

Improving Precision in Forest Inventory Using Small Area Estimation in Loblolly Pine  
Plantations in Coastal Georgia

Bipana Subedi

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

Phil Radtke, Chair

Valerie Thomas

John W. Coulston

12/12/2024

Blacksburg, VA

Keywords: forest inventory, remote sensing, small area estimation, unit-level SAE

# Improving Precision in Forest Inventory Using Small Area Estimation in Loblolly Pine Plantations in Coastal Georgia

Bipana Subedi

## ABSTRACT

The use of small area estimation (SAE) in forest inventory has shown promise for improving the precisions of estimates needed for informed decision-making when sample data are sparse. We evaluated the potential of unit-level SAE for increasing the precision of stand-level estimates of basal area, volume, and above-ground biomass estimates in loblolly pine plantations in coastal Georgia. Following the unit-level approach, field plots sampled in plantations owned by Rayonier Inc. were georeferenced to aerial lidar data using high-quality GPS field coordinates. Results focused on A) gains in precision for stand-level basal area, volume, and above-ground biomass estimates achieved by combining data from field plots with lidar-derived canopy height models in a SAE framework, B) impacts of small sample sizes on the precision of estimated stand level attributes, and C) the effects of nonrandom field plot placement in stands of interest when using unit-level SAE. Findings indicate that higher precision is achievable with greater variance stability than what is possible from very small samples of field data alone. This was true for all three attributes of interest. With careful attention to checking assumptions of the unit-level SAE approach, the use of non-random sampling does not appear to impair SAE's ability to deliver unbiased estimates for forest plantation stands. Simulating the entire population's basal area to test for the effects of non-random plot placement showed that SAE is robust to the type of sampling technique used. However, results can be affected when sampling is intentionally biased. This work can be useful to landowners and forest managers working with southern loblolly pine plantations. By leveraging simulation techniques to generate non-random sampling data from the available random sampling data, this study attempted to bridge the gap between the available empirical data and the desired sampling framework, ultimately widening the applicability of SAE in forest inventory settings.

# Improving Precision in Forest Inventory Using Small Area Estimation in Loblolly Pine Plantations in Coastal Georgia

Bipana Subedi

## GENERAL AUDIENCE ABSTRACT

Accurate forest inventory estimates are essential to make important decisions for forest management. Our research explored how advanced statistical methods, specifically small area estimation (SAE), can enhance forest inventories when only limited data is available. We focused on loblolly pine plantations in coastal Georgia, using data from field plots combined with aerial lidar technology to estimate important forest metrics: basal area (tree density), wood volume, and above-ground biomass. By pairing field and lidar data, we found that SAE significantly improved the accuracy of forest estimates, even when the number of field samples was very small. We also tested how different sampling strategies, such as non-random plot election, affected the results. Our results showed that SAE proved resilient to non-random sampling as long as certain assumptions were met. However, deliberate biases in sampling could still lead to less reliable estimates. Our findings provide valuable tools for forest managers and landowners, especially those managing loblolly pine plantations in the Southeastern US. By applying simulation techniques to extend the use of existing data, this study showed how SAE can fill data gaps and provide more accurate forest measurements, helping to guide better management and conservation decisions.

## **Acknowledgements**

Completing this thesis has been a challenging yet immensely rewarding journey, and it would not have been possible without the support, guidance, and encouragement of many individuals and organizations. First and foremost, I am deeply grateful to my advisor, Dr. Phil Radtke, for his unwavering support, insightful guidance, and constant encouragement throughout my graduate studies. Your expertise, patience, and belief in my abilities have been instrumental in shaping my academic journey. I would also like to extend my gratitude to my committee members, Dr. Valerie Thomas and Dr. John W. Coulston for their valuable feedback and advice, which greatly enhanced the quality of this work. Special thanks go to my colleagues and peers in the Forest Biometrics program at Virginia Tech for creating a collaborative and supportive environment that has enriched my learning experience. A heartfelt thank you to Rayonier Inc. for providing the data and resources essential to this research. I am grateful for the opportunity to collaborate with such a forward-thinking organization and for the exposure to practical applications of forestry research. To my family, whose love and sacrifices have been my constant source of motivation, thank you for believing in me even when I doubted myself. To my friends, both near and far, thank you for your encouragement and for keeping me grounded during this journey. The beauty and complexity of our forests remind me of the importance of the work we do to ensure their sustainability for generations to come. This thesis is a testament to the collective efforts of all those who have supported me, and I am profoundly thankful to each of you.

## Table of Contents

<b>Acknowledgements</b> .....	4
<b>Table of Contents</b> .....	5
<b>List of Figures</b> .....	7
<b>List of Tables</b> .....	8
<b>List of Abbreviations</b> .....	9
<b>Introduction</b> .....	10
<b>Material and Methods</b> .....	17
<b>Study Area and Direct Estimates</b> .....	17
<b>Auxiliary Data</b> .....	20
<b>Data Analysis</b> .....	22
<b>Direct estimators</b> .....	22
<b>Unit-level BHF model</b> .....	22
<b>Thinning effect in the model</b> .....	24
<b>Open ground relationship in the model</b> .....	26
<b>Precision gains:</b> .....	27
<b>Population Data Simulation</b> .....	27
<b>Results</b> .....	32
<b>EBLUP Comparisons to Sample Means</b> .....	32
<b>RMSE Comparison</b> .....	36
<b>Discussion</b> .....	39
<b>Conclusion</b> .....	41
<b>References</b> .....	43
<b>Appendices</b> .....	49
<b>Appendix I: Stand CHMs (with heights in meters) where EBLUP underestimated stand volume (+ symbols, figure 5)</b> .....	49
<b>Appendix II: Stand CHMs (with heights in meters) where EBLUP overestimated stand volume (▲ symbols, figure 5)</b> .....	63
<b>Appendix III: Plot level and stand level predictor variables obtained for natural log of number of zeroes for stands and their respective two plots</b> .....	69

**Appendix IV: Fixed effects coefficient estimates obtained for all three models (Volume, AGB, and Basal Area) with the model form: ..... 76**

**List of Figures**

Figure 1. Study area map showing surveyed Rayonier stands (red) across Coastal Georgia. .... 18

Figure 2. Two example stands and sample plot positions in the study area overlain with CHM data (heights in m). ..... 21

Figure 3. Comparison of stand-level attributes to the mean canopy height of the plot for volume, above-ground biomass, and basal area..... 25

Figure 4. Comparison of stand-level attributes to the proportion of zeroes in the canopy for estimates of basal area, above-ground biomass, and volume (top: without transformation; bottom: with natural log transformation). ..... 26

Figure 5. BHF composite estimates (EBLUP) of stand-level attribute totals compared to direct estimates for volume, above-ground biomass, and basal area. .... 32

Figure 6. Stand 076 circular plot boundaries in relation to the stand boundary and auxiliary CHM values (m)...... 33

Figure 7. Stand 2020028 plot locations. .... 34

Figure 8. Standard errors obtained from direct estimates compared to that of the EBLUPs from unit-level SAE for all three attributes of interest. .... 36

Figure 9. Distribution of SER across all three attributes of interest: volume, AGB, and basal area. .... 38

## List of Tables

Table 1. Summary statistics of directly collected sample data for plantation stands in the study area.....	18
Table 2. Plot and stand level predictor variables obtained for stand 028 along with its two plots.....	35
Table 3. Summary statistics of SER for basal area, volume, and above ground biomass estimates.....	37

## List of Abbreviations

AGB	-	Above Ground Biomass
BHF	-	Battese Harter Fuller
CHM	-	Canopy Height Models
Ha	-	Hectares
M	-	Meters
Mg	-	Megagram
RMSE	-	Root Mean Squared Error
SAE	-	Small Area Estimation
SE	-	Standard Error
SER	-	Standard Error Ratio

## Introduction

Accurate forest inventory estimates play a pivotal role in informing sound decision-making processes and guiding policy formulation (Green et al. 2020b). To generate such estimates, foresters often rely on design-based surveys using fixed-area or variable-radius sample plots within areas of interest (Burkhart et al. 2019). This is called direct estimation, where forest parameters are estimated from directly observed ground-based sample units. While direct estimation proves effective when large samples are collected – often over large areas – obtaining reliable direct estimates can be challenging for relatively small areas that may contain a small number of inventory plots (Reich and Aguirre-Bravo 2009; Mauro et al. 2016; Magnussen et al. 2017).

Forest inventories are generally designed and executed with a goal of achieving a targeted precision or “the smallest possible standard error given a fixed level of effort” (Kershaw et al. 2017, 307). The number of sample plots collected in an area of interest plays an important role in determining the allowable error and precision level for estimating forest attributes. Depending on the number of sample plots observed, an area or domain can be considered large or small. A domain is regarded as large if the domain-specific sample is large enough to yield direct estimates of adequate precision; it is regarded as “small” if the domain-specific sample is not large enough to support direct estimates of adequate precision (Rao and Molina 2015). In inventory settings where estimates are mainly needed for extensive geographic regions or other large-area domains, design-based estimation often works well. The situation changes as forest managers or landowners increasingly demand estimates for small area domains needed for decision support (Fuller 1999). This adaptation is particularly relevant in plantation management,

where greater precision in characterizing plantation attributes requires greater sampling effort. At some point, inventory costs may become prohibitive when high-quality information is needed for many small areas at multiple points in time.

Loblolly pine (*Pinus taeda*, L.) and slash pine (*P. elliotti*, Englem.), two members of the Southern Yellow Pine (SYP) species group form the most commonly planted commercial tree species assemblage in the Southeastern United States (Baker and Langdon 1990). Assessing commercial and ecological conditions over time requires that forest inventories provide accurate and timely information at fine scales, especially where stand-level and tract-level management are relevant (Green et al. 2020a). Increasing sample sizes in small areas to achieve necessary precision from direct estimation incurs greater costs, related to added requirements for resources like time, effort, and labor.

To deal with limited sample size problem, supplementing inventory data with auxiliary information can be a useful method to predicting forest attributes at unsampled locations, ultimately enhancing effective sample sizes without requiring greater sampling effort (Rao 2015; Cao et al. 2022). In such cases, Small Area estimation (SAE), using model-based statistical estimators has shown potential in increasing forest inventory precision for small areas by leveraging auxiliary sources of information like remote sensing and combining them with field observations (Green et al. 2020a; Cao et al. 2022).

Increasingly, auxiliary remote sensing information such as lidar or photogrammetric point clouds, or satellite imagery may be available to incorporate into forest inventory in SYP plantations. Light Detection and Ranging (lidar), often acquired by airborne laser scanning (ALS), sensors is increasingly coupled with georeferenced inventory plot measurements to

improve large- and small-scale estimation of forest inventory parameters (Blinn et al. 2019). Pairing aerial lidar with field-based inventory data offers an opportunity to employ SAE techniques, thereby enhancing the precision of estimated forest attributes for plantation stands without necessitating an increase in sample size or field effort.

Seminal work on SAE was presented by Rao and Molina (2003), who presented a comprehensive framework encompassing various techniques within the SAE framework. SAE-related methods are usually categorized as (1) domain-direct estimation, (2) domain-indirect or synthetic estimation, and (3) composite estimators that combine domain direct and synthetic information (Rao and Molina 2015). Domain-direct estimation focuses solely on sample data collected within a small area of interest, both for attribute estimates such as means or total and for estimates of their variances or mean squared error (MSE). Domain-indirect estimation is a model-dependent approach that employs statistical models that “borrow strength” in making estimates for one small area based on sample survey data collected in other areas or from other points in time periods, intending to reduce standard errors of the estimates within targeted small area domains. Indirect estimators leverage observed sample data ( $Y$ ) beyond a specific domain of interest by integrating it with auxiliary data ( $X$ ) through a model, typically described as a synthetic or assistive model (Rao, 2003; Lehtonen & Veijanen, 2009). This approach improves the precision of population parameter estimates. While direct estimators are usually unbiased under their sampling design, they can be unstable or exhibit high variances when sample sizes are small. In contrast, indirect estimators, such as regression models, often yield precise predictions but may introduce substantial biases if the underlying model assumptions are violated. To combine the unbiased nature of direct estimators with the enhanced precision of

indirect (model-based) estimators, a composite estimator is formed by weighting the two. This composite estimator is designed to balance the unbiasedness of the direct component with the low variance characteristic of the synthetic model. Two widely used approaches in small area estimation (SAE) – area-level and unit-level models – rely on a mixed modeling framework. This framework accounts for the variability within small area sample estimates and the residual variation across domains as predicted by the model (Fay & Herriot, 1979; Battese et al., 1988).

Rao (2003, Ch. 5) distinguished between the area-level and unit-level approaches in small-area model-based domain estimation. In the forest inventory literature, both area-level and unit-level small area estimation (SAE) methods have demonstrated their ability to improve the precision of estimates for small area domains while largely preserving unbiasedness (Breidenbach & Astrup, 2012; Goerndt et al., 2013; Magnussen et al., 2017; Mauro et al., 2017; Green et al., 2020a). Area-level approaches are often linked to the Fay and Herriot (1979) estimator, a well-established and widely utilized technique in area-based SAE applications. Fay Herriot methodology has been applied to several instances within forest inventory (Cao et al. 2022; Green et al. 2020b; Temesgen 2022). Area-level approaches explicitly link sample observations in aggregate to auxiliary data at the scale of domains from which they were sampled.

Unlike the area-level approach, unit-level small area estimation (SAE) integrates data from individual sample units with corresponding auxiliary data sources (Rao & Molina, 2015). Unit-level analyses, which involve combining field sample observations with geospatial auxiliary data or digital maps, necessitate accurate coordinates of the field sample plots. These coordinates are usually obtained using global navigation satellite systems, such as GPS, to enable

the precise alignment of field plot data with co-occurring observations from remote sensing or other geospatial data layers. Unit-level small-area estimation methodologies are frequently associated with the Battese-Harter and Fuller (1988; BHF) BHF estimator. BHF used a widely adopted nested error regression model approach for unit-level SAE that can be used when field-plot survey data and auxiliary information from remote sensing are both available and can be paired, typically via precise geospatial coordinates, at the field-plot level.

In forestry research, the use of non-random sampling techniques represents a departure from the traditional reliance on random sampling methodologies. Random sampling, the most fundamental sampling technique used in forest inventories (Kershaw et al. 2017) is a type of sampling where every potential sample plot within an area of interest has a known probability of being selected. Examples of studies involving applications of SAE using random sampling are extensive (Green et al. 2020a, 2020b; Goerndt et al. 2011; Cao et al. 2022). Non-probability sampling designs, like non-random or purposive sample selection, allow for plots to be established without rigorously following probabilistic rules for selecting sample units. This can enable researchers to use subjective judgments or heuristic guidelines to select samples rather than following a strictly random design (Etikan, Musa, and Alkassim 2016), with a possible loss of unbiasedness in estimation. While purposive or non-random sampling is rarely used to obtain inventory estimates, it does have a legitimate role in some resource assessments where designs involving systematic or random sampling fail to efficiently provide the desired information (Goff et al. 1982). Examples include growth and yield studies (Kershaw et al. 2017), studies requiring destructive sampling (Radtke et al. 2017), detection of rare and invasive species (Goff et al. 1982), or studies of species diversity (Elzinga et al. 2001). While random sampling is not a

prerequisite to employing all SAE techniques (Sterba 2009), most studies using SAE in forest inventories have worked with random sampling to take advantage of its statistical robustness. A somewhat novel aspect of this study includes evaluating the effectiveness of SAE techniques in delivering unbiased estimates even while using non-randomly placed sample plots. In this context, the adoption of non-random sampling methods introduces an avenue for exploring the capabilities of unit-level SAE in addressing a broader range of management or research needs.

As far as we are aware, no studies have specifically explored the application of small area estimation (SAE) techniques in the presence of both random and non-random plot placements, especially in the case of commercially managed SYP plantations. Simulation studies in the context of SAE are often presented very briefly (Warnholz and Schmid 2016). By leveraging simulation techniques to generate non-random sampling data from available random sampling data, this study aimed to bridge the gap between the available empirical data and the desired sampling framework, potentially widening the applicability of SAE in forest inventory settings. The primary objective of this work was to assess the effectiveness of unit-level SAE in improving the precision of stand-level volume, above-ground biomass (AGB), and basal area estimates in loblolly and slash pine plantations in Coastal Georgia. A secondary goal was to evaluate unit-level SAE for delivering unbiased results where very small sample sizes or non-random plot placement were employed. Working with industrial plantation forest inventory data from eight counties in Coastal Georgia, BHF unit-level estimation was the primary SAE method evaluated. Three research questions were addressed: (1) What gains in precision are realized for stand-level plantation volume, AGB, and basal area, estimates achieved by integrating field plot data with lidar-derived canopy height models in a unit-level SAE

framework?; (2) How do small sample sizes ( $n \leq 2$ ) impact the precision of estimated forest attributes, and what strategies can be employed to mitigate the effects of limited sample size in SAE?; and (3) Can unit-level SAE techniques deliver unbiased estimates for plantation stands even when non-random sampling is employed, and what are the key considerations for ensuring the validity of these estimates?

## Material and Methods

### Study Area and Direct Estimates

The study area included Loblolly pine plantation stands representative of industrial forest management practices in Coastal Georgia, USA. The data for this study were provided by Rayonier Incorporated (Rayonier Inc., Wildlight, FL). For this study, we used two different sources of data: directly sampled forest inventory data and canopy height model (CHM) data collected from lidar. The study encompassed eight counties in coastal Georgia, namely Brantley, Camden, Charlton, Liberty, Long, McIntosh, Pierce, and Wayne. A map of the study area (Figure 1) illustrates the spatial distribution of the stands of interest. Stands covered a wide range of ages and included thinned and unthinned forests (Table 1). A total of 114 stands were identified and surveyed within the study area with lidar coverage available for 104 of the stands. Among them, 68 were unthinned, and 36 were thinned prior to the period of data collection.

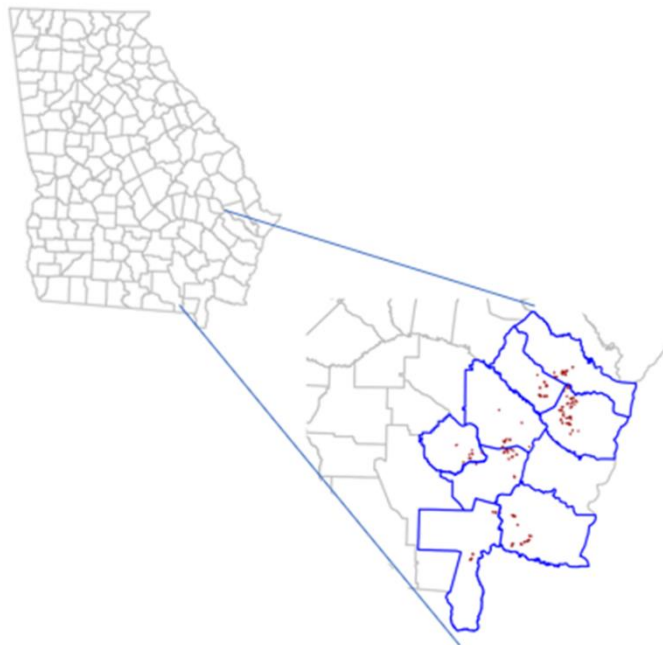


Figure 1. Study area map showing surveyed Rayonier stands (red) across Coastal Georgia.

Table 1. Summary statistics of directly collected sample data for plantation stands in the study area.

<b>Parameters</b>	<b>Min.</b>	<b>Mean</b>	<b>Max.</b>
<b>Area (ha)</b>	2.4	20.3	91.2
<b>Age (yrs)</b>	10	16.8	27
<b>Basal area (m<sup>2</sup>/ha)</b>	10.3	26.6	42.3
<b>AGB (Mg/ha)</b>	24.6	115.2	254.4
<b>Volume (m<sup>3</sup>/ha)</b>	32.6	170.3	375.6

Sample plots used for data collection were circular in shape, each with an area of 0.1 acre (.04047 ha). The geospatial position of each plot center was referenced with sub-meter accuracy using a Trimble (Trimble Inc., Westminster, Colorado) R1GNSS GPS receiver. The type of sampling method used for data collection was random sampling, although field notes indicated a small number of plots may have been purposively located away from stand boundaries, water, and disturbed areas. The number of sample plots in each stand was very small, i.e.  $n = 2$  in each stand, with the exception of four plots in one stand, making for a total of 210 sample plots in the 104 stands making up the study area.

Forest inventory data were collected over four months, from January to April 2020. Sampled stand-level data were available from two plots in each stand to produce estimates of volume, biomass, and basal area with associated sampling errors for each of the 104 stands. Stand estimates and their associated standard errors served as direct forest inventory estimates,

which were evaluated for potential precision improvements through the application of small area estimation (SAE) techniques. Due to the small within-stand sample sizes, it was expected that estimates may have been relatively unstable, especially sample variances based on  $n = 2$ . The attributes of interest estimated for each stand were the means of total stem volume ( $\text{m}^3/\text{ha}$ ), total aboveground biomass ( $\text{Mg}/\text{ha}$ ), and basal area ( $\text{m}^2/\text{ha}$ ). On each field plot, live trees were measured for diameter at breast height (dbh, 1.47 m). Roughly half of all planted SYP trees were measured for total height, with loblolly and slash pines comprising two-thirds and one-third of the sampled trees, respectively. Only live stems with  $\text{dbh} \geq 2.54$  cm (1 inch) were considered here, with the smallest measured tree height of 3.05 m (10 feet). In addition to measured heights, the following measurements were made on all trees: species or species groups (loblolly pine, slash pine, or other); dbh; and crown class (1 for dominant or codominant; 2 for intermediate or suppressed).

Unmeasured planted SYP heights were predicted using a height-diameter equation attributed to M. Näslund by Sharma and Breidenbach (2015)

$$ht = 1.37 + \left( \frac{D}{a + a_i + (b + b_i)D} \right)^3 + e \quad [1]$$

Where  $ht$  is total tree height (m),  $D$  is the diameter at breast height (dbh) in cm,  $a$  and  $b$  are fixed-effects parameters,  $a_1$  and  $b_1$  are random effects parameters to be estimated for each combination of stand ID and crown class, and  $e$  is a tree level residual error. In fitting Eq. [1] parameters  $a$  and  $b$  were treated as fixed-effects parameters with random effects parameters  $a_1$  and  $b_1$  allowed to vary by stand and crown class to capture differences in height versus dbh relationships for the two crown classes and to account for site-specific height, dbh relationships

either due to stand-specific growing conditions or the SYP species (loblolly or slash pine) planted in different stands. Coefficients were estimated using package ‘nlme’ in the R statistical computing environment (Pinheiro et al. 2023, Pinheiro and Bates 2000, R Core Team 2024). The fitted model was subsequently used to impute the height predictions to trees not having measured heights.

Total stem outside-bark volume and aboveground biomass were predicted from published equations for planted loblolly and slash pines developed by Westfall et al. (2023) as a function of dbh, total height, and species.

$$\text{Vol} = \begin{cases} aD^bH^c; & D < k \\ a * k^{(b-b_1)} * D^{b_1} * H^c; & D \geq k \end{cases} \quad [2]$$

Where  $D$  = dbh,  $H$  = total height,  $k = 22.9$  cm (9 inches) is a dbh segmentation point for SYP, and fitted coefficients  $a$ ,  $b$ ,  $b_1$ , and  $c$  are species-specific coefficients given in the publication (Westfall et al. 2023).

### **Auxiliary Data**

Canopy Height Models (CHM) derived from aerial lidar were used as auxiliary information in SAE analyses. Although CHMs were developed prior to their use here and delivered as a 0.5 m raster grid, examination of point clouds used to generate the CHM indicated an average pulse density of 39 returns per square meter. The CHM rasters were clipped to individual stand boundaries using polygon vector data provided by Rayonier for stands in the study area. CHM data were also clipped to the circular sample plots based on their center points and plot geometry (Figure 2).

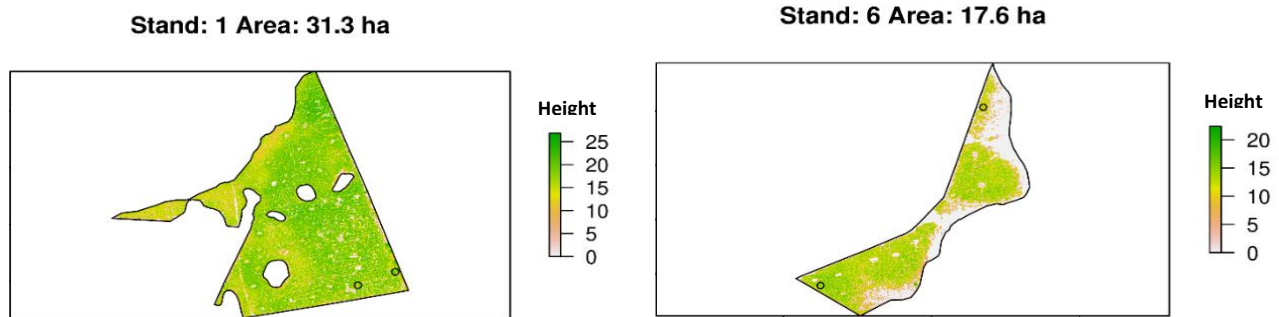


Figure 2. Two example stands and sample plot positions in the study area overlain with CHM data (heights in m).

Plot-level CHM metrics for use as regression predictors were extracted by first binning the raster canopy heights into 1 m discrete classes including a zero class for grid cells classified as uncovered ground. Cell bin heights were summed to calculate the total canopy height (m) for each plot. We denoted the mean height of all cells as  $\overline{ht}_0$  to indicate that zero values from uncovered ground affected its values. A complementary mean height ( $\overline{ht}_*$ ) was calculated with zero grid cells omitted to give a more accurate representation of the canopy heights of trees standing on the plot, regardless of open space among trees. A third possible predictor was formulated as the number of uncovered (zero) pixels in each plot, which we denoted as the proportion of open ground ( $P_o$ ). A preliminary calculation of pairwise correlation coefficients ( $r$ ) between the predictors and the plot level basal area showed moderate correlations ( $r = 0.73$  and  $-0.73$ ) for  $\overline{ht}_0$  and  $P_o$ , respectively. Lower correlations with the basal area were noted for other plot attributes including  $\overline{ht}_*$  at ( $r = 0.37$ ), stem density (trees/ha) at ( $r = 0.54$ ), and a thinning indicator (0 for unthinned; 1 for thinned) at ( $r = -0.34$ ). The same CHM metrics were determined for each stand within its boundaries to serve as stand population values for predictors.

## Data Analysis

### Direct estimators

Forest inventory direct estimates of domain-specific mean stand-level volume, biomass, and basal area were calculated using the conventional formula for the sample mean assuming simple random sampling.

$$\bar{y}_d = \frac{1}{n_d} \sum_{i=1}^{n_d} y_{di} \quad [3]$$

Where  $\bar{y}_d$  = domain (stand) mean of volume, AGB, or basal area per hectare,  $y_{di}$  = attribute of interest observed on plot  $i$  in stand domain  $d$ , and  $n_d$  = number of plots in each domain in all but one stand  $n_d = 2$ . The same formula was used to calculate stand-specific stem volume and AGB per unit area. And the associated standard error for each stand was

$$SE(\bar{y}_d) = \sqrt{\frac{\sum_{i=1}^{n_d} (y_{di} - \bar{y}_d)^2}{n_d(n_d - 1)}} \quad [4]$$

as defined for eq [3].

### Unit-level BHF model

The unit-level small area model developed by Battese, Harter, and Fuller (BHF) (1988) was utilized for conducting SAE analyses. This approach combines two estimates to produce Empirically Best Linear Unbiased Predictions (EBLUPs) for the attribute of interest within each domain. For each stand (domain), the three attributes –total stem volume, AGB, and basal area – were of primary interest.

In the BHF model, a plot observation such as total stem volume ( $\text{m}^3/\text{ha}$ ) is formulated as the sum of a fixed-effects model prediction (also called the synthetic prediction) plus a stand-level random effect, plus a residual term for the plot representing the difference between the plot observation and the mixed effects (fixed plus random) model prediction. Based on the BHF model, the observed response  $y_{di}$  from plot  $i$  in domain  $d$  was composed as

$$y_{di} = \mathbf{x}_{di}^T \boldsymbol{\beta} + u_d + e_{di} \quad [5]$$

and, their means using eq [3]

$$\bar{y}_d = \frac{\sum_{i=1}^{n_d} y_{di}}{n_d} \quad [6]$$

where  $i = \text{plot index (1,2)}$ ,  $d = \text{domain (stand) index}$ ,  $y_{di}$  is the value of a variable of interest  $Y$  (e.g., volume, biomass, or basal areas) for unit  $i$  in domain  $d$ ,  $\mathbf{x}_{di}$  is a p-vector of auxiliary information (lidar data) for sample plot  $i$  in domain  $d$ , and  $\boldsymbol{\beta}$  is a vector of fixed effects associated with the auxiliary information. The model formulation (eq. [5]) assumes there is one set of fixed-effects parameters  $\boldsymbol{\beta}$  that define the relationship between the direct and auxiliary information across all domains, and that  $u_d$  is a domain-specific random effect with  $u_d \stackrel{iid}{\sim} N(0, \sigma_u^2)$ . The error term  $e_{di} \sim N(0, \sigma_e^2)$  is the plot-specific residual error, which is assumed to be independent of the model.

In the BHF nested error regression model, units  $i$  are nested within non-overlapping domains  $d$ . All plots in domain  $d$  are exposed to the same realization of the domain-specific random effect  $u_d$ .

The BHF EBLUP for a particular small area is expressed as

$$\hat{\mu}_d^{EBLUP} = \bar{\mathbf{x}}_{U_d}^T \hat{\boldsymbol{\beta}} + \hat{\gamma}_d (\bar{y}_d - \bar{\mathbf{x}}_{S_d}^T \hat{\boldsymbol{\beta}}) \quad [7]$$

with shrinkage factor

$$\hat{\gamma}_d = \frac{\sigma_u^2}{\sigma_u^2 + \frac{\sigma_e^2}{n_d}} \quad [8]$$

and auxiliary sample mean

$$\bar{\mathbf{x}}_d = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{di} \quad [9]$$

where,  $\bar{\mathbf{x}}_d$  = sample mean (auxiliary)

The BHF fixed effects or synthetic model ( $Y_{syn}$ ) is a linear model denoted as  $\mathbf{x}^T \hat{\boldsymbol{\beta}}$  in [7]

where the base form of the linear model was

$$\hat{y}_{syn} = \mathbf{x}^T \hat{\boldsymbol{\beta}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \quad [10]$$

with  $x_1$  = stand mean canopy height,  $x_2$  = log (stand zero CHM (proportion)),  $x_3$  = thinned (1) or un-thinned (0) indicator, and the model error is e. In fitting the model, observed x consists of a 4-vector with element 1 = 1, and elements 2-4 =  $\{x_1, x_2, x_3\}$  computed from a circular subset of the auxiliary data concomitant with the field plot boundary with which y was observed.

### Thinning effect in the model

The incorporation of thinning status as a predictor variable in our regression model proved to be effective. For volume, a strong negative effect with a coefficient of -213.2 was seen indicating that thinning reduces the volume significantly. It means that when the plot is thinned, observed volume is lowered by around 213.2 ft.<sup>3</sup>/acre. This was consistent for all three attributes

of interest. For above ground biomass, whenever the plot is thinned, the total above ground biomass is lowered by about 4.68 tons/acre. Whereas, for basal area, whenever the plot is thinned, the total basal area, on average, is lowered by about 25.8 ft.<sup>2</sup>/acre. In all three cases, unthinned plots exhibited higher values compared to thinned plots, as shown in Figure 3.

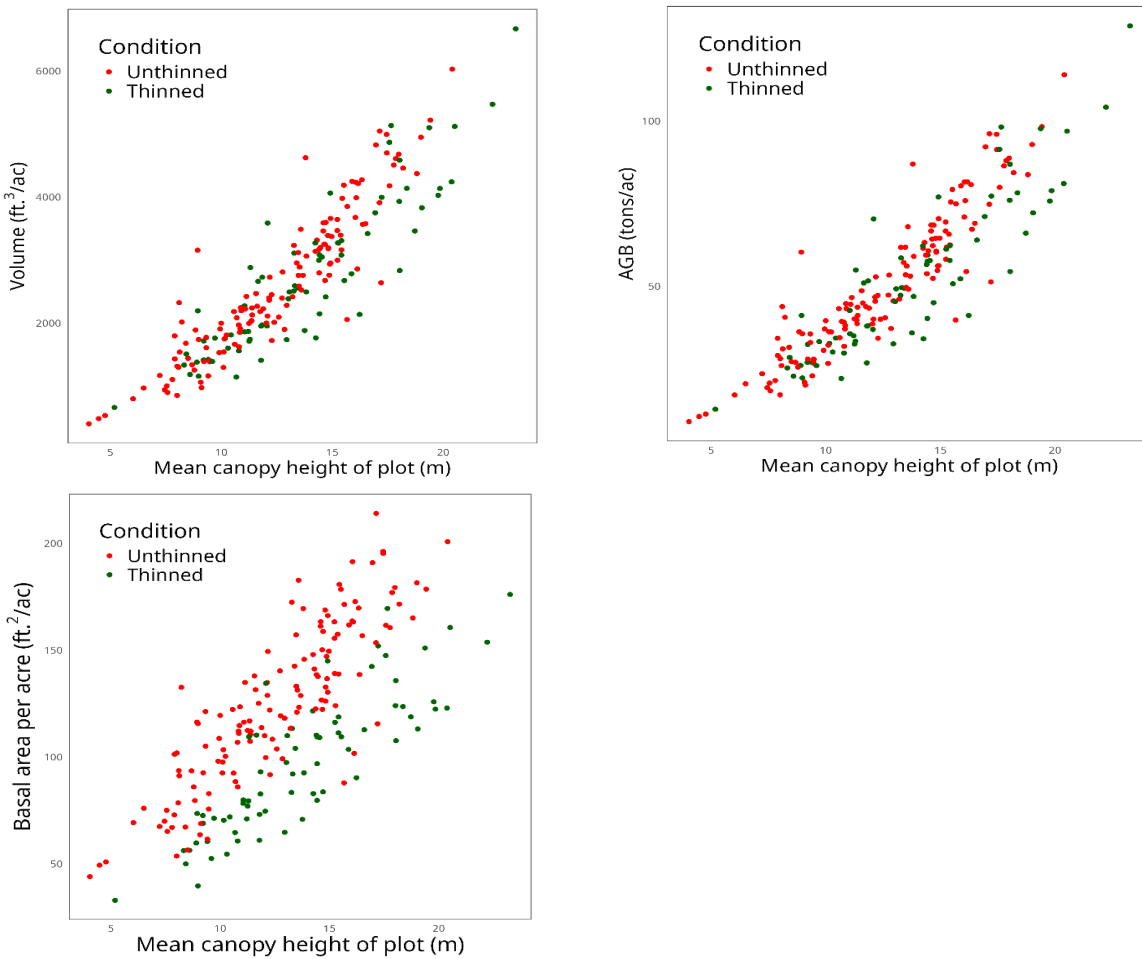


Figure 3. Comparison of stand-level attributes to the mean canopy height of the plot for volume, above-ground biomass, and basal area.

## Open ground relationship in the model

A negative relationship was noted between the amount of open space (zero CHM values) on plots and their observed volume, AGB, and basal area. It indicated that whenever there were open areas in the canopy, the volume, AGB and basal area were lower. The relationship, however, was not exactly linear and was improved using a natural log (ln) transformation (Figure 4.)

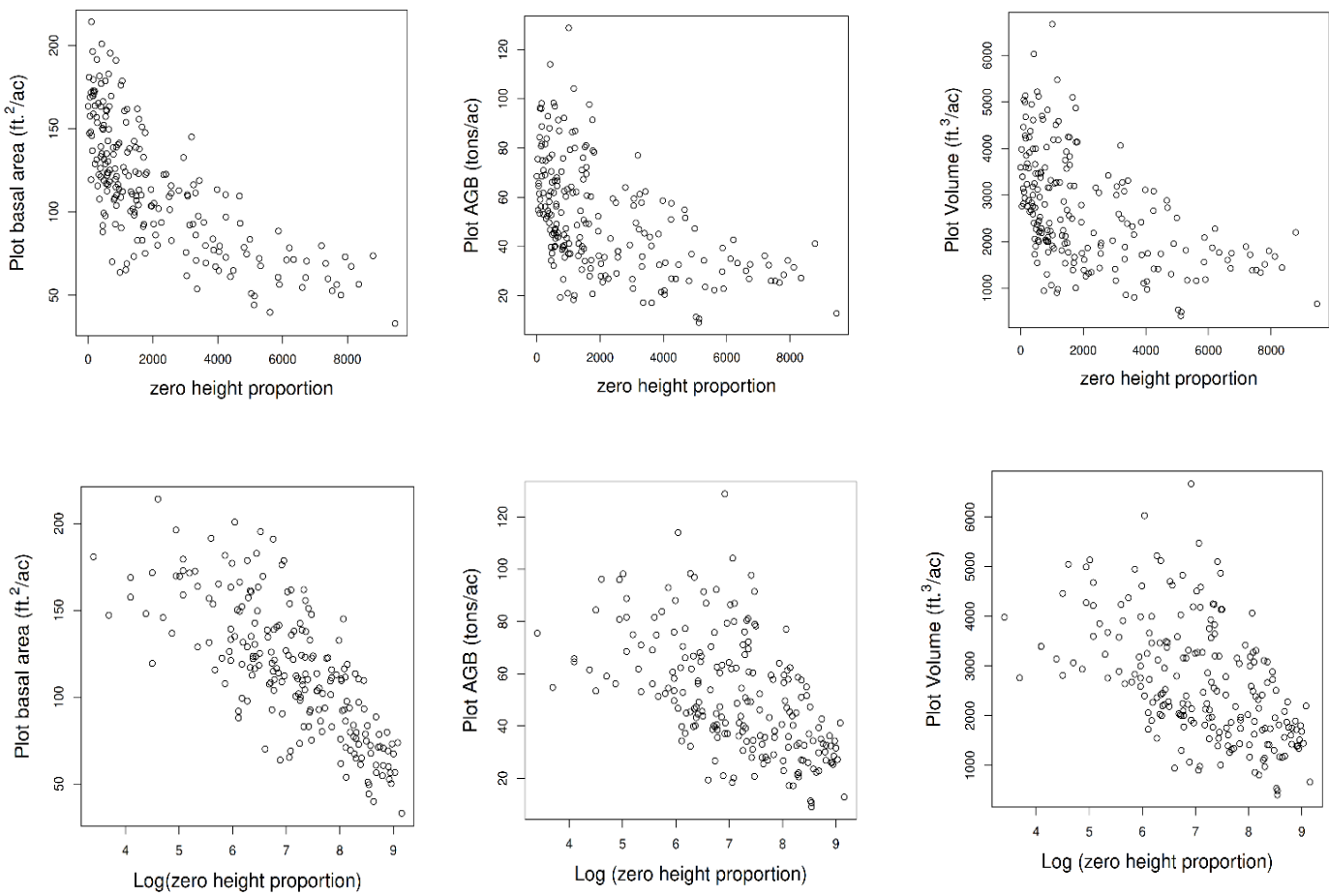


Figure 4. Comparison of stand-level attributes to the proportion of zeroes in the canopy for estimates of basal area, above-ground biomass, and volume (top: without transformation; bottom: with natural log transformation).

### **Precision gains:**

To evaluate gains in the precision of BHF estimators, we used the relative standard errors (RMSE%) for our direct estimates (DIR), and EBLUPs for each stand were calculated as

$$RMSE \%_{EBLUP} = 100 * \sqrt{MSE(\hat{\mu}_d^{EBLUP}) / \bar{y}_d} \quad [11]$$

$$SE \%_{DIR} = 100 * SE(\bar{y}_d) / \bar{y}_d \quad [12]$$

and the standard error ratio (SER) for each stand was calculated as:

$$Standard\ Error\ Ratio\ (SER) = \frac{RMSE\%_{EBLUP}}{SE\%_{DIR}} \quad [13]$$

Unit-level EBLUP estimates were obtained using the package R SAE in R (Molina and Marhuenda 2015).

### **Population Data Simulation**

Simulation was used to generate artificial instances of populations from which sample plots for estimating stand-level EBLUPs were drawn to test assumptions and statistical properties of the BHF estimators. Stands were tessellated into nonoverlapping 20.1168 m x 20.1168 m squares cells matching the field plot size (0.1 acre). In each cell, assumed “known” volume, AGB, and basal area, values were generated to represent forest conditions, with simulated plots used both for generating sample data for use in BHF estimation over differing sampling intensities  $n = \{2, 4, 8\}$  and sampling both by random sampling and a range of nonrandom samples drawn from each basal area, volume and AGB quintile of tessellated plots on a stand.

Calculated EBLUPs and EBLUP MSEs from all stands were compared to simulated population averages and variances on the stands (Warnholz and Schmid 2016).

A theoretical superpopulation model [5] governing all simulations was based on the BHF model fitted to the full set of plot observations and lidar CHMs from plots and stands in the Coastal Georgia study. Components simulating ground plot observations based on the superpopulation model will be generated repeatedly for  $B$  specific population instances  $b = \{1, 2, \dots, B\}$

$$y_{di}^{(b)} = \mathbf{x}_{di}^T \hat{\boldsymbol{\beta}}^{(b)} + u_d^{(b)} + e_{di}^{(b)} \quad [14]$$

with terms defined as follows:  $u_d^{(b)} \sim \mathcal{N}(0, \sigma_u^2)$  is a simulated population random effect for domain  $d$  in the  $b^{\text{th}}$  population based on the superpopulation random-effect variance parameter  $\sigma_u^2$ ;  $e_{di}^{(b)} \sim \mathcal{N}(0, \sigma_e^2)$  is a sample-based residual for plot  $i$  in domain  $d$  for population  $b$ , with  $\sigma_e^2$  the superpopulation residual variance unaccounted for by the mixed effects;  $\mathbf{x}_{di}^T$  is the transposed vector of CHM auxiliary variables for the  $i^{\text{th}}$  tessellated plot in domain  $d$ , which – as a fixed quantity in the BHF framework – is the only term in [14] that did not vary among the simulated populations; and  $\hat{\boldsymbol{\beta}}^{(b)}$  is the vector of fixed effects estimated from simulated sample plot observations and their corresponding auxiliary data.

Domain-specific means were known for each population instance  $b$  based on averages of the tessellated ground conditions

$$\mu_d^{(b)} = \frac{1}{N_d} \sum_{i=1}^{N_d} y_{di}^{(b)} \quad [15]$$

where  $N_d$  is the number of simulated ground plots in the tessellation for stand  $d$ . To test effects of the aforementioned varying sampling intensities from  $n = 2$  through  $n = 24$  in each stand using both random and nonrandom sample selection designs, data will be drawn from each of the 6 sampling intensities for simple random sampling (without replacement) and non-randomly from 5 distinct quintile-based strata, for a total of 36 combinations of sampling intensity by sample selection type in the  $b^{\text{th}}$  population instance. Sampled data will be used to fit the BHF mixed-effects model, producing EBLUP  $(b) = \left(\hat{\mu}_d^{(b)}\right)$  and  $\widehat{MSE}$  (EBLUP  $(b)$ ) estimates for each domain and sample-design combination, then repeated for the  $B$  population instances. Each population instance will also produce estimates of random effect variance  $\hat{\sigma}_u^{2(b)}$  and fixed effects coefficients  $\hat{\beta}^{(b)}$ . Sample-direct means and standard errors will also be calculated for each population instance and sample design to examine relationships between sample design specifications, precision gains in the simulated populations, and biases in estimators.

Simulation-based estimates were summarized to evaluate the effects of sample design specifications on biases in all outputs of the BHF estimator. Quantities of interest included the bias of EBLUPs computed in actual, absolute, and relative units as in [16]-[18] for each sample design specification, which we denote as the mean bias (BIAS), mean absolute bias (MAB), and mean percent bias (MPB)

$$BIAS(\mu_d) = \frac{1}{B} \sum_{b=1}^B \left( \hat{\mu}_d^{(b)} - \mu_d^{(b)} \right) \quad [16]$$

$$MAB(\mu_d) = \frac{1}{B} \sum_{b=1}^B |\hat{\mu}_d^{(b)} - \mu_d^{(b)}| \quad [17]$$

$$MPB(\mu_d) = \frac{100}{B} \sum_{b=1}^B \frac{\hat{\mu}_d^{(b)} - \mu_d^{(b)}}{\mu_d^{(b)}} \quad [18]$$

Where bias calculations [16]-[18] for domain means compare EBLUPs to simulated known values  $\mu_d^{(b)}$  from each population instance, estimated fixed-effects ( $\hat{\beta}^{(b)}$ ) will be compared to the superpopulation model parameters  $\beta$  in similar formulas to compute  $BIAS(\beta)$ ,  $MAB(\beta)$ , and  $MPB(\beta)$ .

Instance-specific estimates of EBLUP variances  $\widehat{mse}(\hat{\mu}_d^{(b)})$  will be compared to an empirical superpopulation variance for each domain to quantify the simulation-level bias of variance estimates

$$BIAS(\widehat{mse}(\hat{\mu}_d)) = \frac{1}{B} \sum_{b=1}^B \widehat{mse}(\hat{\mu}_d^{(b)}) - MSE(\hat{\mu}_d) \quad [19]$$

where the BHF-estimated variances are denoted with lower-case (mse) and the empirical variance is capitalized and defined as

$$MSE(\hat{\mu}_d) = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_d^{(b)} - \mu_d^{(b)})^2 \quad [20]$$

noting that  $MSE(\hat{\mu}_d)$  will serve in place of the unknown superpopulation domain specific BHF variances (Molina et al. 2015). Calculations for  $MAB(\widehat{mse}(\hat{\mu}_d))$  and  $MPB(\widehat{mse}(\hat{\mu}_d))$  will follow

formulas analogous to [17] and [18], substituting estimated mse and empirical MSE terms from [19] in place of estimated and known domain means.

Coverage rates (CR) for confidence intervals (CI) will also be calculated from the simulated population instances, assuming normality of EBLUPs and forming frequentist confidence intervals for EBLUPs from each population instance and sampling scheme.

$$CI(\hat{\mu}_d^{(b)}) = \hat{\mu}_d^{(b)} \pm Z_{\alpha/2} \sqrt{\widehat{mse}(\hat{\mu}_d^{(b)})} \quad [21]$$

An indicator  $I^{(b)} = [0, 1]$  will be set to 1 when  $CI(\hat{\mu}_d^{(b)})$  includes  $\mu_d^{(b)}$ , and 0 otherwise. The proportions (i.e. means) of  $I^{(b)}$  over  $b = \{1, 2, \dots, B\}$  population instances will be compared to the nominal CI coverage  $1 - \alpha$  to identify baseline CR and identify any patterns that arise in how the sampling scheme effects CR (Molina et al. 2015).

## Results

### EBLUP Comparisons to Sample Means

Comparison of stand-level direct estimates to BHF EBLUPs generally showed agreement with each other with some notable exceptions (Fig 5). Out of the 104 stands analyzed, there were 18 stands where the EBLUP estimates for volume and AGB differed by more than 25% from the direct estimates. For basal area, the 25% discrepancy occurred for 9 stands (+ & ▲ symbols, Fig 5). The purpose of examining stands with  $\geq 25\%$  differences between direct and EBLUP estimates was to look for explanations for such differences.

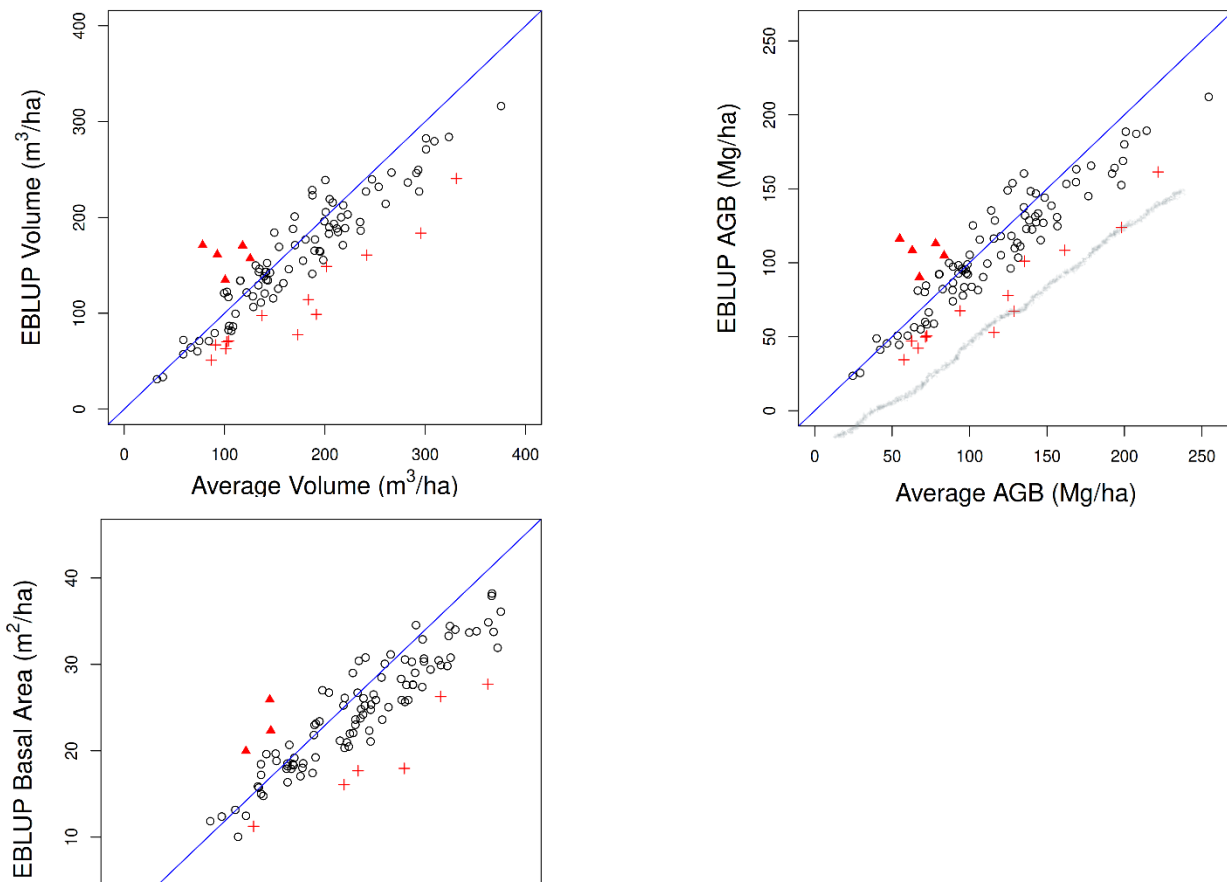


Figure 5. BHF composite estimates (EBLUPs) of stand-level attribute totals compared to direct estimates for volume, above-ground biomass, and basal area (*Note: points with symbol + indicate values with differences  $\geq 25\%$  that fall under the 1:1 line, whereas ▲ indicate values with differences  $\geq 25\%$  lying above the 1:1 line for volume, AGB, and basal area, respectively*).

Closer examination of EBLUPs for stands with large sample means for volume, AGB, and basal area stands (+ symbols, Fig 5) revealed their EBLUPs typically reflected certain conditions for how plot-level auxiliary metrics related to the same metrics compiled for the corresponding stands. In a somewhat extreme example from stand 076 (Fig. 6), approximately 10% of the stand area consisted of gaps or open areas without substantial tree cover; notably, neither sample plot in stand 076 was situated in the open areas of the stand as determined by CHM metrics. This indicated that sample means were likely inflated in relation to overall stand volume, AGB, or basal area. Inspection of CHMs mapped across the 13 stands listed above for volume (+ symbols, Fig 5) showed a possible lack of representativeness of sample plot placement in relation to gaps or open areas – including thinning corridors, small openings, and trails – in these stands (Appendix I).



Figure 6. Stand 076 circular plot boundaries in relation to the stand boundary and auxiliary CHM values (m).

In cases where stand EBLUPs exceeded sample means by > 25%, i.e. stands marked with the ▲ symbol that lay above the line  $Y=X$  (Fig. 5), inspection showed sample plots tended to be situated where small gaps or corridors led to potentially non-representative deflated volume, etc. sample means (Fig 7). For example, stand 028 was populated by tree canopies mainly > 20 m in height with openings (CHM = 0) and corridors covering 38% of the stand, compared to the open areas in sample plots, which averaged 51% (white areas, Fig 7). This indicated that the stand as a whole carried more volume, etc. than the areas within the plots. Presumably, sample means in cases like stand 028 would underestimate whole-stand growing stock or yield, i.e. volume, AGB, and basal area. A closer inspection of CHMs of the 5 stands marked with ▲ symbols in fig 5 showed a similar reason behind this apparent bias (Appendix II).

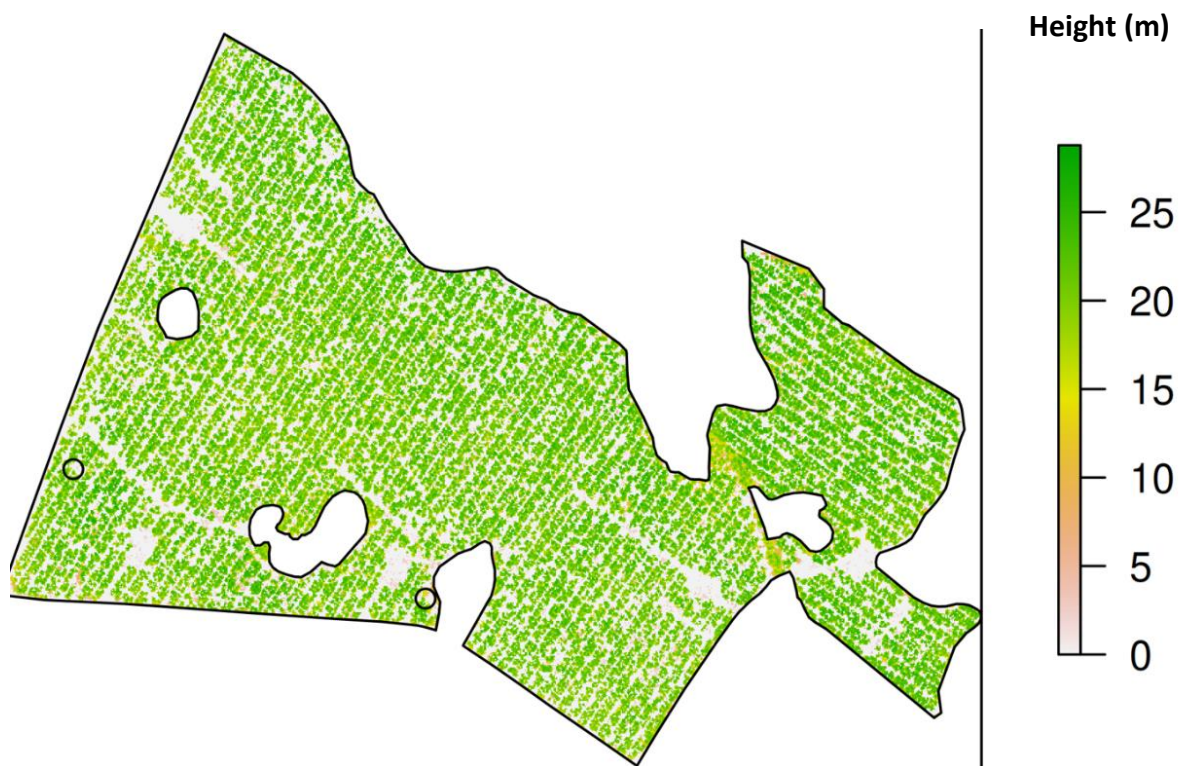


Figure 7. Stand 2020028 plot locations.

To further support this, the predictor values obtained for the number of zeroes per acre for the two plots for stand 028 were found to be consistently smaller than the ones obtained for the whole stand indicating that the EBLUP values should be larger than the plot's average basal area (Table 2). For the stands where the EBLUP seems to be overestimating the average basal area, the predictor values obtained for the number of zeroes per acre for the stands were consistently smaller than the ones obtained for the two plots indicating that the EBLUP values should be larger than the plot's average (Appendix III). In contrast, synthetic model predictions accounted both for differences in average canopy heights and the areas of openings based on model fixed effects coefficients (Appendix IV).

Table 2. Plot and stand level predictor variables for number of zeroes obtained for stand 028 along with its two plots.

	<b>IDs</b>	<b>Thinning status</b>	<b>Ln (number of zeroes)</b>
<b>Stand ID</b>	2020028	1	8.73
<b>Plot ID</b>	2020028A	1	9.08
<b>Plot ID</b>	2020028B	1	8.96

## RMSE Comparison

Relative RMSEs from SAE estimators, compared to the standard errors of direct estimates for stand-level attributes, highlighted the magnitude of differences in estimator errors. (Figure 8). A notable pattern is the lack of correspondence between EBLUP and direct estimation errors. In addition to that, EBLUP RMSEs were more stable than the direct standard errors derived from the small sample plots alone (Figure 8). In contrast, the direct standard errors were highly variable, often fluctuating between very low and very high values, indicating a lack of stability for all three attributes of interest. This instability is likely due to the small sample

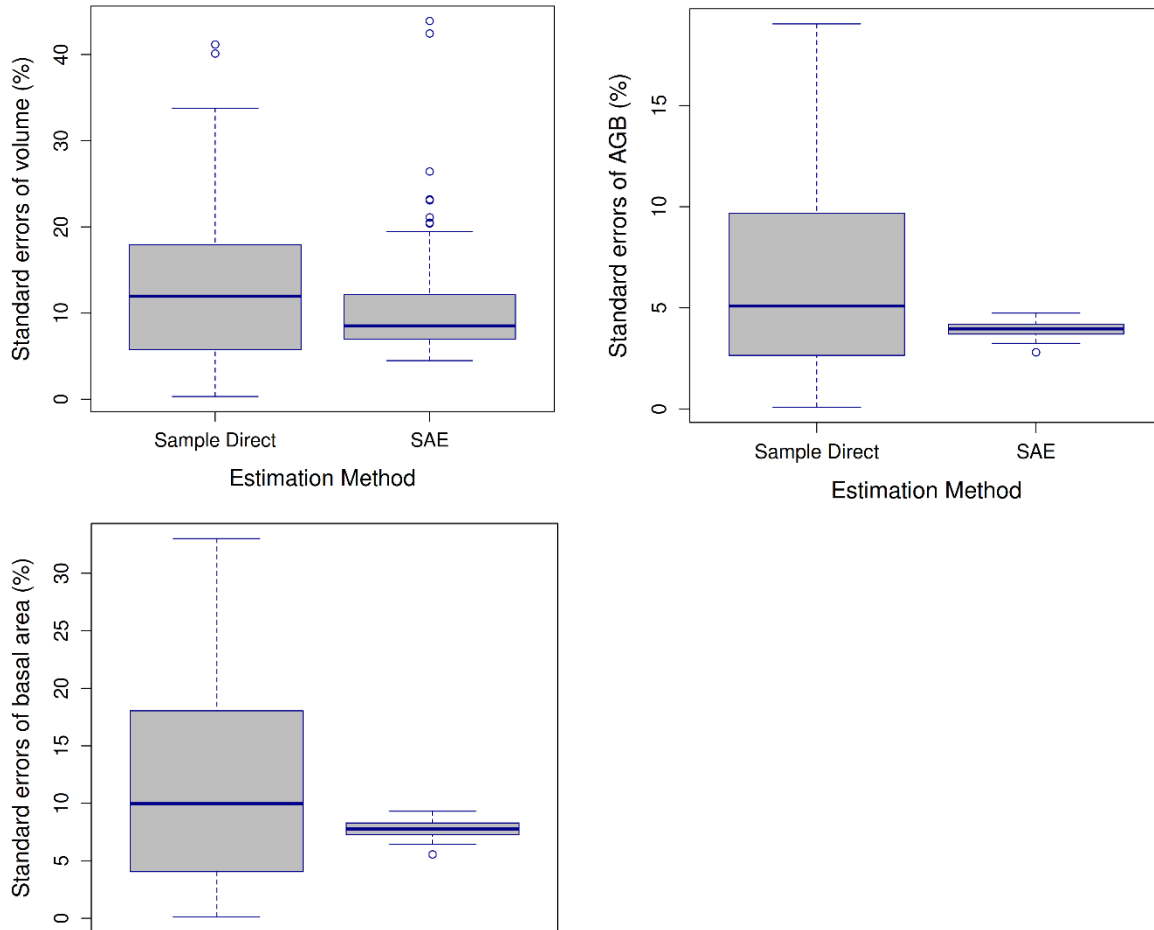


Figure 8. Standard errors obtained from direct estimates compared to that of the EBLUPs from unit-level SAE for all three attributes of interest.

size, which not only affects the reliability of the estimates themselves but also contributes to the instability of their standard errors. There were instances where some stands had very small direct standard error values from their two sample plots, and there were other instances where the direct SER was very high (Fig 8). Across all three attributes, the RMSEs of the EBLUPs demonstrated much less variability and greater consistency throughout the analysis (Fig 8).

For stands with the smallest direct standard errors, inspection showed that the two plots within those stands had nearly the same basal area per acre, causing their standard deviations to be close to zero hence giving such small values for direct standard errors. For basal area, the mean SER was calculated as 2.50 indicating that, on average, the BHF EBLUP RMSEs are two and a half times larger than the sample direct standard errors. For volume, it was 2.18, and for above-ground biomass, it was 2.66 (Table 3).

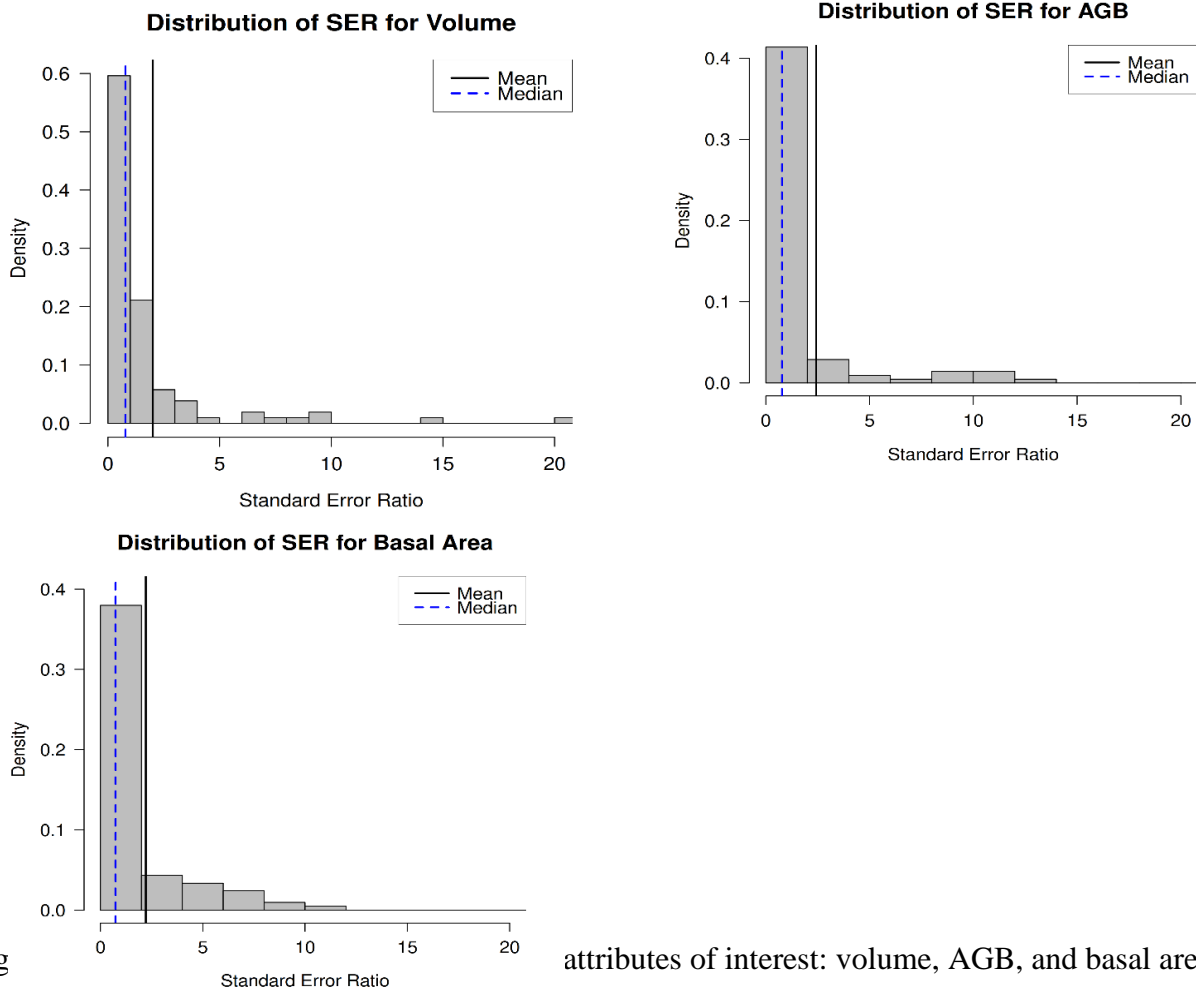
Table 3. Summary statistics of SER for basal area, volume, and above ground biomass estimates.

<b>SER</b>	<b>Min.</b>	<b>1<sup>st</sup> quartile</b>	<b>Median</b>	<b>Mean</b>	<b>3<sup>rd</sup> quartile</b>	<b>Max.</b>
<b>Basal area</b>	0.2207	0.4196	0.7297	2.2146	1.7701	56.6981
<b>Volume</b>	0.1965	0.4113	0.7843	2.0078	1.5364	31.1809
<b>AGB</b>	0.2019	0.4110	0.7792	2.4251	1.5263	40.7456

Results from mean SER suggested that the BHF method did not effectively improve the precision of estimates but closer examination of SERs distribution showed highly skewed distributions (Fig 8). Table 2 shows that the range of SER for basal area, volume and above ground biomass is large, thus making the distribution skewed. This discrepancy can be attributed to a number of stands exhibiting very small direct standard errors while others exhibited large direct standard errors (Figure 9). Because SER is the ratio of EBLUP RMSE to the direct standard error, the very low value of direct standard error is causing the SER to increase because of which the distribution of SER is pulled to the right. In this case a better statistic to look at would be the median. Skewness was evident by the fact that mean (SER)>> median (SER) and

maxima (SER) >> SER means (Table 3). In this case, the median SER was studied as a more suitable measure of the overall SER performance due to its skewedness.

For volume estimates, the median SER was observed to be 0.7843 which tells us that their central tendency as measured by the median showed SERs tended to be <1, indicating precision gains over direct estimate. With this, the apparent sample size for improved volume estimate is 3.25, which means, on average we would have required 1.63 times more plots to achieve the same precision without using SAE. For AGB, it was 0.7792 which also indicate a gain in precision with the apparent sample size being 3.29 and finally, for basal area, the median standard error ratio was found to be 0.7297 with the apparent sample size being 3.75 hence again indicating that we obtained a gain in precision of all three of our attributes of interest.



Fig

attributes of interest: volume, AGB, and basal area.

## Discussion

The findings of this study affirm the potential of small area estimation (SAE) techniques to reduce uncertainties in estimates of volume, above-ground biomass, and basal area. Employing a unit-level SAE approach significantly lowered the root mean squared errors (RMSEs) of estimates for small areas (stands) compared to the less precise direct estimates. Additionally, direct standard errors derived from small-sized samples were found to be highly unstable, whereas RMSEs obtained from Empirically Best Linear Unbiased Predictions (EBLUPs) demonstrated improved stability (Figure 8). This enhanced stability increases confidence in the estimates derived from EBLUPs, made possible through the incorporation of auxiliary data, such as lidar. Consistent with findings by Goerndt et al. (2011), Magnussen et al. (2017), and Mauro et al. (2017), lidar proved to be an effective auxiliary data source for SAE models. Unit-level SAE models, which can utilize finer-scale data, provide greater precision gains compared to area-level models (Molina & Marhuenda, 2015).

Additionally, incorporating information about thinning status as well as addition of variables such as the proportion of zeros in the plot, obtained from lidar led to enhanced precision in the unit-level analysis. It indicated a relationship between the presence of bare ground in the stand and the predicted value for volume, biomass, and basal area estimate. Although auxiliary information like thinning status may not always be available for every plot location (Green et al. 2020a), this study demonstrated that including a thinning indicator variable in the regression model was beneficial. It allowed us to clearly observe the distinct effects of thinning on stand volume, above-ground biomass, and basal area.

In comparing the estimates from EBLUPs and direct field measurements, we observed a general alignment between stand attributes derived from the two methods. By "alignment," we refer to the tendency for estimates to follow a 1:1 relationship, with deviations not necessarily indicating bias. For the most part, the estimates corresponded closely. However, there were notable exceptions where discrepancies exceeded 25% (Fig. 5). At first glance, these discrepancies might suggest the presence of bias in the EBLUP estimates. Yet, upon further analysis, it became evident that these differences were primarily due to the small number of plots

failing to capture the variability of the stand, and hence to estimate the mean and SE appropriately.

In the case of stands where EBLUPs consistently underestimated the mean basal area compared to the direct estimate, it appears that the sample plots were located in areas with taller trees and higher stem density (stems/acre), with minimal open ground leading to few zero pixels. This likely contributed to an overestimation of basal area when relying on direct observation. In contrast, the areas outside these plots contained either shorter trees, more open ground, or both, indicating that the stand, overall, would have a lower average basal area which was better reflected in the EBLUP estimates. This apparent difference suggests that the EBLUP estimates, rather than being biased, effectively accounted for stand variability, resulting in a lower yet more accurate average basal area estimate.

Thus, while the initial apparent underestimation by the EBLUPs may appear to suggest oversaturation or bias, this observation underscores the critical role of sampling design. The placement of just two plots likely led to inflated direct estimates, thereby highlighting EBLUP's strength in producing unbiased stand-level estimates. An important question, however, was whether the observed differences between the EBLUPs and the sample means reflected actual bias or if the EBLUPs remained unbiased. Through simulations, we were able to confirm that the EBLUPs maintained their unbiased properties, effectively capturing stand variability without introducing systematic error.

An important assumption in the application of small area estimation (SAE) techniques is the existence of a linear relationship between the auxiliary data and the variable of interest (Rao & Molina, 2015). SAE methods can produce unbiased estimates only when the auxiliary data is of sufficient quality. For this study, a set of auxiliary variables was selected based on their theoretical relationship with the variable of interest and total volume. The inclusion of thinning status, alongside lidar mean height and the proportion of zeroes in the canopy effect, helped to "adjust" the estimates for volume, biomass, and basal area, as discussed earlier.

The results of this study demonstrate that SAE techniques can reliably improve the precision of volume, basal area, and above-ground biomass estimates. However, these models have

limitations. Specifically, the SAE methods employed in this study are not applicable in situations where a direct estimate is unavailable (Goerndt et al., 2011). In such instances, a model-based, synthetic estimate would be required.

## **Conclusion**

Unit-level SAE models that incorporated lidar-derived canopy heights, the proportion of zero-value pixels, and thinning status as auxiliary information achieved precision improvements for estimates of stand-level forest basal area, above-ground biomass, and volume in commercially planted loblolly pine plantations. These gains were realized when compared to estimates derived from the limited sample data alone. The results have also shown that the precision has increased substantially. For basal area estimates, our efficiency increased substantially by 87.5%, for volume, it increased by 62.5%, and finally, for above-ground biomass, it increased by 64.5%. In other words, for the basal area estimates, we would require an additional 1.87 times more samples to achieve the same precision without using SAE, for volume and above-ground biomass, we would have required an additional 1.62 and 1.64 times more plots, respectively to achieve the same precision without using SAE.

This study has successfully demonstrated the potential of incorporating small area estimation (SAE) techniques into operational forest inventory in commercial loblolly pine plantations. By leveraging auxiliary information from lidar, such as the proportion of zeroes in the canopy, mean canopy height, and thinning status, the uncertainty associated with estimates of total planted volume, above-ground biomass, and basal area was reduced. In response to the research questions outlined in this study: (1) unit-level SAE methods improved the precision of volume, above-ground biomass, and basal area estimates, particularly when direct estimates exhibited high variability; (2) unit-level models demonstrated an increase in the precision of the attributes of interest; and (3) the incorporation of additional auxiliary information, such as lidar and thinning status, further enhanced the precision of estimated stand basal area, volume, and biomass. The auxiliary variables—mean canopy height, the proportion of zeroes in the stand, and thinning status—were found to be particularly effective for estimating forest above-ground biomass, volume, and basal area.

The findings of this study are particularly relevant to forest inventory managers conducting regular inventories in southern pine plantations. As the importance of enhancing and monitoring the productivity of loblolly pine grows, driven by both commercial and ecological interests such as carbon sequestration (Zhao et al., 2016), accurate and precise estimates of stand volume and biomass are crucial. By applying the methods outlined in this study, the precision of inventory estimates can be improved, resulting in greater confidence when making management and planning decisions. The unit-level SAE methods evaluated in this study are broadly applicable to situations where precise GPS locations are available.

## References

- Baker, J. B., and O. G. Langdon. 1990. "Pinus taeda L. Loblolly Pine." In *Silvics of North America*, edited by R. M. Burns and B. H. Honkala, Vol. 1, Conifers: Agricultural Handbook 654, 497–512. U.S. Department of Agriculture, Forest Service.
- Battese, G. E., R. M. Harter, and W. A. Fuller. 1988. "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data." *Journal of the American Statistical Association* 83: 28–36. <https://doi.org/10.1080/01621459.1988.10478561>.
- Bechtold, W. A., and P. L. Patterson. 2005. *The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures*. U.S. Department of Agriculture, Forest Service.
- Blinn, C. E., M. N. House, R. H. Wynne, et al. 2019. "Landsat 8 Based Leaf Area Index Estimation in Loblolly Pine Plantations." *Forests* 10 (3): Article 222. <https://doi.org/10.3390/f10030222>.
- Breidenbach, J., and R. Astrup. 2012. "Small Area Estimation of Forest Attributes in the Norwegian National Forest Inventory." *European Journal of Forest Research* 131 (4): 1255–1267. <https://doi.org/10.1007/s10342-012-0596-7>.
- Burkhardt, H. E., T. E. Avery, and B. P. Bullock. 2019. *Forest Measurements*. 6th ed. Long Grove, IL: Waveland Press.

- Burkhart, H. E., and M. Tomé. 2012. *Modeling Forest Trees and Stands*. Dordrecht, Netherlands: Springer.
- Cao, Q., Dettmann, G.T., Radtke, P.J., Coulston, J.W., Derwin, J., Thomas, V.A., Burkhart, H.E., & Wynne, R.H. (2022). Increased Precision in County-Level Volume Estimates in the United States National Forest Inventory With Area-Level Small Area Estimation. *Frontiers in Forests and Global Change*, 5:769917. <https://doi.org/10.3389/ffgc.2022.769917>
- Elzinga, C. L., D. W. Salzer, J. W. Willoughby, and J. P. Gibbs. 2001. *Monitoring Plant and Animal Populations: A Handbook for Field Biologists*. Malden, MA: Blackwell Science.
- Etikan, I., S. A. Musa, and R. S. Alkassim. 2016. "Comparison of Convenience Sampling and Purposive Sampling." *American Journal of Theoretical and Applied Statistics* 5 (1): 1–4.
- Fay, R. E., and R. A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74: 269–277. <https://doi.org/10.2307/2286322>.
- Goerndt, M. E., V. J. Monleon, and H. Temesgen. 2011. "A Comparison of Small-Area Estimation Techniques to Estimate Selected Stand Attributes Using Lidar-Derived Auxiliary Variables." *Canadian Journal of Forest Research* 41 (6): 1189–1201.
- Goff, F. G., G. A. Dawson, and J. J. Rochow. 1982. "Site Examination for Threatened and Endangered Plant Species." *Environmental Management* 6: 307–316.
- Green, P. C., H. E. Burkhart, J. W. Coulston, and P. J. Radtke. 2020. "A Novel Application of Small Area Estimation in Loblolly Pine Forest Inventory." *Forestry* 93: 444–457. <https://doi.org/10.1093/forestry/cpz073>.
- Green, P. C., Burkhart, H., Coulston, J., Radtke,

- P., and Thomas, V. A. (2020b). Auxiliary information resolution effects on small area estimation in plantation forest inventory. *Forestry* 93, 685–693. doi: 10.1093/forestry/cpaa012
- Kershaw, J. A., Jr., M. J. Ducey, T. W. Beers, and B. Husch. 2016. "Sampling Designs in Forest Inventories." In *Forest Mensuration*, edited by J. A. Kershaw, M. J. Ducey, T. W. Beers, and B. Husch. <https://doi.org/10.1002/9781118902028.ch10>.
- Lehtonen, R., and A. Veijanen. 2009. "Chapter 31 — Design-Based Methods for Domains and Small Areas." In *Handbook of Statistics 29B: Sample Surveys: Inference and Analysis*, edited by D. Pfefferman and C. R. Rao, 219–249. Amsterdam: Elsevier. [https://doi.org/10.1016/s0169-7161\(09\)00231-4](https://doi.org/10.1016/s0169-7161(09)00231-4).
- Magnussen, S., F. Mauro, J. Breidenbach, A. Lanz, and G. Kändler. 2017. "Area-Level Analysis of Forest Inventory Variables." *European Journal of Forest Research* 136: 839–855. <https://doi.org/10.1371/journal.pone.0189401>.
- Maltamo, M., E. Næsset, and J. Vauhkonen. 2014. *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. *Managing Forest Ecosystems* 27: 460.
- Mauro, F., I. Molina, A. García-Abril, R. Valbuena, and E. Ayuga-Téllez. 2016. "Remote Sensing Estimates and Measures of Uncertainty for Forest Variables at Different Aggregation Levels." *Environmetrics* 27: 225–238. <https://doi.org/10.1002/env.2387>.
- Mauro, F., V. J. Monleon, H. Temesgen, and K. R. Ford. 2017. "Analysis of Area-Level and Unit-Level Models for Small Area Estimation in Forest Inventories Assisted with LiDAR

- Auxiliary Information." PLoS ONE 12 (12): e0189401.  
<https://doi.org/10.1371/journal.pone.0189401>.
- Miles, P. D., and W. B. Smith. 2009. "Specific Gravity and Other Properties of Wood and Bark for 156 Tree Species Found in North America." Research Note NRS-38. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station.  
<https://doi.org/10.2737/NRS-RN-38>.
- Molina, I., and Y. Marhuenda. 2015. sae: An R Package for Small Area Estimation.
- Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Models in S and S-PLUS. New York: Springer. <https://doi.org/10.1007/b98882>.
- Pinheiro, J., D. Bates, and R Core Team. 2023. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-164. <https://CRAN.R-project.org/package=nlme>.
- R Core Team. 2024. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rao, J. N. K. 2003. Small Area Estimation. New York: Wiley-Interscience.
- Rao, J. N. K., and I. Molina. 2015. "Small Area Estimation." R Journal 7: 81–98.  
<https://doi.org/10.32614/rj-2015-007>. Reich, R. M., and Aguirre-Bravo, C. (2009). Small-area estimation of forest structure in Jalisco, Mexico. J. For. Res. 20, 285–292. doi: 10.1007/s11676-009-0050-y.
- Richards, F. J. 1959. "A Flexible Growth Function for Empirical Use." Journal of Experimental Botany 10 (29): 290–300. <http://www.jstor.org/stable/23686557>.

- Schumacher, F. X., and S. H. Hall. 1933. "Logarithmic Expression of Timber-Tree Volume." *Journal of Agricultural Research* 47: 719–734.
- Sterba, S. K. 2009. "Alternative Model-Based and Design-Based Frameworks for Inference From Samples to Populations: From Polarization to Integration." *Multivariate Behavioral Research* 44 (6): 711–740. <https://doi.org/10.1080/00273170903333574>.
- Sterba, S. K., M. J. Prinstein, and M. Nock. 2008. "Beyond Pretending Complex Nonrandom Samples Are Simple and Random." Paper presented at the APA Convention, Boston, MA, August.
- Strub, Mike, and Nathaniel Osborne. 2021. "Correcting Tree Count Bias for Objects Segmented from LiDAR Point Clouds." *Mathematical and Computational Forestry and Natural-Resource Sciences* 13: 29–35.
- Wang, H.-J., S. P. Prisley, P. J. Radtke, and J. W. Coulston. 2011. "Errors in Terrain-Based Model Predictions Caused by Altered Forest Inventory Plot Locations in the Southern Appalachian Mountains, USA." *Mathematical and Computational Forestry and Natural-Resource Sciences* 3: 114–123.
- Warnholz, Sebastian, and Timo Schmid. 2016. "Simulation Tools for Small Area Estimation: Introducing the R-Package saeSim." *Austrian Journal of Statistics* 45: 55. <https://doi.org/10.17713/ajs.v45i1.89>.

Westfall, J. A., J. W. Coulston, A. N. Gray, J. D. Shaw, P. J. Radtke, D. M. Walker, A. R. Weiskittel, D. W. MacFarlane, D. L. R. Affleck, D. Zhao, H. Temesgen, K. P. Poudel, J. M. Frank, S. P. Prisley, Y. Wang, A. J. Sánchez Meador, D. Auty, and G. M. Domke. In press. "A National-Scale Tree Volume, Biomass, and Carbon Modeling System for the United States." General Technical Report WO-104. Washington, DC: U.S. Department of Agriculture, Forest Service. <https://doi.org/10.2737/WO-GTR-104>.

White, J. C., M. A. Wulder, A. Varhola, M. Vastaranta, N. C. Coops, B. D. Cook, D. Pitt, and M. Woods. 2013. "A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach." *The Forestry Chronicle* 89 (6): 722–723.

## Appendices

### Appendix I: Stand CHMs (with heights in meters) where EBLUP underestimated stand volume (+ symbols, figure 5)

#### Stand ID's:

2020006

2020007

2020011

2020012

2020037

2020038

2020063

2020076

2020129

2020214

2020215

2020224

2020235

2020006

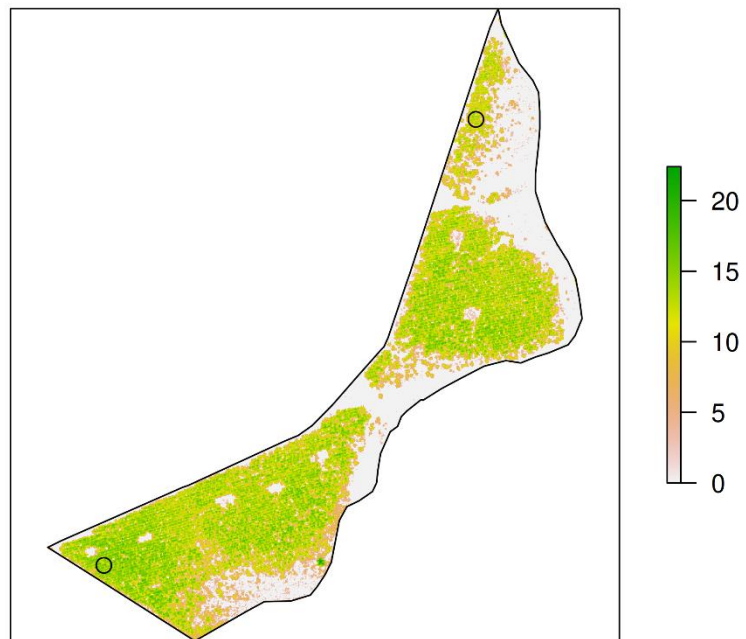


Fig 10. Stand 2020006 map showing the sample plots (two small black circles) located within it.

2020007

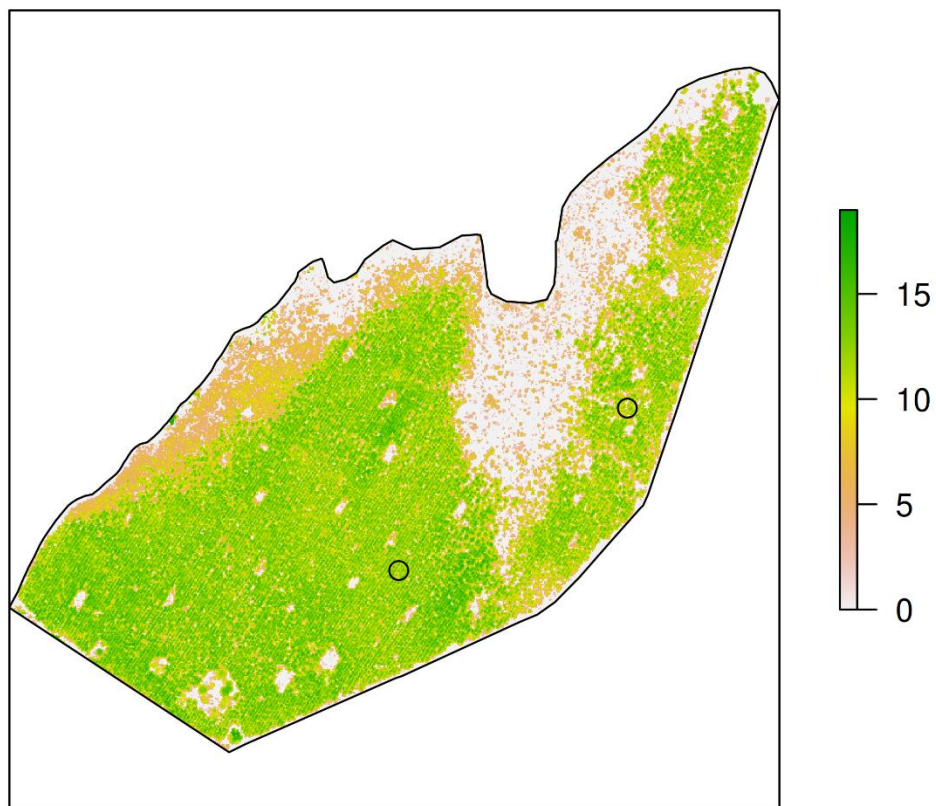


Fig 11. Stand 2020007 map showing the sample plots (two small black circles) located within it.

2020011

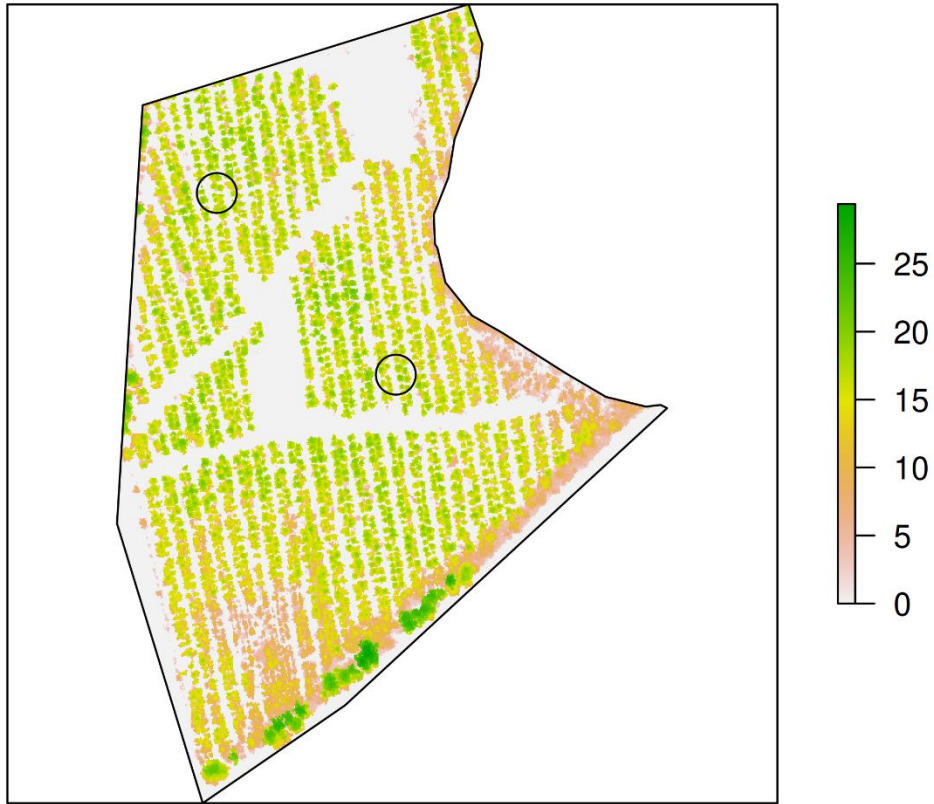


Fig 12. Stand 2020011 map showing the sample plots (two small black circles) located within it.

2020012

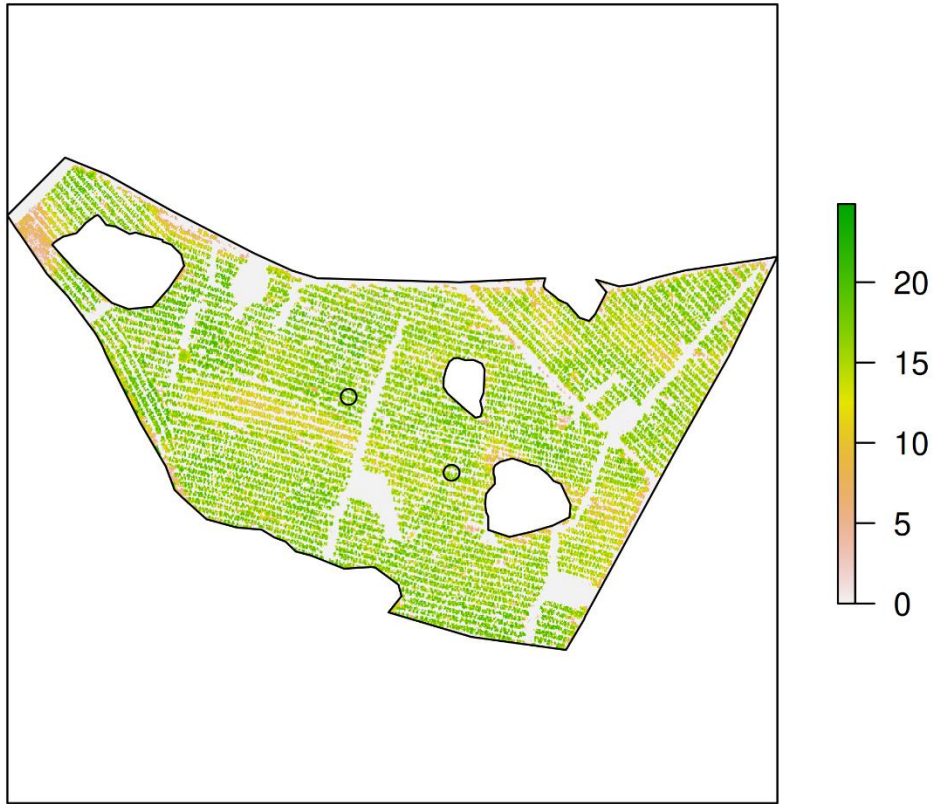


Fig 13. Stand 2020012 map showing the sample plots (two small black circles) located within it.

2020037

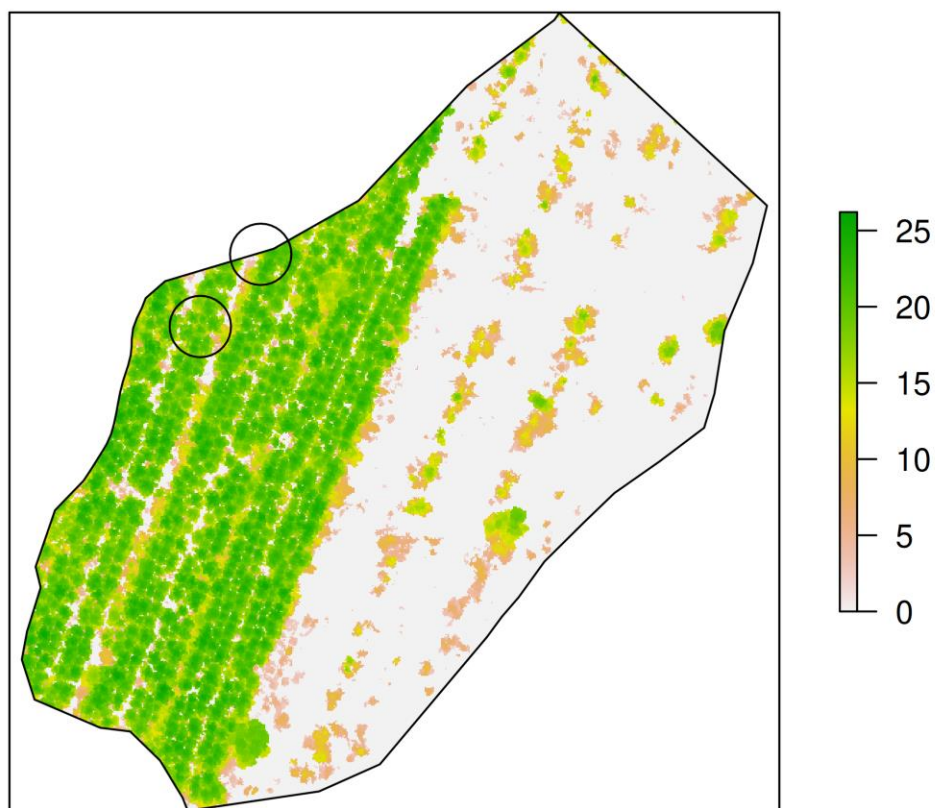


Fig 14. Stand 2020037 map showing the sample plots (two small black circles) located within it.

2020038

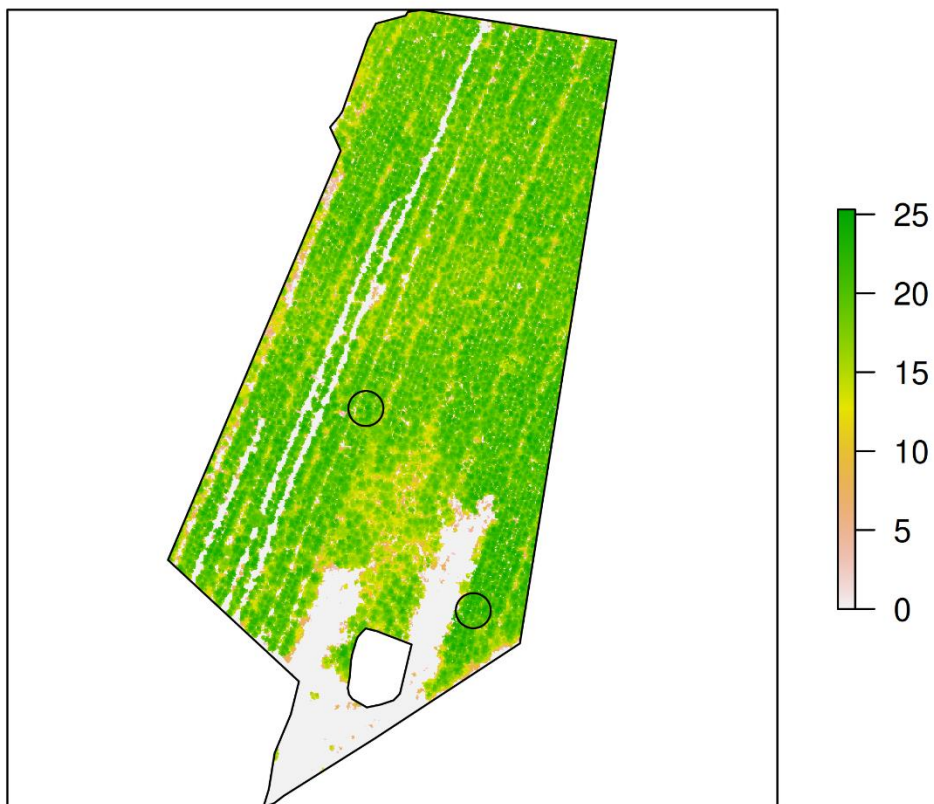


Fig 15. Stand 2020038 map showing the sample plots (two small black circles) located within it.

2020063

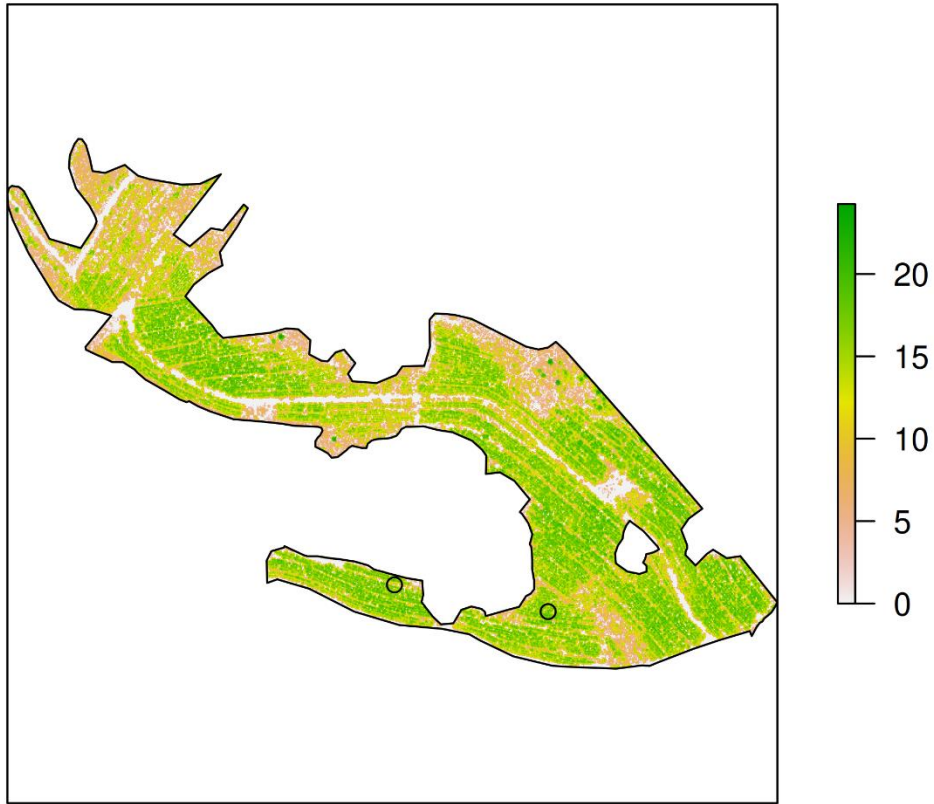


Fig 16. Stand 2020063 map showing the sample plots (two small black circles) located within it.

2020076

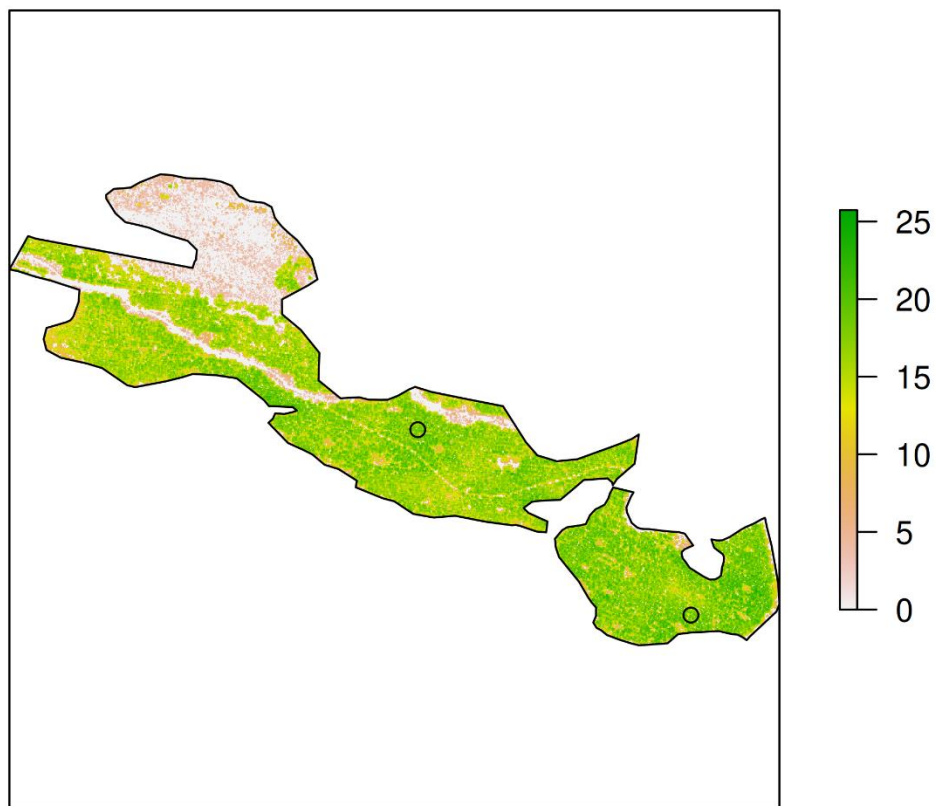


Fig 17. Stand 2020076 map showing the sample plots (two small black circles) located within it.

2020129

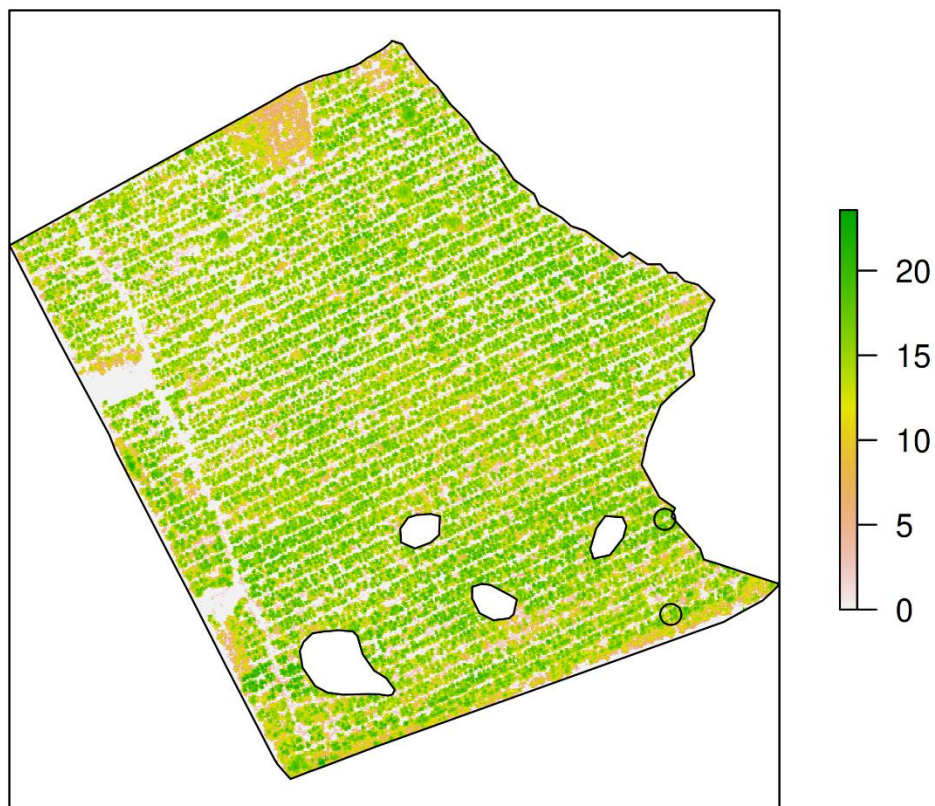


Fig 18. Stand 2020129 map showing the sample plots (two small black circles) located within it.

2020214

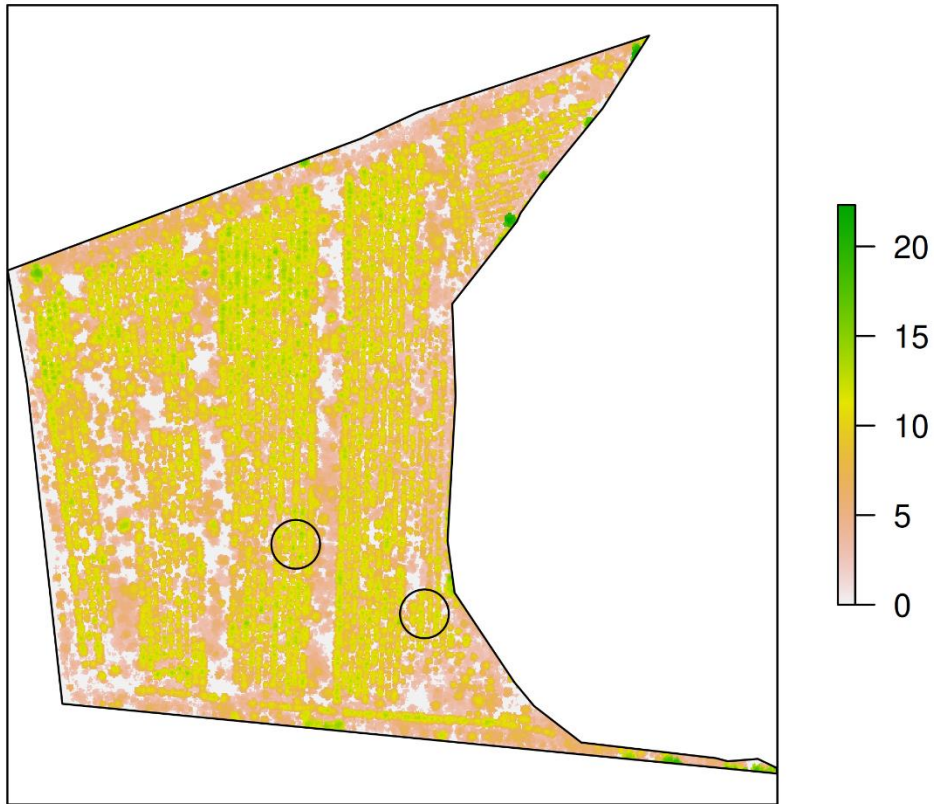


Fig 19. Stand 2020214 map showing the sample plots (two small black circles) located within it.

2020215

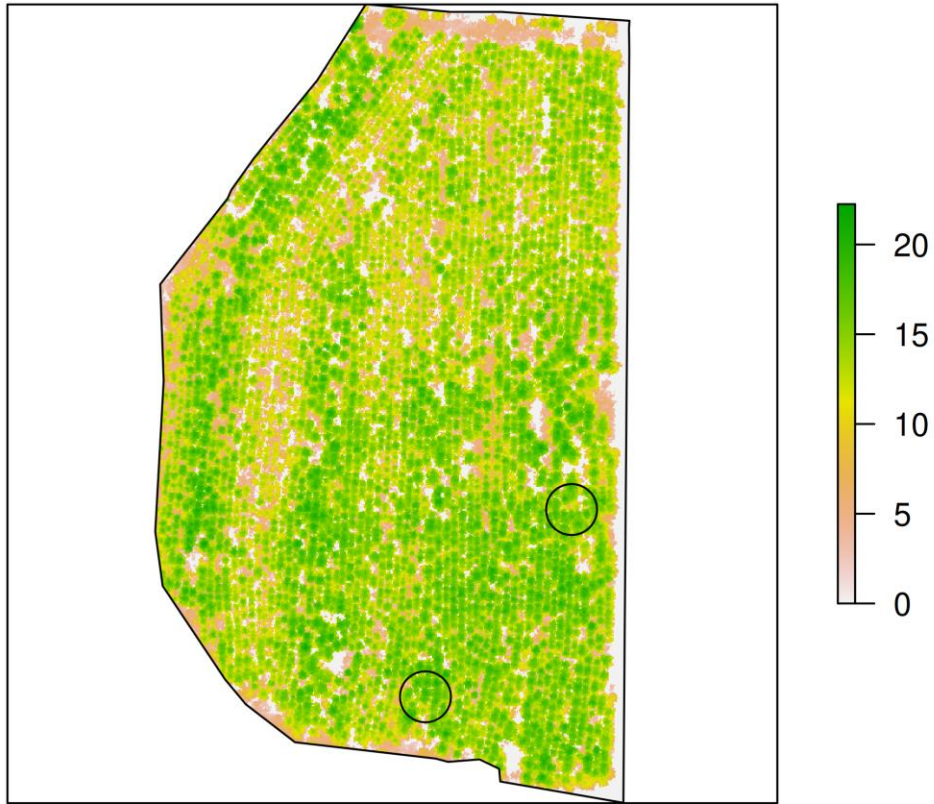


Fig 20. Stand 2020215 map showing the sample plots (two small black circles) located within it.

2020224

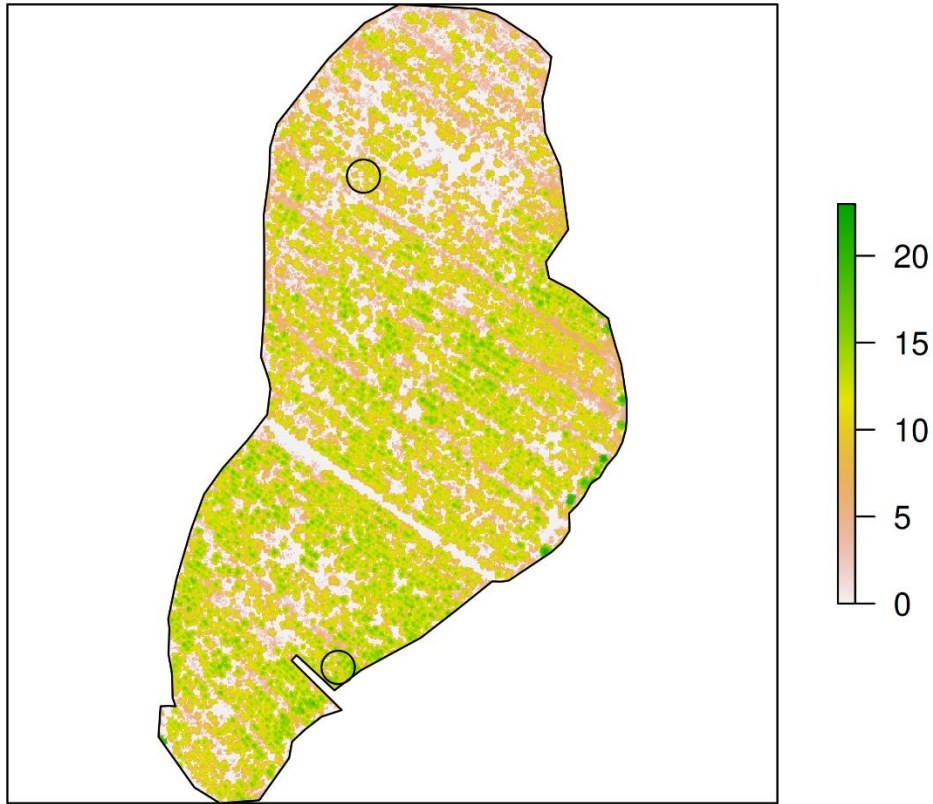


Fig 21. Stand 2020224 map showing the sample plots (two small black circles) located within it.

2020235

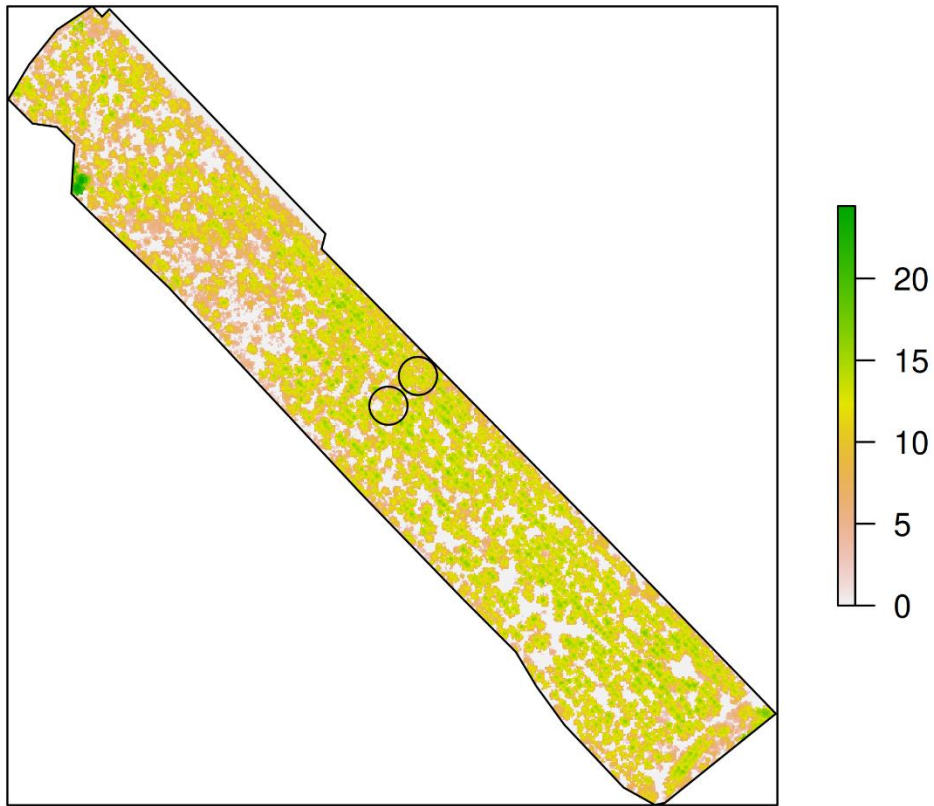


Fig 22. Stand 2020235 map showing the sample plots (two small black circles) located within it.

**Appendix II: Stand CHMs (with heights in meters) where EBLUP overestimated stand volume (▲ symbols, figure 5)**

**Stand ID's:**

2020022

2020028

2020042

2020048

2020056

2020022

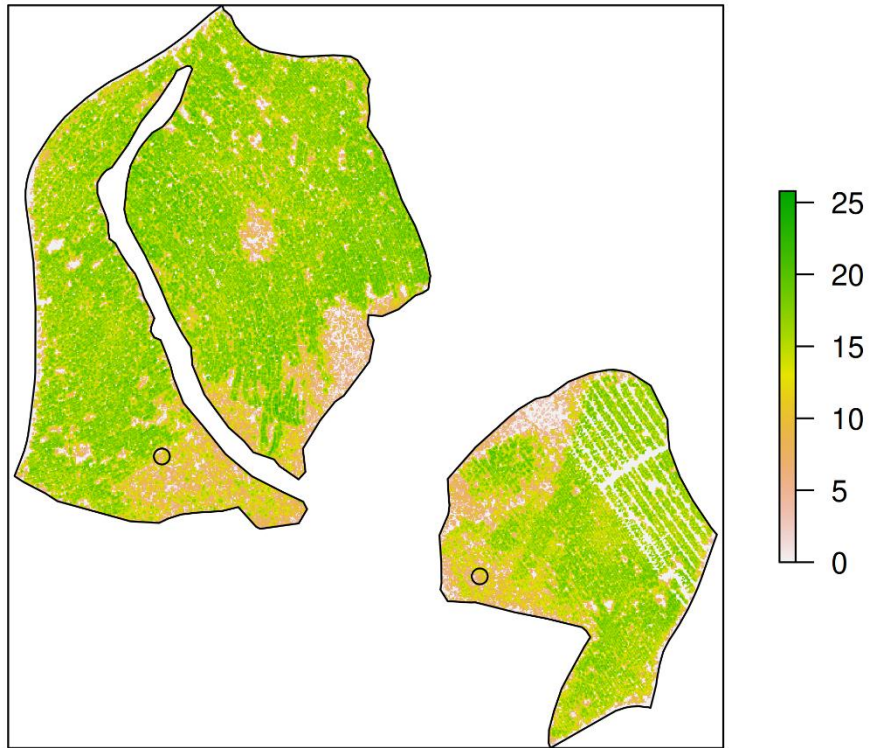


Fig 23. Stand 2020022 map showing the sample plots (two small black circles) located within it.

2020028

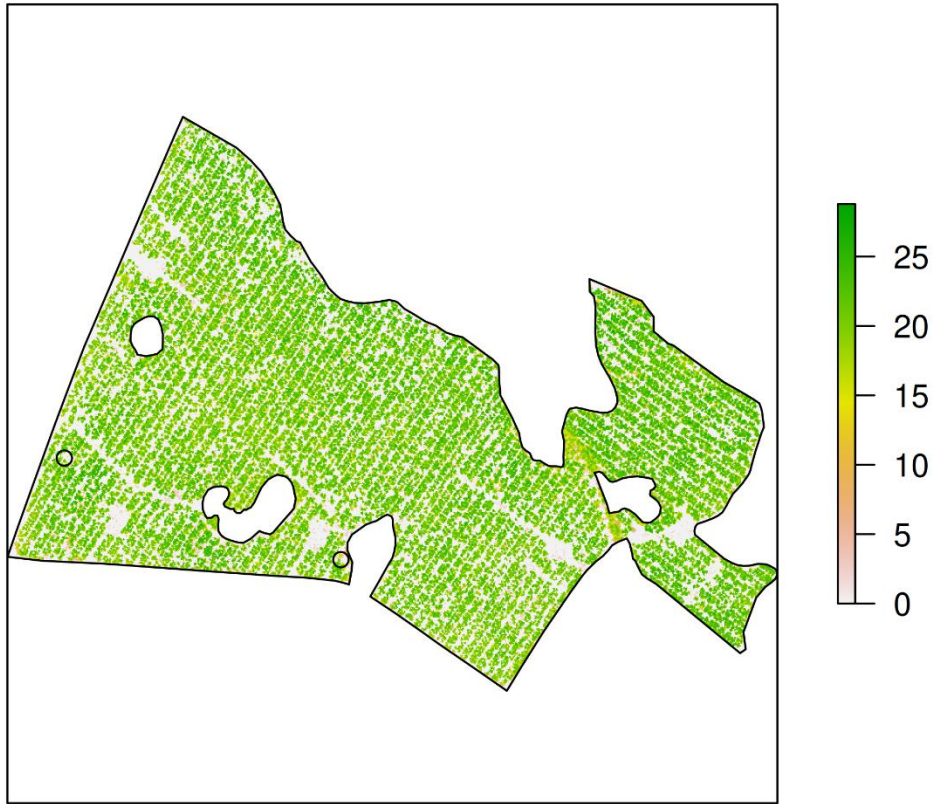


Fig 24. Stand 2020028 map showing the sample plots (two small black circles) located within it.

2020042

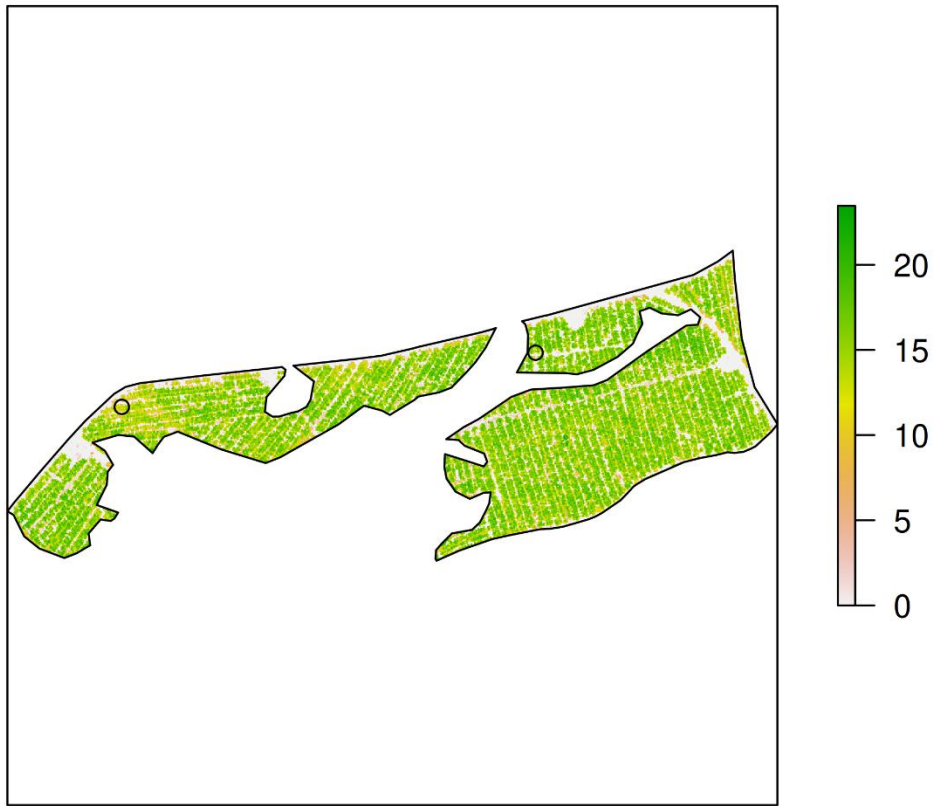


Fig 25. Stand 2020042 map showing the sample plots (two small black circles) located within it.

2020048

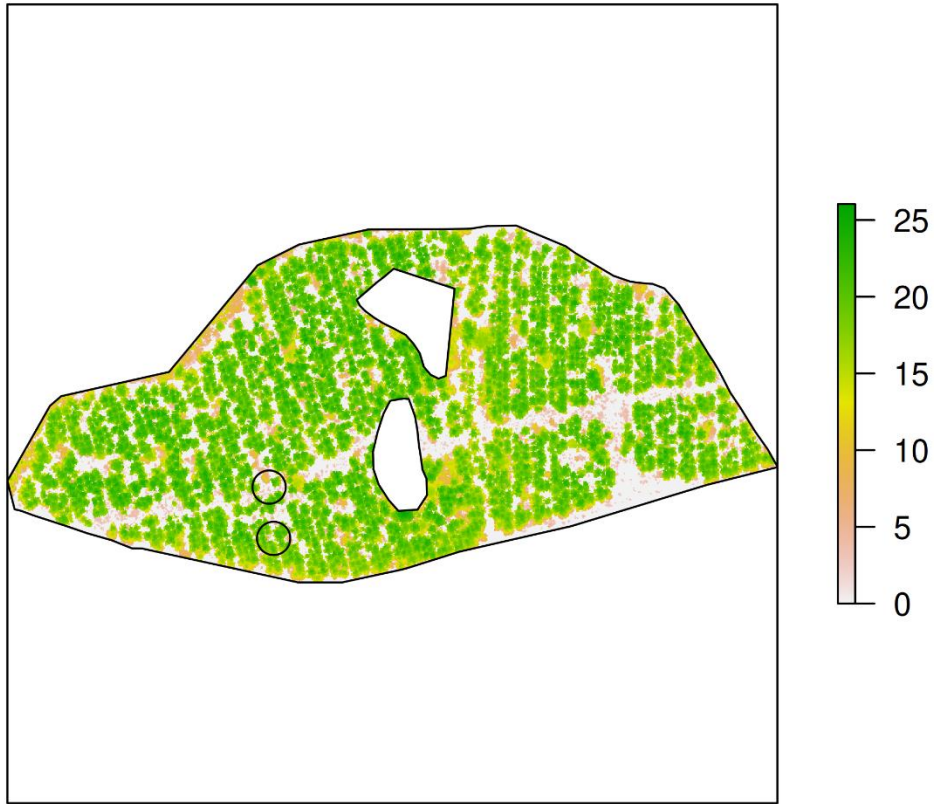


Fig 26. Stand 2020048 map showing the sample plots (two small black circles) located within it.

2020056

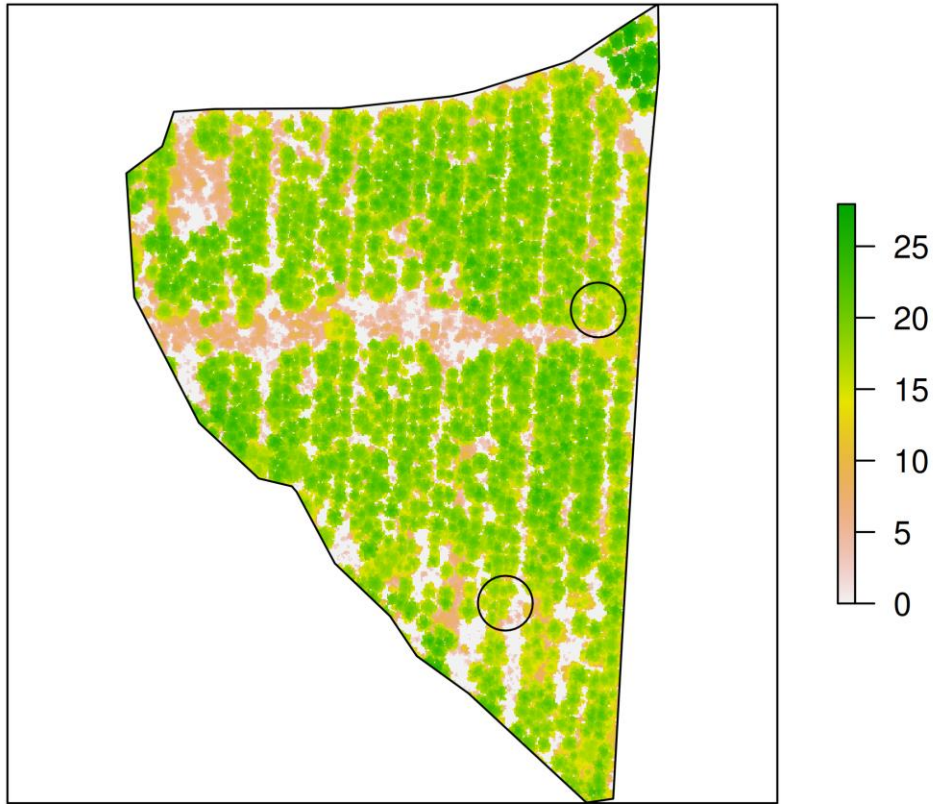


Fig 27. Stand 2020056 map showing the sample plots (two small black circles) located within it.

**Appendix III: Plot level and stand level predictor variables obtained for natural log of number of zeroes for stands and their respective two plots**

<b>StandID</b>	<b>Thinning status</b>	<b>Ln (No. of zeroes)</b>	<b>Plot</b>	<b>Ln (No. of Zeroes)</b>
<b>2020001</b>	0	6.95	A	6.11
			B	5.67
<b>2020006</b>	0	8.55	A	6.68
			B	4.09
<b>2020007</b>	0	8.11	A	6.04
			B	6.92
<b>2020008</b>	0	7.43	A	6.51
			B	6.38
<b>2020011</b>	1	9.00	A	8.95
			B	8.91
<b>2020012</b>	1	8.96	A	8.81
			B	8.82
<b>2020013</b>	1	8.86	A	8.75
			B	8.9
<b>2020014</b>	1	8.80	A	8.68
			B	9.16
<b>2020015</b>	1	8.52	A	8.49
			B	8.73
<b>2020017</b>	0	9.03	A	9
			B	8.98
<b>2020018</b>	0	7.02	A	9.03
			B	8.88
<b>2020022</b>	0	7.39	A	6.09
			B	8.68
<b>2020023</b>	1	8.50	A	6.89
			B	7.42
<b>2020026</b>	0	6.57	A	8.8
			B	8.52
<b>2020028</b>	1	8.73	A	5.86
			B	6.27

<b>2020029</b>	1	6.63	A	9.08
			B	8.96
<b>2020031</b>	0	7.35	A	6.17
			B	5.86
<b>2020032</b>	0	7.36	A	6.76
			B	6.57
<b>2020033</b>	1	7.82	A	7.09
			B	6.52
<b>2020037</b>	0	9.02	A	6.17
			B	6.93
<b>2020038</b>	0	7.84	A	7.23
			B	6.82
<b>2020042</b>	1	8.17	A	5.74
			B	6.04
<b>2020043</b>	0	6.76	A	8.35
			B	8.31
<b>2020045</b>	0	6.86	A	6.13
			B	5.35
<b>2020047</b>	0	7.01	A	5.35
			B	5.56
<b>2020048</b>	1	8.20	A	6.41
			B	6.76
<b>2020049</b>	0	6.47	A	7.57
			B	8.63
<b>2020051</b>	1	7.89	A	5.8
			B	5.19
<b>2020052</b>	0	6.57	A	7.3
			B	6.66
<b>2020054</b>	0	6.35	A	4.09
			B	6.33
<b>2020056</b>	1	7.67	A	5.97
			B	5.56
<b>2020057</b>	0	6.14	A	7.33

			B	8.3
<b>2020062</b>	1	7.89	A	5.6
			B	3.69
<b>2020063</b>	0	7.56	A	7.26
			B	7.35
<b>2020068</b>	0	7.15	A	4.94
			B	6.87
<b>2020070</b>	0	6.44	A	6.11
			B	6.17
<b>2020071</b>	0	6.74	A	4.61
			B	4.38
<b>2020073</b>	0	7.10	A	4.87
			B	3.4
<b>2020075</b>	1	7.03	A	4.5
			B	5.08
<b>2020076</b>	0	7.85	A	6.72
			B	6.35
<b>2020077</b>	0	7.26	A	5.08
			B	5.99
<b>2020083</b>	0	8.49	A	6.36
			B	6.38
<b>2020090</b>	0	6.87	A	8.54
			B	8.2
<b>2020091</b>	0	6.23	A	5.97
			B	6.49
<b>2020123</b>	1	7.74	A	7.02
			B	6.99
<b>2020124</b>	1	8.42	A	7.5
			B	7.36
<b>2020125</b>	1	7.38	A	8.08
			B	8.11
<b>2020129</b>	1	8.39	A	5.01
			B	6.92

<b>2020130</b>	1	8.39	A	7.4
			B	7.13
<b>2020131</b>	1	8.43	A	8.03
			B	7.82
<b>2020132</b>	1	7.87	A	8.13
			B	8.57
<b>2020133</b>	1	7.31	A	7.41
			B	7.48
<b>2020134</b>	1	7.33	A	7.25
			B	7.29
<b>2020137</b>	1	8.62	A	7.35
			B	7.06
<b>2020140</b>	0	8.30	A	8.41
			B	8.1
<b>2020142</b>	0	7.83	A	8.58
			B	7.27
<b>2020144</b>	1	8.32	A	6.77
			B	8.02
<b>2020148</b>	1	8.33	A	8.29
			B	8.03
<b>2020149</b>	0	7.33	A	8.19
			B	8.26
<b>2020150</b>	0	7.39	A	6.13
			B	6.35
<b>2020151</b>	0	7.49	A	6.77
			B	7.04
<b>2020153</b>	1	8.14	A	7.31
			B	7.22
<b>2020154</b>	0	6.51	A	8.05
			B	8.35
<b>2020158</b>	1	8.31	A	4.94
			B	5.08
<b>2020160</b>	1	7.63	A	7.94

			B	8.09
<b>2020161</b>	1	8.33	A	7.47
			B	7.1
<b>2020163</b>	1	8.59	A	8.35
			B	8.14
<b>2020165</b>	0	5.9	A	8.45
			B	8.68
<b>2020167</b>	1	8.21	A	2.3
			B	4.7
<b>2020169</b>	0	6.70	A	8.07
			B	8.45
<b>2020170</b>	0	8.12	A	5.3
			B	6.27
<b>2020171</b>	0	6.72	A	7.44
			B	6.41
<b>2020172</b>	0	6.64	A	5.63
			B	4.5
<b>2020174</b>	0	7.12	A	5.97
			B	5.97
<b>2020175</b>	0	7.28	A	7.19
			B	6.19
<b>2020176</b>	0	8.11	A	6.33
			B	6.73
<b>2020178</b>	1	8.03	A	7.78
			B	7.33
<b>2020179</b>	0	7.84	A	7.67
			B	7.85
<b>2020184</b>	0	8.20	A	6.71
			B	8.11
<b>2020185</b>	0	7.86	A	7.08
			B	8.28
<b>2020186</b>	0	7.88	A	7.72
			B	7.47

<b>2020187</b>	0	7.38	A	8.18
			B	7.29
<b>2020190</b>	0	7.94	A	6.45
			B	7.15
<b>2020191</b>	0	8.05	A	6.86
			B	7.18
<b>2020192</b>	0	8.18	A	7.44
			B	6.96
<b>2020194</b>	0	8.12	A	7.68
			B	8.3
<b>2020195</b>	0	7.92	A	7.85
			B	8.48
<b>2020196</b>	0	8.21	A	7.64
			B	7.61
<b>2020197</b>	0	7.89	A	7.36
			B	7.37
<b>2020205</b>	0	7.89	A	7.75
			B	6.65
<b>2020213</b>	0	7.66	A	6.43
			B	7.58
<b>2020214</b>	0	7.70	A	6.82
			B	7.99
<b>2020215</b>	0	7.40	A	7.83
			B	6.78
<b>2020219</b>	0	7.51	A	6.46
			B	6.41
<b>2020222</b>	0	8.04	A	6.29
			B	6.27
<b>2020223</b>	0	7.73	A	7.85
			B	7.62
<b>2020224</b>	0	8.18	A	7.57
			B	6.48
<b>2020225</b>	0	7.49	A	8.12

			B	6.78
<b>2020230</b>	1	8.42	A	7.45
			B	5.94
<b>2020235</b>	0	8.30	A	8.39
			B	8.26
<b>2020237</b>	1	8.90	A	6.97
			B	8.01
<b>2020238</b>	0	7.47	A	8.72
			B	8.93
<b>2020239</b>	0	7.86	A	7.06
			B	6.61
<b>2020240</b>	0	8.60	A	7.26
			B	6.73

**Appendix IV: Fixed effects coefficient estimates obtained for all three models (Volume, AGB, and Basal Area) with the model form:**

$$\hat{y}_{\text{syn}} = \mathbf{x}^T \hat{\boldsymbol{\beta}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \quad [10]$$

**where,**

$\beta_1 =$  *Sum of height per acre plot (Sum\_ht\_pa),*

$\beta_2 =$  *Log (zero\_per\_acre\_plot) , and*

$\beta_3 =$  *Thinning indicator variable*

<b>Attributes</b>	<b>Coefficients</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>T-value</b>
<b>Volume</b>	Intercept	-8.409e+02	4.316e+02	-1.948
	Sum_ht_pa	1.801e-02	8.526e-04	21.126
	Ln(zero_pa)	-2.826e+01	4.473e+01	-0.632
	Thinned	-2.132e+02	1.085e+02	-1.965
<b>AGB</b>	Intercept	-1.256e+01	8.132e+00	-1.545
	Sum_ht_pa	3.363e-04	1.606e-05	20.941
	Ln(zero_pa)	-7.250e-01	8.427e-01	-0.860
	Thinned	-4.682e+00	2.041e+00	-2.295
<b>Basal Area</b>	Intercept	9.007e+01	1.545e+01	5.829
	Sum_ht_pa	4.306e-04	3.056e-05	14.092
	Ln(zero_pa)	-7.938e+00	1.602e+00	-4.955
	Thinned	-2.588e+01	3.905e+00	-6.629