horizontal extent clipping using the vector files defining the plot (40m × 40m), and vertical extent clipping using the defined understory layer extent of 0 – 3 meters in height.

Figure 4.2 Figure shows the UAV-flight plan to collect lidar data, which was created in UGCS software and projected over google earth pro base map. In the figure, yellow lines, which are 70 meters above ground, are the UAV flight lines, and the red square in the middle represents the sample plot (40m × 40m) boundary.

4.2.2.4 Airborne LiDAR data acquisition and processing

Since 1998, the National Oceanic and Atmospheric Administration (NOAA), in collaboration with the US Army Corps of Engineers (USACE) and US Geological Survey (USGS), has conducted a series of LiDAR mapping surveys of beach and nearshore areas nationally (NOAA 2023). These publicly available datasets have proven to be effective tools for long-term monitoring of morphological changes in barrier islands, frontal beaches, and coastal vegetation patterns. In this study, we downloaded the data that overlapped all our sample plots from NOAA Data Access Viewer website which was collected during the winter of 2018 (NOAA 2023). This dataset exhibited an average

point density of 31 points per square meter, with each pulse yielding an average of five return points.

After downloading the lidar point cloud tiles covering all sixty plots, we transformed the geographic coordinate from the North American Datum of 1983 with the 2011 Adjustment (NAD 83 (2011)) to World Geodetic System 1984 (WGS 84)/ UTM zone 17N, to make it consistent with the other datasets. Similar to UAV LiDAR, other processing steps involved noise filtering, height normalization, and horizontal and vertical extent clipping to sample plot and understory layer extent, respectively. Moreover, the essential parameters of the LiDAR datasets from the three platforms used in this study are presented in Table 4.2, and the detailed point density comparison between UAV and airborne LiDAR data at different height levels is summarized in Table 4.3.

Table 4.2: The essential parameters of the three LiDAR datasets.

Parameters	TLS	UAV	ALS
Sensor	VLP16	VLP16	Leica ALS80
Wavelength (nm)	905	905	1064
Altitude (m above ground level)	--	75	1400
Scan pulse rate (kHz)	300	300	586
Scan frequency (Hz)	10	10	52
Ranging accuracy (cm)	±3	±3	±2
Max number of returns per Pulse	1	2	7
Field of view-horizontal (degrees)	360	360	40
Field of view-horizontal (degrees)	15°	--	--
Nominal flight speed (kts)	--	9.7	150
Measuring range (m)	<100	<100	<5000

Table 4.3: Point density summary at different height levels

LiDAR data Source	Distribution metrics	Height Classes (m)					
		All	**0-1**	**1-2**	**2-3**	**0-3**	**3-6**
UAV-LS	Average	141.6	69.3	10.4	3.1	76.6	3.8
	Min	29.6	17.2	0.3	0.1	17.8	0.1
	Max	511.9	363.1	113.9	27	359.1	21.8
	SD	84.5	50.4	16.4	5.1	50.2	5.8
ALS	Average	30.9	8.1	3.2	2.4	12	5
	Min	25.2	6.7	0.6	0.3	9.2	2.9
	Max	38.5	10.1	5.1	3.1	14.4	7.6
	SD	3.7	0.8	0.8	0.4	1.1	1

4.2.3 Variable Extraction

4.2.3.1 TLS-based variables for understory biomass modeling

Following preprocessing, downscaling, and stem removal, we clipped the LiDAR data using the subplot boundary file. Next, we utilized the clipped $1m^2$ LiDAR point clouds to generate various metrics for modeling understory vegetation. The variables generated for this study include seven statistical variables such as mean, median, and standard deviation of the echo height, nine percentile variables representing echo height at a certain percentage, two variables representing understory cover, and an additional nine variables were incorporated to estimate vegetation volume. These volume estimations were carried out using three distinct methodologies at varying spatial resolutions: voxel count-based volume (Hancock, et al. 2017), volume estimation predicated on alpha-hull fitting (Lafarge, et al. 2016), and a method based on mean height in conjunction with understory cover to calculate vegetation volume. A detailed description of the calculation of these metrics and their statistical summary can be found in Adhikari, et al. (2023b).

4.2.3.2 UAV and Airborne LiDAR-based variables for upscaling from TLS-variable-based model

After the initial processing phase, the UAV and Airborne LiDAR datasets were first projected over Terrestrial Laser Scanning (TLS) data to verify the absence of registration discrepancies (Figure 4.3). This verification process affirmed that the UAV and Airborne LiDAR data spatially overlapped with the TLS data with high accuracy. In instances where minor spatial misalignments were detected, manual adjustments were performed to ensure precise co-registration. Next, the lidar files were clipped horizontally and vertically for plot boundaries and understory levels, respectively, and these spatially corrected LiDAR point clouds were then utilized to extract 40 different pixel metrics at a resolution of 8 meters. Among the 40 metrics extracted, 6 are descriptive statistic metrics such as mean, median, and standard deviation. The other 19 represent the height distribution at different quantiles (percentile), and 11 metrics are the canopy density metrics (Woods, et al. 2008). The remaining four metrics are volume-based and are calculated through the voxelization method employing a voxel size of 1 cubic meter (1 m^3). A detailed description of those 40 metrics is presented in Table 4.4.

Table 4.4: Description of UAV and ALS-based metrics derived for upscaling TLS-based model.

Abbreviations	**Description**
Zmax, Zmean, Zsd, Zskew, Zkurt, Zentropy	Maximum, Mean, Standard deviation, skewness, kurtosis, and entropy of the echo heights distribution
Zq05, Zq10, Zq15, Zq20, Zq25, Zq30, Zq35, Zq40, Zq45, Zq50, Zq55, Zq60, Zq65, Zq70, Zq75, Zq80, Zq85, Zq90, Zq95	Echo height distribution percentiles (5th, 10th, ……, 90th, 95th)

Abbreviations	Description
Zpcum1, Zpcum2, Zpcum3, Zpcum4, Zpcum5, Zpcum6, Zpcum7, Zpcum8, Zpcum9	Cumulative proportion of return in the 1th, 2th, 3th, ……, 9th layer
PzaboveZmean, Pzabov2	percentage of returns above mean height and 2 meters
Voxvol, Wt_voxvol	Volume based on voxel count, weighted voxel volume by height
Log_voxvol, Log_wt_voxvol	Natural log of voxel count based on volume and height-weighted volume

It is noteworthy that, despite the UAV-mounted LiDAR system possessing superior point density when compared to ALS point clouds, the limitations stemming from a low return per pulse configuration (specifically, with only two returns per pulse) resulted in a concentration of LiDAR points primarily at either the upper tree canopy or ground levels (Fig. 4.3). This limitation reflects a reduced representation of the intermediate and understory profiles within the forest, thereby imposing constraints on achieving higher-resolution metrics generation. Given the challenges of limited representation of the forest vertical profile by the UAV LiDAR and the relatively low point density of the Airborne LiDAR data in combination with a plot size of 1600 square meters, we found that to address the data constraints and attain a meaningful geospatial representation effectively, an optimal pixel resolution of 8 meters for metrics generation (equivalent to 64 squared meters) was required.

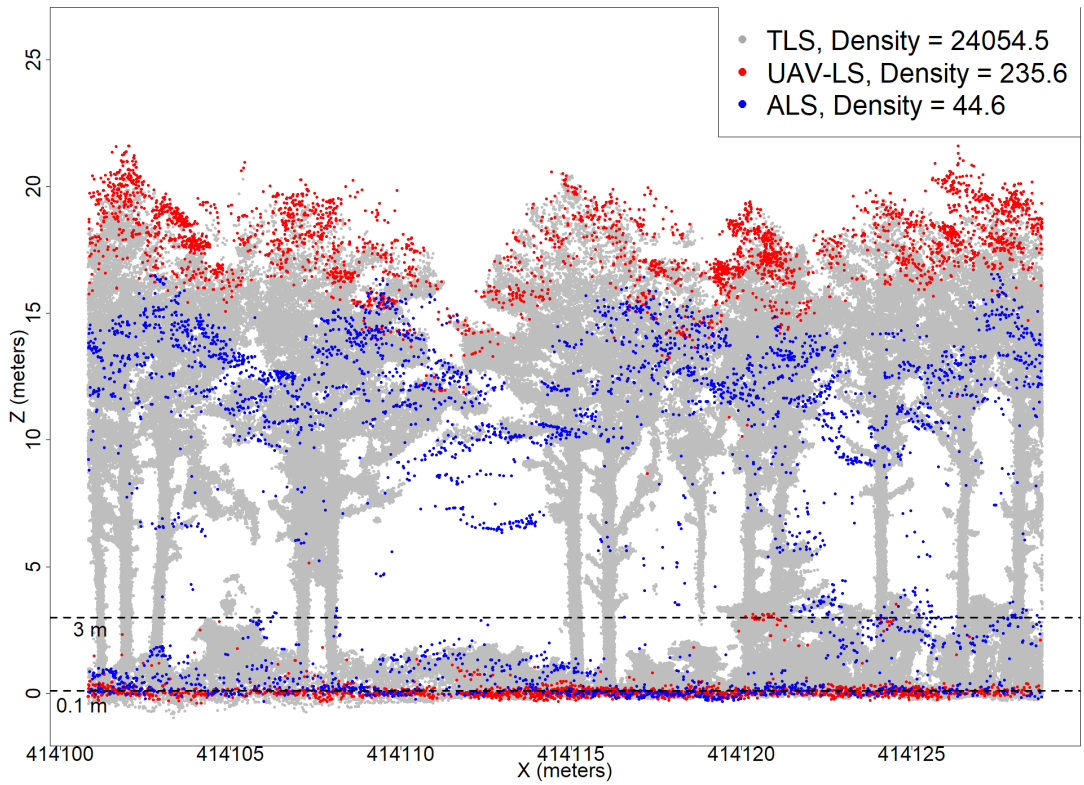

Figure 4.3: Forest vertical profile showing LiDAR point cloud density and distribution comparison plot for TLS, UAV-LS, and ALS for one of the sample plot.

4.2.4 Understory biomass modeling and upscaling

4.2.4.1 Modeling method

For the understory biomass and lidar metrics modeling, we used the Adaptive LASSO (Least Absolute Shrinkage and Selection Operator) method (Zou 2006). While linear regression stands as the widely accepted approach for modeling, it is noteworthy that it confronts a significant challenge when dealing with high-dimensional data, notably the issue of multicollinearity. Multicollinearity denotes the presence of strong correlations among predictor variables. This condition ultimately results in unstable and unreliable model estimates (Holcomb 1999). In this study, where we incorporated various LiDAR metrics as predictors, multicollinearity may limit the model accuracy due to their inherent

relationships stemming from their common dependence on LiDAR echo heights. Adaptive LASSO solves these issues by automatically selecting variables while simultaneously addressing multicollinearity. Notably, it outperformed alternative methods such as least square regression and random forest for variable selection and modeling, specifically LiDAR metrics-based modeling (Adhikari, et al. 2023a). Moreover, the Adaptive LASSO addresses the limitations of traditional LASSO by providing oracle properties, which means it selects the meaningful variables with higher probability, even in the presence of correlated predictors. In this study, the application of Adaptive LASSO allowed us to address multicollinearity challenges associated with incorporating highly correlated TLS metrics as predictors. By automatically selecting the most relevant variables, Adaptive LASSO provided a more robust and accurate model for biomass estimation, which we further validated using a 10-fold cross-validation method.

4.2.4.2 Understory Biomass Modeling with TLS-derived Variables

For modeling understory biomass with TLS variables, we used measured dry understory weight as our dependent variable and the TLS-derived 27 variables as independent variables. The modeling approach employed the Adaptive Lasso technique, which involved a two-step regression procedure. First, we applied Ridge Regression to obtain preliminary coefficient estimates. This method is characterized by a penalty term that reduces the magnitude of the coefficients through shrinkage, though it does not eliminate any variables from the model by setting their coefficients to zero (Hoerl, et al. 1970). The selection of an appropriate penalty strength, denoted as lambda (λ), was determined through 10-fold cross-validation to identify the λ that minimized the prediction error of the model.

Subsequent to Ridge Regression, we used the derived coefficients to determine the penalty weights in the Adaptive Lasso method. In this context, the weight for a given coefficient was inversely proportional to its absolute value as determined by Ridge Regression,

i.e., $w_j = 1/\left|\hat{\beta}_j^{Ridge}\right|$

where $\hat{\beta}_j^{Ridge}$ is the jth coefficient estimated by Ridge Regression and w_j is the weight for the jth coefficient for ALASSO regression.

This weighting scheme ensures that coefficients considered less significant by the Ridge model are subject to a higher penalty. Following this, the Adaptive LASSO method was implemented, utilizing the calculated weights to impose variable penalties across different coefficients, as opposed to a singular penalty in traditional LASSO. The Adaptive LASSO penalty parameter (λ) was also optimized using 10-fold cross-validation to minimize the prediction error.

Finally, the Adaptive Lasso model was fitted with these weights. One of the notable features of Adaptive Lasso is its ability to reduce certain coefficients to zero, thereby facilitating variable selection. Through this process, only a subset of the original 27 variables, those that provided the most predictive power without multicollinearity, were retained to estimate understory biomass accurately. The final model with the selected variables was then used to predict plot-level biomass at a 1-meter resolution.

4.2.4.3 Upscaling TLS variables-based model to UAV-LS and ALS variables

For upscaling, we used plot-level (1600 m^2) understory biomass maps for all sixty plots predicted at the spatial resolution of 1 meter using the model resulting from section 2.4.2 to model with UAV-LS-derived and ALS-derived variables. Due to the limited vertical representation of the forest structure by UAV-LS and the relatively sparse point density provided by the ALS data, as shown in Figure 4.3, we faced challenges in modeling at the 1-meter resolution level. Thus, we adjusted the resolution of the predicted biomass map to 8 meters (64 m^2) for metric generation to mitigate these issues and ensure an effective geospatial representation of the data. We resampled the 1 m^2 TLS-based biomass prediction maps to 8 m^2 spatial resolution. Furthermore, regions with insufficient TLS lidar coverage within the plot were filtered out to enhance data quality and consistency. Additionally, four plots intended for Unmanned Aerial Vehicle (UAV) modeling were excluded due to low data quality. Consequently, this data filtration process resulted in a final dataset comprising 850 pixels suitable for UAV-LS-based modeling and 936 pixels for ALS-based modeling.

After resampling and data filtration, we generated the forty metrics outlined in section 4.2.3.2 for each biomass pixel, utilizing point cloud data from both UAV-LS and ALS platforms. These metrics were modeled using the ALASSO regression method, as detailed in section 4.2.4.2. In the initial ALASSO regression model, the default selection of the lambda minimum value resulted in the retention of several variables with comparatively low coefficient values and added complexity for model application. We opted for a slightly larger lambda value to refine the model further. This adjustment imposed more strict penalties, effectively reducing the number of variables retained in the

model. Although this led to a slight decrease in accuracy, it significantly enhanced the model parsimony, balancing the complexity of the model against the precision of the predictions.

4.2.4.4 Upscaling UAV-LS variables-based model to ALS variables

After upscaling the TLS-based biomass estimation model to UAV-LS derived variables, to exploit the broader spatial coverage provided by UAV-LS compared to TLS, we expanded the upscaling further from UAV-LS to ALS-based variables. Employing the final UAV-LS model established in section 4.2.4.3, we expanded the understory biomass prediction to a larger area outside the plot boundary. Specifically, we incorporated a 2500 m^2 area for each plot, which expanded each plot used for TLS modeling by 5 meters on each plot side. This biomass map was generated using the UAV-LS model and at a spatial resolution of 10 meters (100 m^2), producing a biomass dataset comprising 1500 pixels. Within this dataset, from the sixty sample plots, eight sample plots —equivalent to 200 pixels—were excluded from the analysis. This exclusion was due to various factors affecting data quality and relevance. Three of the plots were omitted due to inferior UAV data quality, which could compromise the reliability of the model output. The other five plots were disregarded because the low forest stand age and stand height to avoid the effect of temporal inconsistency and vegetation growth from the data acquisition gap between UAV-LS and ALS. For those filtered 1300 pixels with a UAV-LS-based model predicted biomass value, we calculated 40 metrics explained in section 4.2.3.2 using the ALS point clouds and fitted ALASSO regression to generate a prediction model.

4.2.4.5 Accessing Accuracy and Performance of the Model

In this study, we evaluated the accuracy of models with different explanatory variables using the k-fold cross-validation method. This approach involves randomly dividing a set of observations into k subsets, or folds, of equal size. The first fold is then utilized as a validation set, while the model is fitted on the remaining k-1 folds. We used the caret package within the R environment to implement this method, which allowed us to split the datasets into ten folds (Kuhn 2015). Subsequently, the model was fitted on nine of the ten folds, while the first fold served as the validation set. This process was repeated ten times with distinct validation subsets. To evaluate the performance of the regression models using this validation method, we employed various metrics, including the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Bias Error (MBE).

4.2.4.6 Uncertainty associated with the model

The study of uncertainty in statistical modeling is essential in understanding the model prediction reliability and precision. This uncertainty arises from various sources, including measurement errors, inaccuracies in model specification, and the innate variability present in the data (Zhen, et al. 2022). Comprehending and quantifying this uncertainty is vital for informed decision-making based on model outputs. Error propagation, another critical aspect of modeling, refers to how uncertainties in input data or model parameters affect the output through the modeling process. This concept becomes particularly significant in a cascaded modeling scenario, where the output of one model serves as the input for another (Molto, et al. 2013). To understand this in our

study, we adopted a three-step Monte Carlo simulation approach to evaluate error propagation across a series of models, each predicated upon the output of its predecessor. Step 1: Involves simulating understory biomass using TLS variables. We identified gamma and normal distributions as representative of the LiDAR metrics and utilized these distributions accordingly. For this phase, we generated 5000 simulated datasets, where the simulated biomass $y_{\text{pre, TLS}}$ is computed as:

$$y_{\text{pre, TLS}} = \beta_{0,\text{TLS}} + \beta_{1,\text{TLS}} \times x_1 + \beta_{2,\text{TLS}} \times x_2 + \epsilon_{\text{TLS}} \quad (4.1)$$

Here, $\beta_{0,\text{TLS}}$ is the intercept, $\beta_{1,\text{TLS}}$ and $\beta_{2,\text{TLS}}$ are the coefficients for x_1 and x_2, respectively, and ϵ_{TLS} represents the normally distributed noise values.
We introduced a realistic level of uncertainty by adding normally distributed noise to the predicted values, with a mean of zero and a standard deviation equal to the actual model residual standard deviation (σ). This process created a set of simulated observed (after adding normally distributed random noise in the predicted values) and predicted values(values before adding noise) of understory biomass.

Step 2: Utilizes the simulated biomass predictions from Step 1 in addition to all other UAV-LS-based and ALS based variables in the UAV-LS-based and ALS based upscaled models, respectively. The predicted output from these models $Z_{\text{pre, UAV}}$ and $Z_{\text{pre, UAV}}$ are given by:

$$Z_{\text{pre, UAV}} = \beta_{0,\text{UAV}} + \beta_{\text{Y, UAV}} \times y_{\text{pre, TLS}} + \sum \left(\beta_{\text{i, UAV}} \times \text{var}_{\text{i, UAV}}\right) + \epsilon_{\text{UAV}} \quad (4.2)$$

$$Z_{\text{pre, ALS}} = \beta_{0,\text{ALS}} + \beta_{\text{Y, ALS}} \times y_{\text{pre, TLS}} + \sum\left(\beta_{\text{i, ALS}} \times \text{var}_{\text{i, ALS}}\right) + \epsilon_{\text{ALS}} \quad (4.3)$$

where $\beta_{0,\text{UAV}}$ and $\beta_{0,\text{ALS}}$ are the intercepts, $\beta_{\text{Y, UAV}}$ and $\beta_{\text{Y, ALS}}$ are the coefficient for the TLS models biomass prediction and ϵ_{UAV} and ϵ_{ALS} are the normally distributed error terms (mean of zero and a standard deviation equal to the actual models residual standard deviation (σ)) for the UAV-LS and ALS variables-based models, respectively.

Step 3: Incorporate predictions from the UAV-LS model ($Z_{\text{pre, UAV}}$) in a UAV to ALS variables-based model (UAV-ALS) with other ALS-specific variables in the model. The output for this model ($W_{\text{pre, UAV-ALS}}$) is calculated as:

$$W_{\text{pre, UAV-ALS}} = \beta_{0,UAV-\text{ALS}} + \beta_{\text{Y, UAV-ALS}} \times Z_{\text{pre, UAV}} + \sum\left(\beta_{\text{i, UAV-ALS}} \times \text{var}_{\text{i, UAV-ALS}}\right) + \epsilon_{\text{UAV-ALS}} \quad (4.4)$$

Here, $\beta_{0,UAV-\text{ALS}}$ is the intercept, $\beta_{\text{Y, UAV-ALS}}$ is the coefficient for the UAV model prediction from equation 4.2 and $\epsilon_{\text{UAV-ALS}}$ is the normally distributed error term with a mean of zero and a standard deviation equal to the actual UAV-ALS model residual standard deviation (σ).

Throughout all three steps, new simulated data for variables in each model were generated, aligning closely with their real-data distributions. The first model predicted values were used as inputs for the subsequent UAV and ALS models, and similarly, the UAV model outputs were employed as inputs for the UAV-ALS model. This cascading

approach, incorporating the Monte Carlo method, enabled the quantification of error propagation, thereby offering a comprehensive estimation of cumulative error across the modeling chain. We quantified this by calculating the absolute differences between the predicted and observed values (generated by adding normally distributed noise to the predicted values) for the UAV model. This procedure provided insights into how initial errors from the TLS-based model predictions could influence the accuracy and reliability of the UAV-based model predictions.

Finally, we computed and summarized the propagated error mean and standard deviation. These metrics offered a comprehensive understanding of the error transmission from the first to the second and third models, highlighting the potential impact of initial prediction inaccuracies on subsequent modeling stages.

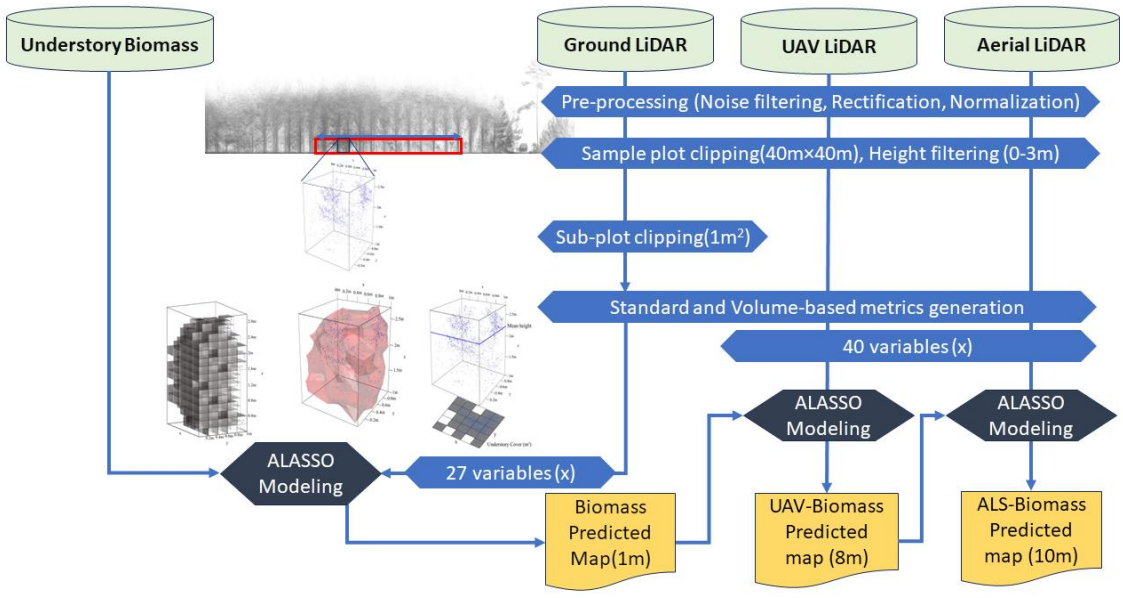

Figure 4.4: Flow chart of the data processing and modeling

4.3. Result

4.3.1 Summary statistics

In the destructively sampled 100 subplots of 1 m^2 from 60 sample plots, the understory vegetation biomass varied between 0.08 kg to 2.87 kg, with a mean of 0.85 kg (81 kg to 28,780 kg, with a mean of 8561 kg per hectare). Similarly, the maximum mean height of the TLS point clouds within a subplot was 2.47 m, the minimum was 0.26 m, and the mean was 0.97 m. Likewise, the understory vegetation biomass in a predicted map based on a TLS-derived model ranged between 0 kg to 1.58 kg, with a mean of 0.56 kg (0 kg-15,778 kg and a mean of 5,625 kg per hectare).

Additionally, the predicted map of understory vegetation biomass, derived from a TLS-based model, displayed a range from 0 kg to 1.58 kg, with a mean of 0.56 kg (ranging from 0 kg to 15,778 kg per hectare and a mean of 5,625 kg per hectare). Furthermore, the average height observed from the ALS point clouds within the understory layer (0 - 3 meters), was 0.7 meters, with a maximum height of 2.9 meters. For UAV-LS, the average height was 0.5 meters, with a maximum of 2.7 meters. Moreover, table 5 below provides a concise overview of the statistical parameters related to the major ALS and UAV-LS-based variables utilized for model upscaling. For a detailed statistical summary of field-measured understory biomass and TLS metrics, refer to Adhikari, et al. (2023b).

Table 4.5: Summary statistics of TLS predicted biomass map and major ALS and UAV-LS variables at 8 m resolution. The difference in TLS predicted biomass summary statistics for ALS and TLS is due to the different sample sizes (no. of pixels).

	Mean		**Median**		**SD**		**Minimum**		**Maximum**	
	ALS	**UAV**	**ALS**	**UAV**	**ALS**	**UAV**	**ALS**	**UAV**	**ALS**	**UAV**
Biomass (kg) (TLS predicted)	35.9	40.3	35.1	39.8	19.6	18.4	2.4	3.0	98.9	89.0
Zmax (m)	2.9	2.7	3.0	2.8	0.1	0.3	2.0	2.0	3.0	3.0
Zmean (m)	0.7	0.5	0.6	0.4	0.3	0.3	0.1	0.1	2.0	1.6
Zsd (m)	0.6	0.4	0.6	0.4	0.2	0.2	0.2	0.1	1.1	1.1
PzaboveZmean (%)	35.5	37.1	37.0	37.5	12.2	10.2	9.9	8.1	60.7	67.3
Zpcum2 (%)	59.9	71.9	62.4	77.8	25.1	24.0	3.0	0.5	99.1	100.0
Zpcum6 (%)	89.8	94.8	92.0	97.7	8.9	7.8	28.1	38.5	100.0	100.0
Voxelvol (m^3)	114.2	122.1	116.0	123.0	31.3	34.2	41.0	35.0	192.0	192.0
Wt_voxelvol	20.0	15.9	17.5	13.4	11.7	11.4	0.7	0.5	65.2	73.4
log_wt_voxelvol	2.8	2.5	2.9	2.6	0.7	0.8	-0.4	-0.7	4.2	4.3

4.3.2 TLS-variable-based biomass estimation model

The Adaptive LASSO method identified two out of the twenty-seven variables as significant contributors to the model (eq. 2). The adaptive LASSO-based model exhibited favorable performance measures in calibration and cross-validation datasets.

$$Y = 0.057 + 0.885 \times x_1 + 0.169 \times x_2 \quad (4.5)$$

Where Y is understory biomass in kilograms per square meters, x_1 is TLS point cloud-based volume calculated based on the mean height and understory cover method

described by Adhikari, et al. (2023b) and x_2 is the 20th percentile of echo height calculated from a TLS point cloud covering 1 m^2 horizontal and 3 m vertical extent. This model achieved an Adjusted R^2 of 0.84 in the calibration dataset with an RMSE of 233 g. The model yielded an Adjusted R^2 of 0.8 and an RMSE of 235 g in the k-fold cross-validation dataset (k=7). These results highlight the predictive accuracy of the Adaptive LASSO model, suggesting its effectiveness in capturing the relationships between the selected variables and the biomass of the understory. The cross-validation result shows the model has a slight positive bias of 0.44, suggesting a tendency for the model to overestimate the understory biomass when it falls below 500 g and underestimate it when it exceeds 1500 g. However, the model demonstrates reasonably accurate predictions within the 500 to 1500 g range. Additionally, the residuals are randomly distributed, indicating that the model adequately captures the variability in the data with no apparent patterns or systematic deviations. For further details and plots for this model performance, refer to Adhikari, et al. (2023b).

4.3.3 Variable correlation

The Pearson correlation between TLS-predicted understory vegetation biomass with the highest correlated ten metrics generated using UAV-LS is presented in Fig. 4.5. For UAV-LS derived metrics, the standard deviation of the echo height had the highest correlation coefficient (r=0.75) with the TLS-predicted understory biomass, followed by 95th percentile (r=0.68). Similarly, in case of ALS derived metrics volume based on voxelization method at a 1 m voxel resolution showed the highest correlation coefficient (r=0.76) followed by its log transformed value (r=0.75) with the TLS-predicted understory biomass.

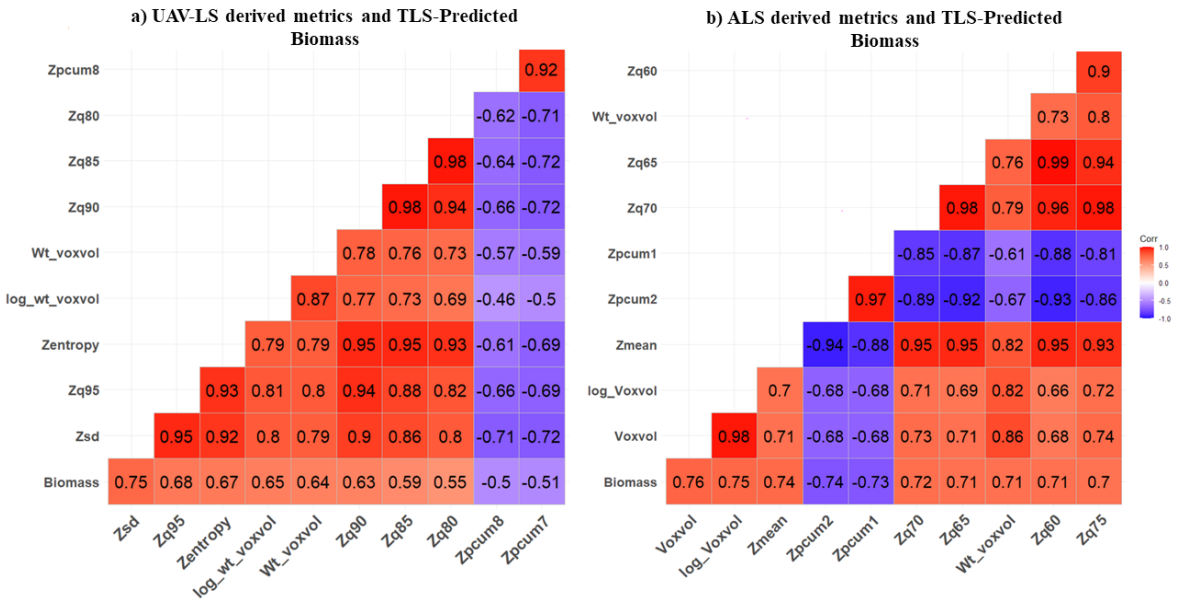

Figure 4.5: Pearson correlation coefficients between TLS-predicted understory biomass and the ten highest correlated UAV-LS derived metrics (a) and ALS-derived metrics (b).

4.3.4 UAV-LS variable-based upscaled model

In the UAV-LS upscaled model, out of the 40 variables, only eight were found most significant and retained in the model, which is shown in Table 4.6 with their coefficient values and intercept. Among those eight, echo height entropy, cumulative percentile proportion at the ninth layer, height standard deviation, and weighted voxel volume were particularly influential in predicting biomass.

The model achieved the adjusted R^2 value of 0.61 and a Root Mean Square Error (RMSE) of 11.37 kg with the calibration dataset. This performance was consistently reflected in the k-fold cross-validation, where we obtained an adjusted R^2 of 0.6 and an RMSE of 11.46 kg across the 10-fold validation set. Notably, the model exhibited a

negligible bias (−0.0166), underscoring the absence of systematic error and confirming the predictive reliability of the model.

Figure 4.6 shows the association between observed and predicted biomass values, illustrating a satisfactory agreement despite some scatter, indicative of the model accuracy in capturing the majority of the variability in the data. Furthermore, Figure 4.7 displays the residual plot, which shows no distinct patterns, indicating that the residuals are randomly distributed. These findings underscore the utility of the selected variables and the potential of UAV Lidar technology in estimating understory biomass at a fine spatial resolution.

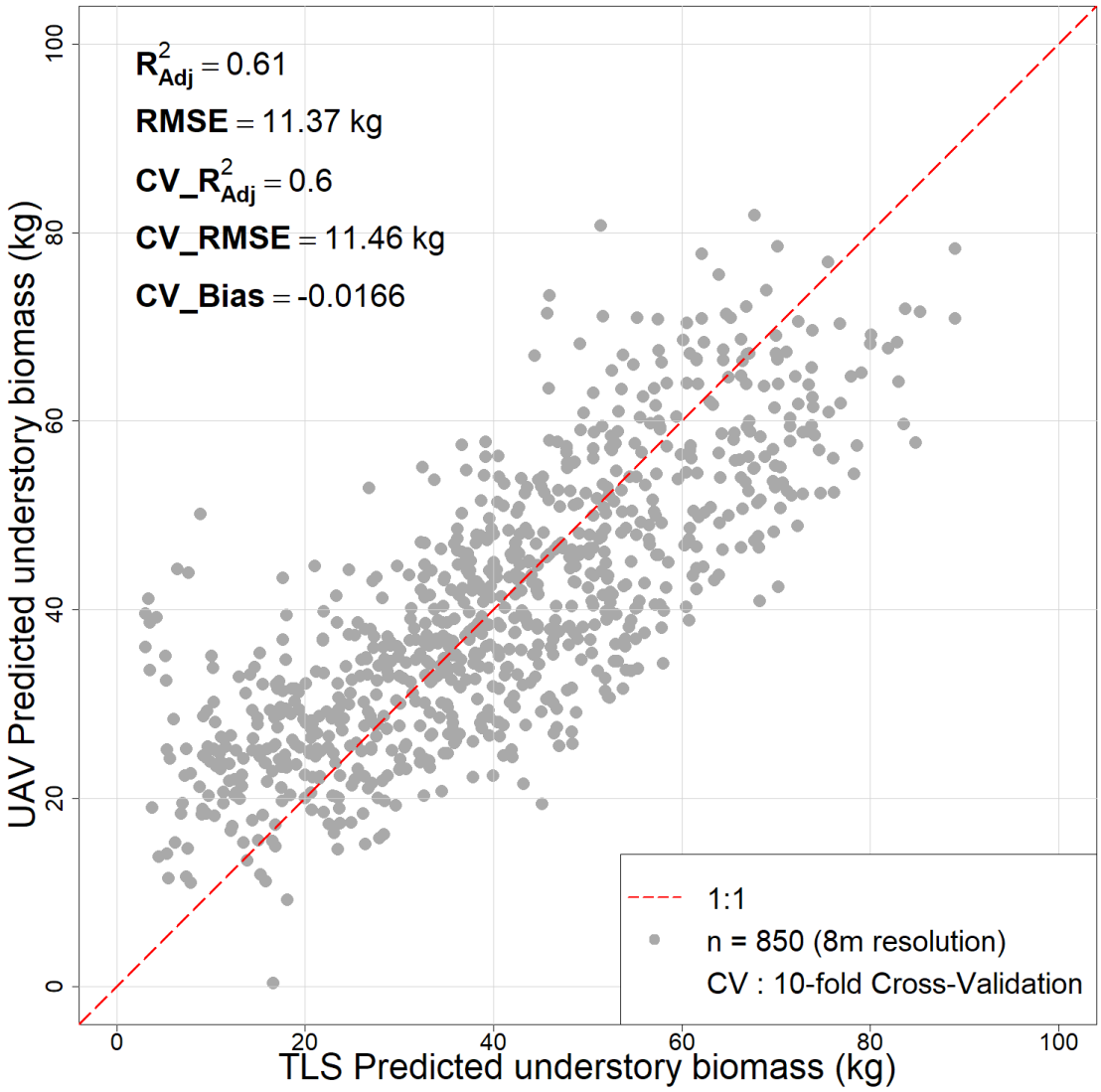

Figure 4.6 Observed vs Predicted plot comparing UAV Lidar-based predicted understory biomass (kg) against TLS-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.61 and RMSE of 11.37 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.6 and RMSE of 11.46 kg).

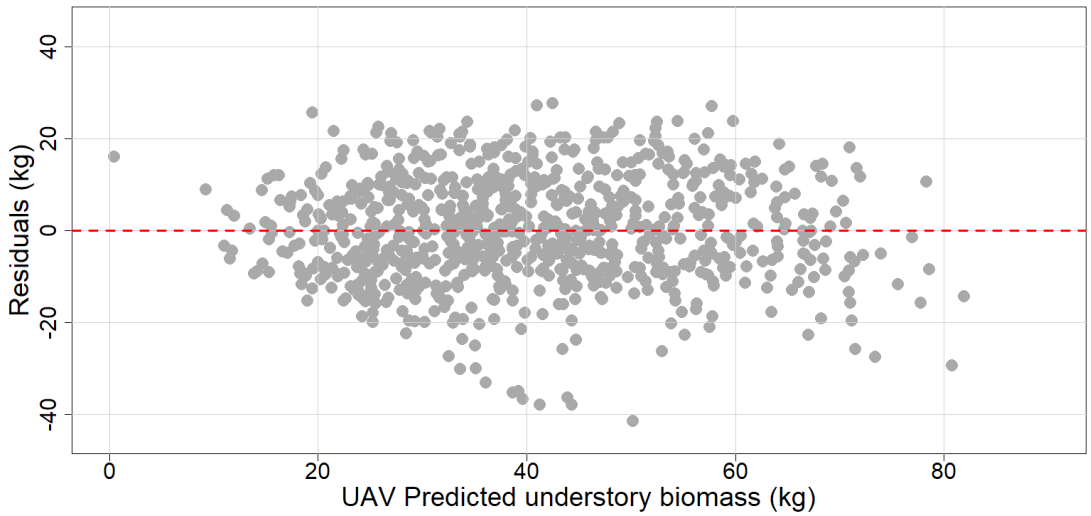

Figure 4.7 Residuals plot for the UAV Lidar-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions.

4.3.5 ALS variable-based upscaled model

In comparison to the variables in the UAV-LS based model, the ALS variable-based model demonstrates superior accuracy and robustness in predicting understory biomass, despite the low density point cloud. Where the UAV-based model required eight variables to achieve its predictive accuracy, the ALS model retained only five variables. This reduction in complexity not only streamlines the modeling process but also suggests that the selected variables in the ALS model are more influential, capturing the essential information needed for accurate biomass estimation. All five variables included in the model and their coefficient are listed in Table 4.6. Among those five, the voxel-based volume and cumulative proportion at the second layer were found to be the most

influential variables, underscoring their importance in predicting understory biomass using ALS data.

Figure 4.8 shows the observed versus predicted plot between TLS-predicted and ALS-predicted biomass, demonstrating the strong agreement between the TLS and ALS-predicted biomass values. A higher adjusted R-squared value of 0.69 in the ALS model indicates that it explains a significant proportion of the variability in the data, which is an improvement over the UAV-based model. Moreover, the near-identical cross-validation metrics (Adj. R^2 of 0.67 and RMSE of 10.75 kg) to the calibration statistics indicate that the ALS model predictive performance is consistent and reliable across different subsets of data, pointing towards its robustness in various scenarios. Moreover, the residual plot in Figure 4.9 further confirms the precision of the ALS model. The residuals, which represent the discrepancies between observed and predicted values, are scattered around the zero line with no apparent patterns that would indicate systematic errors. The randomness in the distribution of residuals indicates the model robustness. It suggests that the ALS model predictions are not biased.

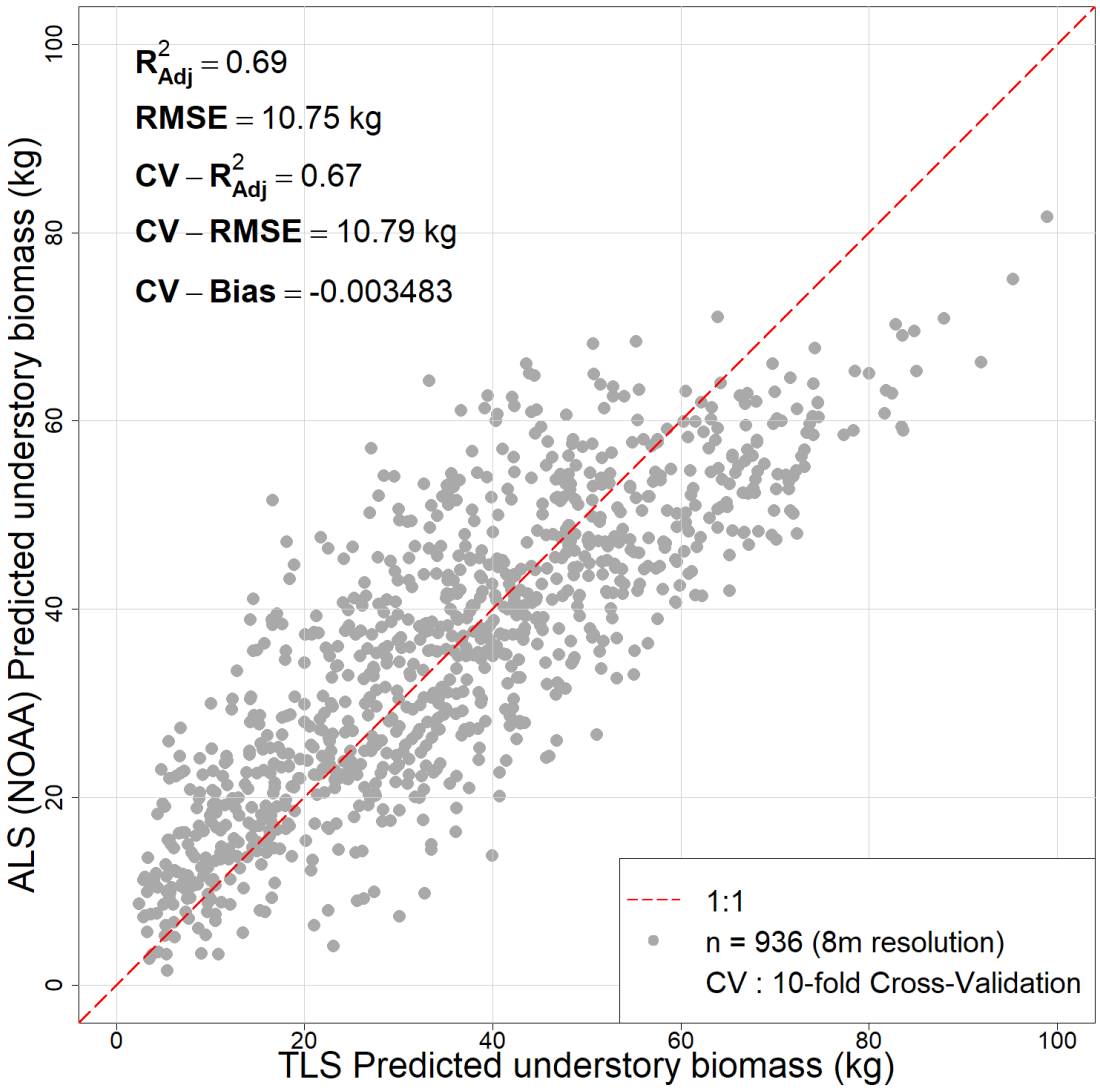

Figure 4.8 Observed vs Predicted plot comparing ALS-based predicted understory biomass (kg) against TLS-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.69 and RMSE of 10.75 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.67 and RMSE of 10.79 kg).

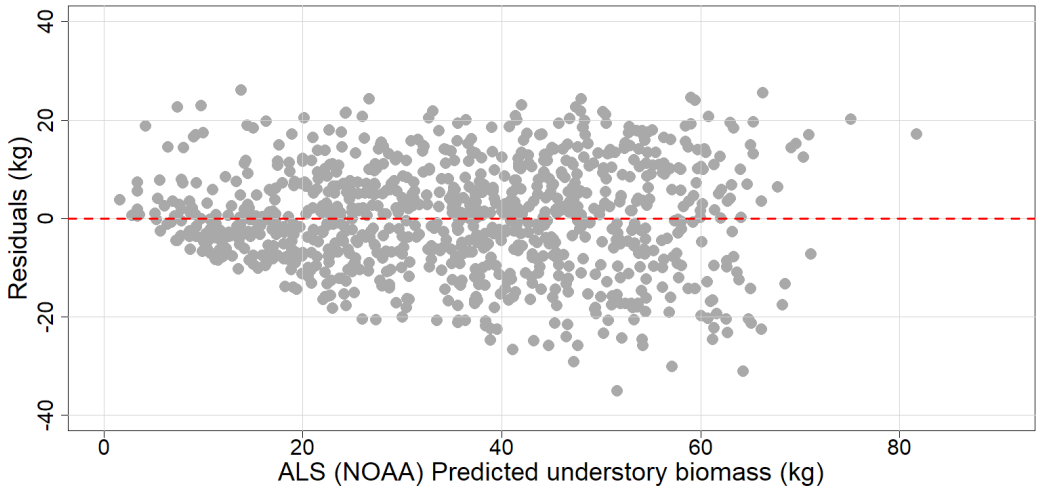

Figure 4.9 Residuals plot for the ALS-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions.

4.3.6 UAV-LS to ALS upscaled model

The two-step upscaling model, incorporating TLS, UAV-LS, and ALS data, demonstrated satisfactory predictive performance. The model achieved an adjusted R^2 value of 0.71 and a RMSE of 9.607 kilograms in the calibration data, indicating a strong correlation between the observed and predicted values. Similarly, during the cross-validation process, which employed a 10-fold technique, the model consistently displayed an adj R^2 of 0.70 and an RMSE of 9.641 kilograms, highlighting the reliability of the model across diverse datasets.

Of the six retained variables within this model, the most influential were the 60^{th} percentile of height, height entropy, and the cumulative proportion of height at the seventh layer. These parameters describe the vertical structure of the understory biomass and provide significant predictive power. Additionally, height entropy and volume based

on voxel count were identified as common variables in both the TLS to UAV-LS and UAV-LS to ALS models, establishing them as consistent indicators across multiple scales of laser scanning technologies. A list of all selected variables and their coefficients is shown in Table 4.6.

Furthermore, Figure 4.10 shows the observed vs predicted plot that compares the understory biomass predicted by the UAV (on the x-axis) against the ALS-predicted understory biomass (on the y-axis). Moreover, the residual plot in Figure 4.11 shows residuals scattered around the zero line with no apparent patterns that would indicate systematic errors. The randomness in the distribution of residuals indicates the robustness of the model.

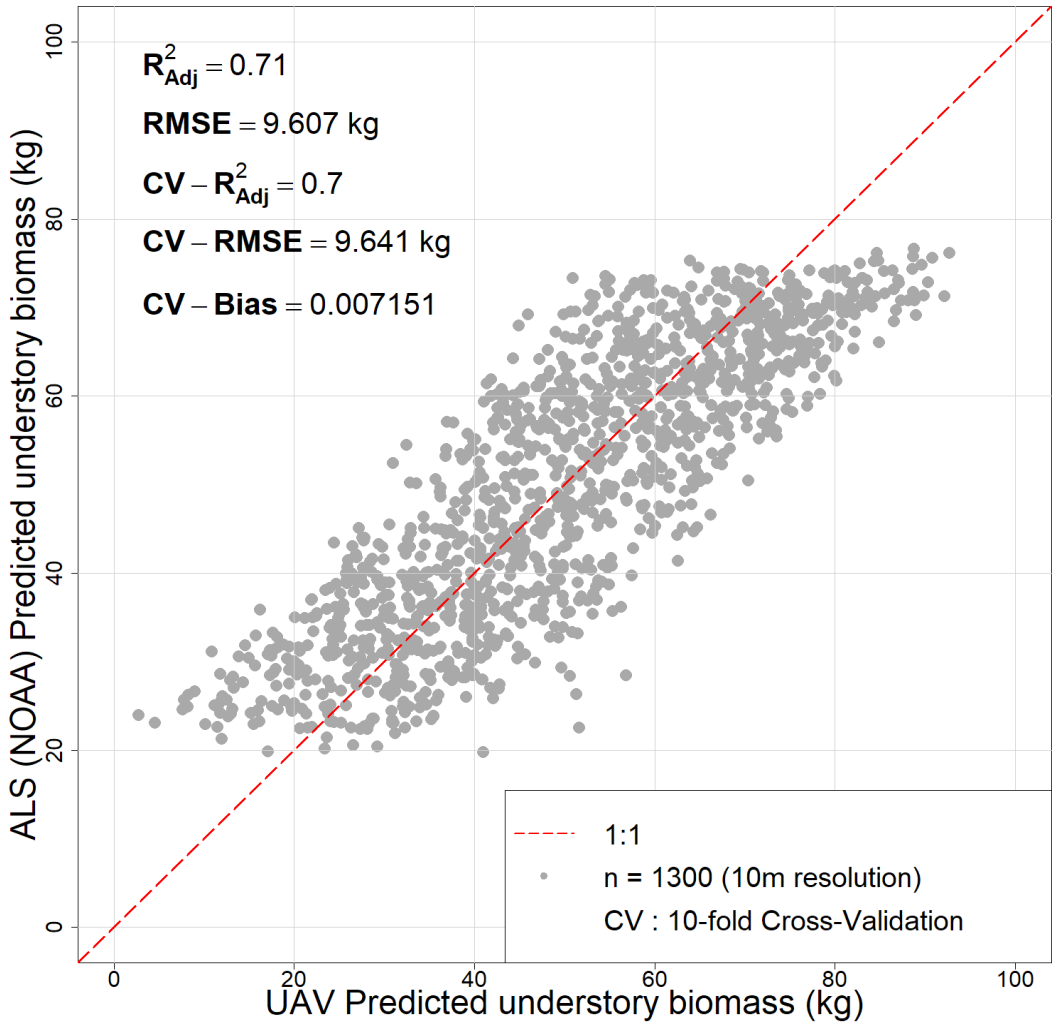

Figure 4.10 Observed vs Predicted plot showing ALS-based predicted understory biomass (kg) against UAV-derived estimates, demonstrating the model prediction accuracy with an adjusted R^2 of 0.71 and RMSE of 9.607 kg on the calibration dataset, along with the results of 10-fold cross-validation (adjusted R^2 of 0.70 and RMSE of 9.641 kg).

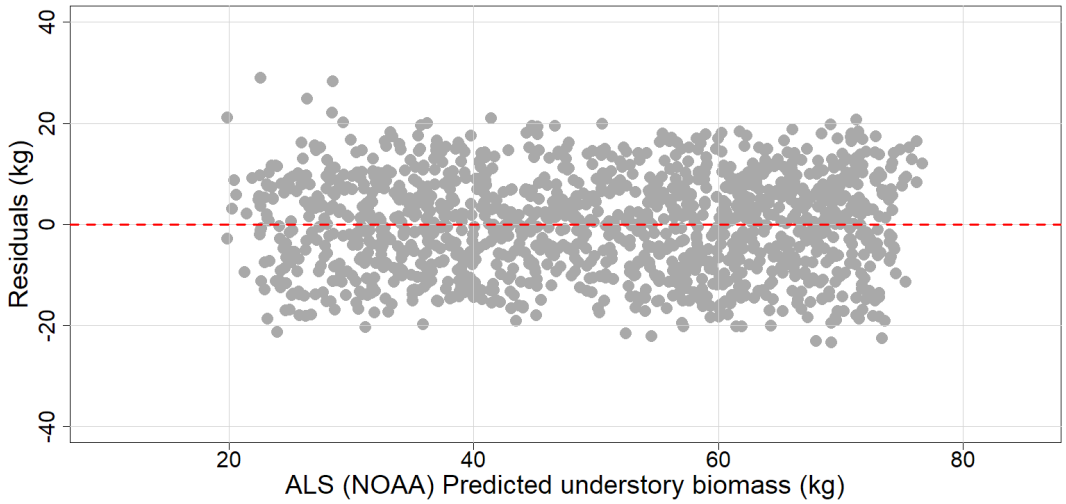

Figure 4.11 Residuals plot for the UAV-LS to ALS-based understory biomass prediction model, displaying the distribution of residuals against the predicted values, with the dashed red line indicating the zero residual level to highlight the random scatter and absence of systematic bias in model predictions.

4.3.7. Selected variables and their correlation

The comparative analysis of the selected variables to biomass in all three models indicates a more robust correlation and lower dispersion of data points when utilizing ALS-based metrics, as shown in Figure 4.13. This is particularly evident with voxel-based volume metrics and the cumulative height proportions at the first and second layers. Conversely, metrics derived from Unmanned Airborne Vehicle Laser Scanning (UAV-LS) exhibit a comparatively weaker correlation with understory biomass, accompanied by a higher scatter, as depicted in Figure 4.12. Furthermore, the ALS-derived metrics demonstrate a moderate correlation with the understory biomass, signifying that while there is some degree of relationship, the predictive capacity may be variable and subject to influencing factors not captured within this dataset. ALS high positioning accuracy and multiple returns in capturing complete 3D profile structural

details may account for the enhanced correlation observed in ALS metrics. In contrast, the UAV-LS may be limited in detecting subtle variations within understory biomass due to its dual return sensor not sufficiently capturing a complete 3D profile and low positioning accuracy, as shown in Figure 4.14.

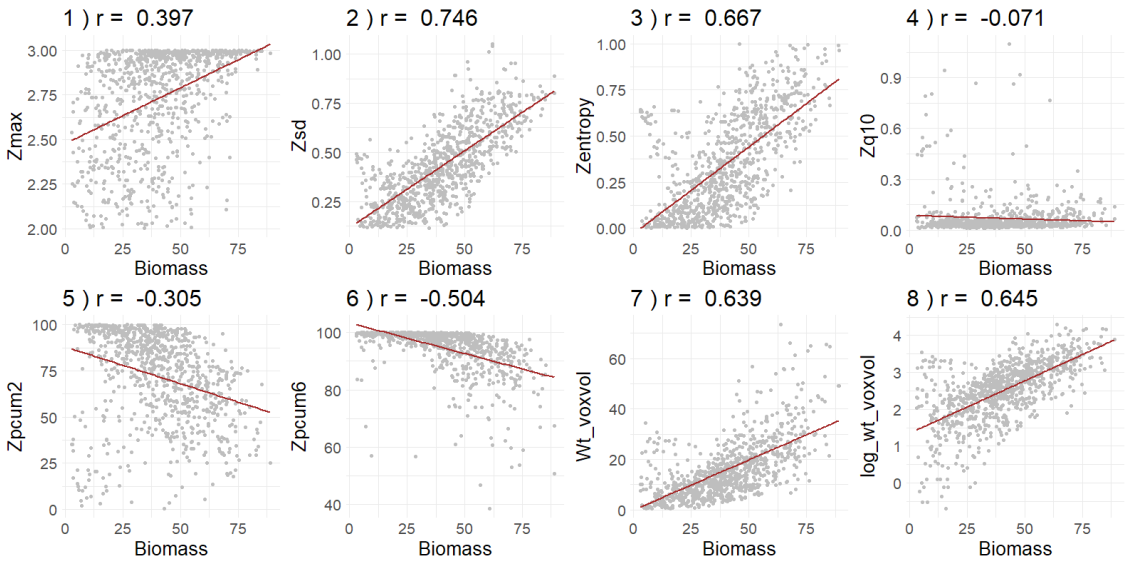

Figure 4.12: Variables selected in the UAV-LS-based model and their correlation with the TLS-based biomass.

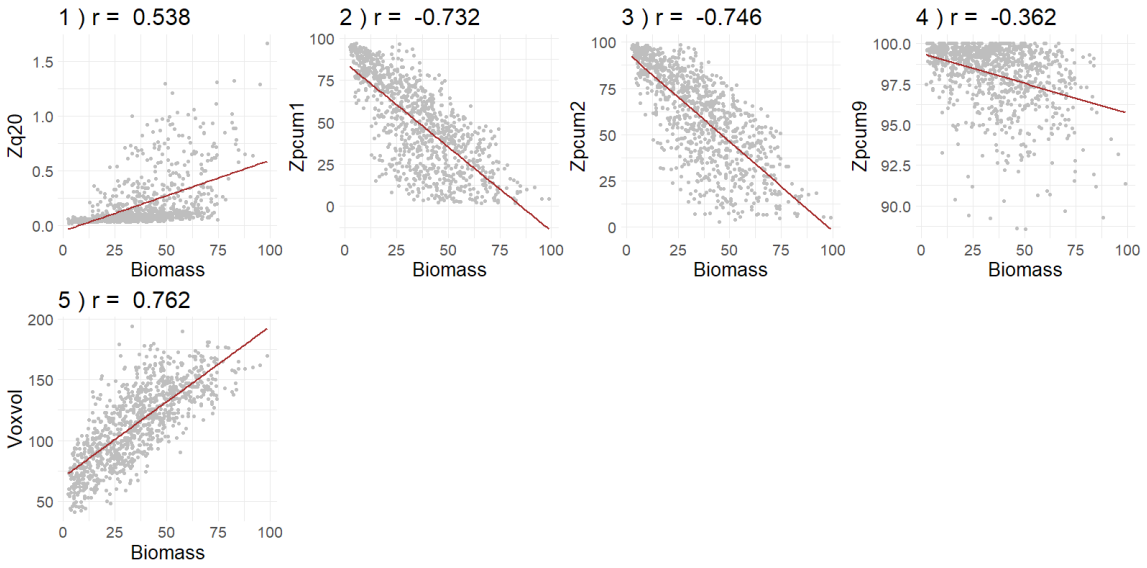

Figure 4.13 Variables selected in the ALS-based model and their correlation with the TLS-based biomass.

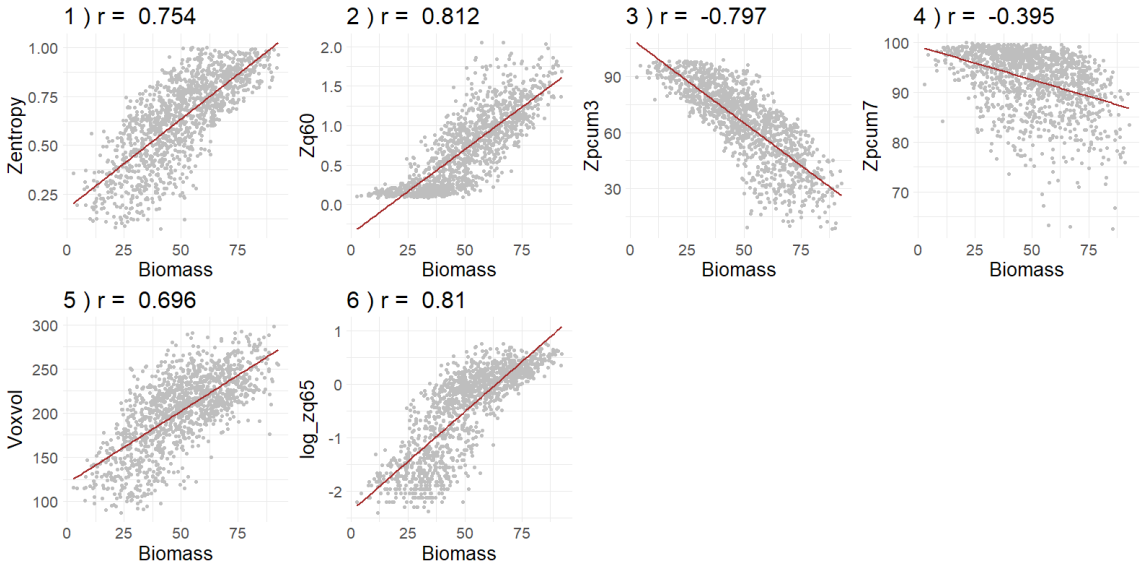

Figure 4.14 Variables selected in the ALS-based model and their correlation with the UAV-LS-based biomass.

	Coefficient		
Variables	TLS-UAV	TLS-ALS	UAV-ALS
Intercept	40.2749	35.5461	49.98
Heigh maximum	0.8956		
Heigh standard deviation	3.9823		

Height entropy	8.7523		5.9979
Height 10th percentile	-1.8958		
Height 20th percentile		1.5696	
Height 55th percentile			
Height 60th percentile			7.7959
Cumulative proportion at 1st layer		-1.6866	
Cumulative proportion at 2nd layer	6.0137	-4.9306	
Cumulative proportion at 3rd layer			-0.3253
Cumulative proportion at 6th layer	-1.5463		
Cumulative proportion at 7th layer			3.0756
Cumulative proportion at 9th layer		-1.8606	
Volume based on voxel count		8.9753	1.1299
Weighted voxel volume by count	0.8019		
Natural log of weighted voxel volume	3.8849		
Natural log of Height 65th percentile			2.5852
Total no. of variables	8	5	6

4.3.9. Uncertainty of understory biomass estimation

The uncertainty in the TLS-based model and the subsequent three upscaled models was assessed using a Monte Carlo simulation conducted 5000 times. This number of simulations was adequate for understory biomass uncertainty estimation, as the estimates stabilized after about 2800 simulations.

In the TLS model, biomass predictions ranged from 0.71 mg/ha to 26.02 mg/ha, with an average of 8.603 mg/ha. The associated uncertainty in this model was, on average, 1.62 mg/ha, representing approximately 14.3% of the mean TLS biomass predictions.

For the TLS to UAV-LS model, biomass predictions varied between 0.08 mg/ha and 12.81 mg/ha, with an average of 6.30 mg/ha. The uncertainty in these predictions had an average of 1.50 mg/ha, making up 17.9 % of the UAV-LS biomass predictions, showing an increase in uncertainty from the TLS model.

The TLS to ALS model displayed biomass predictions ranging from 0.26 mg/ha to 12.07 mg/ha, with an average of 6.08 mg/ha. The uncertainty in this model had an average of 1.35 mg/ha, representing 16.7 % of the mean ALS biomass predictions, slightly better than the UAV model.

In the two-step upscaled model (UAV-LS) to the ALS model, the biomass predictions averaged 5.56 mg/ha, ranging from 0.25 mg/ha to 12.79 mg/ha. The uncertainty in this combined model was higher, with an average of 1.46 mg/ha, accounting for 19.83% of the mean biomass predictions, the highest among all the models.

These results show the increase in uncertainty through each stage of the modeling process, illustrating the complexities of error propagation in cascaded modeling frameworks. This analysis is essential for evaluating the reliability and constraints of predictive models in environmental and ecological research. A summary of the MC simulation result is shown in Table 4.7.

Table 4.6. The estimated understory biomass and its uncertainty after Monte Carlo simulation (Unit: $Mg.ha^{-1}$)

Estimate	Min	Mean	Median	Max	SD	%
TLS_biomass	0.71	8.60	7.88	26.02	5.71	
TLS_uncertainty	0.00	1.62	1.37	9.23	1.23	14.3
UAV_biomass	0.08	6.30	6.04	12.81	2.24	
UAV_uncertainty	0.00	1.50	1.28	8.18	1.13	17.9
ALS_biomass	0.26	6.08	5.77	12.07	2.51	
ALS_uncertainty	0.00	1.35	1.14	6.82	1.02	16.7
UAV_ALS biomass	0.25	5.56	5.65	12.79	2.50	
UAV_ALS uncertainty	0.00	1.46	1.55	6.82	1.10	19.8

4.4. Discussion

This study generated three efficient models for predicting understory biomass, employing LiDAR data from terrestrial (TLS), UAV, and airborne (ALS) platforms. These models were developed using a combination of upscaling approaches 1) from TLS-derived estimates to UAV-lidar, 2) from TLS-derived estimates to ALS-lidar, and 3) from UAV-derived estimates to ALS-lidar. They all exhibited commendable levels of accuracy, as indicated by adjusted R-squared (Adj. R^2) values of 0.61, 0.69, and 0.71, and RMSE of 1.78 $mgha^{-1}$, 1.68 $mgha^{-1}$, and 1.50 $mgha^{-1}$, respectively with acceptable range of uncertainty. The utility of these models lies in their precise quantification and prediction capabilities for understory biomass across different plantation stands, offering resolutions from eight to ten meters. Employing these models allows for an effective, non-destructive method of assessing understory vegetation, which proves advantageous over traditional, labor-intensive, and costly field measurements, particularly within the understory complex structure in the coastal plains of the southeastern United States.

In the pine plantations of this region, understory vegetation predominantly consists of evergreen shrubs combined with hardwood trees. This composition has presented notable assessment challenges with conventional field methods and remote sensing techniques, including Landsat imagery and airborne LiDAR (Campbell, et al. 2018; Peduzzi, et al. 2010). However, as documented in the literature, the TLS-based approach has been recognized for its superior accuracy, speed, and precision over field-based methods. Data collection, including initial sensor calibration and post-scan data logging, can be completed in 10 to 30 minutes, depending on the understory density (Li, et al. 2021). The scan range of up to 100 meters and the very high point density provided by TLS

contribute to its ability to accurately detect subtle variations in understory vegetation, enabling detailed predictions at a one-meter resolution at the stand level.

Despite the proven accuracy of TLS-based predictions in this study and others, its application over extensive areas remains impractical. This limitation highlights the significance of the upscaled models proposed herein, which enhances the scale of LiDAR-based assessments of understory vegetation. The models presented in this research extend benefits beyond the domain of plantation management, where they facilitate the monitoring of interspecific competition between overstory and understory vegetation using either UAV or ALS data. Furthermore, results from this study have the potential to contribute to ecological research efforts by providing valuable tools for conducting understory carbon evaluations and wildlife habitat analysis. These models, therefore, represent a substantial progression in our capability to manage and study forest ecosystems with greater efficacy and ecological sensitivity.

To develop a more generalizable and robust model, meaningful variables must be generated and fed into the model. In this study, we carefully selected the variables suggested by the literature, such as voxel and alpha-hall fitting-based volume (Cooper, et al. 2017). Considering the nature of the data, we generated a few new variables, such as mean height understory cover-based volume and weighted voxel count. Broadly, we incorporated three types of data: first, statistical summary metrics and percentiles; second, canopy density metrics; and third, volume-based metrics. Our result shows that volume-based metrics are more correlated when estimating understory biomass. That is because the high data density approximates an understory structure size and near-to-exact volume. This observation holds consistency across UAV-LS and ALS data, with volume-

based metrics emerging as substantial contributors to model performance. Furthermore, canopy density and percentile metrics have been identified as vital across all models, enhancing the comprehensiveness and predictive accuracy of the study.

The findings from our study underscore the importance of estimating uncertainty, particularly in upscaled models used for understory biomass prediction. The assessment of uncertainty in the TLS-based model and its subsequent upscaled iterations using a Monte Carlo simulation reveals a progressive increase in uncertainty at each modeling stage. This heightened uncertainty in the upscaled models—ranging from 14.3% in the TLS model to 19.83% in the combined UAV-LS to ALS model—highlights the challenges inherent in cascading modeling frameworks.

A notable contributor to this increased uncertainty is the absence of species-specific allometric models (Zhen, et al. 2022). Our approach, which primarily involves comparing the volume and cover of understory vegetation with its weight, does not account for the variability in weight among different species despite similar volumes and sizes. This lack of specificity likely contributes to the observed discrepancies in biomass predictions. Moreover, discrepancies of up to 30% are common in other similar studies utilizing lidar for accessing understory vegetation ((Li, et al. 2021; Venier, et al. 2019).

However, it's crucial to recognize that the primary goal in estimating understory biomass in plantation forests is not precise quantification but rather a rough estimation. This estimation is vital for identifying areas with substantial understory growth, which can then be targeted for cleaning operations. Thus, while accuracy in growth estimation is valuable, the emphasis in this context is more on the utility of the models for practical

forest management, where identifying high understory areas is more critical than precise biomass quantification.

Our study examines the four-year temporal gap between ALS data and other datasets, focusing on the consistent nature of understory vegetation in southeastern plantation forests, particularly after silvicultural activities. In intensively managed loblolly pine plantation forests, the canopy typically closes around mid-rotation age, which aligns with the minimum age of our sampled stands (Campbell, et al. 2013). Research in this area indicates that both the growth and coverage of understory vegetation tend to stabilize post mid-rotation.

Specifically, studies by Archer, et al. (2007) on understory vegetation changes after clearcut regeneration in similar forests reveal that the cover and composition of shrub and herbaceous species remain largely constant beyond mid-rotation age. This consistency is a key factor in our decision to include a four-year gap in our analysis. The growth and changes in understory vegetation cover are generally minimal during this period, especially in the later stages of clearcut regeneration.

Moreover, the use of ALS data in our research comes with inherent limitations in temporal resolution. In studies of overstory biomass, it is common to see a temporal gap of up to four years, mainly due to data availability constraints (Zhen, et al. 2022). Although decreasing this gap could improve the accuracy of our predictive maps, it's essential to balance this potential improvement against the practical challenges and the known stability of understory vegetation in these specific contexts.

Estimating above-ground biomass has traditionally prioritized overstory trees at the expense of understory vegetation, thus neglecting a critical ecosystem component (Gonzalez, et al. 2013). In the context of accelerating environmental concerns, understanding and quantifying the total above-ground biomass (AGB) is more crucial than ever. Biomass, particularly in forest ecosystems, plays a significant role in the global carbon cycle, acting as a major carbon sink and thus mitigating the effects of climate change. Accurate estimation of global biomass is essential for monitoring carbon dynamics, assessing the health of ecosystems, and implementing effective climate change mitigation strategies. Underestimating or overlooking components such as understory vegetation can lead to significant gaps in our understanding of the total carbon storage in forests.

Therefore, developing an efficient methodology for estimating biomass within the understory layer is crucial in above-ground biomass estimation. This is particularly important considering recent and forthcoming satellite missions, such as BIOMASS, GEDI, NISAR, ICESat, and ALOS-4, based on forest AGB estimation (Duncanson, et al. 2019). They represent a concerted effort to provide detailed and accurate measurements of forest biomass on a global scale. The data from these missions are vital for enhancing our understanding of forest carbon stocks and dynamics. The need for accurate calibration and validation across hectare-scale plots becomes imperative to align with these sensors footprints, ensuring the data collected from space accurately reflects ground realities (Brede, et al. 2022; Duncanson, et al. 2019).

The multi-scale lidar-based models developed in this study, incorporating traditional destructive sampling, TLS, UAV-LS, and ALS point clouds, offer an advanced, non-destructive possibility to accurately estimate understory biomass and include it in forest AGB assessment. These technologies provide a means to bridge the gap between ground-based measurements and satellite observations, ensuring the comprehensive inclusion of all forest layers in biomass estimations. Significantly, applying our upscaled models to UAV or ALS data is an ideal calibration tool for these global sensors, thereby contributing to the refinement of biomass estimation at a global scale.

4.5. Conclusion

In conclusion, while TLS offers precise understory biomass estimates, its practicality is hindered by its limitation to small areas. As explored in this study, upscaling models hold significant potential for informing plantation forest management and decision-making through more accessible platforms like UAV and ALS. Although both platforms yielded satisfactory accuracy, the ALS-based models were especially noteworthy for their efficiency, relying on only five variables to deliver high accuracy and robustness. This enhances the ALS model user-friendliness and highlights its capture of critical biomass-influencing variables. UAV applicability, by contrast, is restricted by lower pulse return, which affects understory assessment. An additional limitation of this study is the four-year data acquisition gap between ALS and other LiDAR and biomass data, which likely introduced some bias. Despite this, ALS demonstrated superior performance, suggesting that its accuracy could be further enhanced with reduced temporal gaps between data collections. This aspect warrants further investigation.

4.6 References

1. Adegbidi, H., Jokela, E., Comerford, N. and Barros, N. 2002 Biomass development for intensively managed loblolly pine plantations growing on Spodosols in the southeastern USA. *Forest Ecology and Management*, **167** (1-3), 91-102.

2. Adhikari, A., Montes, C.R. and Peduzzi, A. 2023a A Comparison of Modeling Methods for Predicting Forest Attributes Using Lidar Metrics. *Remote Sensing*, **15** (5), 1284.

3. Adhikari, A., Peduzzi, A., Montes, C.R., Osborne, N. and Mishra, D.R. 2023b Assessment of understory vegetation in a plantation forest of the southeastern United States using terrestrial laser scanning. *Ecological Informatics*, **77**, 102254.

4. Allen, H.L., Fox, T.R. and Campbell, R.G. 2005 What is ahead for intensive pine plantation silviculture in the South? *Southern Journal of Applied Forestry*, **29** (2), 62-69.

5. Archer, J.K., Miller, D.L. and Tanner, G.W. 2007 Changes in understory vegetation and soil characteristics following silvicultural activities in a southeastern mixed pine forest1. *The Journal of the Torrey Botanical Society*, **134** (4), 489-504.

6. Ashcroft, M.B., Gollan, J.R. and Ramp, D. 2014 Creating vegetation density profiles for a diverse range of ecological habitats using terrestrial laser scanning. *Methods in Ecology and Evolution*, **5** (3), 263-272.

7. Bonham, C.D. 2013 *Measurements for terrestrial vegetation*. John Wiley & Sons.

8. Brede, B., Terryn, L., Barbier, N., Bartholomeus, H.M., Bartolo, R., Calders, K. *et al.* 2022 Non-destructive estimation of individual tree biomass: Allometric models, terrestrial and UAV laser scanning. *Remote Sensing of Environment*, **280**, 113180.

9. Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M. *et al.* 2015 Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution*, **6** (2), 198-208.

10. Campbell, M.J., Dennison, P.E., Hudak, A.T., Parham, L.M. and Butler, B.W. 2018 Quantifying understory vegetation density using small-footprint airborne lidar. *Remote sensing of environment*, **215**, 330-342.

11. Campbell, T.N., Jones, P.D., Ezell, A.W. and Demarais, S. 2013 Growth and competition response in intensively established loblolly pine plantations at crown closure. *Journal of Forestry*, **111** (5), 313-318.

12. Cooper, S.D., Roy, D.P., Schaaf, C.B. and Paynter, I. 2017 Examination of the Potential of Terrestrial Laser Scanning and Structure-from-Motion Photogrammetry for Rapid Nondestructive Field Measurement of Grass Biomass. *Remote Sensing*, **9** (6), 531.

13. Dumelle, M., Higham, M., Ver Hoef, J.M., Olsen, A.R. and Madsen, L. 2022 A comparison of design-based and model-based approaches for finite population spatial sampling and inference. *Methods in Ecology and Evolution*, **13** (9), 2018-2029.

14. Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K. *et al.* 2019 The importance of consistent global forest aboveground biomass product validation. *Surveys in geophysics*, **40**, 979-999.

15. Eskelson, B.N., Madsen, L., Hagar, J.C. and Temesgen, H. 2011 Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, **57** (3), 212-221.

16. Fortson, J.C., Shiver, B.D. and Shackelford, L. 1996 Removal of competing vegetation from established loblolly pine plantations increases growth on Piedmont and Upper Coastal Plain sites. *Southern Journal of Applied Forestry*, **20** (4), 188-193.

17. Gonzalez, M., Augusto, L., Gallet-Budynek, A., Xue, J., Yauschew-Raguenes, N., Guyon, D. *et al.* 2013 Contribution of understory species to total ecosystem aboveground and belowground biomass in temperate Pinus pinaster Ait. forests. *Forest Ecology and Management*, **289**, 38-47.

18. Hancock, S., Anderson, K., Disney, M. and Gaston, K.J. 2017 Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. *Remote Sensing of Environment*, **188**, 37-50.

19. Hoerl, A.E. and Kennard, R.W. 1970 Ridge regression: applications to nonorthogonal problems. *Technometrics*, **12** (1), 69-82.

20. Holcomb, J.P. 1999 Applied Regression Analysis: A Research Tool. *The American Statistician*, **53** (2), 170.

21. Hopkinson, C., Chasmer, L., Young-Pow, C. and Treitz, P. 2004 Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research*, **34** (3), 573-583.

22. Jupp, D.L., Culvenor, D., Lovell, J., Newnham, G., Strahler, A. and Woodcock, C. 2009 Estimating forest LAI profiles and structural parameters using a ground-based laser called 'Echidna®. *Tree Physiology*, **29** (2), 171-181.

23. Kuhn, M. 2015 A Short Introduction to the caret Package. *R Found Stat Comput*, **1**, 1-10.

24. Lafarge, T., Pateiro-Lopez, B. and Pateiro-Lopez, M.B. 2016 Package 'alphashape3d'. *Recuperado el*, **26**.

25. Lague, D., Brodu, N. and Leroux, J. 2013 Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). *ISPRS Journal of Photogrammetry and Remote Sensing*, **82**, 10-26.

26. Li, L., Chen, J., Mu, X., Li, W., Yan, G., Xie, D. *et al.* 2020 Quantifying Understory and Overstory Vegetation Cover Using UAV-Based RGB Imagery in Forest Plantation. *Remote Sensing*, **12** (2), 298.

27. Li, S., Wang, T., Hou, Z., Gong, Y., Feng, L. and Ge, J. 2021 Harnessing terrestrial laser scanning to predict understory biomass in temperate mixed forests. *Ecological Indicators*, **121**, 107011.

28. Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C. and Culvenor, D.S. 2005 Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, **214** (1), 398-412.

29. Michael, J. 1980 Long-term impact of aerial application of 2, 4, 5-T to longleaf pine (Pinus palustris). *Weed Science*, **28** (3), 255-257.

30. Miller, J.H., Zutter, B.R., Zedaker, S.M., Edwards, M.B. and Newbold, R.A. 2003 Growth and yield relative to competition for loblolly pine plantations to midrotation—a southeastern United States regional study. *Southern Journal of Applied Forestry*, **27** (4), 237-252.

31. Molto, Q., Rossi, V. and Blanc, L. 2013 Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution*, **4** (2), 175-183.

32. Momo Takoudjou, S., Ploton, P., Sonké, B., Hackenberg, J., Griffon, S., De Coligny, F. *et al.* 2018 Using terrestrial laser scanning data to estimate large tropical trees biomass and calibrate allometric models: A comparison with traditional destructive approach. *Methods in Ecology and Evolution*, **9** (4), 905-916.

33. Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B. *et al.* 2015 Terrestrial laser scanning for plot-scale forest measurement. *Current Forestry Reports*, **1** (4), 239-251.

34. NOAA. 2023 2018 - 2020 USGS Lidar: Florida Peninsular FDEM from 2010-06-15 to 2010-08-15. NOAA National Centers for Environmental Information. https://www.fisheries.noaa.gov/inport/item/69496 (June 2023, 2023).

35. Peduzzi, A., Allen, H.L. and Wynne, R.H. 2010 Leaf area of overstory and understory in pine plantations in the flatwoods. *Southern Journal of Applied Forestry*, **34** (4), 154-160.

36. Popescu, S.C. and Wynne, R.H. 2004 Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, **70** (5), 589-604.

37. Reutebuch, S.E., Andersen, H.-E. and McGaughey, R.J. 2005 Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. *Journal of Forestry*, **103** (6), 286-292.

38. Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R., Meador, A.S. *et al.* 2020 lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, **251**, 112061.

39. Singh, K.K., Chen, Y.-H., Smart, L., Gray, J. and Meentemeyer, R.K. 2018 Intra-annual phenology for detecting understory plant invasion in urban forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, **142**, 151-161.

40. Strahler, A.H., Jupp, D.L., Woodcock, C.E., Schaaf, C.B., Yao, T., Zhao, F. *et al.* 2008 Retrieval of forest structural parameters using a ground-based lidar instrument (Echidna®). *Canadian Journal of Remote Sensing*, **34**, S426-S440.

41. Team, R.C. 2013 R: A language and environment for statistical computing.

42. Venier, L.A., Swystun, T., Mazerolle, M.J., Kreutzweiser, D.P., Wainio-Keizer, K.L., McIlwrick, K.A. *et al.* 2019 Modelling vegetation understory cover using LiDAR metrics. *Plos One*, **14** (11), e0220096.

43. Wagner, R.G., Little, K.M., Richardson, B. and Mcnabb, K. 2006 The role of vegetation management for enhancing productivity of the world's forests. *Forestry*, **79** (1), 57-79.

44. Watts, F.C. 1991 *Soil Survey of Nassau County, Florida*. Soil Conservation Service: Washington, D.C.

45. Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., Gitelman, A. and Olsen, M.J. 2012 Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. *Remote Sensing of Environment*, **124**, 730-741.

46. Woods, M., Lim, K. and Treitz, P. 2008 Predicting forest stand variables from LIDAR data in the Great Lakes St. Lawrence Forest of Ontario. *The Forestry Chronicle*, **84** (6), 827-839.

47. Zhang, J., Powers, R.F., Oliver, W.W. and Young, D.H. 2013 Response of ponderosa pine plantations to competing vegetation control in Northern California, USA: a meta-analysis. *Forestry*, **86** (1), 3-11.

48. Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X. *et al.* 2016 An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sensing*, **8** (6), 501.

49. Zhen, Z., Yang, L., Ma, Y., Wei, Q., Jin, H.I. and Zhao, Y. 2022 Upscaling aboveground biomass of larch (Larix olgensis Henry) plantations from field to satellite measurements: a comparison of individual tree-based and area-based approaches. *GIScience & Remote Sensing*, **59** (1), 722-743.

50. Zou, H. 2006 The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** (476), 1418-1429.

CHAPTER 5

CONCLUSION

Efficient methods for estimating forest attributes are essential to keep pace with the expanding plantation forests driven by increasing demand due to rapid urbanization. Additionally, addressing monitoring methods for forest growth, such as forest inventory attributes, and major growth limitations such as nutrient competition from understory vegetation, requires spatial distribution information for improved forest management decisions.

Chapter 2 assessed four modeling methods for forest attribute estimation using high-dimensional LiDAR metrics. The results revealed that while no method displayed a significant advantage over the others, the ALASSO achieved slightly higher Adjusted R^2 values of 0.88, 0.83, and 0.87 for estimating volume, basal area, and dominant height, respectively. GAMSEL and LSR also demonstrated strong performance and limited bias. In light of these findings and the ease of implementation of the method, we recommend the ALASSO approach for forest attribute estimation. Nevertheless, further research is needed to explore the generalizability of these results to other forest types and conditions.

Chapter 3 introduced an efficient model based on TLS-derived metrics capable of predicting understory biomass at one-meter resolution, addressing cost and vegetation complexity challenges in assessing understory vegetation in coastal forests in the southeastern United States. Volume-based metrics, particularly mean height-based volume combined with the 20th percentile of echo height, showed robust and superior correlations with understory biomass, achieving an R^2 of 0.80 and RMSE of 234.8 grams.

However, a limitation of this method is its suitability for small areas or stand-level assessments, requiring ground-level LiDAR data collection.

Chapter 4 focused on upscaling the TLS-derived metrics-based model to UAV and ALS-derived metrics. This enables efficient assessment of understory vegetation over larger areas with high accuracy, achieving an R^2 of 0.60 with UAV and 0.67 with ALS.

In conclusion, this research provides valuable insights for researchers, plantation forest managers, investors, and stakeholders. It underscores the potential of high-resolution remote sensing methods to enhance data-driven decision-making processes. By utilizing LiDAR from any accessible platforms, this research equips forest managers with a tool to better understand plantation forest growth, the abundance of competing vegetation, and their spatial distribution. Such information can support the development of site-specific management plans and improve overall yield.

Appendix 1.A

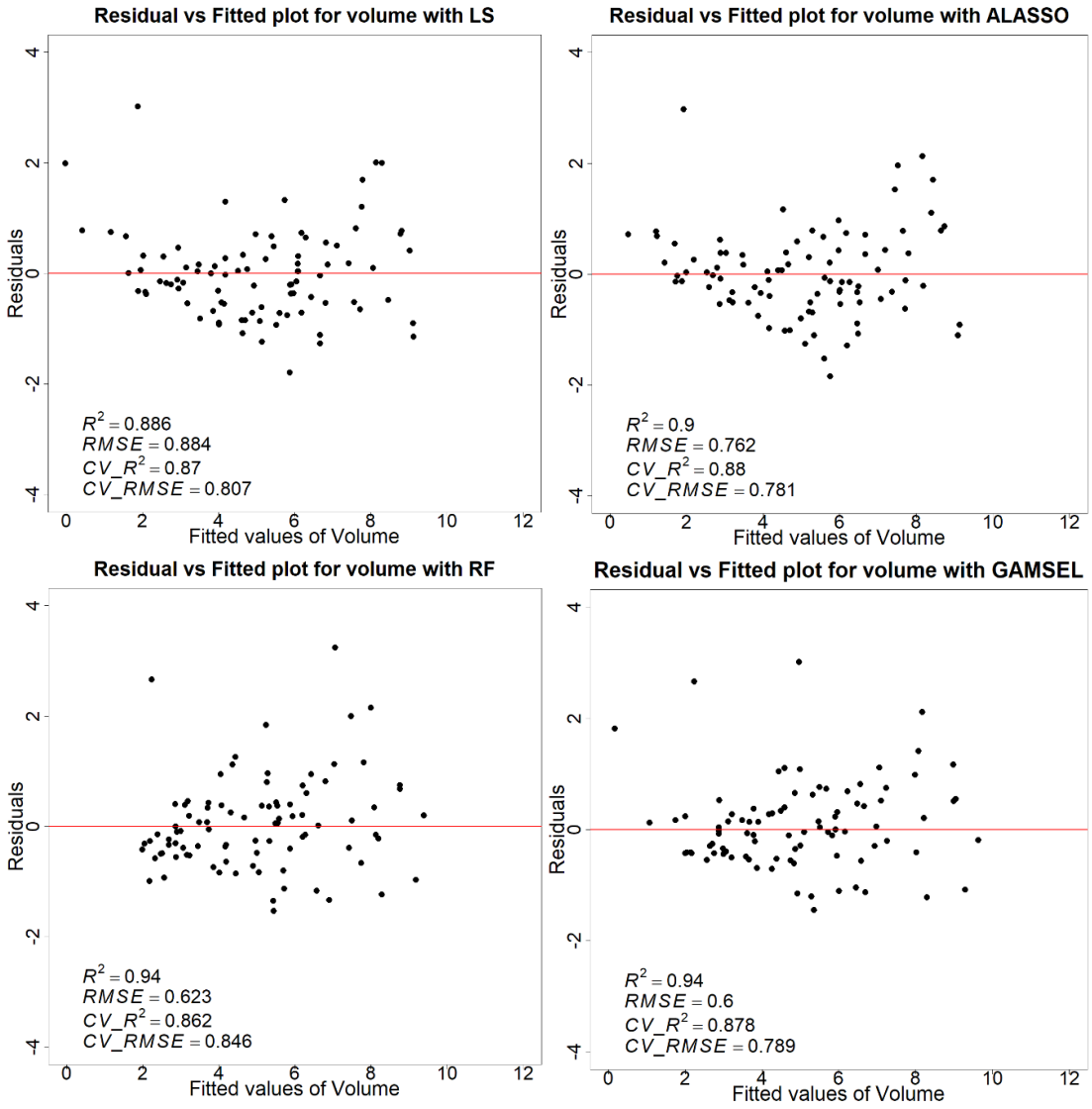

Figure 1.A1: Each panel shows the residual vs. fitted plot for volume with LS, ALASSO, RF, and GAMSEL models. R^2 and RMSE are the model accuracy before cross-validation, and CV_R^2 and CV RMSE show the model accuracy result of 9-fold cross-validation.

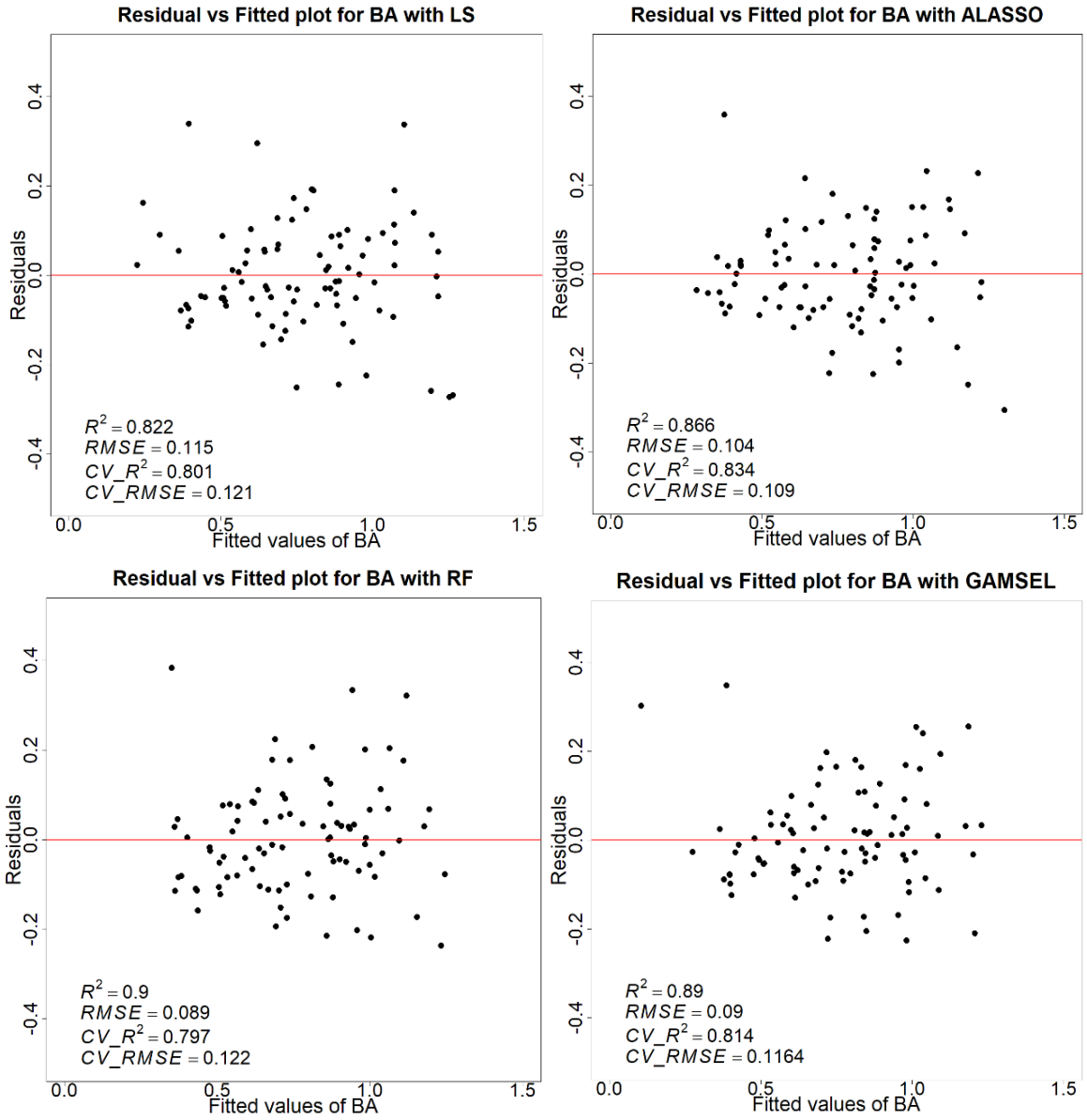

Figure 1.A2. Each panel shows the residual vs. fitted plot for the basal area with LS, ALASSO, RF, and GAMSEL models. R^2 and RMSE are the model accuracy before cross-validation, and CV_R^2 and CV RMSE show the model accuracy result of 9-fold cross-validation.

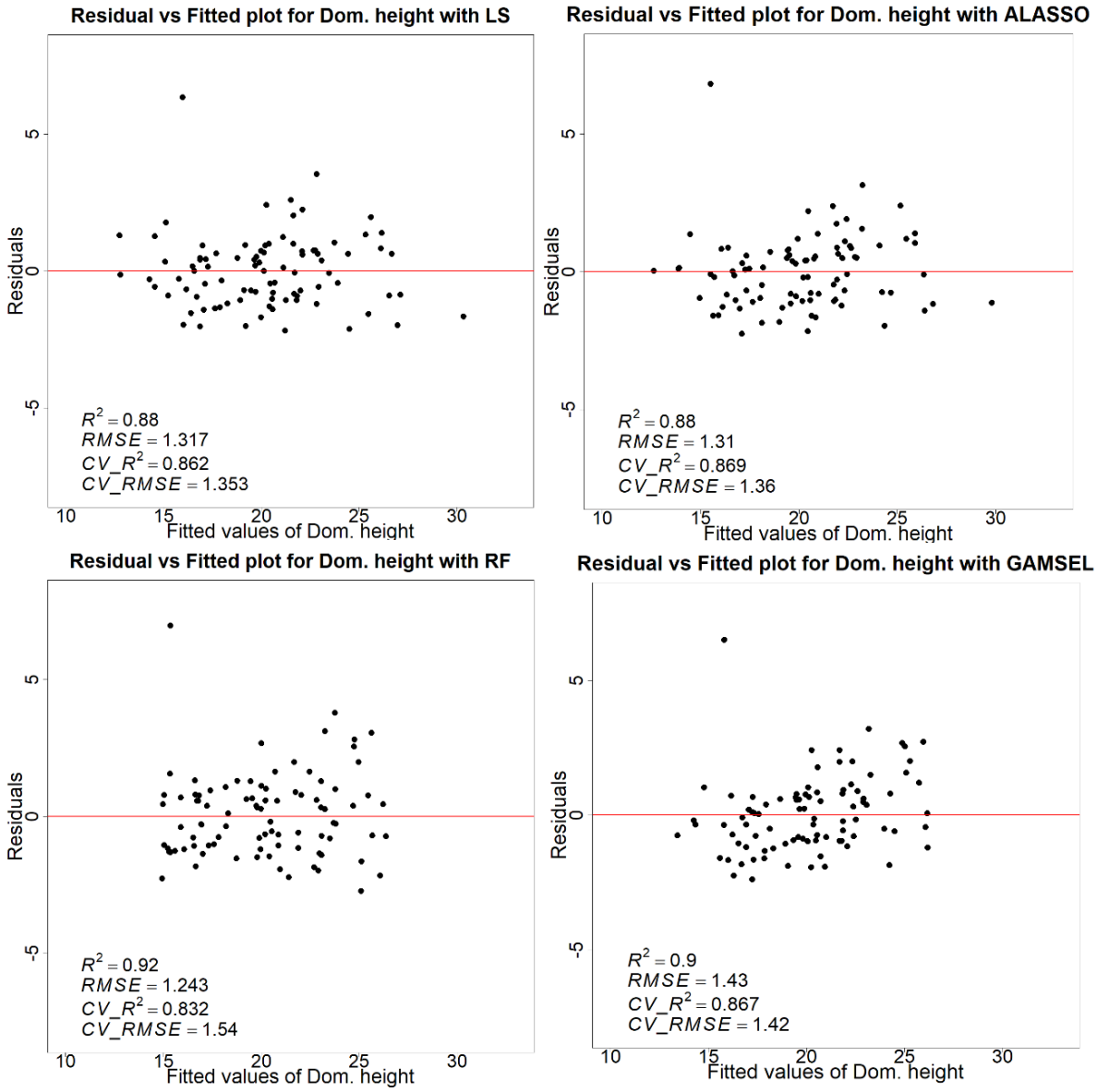

Figure 1.A3. Each panel shows the residual vs. fitted plot for dominant height with LS, ALASSO, RF, and GAMSEL models. R^2 and RMSE are the model accuracy before cross-validation, and CV_R^2 and CV_RMSE show the model accuracy result of 9-fold cross-validation.