Auxiliary information resolution effects on small area estimation in plantation forest inventory

P. Corey Green1*, Harold E. Burkhart1, John W. Coulston2, Philip J. Radtke1 and Valerie A. Thomas1

1Virginia Polytechnic Institute and State University, 310 W. Campus Drive, Blacksburg, VA 24061, USA
2United States Forest Service, 1710 Research Center Drive, Blacksburg, VA 24060, USA

*Corresponding author: E-mail: pcgreen7@vt.edu

Received 13 January 2020

In forest inventory, traditional ground-based resource assessments are often expensive and time-consuming forcing managers to reduce sample sizes to meet budgetary and logistical constraints. Small area estimation (SAE) is a class of statistical estimators that uses a combination of traditional survey data and linearly related auxiliary information to improve estimate precision. These techniques have been shown to improve the precision of stand-level inventory estimates in loblolly pine plantations using lidar height percentiles and thinning status as covariates. In this study, the effects of reduced lidar point-cloud densities and lower digital elevation model (DEM) spatial resolutions were investigated for total planted volume estimates using area-level SAE models. In the managed Piedmont pine plantation conditions evaluated, lower lidar point-cloud densities and DEM spatial resolutions were found to have minimal effects on estimates and precision. The results of this study are promising to those interested in incorporating SAE methods into forest inventory programs.

Introduction

Accurate, precise and up-to-date forest resource assessments are important for informed management decisions. Foresters have often relied on design-based surveys using fixed- and variable-radius sample plots to estimate parameters for the areas of interest (Burkhart et al., 2019). Traditional, ground-based forest inventory provides reliable, unbiased estimates assuming a properly executed, valid sample design. In many cases, however, budgets and logistical issues limit the sample intensity leading to an estimate that does not meet the precision specified in the original design. In recent years, advances in remote sensing, computing capability and statistical methods have afforded many new options for resource evaluation (Coops, 2015). Small area estimation (SAE) is a class of statistical estimators that can leverage auxiliary information derived from, for example, remote sensing products to be used in a composite estimator with ground-based samples for increasing inventory precision.

In the southeastern US, loblolly pine (Pinus taeda L.) is the most widely planted and intensively managed commercial tree species (Baker and Langdon, 1990). Improving the efficiency and reliability of loblolly pine forest inventory is important due to the increasing frequency of land ownership changes (Fox et al., 2007; Jokela et al., 2010; Green et al., 2019) and interest in estimating stand characteristics for both commercial and ecological considerations (Andreu et al., 2008; Zhao et al., 2016; Green et al., 2019). For areas where loblolly pine is grown for commercial use, auxiliary data such as lidar or photogrammetrically derived point clouds are often available that can be used with SAE techniques.

Point-cloud elevation products

Point clouds are three-dimensional spatial data that characterize an area both horizontally and vertically. Light detection and ranging (lidar) is an active sensor often used to generate point clouds that contain elevation information (Campbell and Wynne, 2011). Advances in software and computing resources have made point clouds generated through photogrammetric techniques operationally possible and comparable with lidar (Goodbody et al., 2019).

The use of point clouds requires ground elevation models to normalize the canopy characterized by the three-dimensional data. Generally, elevation models generated from lidar data have been found to be accurate compared with other standard models such as the United States Geological Survey (USGS) 30- and 10-m products (e.g. Reutebuch et al., 2003; Hodgson and Bresnahan, 2004) and well suited to use in forest inventory applications (e.g. Tinkham et al., 2012). To our knowledge, no study has assessed the effects of elevation model spatial resolution for normalizing lidar point clouds when used as auxiliary data in SAE.

Point clouds vary in density depending on acquisition parameters such as flight altitude and sensor specifications. There have been multiple reports in the literature that lower point density lidar results in similar estimates as higher density point clouds.
in area-based analysis. Holmgren (2004) found small differences in prediction errors for tree height, basal area and stem volume using lidar with point densities ranging from 0.10 to 4.29 pulses m\(^{-2}\). Similar results were found in an investigation of plot-level volumes with point densities ranging from 0.13 to 12.7 pulses m\(^{-2}\) (Maltamo et al., 2006). In a model-assisted framework, Strunk et al. (2012) found minimal losses in estimate precision of basal area, volume, biomass, stem density and Lorey’s height with point densities as low as 0.13 pulses m\(^{-2}\) compared with 3 pulses m\(^{-2}\). In a study investigating a variety of stand variables including basal area, height, volume, biomass and crown attributes in *Pinus radiata* D. Don, point densities ranging from 0.5 to 9 pulses m\(^{-2}\) were found to provide comparable model precision (González-Ferreiro et al., 2012). Gobakken and Næsset (2008) found that the point density required for an operational inventory in Sweden could be reduced to as low as 0.25 pulses m\(^{-2}\) without a significant increase in estimate uncertainty of height, basal area and volume. Some studies have pointed to reduced precision with thinned point clouds. Magnussen et al. (2010) found reliability ratios decreased for common inventory attributes when predicted with progressively more thinned point clouds. In a study investigating the effects of plot size and point density, the strength of prediction models of forest structure attributes was found to be greatest with the highest density point clouds and largest plots; however, the impact of plot size was found to have more impact on predictions (Ruiz et al., 2014).

### Small area estimation

SAE models can be broadly classified as either ‘area-level’ or ‘unit-level’. Area-level models relate estimates of the area of interest to area-level covariates derived from the auxiliary data. Unit-level models relate sample unit direct estimates to the corresponding sample unit auxiliary data (Rao and Molina, 2015). SAE has been demonstrated to improve inventory precision in the forestry literature through both area-level (Goerndt et al., 2011; Magnussen et al., 2017; Green et al., 2019) and unit-level (Breidenbach and Astrup, 2012; Goerndt et al., 2013; Green et al., 2019) approaches. Comparisons between the two SAE methods have found that while unit-level models offer similar or greater increases in precision (Mauro et al., 2017; Breidenbach et al., 2018), area-level models are flexible and can be widely applied using a variety of data sources (Breidenbach et al., 2018). In linear formulations of both unit- and area-level models, the strengths of the correlations between the auxiliary data and parameters of interest are strongly related to their ability to reduce estimate uncertainty. To our knowledge, the effects of auxiliary data resolution on this relationship in managed plantations have not been explored in the SAE context.

### Research objectives and questions

While high-resolution elevation products are generally effective for characterizing forest conditions, there are several disadvantages that often limit their practical application: (1) High resolution data comes at the expense of time and budgetary resources. For example, the cost of high density lidar may be prohibitive due to longer, lower altitude flights. (2) Auxiliary data such as photogrammetric point clouds require complementary ancillary information (digital elevation models (DEMs) in the case of photogrammetric point clouds) for their use. In some situations, these data may not be available. (3) High resolution data are cumbersome. Despite modern computational resources, lower resolution products are easier to store, transfer and analyze. The advantages of incorporating auxiliary data must outweigh their costs and computational demands.

The overall objective of this work was to expand the investigations into the area-level SAE methods reported by Green et al. (2019) with regard to the following research questions:

1. What are the impacts on area-level SAE derived estimates and precision using DEMs with lower spatial resolution than those derived from lidar?
2. What are the impacts on area-level SAE derived estimates and precision using lidar point clouds with reduced point density?

These objectives broadly seek to provide guidance regarding the impacts of using lower resolution auxiliary data in the SAE framework. If lower resolution data can be used with similar efficacy as higher resolution data, SAE methods can potentially be more broadly applied.

### Data and methods

#### Study locations

This study focused on investigating the effects of auxiliary data precision in areas that loblolly pine is grown commercially. Forty managed loblolly pine plantations located in three Virginia State forests were used for this study (Figure 1). These stands are managed with silvicultural prescriptions common for the region with a goal of sustainably producing a mixture of fibre and solid wood products. Stands ranged in age from 9 to 43 years old. Of the 40 stands, 14 had been operationally thinned at least once. The selected stands cover a wide range of growing conditions and management scenarios typical for the region. For additional detail regarding the management and spatial distribution of the stands evaluated, readers are directed to Green et al. (2019).

#### Ground data

In the winter and early spring of 2019, temporary inventory plots were installed in every stand. One sample unit per 1.2 ha with a minimum distance of 70–80 m between plot centers was used to produce an assumed equal probability, simple random sample. Time constraints limited the sample intensity in some stands. 0.13 ha fixed-radius plots were used in unthinned stands and 0.02 ha fixed-radius plots were used in thinned stands. Unthinned stands with excessive natural regeneration were sampled using 0.01 ha fixed-radius plots.

On each sample unit, every living, planted stem was tallied and diameter at breast height (DBH) was measured using a diameter tape. A subset of trees per plot was measured for total height (Ht) using either a laser hypsometer or an electronic clinometer. Height trees were selected across the diameter distribution and at least one height per plot was measured. A minimum of 25 height trees per stand was targeted. All living natural trees with DBH ≥ 7.62 cm were tallied and measured for DBH. A subset...
of heights across the diameter distribution was measured for natural loblolly pine, Virginia pine (Pinus virginiana Mill.), and shortleaf pine (Pinus echinata Mill.). In addition, a subset of hardwood heights was measured. All pine (both planted and natural) heights not measured in the field were predicted using the model form in equation (1). Heights of planted pines were predicted using stand-level relationships while natural pine heights used a pooled, ‘region-wide’ relationship developed from all the field measured total heights.

\[
\ln (H_t) = b_0 + b_1 \left( DBH^{-1} \right)
\]

(1)

Plot-level information including thinning status, competing vegetation and other notes were also recorded. For all trees, volume was estimated using the allometric equations presented in Table 1. All ground data processing was performed using R (R Core Team, 2018). Additional packages used include the following: xlsx (Dragulescu and Arendt, 2018) and reshape2 (Wickham, 2007).

### Auxiliary information

For the entire study area, lidar and the associated 1-m DEMs are available from the United States Geological Survey (USGS). The 2015 ‘Chesapeake Bay’ lidar collection encompassed Appomattox-Buckingham (ABSF) and Cumberland (CUSF) State Forests while the 2014 ‘Sandy’ collection covers the Prince Edward-Gallion State Forest (PESF). These data were delivered in \(~1.5 \times 1.5\) km, non-overlapping tiles and were merged using the LASTools (LASTools, 2018) lasmerge function. This was to account for stands that overlapped multiple tiles. Details of the lidar collections are summarized in Table 2.

Lidar and the associated DEM processing was performed with a combination of R (R Core Team, 2018), FUSION (McGaughey, 2018) and LASTools (LASTools, 2018). Additional R packages used for spatial data processing include: Raster (Hijmans, 2019), sp (Pebesma and Bivand, 2005; Bivand et al., 2013) and rgdal (Bivand et al., 2013). This study evaluated the impact of three levels of reduced density point clouds. Using the lasthin function in the LASTools suite, the original point clouds were randomly sampled once to 50 per cent, 10 per cent and 1 per cent of
Lidar specifications for the Chesapeake Bay and Sandy projects

<table>
<thead>
<tr>
<th></th>
<th>Chesapeake Bay mission</th>
<th>Sandy mission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lidar sensor</td>
<td>Riegl 680i</td>
<td>Leica ALS60 or Leica ALS70</td>
</tr>
<tr>
<td>Scan angle (degrees)</td>
<td>60</td>
<td>Not available</td>
</tr>
<tr>
<td>Lidar density (pulses m⁻²)</td>
<td>2.3</td>
<td>Not available</td>
</tr>
<tr>
<td>Nominal pulse spacing (m)</td>
<td>0.66</td>
<td>0.7</td>
</tr>
<tr>
<td>Flight line overlaps</td>
<td>55%</td>
<td>30% (ALS60) or 20% (ALS70)</td>
</tr>
<tr>
<td>Pulse rates (kHz)</td>
<td>200</td>
<td>154.3 (ALS60) or 301.6 (ALS70)</td>
</tr>
</tbody>
</table>

Adapted from Green et al. (2019).

The empirical best linear unbiased predictor (EBLUP) is a weighted estimate based on the sampling estimate variance ($\Psi_i$) and the random error variance ($\sigma_v^2$). As the direct estimate (i.e. the ground sample) becomes more reliable, the estimate is weighted towards it. Conversely, the EBLUP weights more towards the synthetic estimate for unreliable (i.e. relatively high sample variance) samples. The form of the estimator utilized is:

$$
\hat{y}_i^n = \tilde{y}_i + (1 - \tilde{y}_i) z_i^T \hat{\beta} 
$$

(3)

where $\tilde{y}_i$ is the weight that uses both sources of error ($\Psi_i$ and $\sigma_v^2$) and is given by equation (4).

$$
\tilde{y}_i = \sigma_v^2 / (\sigma_v^2 + \Psi_i) 
$$

(4)

The area-level SAE models were fit using the SAE package in R (Molina and Marhuenda, 2015) using the 'mseFH' function. The REML method was used with all default parameters. Details of the REML procedure can be found in Rao and Molina (2015). In the SAE package (Molina and Marhuenda, 2015), the mean squared error (MSE) of an EBLUP is estimated with equation (5):

$$
\text{MSE}(\hat{y}_i^n) = g_1(\sigma_v^2) + g_2(\sigma_v^2) + 2g_3(\sigma_v^2) 
$$

(5)

where

$$
g_1(\sigma_v^2) = \tilde{y}_i \Psi_i 
$$

(6)

$$
g_2(\sigma_v^2) = (1 - \tilde{y}_i)^2 z_i^T \left( \frac{\bar{Z}_i^T \bar{Z}_i}{\sigma_v^2} \right) \bar{Z}_i 
$$

(7)

$$
g_3(\sigma_v^2) = (1 - \tilde{y}_i)^2 \frac{2}{\sum \left( \frac{1}{\sigma_v^2} \right) \Psi_i} 
$$

(8)

where $\bar{Z}$ is an $I \times m$ matrix of $z_i^T$ for each domain and $\Psi$ is an $I$-dimensional vector of $\Psi_i$ for each domain. Datta and Lahiri (2000) provide a detailed description of the MSE estimation.
Points falling below the 1–1 relationship indicate a lower relative error ratio (RER) for the direct estimate is

\[ RER_i(\%) = \frac{\sqrt{\text{MSE}(\text{EBLUP}_i)}}{\text{EBLUP}_i} \times 100 \]  

and the RER for the direct estimate is

\[ RER_i(\%) = \frac{\sqrt{\hat{\psi}_i}}{\hat{\theta}_i} \times 100 \]  

where

\[ \hat{\theta}_i = \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij} \]  

and

\[ \hat{\psi}_i = \text{Var}(\hat{\theta}_i) = n_i^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \]  

In addition, estimates were compared using 1–1 scatterplots. Points falling below the 1–1 relationship indicate a lower y-axis prediction compared with the x-axis prediction. Points falling above the 1–1 line indicate a higher y-axis prediction compared with the x-axis prediction.

The lidar height percentiles evaluated resulted in very similar SAE models. Due to the similarities, only results using the 80th percentile lidar height are presented. As reported in Green et al. (2019), area-level SAE models using both lidar and thinning status often improved the total volume estimate precision (Figure 2a). While bias cannot be directly assessed because the true population values are not known, there are no patterns suggesting a systematic under or over prediction by the SAE model evaluated (Figure 2b). The effects on the total volume estimate and its precision from reducing the spatial resolution of the DEM were minimal. As can be seen in Figure 3a, b, both the 30- and 10-m DEM resolutions resulted in very similar precision compared with the lidar-derived 1-m DEM when using the original, unthinned point cloud. In addition, the estimates are essentially unchanged using either the 30-m (Figure 4a) or 10-m (Figure 4b) DEM. When compared with the full point cloud using the 1-m DEM, thinned point clouds using either the 30- or 10-m DEM resulted in similar or slightly improved precision in most cases (Figure 3c–h). In addition, the estimate remained similar in most cases (Figure 4c–h).

### Discussion

The similar estimate precision using thinned point clouds does not agree with our initial expectation of lower estimate precision with lower density point clouds. The findings of this study are promising for the future of using point-cloud data from a variety of sources in area-based forest inventory. A common difficulty with using lidar in operational forest inventory is the cost of acquisition (McRoberts et al., 2018). Point-clouds derived from DAP have emerged as a promising, lower cost alternative with a growing body of literature demonstrating their potential (e.g. White et al., 2013; Goodbody et al., 2019). DAP generates point clouds that characterize the outer canopy envelope necessitating the need for an elevation model generated from other sources. It has been commonly assumed that lidar derived DEMs were necessary to effectively use DAP point clouds in operational settings (White et al., 2013). The results of this work suggest that in areas without significant topographic relief such as the southeastern US Piedmont, a DEM with lower spatial resolution will generate similar estimates with comparable precision in the SAE framework. Further, the results from this work suggest sparse point clouds are useful for improving the precision of inventory estimates. This is not to say that lower resolution point clouds are ‘better’, rather, they provide similar results using the methods described in the conditions evaluated. Past research investigating the effects of point density (e.g. Gobakken and Næsset, 2008; Strunk et al., 2012), support this finding. In planning future lidar missions, common constraints such as cost could be reduced by higher collection altitudes. It is unlikely that similar results would be found if individual tree detection methods were utilized. Increased point-cloud density has been shown to be important.
Figure 2 Area-level SAE results. (a) Model with lidar 80th height percentile and thinning status as auxiliary information relative error comparison and (b) estimate comparison with lidar 80th height percentile and thinning status as auxiliary information. Error bars represent one standard error in the x direction and the root mean squared error (RMSE) in the y-direction (Adapted from Green et al., 2019).

Figure 3 Relative RMSE comparison between (a) 1-m DEM with full point cloud and 30-m DEM with full point cloud, (b) 1-m DEM with full point cloud and 10 m DEM with full point cloud, (c) 1-m DEM with full point cloud and 30-m DEM with 50 per cent point cloud, (d) 1-m DEM with full point cloud and 10-m DEM with 50 per cent point cloud, (e) 1-m DEM with full point cloud and 30-m DEM with 10 per cent point cloud, (f) 1-m DEM with full point cloud and 10-m DEM with 10 per cent point cloud, (g) 1-m DEM with full point cloud and 30-m DEM with 1 per cent point cloud and (h) 1-m DEM with full point cloud and 10-m DEM with 1 per cent point cloud. Relative RMSE is calculated as: 100*(RMSE/EBLUP) where RMSE is the RMSE for the model estimate and EBLUP is the model estimate.

for accurate tree delineation and calculation of lidar cover metrics (Jakubowski et al., 2013).

An essential requirement when using SAE is a sufficient linear relationship between the variable of interest and the ancillary information (Rao and Molina, 2015). The strength of this relationship based on the coefficient of variation ($R^2$), was found to remain similar, or even slightly improve, as auxiliary data resolution decreased (Table 4). The strength of these linear relationships is major contributors to the similar SAE results found amongst the data evaluated. Additionally, the reduction of auxiliary data resolution did not significantly increase the contribution to the MSE estimation (equation 5) through $g_2(\sigma^2)$ and $g_3(\sigma^2)$ (equations 7 and 8).

Conclusions
This work demonstrated the effects of using lower resolution elevation models and lower density lidar point clouds as covariates...
Auxiliary information resolution effects on SAE

Figure 4 Estimate comparison between (a) 1-m DEM with full point cloud and 30-m DEM with full point cloud, (b) 1-m DEM with full point cloud and 10-m DEM with full point cloud, (c) 1-m DEM with full point cloud and 30-m DEM with 50 per cent point cloud, (d) 1-m DEM with full point cloud and 10-m DEM with 50 per cent point cloud, (e) 1-m DEM with full point cloud and 30-m DEM with 10 per cent point cloud, (f) 1-m DEM with full point cloud and 10-m DEM with 10 per cent point cloud, (g) 1-m DEM with full point cloud and 30-m DEM with 1 per cent point cloud and (h) 1-m DEM with full point cloud and 10-m DEM with 1 per cent point cloud.

Table 4 Comparison of strength of linear relationship between levels of auxiliary data precision and total planted volume using linear model form that includes the lidar 80th height percentile and thinning status

<table>
<thead>
<tr>
<th>Auxiliary data</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-m DEM, full point cloud</td>
<td>0.64</td>
</tr>
<tr>
<td>10-m DEM, 50% of original point cloud</td>
<td>0.63</td>
</tr>
<tr>
<td>10-m DEM, 10% of original point cloud</td>
<td>0.65</td>
</tr>
<tr>
<td>10-m DEM, 1% of original point cloud</td>
<td>0.69</td>
</tr>
<tr>
<td>30-m DEM, 50% of original point cloud</td>
<td>0.64</td>
</tr>
<tr>
<td>30-m DEM, 10% of original point cloud</td>
<td>0.66</td>
</tr>
<tr>
<td>30-m DEM, 1% of original point cloud</td>
<td>0.69</td>
</tr>
</tbody>
</table>

in loblolly pine plantations SAE models. To specifically address our research objectives and questions: (1) DEM resolution had minimal effects on estimates and their precision using the SAE methods evaluated and (2) Reduced lidar point-cloud densities resulted in similar total volume estimates and precision. Lower density point clouds were as useful as the full density point clouds evaluated. The results from this work are promising to inventory managers interested in using SAE techniques in loblolly pine plantations and indicate that lower resolution auxiliary information can be utilized with confidence in area-based SAE applications for the conditions evaluated in this study. The benefits of incorporating these lower resolution data may outweigh the acquisition and analysis costs leading to similar or lower estimate uncertainty at lower or similar costs respectively. Future work should investigate the implications of using lower resolution elevation products in areas with greater topographic diversity and with unit-level model forms. Further, elevation products derived from photogrammetric methods should be explored.

Acknowledgements

The authors would like to acknowledge John Peterson, Sheng-I Yang, Ralph Amateis and the Virginia Department of Forestry for their invaluable help with field work. Finally, Randy Wynne and the reviewers are acknowledged for their valuable insight and suggestions.

Conflict of interest statement

None declared.

Funding

This work was supported by the Forest Modeling Research Cooperative; the Department of Forest Resources and Environmental Conservation; Virginia Tech; and the USDA McIntire-Stennis program (Project No. VA-136630).

References


Forestry


Warner, J.R. and Goebel, N.B. 1963 Total and Bark Volume Tables for Small Diameter Loblolly, Shortleaf, and Virginia Pine in the Upper South Carolina Piedmont. Department of Forestry, South Carolina Agricultural Experiment Station, Forest research series no. 9.

