LARGE SCALE AIRBORNE LASER SCANNING OF FOREST RESOURCES IN SWEDEN

J. Holmgren[a, *] and T. Jonsson[b]
[a] Department of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences,
SE-90183 Umeå, Sweden - johan.holmgren@resgeom.slu.se
[b] The Regional Forestry Board of Dalarna-Gävleborg, Nygatan 3, SE-82683 Söderhamn, Sweden -
thomas.jonsson@svswx.svo.se

KEY WORDS: Basal area, Forest inventory, LIDAR, Stem diameter, Stem volume, Tree height.

ABSTRACT:

The first large scale laser scanning project in Sweden for the purpose of forest inventory was started by a regional forestry board in central Sweden. The objective was to compare laser scanning with traditional operational methods for large area forest variable estimations. Laser data were acquired for a 50 km^2 forest area in central Sweden with approximately 1.2 laser measurements per square meter. One field plot within each of 122 forest stands representing different forest types were used as training data. Mean tree height, mean stem diameter, basal area, and stem volume were predicted using regression functions with variables extracted from the laser canopy height distribution. Separate regression functions were built for Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*) and deciduous trees. The functions were applied at grid nodes for a grid covering the whole forest area and mean values of laser estimated forest variables were calculated within each forest stand. A special validation inventory of 29 forest stands was performed. The size of the forest stands used for validation ranged from 0.5 to 12 ha (median forest stand size was 1.3 ha). The relative RMSE at stand level was 5.0% (0.8 m) for mean tree height, 8.9% (1.9 cm) for mean stem diameter, 12.5% (3.0 m^{2}ha^{-1}) for basal area, and 14.1% (28 m^{3}ha^{-1}) for stem volume estimations. The results imply that estimations of forest variables using laser scanning give higher accuracies compared with using traditional methods.

1. INTRODUCTION

In Sweden, different information for planning forest management is required at different levels. The tactical level deals with allocation of forest operations geographically for certain time periods (Jonsson et al., 1993). For tactical planning, estimates of forest variables are needed for all forest land at the stand level. The objective field methods that exist today for estimation of forest variables are too expensive to use for all forest land. Therefore, forest companies are using several subjective estimation methods. The subjective field inventory methods that are commonly used in Sweden today at stand level typically give standard errors of about 10% (proportion of mean value) for mean tree height estimation, 6 to 17% for mean diameter estimation, and 15 to 25% for stem volume and basal area estimations (Ståhl, 1992). One common method is to use manual interpretation and tree height measurements in stereo models of aerial photos combined with a field survey (Åge, 1985). Several studies have indicated that mean tree height (Næsset, 1997; Magnussen & Boudewyn, 1998; Means et al., 2000), and basal area and stem volume (Means et al., 2000) can be empirically related to the height distribution of laser measurements in the canopy and the proportion of laser returns reflected in the canopy out of all laser returns. A two-staged procedure, where objectively field-measured sample plots are used for building regression functions and the functions then applied for predictions for a raster covering the whole laser scanned area, has been developed in Norway (Næsset & Bjerknes, 2001; Næsset, 2002). In this paper, results are reported from the first large area laser scanning based forest inventory in Sweden. The objective was to compare laser scanning based estimation with traditional methods. A regional forestry board (Dalarna-Gävleborg forestry board, www.svo.se) bought the laser data and performed all field inventories and GPS-measurements. Because GIS is used on a daily basis in forestry, the highly automatic laser data processing can smoothly be integrated as a tool for estimating forest variables.

2. MATERIAL AND METHODS

The test site was a 5000 ha large area located in central Sweden (60°59′ N, 15°14′ E). The forest was dominated by Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), and with a lower abundance of deciduous trees, mainly birch (*Betula* spp.).

2.1 Laser data

The laser scanning system (Optech ALTM 2033) was operated from a flight altitude of 800 m above ground level and with a flight speed of 75 m s^{-1} on the 28th of May 2003. The divergence of the laser (1064 nm) was 0.3 mrad. The first and last return pulse were registered. The pulsing frequency was 33 000 Hz, the scan frequency 50 Hz, and the scan angle ±11 degrees. The average measuring density was 1.2 laser measurements per square meter. For quick access, ascii-files (x, y, z-coordinates), each containing laser data from one square kilometre, were converted to binary files. Also, one text-file, containing information about where to find data in a binary file for a specific 10×10 m^2 area, was created for each binary file.

2.2 Field data

All field inventories were based on 10 m radius circular field plots and used the field inventory procedures of the Forest Management Planning Package (FMPP) (Jonsson et al., 1993). Within each plot all trees with a stem diameter $\geq$ 5 cm were callipered and tree species recorded. Tree height was measured for a sample of trees using an angle measuring hypsometer. Tree height and form height of all callipered trees on the plots were estimated by means of static functions (Söderberg, 1992), having stem diameter as the most influencing variable. Stem volume for single trees was estimated by multiplication of basal area and form height. Functions for tree height and form height are calibrated using the measurements of sample trees to eliminate local systematic errors (Jonsson et al., 1993).

Field data for building regression models were collected from field plots for which the position of the plot centre was measured using DGPS. The inventory was performed in the autumn of 2003. A stratified sampling design based on photo interpretation was used in order to have all forest types represented in the training dataset. The forest had been delineated into 3321 forest stands by interpretation based on stereo models of black and white aerial photos at a scale of 1:30 000 and several forest variables had been interpreted for each forest stand. Of these, 122 forest stands were selected for being used for model selection and parameter estimations and one field plot was therefore placed in the centre of each of the selected forest stands.

Field data for validation were colleted from 29 randomly selected forest stands with a stand size ranging from 0.5 to 12 ha and with a median stand size of 1.3 ha. For the selection two restrictions were applied: (1) forest stands with clear cutting or thinning for the period between photo interpretation (Summer 2002) and the field inventory (Spring 2004) were excluded, and (2) forest stands containing seed trees were excluded because these stands had two separate canopy layers. Within each of the 29 forest stands, approximately 10 field plots were systematically placed using a grid. The positions of the field plots were measured using DGPS within 12 of the 29 forest stands used for validation. These field plots with measured position were also used for estimating parameters of the regression functions. However, prediction of forest variables were performed for each of the 12 stands at a time without using plots from the specific forest stand (cross validation).

2.3 Laser data processing

The coordinates of the reflection surfaces for the laser pulses were classified as ground or vegetation laser points using the TerraScan software (www.terrasolid.fi). For each field plot, two rasters were created: one ground raster and one vegetation raster, each with a size of 30×30 m^2 and with a cell size of 0.5 m. The ground level was estimated for each cell in the ground raster by assigning the z-value of the closest (horizontal distance) ground laser point located within 4 m of the cell centre. If no ground laser points were assigned, the raster cell was labelled as a missing value. For a ground raster cell with a missing value, the value was set to the mean of the ground estimates within the smallest window covering at least one ground level estimate. The height value of a vegetation laser point within a raster cell was computed as the difference between the z-value and the estimated ground level for the raster cell. The height value of a vegetation raster cell was assigned the greatest height value from the vegetation laser points within 0.5 m (horizontal distance) from the centre of the raster cell if the value was above a height threshold of 3 m. The height threshold was applied for eliminating laser measurements from low vegetation and other low objects. On plot level, several variables were computed based on the vegetation raster cell values that had their centre within the plot radius. The variables were 10th percentile (h_{10}), 20th percentile (h_{20}), …, 90th percentile (h_{90}), 95th percentile (h_{95}), and 100th percentile (h_{100}). The vegetation-ratio (D_v) was calculated as the ratio between number of laser returns above the height threshold and total number of returns from the plot. In order to apply regression functions on all forest land, laser variables were extracted at grid nodes according to the following steps: (1) 20 m inter-node distance and 10 m radius, (2) 10 m inter-node distance and 10 m radius, (3) 5 m inter-node distance and 10 m radius, (4) 5 m inter-node distance and 5 m radius, and (5) 2 m inter-node distance and 4 m radius. Predictions for a node were only included when calculating the mean values of forest variables within a forest stand if the distance to the nearest stand boundary was greater than the radius. For a forest stand, the first step that included at least 10 nodes was applied. One advantage of using these steps was that extraction was needed only for whole circles. Variable extractions from parts of circles are less reliable and need to be weighted differently according to the area that is within the stand. One disadvantage of only using grid nodes with a certain distance to the stand boundary could be that the area close to stand boundaries will be less represented when deriving a mean value for the stand.

2.4 Statistical analysis

Separate regression functions were built for Norway spruce, Scots pine and deciduous tree dominated forest stands. The tree species proportion as estimated by photo interpretation was used for grouping the forest stands. The parameters β_0, β_1, β_2, and β_3 were derived separately for each model and dataset (Scots pine, Norway spruce or Deciduous trees) using the least-squares method. For a model and dataset combination, a term was excluded if its coefficient was not significant ($p>0.05$). High correlations were found between basal area weighted mean tree height (h_L) and the laser height percentiles. The following model (Eq. 1) was therefore used:

$$h_L = \beta_0 + \beta_1 h_X + \varepsilon \tag{1}$$

where h_X is a laser height percentile and ε is random error. All extracted percentiles were tested and the percentile yielding the lowest residual sum of squares was selected. Studies of scatter plots showed that height percentiles could be used for estimating basal area weighted mean diameter (d_L). However, the relationship was not always linear. Also, it was suspected that forest density could explain some of the mean diameter variation and the following model (Eq. 2) was therefore used:

$$d_L = \beta_0 + \beta_1 h_X^p + \beta_2 D_v + \varepsilon \tag{2}$$

where h_X is a height percentile raised to the power of p and D_v is the vegetation ratio. Extracted height percentiles and values of p between 1.0 and 2.5 with a 0.1 interval were tested and the combination yielding the lowest residual sum of squares was selected. It was assumed that laser canopy height percentiles were related to mean tree height and the vegetation-ratio (D_v) was related to crown area. Given that stem cross-sectional area is related to crown volume, a multiplicative model with the vegetation-ratio (D_v) and one laser canopy height percentile would be suitable for estimation of both basal area and stem volume. However, in order to apply linear regression, an additive model must be implemented. Therefore, basal area, stem volume, laser canopy height percentiles, and the vegetation-ratio (D_v) were transformed with the natural logarithm before being used in the model. The residuals were correlated with the ratio between a high and a low height percentile (r_X) and this term was therefore also included in the model. For basal area (G) estimations, the following model (Eq. 3) was used:

$$\ln(G) = \beta_0 + \beta_1 \ln(h_X) + \beta_2 \ln(D_v) + \beta_3 r_X + \varepsilon \qquad (3)$$

The laser height percentile and ratio of laser height percentiles yielding the lowest residual sum of squares were selected. For stem volume (*V*) estimations, a similar model (Eq. 4) was used:

$$\ln(V) = \beta_0 + \beta_1 \ln(h_X) + \beta_2 \ln(D_v) + \beta_3 r_X + \varepsilon \qquad (4)$$

The laser height percentile and ratio of laser height percentiles yielding the lowest residual sum of squares were selected. Both basal area and stem volume were transformed to original scale by using the exponential function. Correction was then done for logarithmic bias by multiplying the predicted values by the ratio between field-measured values and predicted values.

3. RESULTS AND DISCUSSION

The RMSE at stand level was 5.0% (0.8 m) for mean tree height (Figure 1), 8.9% (1.9 cm) for mean stem diameter (Figure 2), 12.5% (3.0 m^{2}ha^{-1}) for basal area (Figure 3), and 14.1% (28 m^{3}ha^{-1}) for stem volume estimation (Figure 4).

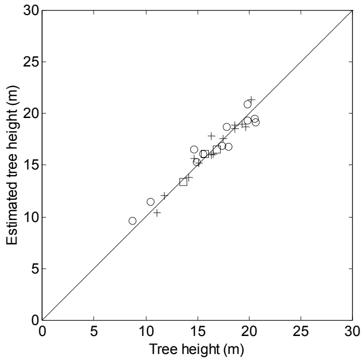

Figure 1. Estimated mean tree height (h_L) versus field measured mean tree height, 29 forest stands, Norway spruce (o), Scots pine (+), and deciduous trees (□).

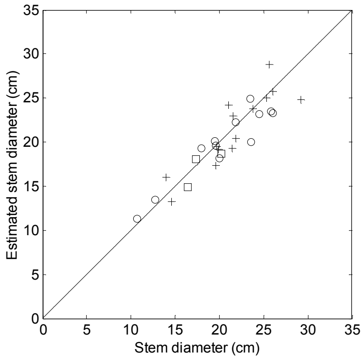

Figure 2. Estimated mean stem diameter (d_L) versus field measured mean stem diameter, same labels as in Figure 1.

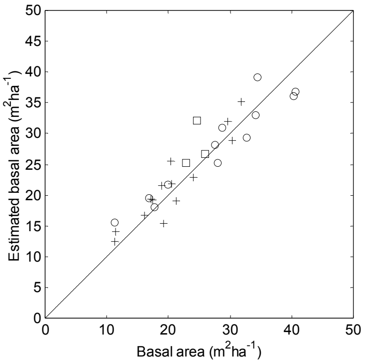

Figure 3. Estimated basal area (*G*) versus field measured basal area, same labels as in Figure 1.

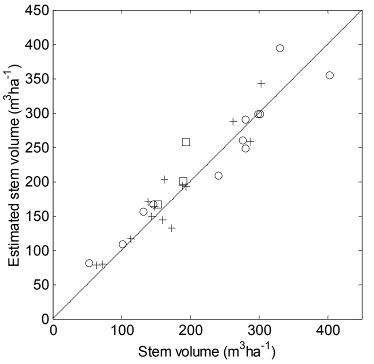

Figure 4. Estimated stem volume (*V*) versus field measured stem volume, same labels as in Figure 1.

The RMSE values are affected by the sampling error of the field inventory. The random error of the remote sensing method (denoted as *s*) can be estimated if one assumes that the errors of the remote sensing method and the field inventory are independent. The variance of *s* is calculated as the variance of the differences between remote sensing and field inventory estimates minus the variance of the random error from the field inventory (Ståhl, 1992). The latter variance is estimated as the mean of the variances to calculated forest stand mean values and multiplied by 0.5 to account for the lower errors of a systematic sampling design according to empirical studies by Lindgren (1984). The random error *s* was estimated to be 4.2% (0.7 m) for mean tree height, 8.1% (1.7 cm) for mean stem diameter, 10.4% (2.5 m^{2}ha^{-1}) for basal area, and 10.9% (22 m^{3}ha^{-1}) for stem volume.

4. CONCLUSIONS

The results from this study imply that mean tree height, mean stem diameter, basal area, and stem volume can be estimated using laser scanning with higher accuracies than are usually achieved using traditional methods.

REFERENCES

Åge, P. J. 1985. Forest inventory: photo interpretation. National Land Survey, Report 1985:13. Gävle, Sweden.

Jonsson, B., Jacobsson, J. & Kallur, H. 1993. The Forest Management Planning Package. Theory and application. Swedish University of Agricultural Sciences, Studia Forestalia Suecica 189. Uppsala, Sweden.

Lindgren, O. 1984. A study of circular plot sampling of Swedish forest compartments. Swedish University of agricultural Sciences, Section of Forest Mensuration and Management, Report 11. Umeå, Sweden.

Magnussen, S. & Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28, pp. 1016-1031.

Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L. & Hendrix, C. J. 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66, pp. 1367-1371.

Næsset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52, pp. 49-56.

Næsset, E. & Bjerknes, K. O. 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 78, pp. 328-340.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, pp. 88-99.

Söderberg, U. 1992. Functions for forest management, height, form height and bark thickness of individual trees. Swedish University of Agricultural Sciences, Dept. of forest survey, Report 52. Umeå, Sweden.

Ståhl, G. 1992. A study on the quality of compartment-wise forest data acquired by subjective inventory methods. Swedish University of Agricultural Sciences, Dept. of Forest Mensuration and Management, Report 24. Umeå, Sweden.

ACKNOWLEDGEMENTS

Fotonor (now Blom Norkart Mapping) delivered the laser data. The authors would like to thank Heather Reese for improving the language of the paper. We would like to thank the field personal at Dalarna-Gävleborg forestry board. The project was financed by Dalarna County's fund for recession compensation within private forestry.