Model-Assisted Forest Yield Estimation with Light Detection and Ranging

Jacob L. Strunk, Stephen E. Reutebuch, Hans-Erik Andersen, Peter J. Gould, and Robert J. McGaughey

ABSTRACT

Previous studies have demonstrated that light detection and ranging (LiDAR)-derived variables can be used to model forest yield variables, such as biomass, volume, and number of stems. However, the next step is underrepresented in the literature: estimation of forest yield with appropriate confidence intervals. It is of great importance that the procedures required for conducting forest inventory with LiDAR and the estimation precision of such procedures are sufficiently documented to enable their evaluation and implementation by land managers. In this study, we demonstrated the regression estimator, a model-assisted estimator (approximately design-unbiased), using LiDAR-derived variables for estimation of total forest yield. The LiDAR-derived variables are statistics associated with vegetation height and cover. The estimation procedure requires complete coverage of the forest with LiDAR and a random sample of precisely georeferenced field measurement plots. Regression estimation relies on sample-based ordinary least squares (OLS) regression models relating forest yield and LiDAR-derived variables. Estimation was performed using the OLS models and LiDAR-derived variables for the entire population. Regression estimates of basal area, volume, stand density, and biomass were much more precise than simple random sampling estimates (design effects were 0.25, 0.24, 0.44, and 0.27, respectively).

Keywords: forest inventory, design-based, LiDAR, model-assisted, regression estimation

The feasibility of modeling forest yield variables, such as basal area, volume, and biomass, with light detection and ranging (LiDAR)-derived variables (or LiDAR metrics) has been demonstrated in a variety of studies (Næsset 1997, Means et al. 2000b, Andersen et al. 2005). These and other studies have found that LiDAR-derived height metrics are highly correlated with forest variables; the coefficient of determination (R^2) in studies modeling structural forest variables with LiDAR often exceeds 0.9 (Lefsky et al. 1999, Riaño et al. 2004, Næsset et al. 2005). Although models developed to relate forest inventory and LiDAR-derived variables can play an important role in forest inventory with LiDAR, it is also important to make inference about the population. The precision of population estimates of forest yield variables depend on the (1) size of the population, (2) the size of the sample, and (3) the precision of the model used.

Estimates of total forest yield and reliable confidence intervals for total estimates are required inferences in forest inventory. In most traditional forest inventories (e.g., a stratified sample with plots arranged on systematic grids within strata), inference is made in the design-based paradigm. Design-based inference, in contrast to model-based inference, is an approach that allows inference to be made about the population from a random sample of data without need to make assumptions about the distribution of the population (Gregoire 1998). There are few examples of design-based estimation in studies on estimating forest yield variables with LiDAR-derived auxiliary variables. Auxiliary variables are variables that are used to aid in estimation but are not necessarily of interest by themselves. Studies that discuss design-based estimation include the studies by Parker and Evans (2004, 2007), Corona and Fattorini (2008), and Gregoire et al. (2011). In these studies, model development was extended with model-assisted (design-based paradigm) estimators to enable inference about the population.

The studies by Parker and Evans (2004, 2007) and Gregoire et al. (2011) are examples of two-phase or double-sampling estimation. In this context, ordinary least squares (OLS) regression is used to relate predictors, such as remote sensing variables, to response variables, such as forest variables. The modeled relationship is then used to adjust the sample mean. The adjustment is based on the difference between the means of the explanatory variables for the sample relative to the mean for the entirety of the auxiliary data. The ratio estimators used in Corona and Fattorini (2008) also uses a model and auxiliary data to adjust the sample mean. This example of the ratio estimator is appropriate when there is complete coverage of the population with the auxiliary data. The linear relationship between the response and predictor should be positive and approximately intersect zero. The ratio estimator of the total uses a ratio model to adjust the sample estimate of the total. The ratio estimator has the advantage of being a simple estimator and has been shown to work well in a variety of scenarios (Gregoire and Valentine 2004, Särndal

Manuscript received December 29, 2010; accepted December 15, 2011. http://dx.doi.org/10.5849/wjaf.10-043.

Jacob L. Strunk (jacob.strunk@oregonstate.edu), College of Forestry, Oregon State University, 204 Peavy Hall, Corvallis, OR 97330. Stephen E. Reutebuch, Hans-Erik Andersen, and Robert J. McGaughey, US Forest Service, Pacific Northwest Research Station, Seattle, WA 98195-2100. Peter J. Gould, US Forest Service, Pacific Northwest Research Station, Olympia, WA 98512. *We appreciate the work of Dr. Jeff Foster from the Environmental Division, Public Works, Joint Base Lewis-McChord, providing data and resources that made this project possible and for providing feedback on an early draft. We would like to thank Dr. David Briggs, Dr. Peter Schiess, and Dr. Temesgen Hailemariam for revisions and technical recommendations on drafts. Finally, we thank three excellent reviewers whose thorough critiques resulted in a much improved final product.*

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^2): 1 m^2 = 10.8 ft^2; cubic meters (m^3): 1 m^3 = 35.3 ft^3; kilometers (km): 1 km = 0.6 mi; hectares (ha): 1 ha = 2.47 ac; milligram (mg): 1 mg = 0.015 grain; megagrams (Mg): 1 Mg = 2204.62 lbs.

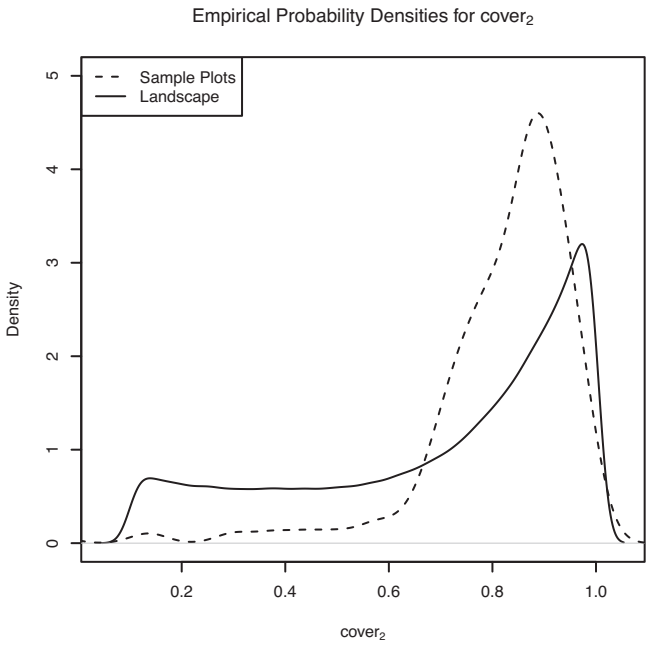

Figure 1. Empirical densities for $cover_2$ for the entire forest and for the sample plots alone. $Cover_2$ is the ratio of the number of first returns above 2 m to the total number of first returns.

et al. 2010). However, this estimator is generally design-biased and can incorporate only a single auxiliary variable.

To facilitate evaluation and adoption of advanced forest inventory techniques with LiDAR by a wide audience, it behooves the LiDAR community to document and vet all of the steps required for implementation in peer-reviewed literature. Currently, the process and performance (precision) of estimation with LiDAR-assisted techniques are not sufficiently documented and demonstrated to encourage general use. The objectives for this study are to demonstrate the steps required for model-assisted estimation of forest yield variables when LiDAR is obtained over the entire forest and to provide an indication of the precision that can be achieved.

Methods

Study Site

The forests considered in this study are located on the Joint Base Lewis-McChord in western Washington State (Figure 1). The Lewis portion of the reservation (formerly the Fort-Lewis Military Reservation) spans 35,052 ha, of which 23,308 ha are forested (Figure 1). The predominant forest type on the installation is mixed conifer dominated by Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) forest with western redcedar (*Thuja plicata* Donn ex D. Don), western hemlock (*Tsuga heterophylla* [*Raf.*] Sarg.), red alder (*Alnus ruba* Bong.), bigleaf maple (*Acer macrophyllum* Pursh), black cottonwood (*Populus balsamifera* L.), and other minor deciduous and coniferous species. There are also scattered Oregon white oak (*Quercus garryana* Douglas ex Hook) and ponderosa pine (*Pinus ponderosa* C. Lawson) woodland and grassland areas interspersed with the coniferous forests. Our target population is managed forests on the military base. We were not able to access this entire population because of censoring of the sample data, but we were able to mitigate this censoring by restricting our target population. We elaborate further on the nature of the censoring of our sample when we discuss our sample, and we discuss our approach to mitigation when we describe estimators.

Table 1. Summary of forest yield variables measured on field plots.

Variable[a]	Mean	Minimum	Maximum	SD
ba	35	1	130	21
vol	457	9	2,376	332
stems	220	12	865	155
bm	294.8	7.5	1,508.9	204.0

[a] Forest yield variables are basal area (ba) (m^2/ha), total stem volume, including top and stump (vol) (m^3/ha), number of stems/ha (stems), and total aboveground biomass (bm) (Mg/ha).

Fort Lewis Forest Inventory Plots

Field measurements were collected at 128 forested continuous forest inventory (CFI) plots established and periodically remeasured by Fort Lewis forestry staff. The fixed-area, 0.081-ha (16.05-m radius) circular plots are located on a regular 1.3 × 1.3-km grid embedded in the forested portions of the Fort Lewis Military Installation. Plots were remeasured every 5 years, on average, from the time that the inventory was initiated in 1973 until the most recent measurement in 2005. Subsequent remeasurements are scheduled for 10-year intervals. Plot measurements for all trees with dbh greater than 20.3 cm include dbh, species, and condition class. Tree heights were measured on a subset of trees, on average 2.5 per plot. Plot positions were accurately georeferenced using dual frequency survey-grade GPS units followed by differential postprocessing of raw GPS data (Clarkin 2007, Andersen et al. 2009).

Missing tree heights were estimated with allometric height-diameter equations. A site-specific equation with random effects for each plot (Temesgen et al. 2008) was used for Douglas-fir. Equation 1 was fitted to trees with measured heights, and then the fitted parameters (included plot-specific random coefficients) were used to estimate the remaining heights.

$$ht_{ij} = 4.5 + \exp\left((b_1 + e_{1,i}) + \frac{(b_2 + e_{2,i})}{dbh_{ij} + 1}\right) + e_{0,j} \qquad (1)$$

where ht_{ij} and dbh_{ij} are height and diameter of jth tree on ith plot; b_1 and b_2 are fixed effects; $e_{1,i}$ and $e_{2,i}$ are random effects for plot i; and $e_{0,j} \sim N(0, \sigma^2)$, $e_{1,i} \sim N(0, \sigma_1^2)$, and $e_{2,i} \sim N(0, \sigma_2^2)$, where $\sim$ indicates distributed as.

Tree heights for other species were estimated using regional coefficients for Equation 1 provided for the Pacific Northwest in the Forest Vegetation Simulator documentation (Keyser 2010). Four forest yield variables (Table 1) were calculated on plots for this study: basal area (ba) (m^2/ha), total volume (vol) (m^3/ha), stand density (stems) (trees/ha), and biomass (bm) (mg/ha). ba and stand density were calculated directly from tree measurements. Total stem volume per tree was predicted with the USDA National Volume Estimator Library add-in for Microsoft Excel for Windows (US Forest Service 2008) (Table 1) and aggregated to the plot level. Biomass for individual trees was predicted using national-scale biomass equations (Jenkins et al. 2003).

LiDAR Acquisition, Variables, and Processing

Airborne discrete-return, near-infrared LiDAR data were collected over our study site between Sept. 19 and Sept. 21, 2005, prior to loss of leaves on deciduous plants (leaf on). A fixed wing aircraft was used for the acquisition. The aircraft flew over the study site at a height of approximately 1,000 m. The LiDAR sensor used was an Optech ALTM 3100. The sensor can detect up to four returns per pulse. The scan angle was set to ±14° and the beam divergence to

Table 2. Summary of several light detection and ranging-derived variables for our field plots.

Variable[a]	Mean	Minimum	Maximum	SD
ht_{40}	20.06	1.416	41.39	9.21
ht_{60}	24.63	1.552	49.12	9.66
ht_{80}	29.35	1.62	54.12	9.58
$Cover_2$	0.74	0.00	0.94	0.19
$Return_1$	3,392	1,671	6,701	940.45

[a] ht_{40} indicates 40th percentile height of light detection and ranging point data falling within a given pixel or plot. $Cover_2$ is the proportion of LiDAR point data within a pixel or plot with height attributes smaller than the height (m) indicated by the subscript. $Return_1$ is the number of first returns that intersected a plot.

Table 3. Summary of several light detection and ranging-derived variables for the target population.

Variable[a]	Mean	Minimum	Maximum	SD
ht_{40}	16.68	1.00	54.63	10.22
ht_{60}	20.55	1.01	55.44	11.15
ht_{80}	24.48	1.02	58.15	12.00
$Cover_2$	0.51	0.00	1.00	0.29
$Return_1$	3,537	0	12,030	1,239.63

[a] ht_{40} indicates 40th percentile height of light detection and ranging point data falling within a given pixel or plot. $Cover_2$ is the proportion of LiDAR point data within a pixel or plot with height attributes smaller than the height indicated by the subscript. $Return_1$ is the number of first returns that intersected a 28.45 × 28.45-m pixel.

0.3 mrad. The scan pulse rate was 71 kHz. The nominal pulse density was 4 pulses/m^2. A discussion of how to select appropriate LiDAR acquisition parameters may be found in Reutebuch et al. (2005).

Summary statistics computed from LiDAR point data (LiDAR metrics) were calculated in two instances for this study. In the first instance, LiDAR metrics were calculated for areas corresponding precisely to the locations and bounds of each circular 0.081 ha CFI plot (Table 2). The set of metrics calculated for plots was used to model the relationship between LiDAR-derived variables and forest inventory variables. In the second instance, LiDAR metrics were calculated for a grid of 28.45 × 28.45-m (0.081 ha) square pixels covering the entire forest (Table 3). LiDAR processing was carried out using FUSION (McGaughey 2009). Two types of LiDAR-derived variables were used that are referred to subsequently as height and cover variables. Height variables (e.g., ht_{10} and ht_{95}) are percentiles calculated from the height attributes of LiDAR point data, with the subscript denoting the percentile. For example, ht_{10} corresponds to the 10th percentile height of LiDAR point data falling within a given pixel or a plot. Cover variables (e.g., $cover_1$ and $cover_2$), indicate the proportion of LiDAR point data within a pixel or plot with height attributes smaller than the height (m) indicated by the subscript.

Model Development

OLS regression was used to relate forest yield variables to LiDAR-derived variables with R (R Development Core Team 2008). The LiDAR-derived variables evaluated as predictors fall into two general groups that describe either vegetation height (e.g., ht_{80}) or the rate at which LiDAR is intercepted by vegetation above some height (e.g., $cover_1$). These LiDAR-derived variables provide measures of stand height and the spread of vegetation, but they are not the same as field-based measures. When forest yield is calculated (estimated) for a plot, it is most often a function of tree diameters and heights. LiDAR does not provide a surrogate for field-measured tree diameters. However, there is a sufficiently strong association between LiDAR metrics and forest yield variables that the models provide significant explanatory power. Conceptually, we undertook model development with the idea that forest yield is in large part a functions of the height of the trees on a plot and the proportion of the plot that is occupied by vegetation. We also hypothesized that there should be an interaction between LiDAR-derived measures of tree height (e.g., ht_{50}, ht_{70}, ht_{90}, which are not the same as field-measured tree heights) and LiDAR-derived measures of the proportion of the plot that is occupied by vegetation. The expected difference in forest yield between two sites with different stand heights depends on how fully the stand is occupied by trees and vice versa. For example, the increase in biomass associated with going from 10% of the horizontal growing space occupied to 70% occupation is different for stands that are 13 m tall than for stands that are 30 m tall—this indicates that there is a need to recognize the interaction between these variables. As a result, we investigated models with LiDAR height and LiDAR occupancy variables, as well as interactions between these continuous variables. Accounting for the interaction between continuous height and occupancy variables enables nonlinear change in expected forest yield relative to changes in multiple predictors.

Estimation

Estimation of forest yield variables was performed with a regression estimator approach. The regression estimator is a model-assisted estimator, approximately design-unbiased, and takes advantage of a model to (potentially) increase estimation precision over that of the simple random sampling (SRS) estimator. This estimator is appropriate when the auxiliary variables, in this case LiDAR-derived auxiliary variables, are measured for the entire population and a sample of field measurement plots is obtained using an SRS (Lohr 1999, Gregoire and Valentine 2004, Särndal et al. 2010). The regression estimator may be used with variable probability sampling, but because our sample was an equal probability sample, we present the simplified total (Equation 2) and total variance (Equation 3) estimators (adapted from Gregoire and Valentine 2004 and Lohr 1999), which do not incorporate variable probabilities.

$$\hat{T}_{y,reg} = N^* (\hat{\boldsymbol{\beta}}\boldsymbol{\mu}_x) \tag{2}$$

where $\hat{T}_{y,reg}$ = regression estimate of total forest yield; $N = A_f/A_p$; N = number of observations in the population; A_f and A_p represent the area of the forest and area of a plot, respectively; $\hat{\boldsymbol{\beta}}$ = vector of OLS regression coefficients, including the intercept estimated from sample observations; and $\boldsymbol{\mu}_x$ = vector of population means for auxiliary variables.

$$SE_{\hat{T}_{y,reg}} = \sqrt{N^2\left(\frac{N-n}{N}\right)\frac{s^2_{reg}}{n}} \tag{3}$$

where $SE_{\hat{T}_{y,reg}}$ = approximate standard error of the regression estimate of the total on the original scale; s^2_{reg} = variance of regression-model residuals on original scale of data; n = number of observations in observed sample; and $(N - n)/N$ = finite population correction.

We also provide SRS estimators (Lohr 1999) of the population total (Equation 4) and population total variance (Equation 5). SRS estimates served as a baseline for comparison with regression estimates. Although our CFI plots are actually arranged on a fixed grid,

Table 4. Ordinary least squares models fitted to relate light detection and ranging-derived variables to each of the forest yield response variables.

Model form and parameter estimates	Model root mean square error	R^2
$ba = 10.0296 - 1.1924 \times ht_{60} - 14.7749 \times cover_2 + 3.5418 \times cover_2 \times ht_{60}$	9.9 m^2/ha	0.78
$vol = 196.680 - 13.947 \times ht_{60} - 618.603 \times cover_2 + 56.367 \times cover_2 \times ht_{60}$	173.7 m^3/ha	0.74
$stems = 4.981 + 10.233 \times ht_{40} - 14.073 \times ht_{80} + 615.178 \times cover_2 + 56.367 \times cover_2 \times ht_{80}$	104.9 stems/ha	0.54
$bm = 116.815 - 7.247 \times ht_{60} - 341.995 \times cover_2 + 32.374 \times cover_2 \times ht_{60}$	111.9 Mg/ha	0.71

ba, basal area, vol, total stem volume, including top and stump; stems, number of stems/ha; bm, total aboveground biomass. ht_{60} indicates 60th percentile height of light detection and ranging point data falling within a given pixel or plot. $Cover_2$ is the proportion of LiDAR point data within a pixel or plot with height (m) attributes smaller than the height indicated by the subscript.

SRS estimators were used as a conservative approximation (Bechtold and Patterson 2005). Regression estimates were compared with SRS estimates both by looking at their comparative magnitudes (variances and totals) and by taking the ratio of the regression estimation variance to SRS estimation variance (design effect). The design effect also indicates the number of additional plots that would be required in an SRS to achieve the same performance as the alternate design or estimator.

$$\hat{T}_y = N\bar{Y} \tag{4}$$

where $\hat{T}_y$ = SRS estimate of total; and $\bar{Y}$ = mean of observed sample.

$$SE[\hat{T}_y] = \sqrt{N^2\left(\frac{N-n}{N}\right)\frac{s_y^2}{n}} \tag{5}$$

where $SE[\hat{T}_y]$ = standard error of total estimate; and s_y^2 = sample variance.

The variance estimators listed are based on the assumption that the finite population is tessellated (cut into nonoverlapping regions, like pixels or strata) into discrete units and that our sample is drawn from these discrete units. In this scenario, the finite population correction (FPC) arises as an adjustment because the number of population elements that one must estimate is reduced by the size of the sample. The number of elements that must be estimated is $N - n$. Our estimation procedure did not precisely match with this theory. In our estimation procedure, we did tessellate our population (into square 28.45 × 28.45-m pixels), but our sample was not taken from these tessellations. Instead, our sample was composed of circular regions that were not aligned with the tessellations. Although the variance from our sample plots is still a reasonable estimate of variance, it is probably not appropriate to use the FPC because we do not actually have measurements of the response for any of the tessellations of our population. The FPC should instead be replaced by the value 1. In most instances, the sample size will be very small relative to the size of the population, and the effect of the FPC will be negligible.

Our population size, N, is derived from a thematic polygon GIS layer and from LiDAR raster layers (28.45 × 28.45 m). N is the area of Fort Lewis that we consider to be forested divided by the area of a plot. Forested areas (pixels) with regard to this study met the following three criteria: pixels fell within areas designated as forested within the thematic GIS polygon layer (Fort Lewis issued), pixels had at least 5% LiDAR cover ($cover_2$), and pixel vegetation heights (ht_{95}) were greater than 5 m. The thematic layer was used to omit areas outside of managed forest (e.g., housing developments that registered as forested according to the LiDAR criteria). The vegetation height criterion was introduced to exclude areas poorly represented by our sample data. Our data set did not include plots for which no tree had a dbh larger than 20.32 cm.

Results

Regression Results

In the development of regression models for estimation of forest yield variables from LiDAR-derived variables, we relied on both an exhaustive search algorithm (Lumley 2009) and nonautomated (hand fitted) backward selection based on model Bayesian information criterion (BIC) values. A hybrid approach was used because the search algorithm often returned models that did not include a cover or height predictor or contained seemingly redundant terms (e.g., ht_{90} with ht_{80}).

The hybrid procedure resulted in identical sets of predictor variables selected to model the three forest variables ba, vol, and bm. The models explained more than 70% of variability in the forest yield variables (Table 4). The models for these three yield variables included a single height variable, ht_{60}, a single cover variable, $cover_2$, and the multiplicative interaction between these two variables. The model for stems included ht_{40} and ht_{80}, $cover_2$, and an interaction between $cover_2$ and ht_{80}. The model for stems explained less variability in the response variable, 54%, than the models for the other three forest yield variables. No transformations of response variables were used, as modeling was performed for model-assisted estimation, which does not depend on linearity and equal variance for asymptotic unbiasedness of the total and total variance estimates.

Landscape Total Estimates

Total forest yield estimates (Table 5) were obtained using both regression and SRS estimators. The regression estimate of the population mean for a response is the predicted value from an OLS regression model for the population means of LiDAR-derived predictor variables. The regression estimate of the total (Equation 2) of the response for the population is the regression estimate of the mean for the population multiplied by the size of the population (total area). The SRS estimate of the total (Equation 4) is obtained by multiplying the sample mean by N, the size of the population. The variance (SE) estimator of the regression estimate of the total (Equation 3) is a function of the variability around the fitted regression line, the size of the sample, and the population size. Similarly, the variance (SE) of the SRS estimate of the total (Equation 5) is a function of the variability around the sample mean, the size of the sample, and the population size.

The difference between regression and SRS estimates of the forest yield variable stems was very small. The differences between SRS and regression estimates for the other variables, in contrast, were quite large. The SRS estimate of basal area, for example is 23% larger than the regression estimate. Confidence intervals (95%) for the two estimators cover the opposing estimates, but SRS estimates are very

Table 5. Regression (light detection and ranging) and simple random sampling (SRS) estimates of forest yield with respective estimates of confidence, precision, and relative precision.

Estimator	Resp.	Estimate	95% L CI	95% U CI	SE	DE
					.....(%).....	
Regression[a]	Basal area (m^2)	662,713	491,204	834,222	11	22
SRS	Basal area (m^2)	818,116	451,165	1,185,067	19	—[b]
Regression	Volume (m^3)	8,089,549	5,102,920	11,076,179	15	27
SRS	Volume (m^3)	10,624,926	4,832,758	16,417,094	23	—[b]
Regression	Stems	5,047,900	3,244,050	6,851,750	17	47
SRS	Stems	5,264,714	3,339,908	7,189,520	22	—[b]
Regression	Biomass (Mg)	4,983,460	2,362,792	7,604,127	15	29
SRS	Biomass (Mg)	6,843,596	3,281,361	10,405,830	22	—[b]

DE, design effect; L CI, lower boundary of confidence interval; Resp., response variable; SE, standard error; U CI, upper boundary of confidence interval.
[a] Regression indicates the regression estimator of the total.
[b] No DE% values are provided for the SRS estimator because the variances of the SRS estimates are the denominator when estimating design effects.

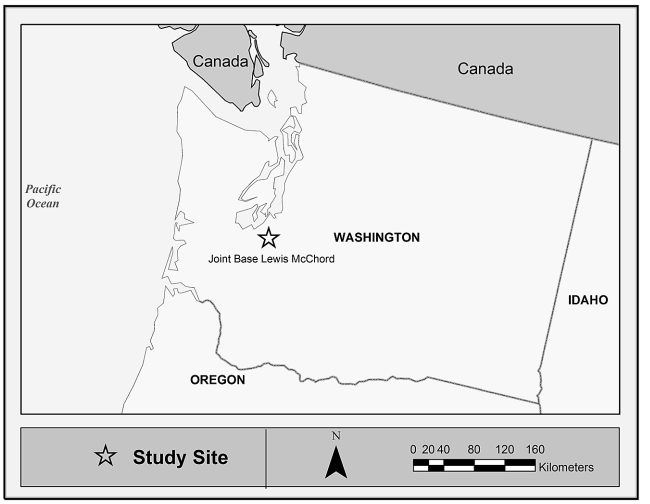

Figure 2. Location of Joint Base Lewis-McChord, the location of our study site, in western Washington state.

close to the boundaries of the confidence intervals for the regression estimates (except stems). As a result of the strong linear association between predictor and response variables, the regression estimates of total forest yield were much more precise than SRS estimates. Design effects for ba, vol and bm were all less than 30%, and the design effect for stems, the variable with the least association with LiDAR-derived variables, was 47%. Although there is no reason that model-assisted and SRS estimates of forest yield cannot diverge because of random chance, as part of due diligence, we examined the issue further using LiDAR-derived variables as surrogates for field measured variable. We looked at the marginal distributions of the predictor variables included in the regression models for the sample and the population. The distribution of the predictor variable $cover_2$ was centered on a mean value approximately 10% higher than the distribution of $cover_2$ for the entire study area (Figure 2). Visually, the sample and population distributions of the other LiDAR-derived variables closely matched. The difference ($cover_2$) indicates that CFI plots were placed in patches of forest that were comparatively dense relative to the rest of the landscape. The difference in the distribution of $cover_2$ appears to be larger in magnitude than what would be expected from random chance. We used a randomization test to evaluate the probability of witnessing a difference as large as the one observed between the sample and population for $cover_2$. We selected 5,000 random samples of size 118 from the LiDAR for the whole forest, and 0.04% of samples had a difference from the population mean for $cover_2$ as large as the difference observed for our sample ($P = 0.0004$). This appears to indicate that all sampling units were not given equal probability of selection.

Discussion

Sampling Bias

As a result of evidence that plots were placed in locations with higher density, we cannot assert that either the SRS or model-assisted estimates were unbiased. We are not sure what the level of bias is for the model-assisted estimators, but there is clearly an effect from the biased sample on the SRS estimator. This scenario is an indication that the model-assisted estimator performs better than the SRS estimator. The model-assisted estimator may not be unbiased in this case (a defect in the sample not in the estimator), but the estimator is able to recover in some sense because it is able to adjust the estimate according to the differences observed between values for explanatory variables observed on field plots and values observed for the entire population. This essentially provides a model-based fallback position for the model-assisted estimator. The theory that we presented for the model-assisted estimator was for design-based estimation, but the model-assisted total estimator remains unbiased for the total in the model-based sense even when the sampling probabilities are not known and, as in this case, are probably not equal.

Ideally, the observed difference between the plots and the landscape would not occur because randomization is the basis for inference in the model-assisted paradigm. However, there are many ways for bias to be introduced in the distribution of plot locations. Examples of ways that bias could be introduced into a design include through subtle preference toward a forest type when ground crews navigate approximate coordinates for CFI plot monumentation, destruction of CFI plot monumentation on harvested sites, or insertion of additional plots to represent desirable types. Although there is evidence to suggest that there was some subjectivity with regard to the establishment of the plots—which would introduce bias into the field-based sample—for the majority of our discussion, we treat our results as if the field sample represents a real equal-probability sample to demonstrate the model-assisted approach and to provide an indication of the precision of the estimator relative to the SRS estimator.

Other Sources of Error

During the plot measurement and aggregation phase of the analysis, three sources of error were identified that likely introduced variance into our response variables and hence our predictions. The

first source of error was estimation of biomass and volume using allometric equations. This source of error is unavoidable in our situation as it is not practical to improve these estimates with typical inventory data. The other two sources of error are in the estimation of tree heights and omission errors resulting from the use of a 20.3-cm threshold for minimum tree diameter. It is not practical to measure all tree heights in every case, as here, but estimates can likely be improved by measuring some trees on each plot, measuring trees that have broken tops so that their heights are not overestimated, and ensuring that the sample of trees is well distributed among diameters and species (e.g., Næsset 2002). A lower diameter threshold for measured trees would also likely be of benefit. These changes would result in more precise measures of forest yield per plot. Improved measurement precision of the response variable would reduce the variability of residuals around the regression line and increase estimation precision (Kmenta 1997, pp. 346–348).

Estimation

A large portion of the variability in forest yield was accounted for by LiDAR-derived variables. This resulted in much smaller variance around the regression model for forest yield then was observed around the sample mean. As a result, the precisions of regression estimates were much improved over SRS estimates. Our design effects ranged from 22% to 47%, indicating that between 2 and 5 times the number of field plots would be necessary to achieve the same precision with an SRS estimator as with the model-assisted estimator. This level of precision, or perhaps even better precision with more precise field measurements, can be expected by leveraging the linear relationship between forest yield and LiDAR-derived variables in a model-assisted forest inventory, although sample size and forest size are also key components of the variance.

Previous studies have demonstrated strong associations between forest yield and LiDAR-derived variables. As previously mentioned, studies that publish R^2 (or adjusted R^2) values exceeding 0.90 are not uncommon. The strength of the associations was leveraged in this study to greatly increase estimation precision over the SRS estimator. However, direct comparison between the precision observed here and in other studies that model forest yield with LiDAR is not readily feasible. We are not aware of any other studies that used the regression estimator for estimation of forest yield. The actual precision that can be achieved with a model-assisted approach for other areas is not yet clear, but the very high ($>$0.90) R^2 values found in studies on transformed scales still indicate a strong association between forest yield and LiDAR-derived variables. This is evidence that the model-assisted approach to estimation is broadly applicable and can be used with good result in a variety of forest types, including, for example, boreal mixed forests (Thomas et al. 2006), northern broadleaf forests (Lim et al. 2003), temperate conifer and deciduous forests (Lefsky et al. 2002), and even tropical rainforests (Drake et al. 2003).

If model R^2 values or model root mean square error values published in studies are used as the basis for evaluation of model-assisted approaches prior to implementation, it is important to recognize that the published model precisions may not be observed in practice. For example, several studies in the US Pacific Northwest achieved R^2 values greater than 0.90 for forest yield variables (Means et al. 1999, 2000a, Lefsky et al. 2005, Goerndt et al. 2010), whereas we observed much reduced model R^2 values. Studies are often performed for very small areas with limited numbers of field plots, and the plots are often measured specifically for the study. The level of care taken in measuring field plots for a research study may exceed that of an operational forest inventory, and the precision observed for models developed for small forests is likely to differ from that observed for large forests.

Relative Precision

Forest managers may have an interest in using LiDAR for forest inventory, but prior to selecting this tool they will probably need some indication of the performance of estimation. A common measure used to evaluate the merits of a design is design effect.

This statistic is a basis for comparison of the precision of a selected design or estimator with SRS estimation. Based on the central limit theorem, design effect indicates how many observations are necessary to achieve the same precision with an alternative design as was achieved with the SRS design. In this study, the regression estimator with LiDAR-derived predictor variables was much more precise than the SRS estimator. A land manager is unlikely to use an SRS design for forest inventory in practice, but by comparing the design effect achieved in this study with the design effect of another design and estimator, the efficiency (precision relative to number of sample plots) of the approach described in this study can be compared with another design and estimator. The analytical design effect estimator is a function of sample size, so it is also possible to look at relative performance for a variety of sample sizes.

Other Considerations with LiDAR

LiDAR constitutes an additional cost that may deter some users. However, model-assisted estimation may result in a decrease in the number of plots that are necessary to achieve a given precision, and there are additional uses for LiDAR that should be considered. LiDAR can also be used to develop maps of forest variables, create bare-earth digital terrain models, perform slope stability analyses (Schulz 2007), perform probable stream channel delineations (Murphy et al. 2008), and (in conjunction with other LiDAR-derived products) develop ecological models (Vierling et al. 2008) and fire fuel models (Morsdorf et al. 2004, Andersen et al. 2005, Mutlu et al. 2008). Alternative uses for LiDAR products should be considered when evaluating LiDAR for forest inventory and when attempting to find parties interested in sharing the cost of a LiDAR acquisition—especially on public or other lands that are likely to have multiple stakeholders.

Conclusions

In this study, we used the regression estimator, a model-assisted estimator, with LiDAR-derived auxiliary variables to estimate forest yield. The precision of the regression estimator (assuming a random sample) was better than that of the SRS estimator, with design effects ranging from 0.22 for ba to 0.49 for stems. Depending on the forest, the precision achieved with LiDAR in this study relative to the SRS estimator may stimulate interest in implementing a forest inventory with LiDAR. It should be noted that there are difficulties in attempting to distinguish species with LiDAR. Also, the particular model-assisted estimator described in this study is applicable only to areas with wall-to-wall LiDAR coverage; other arrangements will require alternative approaches to estimation. Inference for the model-assisted estimator is based on the design-based paradigm. In contrast to the model-based paradigm, we did not have to pay explicit attention to OLS modeling assumptions, such as equal variance and linearity. This simplifies the modeling process. However, it

should be noted that to obtain valid inferences with a model-assisted estimator, there is an additional assumption that the sample used to calibrate the regression model is taken randomly from the population of interest.

Literature Cited

ANDERSEN, H.E., T. CLARKIN, K. WINTERBERGER, AND J.L. 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.

ANDERSEN, H.E., R.J. MCGAUGHEY, AND S.E. REUTEBUCH. 2005. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sens. Environ.* 94(4): 441–449.

BECHTOLD, W.A., AND P.L. PATTERSON. 2005. *The enhanced forest inventory and analysis program: National sampling design and estimation procedures.* US Department of Agriculture Forest Service, Southern Research Station, Asheville, NC. 85 p.

CLARKIN, T. 2007. *Modeling global navigation satellite system positional error under forest canopy based on LIDAR-derived canopy densities*. MSc thesis, Univ. of Washington, Seattle, WA. 99 p.

CORONA, P., AND L. FATTORINI. 2008. Area-based lidar-assisted estimation of forest standing volume. *Can. J. For. Res.* 38(11):2911–2916.

DRAKE, J.B., R.G. KNOX, R.O. DUBAYAH, D.B. CLARK, R. CONDIT, J.B. BLAIR, AND M. HOFTON. 2003. Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. *Global Ecol. Biogeogr.* 12(2):147–159.

GOERNDT, M.E., V.J. MONLEON, AND H. TEMESGEN. 2010. Relating forest attributes with area-and tree-based light detection and ranging metrics for western Oregon. *West. J. Appl. For.* 25(3):105–111.

GREGOIRE, T.G. 1998. Design-based and model-based inference in survey sampling: Appreciating the difference. *Can. J. For. Res.* 28(10):1429–1447.

GREGOIRE, T.G., AND H.T. VALENTINE. 2004. *Sampling strategies for natural resources and the environment,* 1st ed. Chapman and Hall/CRC, New York. 474 p.

GREGOIRE, T.G., G. STAHL, E. NAESSET, T. GOBAKKEN, R. NELSON, AND S. HOLM. 2011. Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Can. J. For. Res.* 41(1):83–95.

JENKINS, J.C., D.C. CHOJNACKY, L.S. HEATH, AND R.A. BIRDSEY. 2003. National-scale biomass estimators for United States tree species. *For. Sci.* 49(1):12–35.

KEYSER, C.E. 2010. *49 Pacific Northwest Coast (PN) variant overview - Forest Vegetation Simulator.* US For. Serv. Internal Report. WO-Forest Management Service Center, US For. Serv., Fort Collins, CO. 49 p.

KMENTA, J. 1997. *Elements of econometrics,* 2nd ed. University of Michigan Press.

LEFSKY, M.A., W.B. COHEN, S.A. ACKER, G.G. PARKER, T.A. SPIES, AND D. HARDING. 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sens. Environ.* 70(3):339–361.

LEFSKY, M.A., W.B. COHEN, D.J. HARDING, G.G. PARKER, S.A. ACKER, AND S.T. GOWER. 2002. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecol. Biogeogr.* 11(5):393–399.

LEFSKY, M.A., A.T. HUDAK, W.B. COHEN, AND S.A. ACKER. 2005. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. *Remote Sens. Environ.* 95(4):532–548.

LIM, K., P. TREITZ, K. BALDWIN, I. MORRISON, AND J. GREEN. 2003. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Can. J. Remote Sens.* 29(5):658–678.

LOHR, S.L. 1999. *Sampling: Design and analysis.* Duxbury Press, Pacific Grove, CA. 494 p.

LUMLEY, T. 2009. *Regression subset selection.* Available online at cran.r-project.org/web/packages/leaps/index.html; last accessed June 28, 2010.

MCGAUGHEY, R.J. 2009. *FUSION/LDV: Software for LIDAR data analysis and visualization, version 2.9.* US For. Serv. Available online at www.fs.fed.us/eng/rsac/fusion; last accessed Apr. 3, 2010.

MEANS, J.E., S.A. ACKER, B.J. FITT, M. RENSLOW, L. EMERSON, AND C. HENDRIX. 2000a. Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sens.* 66(11):1367–1372.

MEANS, J.E., S.A. ACKER, B.J. FITT, M. RENSLOW, L. EMERSON, AND C. HENDRIX. 2000b. Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sens.* 66(11): 1367–1371.

MEANS, J.E., S.A. ACKER, D.J. HARDING, J.B. BLAIR, M.A. LEFSKY, W.B. COHEN, M.E. HARMON, AND W.A. MCKEE. 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sens. Environ.* 67(3):298–308.

MORSDORF, F., E. MEIER, B. KÖTZ, K.I. ITTEN, M. DOBBERTIN, AND B. ALLGÖWER. 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sens. Environ.* 92(3):353–362.

MURPHY, P.N.C., J. OGILVIE, F. MENG, AND P. ARP. 2008. Stream network modelling using lidar and photogrammetric digital elevation models: A comparison and field verification. *Hydrol. Process*. 22(12):1747–1754.

MUTLU, M., S.C. POPESCU, C. STRIPLING, AND T. SPENCER. 2008. Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sens. Environ.* 112(1):274–285.

NÆSSET, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ.* 61(2):246–253.

NÆSSET, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* 80(1):88.

NÆSSET, E., O.M. BOLLANDSÅS, AND T. GOBAKKEN. 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sens. Environ.* 94(4): 541–553.

PARKER, R.C., AND D.L. EVANS. 2004. An application of LiDAR in a double-sample forest inventory. *West. J. Appl. For.* 19(2):95–101.

PARKER, R.C., AND D.L. EVANS. 2007. Stratified light detection and ranging double-sample forest inventory. *South. J. Appl. For.* 31(2):66–72.

R. DEVELOPMENT CORE TEAM. 2010. *R: A Language and Environment for Statistical Computing.* Available online at www.R-project.org; last accessed Jun. 12, 2010.

REUTEBUCH, S.E., H.E. ANDERSEN, AND R.J. MCGAUGHEY. 2005. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. *J. For.* 103(6):286–292.

RIAÑO, D., E. CHUVIECO, S. CONDÉS, J. GONZÁÉLZ-MATESANZ, AND S.L. USTIN. 2004. Generation of crown bulk density for *Pinus sylvestris* L. from lidar. *Remote Sens. Environ.* 92(3): 345–352.

SÄRNDAL, C.E., B. SWENSSON, AND J. WRETMAN. 1992. *Model assisted survey sampling.* Springer Verlag, New York. 694 p.

SCHULZ, W.H. 2007. Landslide susceptibility revealed by LIDAR imagery and historical records, Seattle, Washington. *Eng. Geol.* 89(1–2):67–87.

TEMESGEN, H., V.J. MONLEON, AND D.W. HANN. 2008. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. *Can. J. For. Res.* 38(3):553–565.

THOMAS, V., P. TREITZ, J.H. MCCAUGHEY, AND I. MORRISON. 2006. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: An examination of scanning density. *Can. J. For. Res.* 36(1):34–47.

US FOREST SERVICE. 2008. National Volume Estimator Library (NVEL). Available online at www.fs.fed.us/fmsc/measure/volume/nvel/index.php; last accessed Aug. 19, 2009.

VIERLING, K.T., L.A. VIERLING, W.A. GOULD, S. MARTINUZZI, AND R.M. CLAWGES. 2008. Lidar: Shedding new light on habitat characterization and modeling. *Front. Ecol. Environ.* 6(2):90–98.