LiDAR forest inventory with single-tree, double- and single-phase procedures

Robert C. Parker and David L. Evans

Department of Forestry, Forest and Wildlife Research Center, Mississippi State University, Mississippi State, MS 39762
rparker@cfr.msstate.edu, devans@cfr.msstate.edu

Abstract.

Light Detection and Ranging (LiDAR) data at 0.5-2m postings were used with double-sample, stratified inventory procedures involving single-tree attribute relationships in mixed, natural, and planted species stands to yield sampling errors (one half the confidence interval expressed as a percentage of the mean) ranging from ±2.1% to ±11.5% at α=0.05. LiDAR sample trees were selected with focal filter procedures and heights computed as the difference between interpolated canopy and DEM surfaces. Tree dbh and height data were obtained on LiDAR:ground samples ranging from a 5:1 ratio on 0.08 ha rectangular strips to a 10:1 ratio on 0.02 ha circular plots established with a real-time Differential Global Positioning System. Dbh-height and ground-LiDAR height models were used to predict dbh from adjusted LiDAR height and compute Phase 2 ground and LiDAR estimates of basal area and volume. Phase 1 LiDAR estimates were computed by randomly assigning heights to species classes using the probability distribution from ground plots in each inventory strata. Phase 2 LiDAR estimates were computed by randomly assigning heights to species-product groups using a Monte Carlo simulation for each ground plot. There was no statistical difference between double-sample, mean volume estimates from 0.5m and 1m LiDAR posting densities with and without height bias adjustment or on smoothed and unsmoothed LiDAR canopy surfaces. Volume estimates from single-phase LiDAR inventory procedures utilizing existing tree attribute and LiDAR-ground height bias relationships were obtained with sampling errors of 1.8% to 5.5% for full and minimized data sets to test minimum LiDAR inventory requirements.

1. Introduction

Light detection and ranging (LiDAR) is a relatively new remote sensing tool that has the potential for use in the acquisition of measurement data for inventories of standing timber. LiDAR systems have been used in a variety of forestry applications (Magnussen and Boudewyn 1998, Lefsky et al. 1999, Means et al. 2000) for the quantification of biomass (Nelson et al. 2003), basal area, and tree and stand height estimates. Stand level, LiDAR inventory procedures involving average values of tree attributes such as dominant height, mean diameter, basal area, and volume have been applied to obtain unbiased stand level predictions (Naesset 2002, Naesset 2004, Popescu et al. 2002). Since LiDAR has the capability to detect individual trees and measure tree height with predictable bias when correlated with ground measurements (Persson et al. 2002, Holmgren 2004), strata-level inventory estimates involving individual tree, double-sample inventory procedure have been used by researchers from Mississippi State University in conifer and mixed hardwood stands in the Northwest and Southeast (Collins 2003, Parker and Evans 2004, Parker and Glass 2004, Parker and Mitchel 2005, Parker and Evans 2006, Williams 2006). The individual tree approach to stand inventory when combined with double-sample, ground procedures permits relatively precise estimates of volume with a simple prediction function for ground-LiDAR height bias and ground-based attribute relationship functions for tree diameter and total height which can be used with any standard, standing tree volume function. Stand level approaches involving average tree attribute values for sampling
units require more sophisticated prediction models than an individual tree approach and procedures that differ radically from traditional ground-based inventory methods. The objective of this paper is to summarize and discuss the procedures, models, and advantages/disadvantages of the single-tree approach to using LiDAR data in double- and single-phase forest inventory methods.

2. Methods

2.1 Flight Planning for a LiDAR Inventory

Small-footprint, multi-return LiDAR data have been acquired with various sensors to attain nominal posting spacings of 0.5 -2.0 m, 0.25- 4 point/m², and footprint sizes of 0.122 - 0.330 m for two returns per pulse (Table 1). Aircraft altitudes of 600-1,000 m and swath widths of 189 - 609 m were used. The minimum required density of LiDAR hits is a function of the crown size, average height, and spatial density of the sample trees in the primary canopy. Acceptable sampling statistics were attained for sparse densities of large crown conifers in Idaho (± 11.5% sampling error at the 95% confidence level with a standard error of ± 5 m³, Parker and Evans 2004) with 0.25 points/m²; however, 1 point/m² was required to achieve acceptable inventory results in natural pine and mixed pine-hardwood stands (± 7.6% sampling error, Parker and Glass 2004 ) and 2 points/m² in young (6+ years) pine plantations (± 2.2% sampling error, Parker and Evans 2006) in the Southeast. Increasing LiDAR density from 2 to 4 points/m² did not statistically improve the volume estimation precision and the increased “noise” in the high-density LiDAR data translated into additional sampling error about the volume estimate.

<table>
<thead>
<tr>
<th>LiDAR specification</th>
<th>Example 1 (Idaho)</th>
<th>Example 2 (CP LA)</th>
<th>Example 3 (FW LA)</th>
<th>Example 3 (CP LA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points per m²</td>
<td>0.25</td>
<td>1</td>
<td>1.9</td>
<td>4</td>
</tr>
<tr>
<td>Nominal spacing</td>
<td>2.0 m</td>
<td>1.0 m</td>
<td>0.7 m</td>
<td>0.5 m</td>
</tr>
<tr>
<td>Footprint size</td>
<td>0.330 m</td>
<td>0.213 m</td>
<td>0.250 m</td>
<td>0.122 m</td>
</tr>
<tr>
<td>Aircraft altitude</td>
<td>1,000 m</td>
<td>1,067 m</td>
<td>1,000 m</td>
<td>610 m</td>
</tr>
<tr>
<td>Swath width</td>
<td>600+ m</td>
<td>609 m</td>
<td>243 m</td>
<td>189 m</td>
</tr>
<tr>
<td>Tract size</td>
<td>2,023 ha</td>
<td>485 ha</td>
<td>18,000 ha</td>
<td>485 ha</td>
</tr>
<tr>
<td>Percent LiDAR Coverage</td>
<td>1.43%</td>
<td>100%</td>
<td>10%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Target aircraft altitude is a function of the desired swath width and scan angle for the LiDAR pulse generator and sensor and the technical ability of the sensor to achieve the desired posting density. The swath width diminishes as the desired posting density increases, but reasonable swath widths can be achieved with 2-4 points/m². An important factor influencing desired swath width was the size of the ground-based sample plots used in the inventory procedure. The swath should be sufficiently wide to encompass the sample ground and LiDAR plots within the center one-third of the swath so as to minimize the “edge effects” of the LiDAR data. Tree attribute measurements are severely compromised at the extremes edges of the swath and scan angle.

Percent LiDAR coverage is a function of economics and inventory design. LiDAR data is relatively expensive to obtain and complete area coverage is normally not required for most timber inventory designs. In some instances, the cost of complete LiDAR coverage to produce an accurate, up-to-data Digital Elevation Model (DEM) may be more justifiable than the expense for a timber inventory. The use of a current Geographical Information System (GIS) to
locate flight lines that cross the desired sampling strata can minimize the percent coverage of the LiDAR area. Most forested areas to be inventoried can be flown with ten percent or less LiDAR area coverage by orienting flight lines so as to cross the target inventory strata at desired flight line intervals.

2.2 Field and LiDAR Plot Design and Procedures

Inventory design for the single-tree LiDAR applications involved the use of circular (Parker and Evans 2006, Parker and Glass 2004) or rectangular plots (Parker and Evans 2004) with all plots being Phase 1 LiDAR plots and every 5th plot as a Phase 2 ground plot. Field designs varied from a 9:1 ratio of LiDAR to ground circular plots in a nested arrangement to a 10:1 ratio with rectangular or circular plots along a flight line (Figure 1). UTM coordinates were established at the center of each circular Phase 2 plot or at the endpoints of rectangular plots for navigation with a real-time Differential Global Positioning System (DGPS). Differential corrections from either the U.S. Government WAAS or private enterprise OmniStar geo-stationary satellite were obtained satisfactorily under tree canopies by using a large dome antenna. Based on informal field tests on surveyed bench marks, field locations were obtained with approximately 1 m accuracies with both systems.

2.3 LiDAR Surfacing for Tree Location and Height Determination

The LiDAR data were processed to produce a ground surface or digital terrain model (DTM) and a tree surface for determination of sample tree locations and tree heights within the sample field and LiDAR plot areas. LiDAR data sets were surfaced to produce 1st return canopy and last return DTM with 0.2 m cell sizes using a linear interpolation technique. Tree locations and heights were determined with algorithms and focal filter procedures developed by McCombs et al. (2003) that used a variable search window radius based on relative density. These procedures used moving 2.5, 4.0, or 5.5 ft radius search windows to identify each tree peak as the point that is higher than 85% of the surrounding maxima from one of the three search window, radius files. Tree height was interpreted as the difference between canopy and DTM z-values at each tree peak location. Tree heights were converted to point coverages and clipped to sample area boundaries using UTM coordinates to describe sample plot locations and sizes.

A spatial filtering technique derived from image analysis called smoothing was used to reduce commission errors by minimizing the abrupt elevation changes in the initial canopy surface. The Focal Analysis option in ERDAS' Imagine software performed smoothing based on user-defined inputs for window size and preferred statistical procedure. A 5-by-5 pixel window was used to create a 1 m² filter that would avoid removal of small peaks in the canopy surface (small trees), while maximizing the smoothing function. The filter moved across the LiDAR canopy surface, pixel by pixel, averaged the values within the window, and placed the result in the center pixel.

Smoothing heights on LiDAR surfaces improved the relationship between LiDAR and ground tree heights in terms of $R^2$, reduced height biases for hardwoods, increased height biases for pines, and improved target recognition in terms of trees/ac estimates. There were however no statistical differences ($\alpha =0.05$) between double-sample regression volume estimates with smoothed versus unsmoothed LiDAR surfaces from low- or high-density LiDAR. Standard errors and sampling errors of the regression estimates were lower for all unsmoothed LiDAR data models than with smoothed data models. Thus, smoothing heights on LiDAR surfaces did not produce a statistical gain for volume estimation using double-sample procedures.
Figure 1. Sample designs for 0.02 ha plots with 9:1 and 10:1 ratios of LiDAR plots (Phase 1) to ground (Phase 2) circular plots in a nested arrangement and plots along a flight line.

3. Results

3.1 Double-Sample, Regression Estimator Procedures

The double-sample model widely used with ground-based point sampling (Avery and Burkhart 2002) and aerial photogrammetric inventories and adapted for these studies was:

\[
\bar{Y}_b = \bar{y}_2 + \beta(X_1 - \bar{x}_1)
\]  

(1)

With traditional aerial photogrammetric inventories, the \(X_{1i}\) and \(X_{2i}\) variables are photographic volume/unit area and ground volume/unit area from Phase 1 and Phase 2 plots, respectively, and \(\beta\) is the regression slope coefficient for \(y_i\) (ground volume) over \(x_{2i}\) (photo volume) on ground plots. Thus, a Phase 1 (large sample) variable such as remotely sensed (i.e. photographic or LiDAR derived) volume has a strong, identifiable relationship with a Phase 2 (small sample) variable such as ground volume.

In applications of the double-sample model with single-tree LiDAR data, Phase 2 sample tree measures of dbh and height were used to derive heightdbh and dbh-height equations of the model type:

\[
H_{gr} = b_0 + b_1[\ln(\text{dbh})]^{b_2}[\text{age}]^{b_3} + \varepsilon
\]  

(2)

33
dbh = b_0 + b_1[\ln(H_{gr})]b_2[\text{age}]b_3 + \varepsilon \tag{3}

where \(H_{gr}\) was ground measured height, dbh was ground measured dbh, and age was average stand age (years) from GIS data. The age variable in Models (2) and (3) was removed when age did not contribute significantly to the relationship. Models (2) and (3) were derived from the ground-measured sample trees, but one is not a back transformation of the other. Model (2) was applied to ground-plot trees where dbh was measured on all trees and heights on a sub-sample. Model (3) was applied to LiDAR derived tree heights to obtain a dbh for single-tree volume computation.

Generally only 2 trees per ground plot were measured for height; dbh was measured on all trees. The height-dbh Model (2) was applied to trees on the ground plots for which height was not measured to obtain a height for single-tree volume computation. Sample tree heights from the Phase 2 ground plots were used to predict ground height of target trees identified on LiDAR surfaces. The dbh-height Model (3) was applied to the bias-adjusted, single-tree LiDAR height from the ground-LiDAR height bias Model:

\[ H_{gr} = b_0 + b_1(H_{Li}) + \varepsilon \tag{4} \]

where \(H_{gr}\) was measured ground height of trees on Phase 2 plots and \(H_{Li}\) was interpolated height of the same trees from the LiDAR surface.

Derived dbh on LiDAR plots and derived height on ground plots permitted the use of a standard, standing-tree volume equation with dbh and height as variables to predict volume. Thus, the double-sample models used in this study involved LiDAR mean estimates of basal area (LiBA from Phase 1 and liba from Phase 2 with matching ground plot) and volume (LiVOL from Phase 1 and livol from Phase 2 with matching ground plot) for the x-variables as:

\[ \overline{Y}_{br} = \overline{y} + \beta(LiBA - liba) + \varepsilon \tag{5} \]
\[ \overline{Y}_{br} = \overline{y} + \beta(LiVOL - livol) + \varepsilon \tag{6} \]

with variance:

\[ s_{\overline{y}\varepsilon}^2 = \frac{s_{\overline{y}}^2}{n_2} + \frac{s_{\overline{y}}^2 - s_{\overline{x}}^2}{n_1} \tag{7} \]

where \(\overline{y}\) was Phase 2 mean ground volume, \(\beta\) was the regression slope coefficient for \(y_i\) (ground volume/unit area) over \(x_3\) (LiDAR volume/unit area or basal area/unit area on ground plot) and \(x_{1i}\) was volume or basal area on the LiDAR plot. Data were fitted to Models (5) and (6) for all data combined (i.e. non-stratified), each ages-class strata, and combined strata. Combined strata, linear regression estimates of volume and associated standard error of each double-sample model were obtained by:

\[ \overline{Y}_{br,i} = \sum \frac{(n_{ij} + n_{ik})}{N}(\overline{Y}_{br,i}) \tag{8} \]
where \( n_{1i} \) and \( n_{2i} \) were Phase 1 and 2 sample sizes respectively for stratum \( i \), \( i = 1 \) to \( X \) strata.

All double-sample volume computations were performed with the Windows software program LIDARDS (LiDAR Double-Sample) developed by Parker (2005). The software allowed the user to specify dbh limits for species-product classes, regression coefficients for the dbh-height and ground height-LiDAR height models, stratum definitions of beginning and ending plot numbers and average age, and to enter comma delimited data files of Phase 1 LiDAR heights and Phase 2 ground-plot trees (species, product, dbh, and height of sample trees). LiDAR heights in the Phase 1 data were allocated in a Monte Carlo simulation to species-product classes on each matching Phase 2 ground plot on the basis of percent distribution by numbers on the ground plot. Since species and dbh of the LiDAR trees are unknown, the Monte Carlo simulation (50 iterations) would randomly allocate the LiDAR derived trees (dbh predicted from adjusted LiDAR-to-ground height) to species-product classes and obtain a mean basal area and volume estimate for the species-product class. Thus, basal area and volume estimates from Phase 1 LiDAR plots that had a matching Phase 2 ground plot became Phase 2 LiDAR plots. Phase 1 LiDAR heights that did not have a matching Phase 2 ground plot were randomly allocated to encountered species classes in each stratum in a single iteration and used to compute mean estimates of numbers of trees, basal area, and volume. Phase 2 tree measures of dbh and height were used to compute LiDAR estimates of mean basal area and volume by using field-derived dbh-height equations to predict dbh from LiDAR height and volume. Predicted dbh and height were used in a single-tree volume function to predict individual/single tree volume. Double-sample volume estimates and associated precision statistics were computed with Models (5) and (6) for each stratum and with Models (8) and (9) for combined strata.

### 3.2 Single-Phase Inventory Procedures

A recent study by Williams (2006) investigated the minimal data inputs for a LiDAR based timber inventory with single-phase procedures. Since the ground phase of a double-sample field procedure is both expensive and time consuming, Williams investigated the feasibility of using LiDAR data in single-tree approach to obtaining volume estimates by stratum with a single-phase inventory procedure. Previous studies have shown that LiDAR can provide precise, but biased estimates of tree numbers and heights. If the assumptions are made that (1) the LiDAR height bias is known and relatively constant for a given species-origin class (i.e. pine plantations) and (2) previously established tree attribute relationships are also known, inventory estimates of volume can be obtained with a single-tree approach and single-phase procedures from LiDAR data only.

The tree attribute relationship from Model (3) was developed from ground measurements of dbh and height within Continuous Forest Inventory (CFI) plots and from the Phase 2 ground plots in the double-sample approach by Parker and Evans (2006). Sample trees were randomly selected from the original data sets in groups of 75 and fitted to tree attribute Model (3) under the assumption that ground data were available from previous studies. The ground-LiDAR height bias equation obtained from Model (4) (Parker and Evans 2006) was assumed to be constant and known. LiDAR derived heights from Phase 1 plots were adjusted for bias with Model (4) then used with Model (3) to obtain single-tree dbh estimates for use with a single-tree volume function in a conventional, single-phase inventory processor. The stratum volume estimates and precision statistics were compared to estimates obtained from the Phase 2 ground plots by Parker and Evans (2006).
The single-tree, single-phase volume estimates from LiDAR data compared favorably with ground plot estimates (Table 2). At the tract level for 20 age class strata on 10,443 ha, there was no statistical difference between the single-phase LiDAR estimates and the ground plot estimate. Single-phase volume estimates were obtained for the full data set of Phase 1 LiDAR plots, the Phase 1 LiDAR plots that had a matching Phase 2 ground plot, and 5 iterations of reducing the LiDAR data set to a 5:1 ratio with ground plots within each stratum. The Williams (2006) study found no statistical difference (α=0.05) between tract-level, single-phase volume estimates where the single-tree relationship model was developed with 1,539 sample trees from the Phase 2 ground plots by Parker and Evan (2006), 1,509 trees from the regional CFI plots, or 5 iterations of 75 randomly selected trees from the Phase 2 data set. The study concluded that precise single-phase LiDAR inventory estimates are feasible with minimal inputs of ground data for establishing tree attribute relationships. A potential application of the single-tree, single-phase inventory procedure would be the rapid post-thinning inventory of pine plantations and periodic inventories of forested holdings.

Table 2. Comparison of single-tree, single-phase, volume predictions from LiDAR data using 0.02 ha circular plots on 10,443 ha of pine plantations with ground plot volume estimates where sampling error was half the (1-α) confidence interval expressed as a percentage of the mean.

<table>
<thead>
<tr>
<th>Data set description</th>
<th>No. of plots</th>
<th>Sampling error%</th>
<th>Tract volume vs. control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground plots – Phase 2</td>
<td>842</td>
<td>2.8%</td>
<td>Control</td>
</tr>
<tr>
<td>LiDAR plots – Phase 1</td>
<td>7,562</td>
<td>1.8%</td>
<td>Not statistically different</td>
</tr>
<tr>
<td>LiDAR plots – Phase 2</td>
<td>842</td>
<td>5.2%</td>
<td>Not statistically different</td>
</tr>
<tr>
<td>LiDAR plots – Phase 1:2</td>
<td>5:1 ratio</td>
<td>5.5%</td>
<td>Not statistically different in 3 of 5 iterations</td>
</tr>
<tr>
<td>LiDAR plots – Phase 1</td>
<td>Phase 1:2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Discussion about Single-Tree LiDAR Inventory Procedures

LiDAR provides precise x, y, and z coordinate data that can be used to extract tree heights and locations; however, there are several sources of bias that can impact the accuracy of a per-unit area volume estimate. Height bias is primarily caused by the failure of the laser pulse to hit the terminal leader, but this bias can be predicted with acceptable success in conifers but not in hardwoods with broad rounded crowns. Height bias can also be introduced by the interpolation of tree heights with mixed linear and nonlinear procedures within the same data set. Tree count bias has at least two sources of origin; trees in the mid and lower canopy layers are hidden from the laser pulse by a dominant canopy and tree maxima locations may not be interpreted correctly during the LiDAR surfacing and height extraction process.

The precision of volume estimates with single-tree LiDAR procedures in a double-sample process is not affected by the height or tree count bias inherent in LiDAR data. These biases are effectively adjusted during the double-sample inventory procedures. The height bias can be adjusted prior to single-tree volume computations with LiDAR derived heights or afterwards during the double-sample volume adjustment process. Height bias correction prior to volume computation improves the accuracy of the resulting per-unit area volume estimate. The tree count bias caused by canopy coverage or LiDAR surfacing/processing is also effectively adjusted through the double-sample volume computations. Tree count bias however has a major impact on the accuracy of volume estimates in a single-tree procedure computed with single-phase inventory methods. Thus, if there is any doubt about the validity of the tree counts or height bias during LiDAR processing, a double-sample volume computation process should be used.
Ground sample tree measurements are needed to establish the ground-LiDAR height bias and the relationship between standing tree height and dbh. The number of sample trees is dependent upon the number of parameters in the regression models and variation in the data. A reasonable rule of thumb is 25 samples per parameter estimated in a regression model. Since the LiDAR height bias is relatively constant for a species in a local area and the dbh-height relationship for a give set of species-site conditions is also relatively stable, the sample trees can be obtained from either the LiDAR inventory or surrounding forested areas.

Establishing the LiDAR to ground tree height bias requires the matching of trees on the ground and on the LiDAR surface. Past experience has shown that the location of a plot center must be done with a real time, DGPS and the distance and direction to the sample trees from the plot center should be obtained with a laser so that the x, y coordinates of the sample trees can be located on the LiDAR surface.

Sample plot size and shape on the ground and on the LiDAR surface should be a function of tree density on the ground and the LiDAR processing procedures employed. Experience has shown that the ground plot size should be adjusted such that a minimum of 6 and a maximum of approximately 15 trees should be selected. The minimum number is associated with the within and between-plot variation and the maximum is a logistical consideration for minimizing omission/commission errors in tallying trees. Rectangular plots are easier to handle during the LiDAR processing, but more difficult to establish in the field and to use in establishing distance and direction to sample trees from a DGPS location. The cost of a LiDAR inventory can be minimized by flying only a portion of the desired inventory area. As long as the LiDAR swaths cover the desired strata, ground plots could be located randomly or systematically within strata within swaths. DGPS permits the location of sample plots and trees with relative ease and precision.

References


Parker, R.C. 2005. Computer automation of a LiDAR double-sample forest inventory. Forest and Wildlife Research Center Publication No. FO275, Mississippi State University, Starkville, MS.
Parker, R.C.; Evans, D.L. 2006. Industrial application of a LiDAR double-sample forest inventory. Forest and Wildlife Research Center Publication No. FO317, Mississippi State University, Starkville, MS.