

# Using Airborne Light Detection and Ranging as a Sampling Tool for Estimating Forest Biomass Resources in the Upper Tanana Valley of Interior Alaska

Hans-Erik Andersen, Jacob Strunk, and Hailemariam Temesgen

## ABSTRACT

Airborne laser scanning, collected in a sampling mode, has the potential to be a valuable tool for estimating the biomass resources available to support bioenergy production in rural communities of interior Alaska. In this study, we present a methodology for estimating forest biomass over a 201,226-ha area (of which 163,913 ha are forested) in the upper Tanana valley of interior Alaska using a combination of 79 field plots and high-density airborne light detection and ranging (LiDAR) collected in a sampling mode along 27 single strips (swaths) spaced approximately 2.5 km apart. A model-based approach to estimating total aboveground biomass for the area is presented. Although a design-based sampling approach (based on a probability sample of field plots) would allow for stronger inference, a model-based approach is justified when the cost of obtaining a probability sample is prohibitive. Using a simulation-based approach, the proportion of the variability associated with sampling error and modeling error was assessed. Results indicate that LiDAR sampling can be used to obtain estimates of total biomass with an acceptable level of precision ( $8.1 \pm 0.7$  [8%] teragrams [total  $\pm$  SD]), with sampling error accounting for 58% of the SD of the bootstrap distribution. In addition, we investigated the influence of plot location (i.e., GPS) error, plot size, and field-measured diameter threshold on the variability of the total biomass estimate. We found that using a larger plot (1/30 ha versus 1/59 ha) and a lower diameter threshold (7.6 versus 12.5 cm) significantly reduced the SD of the bootstrap distribution (by approximately 20%), whereas larger plot location error (over a range from 0 to 20 m root mean square error) steadily increased variability at both plot sizes.

**Keywords:** LiDAR, biomass, forest inventory, sampling

Rural communities in interior Alaska are highly vulnerable to the ever-increasing cost of nonrenewable fossil fuels. These communities are typically relatively small, very remote, and heavily dependent on diesel fuel to provide for electric power and heating needs. This combination of factors places tremendous economic pressure on the limited resources of these remote rural communities, where diesel fuel prices have increased 83% from 2000 to 2005, and utility costs can often amount to more than a third of a household's income, as opposed to the 2–4% that is typical in the urban areas of Alaska (Haley and Saylor 2007). Higher utility costs can have a significant impact on the economic well-being and quality of life in rural Alaska. Because demand for energy use in the United States is typically inelastic (not particularly sensitive to fuel costs), higher energy costs in these communities are often offset by reducing expenditures on other needed products and services. In some cases, rural Alaskans simply cannot afford to pay energy bills, which will have repercussions across the broader economy of both rural and urban Alaska (Haley and Saylor 2007).

Increasing economic pressure due to higher diesel fuel costs has led many rural Alaska communities to consider alternative and renewable sources of energy, including small hydroelectric, wind, and biomass-fired combined heat and power systems. Wind and hydroelectric power are viable options in coastal regions and near large rivers in Alaska, whereas wood biomass is the most abundant and lowest-cost alternative fuel source available to rural communities in the interior region of Alaska (Crimp et al. 2008). According to a recent analysis, wood biomass resources in Alaska are capable of providing 48.4 trillion British thermal units/year, equivalent to 1.32 billion liters of no. 2 diesel fuel, or 12 times the amount of diesel fuel currently consumed in rural Alaska communities (Crimp et al. 2008). Obviously, much of this biomass resource is not available to communities because of complex ownership patterns, inaccessibility, and the relatively high costs of extraction. Another obstacle to the development of bioenergy systems is the relatively imprecise inventory information currently available characterizing the biomass resources available to rural communities (J. Hermanns, Tok Area Forester, Alaska Department of Natural Resources, personal communication, 2008).

Manuscript received May 28, 2010, accepted February 10, 2011.

Hans-Erik Andersen ([handersen@fs.fed.us](mailto:handersen@fs.fed.us)), USDA Forest Service, Pacific Northwest Research Station, Seattle, WA 98195. Jacob Strunk and Hailemariam Temesgen, Department of Forest Engineering, Resources and Management, Oregon State University, LORVALLIS, OR 97331. The authors thank Jeff Hermanns of the Division of Forestry, Alaska State Department of Natural Resources, and Jamie Hollingsworth of the University of Alaska-Fairbanks for their tremendous assistance in carrying out the fieldwork for this project. In addition, they thank Marc Much, Will Buntin, Dan Irvine, Seth Ayotte, and Kevin Dobelbower of the Anchorage Forestry Sciences Laboratory for their assistance in collecting the field data.

This article uses metric units; the applicable conversion factors are: centimeters (cm): 1 cm = 0.39 in.; meters (m): 1 m = 3.3 ft; square meters (m<sup>2</sup>): 1 m<sup>2</sup> = 10.8 ft<sup>2</sup>; kilometers (km): 1 km = 0.6 mi; hectares (ha): 1 ha = 2.47 ac.

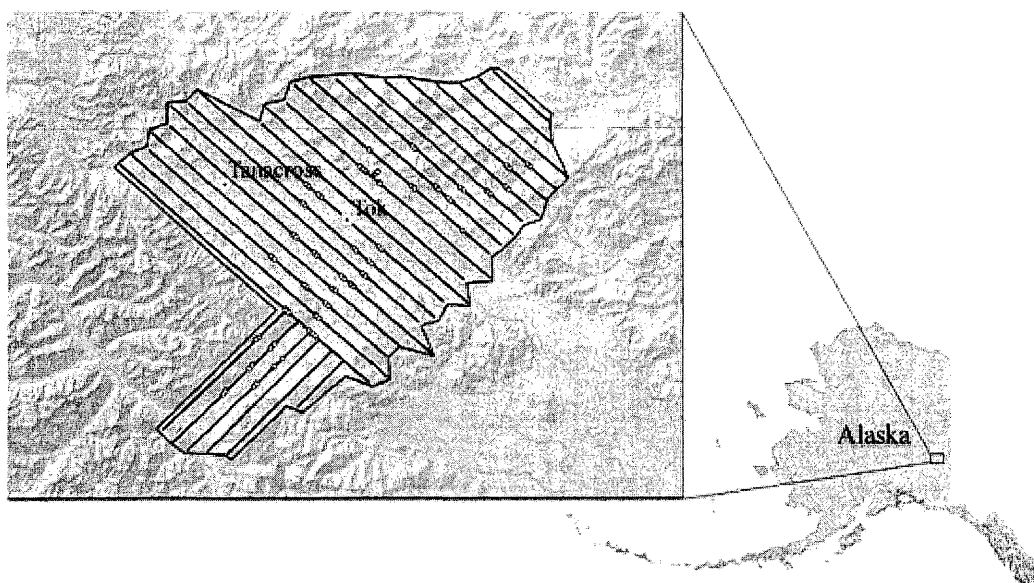


Figure 1. Location of study area (black outline) in the upper Tanana valley of interior Alaska. Black lines indicate location of LiDAR flight lines, and white circles indicate locations of field plots.

### Assessing the Biomass Resources in Interior Alaska Using Airborne LiDAR

The emergence of airborne laser scanning (LiDAR) as a highly effective and economical tool for measuring biomass over extensive, remote areas of forest has provided a means of acquiring much more accurate information on the spatial distribution and quantity of biomass surrounding these communities in rural Alaska than would be available using traditional field-based inventory methods. LiDAR provides direct, three-dimensional measurements of forest structure and the underlying terrain, and it can be used to develop highly accurate estimates of biomass (Means et al. 2000, Andersen et al. 2006, Zhao et al. 2009). Although the cost of acquiring a comprehensive ("wall-to-wall") coverage of airborne LiDAR data can often be prohibitive for a single agency or landowner, the cost of acquiring LiDAR in a sampling mode (i.e., single strips/swaths spaced several kilometers apart) is often affordable. Although the use of airborne laser scanning in a double sampling mode has been successfully demonstrated in previous studies (Næsset 2002, Parker and Evans 2004), the use of airborne LiDAR scanning as a sampling tool and the statistical properties of the resulting estimators have received relatively little attention until recently. Given the logistical difficulties and high costs associated with establishing field plots in rural Alaska (where the cost of establishing a single plot can exceed \$8,000), the use of LiDAR in a model-based sampling framework could potentially provide resource managers and energy developers with critically important and affordable information regarding the biomass resources available to support bioenergy systems in interior Alaska.

In this project, we investigated the combined use of a systematic sample of high-density LiDAR data and colocated representative field plot data to quantify aboveground tree biomass resources available to support the development and operation of bioenergy systems in two rural communities, Tok and Tanacross, located in the upper Tanana valley of interior Alaska. Tok is a small town with a human population of 1,393 (as of 2000) and is currently using a biomass-fired heating system for its primary and secondary schools (kindergarten to 12th grade, approximately 200 students total). Tanacross is a smaller native community (population 140 as of 2000) that is

currently using a small-scale wood-fired heating system for its school (kindergarten to 8th grade) and is also considered a good candidate for further development of bioenergy projects. We describe here the statistical properties of this LiDAR-based estimate of total biomass, including variance estimation via a resampling technique, and assess the influence of plot size, GPS plot location error, and diameter threshold on the variability of LiDAR-derived total biomass estimates in this area.

## Data and Methods

### Study Area

This study was conducted on a 201,226-ha area surrounding the communities of Tok and Tanacross in the upper Tanana valley of interior Alaska (see Figure 1). The forests in this area are characteristic of the boreal forests of interior Alaska, with lowland forests primarily composed of white spruce (*Picea glauca*) in well-drained areas and black spruce (*Picea mariana*) in poorly drained areas. Upland forests are predominantly composed of paper birch (*Betula papyrifera*) and quaking aspen (*Populus tremuloides*). Recently burned areas are composed of remnant spruce snags and, in some cases, a blanket of regenerating young aspen.

### LiDAR Data

High-density airborne laser scanning data were collected in June 2009 along single swaths (strips), regularly spaced approximately 2.5 km apart (see Figure 1). Because of flight safety considerations, the strips in the northern part of the area were oriented in a north-west-southeast direction, while strips in the southern part of the area were oriented in a southwest-northeast direction (so as to avoid flying perpendicular to steep mountainous slopes). Specifications for the LiDAR acquisition are shown in Table 1. The total cost of the LiDAR data acquisition was approximately \$61,000. Previous experience has shown that approximately 10% of this total is spent on fixed costs (mobilization, etc.), and the remainder of the cost is linearly related to flight time.

**Table 1. Specifications for light detection and ranging (LiDAR) strip sampling flights over the study area, located in the upper Tanana valley of interior Alaska.**

|                            |                                    |
|----------------------------|------------------------------------|
| LiDAR system               | Optech Gemini                      |
| Flying height              | 750 m                              |
| Pulse repetition frequency | 125 kHz                            |
| Scan angle                 | ± 9°                               |
| Scan rate                  | 103 Hz                             |
| Speed                      | 160 knots                          |
| Swath width                | 240 m                              |
| Point spacing              | ≤ 0.4 m cross-track and down-track |
| Beam divergence            | 0.3 mRad                           |

**Table 2. Summary of field plot data on forested plots (78 field plots)<sup>a</sup> established in the study area, located in the upper Tanana valley of interior Alaska.**

| Plot variable        | Minimum | Maximum | Mean  | Standard deviation |
|----------------------|---------|---------|-------|--------------------|
| Biomass (Mg/ha)      | 1.1     | 253.5   | 78.1  | 65.9               |
| Mean dbh (cm)        | 3.3     | 26.8    | 9.1   | 4.4                |
| Mean tree height (m) | 2.9     | 17.2    | 7.8   | 3.1                |
| Trees/ha             | 59      | 18,078  | 3,233 | 3,302              |

<sup>a</sup> One field plot was nonforested and was excluded from this summary table because all values were zero. This nonforested plot was included in the analysis, however, because it was a valid observation.

### Field Plot Data

In August and September 2009, 79 field plots were established within the coverage of the LiDAR strips (Figure 1). At each plot, each tree with dbh > 7.62 cm (3 in.) was measured within a circular 1/30 ha (1/12 ac.) plot with a fixed radius of 10.36 m (34 ft). Trees with dbh between 2.5 and 7.62 cm (1 and 3 in.) were measured within a smaller 1/422 ha (1/171 ac.) circular plot with a fixed radius of 2.74 m (9 ft). To maximize the efficiency of the field data collection, plots were collected in pairs spaced approximately 600 m apart. This spacing was chosen so as to increase efficiency of data collection while at the same time minimizing the correlation between plots. The following variables were recorded for each measured tree on the plot: (1) species, (2) dbh, (3) actual height, (4) estimated height (if top was broken), (5) uncompacted crown ratio, (6) percentage of rotten or missing cull, (7) crown class, and (8) crown radius (a selection within crown and dbh classes). A summary of the field plot data is shown in Table 2.

The location of the plot center was measured with submeter error using a survey-grade GPS+GLONASS receiver (Javad Maxor GGD), and data were postprocessed using a dedicated local base station (Andersen et al. 2009). Because most of the area is inaccessible without a helicopter, all field plots were located within reasonable hiking distance (1 km) of a road, trail, or river. Within these accessible areas that overlapped the coverage of the LiDAR strips, the location of plots were randomly located across different forest stand types (based on a polygon GIS layer provided by the Alaska Department of Natural Resources) and historical forest fire perimeters. The field plots were collected within the following classes (followed by the number of plots): white spruce (poletimber [7], sawtimber [5]), birch-poplar-white spruce (poletimber [2]), black spruce-white spruce (poletimber [1], regeneration [4], tall shrub [3], low shrub [2]), mixed birch-aspen (poletimber [3], regeneration [5]), mixed spruce-birch-aspen (poletimber [6]), and historical fire perimeters from 1986 (4), 1990 (8), 2001 (2), 2003 (7), and 2004 (12). Figure 2 shows the observed forest structure and corresponding vertical distribution of LiDAR returns at four plots in different

forest types. The total number of plots (79) was largely determined by budget limitations alone, and although we had planned to obtain at least 5 plots in each class, several classes were underrepresented because of accessibility issues that arose during the data collection. In addition, the forest stand type map did not cover the entire extent of the LiDAR data, so it was impossible to assess the proportion of the total study area within each class. However, in general it was felt that the collection of plots adequately represented the range of conditions present within the study area. Using individual tree-level data collected for all common species by researchers at the University of Alaska-Fairbanks (Yarie et al. 2007), we developed allometric equations to estimate total aboveground biomass for each tree species based on measurements of dbh, squared dbh, and height, using a stepwise regression model development procedure in the R statistical software environment (R Development Core Team 2008). These equations were applied to the trees measured on each plot to obtain a plot-level estimate of aboveground tree biomass.

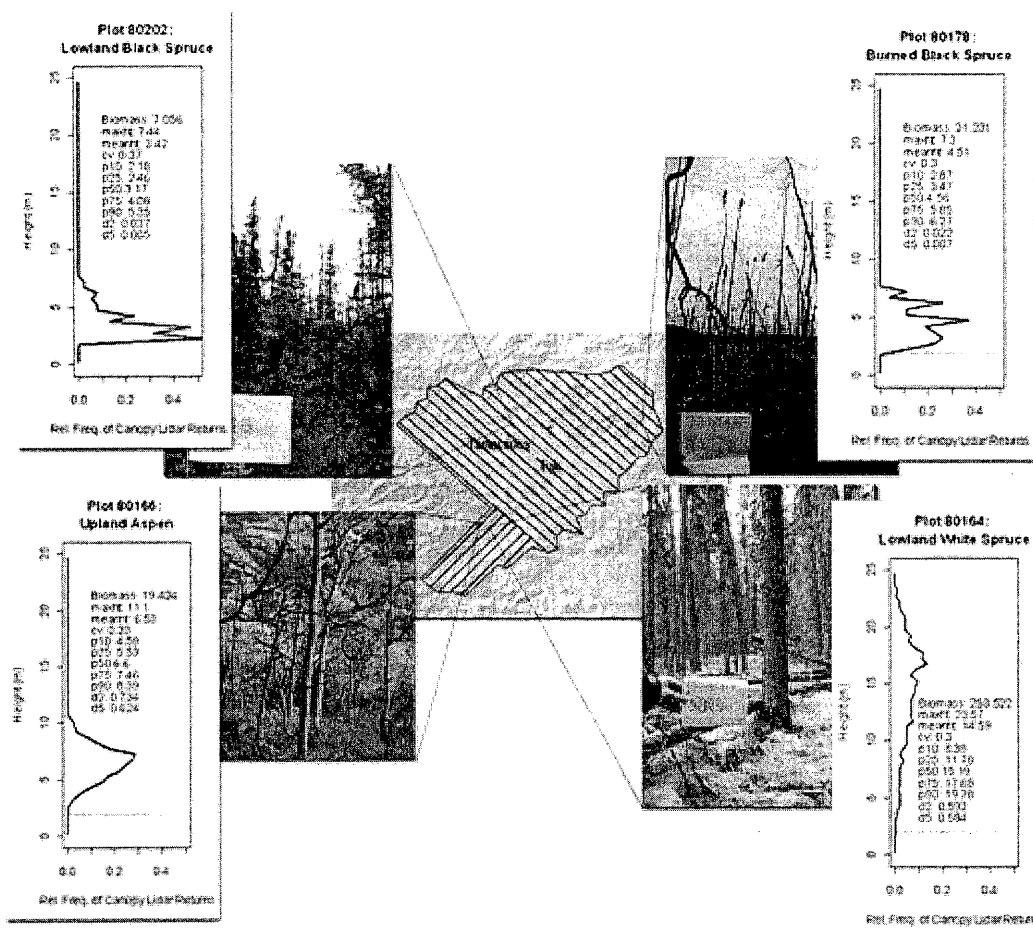
### LiDAR-Based Biomass Estimation

Previous studies have shown that strong allometric relationships exist between three-dimensional canopy structure and biomass; therefore, LiDAR-based forest structural metrics, including the vertical distribution of LiDAR returns and canopy cover, are typically highly correlated with biomass and stem wood volume (Means et al. 2000, Næsset 2002, Andersen et al. 2006, Li et al. 2008). Using regression analysis, predictive models can be developed to estimate biomass using this set of LiDAR-derived structural metrics, and biomass can then be estimated over the coverage of each LiDAR strip sample (see Figure 1).

Because of the inaccessibility and very limited transportation infrastructure throughout much of the study area—a very common constraint in interior Alaska—we did not collect a probability sample over the entire area. In addition, the number of plots was necessarily limited because of time and cost constraints. Therefore, we use a model-based, as opposed to a model-assisted, approach to estimation of biomass within the LiDAR coverage area. According to Schreuder et al. (1993), a model-based approach assumes that the population values  $y_i$  ( $i = 1, \dots, N$ ) are seen as realizations of random variables,  $Y_i$ . In the context of this analysis, the so-called superpopulation is distributed according to the model,

$$\sqrt{Y_i} = \alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + e_i,$$

where  $\text{Variance}_m(e_i) = \sigma^2$ , where the subscript  $m$  indicates conditioning on the underlying superpopulation model,  $Y$ , denotes the total aboveground tree biomass for a given grid cell area,  $X$  indicates the various LiDAR structural parameters (maximum height, canopy cover, etc.) that are correlated with biomass, and  $\alpha$  and  $\beta$  indicate the coefficients of this superpopulation model. The square-root transform has been applied in the past to linearize the relationship between LiDAR structural metrics and aboveground biomass (Andersen and Breidenbach 2007). As Schreuder and others mention, sample selection in a model-based context is usually designed so as to increase the precision of the estimates of  $\alpha$  and  $\beta$ , which usually involves collecting a representative sample that ensures a large spread over possible  $X_i$  values. In our study, the location of the sample plots were designed to provide data across the full range of forest structural and cover types that are representative of this interior boreal forest, using a forest stand type polygon GIS layer (species and timber size class) provided by the Alaska Department of Natural



**Figure 2.** Photographs show several representative forest types in the study area, located in the upper Tanana valley of interior Alaska: lowland black spruce, burned black spruce, lowland white spruce, and upland aspen. Insets show a graphical representation of the height distribution of canopy-level (> 2 m) LiDAR returns within the plot, the estimated biomass, and the LiDAR-based structural metrics for each plot.

Resources and historical fire perimeters. In the context of this study, the population elements were  $18 \times 18$ -m grid cells  $i$  with area  $324 \text{ m}^2$ , or  $0.0324$  ( $1/31$ ) ha, covering all forestland within the study area (note that this grid cell size corresponds to the size of the field plots [ $1/30$  ha]), and although the plots and grid cells are different shapes [circles and squares, respectively], they are the same grain size). To develop the model relating LiDAR metrics to aboveground tree biomass at each grid cell, a pool of LiDAR-derived structural metrics was generated from “canopy-level” LiDAR returns, defined as all LiDAR returns above a 2-m height threshold. These height-based metrics include maximum height, mean height, coefficient of variation of heights, and several height percentiles (10th, 25th, 50th, 75th, and 90th percentiles). In addition, two other metrics were used that represented the density of the entire canopy (percentage of first returns above 2 m height) and the density of the overstory canopy (percentage of first returns above 5 m height). A stepwise regression variable selection routine was implemented in the R statistical package to select the predictive model for biomass. As Miller (1984) states, backtransforming the predictive regression model leads to the following model.

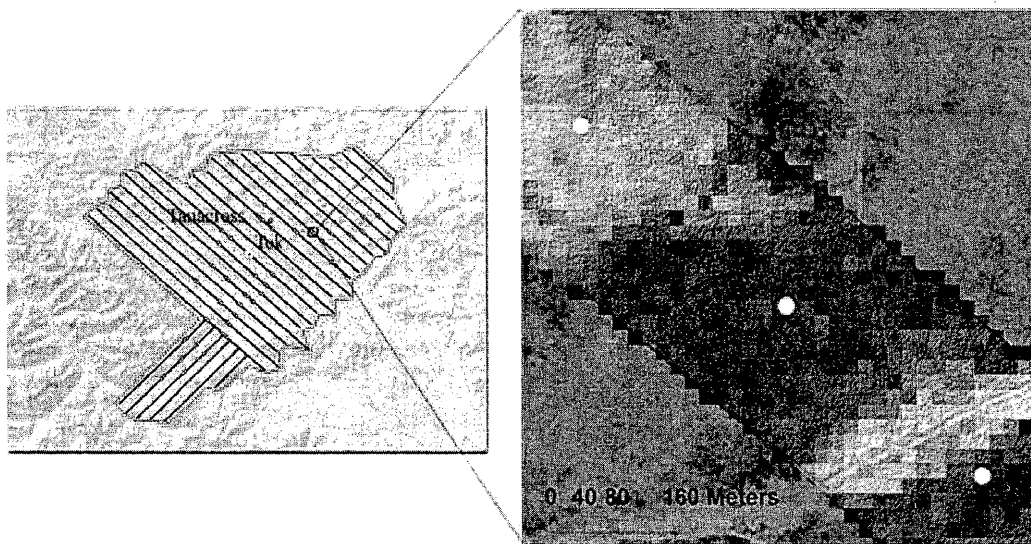
$$Y_i = (\alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + e_i)^2.$$

Therefore, after backtransformation,  $Y$  is distributed as the second moment of a normal random variable with mean  $(\alpha + \beta_1 X_{1i} + \dots$

$+ \beta_p X_{pi})$  and variance  $\sigma^2$ , with expected value  $E(Y) = (\alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi}) + \sigma^2$ . Therefore, a correction factor  $\hat{\sigma}^2$  should be added to each (backtransformed) prediction  $(\hat{\alpha} + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_p X_{pi})$ , to reduce bias. This model (backtransformed and adjusted for bias) was applied to the LiDAR metrics within each 18-m grid cell within the LiDAR coverage, resulting in a map of predicted biomass within each LiDAR swath (see Figure 3). It should be noted that there is an additional source of variability—model selection uncertainty—that should be accounted for when quantifying the precision of total biomass estimates (Buckland et al. 1997).

### Statistical Properties of LiDAR-Based Total Biomass Estimator

The sampling design is a single-stage cluster sample with a model-based estimate of biomass within each cluster, where the model was developed from subsample of plots collected across a representative range of forest conditions. In this design, the LiDAR strips/swaths can be seen as clusters distributed systematically over the entire study area (with a regular 2.5-km interval between strips). Coburn et al (2009) provide an overview of the mathematical principles behind the strip sampling approach in the context of locating unexploded ordnance but do not address strips of differing length (i.e., clusters of varying size). A secondary sample of field plots was collected within the accessible areas covered by the LiDAR strips, with data



**Figure 3.** LiDAR-based estimation of biomass within a single strip, located in the upper Tanana valley of interior Alaska. Left panel shows LiDAR strip samples as black lines; red outline is area of interest for biomass study. Right panel shows LiDAR-predicted biomass values within the LiDAR swath (black, low biomass; white, high biomass [ $\sim 200$  M/ha]), overlaid on SPOT color-infrared image (2.5-m resolution). Field plots are shown as white circles in both images.

collected so as to obtain an adequate sample size ( $n > 5$ ) within each forest condition class (forest stand type and historical fire perimeter). The total biomass for all forested area within the study area,  $\hat{Y}_{\text{total}}$ , was estimated by applying the ratio-to-size estimator as described by Cochran (1977) as

$$\hat{Y}_{\text{total}} = M_0 \frac{\sum^n \hat{Y}_k}{\sum^n M_k},$$

where  $n$  is the number of LiDAR strips,  $\hat{Y}_k$  is the estimated total biomass for the  $k$ th LiDAR strip,  $M_k$  is the total number of elements in the  $k$ th LiDAR strip, and  $M_0$  is the total number of elements in the population (total number of forested grid cells within the study area); also, LiDAR strips were randomly distributed and selected with equal probabilities (although the LiDAR strips were actually systematically [regularly] distributed in our case, for the purposes of variance calculations we assume a random distribution). In this study, the total size of the study area was 201,227 ha (see Figure 1), and the total forested area was estimated to be 163,913 ha. Determination of forestland was based on the LANDFIRE classification product for this area, where forest was defined to be areas with canopy cover  $> 10\%$  (Rollins and Frame 2006).

Also following Cochran, the estimated SD of the single-stage ratio-to-size estimate of total biomass is given by the equation,

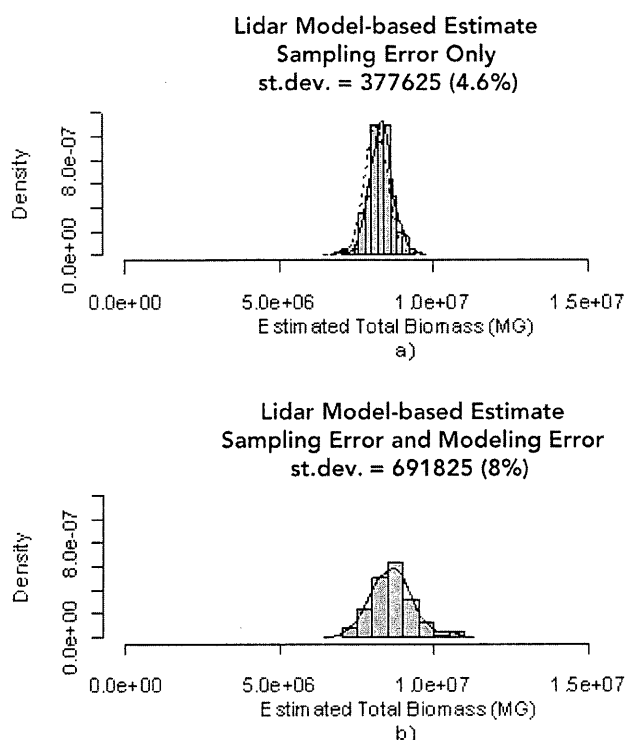
$$\widehat{\text{SD}}(\hat{Y}_{\text{total}}) = \sqrt{\frac{N^2(1-f)}{n} \frac{\sum^n M_k^2 (\hat{Y}_k - \hat{\bar{Y}})^2}{n-1}}$$

where  $\hat{Y}_k$  is estimated mean biomass for the  $k$ th LiDAR strip,  $N$  is the total number of strips covering the entire study area,  $\hat{\bar{Y}}$  is the estimated mean biomass across LiDAR strips, and  $f$  is the sampling fraction for the cluster sample ( $n/N$ ).

#### Assessing the Combined Influence of Sampling and Modeling Error on Biomass Estimates

A resampling (bootstrapping) approach was developed to quantify the variability of the LiDAR-based biomass estimates, which incorporated the variability due to both sampling error and LiDAR model error (Efron and Tibshirani 1994). Because traditional bootstrapping typically assumes an infinite population, which is not the case here, we used the modified version of the bootstrap, the without-replacement bootstrap (BWO), developed by McCarthy and Snowdon (1985) and extended by Booth et al. (1994) to account for sampling from a finite population. In this approach, if  $f$  is the single-stage sampling fraction, then the each cluster is replicated  $1/f$  times to obtain a pseudopopulation of (approximately) size  $n/f = N$ . As McCarthy and Snowdon describe, if samples of size  $n$  are drawn without replacement from this pseudopopulation, then a sample is produced that might have been drawn from the original (finite) population (although, as is usual in a bootstrapping context, a single transect can be selected multiple times).

To incorporate the additional variability due to LiDAR-based modeling of biomass, we use a modified form of the bootstrapping approach proposed by Sitter (1997) for estimating the variance of a two-phase regression estimator. In the approach described by Sitter, the data are split into the first- and second-phase components, then the usual bootstrap (i.e., sampling with replacement) is applied each component separately. The estimator is calculated from this bootstrap sample, and this procedure is repeated a large number of times. Although the design in our case is not a true two-phase sample within the LiDAR strips, because the field plots were not a probability sample, this resampling approach can still be used to estimate variability due to the field plot selection and regression modeling. To include variability associated with the model selection process, a stepwise variable selection procedure was applied at each iteration (i.e., selected predictor variables can vary) in the spirit of Buckland et al. (1997). The variability of the total biomass estimate is then characterized by the variance (or SD) of the collection of bootstrap



**Figure 4.** Bootstrapped estimation of variance for total aboveground tree biomass within the study area, located in the upper Tanana valley of interior Alaska. (a) Sampling error only (theoretical sampling distribution of ratio-to-size estimator [from Cochran 1977] shown by dashed line). (b) Sampling error and modeling error, with subsample stratified by forest type and fire history classes. The number of bootstrapping replications in both cases was 100.

estimates. In our case, because we had LiDAR measurements at each grid cell (i.e., a total census) within the LiDAR strips, there was no sampling error within the clusters/strips; therefore, we applied the bootstrap only to the field plot sample, and we held the LiDAR strip data fixed. In addition, because we randomly selected field plots within forest type/fire history strata, we applied the bootstrap independently within each stratum (to ensure a representative sample for model development). Using this approach, we were able to estimate the additional variability due to the use of a LiDAR-based regression model to estimate biomass within each cluster.

#### Influence of Plot Location Error and Diameter Threshold on Variability of Total Biomass Estimates

This resampling technique described previously can be expanded to assess additional sources of error, such as the influence of GPS plot location error, differing plot sizes, and differing diameter thresholds on the variability of total aboveground biomass estimates. Determination of appropriate plot size, types of GPS receivers to use, and diameter thresholds are important considerations in designing an inventory system in interior Alaska. In this study, we were particularly interested in assessing the variability of LiDAR-derived biomass estimates obtained using the standard Forest Inventory and Analysis (FIA) subplot design (single 1/59 ha subplot, 1/741 ha microplot, 12.7-cm-diameter threshold) in comparison with the research plot design (1/30 ha, 1/423 ha microplot, 7.6-cm-diameter threshold), across a range of plot location errors (Bechtold and Scott 2005). To carry out this comparison, the tree data asso-

ciated with the standard FIA subplot and microplot areas was extracted from the Tok plots and used to generate a plot-level biomass estimate. In addition, a random error was applied to the plot coordinate, based on a Gaussian distribution centered on the true coordinate with a specified SD (root mean square error). This simulated plot location error (root mean square error) was increased by 2.5-m increments from 2.5 to 20 m, with 100 bootstrapped iterations at each level of error. The variability in the total biomass estimate for each plot size was then calculated as the SD of these bootstrapped estimates at each level of GPS error. These estimates of variability incorporate error due to (1) sampling error due to using LiDAR strips, (2) regression modeling (model fit and variable selection process), (3) plot size (1/30 ha versus 1/59 ha), (4) diameter threshold (7.5 versus 12.7 cm), and (5) GPS plot location error (2.5–20 m).

#### Results

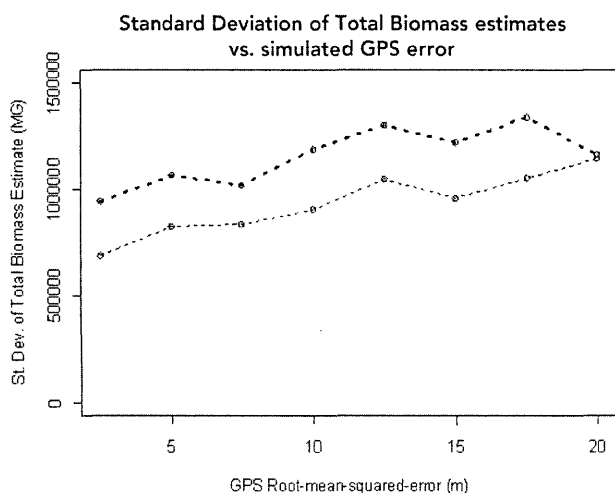
The estimate of total aboveground tree biomass in the forested area ( $\hat{Y}_{\text{total}}$ ) within the study area, based on the ratio-to-size estimator, is 8,138,278 Mg. In this study, there were 27 sampled LiDAR strips ( $n$ ), the total number of potential strips ( $N$ ) was 212, and the sampling fraction ( $f$ ) was 0.127, leading to an estimated SD of 361,309 Mg (4.4%).

The BWO approach developed by McCarthy and Snowdon (1985) was used to quantify the variability of the total biomass estimate  $\hat{Y}_{\text{total}}$ , and the SD of the bootstrap distribution (377,626 Mg) was very close to the result obtained using the analytical formulation provided by Cochran (1977) (361,309 Mg) described previously (see Figure 4a). (It should be noted that McCarthy and Snowdon suggest applying a additional multiplier,  $[n - 1/k]/[n - 1]$ , to obtain a slightly more accurate estimate of the bootstrap variance. However, in our case, this multiplier did not improve the correspondence between the analytical and BWO estimate and was therefore not applied.) As shown by Figure 4b, introducing the variability due to modeling increased the bootstrap standard error of the total biomass estimate significantly, from 377,625 Mg (4.6%) to 691,825 Mg (8%).

As expected, Figure 5 indicates that the precision of the total biomass estimates decreased significantly when the smaller 1/59 ha plots (comparable to FIA subplots) were used. This difference in precision between the 1/30 ha Tok research plots and simulated FIA subplots was significant (approximately 20%) and fairly consistent across the varying levels of simulated GPS error, until the standard deviations converged at 20 m. The precision of the estimates also steadily decreased with increasing levels of GPS error for each plot size.

#### Discussion and Conclusion

The results of this study indicate that LiDAR sampling can be used to produce estimates of total aboveground tree biomass in areas surrounding rural communities of interior Alaska at a reasonable cost and acceptable level of precision (8%). These results suggest that use of airborne LiDAR, collected in a sampling mode, could have a role to play in supporting bioenergy development in interior Alaska. Using a bootstrapping approach to quantify the precision of model-based estimates, we have shown that sampling error (due to using LiDAR collected in single swaths spaced 2.5 km apart instead of wall-to-wall LiDAR) represents the largest component of the error budget (4.6%), but variability due to model selection contributes a significant additional source of error (3.4%).



**Figure 5.** Bootstrapped SD of total biomass estimates versus simulated GPS plot location error for two different plot types, for study area in the upper Tanana valley of interior Alaska. Bold dashed line indicates standard FIA subplot (1/59 ha subplot, 1/741 ha microplot, 12.7-cm-diameter threshold). Lighter dashed line indicates Tok research plot (1/30 ha, 1/423 ha microplot, 7.6-cm-diameter threshold).

Even in cases where much of the area is inaccessible and the sample size for field plot data is necessarily limited, a stratified approach to field plot sampling within accessible areas can be used to ensure that a representative sample of field plots is obtained to develop the LiDAR biomass models. Although the use of a model-based LiDAR sampling approach introduces variability due to sampling error and modeling error, we have shown that this approach can be used to generate estimates of total aboveground tree biomass over a very large area (201,226 ha) at a reasonable cost (27 LiDAR strips + 79 field plots) and with acceptable precision (relative standard error of 8%; see Figure 4). However, it should be noted that a model-assisted approach—based on a true probability sample of field plots over the entire LiDAR coverage area—would be advisable if design-unbiased estimates are required (Särndal et al. 1992).

For this study, several modifications were made to the standard FIA subplot measurement protocol, including increased plot (and microplot) size and a reduced diameter threshold. The results obtained from applying the resampling-based variance estimator to LiDAR-based estimates obtained using the two alternative plot designs, one arguably optimized for double-sampling with LiDAR and the other matching the standard FIA subplot design, indicates that using larger plots with a lower diameter threshold could substantially increase the precision of the total biomass estimates in the boreal forest types observed in this study. The results also indicate that plot location error—usually resulting from the use of less-sophisticated GPS receivers on the plots—can have a sizable effect on the precision of LiDAR-based total biomass estimates. Using a larger plot captures more of the variability within a given forest area, and lowering the diameter threshold will reduce the sampling error associated with measuring only small trees on a smaller microplot and will increase the correspondence between the forest structure measured in the field and by the LiDAR within the larger plot area. Previous studies have suggested varying the plot size based on the different canopy structures and densities, but this is often not feasi-

ble in an operational forest inventory with a standard plot protocol (Gobakken and Næsset 2008). Although detailed analysis of each type of error was outside the scope of this report, this approach provided a comparison of the trade-offs in using a standard FIA plot versus a larger plot with a lower diameter threshold; it also indicated the decreased accuracy that can be expected if less sophisticated GPS equipment and techniques are used to establish plot positions. Although these results indicate that increasing the plot size and decreasing the diameter limit can lead to more precise estimates of biomass using LiDAR, there are many other factors to consider when determining the appropriate plot size for a forest inventory program, including consistency, time spent on plot, and statistical efficiency for estimation of many other inventory parameters. The results presented here are intended to show that there are advantages to using a larger plot and lower diameter threshold for LiDAR model-based estimation of biomass in these boreal forest types, an important consideration for planning future projects.

## Literature Cited

- ANDERSEN, H.-E., S.E. REUTEBUCH, AND R.J. MCGAUGHEY. 2006. Active remote sensing. In *Computer Applications in Sustainable Forest Management*, Shao, G., and K. Reynolds (eds.). Springer-Verlag, Dordrecht. 43–66 p.
- ANDERSEN, H.-E., AND J. BREIDENBACH. 2007. Statistical properties of mean stand biomass estimators in a LiDAR-based double sampling forest survey design. *Int. Arch. Photogramm. Remote Sens.* 36(3/WS2):8–13.
- ANDERSEN, H.-E., T. CLARKIN, K. WINTERBERGER, AND J. STRUNK. 2009. An accuracy assessment of positions obtained using survey- and recreational-grade global positioning system receivers across a range of forest conditions within the Tanana valley of interior Alaska. *West. J. Appl. For.* 24(3):128–136.
- BECHTOLD, W., AND C. SCOTT. 2005. The Forest Inventory and Analysis plot design. In *The Enhanced Forest Inventory and Analysis Program: National design and estimation procedures*, Bechtold, W., and P. Patterson (eds.). US For. Serv. Gen. Tech. Rep. SRS-80. US For. Serv. South. Res. Stn., Asheville, NC.
- BOOTH, J., R. BUTLER, AND P. HALL. 1994. Bootstrap methods for finite populations. *J. Am. Statist. Assoc.* 89(428):1282–1289.
- BUCKLAND, S., K. BURNHAM, AND N. AUGUSTIN. 1997. Model selection: An integral part of inference. *Biometrics* 53:603–618.
- COBURN, T., S. MCKENNA, AND H. SAITO. 2009. Strip transect sampling to estimate object abundance in homogeneous and non-homogeneous Poisson fields: A simulation study of the effects of changing transect width and number. *Math. Geosci.* 41(1):51–70.
- COCHRAN, W. 1977. *Sampling techniques*. Wiley and Sons, New York, NY. 428 p.
- CRIMP, P., S. COLT, AND M. FOSTER. 2008. Renewable power in rural Alaska: Improved opportunities for economic development. *Proceedings of the 2007 Arctic Energy Summit, Anchorage, Alaska*, Institute of the North, Anchorage, AK. 15 p.
- EFRON, B., AND R. TIBSHIRANI. 1994. *An introduction to the bootstrap*. Chapman and Hall/CRC, Boca Raton, FL.
- GOBAKKEN, T., AND E. NÆSSET. 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can. J. For. Res.* 38:1095–1109.
- HALEY, S., AND B. SAYLOR. 2007. *Effects of rising utility costs on household budgets, 2000–2006*. Institute of Social and Economic Research, Anchorage, AK. Available online at [www.iser.uaa.alaska.edu/publications/risingutilitycosts\\_final.pdf](http://www.iser.uaa.alaska.edu/publications/risingutilitycosts_final.pdf); last accessed Apr. 26, 2011.
- LI, Y., H.-E. ANDERSEN, AND R. MCGAUGHEY. 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West. J. Appl. For.* 23(4):223–231.
- MCCARTHY, P., AND C. SNOWDON. 1985. The bootstrap and finite population sampling. *Vital and Health Statistics, Series 2, No. 95*, DHHS Pub. No. (PHS) 85-1369. Public Health Service, US Government Printing Office, Washington DC.
- MEANS, J., S. ACKER, B. FITT, M. RENSLOW, AND L. EMERSON. 2000. Predicting forest stand characteristics with airborne scanning LiDAR. *Photogramm. Eng. Remote Sens.* 66(11):1367–1371.
- MILLER, D. 1984. Reducing transformation bias in curve fitting. *Am. Statist.* 38(2):124–126.
- PARKER, R., AND D. EVANS. 2004. An application of LiDAR in a double-sample forest inventory. *West. J. Appl. For.* 19(2):95–101.

- NÆSSET, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* 80:88–99.
- R DEVELOPMENT CORE TEAM. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- ROLLINS, M., AND C. FRAME, EDS. 2006. *The LANDFIRE Prototype Project: Nationally consistent and locally relevant geospatial data for wildland fire management*. US For. Serv. Gen. Tech. Rep. RMRS-GTR-175. US For. Serv., Rocky Mount. Res. Stn., Fort Collins, CO.
- SÄRNDAL, C.-E., B. SWENSSON, AND J. WRETMAN. 1992. *Model-assisted survey sampling*. Springer-Verlag, New York, NY.
- SCHREUDER, H., T. GREGOIRE, AND G. WOOD. 1993. *Sampling methods for multiresource forest inventory*. Wiley, New York, NY.
- SITTER, R. 1997. Variance estimation for the regression estimator in two-phase sampling. *J. Am. Statist. Assoc.* 92(438):780–787.
- YARIE, J., E. KANE, AND M. MACK. 2007. *Aboveground biomass equations for the trees of interior Alaska*. Agricultural and Forestry Experiment Station Bulletin: 115. University of Alaska-Fairbanks. 15 p.
- ZHAO, K., S. POPESCU, AND R. NELSON. 2009. Lidar remote sensing of biomass: a scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* 113(1):182–196.