Assessing the utility of NAIP digital aerial photogrammetric point clouds for estimating canopy height of managed loblolly pine plantations in the southeastern United States

Alison Leigh Ritz

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Forestry

Valeria A. Thomas (Co-advisor)

Randolph H. Wynne (Co-advisor)

Todd A. Schroeder

May 10, 2021

Blacksburg, Virginia

Keywords: remote sensing, NAIP, height estimation, point clouds, loblolly pine

Assessing the utility of NAIP digital aerial photogrammetric point clouds for estimating canopy height of managed loblolly pine plantations in the southeastern United States

Alison Leigh Ritz

ABSTRACT

Remote sensing offers many advantages to previous forest measurements, such as limiting costs and time in the field. Light detection and ranging (lidar) has been shown to enable accurate estimates of forest height. Lidar does produce precise measurements for ground elevation and forest height, where and when it is available. However, it is expensive to collect and does not have wall-to-wall coverage in the United States. In this study, we estimated height using the National Agricultural Imagery Program (NAIP) photogrammetric point clouds to create a predicted height map for managed loblolly pine stands in the southeastern United States. Recent studies have investigated the ability of digital aerial photogrammetry (DAP), and more specifically NAIP, as an alternative to lidar as a means of estimating forest height due to its lower costs, frequency of acquisition, and wall-to-wall coverage across the United States. Field-collected canopy height for 534 plots in Virginia and North Carolina were regressed against the 90^{th} percentile derived from NAIP point clouds. The model for predicted pine height using the 90^{th} percentile of height (P90) is predicted pine height = $1.09(\text{P90}) - 0.43$. The adjusted R^2 is 0.93, and the RMSE is 1.44 m. This model is being used to produce a 5 m x 5 m canopy height model for all pine stands across Virginia, North Carolina, and Tennessee. NAIP-derived point clouds are thus a viable means of predicting canopy height in southern pines.

Assessing the utility of NAIP digital aerial photogrammetric point clouds for estimating canopy height of managed loblolly pine plantations in the southeastern United States

Alison Leigh Ritz

GENERAL ABSTRACT

Collecting accurate measurements of pine plantations is essential to managing them to maximize various ecosystem goods and services. However, it can be difficult and time-consuming to collect these measurements by hand. This study demonstrates that point clouds derived from digital stereo aerial photograms enable calculating forest height to an accuracy sufficient for pine plantation management. We developed a simple linear regression model to predict forest canopy height using the 90th percentile of the photo-derived heights above the ground in a given area. With this model, we created a map of pine plantation canopy heights (consisting of 5 m x 5 m grid cells, each containing a canopy height estimate) for pine forests in Virginia, North Carolina, and Tennessee. Digital aerial photography from the National Agricultural Imagery Program (NAIP) is repeated every three years for a given state, allowing growth to be mapped over time. Photography collected by NAIP and similar programs also has uniform acquisition parameters in a given year applicable over large regions. State- and national photography programs like NAIP are also less expensive than other data sets, like airborne laser scanning data, that enable estimation of tree height.

Table of contents

List of figures

List of tables

ATTRIBUTION

Valerie A. Thomas and Randolph H. Wynne are the advisors to this study. They oversaw the process and were editors on the thesis.

Todd A. Schroeder is a third committee member who provided some advice on the study. He also provided feedback on the thesis near the end of the editing process.

Patrick C. Green and Harold E. Burkhart provided the field data for the three Virginia state forests.

Timothy J. Albaugh, David Carter, and Rachel Cook provided the field data for Patrick County in Virginia and Bladen, Jones, and Brunswick counties in North Carolina.

1. Introduction

The advancements in remote sensing have grown significantly over the past few decades, allowing for new and revolutionary ways to measure forest height. Maintaining an accurate forest height model is essential for forest management, especially in productive timber stands (Schultz 1999; Green et al., 2020; Navarro et al., 2020; Næsset 2002). Airborne laser scanning (ALS) has led this revolution for some time. However, several impediments to the use of ALS data exist, including high costs and limited availability across the United States (and elsewhere). Digital aerial photogrammetry has been studied as an alternative and found to be comparable to ALS in forest height estimation when utilizing a precise digital elevation model (DEM) obtained from ALS point clouds (Goodbody et al., 2019; Strunk et al., 2020; Lebarl et al., 2010; White et al., 2015; Zimmerman et al., 2017; Bohlin et al., 2012).

Measuring and predicting forest height is an easier task in managed stands due to the species' homogeneity and age class. Forest height is better predicted in a homogenous stand of trees than in natural stands with varying species and age classes when using remote sensing data (Gopalakrishnan et al., 2019; Rahlf et al., 2017). Managed stands also provide an excellent foundation for studies utilizing an area-based approach (ABA) to process the remote sensing data, which relies on the relationship between the field measurements and the remote sensing predictor. This relationship is much harder to establish in natural stands (Kwong & Fung 2020; Gobakken et al., 2014; Rahlf et al., 2017; Green et al., 2020). Previous studies have shown that the correlation between field measurements and DAP is stronger in an ABA study (Moe et al., 2020; Goodbody et al., 2019; White et al., 2015; Gobakken et al., 2014). Other studies have investigated and demonstrated the utility of DAP to collect forest attributes at a reasonable accuracy, such as height (Næsset et al., 2002; White et al., 2015; Bohlin et al., 2012; Zimmerman & Hoffmann 2017; Mielcarek et al., 2020; Strunk et al., 2020; Michez et al., 2020; Maimaitijiang et al., 2020; Zagalikis et al., 2005; Gobakken et al., 2014; Kwong & Fung 2020; Noordermeer et al., 2019; Shen et al., 2019; Moe et al., 2020; Nurminen et al., 2013), volume (Green at al., 2020; Navarro et al., 2020; White et al., 2015; Bohlin et al., 2012; Strunk et al., 2020; Rahlf et al., 2017; Noordermeer et al., 2019; Shen et al., 2019; Strunk et al., 2019; Nurminen et al., 2013), density (Gobakken, et al., 2014), basal area (White et al., 2015; Bohlin et al., 2012; Strunk et al., 2020; Rahlf et al., 2017; Noordermeer et al., 2019), leaf area index (LAI) (Maimaitijiang et al., 2020), site index (Noordermeer et al., 2021; Gopalakrishnan et al., 2019), ecological attributes (Masek et al., 2015), and tree detection (Strunk et al., 2020; Thiel & Schmullius 2016; Noordermeer et al., 2019).

The National Agriculture Imagery Program (NAIP) is aerial imagery collected by the USDA's Farm Service Agency (FSA). For a given state, NAIP was acquired every five years from 2003 to 2009 and is now acquired every three years during the agricultural growing season. NAIP is collected using

overlapping photographs, creating wall-to-wall imaging of the country (Strunk et al., 2019; Strunk et al., 2020). The routine collection of the NAIP program allows for easy modeling of forest growth and change across the country from homogenous acquisition parameters. The homogeneity of the NAIP acquisition is also reflected in its ability to capture more of the surface in a smaller amount of time, leaving less room for changes to occur between flight lines (Strunk et al., 2020 & 2019). In addition, NAIP photos can generate 3D point clouds of the surface comparable to lidar returns of the upper canopy.

Not only is NAIP wall-to-wall, but it can be acquired for 1/3 of the costs of lidar with routine data acquisitions (Goodbody et al., 2019; Michez et al., 2020; White et al., 2015; Navarro et al., 2020). In addition, some studies have shown that NAIP imagery and point clouds are not significantly impacted by acquisition parameters within the typical constraints, including as sun position, terrain, and shadow (Rahlf et al., 2017; White et al., 2015; Kwong & Fung 2020). However, NAIP cannot penetrate the canopy of a forest structure and only produces a digital surface model (DSM) of the stand. A solution to this is the use of lidar created DEMs, which is considered the "best available data products" by Goodbody et al. (2019), with the NAIP point clouds to calculate a normalized canopy height model (Michez et al., 2020; Mielcarek et al., 2020; White et al., 2015; Navarro et al., 2020; Bohlin et al., 2012). After using a high-quality DEM, studies have found that the strongest correlation in modeling forest height between lidar and DAP methods was for pine height metrics in western Europe, with an R^2 greater than 0.9 (Mielcarek et al., 2020; Michez et al., 2020).

These early studies highlight the potential for DAP to predict pine or other forest heights in Europe and western North America. However, they have not been widely explored in the heavily managed pine forests of the southeastern United States. Thus, there is a need to explore this potential for managed southern pines, which span a large geographic region and are a vital economic resource for the United States.

2. Objective

The objective of this study was to develop a model that accurately predicts the canopy height of actively managed loblolly pine plantations from NAIP photogrammetric point clouds. This model was then utilized to create a pine canopy height map for Virginia, North Carolina, and Tennessee.

3. Materials and methods

3.1 Study area

The overall study area is the evergreen, mixed and woody wetland forests of Virginia, North Carolina, and Tennessee, based on the National Land Cover Dataset. Field data containing measured heights were used for 534 plots across seven loblolly forests (the training regions) in North Carolina and

Virginia. All these training regions are in humid, temperate climates with hot summers and mild winters. The study area ranges from coastal to mountainous, with soils varying from sandy to clay. More information on the loblolly pine forests comprising the training regions can be found in Table 1.

3.1.1 Training Regions

The seven training regions were spread out over Virginia and North Carolina with three in Virginia state forests, one in a Virginia region-wide 13 study site (Forest Productivity Cooperative; Scolforo et al., 2020), and three in North Carolina that are also part of the region-wide 13 study (Figure 1a). The three Virginia state forests are Appomattox-Buckingham (ABSF), Cumberland (CUSF), and Prince Edward (PESF) state forests. These state forests reside in bordering counties in the state's center with a combined area of 42,537 acres. These three state forests maintain numerous actively managed loblolly pine stands, in which 142 plots, 84 plots, and 34 plots were measured in the ABSF, CUSF, and PESF, respectively. Plots were established with a Trimble Geo 7x GPS, are managed by the Virginia Department of Forestry (VDoF), and were measured in 2018-2019 as part of a study conducted by the Forest Modeling Research Cooperative (FMRC; Green et al., 2020). The other Virginia training region was the largest, comprised of 144 plots. The next largest training region was in Bladen County, North Carolina, containing 108 plots. Both stands were planted between 1984 and 1987. The remaining two training regions were in Jones and Brunswick counties in North Carolina and were planted in 2017. These two regions were each comprised of two plots that were broken into six subplots with 12 plot measurements each. All the plots in Jones and Brunswick counties contained various fertilization prescriptions for the intent of the original study, California Fertilizer (CaFe). Since the training regions in Jones and Brunswick counties were planted in 2017 and measured in 2018, the trees were short and relatively homogeneous in height.

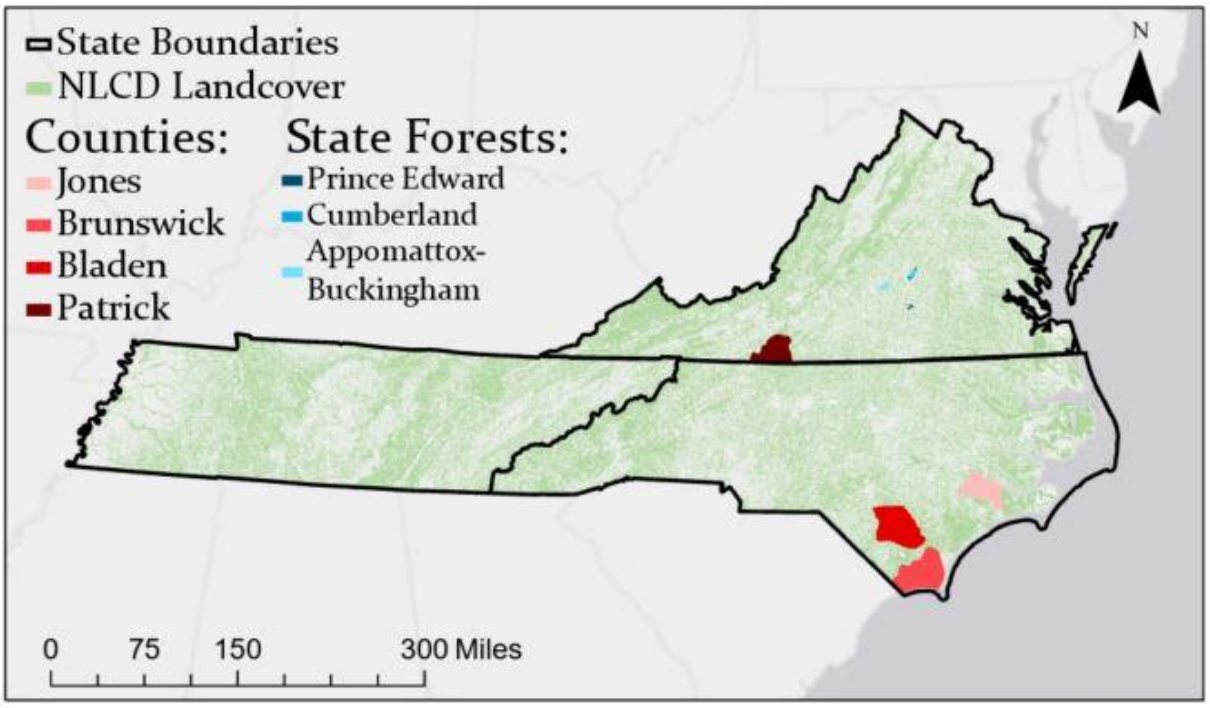
3.1.2 Mapping Regions

The model was initially applied to the ABSF to test the mapping approach and to ensure the spatial coherence of the model across the landscape (Figure 1b). Since these plots did not have field measurements and were a non-mixed pine forest, a 5 m x 5 m grid was used to demonstrate the application of the model.

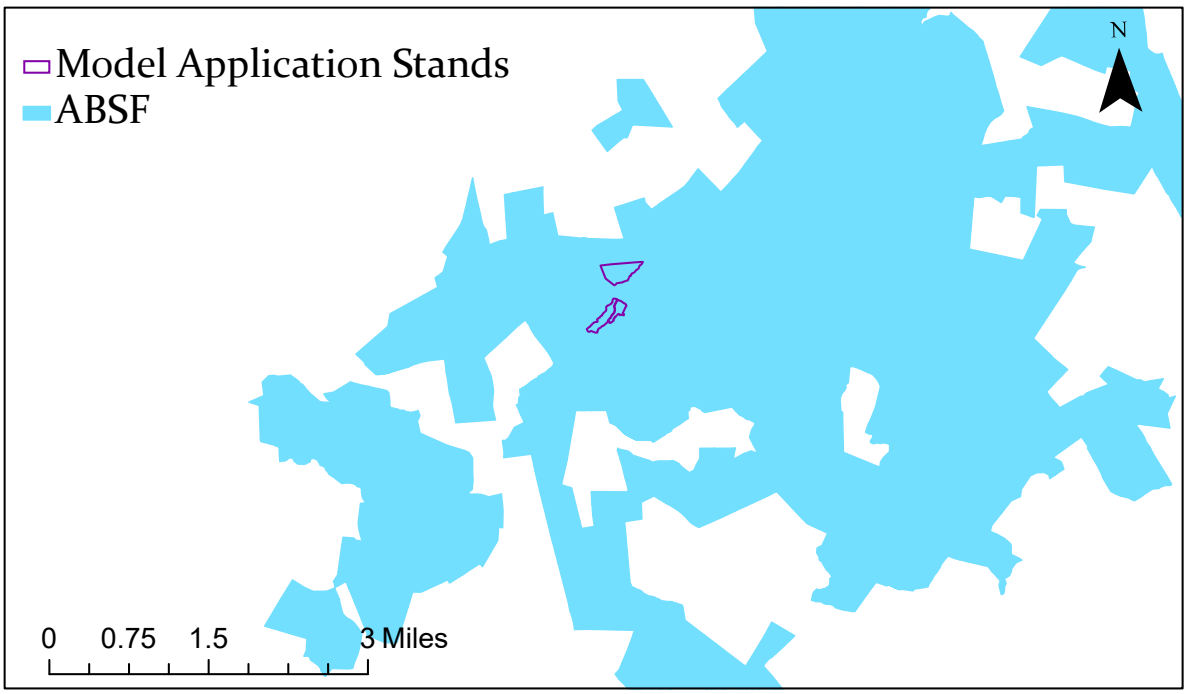
The resulting model is applied to all areas likely to contain pines in Virginia, North Carolina, and Tennessee (Figure 1a). Pine pixels are defined as those classified as evergreen forests (42), mixed forests (43), and woody wetlands (90) in the 2018 national landcover dataset (NLCD).

Table 1. Geographic information for the field plots

Stand	Coordinates	Geographic Region	Plot size	#Plots #Thinned	Notes
ABSF	37˚25'21"N 78˚40'46"W	Virginia Piedmont	0.013 – 0.02 ha fixed radius	140 27	Region-wide19 study. More information can be found in Green et al. 2020.
CUSF	37˚32'53"N 78˚15'12"W	Virginia Piedmont	0.013 – 0.02 ha fixed radius	84 56	Region-wide19 study. More information can be found in Green et al. 2020.
PESF	37˚10'12"N 78˚17'34"W	Virginia Piedmont	0.013 – 0.02 ha fixed radius	34 17	Region-wide19 study. More information can be found in Green et al. 2020.
Patrick	36˚38'35"N 80˚9'17"W	Virginia Piedmont/ Blue Ridge	0.05 – 0.15 ha	144 0	Region-wide 20 study. Trees were nine years old at the time of measurement.
Bladen	34˚49'50"N 78˚35'19"W	North Carolina Coastal Plain	0.05 – 0.15 ha	108 0	Region-wide 20 study. Trees were nine years old at the time of measurement.
Jones	35˚0'29"N 77˚22'52"W	North Carolina Tidewater/ Coastal Plain	0.1 ha plots	12 0	Region-wide 20 study. Trees were one year old at the time of measurement.
Brunswick	34˚9'45"N 78˚15'52"W	North Carolina Tidewater	0.1 ha plots	12 0	Region-wide 20 study. Trees were one year old at the time of measurement.



a)



b)

Figure 1. Virginia, North Carolina, and Tennessee: (a) training regions and the extent of statewide modeling with landcover classes from NLCD 2016, where class codes are 42 = evergreen forest, 43 = mixed forest, and 90 = woody wetlands, and (b) the stands in which the model was applied in Appomattox-Buckingham state forest (ABSF).

3.2 R

To facilitate a repeatable workflow, the entire analysis for this research was done using R. R is statistical modeling software that includes an expansive list of packages for various modeling techniques. For this study the following packages were used: `sf`, `rgdal`, `raster`, `elevatr`, `ggplot2`, `lidR`, `devtools`, `rlas`, `stats`, `metrics`, `base` (Table 2). The use of R in this research was integral to the ease of modeling and communicating the data for reproduction purposes.

Table 2. Definitions of R packages used in the analysis

Package	Description	Version	Citation
sf	Simple features for R	0.9-5	Pebesma 2018
rgdal	Bindings for the 'Geospatial' data abstraction library	1.5-10	Bivard 2020
raster	Geographic data analysis and modeling	3.1-5	Hijmans 2021
elevatr	Access elevations data from various APIs	0.2.0	Hollister 2020
ggplot2	Create elegant data visualizations using the grammar of graphics	3.3.1	Wickham 2016
lidR	Airborne lidar data manipulation and visualization for forestry applications	3.0.4	Roussel 2020
devtools	Tools to make developing R packages easier	2.3.1	Hester 2020

rlas	Read and write 'las' and 'laz' binary file formats used for remote sensing data	1.3.6	Roussel 2021
stats	R statistical functions	4.0.1	R Core Team, 2020
metrics	Implementation of evaluation metrics in R that are commonly used in supervised machine learning	0.1.4	Frasco, M., 2018
base	The R base package	4.0.1	R Core Team, 2020

3.3 Field measurements

The plots within the three Virginia state forests were established with a Trimble Geo 7x GPS and are managed by the Virginia Department of Forestry (VDOF). They were measured in 2018-2019 as part of a study conducted by the FMRC (Green et al., 2020). As noted in Section 3.1, the training region in Patrick County, Virginia and the training region in Bladen County, North Carolina were planted in a split-block design where some plots received treatment from a fertilizer study (Albaugh et al., 2017; Albaugh et al., 2018). The last two stands in Jones and Brunswick counties were planted in a randomized complete block design. A hypsometer was used to measure tree heights in all plots.

3.4 Remote sensing data

3.4.1 NAIP

The USDA FSA APFO NAIP program collected NAIP imagery in 2018-2019 for all three states (USDA-FPAC-BC-APFO Aerial Photography Field Office, 2019). The acquisition parameters are available online (e.g., OCM Partners, 2021) and were the same for all three states. There were multiple sensors used in this acquisition: serial numbers 10510, 10515, 10522, 10541, 10542, 10548, 10552 from Leica ADS-100. There was 30% side lap of the flight lines, allowing for 0.4 m sampling distance resolution, and image acquisition used +/- 6 m to ground specification. The Lexica XPro SGM Module was used to create point clouds (.laz format) generated from the ADS 100 imagery. Imagery for each flight line was 20,000-pixels in width. Autocorrelation algorithms were then used to identify image tie points and then mosaiced into one image. Four band images, with blue (435-495 nm), green (525-585 nm), red (619-651 nm), and near infrared (808-882 nm) spectral bands, were collected and compiled to create the point clouds. The data provided by NAIP also included thinned point clouds, where over 97% of the points were removed. Initial exploration showed that this was not sufficient for modeling tree heights, with some plots having no points in the upper canopy, or no points at all. Therefore, only the un-thinned point clouds are discussed further.

3.4.2 Lidar and DEM

ALS data were used in this study for their (1) accurate DEMs needed for normalizing the NAIP point clouds to height above ground, and (2) potential ability to enable the estimation of 2018 tree heights

(with de facto constraints associated with the inherent variability of acquisition parameters and timing). Lidar point clouds and associated 1-m digital terrain models were acquired from publicly available data from the United States Geological Survey (USGS) 3D Elevation Program (e.g., USGS, 2021). Since the terrain is unlikely to change much over time, older DEMs were used for the states that did not have recent lidar acquisitions. Patrick County was part of a 2018 lidar campaign, ABSF and CUSF were part of a 2015 lidar campaign, and PESF was from a 2014 campaign. Details of the lidar specifications can be found in Table 3 (USGA, Green et al., 2020). The USGS National Map (USGS 2017 a, b) holds the DEMs from the associated lidar point clouds.

Table 3. Lidar collection specifications for the four Virginia training regions. Note the varying acquisition parameters and collection dates.

	ABSF and CUSF	**PESF**	**Patrick Co.**
Campaign name	Chesapeake Bay	Sandy	South Central
Collection dates	November 15, 2015 – March 30, 2016	March 24, 2014 – April 21, 2014	April 14, 2017 – May 24, 2018
Sensor	Riegl 68oi	Leica ALS60 or Leica ALS 70	Riegl LMS-Q1560 or Riegl VQ-1560i or Riegl VQ-780i
Scan angle (degrees)	60	Unreported	58.52 (Serial numbers: Q1560 & 1560i) or 60 (780i)
Point density (pulse/m^2)	2.3	2.5	5.2
Nominal pulse spacing (m)	0.66	0.7	2.0 (Serial numbers: Q1560 & 1560i) or 1.0 (780i)
Flight line overlap	55%	30% (ALS60) or 20% (ALS70)	15% (Serial numbers: Q1560 & 1560i) or 55% (780i)
Pulse rate (kHz)	200	154.3 (ALS60) or 301.6 (ALS70)	687 (Serial numbers: Q1560 & 1560i) or 300 (780i)

3.5 Analyzing data

Data analysis was conducted in R and began by clipping the NAIP and lidar point clouds and the DEM to the extent of the plots in all three state forests (Table 4: `lasclip`). This clipping allows for measuring to occur only in the areas of study so that points outside the measurement area do not confound the statistical analysis. Then the DEM was subtracted from both the NAIP and lidar point clouds to create a point cloud of heights above the ground (Table 4: `normalize_height`) for each plot in the training area. Creating a canopy height model removes the impact of terrain on the height above the earth's ellipsoid that NAIP and lidar are measured in, converting the data to heights above the ground.

Table 4. R functions used in data analysis

Function	**Package and Description**	**Parameters**
`lasclip`	*lidR*: clip points within a given geometry from a point cloud ('LAS' object) or a catalog ('Lascatalog' object)	No extra parameters necessary.

`normalize_height`	*lidR*: subtract digital terrain model (DTM) from point cloud to create a dataset normalized with the ground at 0	No extra parameters necessary.
`function`	*sf:* creating user-defined functions in R for repeat use. Function created for 90th percentile.	`function`(Z) {p90 = quantile(Z, probs = c(0.9))}
`cloud_metrics`	*lidR*: computes a series of user-defined descriptive statistics for a dataset. The 90th percentile of height was used.	No extra parameters necessary.
`lm`	*stats*: fits linear models between variables	No extra parameters necessary.
`cor`	*stats*: if *x* and *y* are matrices then the covariances (or correlations) between the columns of **x** and the columns of **y** are computed	Default is Pearson correlation.
`rmse`	*metrics*: computes the root mean squared error between two numeric vectors	No extra parameters necessary.
`summary`	*base*: a generic function used to produce result summaries of the results of various model fitting functions	Default quantile type is all quantiles.
`grid_metrics`	*lidR*: computes a series of user-defined descriptive statistics for a dataset within each pixel of a raster (area-based approach)	No extra parameters necessary.

After the data were clipped and normalized, the 90th percentile of height for each plot was calculated to ensure that the top of the canopy was represented (Table 4: `cloud_metrics`) (Maimaitijiang et al. 2020). However, for the three stands in North Carolina, there were no lidar data available for the year of the field measurements so lidar analysis was only done for the Virginia stands and NAIP analysis was done for both North Carolina and Virginia.

Once the lidar data were plotted, it was clear that the data could not serve as a comparison in this study due to the time between the acquisition date and the field measurement date (noted in Table 3) and the height difference this creates between the field collection dates in 2018-2019. Lidar data for the four stands for which it was available were still analyzed, but purely on a general statistical summary basis.

The training region metrics were then combined so that all the NAIP and field metrics were in one large data frame. A model was then computed using linear regression techniques, such as correlation, RMSE, adjusted R^2, intercept, and the coefficient of *x* for the remote sensing measurement against the field measurements. The intercept and the coefficient of *x* were created by modeling the field measured height against the NAIP 90th percentile of height using the lm function (Table 4) in R, creating a simple linear regression relationship. Once this model was created the summary function (Table 4) in R provided the coefficient of *x* and the intercept associated with these two measurements. The coefficient of *x*, which

can be interpreted as the under or over estimation of height from the NAIP data, and the intercept, which can be interpreted as the adjustment required to predict the true height of the top of the canopy, together created the NAIP model. The NAIP model was then ready for visualization and analysis between the field canopy height and the NAIP collected heights for all seven stands.

Plotting of the NAIP data presented several points as outliers in the dataset. These outliers are not unexpected as NAIP has issues with open canopy plots with recent thinning (Thomas et al., 2019). Plots with potential recent or full harvests were determined to be the cause of these outliers, and they were flagged by having a NAIP height value greater than one standard deviation away from the field height value. These points were then visually assessed using historical aerial photography or very high spatial resolution satellite data available through Google Earth Pro (Figure 2). Six flagged plots that were visually verified as having recent full or partial harvests were removed.

Figure 2. A sample of flagged NAIP points that were visually assessed as being in recent full or partial harvests in Appomattox-Buckingham and Cumberland state forests.

To calculate the equation for predicting height, several statistical tests were run on the NAIP versus field model that produced correlations, summary statistics, and RMSE values (Table 4: `cor`, `summary`, `rmse`). These tests were run to produce general statistical information about the model and to interpret error associated with NAIP so that the predicted height equation can have a strong R^2.

3.6 Applying the model

A canopy height model for the ABSF application area (Figure 1b) was created using the model for expected pine height. The same process by which metrics were generated for the training plots was applied to 5 m x 5 m grid cells (Table 4: `grid_metrics`). A 5 m x 5 m grid was selected so that at least one tree crown (or portion thereof) would be in a typical grid cell. Once the NAIP point cloud was normalized (to heights above ground) using the lidar-derived DEM, the 90th percentile value for the NAIP points in each grid cell was calculated. The canopy height model was then created by applying the new model for predicted pine height. The number of points in each grid cell used to calculate the 90th percentile was also retained and mapped.

3.7 Statewide modeling

After a successful analysis of the application stand data, the new model was applied to three NLCD forest classes assumed to contain mostly pine in Virginia, North Carolina, and Tennessee in a high-performance computing environment (Virginia Tech Advanced Research Computing (ARC)). Once this process is completed a map will be produced with predicted height values for all pine stands across the three states.

4. Results

The lidar data in this study were inadequate for measuring forest height due to the drastic difference in height from the various years and parameters of acquisition. A visualization of this can be seen in Figure 3 where the lidar measured height falls much lower than the field height, again distinguishing the almost four-year-old height difference. Table 5 provides basic statistical summaries about the lidar and respective field data.

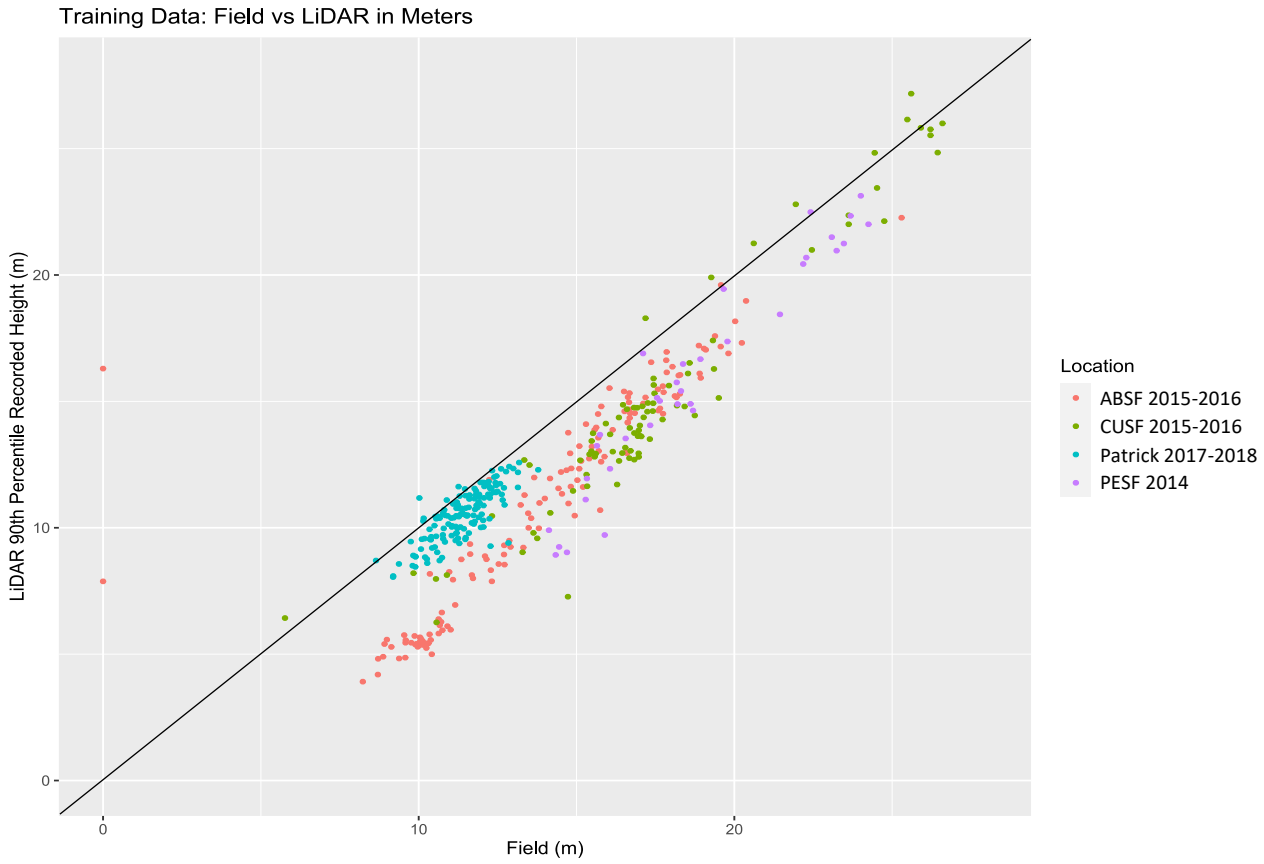


Figure 3. Model of the lidar 90th percentile of height versus the field measured height. The 1:1 line is represented in black. The years noted in the legend are the years the lidar data were acquired.

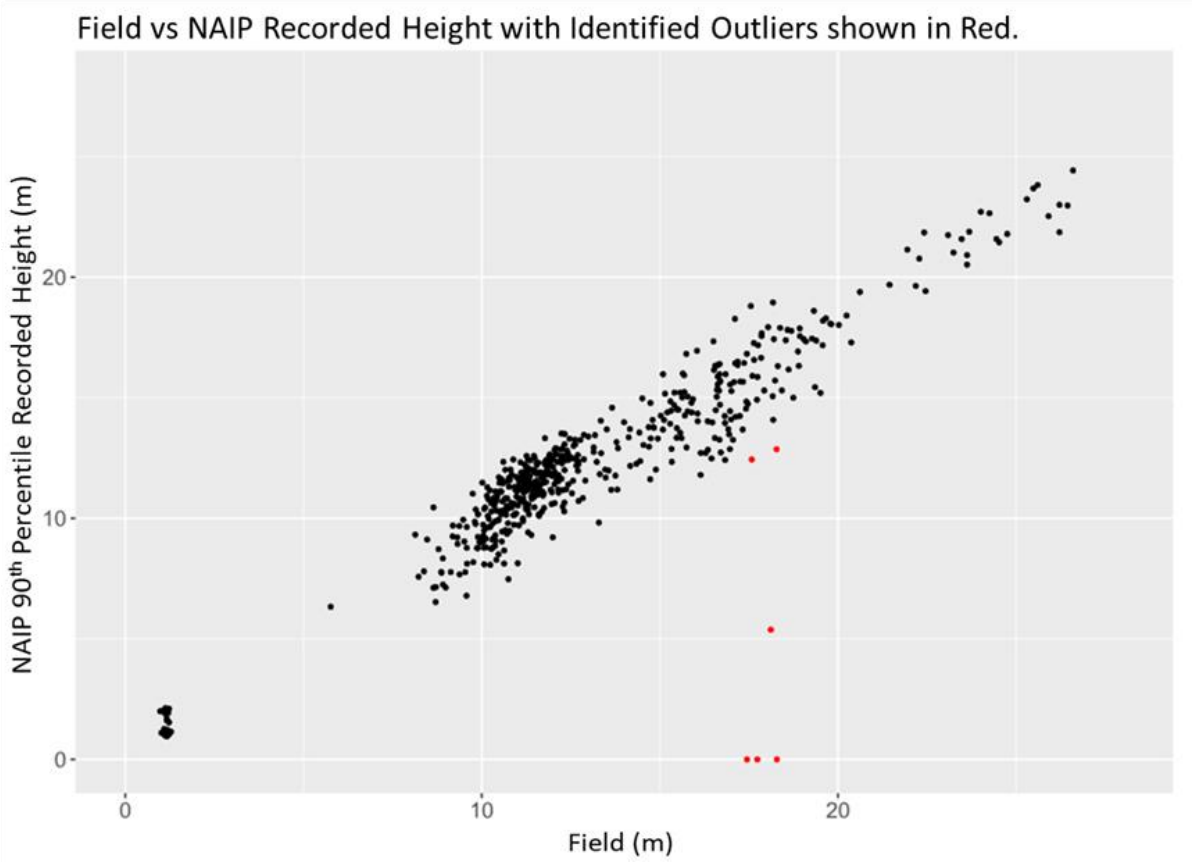
Table 5. Summary statistics of the data in the lidar plots

Data	Min (m)	Max (m)	Median (m)	Mean (m)	Count
Field	5.76	26.59	12.95	14.26	402
Lidar	3.91	27.17	11.42	12.16	402

The combined field data plotted against the NAIP 90th percentile of height provides a strongly related plot of height values, as expected (Figure 4a, b). A distribution of all the thinned plots versus the non-thinned are shown in Figure 4a. The six outliers, that were later removed, are highlighted in Figure 4b. From these plots, it is clear that the NAIP 90th percentile measures tree heights that closely correspond to field-measured tree heights.


Thins in the Training Data
NAIP 90th Percentile of Height (m)
0
10
20
Field Height (m)
0
10
20
Thinned
N
Y

222 a)


Field vs NAIP Recorded Height with Identified Outliers shown in Red.
NAIP 90th Percentile Recorded Height (m)
0
10
20
Field (m)
0
10
20

223 b)

Figure 4. Model of the NAIP 90th percentile of height versus field data height, where (a) shows the distribution of thinned versus the non-thinned plots across the training data and (b) shows the outliers that were removed in the study, represented in red.

The model created for the NAIP point cloud demonstrated strong correlations with the field heights. The correlation, adjusted R^2, Intercept, coefficient of *x* (where x is the height collected by NAIP), and press statistic (shown as the predicted R^2**)** values of these models for the original dataset and dataset after removing outliers can be found in Table 6. The standard deviation, which was calculated from the field data, was 4.55 m (Table 7) and it removed 6 outlier points based off the values in the NAIP dataset.

Table 6. Correlation, summary statistics and RMSE for the training stand's 90th percentile height metrics

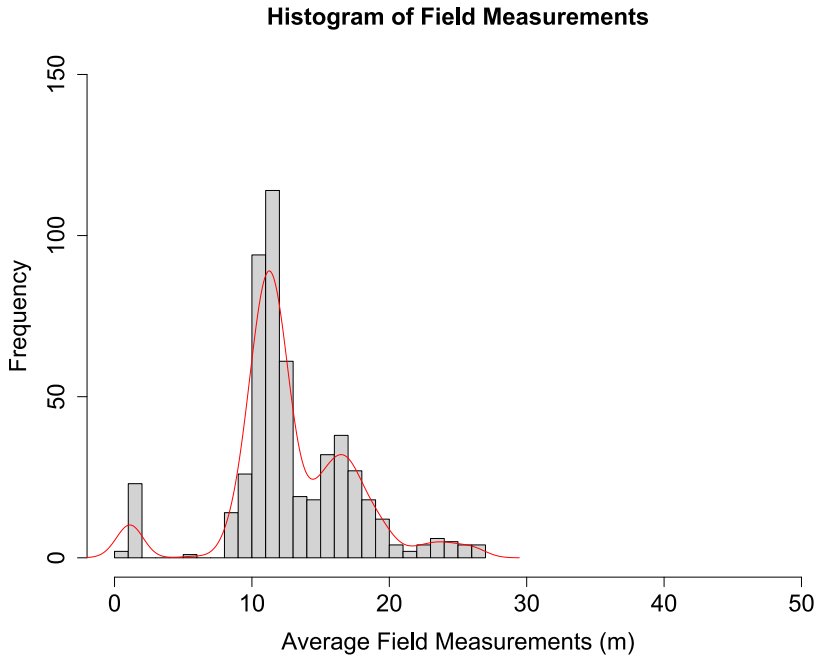
Metrics	Correlation	Adjusted R^2	Intercept	Coefficient of *x*	RMSE	n	Predicted R^2
NAIP vs field with outliers	0.91	0.83	0.76	1.01	2.06	534	0.82
NAIP vs field after outliers are removed	0.96	0.93	-0.43	1.09	1.44	528	0.93

Table 7. Summary statistics of the data in the training regions and the FIA data

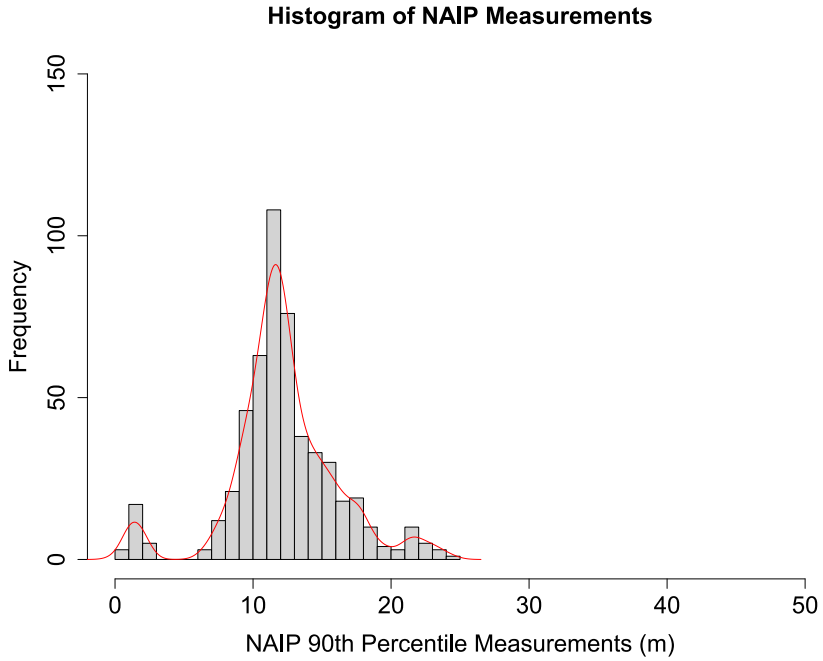
Data	Min (m)	Max (m)	Median (m)	Mean (m)	Standard deviation (m)	Variance (m^2)	Count
Field	0.98	26.59	11.92	12.96	4.55	20.10	528
NAIP with outliers	0.00	24.42	11.88	12.16	4.05	16.37	534
NAIP after outliers are removed	0.96	24.43	11.90	12.24	3.95	15.62	528
FIA data for VA, NC, and TN	1.52	53.95	15.24	15.622	6.21	38.54	124,857

The data distribution is similar with peaks at 10-15 m (Figure 5a, b, and c). This similarity is also reflected in Table 7 where the mean height values for the data are around 12 m, with FIA being slightly higher at 15 m. The NAIP distribution models very closely to the field measurements but is somewhat lacking in the field inventory analysis (FIA) data, collected by the USDA Forest Service. The FIA data was included in this distribution analysis to ensure that the training data used in this study captures the full range of loblolly pine heights in the three states. By comparing Figures 5a and b to Figure 5c, the training data is lacking in the heights above 30 m, however, there are not many FIA heights in this range so overall the training data is considered to capture an appropriate range of heights. Figures 5a and 5b are

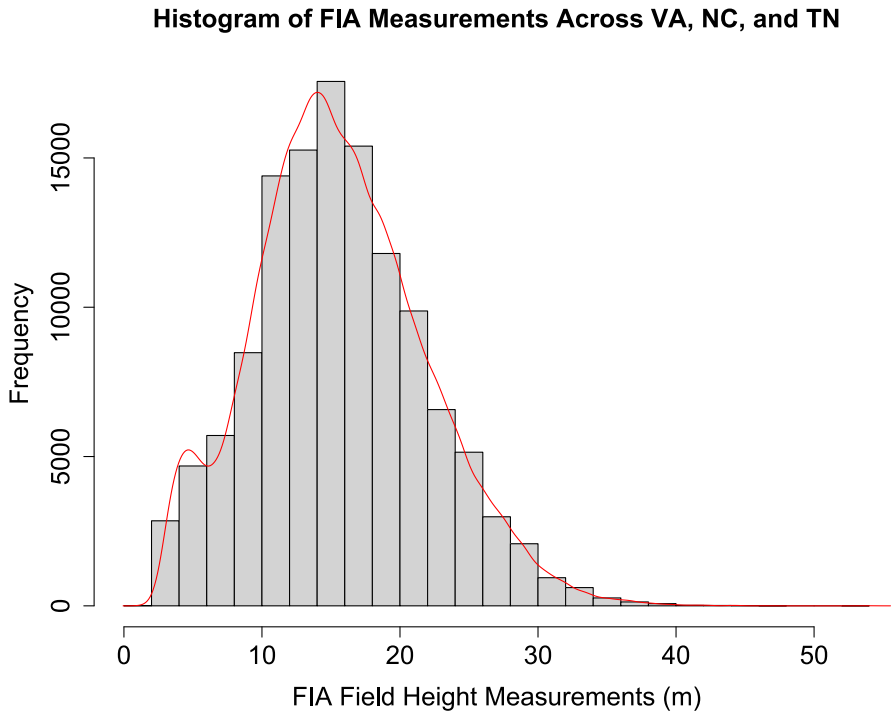
very similar in their distributions, demonstrating only a slightly lower height calculation by the NAIP data compared to the field data, which is expected.



a)



b)



c)

Figure 5. Histograms of the measured data where (a) is the distribution of the field data, (b) is the distribution of the NAIP 90th percentile measurements, and (c) is the distribution of heights from the Forest Inventory and Analysis data for loblolly pines in the states of VA, NC, and TN. Their respective density curves are shown in red.

The resulting simple linear regression model for the predicted pine height is as follows:

$$Predicted\ Pine\ Height\ (PPH) = 1.09X - 0.43, \quad (1)$$

where X is the 90th percentile of height from the normalized NAIP. The 95% confidence interval for the coefficient of x is 1.07 – 1.12 with a standard error of 0.01, and the 95% confidence interval for the intercept is (-0.76) – (-0.09) with a standard error of 0.17. The negative intercept was not expected since it has been demonstrated that DAP tends to underpredict tree height (Michez et al., 2020, Mielcarek et al., 2020, Næsset 2002). A plot of the predicted pine height from the training data versus the field height is displayed in Figure 6.

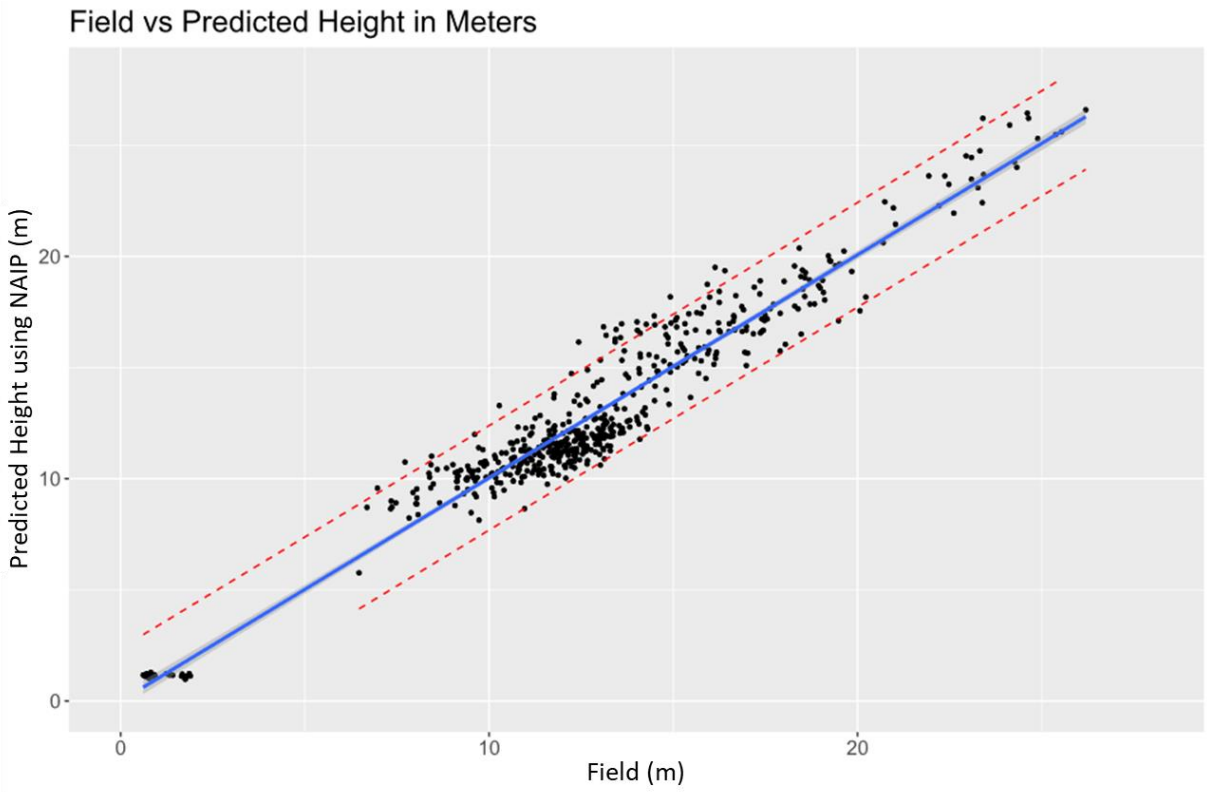
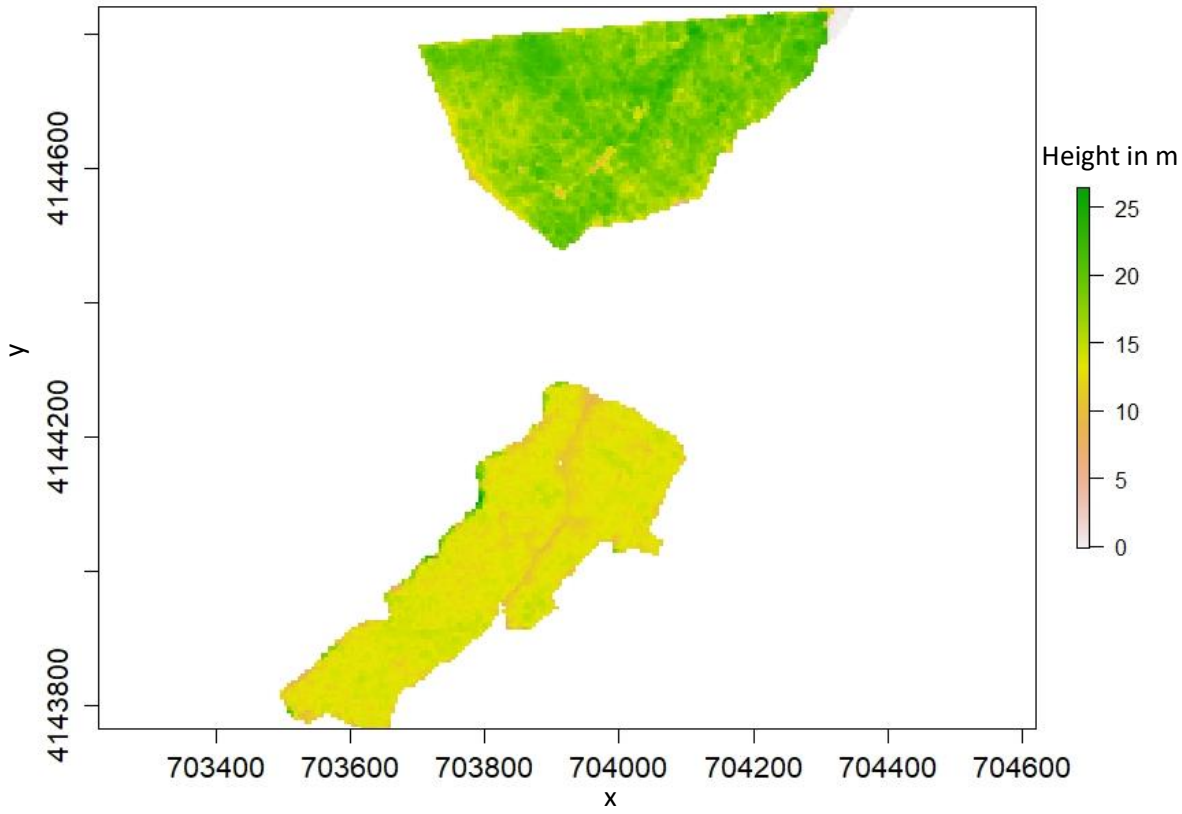
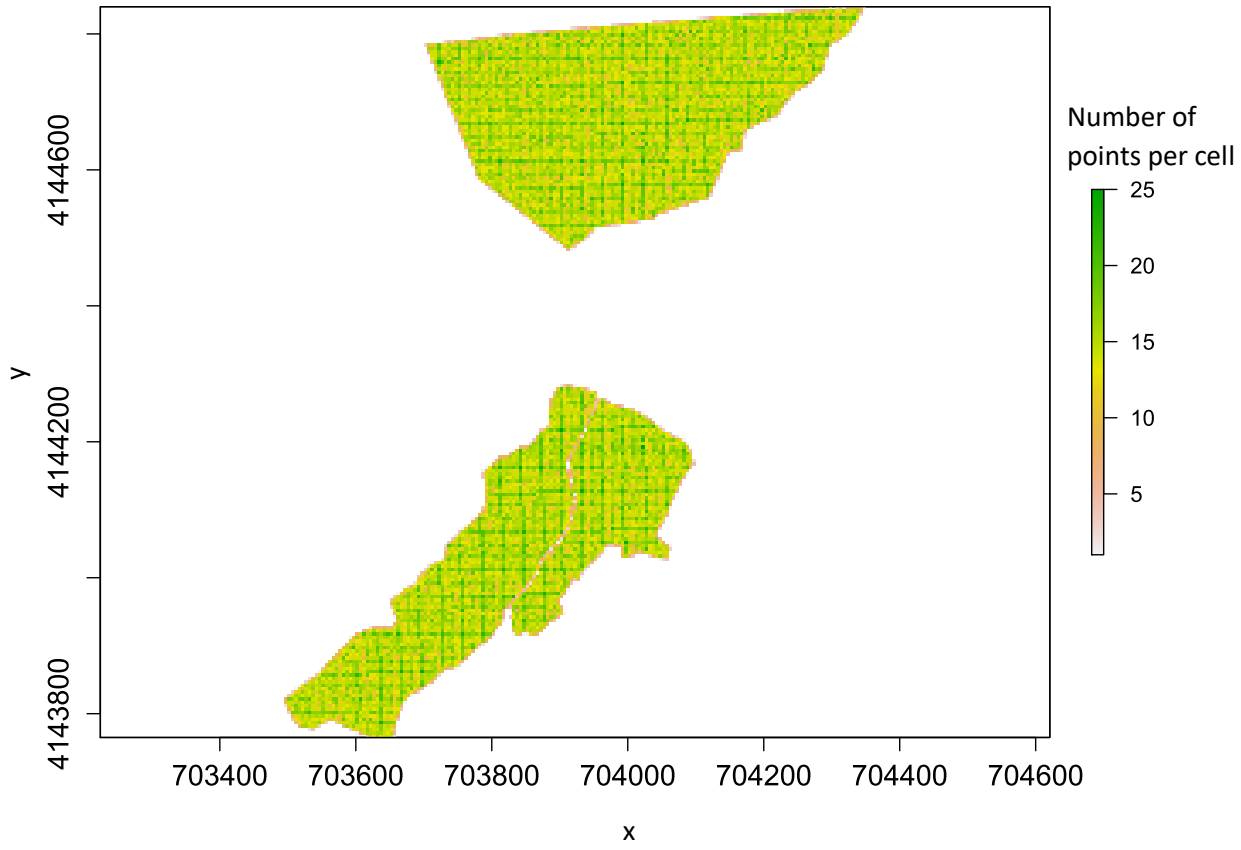


Figure 6. NAIP predicted height and field measured height with the regression line in blue, the confidence band in light grey, and the prediction interval in red dashes.

Equation (1) was then used to create a canopy height model of the three plots in the model application stand. The predicted canopy height model from the NAIP for all three plots in the model application stand is shown in Figure 7a and the point density of the plots is shown in Figure 7b.

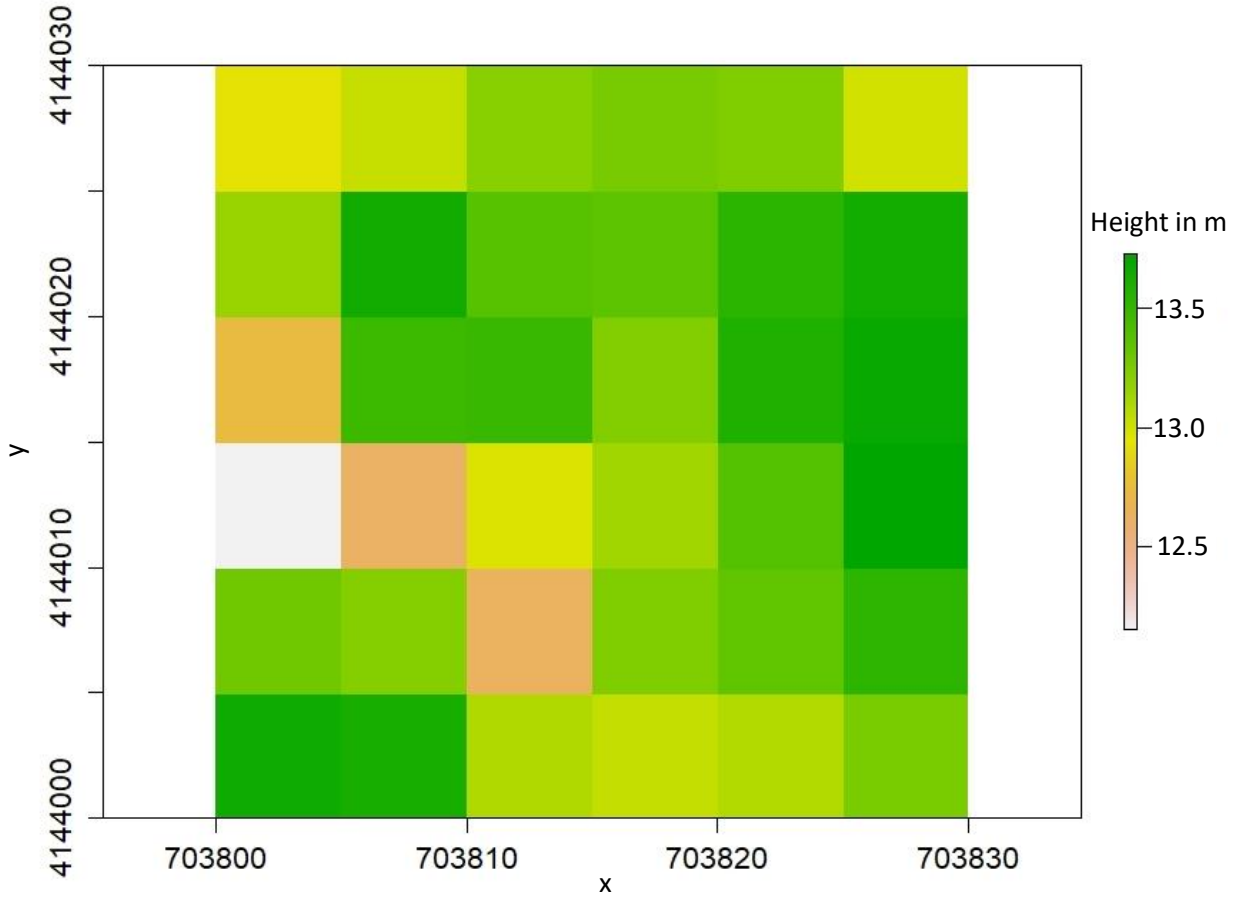


265 a)
266

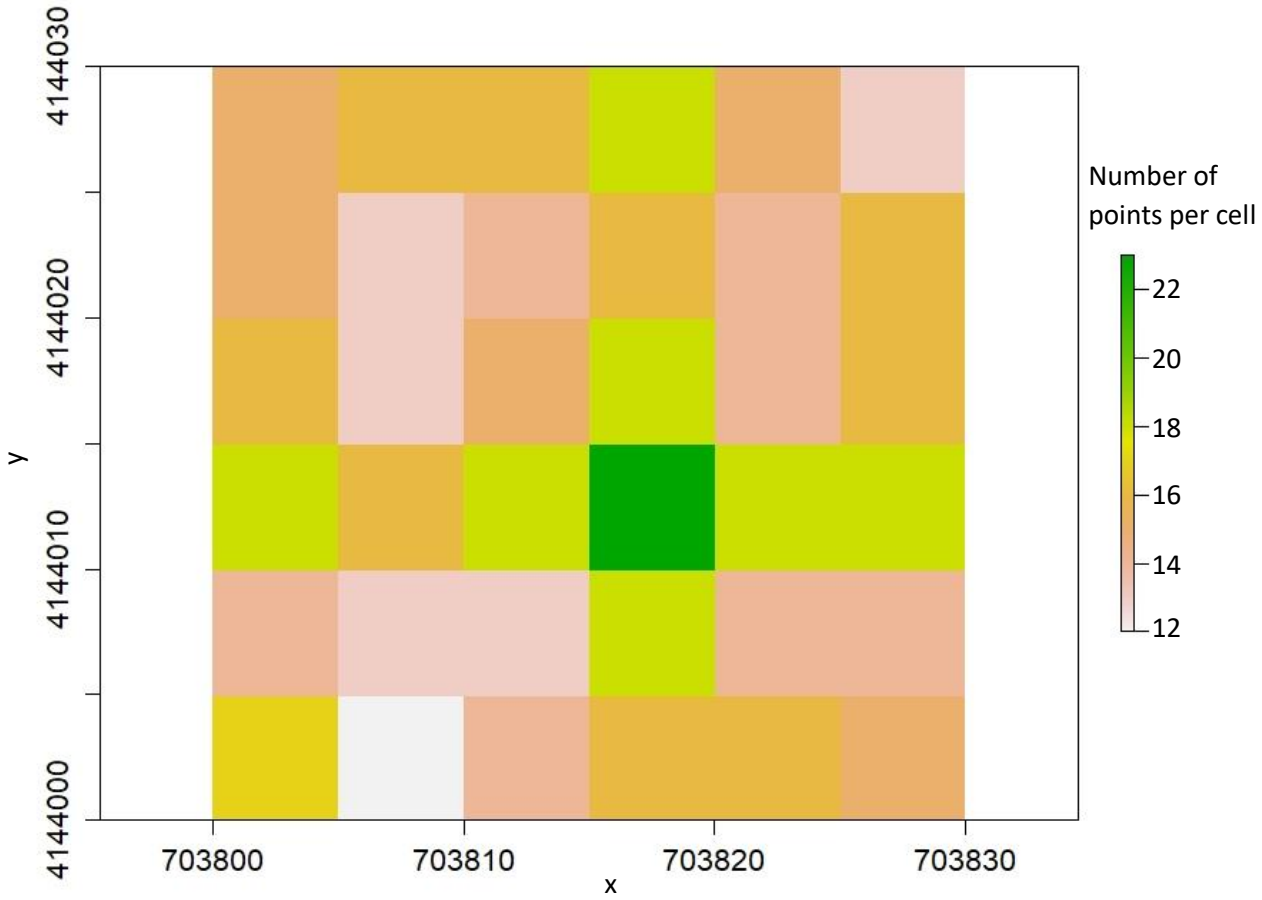


267 b)

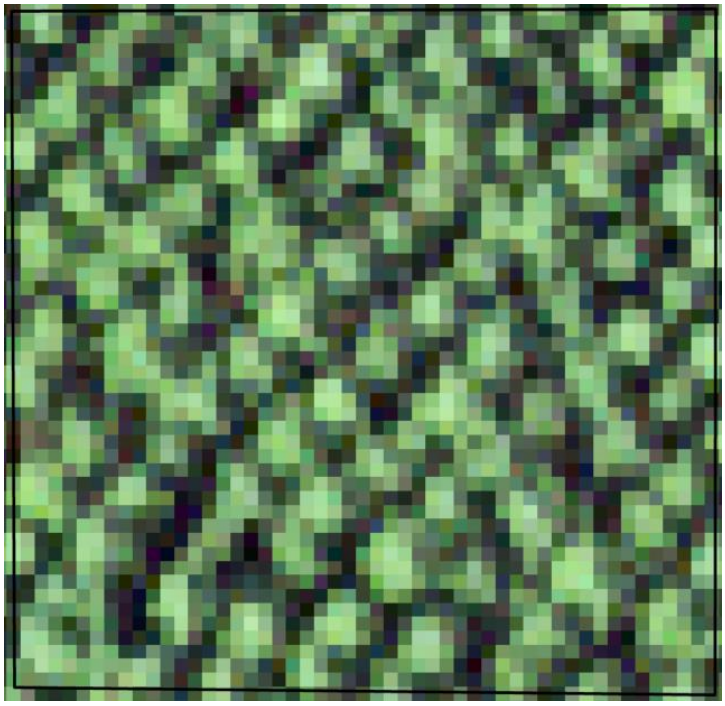
Figure 7. Application plots in Appomattox-Buckingham state forest, where (a) is the canopy height model (CHM) after running the normalized heights through the PPH model and (b) is the point density model. X and Y axis are coordinates in projection GRS80 UTM 18N.



a)



b)



c)

Figure 8. A close-up of the model application site in Appomattox-Buckingham state forest from 703800 to 703830 X and 4144000 to 4144030 Y, where (a) is the canopy height model (CHM) after running the

normalized heights through the PPH model, (b) is the point density model, and (c) is the NAIP imagery. X and Y axes are coordinates in projection GRS80 UTM 18N.

5. Discussion

The future of forest management requires reliable and readily available forest metrics to allow managers to maintain or harvest their stands appropriately. Loblolly pine is one of the most productive and intensely managed timber species, thus presents a necessity for accurate forest metrics to maintain proper management techniques (Fox et al., 2007; Zimmerman et al., 2017; Navarro et al., 2020; Green et al., 2020; Schultz 1999). The advancements of remote sensing provide a faster and cheaper method for forest metric collection compared to previous field collection methods (Noordermeer et al., 2021; Næsset 2002). However, there is the risk of an exchange between reliable data and the cost of data collection (Noordermeer et al., 2021).

Since the beginning of point cloud remote sensing, lidar has been shown by numerous authors to provide accurate and reliable stand height measurements in many forest environments (Masek et al., 2015; Michez et al., 2020; Strunk et al., 2020; Lebarl et al., 2010; Maimaitijiang et al., 2020; Gopalakrishnan et al., 2019). Therefore, lidar has been heavily utilized as a reliable source for forest attribute calculation. Still, there are two large challenges with this form of remote sensing: the costs are quite high, and it is not wall-to-wall across the United States. These two limiting factors have led to the investigation of other remote sensing data sources to measure stand metrics, such as DAP.

DAP point clouds have proven to be a promising alternative to lidar in terms of accuracy in measurement, availability, and cost-efficiency (Goodbody et al., 2019; Strunk et al., 2020; Lebarl et al., 2010; White et al., 2015; Maimaitijiang et al., 2020). The accuracy of DAP to lidar has demonstrated to be strongest when using an ALS derived DEM (Goodbody et al., 2019; Michez et al., 2020; Mielcarek et al., 2020; White et al., 2015; Navarro et al., 2020; Bohlin et al., 2012; Zagalikis et al., 2005; Noordermeer et al., 2021). It also has the strongest correlation with lidar, specifically when measuring species of pine (Mielcarek et al., 2020; Michez et al., 2020). The availability of DAP is not an issue due to a large number of routine data acquisitions from various satellite and aircraft missions, including NAIP (USDA 2011). In terms of costs, DAP derived point clouds can be collected at 1/3 of the costs of lidar, and they are shown to save even more when used with already available aerial imagery (Goodbody et al., 2019; Michez et al., 2020; White et al., 2015; Noordermeer et al., 2021).

The performance of NAIP point clouds and their ability to collect pine heights in this study is strong and even performs better than originally anticipated. High correlation and R^2 values demonstrate that NAIP collects height information at a suitable accuracy for creating canopy height models. The excellent model quality aids the argument that NAIP can be a reliable alternative to lidar in predicting the

canopy height of pine trees. Even though lidar is known for being extremely accurate (Strunk et al., 2020; Bohlin et al., 2012; Masek et al., 2015; Noordermeer et al., 2021; Nelson et al., 2007), this study demonstrates the limitations of lidar availability. Because no lidar data was available at a reasonable acquisition date for the Virginia or North Carolina stands, no comparison in height performance could be made in this study. However, this result does enhance the argument for NAIP to be utilized instead of lidar for height measurement since NAIP is acquired every three years in the US and has homogeneous acquisition parameters (Strunk et al., 2019; Strunk et al., 2020). Inferences can be made on the performance of NAIP versus lidar and what we would have expected in this study from these sources: Kim et al. (2020); Strunk et al. (2020); Lebarl et al. (2010); White et al. (2015); Navarro et al. (2020); Maimaitijiang et al. (2020); Noordermeer et al. (2021); Gobakken et al. (2014); Noordermeer et al. (2019); Lebarl et al. (2010).

In this study, NAIP was reviewed as an alternative form of forest height metric collection when lidar data is not readily available. This study monitored NAIP's ability to accurately calculate loblolly pine in seven stands throughout Virginia and North Carolina. The point clouds were tested against field height measurements to determine the error associated with each remote sensing method. In these training regions, we modeled the 90^{th} percentile of height for the NAIP point clouds against field-measured height for all seven stands and created a regression. This regression model proved to be highly correlated with strong R^2 values, aligning with the results of Mielcarek et al. (2020) and Maimaitijiang et al. (2020). The strong relationship confirmed that NAIP is a suitable alternative measurement of forest height, as we expected from previous studies (Kim et al., 2020; Strunk et al., 2020). The purpose of the model application stand was to compose a code that would apply the *predicted pine height* model to a stand of pines and produce a canopy height model of the pines with no field data. The success of this stand verified the model's ability to capture the predicted height of a pine stand and create an effective canopy height model. The *predicted pine height* model was then ready for application to Virginia, North Carolina, and Tennessee. The model was applied to these states because of the availability of DEMs from lidar and their presence in the southern pine belt.

One limitation in this study is the potential of accidentally applying the model to non-pine areas in the three states. This limitation would be the factor of the not completely corresponding NLCD classes used to map pine locations in the states. Another limitation of this study is that it may not capture the full range of heights for loblolly pine in the Southeast. Because of this possible limitation, mapping pine canopy height outside of these three states should only proceed with caution. A third limitation in this study is the lack of useable lidar data in the study areas that could have been used for comparison in the

training plots. This limitation did not allow NAIP-derived pine heights to be directly compared to lidar-derived pine heights.

Future work will include creating canopy height models for the remaining states in the southeastern United States pine belt. In addition, with the repeated acquisition of NAIP imagery every two to three years, there is the possibility of creating forest growth models for the pine stands, which may be especially interesting to stand managers. Other applications that build off this study could include creating canopy height models for deciduous tree species and modeling stand productivity over time.

6. Conclusion

This study was conducted to create a model that will accurately predict loblolly pine height (and other pines) to an acceptable error using NAIP photogrammetric point clouds. This study validated that for areas of pine, specifically loblolly pine, NAIP can be used to produce a reliable predicted canopy height model. This study also shows that NAIP should be considered a means for mapping pine height in other states in the southern pine belt to create a canopy height model of loblolly pine plantations that can be updated routinely from NAIP's routine acquisitions. The performance of NAIP in this study presents a strong argument for its use in addition to lidar in loblolly pine plantation management across the southeast united states.

References

3D elevation program. *USGS*. https://www.usgs.gov/core-science-systems/ngp/3dep

Bivard, R. (2020). Bindings for the 'Geospatial' data abstraction library. *CRAN*, https://cran.r-project.org/web/packages/rgdal/index.html

Bohlin, J., Wallerman, J., Fransson, J. E. S. (2012). Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM, *Scandinavian Journal of Forest Research*, 27:7, 692-699, doi: 10.1080/02827581.2012.686625

Borders, B. E. & Bailey, R. L. (2001). Loblolly pine-Pushing the limits of growth. *Southern Journal of Applied Forestry*, *25*(2), 69-74. doi: 10.1093/sjaf/25.2.69

Bortolot, Z. J. & Wynne, R. H. (2005). Estimating forest biomass using small footprint lidar data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing 59,* 342-360. doi: 10.1016/j.isprsjprs.2005.07.001

Boyd, D., Danson, F. (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography: Earth and Environment 29:1*, 1-26. doi: 10.1191/0309133305pp432ra

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., & Astrup, R. (2020). A century of national forest inventory in Norway - informing past, present, and future decisions. *Forest Ecosystems*, *7*(1), 46. doi: 10.1186/s40663-020-00261-0

Elseberg, J., Barrmann, D., Nüchter, A. (2011). Efficient processing of large 3D point clouds. *XXIII International Symposium on Information, Communication and Automation Technologies,* 1-7. doi: 10.1109/ICAT.2011.6102102

Fox, T. R., Jokela, E. J., Allen, H. L. (2007). The development of pine plantation silviculture in the southern United States. *Journal of Forestry 105:7* 337-347. doi: 10.1093/jof/105.7.337

Frasco, M. (2018). Evaluation metrics for machine learning. *CRAN*, https://cran.rproject.org/we b/ packages/Metrics/index.html

Gobakken, T., Bollandsås, O. M., Næsset, E. (2015). Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data. *Scandinavian Journal of Forest Research 30,* 73-86. doi: 10.1080/02827581.2014.961954

Goodbody, T. R. H., Coops, N. C., White, J. C. (2019). Digital aerial photogrammetry for updating aera-based forest inventories: A review of opportunities, challenges, and future directions. *Curr Forestry Rep 5,* 55-75 doi: 10.1007/s40725-019-00087-2

Gopalakrishnan, R., Kauffman, J., Fagan, M., Coulston, J., Thomas, V., Wynne, R., Fox, T., Quirino, V. (2019). Creating landscape-scale site index maps for the southeastern US is possible with airborne lidar and Landsat imagery. *Forests 10*, 234. doi: 10.3390/f10030234

Green, P. C., Burkhart, H. E., Coulston, J. W., Radtke, P. J. (2020). A novel application of small area estimation in loblolly pine forest inventory. *Forestry: An International Journal of Forest Research 93(3),* 444-457. doi: 10.1093/forestry/cpz073

Hester, J. (2020). Devtools: Tools to make developing R packages easier. *CRAN*, https://cran.rproject.org /web/packages/devtools/index.html

Hijmans, R. (2020). Raster: Geographic data analysis and modeling. *CRAN*, https://cran.rproject.org/web /packages/raster/index.html

Hogland, J., Anderson, N., St. Peter, J., Drake, J., Medley, P. (2018). Mapping forest characteristics at fine resolution across large landscapes of the southeastern United States using NAIP imagery and FIA field plot data. *ISPRS Int. J. Geo-Inf. 7,* 140. doi: 10.3390/ijgi7040140

Hollister, J. (2020). Access elevation data from various APIs. *CRAN*, https://cran.rproject.org/web/pack ages/elevatr/index.html

Hulet, A., Roundy, B., Petersen, S., Bunting, S, Jensen, R., Roundy, D. (2014). Utilizing national agriculture imagery program data to estimate tree cover and biomass of Pinon and Juniper woodlands. *Rangeland Ecol Manage 67*, 563-572. doi: 10.2111/REM-D-13-00044.1

Iqbal, I. A., Osborn, J., Stone, C., Lucieer, A., Dell, M., McCoull, C. (2018). Evaluating the robustness of point clouds from small format aerial photography over a Pinus radiata plantation. *Australian Forestry* doi: 10.1080/00049158.2018.1482799

Kim, M., Park, S., Irwin, J., McCormick, C., Danielson, J., Stensaas, G., Sampath, A., Bauer, M., Burgess, M. (2020). Positional accuracy assessment of lidar point cloud from NAIP/3DEP pilot project. *Remote Sensing 12,* 1974. doi: 10.3390/rs12121974

412 Kwong, I. H. Y. & Fung, T. (2020). Tree height mapping and crown delineation using lidar. Large format
413 aerial photographs, and unmanned aerial vehicle photogrammetry in subtropical urban forest.
414 *International Journal of Remote Sensing 41,* 5228-5256. doi: 10.1080/01431161.2020.1731002

415 Lebarl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., Wiechert, A. (2010). Point clouds:
416 Lidar versus 3D vision. *Photogrammetric Engineering and Remote Sensing 76,*10 1123-1134.

417 Maimaitijiang, M., Sagan, S., Erkbol, H., Adrian, J., Newcomb, M., LeBauer, D., Pauli, D., Shakoor, N.,
418 Mockler, T. (2020). UAV-based, sorghum growth monitoring: A comparative analysis of lidar
419 and photogrammetry. *Remote Sensing and Spatial Information Sciences 5-3*. doi: 10.5194/isprs-
420 annals-V-3-2020-489-2020

421 Masek, J. G., Hayes, D. J. , Hughes, M. J., Healey, S. P., Turner, D. P. (2015). The role of remote sensing
422 in process-scaling studies of managed forest ecosystems. *Forest Ecology and Management 355:1*,
423 109-123. doi: 10.1016/j.foreco.2015.05.032

424 Michez, A., Huylenbroeck, L., Bolyn, C., Latte, N., Bauwens, S., Lejeune, P. (2020). Can regional aerial
425 images from orthophoto surveys produce high quality photogrammetric canopy height model? A
426 single tree approach in Western Europe. *Int J Appl Earth Obs Geoinformation 92,* 102190. doi:
427 10.1016/j.jag.2020.102190

428 Mielcarek, M., Kamińska, A., Stereńczak, K. (2020). Digital aerial photogrammetry (DAP) and airborne
429 laser scanning (ALS) as sources of information about tree height: Comparisons of the accuracy of
430 remote sensing methods for tree height estimation. *Remote Sens. 12,* 1808. doi:
431 10.3390/rs12111808

432 Moe, K. T., Owari, T., Furuya, N., Hiroshima, T. (2020). Comparing individual tree height information
433 derived from field surveys, lidar and UAV-DAP for high-value timber species in Northern Japan.
434 *Forests 11,* 223. doi: 10.3390/f11020223

435 Næsset, E. (2002). Determination of mean tree height of forest stands by digital photogrammetry.
436 *Scandinavian Journal of Forest Research 17:5*, 446-459. doi: 10.1080/028275802320435469

437 Navarro, J. A., Tomé, J. L., Marino, E., Guillén-Climent, M. L., Fernández-Landa, A. (2020). Assessing
438 the transferability of airborne laser scanning and digital aerial photogrammetry derived growing
439 stock volume models. *International Journal of Applied Earth Observation and Geoinformation,*
440 *91,* 102135. doi: 10.1016/j.jag.2020.102135

441 Nelson, R. F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. J., Campbell, R., Edwards, W. (2007).
442 Investigating RaDAR-lidar synergy in a North Carolina pine forest. *ScienceDirect 110,* 98-108.
443 doi: 10.1016/j.rse.2007.02.006

444 Noordermeer, L., Bollandsås, O. M., Økra, H. O., Naesset, E., Gobakken, T. (2019). Comparing the
445 accuracies of forest attributes predicted from airborne laser scanning and digital aerial
446 photogrammetry on operational forest inventories. *Remote Sensing of Environment 226*, 26-37.
447 doi: 10.1016/j.rse.2019.03.027

448 Noordermeer, L., Gobakken, T., Næsset, E., Bollandsås, O. M. (2021). Economic utility of 3D remote
449 sensing data for estimation of site index in Nordic commercial forest inventories: a comparison of

airborne laser scanning, digital aerial photogrammetry and conventional practices. *Scandinavian Journal of Forest Research, 36*(1), 55-67. doi: 10.1080/02827581.2020.1854340

Nurminen, K., Karjalainen, M., Yu, X., Hyyppä, J., Honkavaara, E. (2013). Performance of dense digital surface models based on image matching in the estimation of plot-level forest variables. *ISPRS Journal of Photogrammetry and Remote Sensing 83,* 104-115. doi: 10.1016/j.isprsjprs.20 13.06.005

OCM Partners. (2021). *2018 Virginia NAIP Digital Ortho Photo Imagery.* https://www.fisheries.noa a.gov/inport/item/58386/citation

Pebesma, E. (2018). Simple features for R. *CRAN*, https://cran.r-project.org/web/packages/sf/index.html

R Core Team. (2020). The R base package. *STAT*, https://stat.ethz.ch/R-manual/R-devel/library/base/DE SCRIPTION

R Core Team (2020). The R stats package. *CRAN*, https://cran.r-project.org/web/packages/STAT/index.h tml

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R. (2017). Digital aerial photogrammetry can efficiently support large-area forest inventories in Norway. *Forestry 90,* 710-718. doi: 10.1093/forestry/cpx027

Roussel, J. (2020). Airborne lidar data manipulation and visualization for forestry applications. *CRAN,* https://cran.r-project.org/web/packages/lidR/index.html

Roussel, J. (2021). Read and write 'las' and 'laz' binary file formats used for remote sensing data. *CRAN,* https://cran.r-project.org/web/packages/rlas/index.html

Schultz, R.P. (1999). Loblolly–the pine for the twenty-first century. *New For. 17*(1–3), 71–88. doi: 10.1023/A:1006533212151

Scolforo, H. F., Montes, C., Cook, R. L., Allen, H. L., Albaugh, T. J., Rubilar, R., Campoe, O. (2020). A new approach for modeling volume response from mid-rotation fertilization of *Pinus taeda* l plantations. *Forests 11*, 646. doi.org/10.3390/f11060646

Shen, X., Cao, L., Yang, B., Xu, Z., Wang, G. (2019). Estimation of forest structural attributes using spectral indices and point clouds from UAS-based multispectral and RGB imageries. *Remote Sens. 11,* 800. doi: 10.3390/rs11070800

Strunk, J. L., Gould, P. J., Packalen, P., Gatziolis, D., Greblowska, D., Maki, C., McGaughey, R. J. (2020). Evaluation of pushbroom DAP relative to frame camera DAP and lidar for forest modeling. *Remote Sensing of Environment 237,* 111535. doi: 10.1016/j.rse.2019.111535

Strunk, J., Packalen, P., Gould, P., Gatziolis, D., Maki, C., Anderson, H., McGaughey, R. (2019). Large area forest yield estimation with pushbroom digital aerial photogrammetry. *Forests 10,* 397. doi: 10.3390/f10050397

Thiel, C. & Schmullius, C. (2016). Comparison of UAV photograph-based and airborne lidar-based point clouds over forest from a forestry application perspective. *International Journal of Remote Sensing*. doi: 10.1080/01431161.2016.1225181

Thomas, V. A., Wynne, R. H., Gopalakrishnan, R., Schroeder, T. A., Green, P. C., Carter, D., Albaugh, T. (2019). Changes in site productivity in the dynamic southeastern United States. *Silvilaser Conference.*

Davis, D. (2011). 2017 NAIP information sheet. *US Department of Agriculture.* https://www.fsa.usda.gov /Assets/USDA-FSA-Public/usdafiles/APFO/support-documents/pdfs/naip_infosheet_2017.pdf

US Geological Survey. (2017a). 1-meter digital elevation models (DEMs) - USGS National Map 3DEP downloadable data collection, *US Geological Survey*. https://www.sciencebase.gov/catalog/item /543e6b86e4b0fd76af69cf4c

US Geological Survey. (2017b). Lidar point cloud - USGS National Map 3DEP downloadable data collection, *US Geological Survey*. https://www.usgs.gov/core-science-systems/ngp/3dep

US Geological Survey. (2019). Virginia FEMA NRCS South Central lidar project. *Dewberry.*

USDA-FPAC-BC-APFO Aerial Photography Field Office. (2019). NAIP digital georectified image. *USDA-FPAC-BC-APFO Aerial Photography Field Office*.

White, J. C., Stepper. C., Tompalski, P., Coops, N. C., Wulder, M. A. (2015). Comparing ALS and image-based point cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. *Forests 6,* 3704-3732. doi: 10.3390/f6103704

Wickham, H. (2016). Create elegant data visualizations using the grammar of graphics. *CRAN*, https://cran.r-project.org/web/packages/ggplot2/index.html

Will, R. E., Narahari, N. V., Shiver, B. D., Teskey, R. O. (2005). Effects of planting density on canopy dynamics and stem growth for intensively managed loblolly pine stands. *Forest Ecology and Management. 205*, 29-41. doi: 10.1016/j.foreco.2004.10.002

Zagalikis, G., Cameron, A. D., Miller, D. R. (2005). The application of digital photogrammetry and image analysis techniques to derive tree and stand characteristics. *Can. J. For. Res. 35*, 1224-1237. doi: 10.1139/X05-030

Zimmermann, S., Hoffmann, K. (2017). Accuracy assessment of normalized digital surface models from aerial images regarding tree height determination in Saxony, Germany. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 85*, 257-263. doi: 10.1007/s41064-017-0021-4

Statewide mapping progress

With the successful creation of a canopy height model, the code used to create it in R was transferred to the ARC and adapted to allow for the application of an entire state of NAIP data. NAIP data was uploaded to the ARC for each state and normalized over DEMs created by lidar. The remaining steps and their progress to date are noted in Table 8. The final model created in this study will consist of all pine stands, managed and unmanaged, across Virginia, North Carolina, and Tennessee.

Table 8. Progress of the application of the PPH model to Virginia, North Carolina, and Tennessee

Task	Progress
Upload all NAIP files to ARC	Completed
Normalize all NAIP files with lidar DEMs	In Progress
Clip normalized point clouds to the extent of NLCD fields 42, 43, and 90	Not Completed
Calculate 90th percentile of height metric on a 5 m x 5 m grid over each normalized, clipped point cloud	Not Completed
Apply the PPH model to each 90th percentile metric	Not Completed
Produce a CHM for each point cloud across the three states	Not Completed

Appendix A: Code for data analysis in R for the training regions

The below code is cut from the NAIP analysis in the Bladen training stand using the RGB bands to display the point cloud. This code was repeated for each training region and was, therefore, run seven times.

```
#Bladen processing of all 4 NAIP types
#Loading in all necessary packages

library(sf)
library(rgdal)
library(raster)
library(elevatr)
library(ggplot2)
library(lidR)
library(devtools)
library(rlas)
library(stats)
library(metrics)
library(base)
library(AICcmodavg)

#In the below code I imported the shapefile for Bladen co. NC

Bladen_shp <- readOGR(dsn = path.expand("D:/Summer1/Plots"), layer = "Bladen_plots_new")

#I created a DTM from the lidar raster I made in ArcGIS Pro. For consistency, all others in this file and
the lidar will use this dtm.

new_raster <- raster("D:/Summer1/LiDAR/Bladen/bladen_ground/Bladen_ground.tif")
elevation <- get_elev_raster(new_raster, z = 9)

#Importing the field data

field_Bla <- read.csv("D:/Summer1/final_data_sheets/Bladen/Bladen_field_avg.csv")

#RGB
```

```
#In the below code, I read in the RGB files for Bladen co. VA

Bladen_RGB <-
readLAS("D:/Summer1/NAIP/Bladen/RGB_4057_541_181022_1614__1_30560_2336.las")

#I made sure the data was in the right projection region (The NAIP data was in UTM Z19 so I moved the
other files to match location instead of moving the NAIP and risking the point cloud).

proj4string(elevation) <- CRS("+proj=utm +zone=19 +ellps=GRS80")

#I then used the elevation dtm to normalize the data, creating a CHM

Bladen_normalized_RGB <- normalize_height(Bladen_RGB, elevation)

#I then clipped the RGB to match each plot

Bladen_RGB_plots <- lasclip(Bladen_normalized_RGB, Bladen_shp)

#To visualize the individual plot point cloud, I plotted it in R which opens a new window with the
point cloud.

plot(Bladen_RGB_plots[[1]])

#Testing the printing of the first plots metrics

Bladen_RGB_metrics_1 <- cloud_metrics(Bladen_RGB_plots[[1]], .stdmetrics)
temp_Bladen_RGB <- as.data.frame(Bladen_RGB_metrics_1)

#This code takes the metrics for each plot in the file

Bla_RGB <- c()
for (i in 1:length(Bladen_RGB_plots)) {
        Bla_RGB[[i]] <- cloud_metrics(Bladen_RGB_plots[[i]], .stdmetrics)
}

#Code that takes the metrics created above and put them in a data frame.

Bladen_RGB_metrics_final <- data.frame(do.call(rbind.data.frame, Bla_RGB))

#Adding the original IDs and reordering the metric table

ID <-
c(1110,1111,1120,1121,1130,1131,1210,1211,1220,1221,1230,1231,1310,1311,1320,1321,1330,1311,14
10,1411,1420,1421,1430,1431,1510,1511,1520,1521,1530,1531,1610,1611,1620,1621,1630,1631,2110,2
111,2120,2121,2130,2131,2210,2211,2220,2221,2230,2231,2310,2311,2320,2321,2330,2331,2410,2411,
2420,2421,2430,2431,2510,2511,2520,2521,2530,2531,2610,2611,2620,2621,2630,2631,3110,3111,3120
,3121,3130,3131,3210,3211,3220,3221,3230,3231,3310,3311,3320,3321,3330,3331,3410,3411,3420,342
1,3430,3431,3510,3511,3520,3521,3530,3531,3610,3611,3620,3621,3630,3631)
Bladen_RGB_metrics_final <- cbind(Bladen_RGB_metrics_final, ID)
Bladen_RGB_metrics_final <- Bladen_RGB_metrics_final[, c(57, 1:56)]
Bladen_RGB_metrics_final <- Bladen_RGB_metrics_final[c(2, 1, 3:108),]

#Writing the metric table to a csv file and exporting
```

```
write.csv(Bladen_RGB_metrics_final,
"D:/Summer1/final_data_sheets/Bladen/Bladen_RGB_Metrics.csv")

#Plotting the RGB against the average over all tree heights in the plot

ggplot(data = NULL, aes(y = Bladen_RGB_metrics_final$zq90, x = field_Bla$Ht_in_M)) +
        labs(title = "Bladen Site: RGB ", x = "Field (M)", y = "RGB (M)") +
        geom_point() + geom_smooth(method = 'lm', formula = y~x)

#Getting the correlation and summary statistics on the 1st model

cor(Bladen_RGB_metrics_final$zq90, field_Bla$Ht_in_M)
rmse(Bladen_RGB_metrics_final$zq90, field_Bla$Ht_in_M)
bla_rgb_model <- lm(Bladen_RGB_metrics_final$zq90 ~ field_Bla$Ht_in_M)
summary(bla_rgb_model)
```

Appendix B: Code for data analysis in R for the application area and statewide mapping

The R code used throughout the testing and validating stages was uploaded to ARC. Since the version of R on ARC was slightly different from the version used for initial development, minor code adjustments were required. Trials were conducted to allow for a more straightforward processing flow for the data. Each state has hundreds of folders, and each folder has dozens of files. As such, loops and LAScatalogs were used to improve the processing efficiency. The operational code was implemented on the ARC Cascades cluster, enabling multicore parallel processing.

```
#Testing the models from the field measured forests on pine stands in ABSF.

#Libraries needed

library(sf)
library(rgdal)
library(raster)
library(elevatr)
library(ggplot2)
library(lidR)
library(devtools)

#Import the shape file created in Arc

shape <- readOGR(dsn = path.expand("D:/Summer2/Model_in_cells"), layer = "Polygons1_final")

#Import the DEM then compute the surrounding elevation.

raster <- raster("D:/Summer2/absf_1m_dem.tif")
elevation <- get_elev_raster(raster, z = 9)
```

```
#Zone projection for Z17 then change the projection of raster and shape to Z19 like NAIP

crs <- "+proj=utm +zone=19 +ellps=GRS80 +units=m +no_defs"
proj4string(shape) <- CRS(crs)
proj4string(raster) <- CRS(crs)

#Making a function to collect only the 90th percentile of the data

metric <- function(z) {q90 = quantile(z, probs = c(0.9))}

#The model that was determined from the NAIP imagery

# y = 1.09x -0.43, cor = 0.96, R^2 = 0.93

#RGB

#Import the RGB file

RGB <- readLAS("D:/Summer2/NAIP/mergedRGB.las")

#Clip the data to three stands in ABSF then normalize the heights from the RGB

RGB_clipped <- lasclip(RGB, shape)
rgb_norm <- c()
for (i in 1: length(RGB_clipped)) {
        rgb_norm[[i]] <- normalize_height(RGB_clipped[[i]], raster)
}

#Calculating the 90th percentile of height for each cell in each plot

rgb_90 <- c()
for (i in 1:length(rgb_norm)) {
        rgb_90[[i]] <- grid_metrics(rgb_norm[[i]], ~metric(Z), 2.2361) #2.2361 is the sqrt of 5 for 5m.
}

#Viewing the three plots together and individually

together <- merge(rgb_90[[1]], rgb_90[[2]], rgb_90[[3]])
plot(together, main = "RGB Height")
plot(rgb_90[[1]], main = "RGB Height Plot 1")
plot(rgb_90[[2]], main = "RGB Height Plot 2")
plot(rgb_90[[3]], main = "RGB Height Plot 3")
RGBGrid_metrics_final <- data.frame(do.call(rbind.data.frame, rgb_90))

#This section of code will repeat the above but after applying the model from the original
plots to the normalized heights

#Creating a copy of the normalized data to work with

expected_RGB <- rgb_norm
```

```
#Changing the normalized height(Z) to the expected height based off the model

expected_RGB[[1]]@data[["Z"]] <- (1.09 * expected_RGB[[1]]@data[["Z"]]) - 0.43
expected_RGB[[2]]@data[["Z"]] <- (1.09 * expected_RGB[[2]]@data[["Z"]]) - 0.43
expected_RGB[[3]]@data[["Z"]] <- (1.09 * expected_RGB[[3]]@data[["Z"]]) - 0.43

#Calculating the 90th percentile of height for each cell in each plot

exp_metrics_1 <- grid_metrics(expected_RGB[[1]], ~metric(Z), 2.2361) #2.2361 is the sqrt of 5 for 5m.
exp_metrics_2 <- grid_metrics(expected_RGB[[2]], ~metric(Z), 2.2361)
exp_metrics_3 <- grid_metrics(expected_RGB[[3]], ~metric(Z), 2.2361)

#Viewing each plot

expected_together <- merge(exp_metrics_1, exp_metrics_2, exp_metrics_3)
plot(expected_together, main = "Expected RGB")
plot(exp_metrics_1, main = "Expected RGB Plot 1")
plot(exp_metrics_2, main = "Expected RGB Plot 2")
plot(exp_metrics_3, main = "Expected RGB Plot 3")
----------------------------------------------------------------------------------------------------
```