Silva Fennica
Silva Fennica 46(1) research articles
www.metla.fi/silvafennica · ISSN 0037-5330
The Finnish Society of Forest Science · The Finnish Forest Research Institute

The Suitability of Leaf-off Airborne Laser Scanning Data in an Area-based Forest Inventory of Coniferous and Deciduous Trees

Maria Villikka, Petteri Packalén and Matti Maltamo

Villikka, M., Packalén, P. & Maltamo, M. 2012. The suitability of leaf-off airborne laser scanning data in an area-based forest inventory of coniferous and deciduous trees. Silva Fennica 46(1): 99–110.

This study examined the suitability of airborne laser scanner (ALS) data collected under leaf-off conditions in a forest inventory, in which deciduous and coniferous trees need to be separated. All analyses were carried out with leaf-on and leaf-off ALS data collected from the same study area. Additionally, aerial photographs were utilized in the Nearest Neighbor (NN) imputations. An area-based approach was used in this study. Regression estimates of plot volume were more accurate in the case of leaf-off than leaf-on data. In addition, regression models were more accurate in coniferous plots than in deciduous plots. The results of applying leaf-on models with leaf-off data, and vice versa, indicate that leaf-on and leaf-off data should not be combined since this causes serious bias. The total volume and volume by coniferous and deciduous trees was estimated by the NN imputation. In terms of total volume, leaf-off data provided more accurate estimates than leaf-on data. In addition, leaf-off data discriminated between coniferous and deciduous trees, even without the use of aerial photographs. Accurate results were also obtained when leaf-off ALS data were used to classify sample plots into deciduous and coniferous dominated plots. The results indicate that the area-based method and ALS data collected under leaf-off conditions are suitable for forest inventory in which deciduous and coniferous trees need to be distinguished.

Keywords forest inventory, group-specific estimates, leaf-off data, LIDAR
Addresses University of Easten Finland, Dept. of Forest Sciences, Joensuu, Finland
E-mail petteri.packalen@uef.fi
Received 8 July 2011 **Revised** 12 September 2011 **Accepted** 22 September 2011
Available at http://www.metla.fi/silvafennica/full/sf46/sf461099.pdf

1 Introduction

Airborne laser scanning (ALS) provides an accurate three-dimensional description of the surveyed area. ALS data is particularly useful in terrain modelling (Ahokas et al. 2008) and is also well suited to forest inventory purposes (Næsset 2002). There are two ALS-based approaches that are used for forest inventory. In the individual tree delineation approach, individual trees are first recognised and delineated from high resolution ALS data with multiple laser points per square metre. The tree-level characteristics are estimated and then summed together in order to get stand-level estimates (e.g. Persson et al. 2002). In the area-based method, which relies on the height and density distributions of ALS points, the forest characteristics are estimated directly on the plot or sub-stand level (e.g. Næsset 2002, Holmgren 2004). The area-based method is already in operational use in Nordic countries (e.g. Næsset 2007, Maltamo et al. 2009). The ALS data used in forest inventories have traditionally been collected under leaf-on conditions during the summer although research has also been conducted with leaf-off data (e.g. Brandtberg et al. 2003, Næsset 2005, Hill and Broughton 2009).

The use of ALS data in national level terrain modelling is increasing (Liang et al. 2007). The leaf-off ALS data is commonly preferred in order to minimize bias caused by vegetation. For example, leaves and branches may reduce the accuracy of elevation information (Hodgson et al. 2003). On the other hand, the understorey vegetation is considered to be a more serious problem than tree canopies in Finland (Sirkiä and Laaksonen 2009). The National Land Survey of Finland is also using ALS data in digital terrain model (DTM) production (Ahokas et al. 2008). The technical requirements for the ALS data in national terrain modelling are almost identical to the requirements of forestry applications, but in terrain modelling the primary aim is to acquire leaf-off data (Liang et al. 2007). However, using the same (leaf-off) data in national terrain modelling and forestry applications would mean significant cost savings in both campaigns.

The leaf-off season without snow is relatively short in the boreal conditions in Finland, and its length is not known beforehand, which further increases the uncertainty of leaf-off data collection. From the viewpoint of remote sensing based forest inventory the optimal arrangement is that there is only a very short time period between aerial photograph acquisition, which has to be conducted during the summer, and ALS acquisition. Generally, the shorter the time difference the less likely are harvestings or other silvicultural operations between ALS data and aerial photograph acquisitions.

Canopy conditions remain constant throughout the snowless period of a year for coniferous species. This indicates that, in pure coniferous stands, an ALS-based inventory system should work equally well with either leaf-off or leaf-on data. In the case of deciduous trees, the response differs considerably between leaf-off and leaf-on data. Several studies have reported that, in coniferous forests, ALS data have a tendency to underestimate the tree height (e.g. Maltamo et al. 2004). The underestimation in coniferous boreal forests has usually been attributed to insufficient sampling density, which means that it is unlikely that some ALS point hit the actual tree top (e.g. Næsset and Økland 2002). Gaveau and Hill (2003) showed that, in deciduous forests, laser pulses penetrate into the canopy before an echo is detected, causing the underestimation of canopy height. The amount of penetration depends on the closure of the upper canopy surface, which is affected, for example, by species, crown structure, crown density and leaf area index. Under leaf-off conditions, the laser pulses penetrate deeper into deciduous tree crowns than under leaf-on conditions, and a large number of the laser pulses are reflected directly from the ground and lower vegetation (Hill and Broughton 2009).

Although forestry research has been focused on the utilization of leaf-on data in several applications, there are some studies in which leaf-off data have been tested (e.g. Næsset 2005). Many previous studies support the use of leaf-off data for forest inventory purposes. Most of these studies have concentrated on the tree species classification at individual tree level. Results indicate that leaf-off data discriminates between tree species at tree level better than leaf-on data (e.g. Brandtberg et al. 2003). The potential of leaf-off data in other forestry-related applications, such as mapping the understorey, has also been demonstrated (Hill and Broughton 2009).

Brandtberg et al. (2003) were the first to show the potential of high-density leaf-off ALS data in deciduous forests. The data were evaluated for tree crown detection, tree height measurements and species classification. The overall classification accuracy of three deciduous species (oak, red maple and yellow poplar) was 60%, and height estimates based on ALS data were very accurate. The presence of snow on the ground did not appear to cause problems, although snow has a high reflectance for wavelengths in the near infrared portion of the spectrum. Brandtberg (2007) further improved the classification methodology used in Brandtberg et al. (2003).

Liang et al. (2007) studied the classification of trees into deciduous and coniferous using the range difference between first and last echoes. The study was conducted on a suburban site in Finland using leaf-off ALS data. The classification was based on the assumption that the first echoes are reflected from the tree tops in the case of both deciduous and coniferous trees, and that the last echoes are reflected from the tree tops in the case of coniferous trees, and from the ground below in the case of deciduous trees. The classification was made at an individual tree level and an accuracy of 89% was achieved.

Næsset (2005) seems to be the only study where the leaf-on and leaf-off data have been compared in the estimation of forest attributes by area based method. Mean tree height, basal area and volume were estimated in order to demonstrate the effect of leaf-off conditions. Separate regression models were constructed with data collected under leaf-on and leaf-off conditions. The estimates were of similar accuracy under leaf-on and leaf-off conditions. Næsset (2005) also studied how explanatory variables derived from ALS data were affected by leaf-on and leaf-off canopy conditions. He concluded that the last echoes were more affected by canopy conditions than the first echoes, and that the canopy height derived from ALS data had higher variability under leaf-off conditions.

Full-waveform ALS data collected under leaf-off conditions have also been used in some studies (Reitberger et al. 2008, Hollaus et al. 2009). Hollaus et al. (2009) studied the classification of individual trees in a mixed woodland in the eastern part of Austria. First, deciduous and coniferous trees were classified and then the classification was made by species (spruce, larch, beech). The classification was based on the scattering mechanisms of different species under leaf-off conditions. The classification accuracy of coniferous and deciduous trees was 83%. A species-specific classification had 75% accuracy. In a similar way, Reitberger et al. (2008) classified individual trees into coniferous or deciduous trees using full-waveform data. The classification accuracy was 85% with leaf-on data and 96% with leaf-off data.

Typically, boreal forests are in Finland dominated by coniferous trees. Deciduous species rarely form pure stands, but still usually exist as minor tree species in coniferous forests. In Finland, the information on species-specific stand characteristics is essential in forest management. This also means that, in remote sensing-based forest inventories, it is necessary to distinguish tree species. Usually, the separation of pine, spruce and deciduous trees is required in remote sensing-based forest inventories. For such species-specific inventory purposes, Packalén and Maltamo (2007) developed an inventory system which is based on a combination of ALS data, aerial photographs and field measurements. Aerial photograph were used in order to improve the discrimination between tree species. The estimation was carried out by Nearest Neighbor (NN) imputation method at plot level and applied by grid cells to the whole inventory area (Packalén and Maltamo 2007).

The overall aim of this study was to test the suitability of area-based method and ALS data collected under leaf-off conditions in a forest inventory in which deciduous and coniferous trees need to be separated. All analyses were done with leaf-off and leaf-on data in the same study area. Specific aims were to investigate what is the effect of the dominant species type in terms of accuracy, is it appropriate to mix leaf-off and leaf-on data, and how well leaf-on and leaf-off ALS data with and without aerial photographs suit for the estimation of plot volumes by deciduous and coniferous trees.

2 Materials

2.1 Study Area and Field Data

The study area is located in Finland near to the city of Kuopio (Fig. 1). It is a typical Finnish, managed, boreal forest area which is dominated by Scots pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* (L.) Karst.). Deciduous trees are in a minority in the tree stock. The study area covers approximately 5000 hectares.

Field measurements were carried out in summer 2008. The plot locations were pre-selected using existing stand register data. This stratification took into account the development class, dominant tree species, basal area and mean height. The particular aim was to ensure that there are enough deciduous tree dominated plots for the purposes of this study. The plots that were placed on a stand boundary in a pre-selection phase were moved inside stands. The sample plots were circular, with a radius of 9 metres. A total of 192 sample plots were measured. A survey grade Global Positioning System was used to determine the position of the plot centres, with an accuracy of about 1 metre under canopy cover (the accuracy of the comparable positioning system under canopy cover was tested nearby, unpublished). Each tree, with a diameter at breast height (DBH) of at least 5 cm, was measured. The DBH, tree species, storey class and whether the tree was living or dead, were recorded for each tree. The height of basal area median tree of each species in each storey class was measured in each plot.

Height measurements were used to estimate the parameters of Näslund's (1937) height models by tree species using a mixed effect modelling with a random constant and coefficient for each plot. The models with predicted plot effects were used to predict heights for tally trees. The volumes of individual trees were calculated as a function of DBH and tree height using the species-specific models by Laasasenaho (1982). Finally, the tree volumes were summed together and scaled up to the hectare level. In addition, the dominant height was calculated for each plot.

The plots were divided into two groups in terms of the volume of coniferous and deciduous trees. Plots with a volume of coniferous species ≥ 50% of the total volume were classified as coniferous, and the remaining plots were classified as deciduous-dominated plots. 69% of the plots were classified as coniferous and 31% as deciduous-dominated. This classification is needed since the main interest of this study is to examine how the leaf-off data affect the analysis of deciduous plots. In the deciduous group, the proportion of deciduous trees varied between 50% and 100%, the average proportion of deciduous trees being 85%. In the coniferous group, the proportion of coniferous trees varied between 54% and 100%, with the average proportion of coniferous trees being 90%. The average proportions of pine and spruce were 30% and 60%, respectively. A summary of the ground reference data is presented in Table 1.

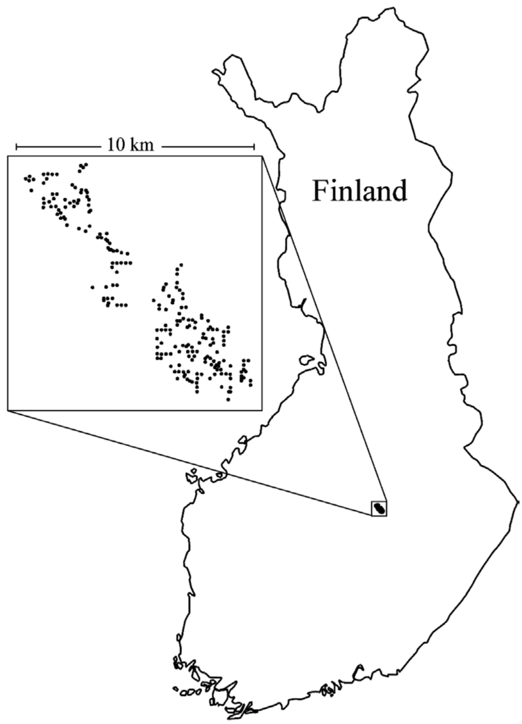

Fig. 1. Location of the test area and placement of sample plots.

Table 1. Summary of sample plot data. Volumes are in m^3ha^{-1}.

	All plots (N=192)		Coniferous plots (N=132)*		Deciduous plots (N=60)**	
	Mean	Range	Mean	Range	Mean	Range
Conif. volume	148.3	0.0–641.3	201.5	0.0–641.3	31.4	0.0–280.4
Decid. volume	56.4	0.0–284.2	22.2	0.0–144.3	131.6	21.2–284.2
Total volume	204.7	24.5–752.7	223.7	28.2–752.7	162.9	24.5–564.6
Dom. height	20.7	8.3–32.1	20.7	8.30–32.1	20.9	11.1–31.6
Proportion of total volume by coniferous and deciduous plots						
Coniferous	66.4	0.0–100.0	89.5	54.3–100.0	15.5	0.0–49.7
Deciduous	33.6	0.0–100.0	10.5	0.0–45.8	84.5	50.3–100.0

*Volume of coniferous species ≥ 50% of total volume
**Volume of deciduous species > 50% of total volume

2.2 Remote Sensing Data

The leaf-off ALS data were collected on 16–17 May, 2008 and the leaf-on data on 31 August and 1 September 2008. Both datasets were collected with the same Optech ALTM Gemini instrument using the same configuration. The test site was measured from an altitude of 2,000 m above ground level (a.g.l.) using a field of view of 28 degrees and a side overlap of 20%. The flight speed was 75 ms^{-1} and the pulse frequency was set to 50 kHz. This resulted in a swath width of approximately 1000 m and a nominal sampling density of about 0.6 m^{-2}.

Optech ALTM Gemini captures four range measurements for each pulse, but here the measurements were reclassified to represent the first and last pulse datasets. The first pulse data contained the echo categories 'first of many' and 'only', while the last pulse data contained 'last of many' and 'only' echoes. Intermediate echoes were ignored. The DTM was created using the data collected under leaf-on conditions. Only the last pulse data were used for DTM creation. First ALS points were classified as ground points and other points as explained in Axelsson (2000). A DTM raster with a cell size of 2 metres was then interpolated by Delaunay triangulation (Fowler and Little 1979).

The aerial photographs were taken with a Vexcel UltraCamD digital aerial camera under leaf-on canopy conditions. Owing to poor weather conditions, aerial photography was carried out in two stages, on 29 July and 8 September 2008. The images were taken at an altitude of 5800 metres a.g.l., with both a sidelap and endlap of 60%. Final multispectral images were produced with the spatial resolution of a panchromatic band by applying a pan-sharpening process. Although the sensor consists of three colours (red, green, blue) and near-infrared (NIR) bands, only pan-sharpened NIR, red and green bands were used in this study. Finally, aerial photographs were orthorectified to a pixel size of 0.5 m using the DTM generated from the ALS data.

3 Methods

3.1 Independent Variables

All ALS metrics were calculated separately for the first and last pulse data. The first step was to convert ALS height to above-ground scale by subtracting the DTM from the orthometric heights. Separate height distributions were calculated for the first and last pulse data for each plot. All the points, including ground points, were included in the analysis. Canopy height percentiles and proportional canopy densities were calculated from the first and last pulse height distributions for both ALS datasets. The calculated canopy height percentiles were 1%, 20%, 40%, 60%, 80%, 95% and 100%. In addition, the average canopy height and standard deviation were calculated. The canopy densities, i.e. proportion of laser points accumulated at certain heights, were calculated for 5%, 30%, 50% and 70% percentiles. Similar types of ALS metrics have been used as independent variables in many previous studies (Næsset 2002; Packalén and Maltamo 2007).

Independent variables representing spectral and textural features were calculated from the aerial photographs. The aim of using aerial photographs was to improve the discrimination of tree species. Spectral features contained mean, median and standard deviation which were calculated by bands for each sample plot. Textural features were calculated from the Grey Level Co-occurrence Matrix (GLCM) using a method presented by Haralick et al. (1973). Only one GLCM was constructed for each sample plot (for each direction by bands), thus the moving window technique, usually applied to calculate textures around the local neighbourhood of each pixel, was avoided. There are a large number of textural features which may be calculated from GLCM and several parameters to consider when selecting how to calculate these features. Here, a preliminary selection of features and their parameters was made based on correlation and discriminant analysis, as explained in Packalén and Maltamo (2007). The selected features were calculated using 10–30 requantification classes as an average of all directions (0°, 45°, 90° and 135°) with lag distances of 1–5 pixels. Altogether, 6 spectral and 7 textural features were selected for subsequent modelling.

3.2 Regression Models

Linear regression models were used to estimate the dominant height and total volume. Volume models were constructed separately for coniferous and deciduous plots in order to examine what is the difference of leaf-off and leaf-on data in the case of coniferous and deciduous forests. A coniferous plot is a plot where coniferous trees are dominant in terms of volume and, correspondingly, deciduous trees are dominant in a deciduous plot. Dominant height was modelled using all the sample plots. Thus, in regression modeling, tree species is taken into account only by stratification; using this arrangement it is possible to examine how the accuracy differs in coniferous and deciduous plots.

Regression models were constructed using only ALS-based independent variables and separate models were created with leaf-off and leaf-on ALS data. Independent variables were chosen on mainly the basis of root mean square errors (RMSE). After preliminary tests it was decided that the maximum number of independent variables is three. Several transformations were tested both to the independent variables and response and it was verified that the variance of the error term is constant. In order to demonstrate the effect of mixing datasets collected under different canopy conditions, the models that were constructed with leaf-off data were applied to leaf-on data and, correspondingly, the models constructed with leaf-on data were applied to leaf-off data.

3.3 Nearest Neighbor Imputation

Plot volumes were estimated by the NN imputation in order to investigate how well species specific plot volumes can be predicted by leaf-on and leaf-off ALS data alone or combined with features of aerial photographs. The distance metric used in NN imputation is based on canonical correlations and the generalized Mahalanobis distance. It is a modification of the original MSN method proposed by Moeur and Stage (1995) such that the number of nearest neighbors may be >1; therefore also referred as k-MSN (Sironen et al. 2001). The MSN distance metric is as follows:

$$D_{uj}^2 = \underset{l \times p}{(X_u - X_j)} \; \underset{p \times p}{\Gamma \Lambda^2 \Gamma'} \; \underset{p \times d}{(X_u - X_j)'} \qquad (1)$$

where X_u is the vector of independent variables from the target observation, X_j is the vector of independent variables from the reference observation, Γ is the matrix of canonical coefficients of the predictor variables and Λ is the diagonal matrix of squared canonical correlations. In the k-MSN estimation, k nearest neighbors for each target observation are searched for from the reference data. The estimate for each observation is calculated as a distance-weighted mean of the selected k nearest neighbors. The weighting is based on the inverse of the MSN distance. The weight W_{uj} of a reference plot u for the target plot j was calculated as follows:

$$W_{uj} = \frac{\left(\frac{1}{D_{uj}^2}\right)}{\sum_{1}^{k}\left(\frac{1}{D_{uj}^2}\right)} \tag{2}$$

where k is the number of nearest observations and $u \neq j$. The volumes of deciduous and coniferous trees were estimated simultaneously and the total volume was calculated as the sum of these. This guarantees the logical cohesion of the results, i.e. the sum of the volume of deciduous and coniferous trees is the same as the total volume.

The final independent (X) variables for the NN model were selected using the algorithm presented in Packalén et al. (2009). The core of the algorithm is to add and remove variables randomly and always to include or exclude a variable, or variables, if it decreases the cost. The cost is defined as the weighted average of the RMSEs, thus the aim is to minimize an error. It was assumed that the ALS-derived variables give information on the size of trees, while the features derived from aerial photographs should give information on tree species. To compare leaf-on and leaf-off data, the following research design was established: 1) the estimation was carried out using only ALS-based independent variables; and 2) the estimation was carried out using both ALS- and aerial photographs-based independent variables. In either case, the use of both leaf-on and leaf-off data were tested.

3.4 Classification by Coniferous and Deciduous Dominated Plots

A linear discriminant analysis was performed in order to examine the separation of coniferous- and deciduous-dominated plots (Venables and Ripley 2002). The selection of independent variables to the discriminant function was based on a stepwise procedure in which the accuracy ratio was used as the performance criterion (Garczarek 2002). The value of improvement criterion was chosen separately for each dataset in order to get approximately the same number of variables for the discriminant models. The maximum number of variables was restricted to 5. The classification was done separately with the leaf-on and leaf-off ALS data.

3.5 Accuracy Assessment

Accuracy was validated at the plot level by means of RMSE and the bias for each continuous variable (see e.g. Packalén et al. 2009). Relative RMSE was calculated by dividing the absolute RMSE by the observed mean value. In the classification by coniferous- and deciduous-dominated plots the confusion matrix, overall accuracy and kappa value were used to assess the goodness of classification (Lillesand et al. 2004).

4 Results

Coefficients and other model parameters used in the regression modelling, NN imputation and classification are case-specific and are, therefore, not presented here. Independent variables are listed in Appendix 1.

The accuracy of the regression estimates are presented in Table 2. In general, leaf-off estimates were slightly more accurate than leaf-on estimates, but typically the difference was minor. An exception was the volume model in deciduous plots where leaf-off data provided considerably more accurate estimates than leaf-on data. The dominant height was modelled with almost equal accuracy by leaf-off and leaf-on ALS data. Volume models constructed on coniferous plots were more accurate than corresponding models in deciduous plots.

Table 2. Accuracies of regression estimates.

	RMSE%	Bias%
Dominant height, all plots		
leaf-off	9.78	–0.07
leaf-on	9.87	–0.05
Volume, coniferous plots		
leaf-off	19.96	–0.02
leaf-on	21.41	–0.17
Volume, deciduous plots		
leaf-off	21.51	–0.06
leaf-on	27.63	–0.31

The regression models were also applied to ALS data collected under different canopy conditions (Table 3). All leaf-off models led to overestimates when the model was applied to leaf-on data, and all leaf-on models applied to leaf-off data caused underestimation (negative bias indicates overestimation). Correspondingly, the RMSEs increased when leaf-on models were applied to leaf-off data and vice versa. The largest increment in RMSE was in the case of deciduous plots. Especially a leaf-off model applied to leaf-on data in the estimation of volume in deciduous plots led to poor estimates; the RMSE being over 70% and negative bias almost 50%. The model-mixing decreased accuracy only slightly and also bias was rather low when dominant height was estimated with all the sample plots.

The accuracies of NN estimates are presented in Table 4. When only ALS data were used the leaf-off estimates were more accurate than leaf-on estimates in terms of all dependent variables. Although the total volume was clearly estimated better by leaf-off than leaf-on data, an even more noticeable difference was in the accuracies of coniferous and deciduous volumes, where leaf-off data produced substantially better estimates. For instance, in the case of coniferous volume the RMSE was 30.63% with leaf-off data and 58.25% with leaf-on data.

The inclusion of image features improved the accuracy considerably in the case of leaf-on data, but only slightly in the case of leaf-off data. The RMSE of total volume even increased slightly when image features were combined with leaf-off data. However, this is partly a consequence of favouring the kind of features in variable selection that increase the accuracy of coniferous or deciduous volume instead of total volume. After the inclusion of aerial features the difference in accuracy between leaf-off and leaf-on data diminished. However, the leaf-off estimates were still superior compared to leaf-on estimates. Note that the estimates obtained by leaf-off ALS data only were already as accurate as the estimates obtained by the combination of leaf-on ALS data and aerial photographs.

A confusion matrix of the classification into coniferous- and deciduous-dominated plots is presented in Table 5. 91% of the plots were classified correctly with the leaf-off data and 76% with the leaf-on data. Since there were over twice as many coniferous-dominated plots than deciduous-dominated plots even an entirely random classification indicates rather good overall accuracy. Therefore, the kappa value is more meaningful measure. The kappa value was 0.79 with the leaf-off data but only 0.39 with the leaf-on data. In particular, in terms of deciduous plots, the difference was significant: with the leaf-off data 80% of the deciduous plots were correctly classified, but with the leaf-on data only 47% were correctly classified.

Table 3. Accuracies of regression estimates when applying leaf-off models on leaf-on data and vice versa.

	RMSE%	Bias%
Dominant height, all plots		
leaf-on model in leaf-off data	10.32	2.86
leaf-off model in leaf-on data	11.36	–5.26
Volume, coniferous plots		
leaf-on model in leaf-off data	27.59	16.51
leaf-off model in leaf-on data	25.41	–11.00
Volume, deciduous plots		
leaf-on model in leaf-off data	33.96	23.98
leaf-off model in leaf-on data	70.1	–47.23

Table 4. Accuracies of volume estimates by NN imputation.

	leaf-off RMSE%	leaf-on RMSE%
ALS data		
Deciduous Volume	47.42	84.09
Coniferous Volume	30.36	58.25
Total volume	22.22	27.70
ALS data and aerial photographs		
Deciduous Volume	41.45	46.22
Coniferous Volume	28.67	39.39
Total volume	22.78	24.23

Table 5. Confusion matrices of the classifications based on dominant tree species.

	Leaf-on case: predicted		Leaf-off case: Predicted	
	coniferous	deciduous	coniferous	deciduous
Observed				
coniferous	118	14	127	5
deciduous	32	28	12	48

5 Discussion

The leaf-off estimates were always more accurate than leaf-on estimates when plot volume was regressed separately in coniferous- and deciduous-dominated sample plots. This indicates that leaf-off data are suitable for the estimation of total volume by using stratification into coniferous and deciduous plots. This conclusion is congruent with Næsset's (2005) study, where leaf-off regression models produced at least as accurate results as leaf-on models. In Næsset's (2005) study, stratification was somewhat similar to that used here; coniferous forest dominated by spruce and pine, and mixed forest with an average proportion of deciduous species of 31–42%.

The leaf-off regression models applied to leaf-on data caused systematic overestimation and, correspondingly, the leaf-on models applied to leaf-off data caused underestimation. The bias was obvious in both coniferous- and deciduous-dominated plots, although in coniferous plots the bias was smaller than in deciduous plots. It is logical to assume that applying a leaf-off model with leaf-on data causes overestimation in the case of deciduous trees since the laser pulses penetrate deeper into the deciduous tree crowns under leaf-off conditions. However, the same tendency was observed with coniferous plots, although canopy conditions should remain fairly constant in coniferous plots throughout the year. This may indicate that even a minor proportion of deciduous tree species, which might exist in coniferous plots, may have a considerable effect on the plot-level models. Model-mixing caused only slightly biased results when dominant height was estimated for all the sample plots.

The volumes of coniferous, deciduous and as their sum all the trees were estimated simultaneously by the NN imputation at the plot level. In general, leaf-off data provided considerably more accurate estimates than leaf-on data. The accuracy was improved with both leaf-off and leaf-on data when image features were combined with ALS data, but in relative terms the accuracy of leaf-on estimates improved much more. However, an important observation is that leaf-off ALS data alone provided more accurate estimates than leaf-on ALS data combined with image features. The RMSE of total volume was about the same, as in the study by Packalén and Maltamo (2007). In terms of the volume of deciduous trees, the accuracy here was substantially better than in Packalén and Maltamo (2007), but this is at least partly caused by the higher proportion of deciduous trees in this study.

Highly accurate results were also obtained when leaf-off ALS data were used to classify the sample plots into coniferous and deciduous plots, based on their dominant tree species. The classification confirms the earlier observation made in the regression modelling and nearest-neighbour imputation: the ability of leaf-off ALS data to differentiate between coniferous and deciduous trees is better with leaf-off than leaf-on data.

The reason for the better discrimination between deciduous and coniferous plots with leaf-off than leaf-on data lies in differences in height distributions. Fig. 2 depicts height distributions of first pulse data as a function of density in pure one-species plots for leaf-off and leaf-on data. A plot was considered to be a pure one species plot if the proportion of certain tree species was over 90% in terms of volume. Laser points from different sample plots are combined in Fig. 2. Especially at lower heights, the distributions between coniferous and deciduous plots differ considerably in leaf-off data. This tendency cannot be seen in the case of leaf-on data. It can also be seen that the distributions of pine and spruce follow each other quite closely and do not differ between leaf-on and leaf-off data.

The combination of leaf-off and leaf-on ALS data might further improve the accuracy obtained in this study. However, collecting both leaf-off and leaf-on datasets is not an option in practical inventories and, therefore, it was not investigated here. The use of LIDAR intensity might also improve the discrimination between coniferous and deciduous plots. At least at the individual tree-level intensity data have provided useful information for tree species discrimination (Ørka et al. 2009, Kim et al. 2009). An advantage of intensity is also that it is not related to the size of trees, unlike most other ALS metrics. However, there are some issues which may make it difficult to utilize LIDAR intensity in operational forest inventories, such as difficulties in calibration and automatic gain control (see e.g. Korpela 2008).

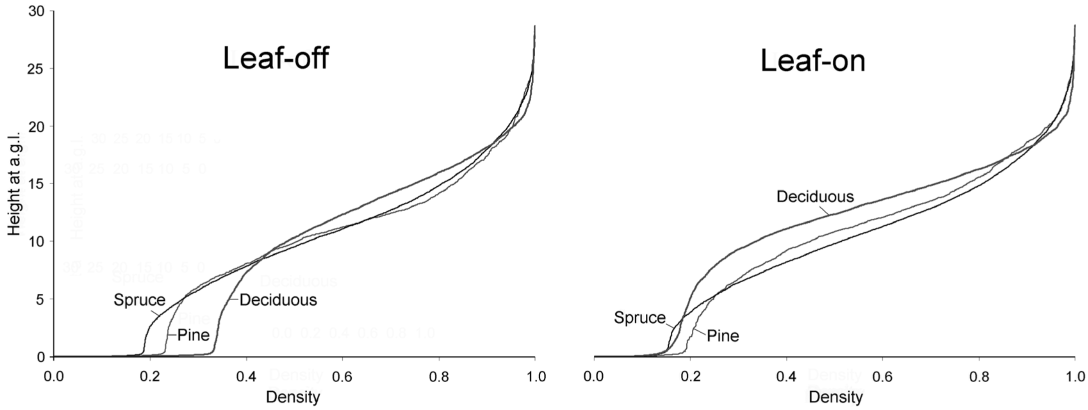

Fig. 2. Height distributions of first pulse data on pure one-species plots for leaf-off and leaf-on data. Laser points from different sample plots are combined.

Typically, the estimation of stand attributes by tree species (pine, spruce and deciduous trees) is required in a remote sensing-based forest inventory in Finland. In this study, we had to merge pines and spruces as the number of sample plots was so low. However, as the canopy conditions remain more or less constant throughout the year in coniferous stands, it can be assumed that the separation of pine and spruce is of similar accuracy as in previous studies done with leaf-on ALS data.

The timing of the leaf-off data collection is critical as the response is affected by snow cover on the ground, as well as by small buds on the trees. Difficulties may arise, especially because the time of snow melting and bud break varies from year to year. Although only a few leaf-off ALS studies have been conducted, some of them have already reported difficulties with timing and possible problems owing to early bud break (Liang et al. 2007, Kim et al. 2009). Mixing leaf-off and leaf-on ALS data may be tempting in certain situations. For instance, if the whole inventory area was not surveyed during the leaf-off season an alluring solution is to patch up those missing areas by leaf-on data. However, our conclusion is that datasets collected under different canopy conditions should not be merged together since this cause serious bias and decreases accuracy.

6 Conclusions

The overall conclusion is that leaf-off ALS data are suitable for an area-based forest inventory in which deciduous and coniferous trees need to be separated. However, the narrow time window when leaf-off ALS data can be collected may restrict the applicability. In general, results were better with leaf-off than leaf-on data. In addition, leaf-off ALS data per se had the ability to discriminate between deciduous and coniferous trees, which may decrease the inventory costs if, therefore, the acquisition of aerial photographs is avoided entirely, and if there is possibility for joint ALS data acquisition between forestry and land survey organizations.

References

Ahokas, E., Kaartinen, H. & Hyyppä, J. 2008. On the quality checking of the airborne laser scanning-based nationwide elevation model in Finland. 21st ISPRS Congress Beijing 2008. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 37(B1/I): 267–270.

Axelsson, P. 2000. DEM generation from laserscanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, Amsterdam 33(B4): 110–117.

Brandtberg, T. 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS Journal of Photogrammetry and Remote Sensing 61(5): 325–340.

— , Warner, T.A., Landenberger, R.E. & McGraw, J.B. 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment 85(3): 290–303.

Fowler, R. & J. Little. 1979. Automatic extraction of irregular network digital terrain models. Computer Graphics 13: 199–207.

Garczarek, U.M. 2002. Classification rules in standardized partition spaces. Doctoral dissertation, University of Dortmund, Germany. 226 p. Available at: hdl.handle.net/2003/2789 [Cited 14 April 2011].

Gaveau, D.L.A. & Hill, R.A. 2003. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Canadian Journal of Remote Sensing 29(5): 650–657.

Haralick, R.M., Shanmugam, K. & Dinsteein, J. 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3(6): 610–621.

Hill, R.A. & Broughton, R.K. 2009. Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: a case study in lowland Britain. ISPRS Journal of Photogrammetry and Remote Sensing 64(2): 223–233.

Hodgson, M.E., Jensen, J.R., Schmidt, L., Schill, S. & Davis, B. 2003. An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. Remote Sensing of Environment 84(2): 295–308.

Hollaus, M., Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., Bauerhansl, C. & Regner, B. 2009. Tree species classification based on full-waveform airborne laser scanning data. Silvilaser 2009, October 14–16, 2009 – College Station, Texas, USA.

Holmgren, J. 2004 Prediction of tree height, basal area and stem volume using airborne laser scanning, Scandinavian Journal of Forest Research 19: 543–553.

Kim, S., McGaughey, R.J., Andersen H-E., & Schreuder, G. 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sensing of Environment 113(8): 1575-1586.

Korpela, I. 2008. Mapping of understory lichens with airborne discrete-return LiDAR data. Remote Sensing of Environment 112(10): 3891–3897.

Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. Communicationes Instituti Forestale Fenniae 108.

Liang, X., Hyyppä, J. & Matikainen, L. 2007. Deciduous-coniferous tree classification using difference between first and last pulse laser signatures. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36(3/W52).

Lillesand, T., Kiefer, R. & Chipman, J. 2004. Remote sensing and image interpretation. Fifth edition. Wiley, New York. 763 p.

Maltamo, M. , Mustonen, K., Hyyppä, J., Pitkänen, J. & Yu, X. 2004. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. Canadian Journal of Forest Research 34(9): 1791–1801.

— , Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L. & Hyvönen, P. 2009. Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. European Journal of Forest Research 128(9): 305–317.

Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science 41(2): 337–359.

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(1): 88–99.

— 2005. Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment 98(2–3): 356–370.

— 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. Scandinavian Journal of Forest Research 22(5): 433–442.

— & Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in boreal nature reserve. Remote Sensing of Environment 79(1): 105–115.

Näslund, M. 1937. Skogsförsöksanstaltens gallringsförsök i tallskog. Meddelanden från Statens Skogsförsöksanstalt 29. 169 p. (In Swedish).

Ørka, H.O., Næsset, E. & Bollandsås, O.M. 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sensing of Environment 113(6): 1163–1174.

Packalén, P. & Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment 109(3): 328–341.

— , Suvanto, A. & Maltamo, M. 2009. A two stage method to estimate species-specific growing stock by combining ALS data and aerial photographs of known orientation parameters. Photogrammetric Engineering and Remote Sensing 75(12): 1451–1460.

Persson, Å., Holmgren, J. & Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing 68(9): 925–932.

Reitberger, J., Krzystek, P. & Stilla, U. 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. International Journal of Remote Sensing 29(5): 1407–1431.

Sirkiä, O. & Laaksonen, H. 2009. Kesäkeilausten soveltuvuudesta korkeusmallituotantoon. Kevätlaser metsävarojen inventoinnissa. MMM:n konserniohjelmahankkeen loppuraportti (proj. n:o 311100). (Final report, in Finnish).

Sironen, S., Kangas, A., Maltamo, M. & Kangas J. 2001. Estimating individual tree growth with k-nearest neighbour and Most Similar Neighbour methods. Silva Fennica 35: 453–467.

Venables, W.N. & Ripley, B.D. 2002. Modern applied statistics with S. Fourth Edition. Springer. 495 p.

Total of 33 references

Appendix 1. Independent variables used in the regression modelling, NN imputation and classification. Prefix *f_* denotes first and *l_* last pulse data, *havg* is the mean height, *hstd* is the standard deviation of the height, *hX* denote the height at which *X* percent of the height distribution of laser pulses has accumulated, *pX* refers to the laser point density below *hX*. Variables not prefixed with *f_* or *l_* are calculated from the aerial photographs. Postfix *_X* denotes band as follows: *1*=blue, *2*=green, *3*=red and *4*=NIR. *avg_X* is the mean DN of the band *X* and *pe95* is the 95th percentile of DN values. Textural variables are (see Haralick et al. 1973.): *savg*=sum average, *tvar*=sum of squares: variance, *asm*=angular second moment, *dent*=difference entropy.

	leaf-on	leaf-off
Regression, dominant height, all plots	*f_havg f_h95*	*f_h95 f_p50 l_p50*
Regression, volume, coniferous plots	*f_h20 l_hstd l_p30*	*l_hstd f_havg*
Regression, volume, deciduous plots	*f_h95 f_p30*	*f_hstd f_h60*
NN imputation, only ALS data	*f_havg f_h40 f_h95 f_p70 l_havg l_h20 l_h60 l_h80 l_p5 l_p70*	*f_havg f_hstd f_h60 f_p5 f_p50 l_hstd l_h20 l_h80 l_p50*
NN imputation, ALS data and aerial photographs	*avg_1 avg_3 pe95_4 savg_1 tvar_2 f_havg f_hstd f_h80 f_h95 f_p30 f_p50 l_h95*	*avg_1 pe95_1 avg_3 avg_4 pe95_4 asm_1 tvar_2 dent_2 f_havg f_h95 f_p50 l_hstd l_h60*
Classification by dominant tree species	*f_h1 f_h60 f_h100 l_h1 l_h100*	*f_h40 l_hstd l_h20 l_p70*