Impact of sample size – empirical results from a hybrid inference two-phase inventory based on dense laser scanning

H. J. Persson[1], K. Olofsson[1], J. Holmgren[1]

[1]Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden
Email: henrik.persson@slu.se

1. Introduction

Airborne laser scanning (ALS) from low flight altitudes will produce high-resolution data but is not efficient for large area mapping. ALS data can therefore be collected in strips to enable an accurate inventory of larger areas. High-resolution data from ALS can provide estimates of tree positions, tree height, and tree species, but reference data are commonly needed from field inventories for estimation of stem attributes. In this study, we apply ALS strip sampling as first phase in forest stands, and use sample trees from a second phase using terrestrial laser scanning (TLS). The TLS measures were used instead of manual field inventories to build regression functions, which were used to predict stem volume of all detected trees in the strips (first phase). The estimated volumes with respect to the number of strips in each forest stand was evaluated.

2. Material

2.1 Study area

The study area is located at 62.9°N 16.9°E in middle Sweden. A subset consisting of ten boreal forest stands covering 207 ha were used in this study. The forest was dominated by Norway spruce (Picea abies (L.) H. Karst.), Scots pine (Pinus sylvestris L.), and birch (Betula spp.)., where pine (50%) and spruce (44%) constituted 94% of the growing stock and birch 6%. The stands had an average VOL of 235 m^3/ha and the stand ages were between 98 and 150 years.

2.2 Harvester data

Two forest harvesters were used to harvest the ten forest stands in 2020, after ALS scanning. The harvester registered the tree species and measured the stem diameter along the trunk. We estimated the stem volume based on the DBH and height, using regional valid functions. The harvested areas were flown with a drone after harvest, carrying an optical camera. Ortho-photographs with 4 cm pixel resolution could be generated, which enabled us to delineate the harvested forest areas accurately.

2.3 Laser scanning data

The ALS system Riegl LMS-Q680i was acquiring data from a helicopter. The nominal flight speed 20 km/h, and the altitude 70 m above ground level. The nominal swath width was 90 m and the nominal point density ranged from 490 points/m^2 to 654 points/m^2, with an average of 593 points/m^2. The laser scanning of the 10 forest stands was performed 3 November 2019.

Terrestrial Laser Scanning (TLS) was conducted at the location of the field plot centers. The Trimble TX 8 instrument scanned in a hemispherical pattern with a point spacing of 11.3 mm at 30 m distance. The wavelength was 1.5 micrometer. The tree stem properties was estimated using the TLS data and an algorithm by Olofsson and Holmgren (2016).

3. Methods

3.1 Processing of ALS data

For automatic delineation of tree crowns, we used an algorithm based on density models of tree crown (Holmgren and Lindberg 2019). It first uses a canopy height model (CHM) to obtain an approximate height of potential tree height positions, at 0.25 m gridding. Template matching was then applied to

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021. Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.1932

create a model similarity surface that was used as input in a watershed segmentation to generate an automatic delineation of all tree crowns.

3.2 Processing of TLS data

The algorithm used to estimate the stem profiles and VOL of the trees from a 3D point cloud was presented in Olofsson and Holmgren (2016). The volume of the stems were estimated as truncated cones connecting the centers of the modeled stem cylinders (Olofsson and Holmgren 2017). The top part of the tree (where the singletree detection algorithm was unable to detect the stem cylinder), was modeled as a complete cone reaching the highest registered laser point of the canopy.

3.3 Two-phase hybrid inference

The first phase consisted of ALS strips across the stands. These were selected to cover the systematically distributed TLS scan locations in phase two. Depending on the stand geometry, the number of strips varied between 3 and 5, but for some stands perpendicular flight lines of nearby stands made even more strips available. The strips (used in the results below) were randomly selected. All trees that were identified in both the TLS and ALS data within the strips were used to estimate model parameters with robust multiple linear regression, using the R-package MASS and the default Huber variance estimator (Huber 1981). The regression model was used to predict the VOL, $\hat{y}$, on all segmented trees, and the model had the form of:

$$ln(\text{VOL}) = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_p X_p \tag{1}$$

where the parameters $[\alpha_0, \alpha_p]$ for the p attributes X_i, $i \in [1, p]$ were the following statistical metrics computed from the ALS point clouds for the single tree segments: height percentile 10, 80, 95, and crown width. Since the dependent variable VOL was transformed using the natural logarithm, a correction for logarithmic bias was applied, by adding $s^2/2$ to (1) before taking the inverse transform of the prediction, s^2 being the residual variance from (1) (Finney 1941).

The population (stand) means ($\hat{\mu}_{Y_{VOL}}$) were estimated with a ratio-to-size estimator, as the size of the strips varied depending on the stand shape. The total volume T_k for each strip k was calculated by summing up the volumes $\hat{y}_j$ predicted with the linear regression function for all trees M_k in the strip:

$$\hat{T}_k = \sum_{j=1}^{M_k} \hat{y}_j \tag{2}$$

Then, the mean ($\hat{\mu}_{Y_{VOL}}$) was estimated as

$$\hat{\mu}_{Y_{VOL}} = \frac{\sum_{k=1}^{M_c} \hat{T}_k}{\sum_{k=1}^{M_c} \hat{a}_k} \tag{3}$$

where $\hat{a}_k$ denotes the total area of strip k, and M_c denotes the total number of strips in the stand.

An approximated variance (Ståhl et al. 2011) of the estimators in (3) and (5) is

$$\widehat{Var}(\hat{\mu}_{Y_{VOL}}) = s^2_{Y_{VOL}} + \sum_{d=1}^{p} \sum_{e=1}^{p} \widehat{Cov}(\hat{\alpha}_d, \hat{\alpha}_e)\, \bar{\hat{T}}'_d \bar{\hat{T}}'_e \tag{4}$$

where the first term represents the variability due to the first-phase sampling and the second term represents the model error due to the uncertainty of the parameter estimates. p is the number of model parameters, $\widehat{Cov}(\hat{\alpha}_d, \hat{\alpha}_e)$ is the estimated covariance between the model parameter estimates, and $\bar{\hat{T}}'_d \bar{\hat{T}}'_e$ are the estimated average values of the first order partial derivatives of the function used to estimate the target variable. The first-phase sampling variability was estimated as:

$$s^2_{Y_{VOL}} = \left(1 - \frac{a}{A}\right) \frac{\sum_{k=1}^{M_c} \left(\hat{T}_k - \hat{\mu}_{Y_{VOL}} \hat{a}_k\right)^2}{M_c \bar{\hat{a}}^2 (M_c - 1)} \tag{5}$$

where a represents the area covered by the strips, A represent the total area of the population (stand), and $\bar{\hat{a}}$ is the mean strip area.

To provide an estimate of mean bias, the estimated stand means ($\hat{\mu}_{Y_l}$) were compared with the reference values μ_Y from the harvester, and the mean population bias for all M_d stands was estimated as

$$\hat{B} = \frac{1}{M_d} \sum_{l=1}^{M_d} (\hat{\mu}_{Y_l} - \mu_{Y_l}) \tag{6}$$

The RMSE was used as accuracy measure for all stands, and it was estimated as

$$\widehat{RMSE} = \sqrt{\frac{1}{M_d} \sum_{l=1}^{M_d} (\hat{\mu}_{Y_l} - \mu_{Y_l})^2} \tag{7}$$

using the same notations as earlier. The standard error was calculated as the square root of the variance:

$$\widehat{SE} = \sqrt{\widehat{Var}(\hat{\mu}_Y)}). \quad (8)$$

4. Results

The results showed that an increasing number of strips lowered the RMSE, both in absolute and relative terms (Table 1), and the bias showed a similar trend. The trend can also be seen in the scatter plots (Figure 1). The proportion sampled area increased with more strips from n=1 until n=3, where it plateaued. The proportion corresponded to about 30%, but higher n decreased the RMSE further, without sampling a larger proportion of the stands. This indicates that it would be more cost efficient to collect more, shorter strips, rather than fewer but larger strips. The standard error was rather stable in the range of 9% to 17%, but without correlation to the number of strips. Since the variance estimator depends on the covered proportion, larger differences and repeated simulations would be needed to investigate this further.

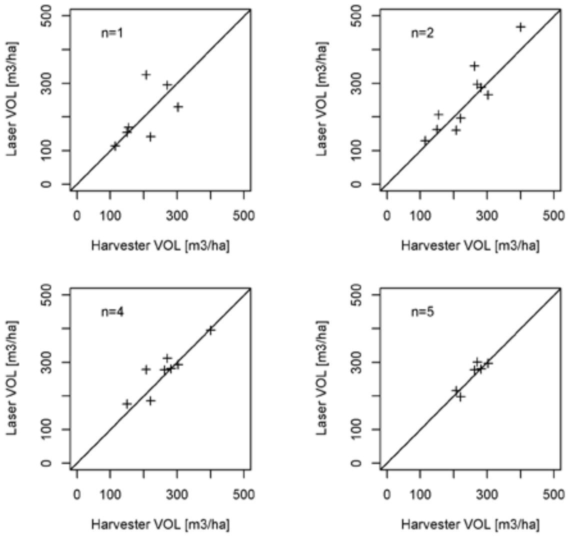

Figure 1: Estimated vs. reference VOL for different number of flight lines.

Table 1. Accuracies results

n	RMSE	Bias	Standard Error	Proportion of Tot Area	Stands
1	77.7 (34.2%)	19.6 (8.64%)	-	10.6%	8
2	45.0 (19.0%)	16.1 (6.82%)	36.2 (15.3%)	21.6%	10
3	45.1 (19.1%)	12.3 (5.18%)	21.4 (9.06%)	32.5%	10
4	33.7 (12.9%)	13.3 (5.09%)	35.2 (13.5%)	29.3%	8
5	17.3 (6.73%)	4.12 (1.60%)	45.0 (17.5%)	30.1%	6

Acknowledgements

The work was funded by the Bo Rydin’s foundation for scientific research (award no F19/17) and Mistra Digital Forest. The HPC2N is acknowledged for providing data resources for the TLS processing.

References

Finney, D. J. 1941. “On the Distribution of a Variate Whose Logarithm Is Normally Distributed.” *Supplement to the Journal of the Royal Statistical Society* 7(2):155–61.

Holmgren, Johan and Eva Lindberg. 2019. “Tree Crown Segmentation Based on a Tree Crown Density Model Derived from Airborne Laser Scanning.” *Remote Sensing Letters* 10(12):1143–52.

Huber, P. J. 1981. *Robust Statistics*. Wiley, New York.

Olofsson, K. and J. Holmgren. 2017. “Tree Stem and Canopy Biomass Estimates from Terrestrial Laser Scanning Data.” Pp. 157–60 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. Vol. 42.

Olofsson, Kenneth and Johan Holmgren. 2016. “Single Tree Stem Profile Detection Using Terrestrial Laser Scanner Data, Flatness Saliency Features and Curvature Properties.” *Forests* 7(9).

Ståhl, Göran, Sören Holm, Timothy G. Gregoire, Terje Gobakken, Erik Næsset, and Ross Nelson. 2011. “Model-Based Inference for Biomass Estimation in a LiDAR Sample Survey in Hedmark County, Norway.” *Canadian Journal of Forest Research* 41(1):96–107.