Assessing effects of sample plot positioning errors on biophysical stand properties derived from airborne laser scanner data

T. Gobakken[1], E. Næsset[1]

[1] Norwegian University of Life Sciences, Dep. of Ecology and Natural Resource Management - (terje.gobakken, erik.naesset)@umb.no

Abstract

Canopy height distributions were created from small-footprint airborne laser scanner data with an average sampling density of 0.9 points m^{-2} collected over 100 georeferenced field sample plots and 57 stands. Height percentiles, mean and maximum height values, coefficients of variation of the heights, and canopy density at different height intervals above the ground were computed from the laser-derived canopy height distributions of the first return data. The plot positions were altered randomly by means of Monte Carlo techniques. The standard deviation (SD) of the differences for various metrics derived from the canopy height distributions between incorrect plot positions and ground-truth positions were compared. The SD increased with increasing plot position error.

The effects of sample plot position error on the accuracy of mean tree height (h_L), stand basal area (G), and stand volume (V) predicted at stand level using a two-stage procedure combining field training data and laser data were assessed. The standard deviation of the differences increased with increasing plot position errors. Except for h_L the largest increase in median SD was found for mature forest on poor sites. The effects of plot position error seem to be more pronounced for G and V compared to h_L.

Keywords: Airborne laser scanner, GPS, position error, sample plot

1. Introduction

The aim of forest inventories at a property/compartment level is to provide data for forest planning and management, and they are often carried out according to an area-based approach, which implies that the individual forest stands are the basic units of the inventories. During the last 15-20 years, several experiments have been carried out in order to determine various biophysical strand properties, such as mean tree height, basal area, and timber volume based on airborne laser scanning (ALS) measurements (Means *et al.* 2000; Næsset 1997, 2002). The operational area-based forest stand inventory method adopted in Scandinavia utilizes mainly ALS data in a two-stage procedure proposed by Næsset and Bjerknes (2001) and Næsset (2002). In a first stage, georeferenced field training plots with corresponding ALS data are used to develop empirical relationships between various metrics derived from the laser data and biophysical properties measured in field. These relationships provide, in the second stage, corresponding predicted values of each stand from the laser data.

Thus, accurate geographical co-registration of ALS data and field plots is essential for accurate predictions of biophysical stand properties. If the remotely sensed data and the field data are poorly co-registered, the basic laser-derived metrics will be subject to errors. If the basic laser-derived metrics are subjected to errors, it is likely that also the resulting stand predictions of the biophysical variables will be affected. However, since the biophysical properties are predicted from equations that are combinations of several laser variables, the effects of position errors can hardly be quantified by just assessing effects of one laser-derived metric at a time.

The Global Position System (GPS) technology is usually applied to obtain the geographical location of the field observations. GPS can provide timely and accurate spatial data under "clear sky" conditions. However, in forested landscapes, biological and topographic obstacles tend to degrade GPS position accuracy. Sophisticated GPS receivers are expensive to acquire and the logistics and data management in differential positioning is time-consuming in forest inventory applications, particularly in remote areas where it takes time to collect data and maintain base reference stations far from the field locations in the forest. In an operational context, there is a trade-off between costs and accuracy. One would often seek an accuracy that is "good enough" in order to save costs and simplify the work. In Norway, for example, the GPS accuracy for the National forest inventory (NFI) plots is expected to be within 10 m for 99% of the plots (Gjertsen 2007).

Gobakken and Næsset (2008b) assessed the effects of positioning errors and sample plot size on biophysical stand properties derived from ALS. They found significant effects of plot position errors and the effects were larger for poor sites with more scattered trees compared to productive sites with denser canopies and more evenly distributed of the trees. However, the study was limited to only one test site. The present study was carried out to verify the main findings based on data from another forest area. Thus, the objectives of this study were to assess the effects of field plot position errors (1) on selected laser-derived metrics and (2) on three important biophysical stand properties of interest in forest inventories predicted from the ALS data, i.e., mean tree height, stand basal area, and timber volume. Nine different levels of field plot position errors were assessed. The position errors were analysed using Monte Carlo techniques. The accuracy of the predicted biophysical properties was evaluated using an independent validation dataset.

2. Method

2.1 Study area

A forest area in the municipality of Krødsherad (60°10'N 9°35'E, 130-660 m a.s.l.), of about 6500 ha was selected for this study. The main tree species in the area were Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.). Further details can be found in Næsset (2004). This study was based on two different field data sets, i.e., (1) sample plots and (2) forest stands. The sample plots were used to assess the effects of different laser point densities on laser-derived metrics and to develop regression models for the three biophysical properties of interest. The forest stands were used to assess how sample plot position error affected the stand predictions of the three biophysical properties.

2.2 Sample plots

In total, 100 sample plots were distributed systematically throughout the entire 6500 ha study area according to a regular grid. The plots were divided into three strata according to age class and site quality of the stands in which they were located. The area of the sample plots was 232.9 m^2. The measurements were carried out during the summer 2001 (Næsset, 2004b). On each plot, all trees with d_{bh} >10 cm were callipered. The d_{bh} was recorded in 2 cm classes. Basal area (G) was computed as the basal area per hectare of the callipered trees. The heights of sample trees were measured by a Vertex hypsometer. Mean height of each plot was computed as Lorey's mean height (h_L), i.e., mean height weighted by basal area. Volume of each tree was computed by means of volume equations of individual trees (Brantseg, 1967; Braastad, 1966; Vestjordet, 1967) with height and diameter as predictor variables. Total plot volume (V) was computed as the sum of the individual tree volumes.

Finally, to synchronize the h_L, G, and V values to the date the laser data were acquired the individual plot values were prorated by means of growth equations (Blingsmo, 1984; Braastad, 1975; Braastad, 1980; Delbeck, 1965). The prorated values were used as ground-truth. A summary of the ground-truth sample plots data is displayed in Table 1. Differential Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) were used to determine the position of the centre of each sample plot.

2.3 Stand inventory

In total, 57 large test plots distributed on the three pre-defined strata located in subjectively selected stands were selected. Ground reference data for the test plots were collected during summer 2001 following similar measurement and computational procedures as for the sample plots, see above. Each plot was a quadrate or had a shape close to a quadrate, and the size ranged from 2869 to 4219 m^2. The large test plots are hereafter denoted as stands. A summary of the ground-truth stand data is displayed in Table 1.

Table1: Summary of field inventory of sample plots and stands [a].

	Small plots (233 m^2)				Stands			
Characteristic	Range			Mean	Range			Mean
Young forest – stratum I				(n=30)				(n=19)
h_L (m)	8.1	-	19.5	13.4	10.5	-	19.7	15.4
G (m^2ha^{-1})	6.4	-	62.4	25.0	12.0	-	41.9	27.3
V (m^3ha^{-1})	26.8	-	617.6	182.4	64.0	-	329.6	212.5
Mature forest, poor site quality – stratum II				(n=37)				(n=19)
h_L (m)	9.9	-	25.4	15.5	12.2	-	20.1	15.6
G (m^2ha^{-1})	5.6	-	42.7	22.6	12.0	-	31.5	21.1
V (m^3ha^{-1})	29.6	-	446.3	173.7	83.0	-	292.7	162.3
Mature forest, good site quality – stratum III				(n=33)				(n=19)
h_L (m)	15.0	-	26.0	21.0	15.7	-	24.4	20.3
G (m^2ha^{-1})	15.5	-	57.0	34.2	21.9	-	37.7	29.8
V (m^3ha^{-1})	116.8	-	674.8	338.0	186.0	-	378.9	286.6

[a] h_L=Lorey's mean height, G=basal area, V=volume.

2.4 Laser scanner data

A fixed-wing aircraft carried the ALTM 1210 laser scanning system (Optech, Canada). The laser scanner data were acquired in the period between 23 July and 1 August 2001 (cf. Næsset 2004). The average sampling density was 0.9 m^{-2}. A complete post-processing of the laser data was undertaken by the contractor (Blom Geomatics, Norway). A triangulated irregular network (TIN) was generated from the planimetric coordinates and corresponding height values of the individual terrain ground points. All the return observations (points) were spatially registered to the DTM according to their coordinates. Terrain surface height values were computed for each point by linear interpolation from the DTM. The relative height of each point was computed as the difference between the height of the return and the interpolated terrain surface height. Only the first returns were used for further analysis. The first return data were spatially registered to the field plots and stands.

2.5 Simulation of plot position error

To investigate the effects of position errors on metrics derived from the laser data for each field plot, the position errors were simulated. This was done by introducing a horizontal shift in the field plot coordinates from the ground-truth positions prior to extracting laser points inside the plots. Horizontal shifts from true positions of 0.5, 1, 2, 3, 4, 5, 10, 15, and 20 m, respectively, were used. For each of these nine fixed levels of position errors, the error-contaminated positions were computed 500 times in a Monte Carlo simulation by using a randomly selected angle. 500 repetitions were used so that we could control the random effects in the simulations.

2.6 Computations of laser metrics

For each sample plot and stand inventoried in field, height distributions were created for those laser points that were considered to belong to the tree canopy, i.e., points with a height value of >2 m. Canopy height percentiles at 10% (h_{10}), 50% (h_{50}), and 90% (h_{90}) were computed. In addition, also the maximum (h_{max}) and mean values (h_{mean}), and the coefficient of variation (h_{cv}) of the canopy height distributions were computed. Furthermore, several measures of canopy density were derived. Canopy density was computed as cumulative densities of 10 different vertical layers of equal height (Næsset 2004). The height of each layer was defined as one tenth of the distance between the 95% percentile and the lowest canopy height (>2 m) (Gobakken and Næsset 2008a). The cumulative canopy densities were then computed as the proportions of laser echoes above layer # 0 (>2 m), 1, . . ., 9 to total number of echoes. The cumulative densities for layer # 1 (d_1), # 5 (d_5), and # 9 (d_9) were selected for further studies.

To assess how sample plot position error influenced on the stability of laser-derived metrics, differences between corresponding metrics derived for the plots with error-contaminated positions and in true positions were computed for each sample plot. The standard deviations of the differences were then calculated for each of the 500 repetitions in the Monte Carlo simulation. Separate comparisons were carried out for the three strata. Further explorative data analysis of the Monte Carlo repetitions was performed using graphical methods, i.e., box-and-whisker plots (R Development Core Team 2006; Tukey 1977).

2.7 Predictions of biophysical stand properties

To assess the accuracy of laser-based predictions of mean tree height, basal area, and volume based on different field plot position errors, we followed the two-step procedure proposed by Næsset & Bjerknes (2001) and Næsset (2002) (1) by relating the three biophysical properties of interest to the laser data of the sample plots using regression analysis, and (2) by applying the estimated regression models to predict corresponding values of the test stands. As an additional step, (3) the differences between predicted values of the biophysical stand properties and ground-truth values were computed. The standard deviations of the differences were also calculated. The predicted values were restricted to predefined reasonable maximum values for the forest area in question.

As a preparation for the simulations, we wanted to determine a fixed set of explanatory variables to avoid effects of altering the variables in the regression models. Thus, variables to be included in the models were determined using the ground-truth field plot positions. The estimation of regression models was based on the height and density-related metrics derived from the first return height distributions as candidate explanatory variables. In the regression analysis, multiplicative models were estimated as linear regressions in the logarithmic variables.

The effects of field plot position error on the estimation and prediction of biophysical stand properties were assessed by means of Monte Carlo techniques as described above. The entire sequence in steps (1)–(3) above was repeated 500 times for each of the nine plot position errors. Thus, 500 × 9 estimates of the mean differences between predicted biophysical stand properties and ground-truth values and corresponding estimates of the standard deviations of the differences were derived. As a reference, the mean differences and the SD values when using ground-truth plot positions were calculated for the respective strata.

3. Result

3.1 Effects of plot position error on the stability of laser-derived metrics

Figure 1 shows the standard deviation (SD) of the differences between corresponding laser-derived metrics computed for the plots in error-contaminated positions and in true positions for different forest types (strata I-III) over the 500 repetitions of erroneous positions using the Monte Carlo procedure. The standard deviation of the differences for the height percentiles (h_{10}, h_{50}, and h_{90}), the three height-related metrics maximum laser canopy height (h_{max}), arithmetic mean laser canopy height (h_{mean}), and coefficient of variation of laser canopy heights (h_{cv}), and for the density-related metrics (d_1, d_5, and d_9) increased with increasing plot position error. The amount of extreme observations increased with increasing plot position error.

3.2 Effects of plot position errors on predicted biophysical stand properties

The effects of using regression models estimated from plots with error-contaminated positions were assessed by using the estimated regressions and the two-step procedure to compute stand mean values of the three biophysical properties in 57 forest stands. As a reference, differences were computed assuming true plot positions. Using ground true plot positions the mean differences for h_L, G, and V were -1.4, 9.7, and 9.5% for stratum I, -1.9, 3.4, and 4.3% for stratum II, and -1.3, 8.8, and 6.9% for stratum III, respectively. The box plots illustrating the results of the simulations with error-contaminated plot locations show that the variation in mean difference between the 500 Monte Carlo repetitions in general increased with increasing plot location errors even if the median mean difference decreased for some of the comparisons (Fig. 2, Left).

Using ground-truth plot positions, the standard deviations (SD) of the differences for h_L were 6.6, 3.5, and 3.2% for strata I, II, and III, respectively (Fig. 2, Right). The SDs were 14.5, 9.3, and 12.4% for G and 19.1, 10.2, and 12.7% for V for strata I, II, and III, respectively, using ground-truth plot positions. The standard deviation of the differences increased with increasing plot position errors. Except for h_L, the largest increase in median SD was found for mature forest on poor sites (stratum II). The effects of plot position error seem to be larger for G and V compared to h_L.

4. Discussion

The results in this study are in line with Gobakken and Næsset (2008b). The main findings in the present study were that on poor sites where there normally are few stems it is important to have accurate plot positions to obtain accurate estimates of *V* and *G*. Improved GPS positioning by e.g. longer time periods of GPS data collection might be considered for more variable forests in order to reduce positional errors. However, there will often be fewer biological obstacles providing good conditions for GPS data collection in open forests and normally relatively precise GPS positions would be expected in such forests compared to dense forests.

Furthermore, cost-plus-loss analyses (cf. Eid *et al.* 2004) where the total costs of the inventory as well as the expected economic losses as a result of future incorrect decisions due to errors in measurements are considered, should be applied to evaluate the effects of plot position error. Cost-plus-loss comparisons between inventories with different positional accuracies might find that the requirements for positional accuracy is lower in variable and open forests compared to fully stocked and more even forests.

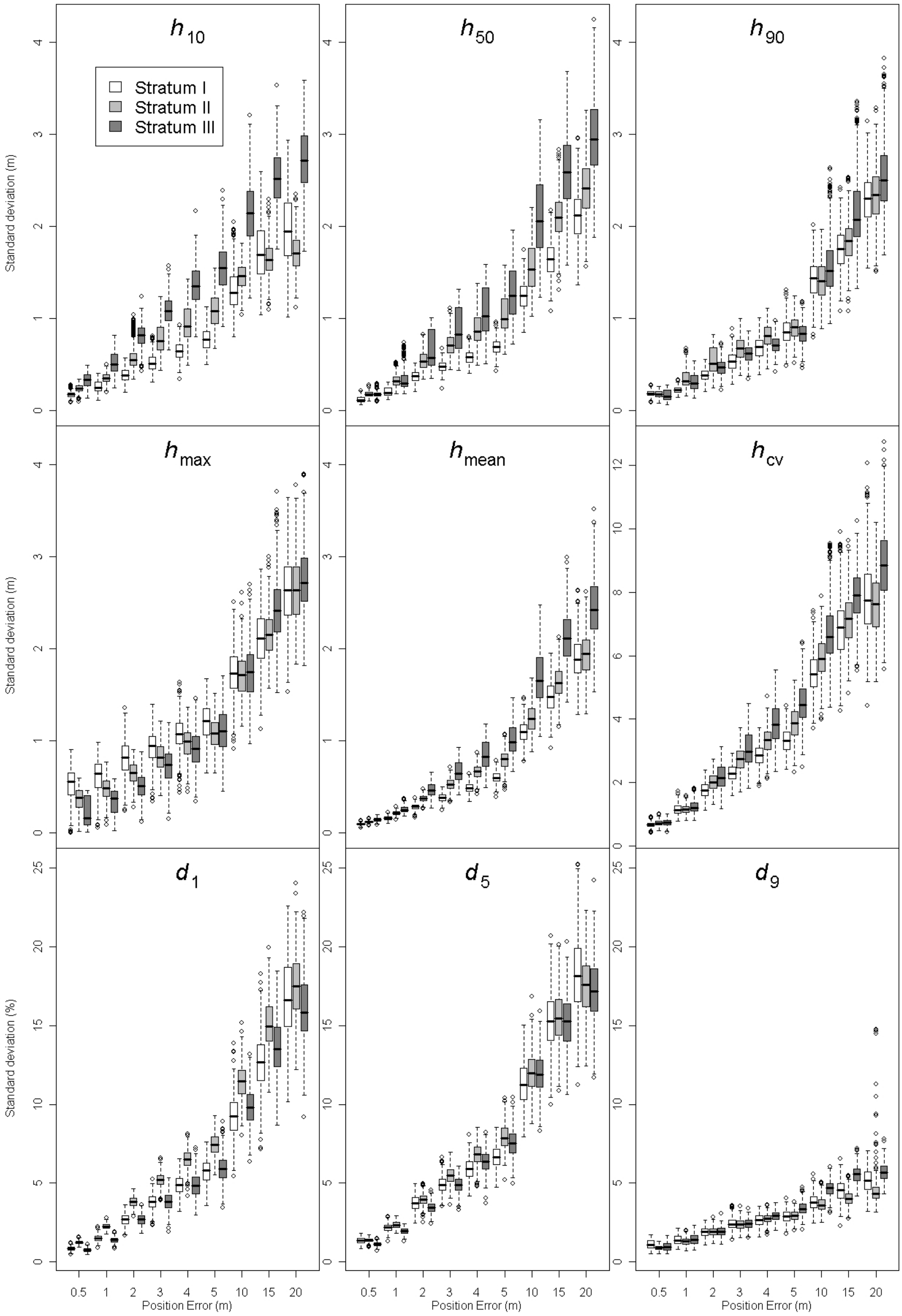

Figure 1: Standard deviation of the differences between laser-derived metrics (see text) of plots with error-contaminated positions generated with Monte Carlo simulation (500 repetitions) and true positions for different forest types (young forest=stratum I; mature forest with poor site quality= stratum II; mature forest with good site quality=stratum III)[1]

[1] The box-and-whisker plots show first and third quartile as the box ("hinges"), the median as the horizontal line dividing the box and extreme values as points.

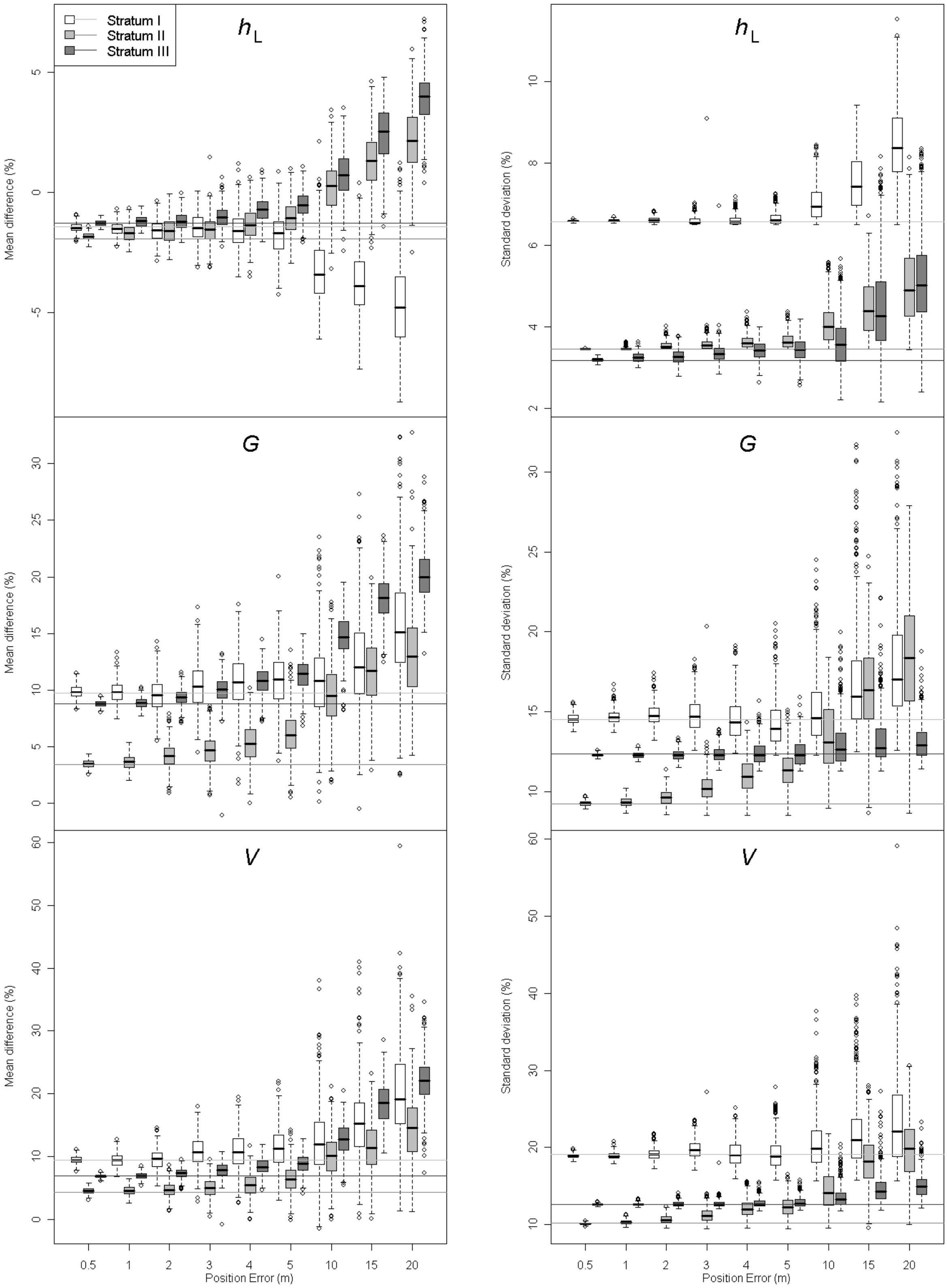

Figure 2: Mean difference (*left*) and standard deviation of the differences (*right*) between predicted and observed values of Lorey's mean height (h_L) (*top*), basal area (*G*) (*middle*), and volume (*V*) (*bottom*) in forest stands of different types (young forest=stratum I; mature forest with poor site quality= stratum II; mature forest with good site quality=stratum III) based on prediction models estimated with sample plots assuming different levels of position errors of the plot locations. Statistics for each level of position error is computed from the outcome of the 500 Monte Carlo repetitions[1]. As a reference, the horizontal lines indicate the results when using ground-truth plot positions for stratum I, II, and III, respectively

The field-measured ground-truth plot positions were treated as if they were free from errors in this study. In fact, the computed ground-truth plot coordinates had an expected average accuracy of approximately 0.2 m (Næsset 2004), however, this miss-location should only have a marginal influence of the major findings.

Gobakken and Næsset (2008b) also found that larger plot sizes to a certain extent can compensate for sample plot position errors. Consequently more research is needed to find the optimal combination of field plot size and requirements for plot position accuracy.

To conclude, the results have shown that the accuracy of positions of the sample plots are an important factor affecting precision of forest inventory based on ALS data.

References

Eid, T., Gobakken, T., & Næsset, E. 2004. Comparing stand inventories based on photo interpretation and laser scanning by means of cost-plus-loss analyses. *Scandinavian Journal of Forest Research*, 19, 512-523

Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment*, 110, 420-430

Gobakken, T., & Næsset, E. 2008a. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 38, 1095-1109

Gobakken, T., & Næsset, E. 2008b. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne Laser scanner data. *In submission*

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

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61, 246-253

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, 88-99

Næsset, E. 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19, 164-179

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, 328-340

R Development Core Team 2006. R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing

Tukey, J.W. 1977. Exploratory data analysis: Addison-Wesley Reading