Tree Species Classification Using Airborne LiDAR – Effects of Stand and Tree Parameters, Downsizing of Training Set, Intensity Normalization, and Sensor Type

Ilkka Korpela, Hans Ole Ørka, Matti Maltamo, Timo Tokola and Juha Hyyppä

Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T. & Hyyppä, J. 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica 44(2): 319–339.

Tree species identification constitutes a bottleneck in remote sensing-based forest inventory. In passive images the differentiating features overlap and bidirectional reflectance hampers analysis. Airborne LiDAR provides radiometric and geometric information. We examined the single-trees-level response of two LiDAR sensors in over 13 000 forest trees in southern Finland. We focused on the commercially important species. Our aims were to 1) explore the relevant LiDAR features and study their dependencies on stand and tree variables, 2) examine two sensors and their fusion, 3) quantify the gain from intensity normalizations, 4) examine the importance of the size of the training set, and 5) determine the effects of stand age and site fertility. A set of 570 semiurban broad-leaved trees and exotic conifers was analyzed to 6) examine the LiDAR signal in the economically less important species. An accuracy of 88–90% was achieved in the classification of Scots pine, Norway spruce, and birch, using intensity variables. Spruce and birch showed the highest levels of confusion. Downsizing the training set from 30% to 2.5% of all trees had only a marginal effect on the performance of classifiers. The intensity features were dependent on the absolute and relative sizes of trees, especially for birch. The results suggest that leaf size, orientation, and foliage density affect the intensity, which is thus not affected by reflectance only. Some of the ecologically important species in Finland may be separable, since they gave rise to high intensity values. Comparison of the sensors implies that performance of the intensity data for species classification varies between sensors for reasons that remained uncertain. Both range and gain receiver normalization improved species classification. Weighting of the intensity values improved the fusion of two LiDAR datasets.

1   Introduction

The conventional way of measuring trees is giving way to applications that combine *in situ* observations and remote-sensing (RS) data. In this process, the introduction of airborne LiDAR was a breakthrough (Næsset 2002). LiDAR permits accurate three-dimensional (3D) probing of the vegetation and terrain, which cannot be done reliably in passive RS data. Prior to the LiDAR era, 3D methods were never used in practice, while method development in imagematching-based reconstruction of canopy surface and crowns was actively pursued (Korpela 2000, Miller et al. 2000, Gong et al. 2002). Since 2002, LiDAR has been used in Scandinavia by private companies, not only for area-based forest inventory but also for single-tree remote sensing (STRS). Trees constitute distinct targets; however, STRS requires high sampling densities and individual crowns are not always detectable in LiDAR or images (Persson et al. 2002, Korpela 2004). In area-based RS, the forest is described by LiDAR point-cloud features that characterize the structure and density of the canopy and the species admixture (Næsset 2002). An *in situ* training set is used for the estimation of imputation models. If tree positions are not needed, the area-based technique may well be more cost-efficient than STRS. The techniques can be combined in a hybrid forest inventory system.

Tree species information is crucial for economic, ecological, and technical reasons. Errors made in tree species identification can result in prominent bias in the estimates of stem biomass and dimensions, because the allometric dependencies are species-specific (Korpela and Tokola 2006). Species inaccuracy in the inventory data may lead to nonoptimal management decisions (Holopainen and Talvitie 2006). STRS research suggests that discrete-return (DR) LiDAR data can potentially be used for tree species discrimination. Species recognition in 1–4-return data was examined by Brandtberg et al. (2003), Holmgren and Persson (2004), Brandtberg (2007), Ørka et al. (2007, 2009), and Kim et al. (2009). Flying altitudes of 100–1200 m and footprint diameters of 10–37 cm were reported. Both geometric and noncorrected intensity features were applied. Holmgren and Persson (2004) extracted LiDAR

points within a crown and computed 20 features. They achieved an accuracy of 95% in separating 562 samples of two conifers, Scots pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* (L.) H. Karst.) in Sweden. Crown base height (CBH), variation in intensity, and the mean intensity of returns near the estimated crown surface were the strongest features. The LiDAR system (TopEye) had a built-in range correction of the intensity data. Brandtberg (2007) provided an elaborate discussion of the LiDAR-crown interactions. He obtained an accuracy of 64% in the classification of three broad-leaved species of West Virginia, using height (h) and intensity distribution metrics. Ørka et al. (2007) examined species identification in southern Norway for 224 samples of birch (*Betula* L. spp.), European aspen (*Populus tremula* L.), and Norway spruce. The LiDAR data were divided into echo types: first, only, and last. The mean and standard deviation (SD) of intensity were computed for these echo categories. The classification accuracy was 60–74%, depending on the number of variables. Coniferous and deciduous trees differ in near-infrared (NIR) reflectance in optical data, but the differences in LiDAR intensity ($\lambda = 1064$ nm) were marginal. Ørka et al. (2009) examined 412 spruce and birch trees of all relative size classes in an unmanaged forest. The distribution moments and percentiles of the raw intensity (I_{raw}) and the point-height distribution were extracted for each crown. Tree height influenced the LiDAR features, of which 52 were explained by species in analysis of covariance (ANCOVA). The classification accuracy was 64% for the nondominant trees and 88% for the dominant trees. The features varied with the echo type and tree height. Combination of optical data with LiDAR improved the species classification results of single trees in Persson et al. (2004), Heinzel et al. (2008), and Holmgren et al. (2008). In Holmgren et al. (2008), accuracies of 84–96% were achieved in the classification of 1711 mature pine, spruce, and deciduous trees in 14 plots. The accuracy varied for different combinations of image and LiDAR variables. The LiDAR data were acquired from a flying altitude of 130 m (50 pulses per m^2) and the DMC aerial images from 1200 m (60-cm red-green-blue-NIR (RGBNIR) pixels), which are extremely costly for practical applications. Korpela et al. (2008)

examined airborne LiDAR and 1-km RGBNIR data in species classification in seedling stands. Range-normalized intensity aided in the detection of the competing vegetation, but the accuracy in the classification of seedlings, competing vegetation, and abiotic material was only 60–80%. Large-leaved shrubs and trees showed the highest intensity. Under mire-type conditions Korpela et al. (2009) classified 10 × 10-m plots according to the dominant species, using area-based LiDAR features. The height distribution features, which characterized the habitats, surpassed the range-normalized intensity metrics in explanatory power of the species admixture.

Full-waveform (FW) LiDAR data were examined in Reitberger et al. (2006, 2008) and Höfle et al. (2008). Waveform data give detailed description of the backscattered pulse. An increase in point density is achieved by waveform decomposition (Wagner et al. 2006). This is advantageous for tree detection and additional structural details are available for species determination (Reitberger et al. 2008). Echo width and amplitude were used to characterize tree species. Amplitude is believed to correspond to intensity in DR sensors. Höfle et al. (2008) used calibrated FW data for discrimination of European larch (*Larix decidua* Mill.), English oak (*Quercus robur* L.), durmast oak (*Q. petraea* (Matt.) Liebl.), and European beech (*Fagus sylvatica* L.). Echo width separated larch from the broad-leaved trees, but oak and beech showed similar responses in backscatter cross-section and echo width.

The illuminated area, bidirectional reflectance distribution function (BRDF) of the illuminated targets, and incidence angles, i.e. the target geometry, are the main factors affecting intensity (Höfle et al. 2008). In addition, the variation in laser power (mainly due to changes in the pulse repetition frequency at different flying altitudes) and the receiver settings affect the intensity observed as does atmospheric attenuation and the scanning range (Ahokas et al. 2006, Kaasalainen et al. 2007). This implies that the retrieval of canopy reflectance must be ill-posed, even in calibrated data (targets larger than the footprint), and that the radiometric LiDAR features will exhibit substantial tree-level variation, due to differences in the illuminated area (foliage density), reflectance (of illuminated scatterers), and the geometry of

the scatterers (leaf orientation). This implies that species determination in LiDAR needs to be based on the analysis of distribution characteristics that are derived from a multitude of pulses.

Scots pine, Norway spruce, downy birch (*Betula pubescens* Ehrh.), and European white birch (*Betula pendula* Roth) constitute 96.5% of the stem volume in Finland (Tomppo et al. 2001). However, in image-based species identification, separation between pine, spruce, and birch has proven difficult, due to the overlapping spectral features (Haara and Haarala 2002). Shading, occlusion, bidirectional reflectance, and the varying phenology of trees cause signal variation (Korpela 2004, Korpela et al. 2008, Packalen and Maltamo 2007). The branching patterns of crowns are utilized in structure-based classification, but this leads to high data costs, since very-high resolution near-nadir data are required (Brandtberg 1999). Korpela (2004) used 3D treetop positioning followed by image-based crown modeling, leading to the sampling of spectral values in several views. The features characterizing self-shading enhanced the spectral classification accuracy of pine, spruce, and birch to 85%. In view of the difficulties and limitations inherent in passive optical data, LiDAR offers an alternative data source for tree species classification. However, many aspects of LiDAR-based species classification remain unexplored, mainly due to the lack of large experiments with variation in forest parameters. Empirical studies are needed, even though LiDAR simulators are being developed for the task (Morsdorf et al. 2007). We used an extensive dataset consisting of 13 890 trees in 118 plots and two LiDAR campaigns to investigate 1) the important LiDAR features used for the separation of the three main tree species in Finland, 2) the performance of two sensors, the Leica ALS50-II (Leica Geosystems AG, Heerbrugg, Switzerland) and the Optech ALTM3100 (Optech Inc., Vaughan, Ontario, Canada) and sensor fusion 3) range- and receiver gain normalization of intensity in these sensors, 4) importance of the size of the training set, 5) possible effects of age and site fertility in the LiDAR signal and classification results, and 6) the ability of LiDAR to separate ecologically important, rare tree species.

We confined our study to single-tree species identification of individual crowns representing

dominant, codominant, and intermediate trees that are visible to the sensor.

2 Material and Methods

2.1 Study Area and LiDAR Data

The experiments were carried out near the Hyytiälä forest station in southern Finland (61°50´N, 24°20´E). The rotation periods are typically 80−130 yr and trees on mineral soils attain levels of h of 21−33 m, depending on site fertility. Scots pine and Norway spruce predominate and form both mixed and pure stands. The birch stands are younger than 40 yr, due to previous periods of conifer-favoring silviculture. Isolated birches are found in coniferous stands. Mineral soils with gentle slopes prevail and the elevation is 135−198 m above sea level (a.s.l.). Lakes, open mires, spruce mires, and pine bogs cover the basins. The mires are largely drained and the state-owned forests are managed for commercial forestry. The study area extends 2×6 km and encompasses a large number of permanent forest plots both in managed and pristine forests, mineral and peat soils, and pristine and drained mires. Aerial images from 1946 to 2009 were available, as were the results of four LiDAR campaigns after 2004. We used LiDAR data from 2006 and 2007 (Table 1). Both campaigns had an average pulse density of 6−8 per m^2 and a 1064 nm wavelength The LiDAR data were stored in 207-byte binary records that contained the full-pulse geometry, including the position and attitude of the LiDAR for accurate range calculations. The ALTM3100 postprocessing software supported this feature (comprehensive format), but the ALS50 data required additional postprocessing to include the trajectory data.

The raw intensity (I_{Raw}) observations in both LiDAR datasets were normalized for the range R, using a reference (average) range, R_{Ref}

$$I_{Ran} = \left(\frac{R}{R_{Ref}} \right)^{2.4} \cdot I_{Raw} \tag{1}$$

The ALS50 data were also compensated for the

Table 1. Characteristics of the LiDAR datasets.

Instrument	ALTM3100	ALS50-II
Date	July 25, 2006	July 4, 2007
Pulse frequency	100 kHz	115.8 kHz
Scanning frequency	70 Hz	52 Hz
Footprint	25−28 cm	17−18 cm
Flying altitudes, nominal	1000 m a.s.l.	930 m a.s.l.
Scanning angle	± 14°	± 15°
Air humidity, 2 m	48−52%	60−75%
AGC (Automatic Gain Control)	-	8 bits

influence of the automatic gain control, AGC that maintains the recorded intensity data in an 8-bit range

$$I_{Ran} = \left(\frac{R}{R_{Ref}} \right)^{2.5} \cdot I_{Raw} + 0.045 \cdot I_{Raw} \cdot (127.5 - AGC) \tag{2}$$

The R- and AGC-normalized intensity values of the ALS50 were normalized to match the ALTM3100 data for data fusion by a simple multiplication

$$I_{Fus} = 0.5578 \cdot I_{Ran} \tag{3}$$

The targets, including gravel, asphalt, an oat field, grass, ground lichens, and shrubs, were used for finding the parameters in Eqs. 1−3 (Korpela 2008). We had I_{Raw} and I_{Ran} data for the ALTM3100 and I_{Raw} and R and AGC-normalized I_{Ran} data for the ALS50 sensor. For the combined sensor consisting of ALTM3100 and ALS50 data, there were I_{Raw}, I_{Ran}, and I_{Fus} intensity variables.

2.2 Reference Trees

Over 15 000 reference trees in 117 plots were measured during 2002−2008 in Hyytiälä. Prior to 2006, a tacheometer was used for the positioning of trees. In 2006−2008, a photogrammetric-geodetic procedure was applied (Korpela et al. 2007). In it, treetops are first positioned, using ray-intersection in multiple aerial images or by a monoplotting procedure that employs images
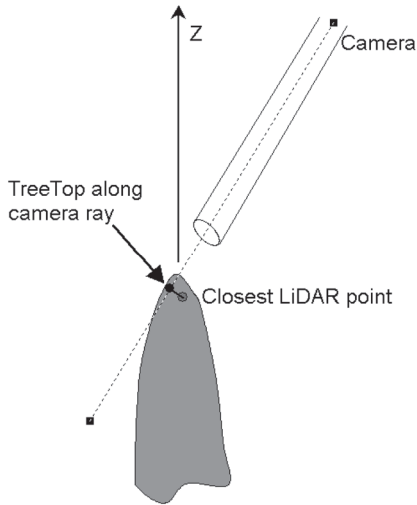
Fig. 1. Principle of the monoplotting method for treetop positioning. A treetop is pointed in one image by the operator, which usually is the backlit view, if available. This establishes a camera ray and the distance to the first LiDAR point intersected by the 0.5–1-m-wide camera tube is forced to coincide with the camera ray, keeping the distance to the camera unchanged. The mean underestimation of Z was 0.33 m in 465 trees measured in the field in 2007. The XY accuracy is better than 0.4 m. The rate was 200–300 measured trees per hour.

and LiDAR (Fig. 1). These trees serve the role of field control points for the positioning of other targets, using trilateration and/or triangulation. Trees in stands with h values below 6–8 m were mapped, using kinematic or Network RTK satel-lite positioning (Wanninger 2005). All trees were recorded for species, diameter-at-breast height (DBH) and crown status, and 30–100% of the trees per plot were measured for CBH and h. The plots were fixed in size, circular, rectangu-lar, or free in shape and 0.04–1.8 ha in area. In 2002–2003, crown width (d_{cr}) was measured in nearly 1000 35–140-yr-old trees. To study LiDAR use in rare tree species, 580 semiurban trees were mapped in August 2007 near the Hyytiälä forest station; this set is referred to as semiurban trees. The status of all trees at the time of the LiDAR campaigns was visually verified, using large-scale aerial images taken in August 2006 and June 2007, and the treetop positions were updated with the monoplotting technique (Fig. 1). Some broken, fallen, or cut trees were rejected during this work and dead standing trees were reclassified. Only trees that were discernible in the images and/or visualized LiDAR data were included in the tree set. Most suppressed and intermediate trees with a relative height (h_{rel}) of below 0.6 were rejected. In all, 13 317 forest trees and 570 semiurban trees were contained in the sets of discernible trees.

For each tree we computed the best available h observation, which was the field measure-ment if made in 2006–2008. For the other trees, LiDAR treetop positioning was applied and for the unseen trees, plot-level regression curves were applied. For each tree the h_{rel} was determined, using a local dominant height. The site-type for each plot or subplot was determined, using maps initially drawn in the field. Site fertility in six classes was derived from the site type descrip-tion (Table 2). The mean age and age variation

Table 2. Number of living pine, spruce, and birch forest trees in site fertility and age classes. For example, fertility class 2 included the *Oxalis-myrtillus* (OMT) site type of mineral soils, the eutrophic, paludified alder-spruce forest, and the herb-rich drained mire. Similarly, the *Calluna* type (CT), drained dwarf-shrub type, as well as rocky or paludified *Vaccinium* sites (VT) were included in fertility class 5.

Site/age	<30	30–55	55–70	70–90	90–110	>110	Sum
OMT-2	170	194	225	-	26	109	724
MT-3	129	4645	2113	522	783	1026	9218
VT-4	729	1204	77	156	111	53	2330
CT-5	-	5	-	251	336	-	592
ClT-6	-	-	15	25	25	-	65
Sum	1028	6048	2430	954	1281	1188	12 929

Table 3. Mean and (SD) of tree variables in living pine (4982), spruce (6066), and birch (1881) forest trees. Echoes refers to the number of first-or-only returns in the crown, both sensors combined.

Var	Pine	Spruce	Birch
DBH, cm	19.8 (6.4)	20.2 (8.0)	15.1 (6.7)
h, m	17.3 (4.0)	18.1 (5.8)	16.4 (5.4)
d_{cr}, m	3.0 (0.70)	3.3 (0.8)	3.0 (1.2)
Age, yr	58 (25)	68 (29)	46 (25)
Echoes	83 (53)	92 (72)	84 (83)

Fig. 2. Map of LiDAR pulses, which have echoed from an h of below 20 m and have been back-projected to a mapping surface 20 m aboveground. The map is from a 25−30-m-high spruce stand (70 × 70 m). The large openings in the periphery are caused by large pine and birch crowns. We refer to the dots as pseudoechoes. The ALTM3100 data from 2006 are in green and blue dots that also depict the ALS50 data. There are six overlapping LiDAR strips, and XY matching is reasonable between sensors and strips.

were estimated, using field measurements, aerial images of 1946–2009, and forest management documentation. In estimating the age for each tree we allowed some random variation around the plot mean age in naturally regenerated stands. The proportion of crown overlap was estimated for each tree by computing the overlap of neighboring pine, spruce, and deciduous crowns as overlapping circles to characterize competition by neighbors of different species. The distribution of trees according to age and site fertility is in Table 2. The distributions are not in balance. Only 1.5% of all trees had h_{rel} below 0.5. Thus, the study applies to the dominant, codominant, and intermediate trees, which range from 2 m to 38 m in h (Table 3).

2.3 Extraction of LiDAR Features for the Reference Trees

Extraction of LiDAR data and derivation of features were incorporated in a crown-modeling procedure, in which a curve of revolution was fitted to the LiDAR points near the treetop (Korpela 2007). To collect the LiDAR returns from a particular tree is an ill-posed task, especially in the lower parts of crowns where neighboring trees overlap. Fig. 2 illustrates the problem in a spruce stand. First, 871 field measurements were used for estimating species-specific regression models that predict d_{cr}, using DBH and h as predictors. The regression estimate of d_{cr} was multiplied by a factor of 1−1.4, depending on stand density, which was visually determined in aerial images and was transformed into an initial approximation (over-
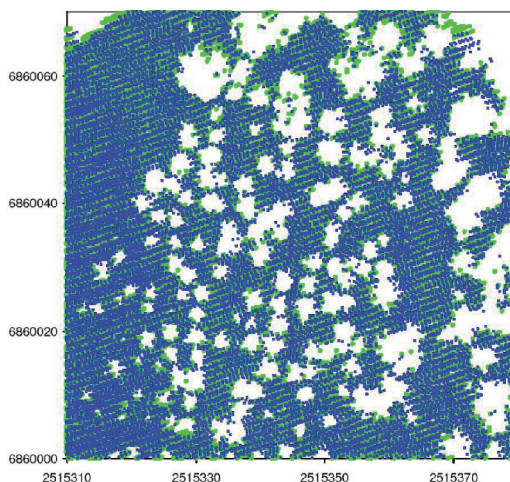
estimate) of the crown model. A three-parameter curve of revolution was fitted to the LiDAR point cloud of combined sensors, using a weighted least squares (WLS) adjustment (regression). In it, LiDAR points were treated as observations of crown radius and observations that fell outside the crown envelope were weighted more (WLS weights = $1/\sigma^2$). The accuracy (σ) of the outer points was 0.44 m, while $\sigma = 1$ m for the other points. WLS reduces the inherent underestimation of the crown envelope in this kind of modeling. The length of the crown model was fixed and CBH was always 40% down from the top. This is clearly a limitation in our analysis; e.g. young and open-grown spruce trees have much deeper crowns. LiDAR points inside or within 1 root-mean-squared-error (RMSE) of the envelope were saved for feature computations. Echoes lower than 40% from the top were excluded.

Both sensors were capable of extracting 1−4 echoes from the reflected pulse. Minimum range

Table 4a. Names of the LiDAR variables consist of five parts: #1_#2_#3_#4_#5. For example, *comb_fo_cr_mean_ I_{Raw}* is the mean raw intensity computed, using first-or-only echoes by combining both sensors.

Part	Explanation
#1	Sensor: *altm*=ALTM3100; *als*=ALS50-II; *comb*=Both, combined
#2	Echo type: *on*=only; *fi*=first-of-many; *fo*=first-or-only; *al*=all
#3	Delineation: *cr*=full crown; *su*=near-surface points only; Vertical: *u1*=upper 10%; *u2*=10–20%; *u3*=20–30%; *u4*=points in 30–40% of the tree
#4	Type of statistics: *mean, sdev, skew, kurt*=distribution moments; *dd#* and *d#*=density deciles, deciles; *p*=proportion; *n*=number
#5	Variable: I_{Raw}=raw intensity data; I_{Ran}=range- or range- and AGC-normalized intensity; I_{Fus}=ALS50 normalized to ALTM3100; *h*=height (0=top, 1=butt)

Table 4b. Features computed for each sensor (*altm, als, comb*) × echo-type (*on, fi, fo, al*).

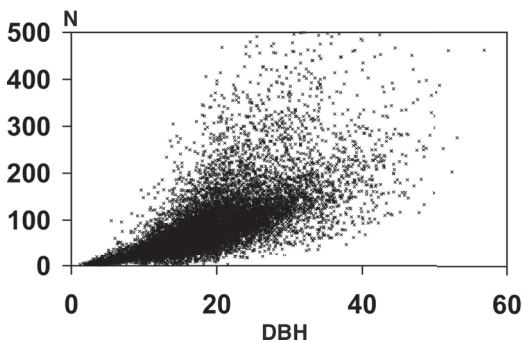| (Delineation) × (Statistic) × (Variable)	n	Description
(*cr*) × (*mean, sdev, skew, kurt*) × (I_{Raw}, I_{Ran}, I_{Fus})	12	Intensity moments in the whole crown
(*su*) × (*p*), (*cr*) × (*n*)	2	Proportion and number of points
(*su*) × (*mean, sdev, skew, kurt*) × (I_{Ran})	4	Intensity moments near the crown surface
(*u1, u2, u3, u4*) × (*mean*) × (I_{Raw}, I_{Ran}, I_{Fus})	12	Intensity means in vertical crown layers
(*cr*) × (*d1,...,d10*) × (I_{Raw}, I_{Ran}, I_{Fus})	30	Intensity deciles in the whole crown
(*cr*) × (*dd1,...,dd10*) × (*h*)	10	Height density deciles, *comb* × *al* data
(*cr*) × (*d1,...,d10*) × (*h*)	10	Height deciles, *comb* × *al* data

Fig. 3. Number of all ALTM3100 and ALS50 echoes in a crown (N) as function of DBH (cm).

differences or blind zones of 3 m and 3.5 m are required in the ALTM3100 and ALS50 sensors, respectively. We labeled the echoes as belonging to the groups first-of-many, only, first-or-only, and/or all. The near-surface echoes formed a separate class (± 1 RMSE from the crown surface). The relative distance from the treetop, 0–0.4, was calculated for each echo. The 40% deep crown

was vertically divided into layers *u1–u4*. Distribution moments and deciles (*d#*) were computed for the intensity variables. Deciles and density deciles (*dd#*) were computed for the height distributions. The *dd#* were calculated by dividing the range in *h* into 10 equal intervals and by taking the cumulative proportion of echoes, starting from the treetop. A total of 60 intensity features were available for each of the 12 sensor × echo-type combinations, in addition to 20 common height variables, giving 740 variables (Tables 4a and 4b).

The average number of first-or-only echoes in a crown was 88 in the combined sensor (Fig. 3, Table 3). There were on average 38 ALTM3100 and 50 ALS50 echoes per crown. Trees higher than 9–10 m could produce more than one echo, considering the nominal range discrimination distances of 3 m and 3.5 m in the sensors used. In a 28-m-high spruce-pine stand, we used 49 000 range differences and observed distances of 2.1 m and 3.1 m in the ALTM and ALS sensors, respectively. Some of the observed second, third,

and fourth returns were undoubtedly caused by oblique (± 15°) pulses hitting neighboring trees.

2.4 Classification Tools

We used the linear discriminant analysis (LDA), k-nearest neighbor (k-NN), k-most similar neighbor (k-MSN), and random forest (RF) algorithms in the classification trials. LDA is a parametric method in which a linear combination of features that best separates the classes is sought. We used the lda-function of the MASS package in R with equal prior probabilities (Venables and Ripley 2002). We did not test the multivariate normality of the data. The data were outlier-free, making LDA applicable. LDA was used in single-tree species classification in LiDAR (Holmgren and Persson 2004; Brandtberg 2007; Ørka et al. 2009). The nonparametric k-NN method has been used in multisource forest inventories to combine different data sources (Tomppo et al. 2008). In k-NN, normality assumptions are not needed. We used the DISCRIM procedure in SAS software (SAS Institute Inc., Cary, NC, USA) with $k = 7$, Mahalanobis distance, and equal priors. A nonparametric method with increasing popularity in forest applications is RF (Breiman 2001). RF is a development of classification and regression trees that grows a large number of such trees (several hundred to several thousand trees) on randomly selected features and identifies the class by a majority vote among these trees. RF was used for classification of forest structure (Falkowski et al. 2009) and mire types (Korpela et al. 2009) in LiDAR. To obtain comparable results with LDA and k-NN with equal priors, a balanced RF procedure was applied (Chen et al. 2004). In balanced RF the number of samples from each class in each tree grown was set equal to the number of objects in the smallest class. RF trials were carried out with the R-package randomForest (Liaw and Wiener 2002). We used the mean decrease in Gini index for feature selection. The k-MSN method is a nearest neighbor algorithm variant that uses canonical correlation analysis between dependent and independent variables to produce a weighting matrix. The MSN distance metric is then derived from the weighting matrix (Moeur and Stage 1995). Finally, estimates for validation

data are calculated from k nearest neighbors that according to independent variables are similar to the target of prediction, employing a weighting based on the inverse of the MSN distance. We used the k-MSN method for classification. Weighted sums of $k = 9$ nearest neighbors of different tree species were calculated and the one with the greatest value was chosen as a tree species estimate. The k-MSN approach is often applied, since it can easily utilize information on large numbers of independent variables (e.g. Packalén and Maltamo 2007).

3 Results

3.1 Analysis of Individual LiDAR Features

The mean intensity of first-or-only returns was lower than the mean intensity of only echoes (Fig. 4). The first returns represent multireturn pulses with enough energy after the first reflection to produce subsequent scattering and registered echoes. The vertical dampening of the intensity down the crown was notably stronger in the first-or-only echo class than in only echoes (Fig. 4). In birch, there was little dampening of intensity in the only echoes, which may indicate that the foliage was homogenous in vertical. In pine and spruce the dampening may be explained by the lower needle density near the crown base, or by vertical changes in the structure and reflectance of the scatterers. Transmission losses are also a possible explanation. The topmost shoots in spruce crowns are often covered by epiphytic lichens (e.g. the tube lichen *Hypogymnia physodes* (L.) Nyl.), which may partly explain the high levels of backscatter in the upper crown (Fig. 4).

The height deciles (Fig. 5) or density deciles were not very strong explanatory variables. In analysis of variance (ANOVA), the species explained merely 10% and 12% of the variation in the 70th and 80th height percentiles, respectively.

The intensity of first-or-only returns for the combined sensor was highest in Norway maple (*Acer platanoides* L.), which has large, favorably oriented, almost horizontal leaves (Table 5). Goat willow (*Salix caprea* L.) and European mountain-ash (*Sorbus aucuparia* L.) have similar
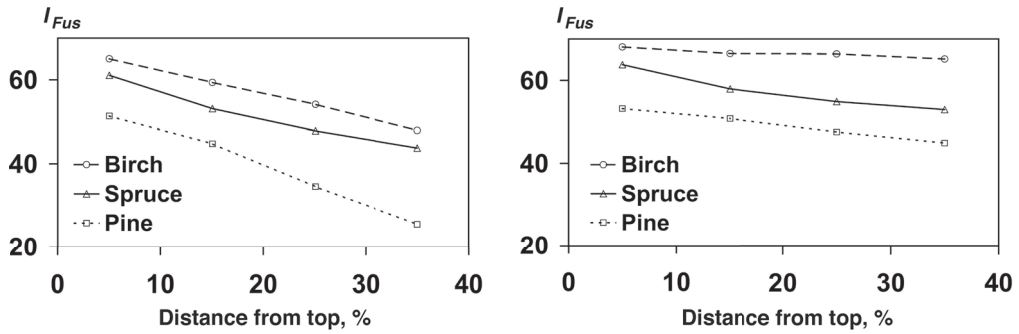
Fig. 4. Per-species mean intensity (*comb_##_al_u#_IFus*) of first-or-only (left) and only echoes (right) in four vertical 10% deep crown layers.
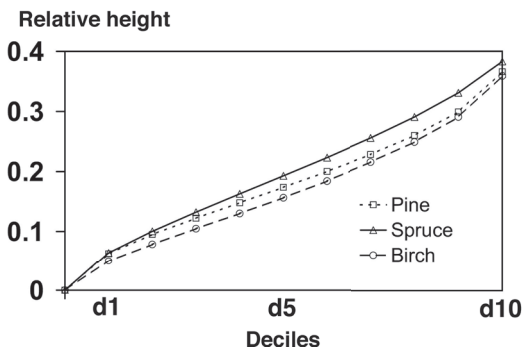

Fig. 5. Average height deciles of pine, spruce and birch (*comb_al_cr_d#_h*).

leaf orientation and their mean intensities were 65–90% that of maple, depending on the growing conditions. In the *Betulaceae* family, gray alder (*Alnus incana* (L.) Moench) and European alder (*Alnus glutinosa* (L.) Gaert.) showed higher intensity values than did *B. pendula* and *B. pubescens*. The size and orientation differences of leaves likely explain this difference, although actual differences in reflectance cannot be ruled out. The mean intensity of *Populus tremula* was close to that of *Betula* (spp.) under both forest conditions and for the semiurban trees near the Hyytiälä forest station. Both species have small leaves with high inclination angles. Siberian fir (*Abies sibirica* Ledeb.) had the highest intensity of the conifers tested. The relatively wide needles form dense, nearly planar surfaces around the shoots in this species. The semiurban Siberian larch (*Larix sibirica* Ledeb.) trees were younger than the forest trees, which all grow in an 85-yr-old stand on semifertile soil. This may explain the large difference in mean intensity between the two groups. Pine had a clearly lower intensity than spruce in forest trees, but in open-grown trees on fertile soil, there was no difference between pine and spruce. The mean intensity of spruce was close to that of aspen. The dataset contained dead defoliated trees, which showed an approximately 40–60% lower intensity than living trees. The trees that were labeled as deteriorated (N=43) in the field also showed 10–20% lower average intensity than normal trees, suggesting that intensity may be useful for the coarse measurement of needle density or mass, even at the tree level. We also tried clustering techniques (k-means, hierarchical clustering, and principal component analysis) with the semiurban tree data, but the resulting grouping was not easy to interpret and explain by morphological features in these species.

We tested the dependencies of intensity variables on tree age and *h*. In spruce, the mean intensity of first-or-only crown echoes was 5–8% lower for the tallest trees than for 5–10-m-high trees. However, the feature *comb_fo_u1_mean_IFus*, which is the mean intensity in the top 25% of the crown, showed no trends. The echoes from the lower parts of the crown caused the dependence. In pine, a contrasting trend was seen, since the tallest pines had a 5–10% higher intensity than the shortest and youngest trees ($R^2=0.033$). In pine, the trend was caused by the echoes in the topmost part of the crown; *comb_fo_u1_mean_IFus* was

Table 5. Mean intensity of first-or-only returns in a crown. Combined sensor. *comb_fo_cr_mean_IFus.*

Species	Forest trees			Semiurban trees		
	N	Mean	SD	N	Mean	SD
Broad-leaved						
Acer platanoides L.	1	90.0	-	33	78.2	11.6
Alnus glutinosa (L.) Gaertner	67	65.6	9.7	16	59.6	12.9
Alnus incana (L.) Moench	22	57.3	16.2	89	62.8	12.1
Betula pubescens Ehrh., *B. pendula* Roth – living	1878	54.9	10.6	101	50.8	11.8
Betula L. spp. – dead standing	6	22.8	8.9	-	-	-
Fraxinus excelsior L.	-	-	-	2	83.7	7.5
Populus tremula L.	34	49.1	11.0	65	55.1	11.3
Prunus padus L.	-	-	-	1	84.4	-
Quercus robur L.	-	-	-	2	73.6	7.4
Salix caprea L.	21	61.8	11.5	67	73.0	11.4
Sorbus aucuparia L.	48	70.3	14.8	31	72.6	14.7
Tilia cordata Mill.	-	-	-	9	65.9	6.4
Ulmus glabra Huds.	-	-	-	8	58.3	10.2
Conifers						
Abies sibirica Ledeb.	7	64.6	5.4	45	72.6	9.6
Larix sibirica Ledeb.	41	45.9	6.2	17	64.3	8.6
Picea abies (L.) H. Karst – living	6062	51.8	6.2	32	50.7	6.3
Picea abies – dead standing	95	31.4	7.7	-	-	-
Picea omorika (Pancic) Purk.	3	59.5	3.4	-	-	-
Pinus cembra L.	-	-	-	9	56.4	5.4
Pinus contorta Douglas ex Louden	-	-	-	2	47.5	5.4
Pinus sylvestris L. – living	4982	41.7	5.6	39	48.9	6.9
Pinus sylvestris – dead standing	25	26.4	9.8	-	-	-
Pseudotsuga menziesii (Mirb.) Franco	6	55.2	3.6	1	66.3	-
Thuja occidentalis L.	-	-	-	1	74.6	-

10–15% higher in the tallest pines ($R^2=0.054$). Again, the intensity of the echoes from the lower part of the crown was not correlated with *h*. The dependencies were similar with the age variable. In birch, the intensity of the first-or-only echoes in the crown was 20–30% lower for the tallest trees ($R^2=0.180$) and the trend was strongest in the echoes that had reflected from the upper half of the crowns. A likely explanation is the lower leaf density of old/large birch trees that have wide crowns. In general, the intensity metrics are correlated with size/age, especially in birch, and these dependencies vary according to the vertical layers in a crown.

We also examined the correlation between h_{rel} and the intensity variables. In pine, the dominant trees had an approximately 5% higher intensity than the intermediate trees ($R^2=0.03$). The echoes from the lowest parts of the crown showed clearly
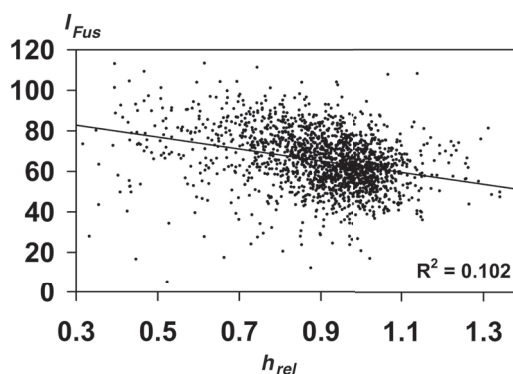

Fig. 6. Dependency of the relative height (h_{rel}) and mean intensity in the top of the crown (*comb_fo_u1_mean_IFus*) in birch.
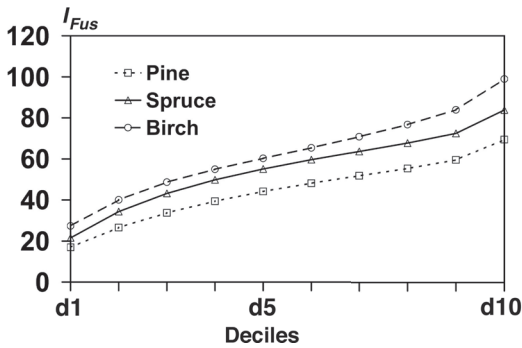
Fig. 7. Intensity deciles (*comb_fo_cr_d#_IFus*) for pine, spruce, and birch.

higher mean intensity in the dominant trees ($R^2 = 0.08$), probably due to the relatively deeper crowns of the dominant trees. No correlation was observed for spruce. In birch, the intensity of the dominant trees was 15–20% higher than in the intermediate trees ($R^2 = 0.08$). The effect was caused by the uppermost echoes in the crown (Fig. 6). This analysis suggests that the intensity features need to be computed in vertical layers and perhaps treated differently, depending on the relative size of the tree.

In pine, the proportion of competing crown coverage by neighbors of other species was correlated with the mean intensity of the crown surface echoes ($R^2 = 0.04$). A 10% higher intensity was observed for pines with overlapping crowns of broad-leaved or spruce trees than in competition-

free pines, suggesting that under dense forest conditions the extraction method used is unable to correctly filter the LiDAR points for each tree. Effects by other factors, such as site fertility, could have also produced the observed dependency.

The intensity deciles separate the three species, as illustrated in Fig. 7. For example, in ANOVA, 59.1% of the variation in the 9th decile was explained by the species.

3.2 Effects of the Size of the Training Set, Stand Age, and Site Fertility

We used RF for selecting the 10 most important LiDAR features, which we applied in k-MSN classification trials. Those features utilizing the first-or-only observations had the least missing data and also performed better than the all echoes. It should be noted however, that the separate use of features computed from the first-and-only returns, which reduced the number of test cases by 50–75%, resulted in higher classification performance, using RF in all sensor combinations. The number of test cases was, however, reduced by 50–75%, due to missing data. Only intensity features were selected by RF (Table 6).

Using random sampling, the living pine, spruce, and birch trees were sampled into 100 training and validation sets at levels of 2.5%, 5%, 15%, and 30% (Tables 7–9). The overall classification accuracy with k-MSN was 87.8–88.9% (Table 7). The downsizing had only a minor effect. Most of

Table 6. Mean decrease in Gini importance ranking of the first-or-only LiDAR variables by Random Forest. We used combined sensors and I_{Fus} data with living pine, spruce, and birch. The R^2 values were obtained in ANOVA by explaining the variation of the feature with the tree species.

Feature	R^2	Explanation
comb_fo_cr_d9_IFus	0.591	90th percentile of intensity distribution
comb_fo_cr_d10_IFus	0.574	Maximum intensity distribution
comb_fo_cr_d8_IFus	0.535	80th percentile of intensity distribution
comb_fo_cr_d7_IFus	0.482	70th percentile of intensity distribution
comb_fo_u4_mean_IFus	0.418	Mean intensity 30–40% down from the top
comb_fo_cr_d6_IFus	0.432	60th percentile of intensity distribution
comb_fo_cr_mean_IFus	0.451	Mean intensity in crown
comb_fo_cr_sdev_IFus	0.295	Standard deviation of intensity distribution
comb_fo_cr_skew_IFus	0.089	Skewness of intensity distribution
comb_fo_u3_mean_IFus	0.376	Mean intensity 20–30% down from the top

Table 7. Mean and (SD) of classification accuracy percentage and kappa in 100 trials with randomly sampled training and validation sets, using k-MSN classification with the 10 features in Table 6.

Training data	All, %	Kappa	Pine	Spruce	Birch
30% of trees	88.9 (0.3)	0.813 (0.004)	92.7	88.5	78.8
15% of trees	88.8 (0.2)	0.812 (0.004)	92.5	88.4	78.8
5% of trees	88.4 (0.4)	0.806 (0.007)	92.3	87.8	79.0
2.5% of trees	87.8 (0.9)	0.795 (0.015)	92.2	86.8	78.4

Table 8. Mean classification accuracy in age classes in 100 trials with randomly sampled training and validation sets, using k-MSN classification with the 10 features in Table 6.

Training data	<30	30–50	50–70	70–90	90–110	110–130	≥ 130
30% of trees	72.4	88.8	93.0	91.1	90.6	83.2	85.5
15% of trees	72.0	88.7	92.9	91.1	90.4	83.1	85.2
5% of trees	71.4	88.4	92.7	90.7	90.1	82.7	84.6
2.5% of trees	70.6	87.7	92.3	89.8	89.4	82.3	83.9

Table 9. Mean classification accuracy in site fertility classes in 100 trials with randomly sampled training and validation sets, using k-MSN classification with the 10 features in Table 6.

Training data	OMT-2	MT-3	VT-4	CT-5	ClT-6
30% of trees	88.5	89.2	86.6	92.9	82.0
15% of trees	88.6	89.1	86.5	92.9	81.8
5% of trees	87.8	88.8	86.2	92.4	81.5
2.5% of trees	87.0	88.3	85.6	91.2	79.4

Table 10. Mean and (SD) of classification accuracy percentage and kappa in 100 trials with randomly sampled training and validation sets. Division between training and validation sets was done according to age and site fertility.

Training data	All, %	Kappa	Pine	Spruce	Birch
30% of <55-yr-old trees	88.0 (0.2)	0.798 (0.004)	91.6	87.2	80.1
30% of trees in fertility classes 2 and 3	87.8 (0.2)	0.752 (0.007)	98.5	68.9	67.1

the confusion occurred between birch and spruce, while pine was classified most reliably at 92.5%. Even with 300 trees in the training set (2.5%), the accuracy was 87.8% in the validation set of 11 700 trees.

The accuracy was lowest in the youngest and oldest trees (Table 8). Nearly 50% of the trees younger than 30 yr were birches, which partly explains the low accuracy in that age class. In addition, the young trees were smaller and thus had smaller numbers of LiDAR hits per tree. The accuracy was lowest in the most barren fertility class (Table 9). However, only two plots and 65 trees belonged to this class.

We also applied divisions according to the site fertility and age variables (Table 10). The accuracy in the trees older than 55 yr was 88% when k-MSN was trained with a sample consisting of 30% of trees younger than 55 yr. However, when the validation set (to be classified) was formed by trees in the three most barren fertility classes, while training with 30% of the trees in the two most fertile classes, low classification accuracy was obtained for spruce and birch. Here, the optimal (yield) site fertility classes of spruce and birch were used for training. The birch trees in the barren sites were on average 15 yr younger and the mean h was 10.5 m in comparison to 17.4 m in the training set. This may explain the poor performance, since the intensity features were correlated with tree size in birch. Similar imbalances of tree variables also applied to spruce. Thus, due to the imbalanced data, we cannot conclude for certain whether site fertility is a strong factor affecting the LiDAR signal. Also, a reliable separation of the effects by site fertility and tree size is impossible.

3.3 Comparison of sensors and analysis of sensor fusion, range, and AGC normalization

We applied the k-NN, LDA, and RF classifiers with 13 intensity variables computed from the first-or-only returns to test the differences between sensors and to examine the gain from intensity normalization (Table 11).

Intensity normalization (Eqs. 1, 2) for the effect of varying range improved the classification performance in the ALTM3100, ALS50, and combined data (Tables 12, 13). It should be noted that the ALS50 data were normalized for the effects of the AGC and range (Eq. 2). The AGC compensation may have enlarged the dynamic range of the ALS50 sensor, since I_{Ran} improved the classification accuracy by 6–8%, while the gain was 2% in the ALTM3100. The ALS50 also performed better than the ALTM3100 with I_{Raw}.

The ALS50 data were captured 100 m lower than the ALTM3100 data and the footprint diameter was 17–18 cm or 240 cm^2 in comparison to the 25–28-cm or 550-cm^2 footprint of the ALTM3100. In theory, the smaller footprint may lead to an improved signal-to-noise ratio in the intensity measurement. The classification performance was better with the combined data and the I_{Fus} intensity observations (weighting of sensors) than the I_{Ran} data. The average number of first-or-only echoes in a crown was 37.5 for the ALTM3100 and 50.5 for the ALS50 data. The ratio of the ALTM3100 and ALS50 echoes varied considerably among the trees and plots, which is explained by the LiDAR data capture in narrow overlapping parallel and perpendicular strips. The assumption that I_{Fus} would compensate for the sensor differences, when the contribution of the sensors varies, was thus supported by the results (Tables 12, 13).

Table 11. Intensity variables used for testing sensor differences. In the variable names X refers to the sensor and $Yyyy$ to the intensity variable: I_{Raw}, I_{Ran}, and I_{Fus}.

Variable	Description
X_fo_cr_mean_IYyyy	Mean intensity of crown returns
X_fo_cr_sdev_IYyyy	Standard deviation of crown returns
X_fo_cr_skew_IYyyy	Skewness of crown returns
X_fo_cr_kurt_IYyyy	Kurtosis of crown returns
X_fo_u1_mean_IYyyy	Mean intensity, topmost 10% of the tree
X_fo_u2_mean_IYyyy	Mean intensity, topmost 10–20% of the tree
X_fo_u3_mean_IYyyy	Mean intensity, topmost 20–30% of the tree
X_fo_u4_mean_IYyyy	Mean intensity, topmost 30–40% of the tree
X_fo_cr_d2_IYyyy	2nd decile of intensity distribution, whole crown
X_fo_cr_d4_IYyyy	4th decile of intensity distribution, whole crown
X_fo_cr_d6_IYyyy	6th decile of intensity distribution, whole crown
X_fo_cr_d8_IYyyy	8th decile of intensity distribution, whole crown
X_fo_cr_d9_IYyyy	9th decile of intensity distribution, whole crown

Table 12. k-NN classification results with 13 intensity variables (Table 11) in the first-or-only data. The number of trees was 10 982, 11 104, and 11 908 for the ALTM3100, ALS50, and combined sensors, respectively. (ASE) is the asymptotic standard error of the simple Kappa coefficient. The 95% confidence limits are ± 2 × ASE. Leave-one-out cross-validation.

Sensor / intensity	All, %	Kappa (ASE)	Pine	Spruce	Birch
ALTM3100 – I_{Raw}	73.2	0.629 (0.006)	87.6	73.8	58.1
ALTM3100 – I_{Ran}	75.3	0.668 (0.006)	88.8	77.8	59.3
ALS50 – I_{Raw}	78.2	0.645 (0.006)	88.8	71.0	74.7
ALS50 – I_{Ran}	85.0	0.770 (0.005)	93.6	82.1	79.4
COMB – I_{Raw}	83.1	0.731 (0.006)	87.7	81.2	80.5
COMB – I_{Ran}	88.4	0.814 (0.005)	91.2	87.3	87.0
COMB - I_{Fus}	88.7	0.828 (0.005)	92.3	88.6	85.3

Table 13. k-NN, LDA, and RF classification results with 13 intensity variables (Table 11) in first-or-only data. The number of trees was 10 337, with 3750 pines, 5385 spruces, and 1202 birches. These trees had data for all of the 7 × 13 variables. Leave-one-out cross/validation in k-NN/LDA and out-of-bag (oob) error estimate in RF.

Sensor / intensity	All, %	Kappa	Pine	Spruce	Birch
		k-NN			
ALTM3100 – I_{Raw}	73.4	0.639	88.6	75.3	56.1
ALTM3100 – I_{Ran}	76.0	0.681	89.3	79.5	59.2
ALS50 – I_{Raw}	78.4	0.644	88.3	71.8	75.0
ALS50 – I_{Ran}	85.1	0.774	93.4	83.4	78.4
COMB – I_{Raw}	83.1	0.715	90.0	78.0	81.2
COMB – I_{Ran}	89.0	0.816	93.7	86.5	87.4
COMB - I_{Fus}	89.4	0.840	93.7	89.2	86.0
		LDA			
ALTM3100 – I_{Raw}	77.1	0.628	88.9	72.7	60.0
ALTM3100 – I_{Ran}	78.9	0.656	90.5	75.2	59.8
ALS50 – I_{Raw}	79.6	0.664	88.7	73.8	77.0
ALS50 – I_{Ran}	87.4	0.785	91.5	86.6	77.7
COMB – I_{Raw}	82.2	0.703	89.2	77.8	79.8
COMB – I_{Ran}	88.5	0.805	91.7	87.5	83.0
COMB - I_{Fus}	90.0	0.830	92.2	90.3	81.7
		RF			
ALTM3100 – I_{Raw}	81.7	0.690	88.4	82.9	55.2
ALTM3100 – I_{Ran}	83.5	0.720	89.7	84.9	57.6
ALS50 – I_{Raw}	82.6	0.706	86.7	81.7	74.0
ALS50 – I_{Ran}	88.5	0.804	92.3	88.4	77.4
COMB – I_{Raw}	85.1	0.750	88.3	83.3	83.2
COMB – I_{Ran}	89.3	0.819	91.5	88.5	86.4
COMB - I_{Fus}	90.8	0.844	93.8	90.4	83.5

The RF method resulted in the best classification performance, although the differences between classifiers were marginal.

The results of the k-NN classification with combined sensor and I_{Fus} variable, which resulted in an overall accuracy of 88.7% (Table 12), were analyzed at the tree and plot levels. The mean values of h_{rel} were 0.88 and 0.91 in the erroneously and correctly classified trees, respectively. The mean proportions of crown coverage by neighboring trees were 45% and 37%, respectively. In logistic regression (GENMOD procedure in SAS 9.13) DBH, h, h_{rel}, crown overlap, and number of echoes explained the binary classification result. These results imply that close neighbors affected the classification as well as the relative and absolute tree size. The classification accuracy was 90%, 85.3%, and 61.7% in the three classes of crown status: normal, defective, and deteriorated. The lowest plot-level accuracy (52%) was observed in a 20-yr-old pine-spruce-birch plot with 246 1–9-m-high trees with an average of 23.2 first-or-only echoes per tree. A low accuracy of 67% was observed in a very dense, unmanaged 35-yr-old birch plot, where the living crowns were narrow and short and had an average of 30 echoes per tree. The RMSE of crown modeling did not explain the classification success, which implies that the point-extraction method (Section 2.3) did not affect the results.

4 Discussion

4.1 Confines of the Study

LiDAR is a widely recognized tool for positioning and measurement of many forest parameters. We examined the potential use of LiDAR in tree species recognition, in the context of STRS, which provides an upper bound for the accuracy achievable in area-based methods. The limits of our study are several. Our experiment was extensive in the quantity of trees and different tree- and stand-level factors that could have affected the performance of the LiDAR data. Our reference data were, however, unbalanced in the main factors of interest: site fertility and age. This could have skewed the results and affected our interpretation, despite our awareness of the dangers. We undertook two comparable LiDAR campaigns; however, variation in important LiDAR parameters (Hopkinson 2007, Næsset 2009), such as the pulse density, pulse-repetition frequency, or flying altitude was lacking. Here the results applied to a 1-km flying altitude only, and we expect that inferior classification results will be obtained for higher flying altitudes. We did not try to estimate the CBH in LiDAR, which is a strong separating factor between pine and spruce, as shown in Holmgren and Persson (2004). However, reliable estimation of CBH in dense canopies or in the presence of understory trees can be difficult. We used LDA, RF, and nearest neighbor classifiers to control the differences and made few attempts to tune the classifiers for an optimal, best-case performance. However, our classification results represent an optimistic STRS scenario, because we collected the point clouds per tree, using manual positioning of treetops and knowledge of the DBH of the tree when crown modeling was applied. In practice, tree detection and segmentation tasks should be carried out automatically, which likely would reduce the species classification accuracy. In addition, we neglected aspen in the classifications, because it is so rare in the study area. Aspen constitutes 1.5% of the stem volume in Finland (Tomppo et al. 2001), although it can locally be an important species where fertile soils prevail. Our results suggest that the intensity features for aspen deviate from those of birch, although these species are often combined in classification studies (Holmgren et al. 2008, Vauhkonen et al. 2009).

4.2 Features with Explanatory Power in Species Classification

The results indicated that the use of intensity of first-or-only echoes is optimal and is explained by the fact that these echoes represent the highest reflections that are less affected by intracrown transmission losses. The use of only returns leads to a reduced number of available measurements. However, if the LiDAR settings are altered, e.g. due to higher flying altitude and pulse-repetition frequency, the proportion of single-echo pulses can be much higher than here (Hopkinson 2007,

Næsset 2009). First-and-only echoes differ conceptually and the only echoes had higher intensity values than the first echoes (Ørka et al. 2009). Hence, we cannot conclude that first-or-only echoes are generally a better choice. However, our results suggest that a strategy in which three separate feature sets are computed, using the first-and-only returns separately and in combination, may be reasonable. We did not test the proportions of the echo types as explanatory variables, since they may be sensitive to LiDAR settings. The proportions were used as explanatory variables in Holmgren et al. (2008).

Feature selection by RF resulted in 10 intensity variables. Similar results were obtained in ANOVA of individual features. However, we did not make an exhaustive test of all 740 variables in each combination of sensor and echo type. The features derived from the height distribution were inferior to intensity features. However, our crown-modeling procedure (Section 2.3, Korpela 2007) may have affected the point-height distribution such that the explanatory power became less. The mean intensity values in the four vertical layers of the living crown showed high explanatory power and interesting associations with the absolute and relative sizes of the tree, especially for birch. This may explain the higher confusion in birch in all classification trials. The mean intensity of the near-surface echoes was not strong variable. Our results differ from those of Holmgren and Persson (2004), who used a surface model approach. The upper percentiles (60th–90th) were crucial features. The maximum intensity was an important feature in Brandtberg et al. (2003) and Ørka et al. (2009), while the percentiles were not tested in these studies.

4.3 What is Intensity Signal Measuring in a Tree?

The intensity of DR sensors is believed to correspond to the maximum backscatter amplitude. Intensity is a measure of reflectance for well-defined surfaces that are larger than the footprint (Kaasalainen et al. 2007). However, canopy reflections give rise to more variable echoes, the structures of which are available in FW data and undoubtedly contain information unavailable in

DR data (cf. Höfle et al. 2008). The footprint or volume that contributes to an echo in the canopy may also contain different scatterers, e.g. bark, needles, cones, flowers, epiphytes, and twigs, all of which vary in reflectance. Our results suggest that the leaf size, orientation, branch structure, and foliage density may determine the intensity in DR sensors. It is, however, impossible to separate these factors. Broad-leaved species such as *Acer platanoides, Salix caprea,* and *Sorbus aucuparia* with large and/or horizontal leaves in a dense and compact configuration had the highest intensity. The depth and density of the foliage layer in these species may have also contributed to the high intensity. This may explain the difference between conifers such as *Abies sibirica*, *Picea abies*, and *Pinus sylvestris*. The intensity difference observed in intermediate/codominant and dominant birches and the dependence on absolute size in birch may be a result of differences in crown architecture. We also observed reduced intensity for echoes in the lower parts of the crown. This may be explained by untraceable transmission losses or by the lower needle density 20–40% down from the treetop. Our results suggest that part of the intraclass variation in the intensity metrics may be due to differences in needle mass and vigor. We did not have increment measurements to test whether intensity metrics would explain intertree variation in growth.

4.4 Intensity Normalization for Noise Removal and Sensor Fusion

In our study area the elevation range of the treetops was 70 m, 147–217 m a.s.l. The flight paths were mostly flown within ± 5% of the nominal flying altitude, 1000 m and 930 m a.s.l for the ALTM3100 and ALS50 sensors, respectively. The effect of the topography and variation in the flying altitude was approximately ± 4.5% in the sensor-to-target range. Similarly, the effect of the scan angle was ± 3.5%. Thus, the range normalization altered the intensities ± 20% ≈ $(1-0.08)^{2.5}$. Range normalization improved the classification accuracy in both sensors, although in the ALS50 we did not separate the AGC and range normalization. Both should be seen as necessary noise removal and preprocessing steps.

Range normalization has, however, been omitted in most of the earlier research dealing with tree species or vegetation classification, except for studies based on the TopEye sensor, which has built-in range correction. Thus, the results obtained in these studies mostly underestimate the true potential. We did not optimize the power factors of 2.4 and 2.5 in Eqs. 1 and 2, which were obtained using nontree targets by minimization of per-target variance (Korpela 2008). Range compensation is an ill-posed task and we believe that our results could be improved by searching for the optimal, sensor-, and species-specific parameters for the task. In addition, no specifications for the functioning of the AGC by the sensor manufacturer of the ALS50 were available during the experiment and our model for the AGC normalization was experimental and derived by the method of reverse engineering. The digital number (DN) values of the AGC in 8 bits were mostly in the range of 118–138, which resulted in approximately ± 30% correction of the intensity. In all, the intratree variance of range- and AGC-normalized intensity values was 57% that of raw data in the ALS50.

We applied sensor fusion to increase the data per tree, particularly to support the crown-modeling procedure. In practice, the fusion of data from different LiDAR sensors or sensor settings can occur, e.g. when the weather does not favor the use of laser scanning and the equipment is taken elsewhere. One requirement would be a good geometric match between the datasets and we propose pseudoecho maps (Fig. 2) for the evaluation of the XY match, which can be complex under forest conditions if artificial linear structures (power lines, buildings, etc.) or sharp edges (e.g. paludified shores) are not available. LiDAR data that contain the trajectory are needed for the maps. The combined sensor data consisted of the per-tree sum distributions of two campaigns. Assuming that the LiDAR settings were unaltered in both sensors during the campaigns and the sampling densities per sensor did not vary spatially, it would have been possible to combine the datasets without weighting of the intensities. However, since our data showed considerable variation in the pulse density (the number of echoes per crown in the two sensors had an R^2 of 0.33) between the two sensors, we attempted to weight the ALS50

intensity observations. This improved the stability of the distribution features and slightly improved the classification results. We used natural and artificial targets for intensity normalization, assuming that the backscatter properties had not changed between campaigns.

4.5 Comparison of the ALTM3100 and ALS50-II Sensors

The best classification performance was obtained, using both the ALTM3100 and ALS50 data together. The overall classification accuracy was 88–90% for pine, spruce, and birch, depending on the classifier, feature set, and validation method. The accuracy was almost at the required level of 95%, which is considered sufficient for STRS applications (Korpela and Tokola 2006). A high number of echoes per tree presumably improved the accuracy by reducing the random noise in the features used. Surprisingly, the Leica ALS50 outperformed the Optech ALTM3100, since in the ALS50 both the AGC and range corrections were needed and in the Optech only the range correction. The 20% higher sampling rate of the ALS50 data may explain the difference, because the distribution metrics are more stable. The 50% smaller footprint area of the ALS50 may also have resulted in a better signal-to-noise ratio. Our results suggest substantial differences in the intensity data between the ALS50 and ALTM3100, especially after range- and AGC normalization. The same performance gap was observed in the classification of reindeer lichens (*Cladina* (Nyl.) Nyl. spp.), using the same first-or-only data (Korpela 2008). However, the differences in the systems may be campaign-dependent, since intensity recording is seldom optimized. More studies are needed to confirm the results.

4.6 Effects of Age and Site Fertility

Some of the intensity variables were dependent on tree size and/or age, especially for birch. These observed effects may be due to changes in the crown architecture, which are related to the age and speed of growth in trees. However, in the classification test where age was used for splitting the

data into training and validation sets, we observed no reductions in classification performance. In a similar test, using site fertility for the division, the accuracy of spruce and birch was reduced when the training set was formed by the most fertile sites, which are favored by these species. Thus, site fertility may affect trees such that it is seen in the intensity signal.

4.7 Size of the Training Set

We were surprised that reduction in the training set to 2.5% (300) of the trees reduced the k-MSN classification accuracy of pine, spruce, and birch to merely 1% of the level achieved, using 30% (3600) of the trees in the training set. On the other hand, similar results were obtained at plot level, using even fewer than 100 training observations (Maltamo et al. 2009). In the case of reduced data sets we applied k-MSN for classification, in which among chosen neighbors there can be wrong tree species but the estimate can still be correct. This emphasizes the fact that the number of observations in certain classes can be lower in training data and differs from the approach of Packalén and Maltamo (2007), in which each wrongly chosen neighbor automatically adds RMSE. Our results show that a rather small amount of *in situ* training data for STRS and species recognition is sufficient, if it represents well the different age classes.

4.8 Ecologically Important Species

The classification of aspen may prove difficult because the intensity metrics overlapped with that of spruce and birch (cf. Ørka et al. 2007, Säynäjoki et al. 2008). However, *Salix caprea*, *Alnus incana,* and *Alnus glutinosa* could potentially be separable based on the high intensity values associated with these species. A simple geographic information system (GIS)-thresholding technique, using the upper intensity deciles and height data combined with a clustering operation, could perhaps unmask the rare occurrences of these species.

4.9 Suggestions for Future Research

We omitted CBH and crown length as predictors, which would likely aid in the classification of spruce (Holmgren and Persson 2004). The use of complex crown shape metrics could also lead to an improvement (Holmgren et al. 2008, Vauhkonen et al. 2009). However, there are practical constraints that limit the available sampling rates and/or flying altitudes of LiDAR. Similarly, the CBH estimation methods should be tested comprehensively in a challenging forest environment. We did not examine the orthogonality of the LiDAR and image features. However, Persson et al. (2004), Heinzel et al. (2008), and Holmgren et al. (2008) showed that auxiliary image features may improve the species classification accuracy of dominant, photovisible, sunlit crowns. The use of NIR image features may provide an aid for the classification of birch, which is differentiated poorly in LiDAR data. Co-use of images and LiDAR adds to the costs and the focus should be on studies that examine the importance of the flying altitude and sampling density. The intensity data in DR LiDAR probably misses some of the explanatory power that is contained in FW data, which has not thus far been examined in Scandinavian forests. An excellent scenario for forest applications would be if DR sensors were developed to register the on-the-fly echo characteristics, which are extracted in postprocessing of FW data. The topographic LiDAR sensors that are currently used in forest applications are primarily optimized for elevation modeling. The range normalization parameters (Eqs. 1, 2) were not optimized in our study and this constitutes a future topic. Based on our preliminary results, the backscatter features may explain needle mass and growth at the tree level, which is an unexplored topic.

5 Conclusions

An accuracy of 88–90% was achieved in the species classification of individual pine, spruce, and birch trees in DR LiDAR intensity data. Spruce and birch showed the lowest classification accuracy. The intensity features were dependent on the

absolute and relative size of trees, especially for birch, which likely explain the lowest accuracy of birch. Among the classifiers tested, RF resulted in the highest classification performance. Our results suggest that leaf size, orientation, and foliage density affect mostly the intensity data, which is not a measure of target reflectance. Some of the rare and ecologically important species in Finland could be separated, since their foliage gave rise to high intensity values. Comparison of the ALS50-II and ALTM3100 sensors suggests that the smaller footprint or the enhanced dynamic range of the ALS50-II may explain the better classification results, using intensity data. Range normalization, i.e. compensation for the two-way losses in the target-to-sensor path, improved the accuracy of species classification and constitutes a necessary data-processing step. Normalization for the receiver's AGC affected the intensity data favorably in the ALS50-II sensor. Fusion of two LiDAR datasets with 1-yr temporal mismatch and different instruments was improved by normalization of the intensity datasets, using natural targets for finding the parameters.

References

Ahokas, E., Kaasalainen, S., Hyyppä, J. & Suomalainen, J. 2006. Calibration of the Optech ALTM 3100 laser scanner intensity data using brightness targets. Revue Française de Photogrammétrie et de Télédétection 182(2): 10–16.

Brandtberg, T. 1999. Automatic individual tree based analysis of high spatial resolution aerial images on naturally regenerated boreal forests. Canadian Journal of Forest Research 29: 1464–1478.

Brandtberg, T., Warner, T.A., Landerberger, 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: 290–303.

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

Breiman, L. 2001. Random forests. Machine Learning 45: 5–32.

Chen, C., Liaw, A. & Breiman, L. 2004. Using random forest to learn imbalanced data. Department of Statistics, University of California, Berkeley. Available at: http://www.stat.berkeley.edu/tech-reports/666.pdf. [Cited 15 August 2009].

Falkowski, M.J., Evans J.S., Martinuzzi, S., Gessler, P.E. & Hudak, A.T. 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. Remote Sensing of Environment 113: 946–956.

Gong, P., Sheng, Y. & Biging, G.S. 2002. 3D model-based tree measurement from high-resolution aerial imagery. Photogrammetric Engineering and Remote Sensing 68: 1203–1212.

Haara, A. & Haarala, M. 2002. Tree species classification using semi-automatic delineation of trees on aerial images. Scandinavian Journal of Forest Research 17: 556–565.

Heinzel, J.N., Weinacker, H. & Koch, B., 2008. Full automatic detection of tree species based on delineated single tree crowns – a data fusion approach for airborne laser scanning data and aerial photographs. In: Hill, R.A., Rosette, J. & Suárez, J. (eds.). Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. September 2008. Edinburgh, UK. p. 76–85.

Höfle, B., Hollaus, M., Lehner, H., Pfeifer, N. & Wagner, W. 2008. Area-based parameterization of forest structure using full-waveform airborne laser scanning data. In: Hill, R.A., Rosette, J. & Suárez, J. (eds.). Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. September 2008. Edinburgh, UK. p. 227–236.

Holmgren, J. & Persson, Å. 2004. Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment 90: 415–423.

— , Persson, Å. & Söderman, U. 2008. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. International Journal of Remote Sensing, 29: 1537–1552.

Holopainen, M. & Talvitie, M. 2006. Effect of data acquisition accuracy on timing of stand harvests and expected net present value. Silva Fennica 40: 531–543.

Hopkinson, C. 2007. The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency

distribution. Canadian Journal of Remote Sensing 33: 312–324.

Kaasalainen, S., Hyyppä, J., Litkey, P., Hyyppä, H., Ahokas, E., Kukko, A. & Kaartinen, H. 2007. Radiometric calibration of ALS intensity. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, Part 3/W52: 201–205.

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: 1575–1586.

Korpela, I. 2000. 3D matching of tree tops using digitized panchromatic aerial photos. Licentiate Thesis. Department of Forest Resource Management, University of Helsinki. 109 p.

— 2004. Individual tree measurements by means of digital aerial photogrammetry. Silva Fennica Monographs 3. 93 p.

— 2007. Incorporation of allometry in single-tree remote sensing with LiDAR and multiple images. In: Heipke, C., Jacobsen, K. & Gerke, M. (eds.). Proceedings of the ISPRS Hannover workshop 2007. High Resolution Earth Imaging for Geospatial Information. Hannover, Germany, May 29–June 1, 2007. ISPRS 36 Part I/W51, ISSN No. 1682–1777. CD-ROM. 6 p.

— 2008. Mapping of understory lichens with airborne discrete-return LiDAR data. Remote Sensing of Environment 112: 3891–3897

— & Tokola, T. 2006. Potential of aerial image-based monoscopic and multiview single-tree forest inventory – a simulation approach. Forest Science 52: 136–147

—, Tuomola, T. & Välimäki, E. 2007. Mapping forest plots: an efficient method combining photogrammetry and field triangulation. Silva Fennica 41: 457–469

—, Tuomola, T., Tokola, T. & Dahlin, B. 2008. Appraisal of seedling stand vegetation with airborne imagery and discrete-return LiDAR – an exploratory analysis. Silva Fennica 42: 753–772.

—, Koskinen, M., Vasander, H., Holopainen, M. & Minkkinen, K. 2009. Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. Forest Ecology and Management 258: 1549–1566.

Liaw, A. & Wiener, M. 2002. Classification and regression by randomForest. R News 2(3): 18–22.

Maltamo, M., Bollandsås, O.M., Næsset, E., Gobakken, T. & Packalén, P. 2009. Different sampling strategies for field training plots in ALS inventory. In: Popescu, S., Nelson, R., Zhao, K. & Neuemshwander, A. (eds.). Silvilaser 2009 Proceedings. College Station, Texas, USA, October 14–16, 2009. ISBN: 978-1-61623-997-8. CD-ROM

Miller, D., Quine, C. & Hadley, W. 2000. An investigation of the potential of digital photogrammetry to provide measurements of forest characteristics and abiotic damage. Forest Ecology and Management 135: 279–288.

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

Morsdorf, F., Frey, O., Koetz, B., & Meier, E. 2007. Ray tracing for modeling of small footprint airborne laser scanning returns. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36, Part 3/W52: 294–299.

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.

— 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment 113: 148–159.

Ørka, H.O., Næsset, E. & Bollandsås, O.M. 2007. Utilizing airborne laser intensity for tree species classification. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36, Part 3/W52: 300–304.

—, 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: 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: 328–341.

Persson, Å., Holmgren, J. & Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineer-

ing and Remote Sensing 68: 925–932.

—, Holmgren, J., Södermann, U. & Olsson, H., 2004. Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near-infrared digital images. In Thies, M., Koch, B., Spiecker, H. & Weinacker, H. (eds.). Proc. of the ISPRS WG VIII/2 'Laser-Scanners for Forest and Landscape Assessment'. IAPRS, XXXVI, Part 8/W2: 204–207.

Reitberger, J., Krzystek, P., Stilla, U. 2006. Analysis of full waveform LiDAR data for tree species classification. In: Förstner, W. & Steffen, R. (eds.). Symposium of ISPRS Commission III Photogrammetric Computer Vision PCV ' 06. Bonn, Germany. 20–22 September 2006. IAPRS, XXXVI, Part 3: 228–233.

—, 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: 1407–1431.

Säynäjoki, R., Packalén, P., Maltamo, M., Vehmas, M. & Eerikäinen, K. 2008. Detection of aspens using high resolution Aerial Laser Scanning data and digital aerial images. Sensors 8: 5037–5054.

Tomppo, E., Henttonen, H. & Tuomainen, T. 2001. Valtakunnan metsien 8. inventoinnin menetelmä ja tulokset metsäkeskuksittain Pohjois-Suomessa 1992–94 sekä tulokset Etelä-Suomessa 1986–92 ja koko maassa 1986–94. Metsätieteen aikakauskirja 1B/2001: 99–248. (In Finnish).

—, Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112: 1982–1999.

Vauhkonen, J., Tokola, T., Packalen, P. & Maltamo, M. 2009. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. Forest Science 55: 37–47.

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

Wagner, W., Ullrich, A., Ducic, V., Melzer, T. & Studnicka, N. 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. ISPRS Journal of Photogrammetry and Remote Sensing 60: 100–112.

Wanninger, L. 2005. Introduction to network RTK. IAG Working Group 4.5.1: Network RTK. Available at http://www.network-rtk.info/intro/introduction.html. [Cited 15 Aug 2009].

Total of 46 references