International Journal of Applied Earth Observations and Geoinformation 100 (2021) 102334

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Classification of tree species classes in a hemi-boreal forest from multispectral airborne laser scanning data using a mini raster cell method

Eva Lindberg *, Johan Holmgren, Håkan Olsson

Swedish University of Agricultural Sciences, Department of Forest Resource Management, 901 83 Umeå, Sweden

ARTICLE INFO

Keywords:
Multispectral LiDAR
Tree species maps
Laser wavelengths
Laser light reflectance

ABSTRACT

Classification of tree species or species classes is still a challenge for remote sensing-based forest inventory. Operational use of Airborne Laser Scanning (ALS) data for prediction of forest variables has this far been dominated by area-based methods where laser scanning data have been used for estimation of forest variables within raster cells. Classification of tree species has however not been achieved with sufficient accuracy with area-based methods using only ALS data. Furthermore, analysis of tree species at the level of raster cells with typical size of 15 m × 15 m is not ideal in the case of mixed species stands. Most ALS systems for terrestrial mapping use only one wavelength of light. New multispectral ALS systems for terrestrial mapping have recently become operational, such as the Optech Titan system with wavelengths 1550 nm, 1064 nm, and 532 nm. This study presents an alternative type of area-based method for classification of tree species classes where multispectral ALS data are used in combination with small raster cells. In this "mini raster cell method" features for classification are derived from the intensity of the different wavelengths in small raster cells using a moving window average approach to allow for a heterogeneous tree species composition. The most common tree species in the Nordic countries are *Pinus sylvestris* and *Picea abies*, constituting about 80% of the growing stock volume. The remaining 20% consists of several deciduous species, mainly *Betula pendula* and *Betula pubescens*, and often grow in mixed forest stands. Classification was done for pine (*Pinus sylvestris*), spruce (*Picea abies*), deciduous species and mixed species in middle-aged and mature stands in a study area located in hemi-boreal forest in the southwest of Sweden (N 58°27', E 13°39'). The results were validated at plot level with the tree species composition defined as proportion of basal area of the tree species classes. The mini raster cell classification method was slightly more accurate (75% overall accuracy) than classification with a plot level area-based method (68% overall accuracy). The explanation is most likely that the mini raster cell method is successful at classifying homogenous patches of tree species classes within a field plot, while classification based on plot level analysis requires one or several heterogeneous classes of mixed species forest. The mini raster cell method also results in a high-resolution tree species map. The small raster cells can be aggregated to estimate tree species composition for arbitrary areas, for example forest stands or area units corresponding to field plots.

1. Introduction

Accurate classification of tree species is still a challenge for remote sensing-based forest inventory. Estimates of forest variables related to height and density such as tree height and stem volume can be derived with high accuracy from Airborne Laser Scanning (ALS) (Næsset 2002, Persson et al. 2002, Maltamo et al. 2006, Wulder et al. 2012, Nilsson et al. 2015). Passive multispectral optical sensors provide additional information about tree species (Lillesand et al. 2007, Fassnacht et al. 2016). The combination of ALS data and passive optical sensors is used operationally for forest inventory for example in Finland (Packalén and Maltamo 2006, Packalén and Maltamo 2007, Packalén and Maltamo 2008, Packalen et al. 2009), but the tree species information in the current automated inventories should preferably be more accurate to fulfil the needs in operational forestry (Maltamo et al. 2014).

Operational use of ALS data for prediction of forest variables has this far been dominated by area-based approaches (ABA). Area-based methods are based on correlations between statistical features from ALS data and forest inventory data over certain area units with typical size of 15 m × 15 m. Statistical features are derived from the ALS data

* Corresponding author.
E-mail addresses: eva.lindberg@slu.se (E. Lindberg), johan.holmgren@slu.se (J. Holmgren), hakan.olsson@slu.se (H. Olsson).

https://doi.org/10.1016/j.jag.2021.102334
Received 5 May 2020; Received in revised form 28 March 2021; Accepted 30 March 2021
Available online 15 April 2021

inside raster cells and used as independent variable in models for field-measured forest variable such as mean tree height and stem volume per hectare (Vauhkonen et al. 2014). Area-based methods require less dense ALS data than individual tree methods. To the best of our knowledge, area-based methods using only ALS data have not been shown to provide information to separate tree species, since species identification requires more complex methods than those used to estimate stem volume or basal area with area-based analysis (White et al. 2016). Since most ALS systems that have been used are not designed to provide information about radiant flux, the recorded intensity should not be interpreted physically and has rarely been used for tree species identification from ALS data. Additionally, a mix of tree species can complicate the species classification since the features will be influenced by different tree species and several mixed species classes with different tree species compositions might be needed.

Species classification of individual trees from ALS data has the advantage that features describing the 3D structure of the tree crown can be included (Brandtberg et al. 2003, Vauhkonen et al. 2009, Heinzel and Koch 2012, Dalponte et al. 2014) as well as features describing spectral properties of the tree crowns (Holmgren and Persson 2004, Budei et al. 2017, Yu et al. 2017, Axelsson et al. 2018) and the full waveform if such information is available (Reitberger et al. 2008). Additionally, no mixed species classes are needed. However, individual tree approaches require high density ALS data (i.e., point density at least in the order of 5 per m^2) and more advanced algorithms for processing the ALS data and delineating individual tree crowns than area-based methods (Hyyppä et al. 2006). Additionally, the parameters of the delineation method must be adapted to the local forest condition, which is a non-trivial task and often requires field data with positions of individual trees as training data (e.g., Holmgren and Lindberg 2013).

Passive optical sensors have the disadvantage that the recorded spectral values depend on the sunlight conditions, atmosphere, and shadowing, which means that the spectral values for a given species are not comparable between different images and can even vary within an image (e.g., Korpela 2004). Classifying tree species based on spectral information alone gives different results depending on sensors and forest types (e.g., Olofsson et al. 2006). The variation caused by directional reflectance anisotropy in combination with varying view angles makes automatic classification challenging, in particular when using passive airborne sensors (Korpela et al. 2011). Research has been aimed at performing radiometric correction to remove the brightness variations (Pellikka et al., 2000). From imagery with high resolution (e.g., in the order of 1 m), information about tree species can be derived through texture features, in particular for deciduous trees (Franklin et al. 2000). For very high resolution multispectral stereo-imagery (VHRSI; less than1 m pixels), individual tree crowns can be delineated (Lindberg and Holmgren 2017) and their species classified (Fassnacht et al. 2017). Classification of individual tree crowns delineated from aerial images has been done based on spectral variables for pine, spruce, and deciduous trees (Haara et al., 2002) and geometrical properties for pine, spruce, birch and aspen (Brandtberg 2002).

Active sensors such as ALS are more independent of light conditions since the sensor itself emits the light for which the reflection is measured. Another advantage of ALS data is that data from the tree

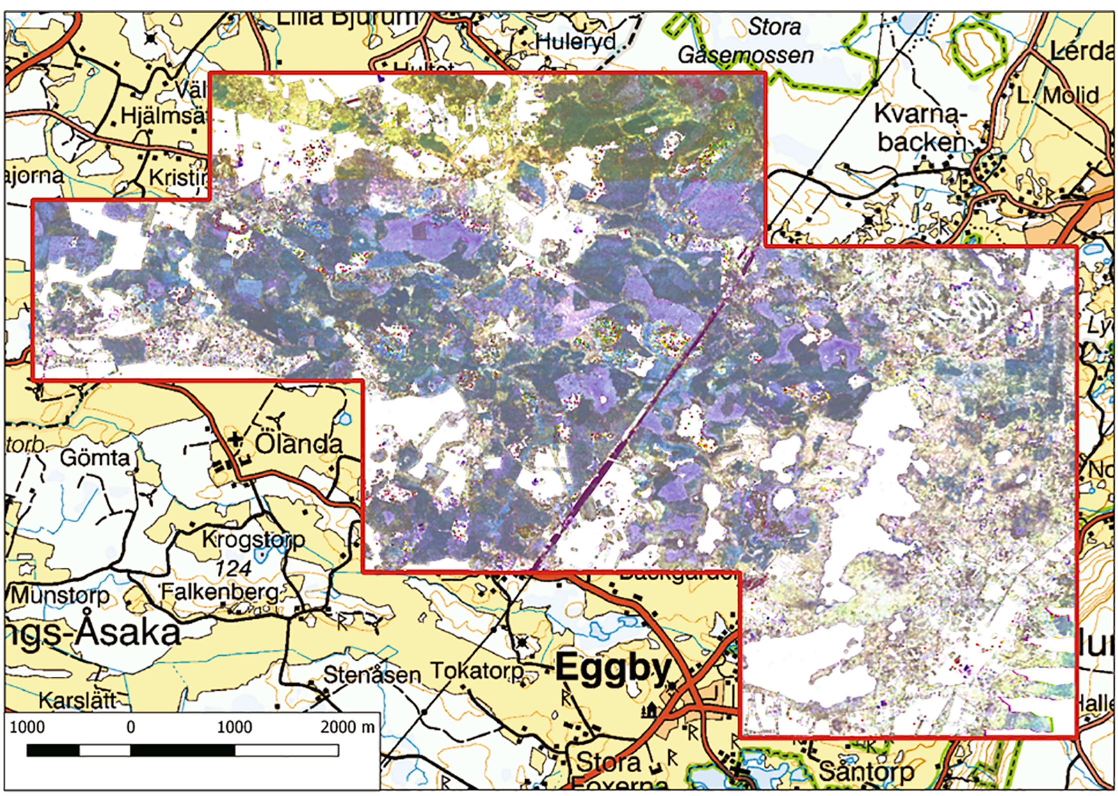

Fig. 1. The study area with the laser-scanned area inside the red polygon. The laser-scanned area is a false color composite image of raster from range corrected laser intensities for returns from the tree canopies with the wavelength of the Optech Titan system 1550 nm displayed as red, 1064 nm as green, and 532 nm as blue. The background (i.e., areas with no ALS returns at least 2 m above the DTM) inside the red polygon is white. The line from south-west to north-east is a power line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

canopy and the ground can be separated to avoid mixed signals, a common problem for passive optical sensors (Li and Strahler 1985).

The ALS systems that have been commercially available for terrestrial mapping since the mid 1990s use only one wavelength of light and most sensors are not designed to measure differences in radiant flux. Attempts to use intensity from ALS data for tree species classification have shown that the performance of intensity from different ALS systems varies. Even the intensity from the same ALS system varies due to differences in reflectance of different scatterers (e.g. bark, needles, cones, flowers, epiphytes, and twigs) as well as differences in leaf size, orientation, branch structure, and foliage density, depth and density of the foliage layer, and crown architecture (Korpela et al. 2010). In the case of multiple returns for one emitted pulse, parts of the energy of the emitted pulse is reflected in the first return and the intensity of the second return depends on the intensity of the first return (Wagner et al. 2008). Additionally, the incident angle of the pulse affects the intensity (Wagner 2010).

Since the reflectance for different wavelengths differ between tree species, ALS data with more than one wavelength have the potential to characterize tree species better (Lindberg et al. 2015). New multispectral ALS systems have recently become operational, such as the Optech Titan system with wavelengths 1550 nm, 1064 nm, and 532 nm and the Riegl VQ-1560i-DW system with wavelengths 1064 nm and 532 nm. The wavelengths differ in the proportions of returns at ground level, vertical foliage distributions, and gap probability across wavelengths. These differences could be exploited in intensity-based land cover classification (Hopkinson et al. 2016). The use of data from the Optech Titan system has been demonstrated for land cover classification with spectral and geometrical features (Teo and Wu 2017) and spectral, elevation and texture features (Bakuła et al. 2016) as well as for object-based classification (Matikainen et al. 2017). Multispectral ALS data have been shown to provide similar accuracy for classification of dominant tree species classes at plot level as a combination of ALS data with one wavelength and aerial images (Kukkonen et al. 2019). A few studies have used data from the Optech Titan system for tree species classification based on structural and intensity features with promising results (Ahokas et al. 2016, Budei et al. 2017, Yu et al. 2017, Axelsson et al. 2018). Ahokas et al. (2016) classified individual trees in a boreal forest into pine, spruce and birch with a maximum overall accuracy of 93.5%. Budei et al. (2017) classified individual trees of ten species of coniferous and deciduous trees in a temperate forest with an overall accuracy of 76%. Yu et al. (2017) classified the species of individual trees of *Pinus Sylvestris*, *Picea Abies*, and *Betula* sp. in a hemi-boreal forest with an accuracy of 90.5% for isolated trees. Axelsson et al. (2018) classified the species of solitary individual trees of nine genera of coniferous and deciduous trees in the hemi-boreal region with an accuracy of 76.5% for a combination of spectral ALS features and structural ALS features compared to 43.0% when using only structural features. The wavelengths of the multispectral ALS systems that are commercially available have not been selected specifically for tree species mapping, but have still proven to be useful for this. Even more accurate tree species characterization could potentially be obtained by selecting wavelengths based on their ability to separate tree species (Vauhkonen et al. 2013).

The aim of this study is to present and evaluate an alternative type of area-based method using small raster cells for classification of tree species classes, in this case Scots pine, Norway spruce, and deciduous species from multispectral ALS data. The method is from hereon called the mini raster cell method. In the mini raster cell method, statistical features are derived from the ALS data inside raster cells and used as independent variable in models to estimate forest variables, but the raster cells are smaller than what has previously been used to allow for more spatial heterogeneity (i.e., mixed species classes). The features used for the classification are derived from the intensity of the different wavelengths inside small raster cells using a moving window average approach to allow for a heterogeneous tree species composition. The mini raster cell method results in a high-resolution (i.e., raster cell size 0.5 m) tree species raster map. The results are validated at plot level with the tree species composition defined as the proportion of basal area of the tree species classes.

2. Material and methods

2.1. Study area

The study area is located in hemi-boreal forest in the southwest of Sweden (N 58°27′, E 13°39′) in the Remningstorp forest estate with an area of 1602 ha (Fig. 1). The western part of the area consists of managed, mostly planted coniferous-dominated forest in homogenous forest stands. A forest management plan created from manual interpretation of aerial images is available for this part of the area including a map created by manual interpretation of aerial images in a photogrammetric work station. The map shows forest stand boundaries and tree species proportions in each stand. The eastern part of the area consists of mostly naturally generated deciduous and mixed species forest and pasture land. The main tree species are Norway spruce [*Picea abies (L.)*], Scots pine [*Pinus sylvestris (L.)*], and birch [*Betula pendula (L.)* and *Betula pubescens (L.)*]. Other tree species are oak [*Quercus robur (L.)*], black alder [*Alnus glutinosa (L.)*], maple [*Acer platanoides (L.)*], aspen [*Populus tremula (L.)*], rowan [*Sorbus aucuparia (L.)*], as well as other broadleaved trees and shrubs, primarily hazelnut [*Corylus avellana (L.)*].

2.2. Field data

Two hundred and fifty-one circular field plots with 10 m radius were

Table 1
Summary of all field plots and subsets of field plots.

	Number of field plots	Number of trees per hectare (mean / standard deviation)	Mean DBH (cm, mean / standard deviation)	Basal area (m^2/ha, mean / standard deviation)
All field plots	251	745 / 605	22.5 / 11.6	20.2 / 13.8
Field plots with mature forest	186	892 / 516	26.0 / 8.1	26.1 / 10.7
Field plots dominated by pine	22	606 / 280	30.2 / 4.7	26.1 / 6.5
Field plots dominated by spruce	89	757 / 342	25.8 / 7.0	29.0 / 11.9
Field plots dominated by deciduous trees	23	920 / 543	27.0 / 12.5	20.7 / 7.8
Field plots with mixed forest	52	1232 / 646	24.1 / 8.1	23.4 / 9.5

allocated during August-October 2016 and March 2017. The positions of the field plot centers were measured using a Trimble GeoXR 6000 differential GPS (Trimble 2013) with accuracy below 1 m under canopy cover. Within the field plots, the diameter at breast height (DBH) of all trees and shrubs with DBH ≥ 4 cm was measured using a caliper and the species was recorded (The Heureka project). The positions of the trees and shrubs were measured relative to the plot center using a PosTex ultrasound instrument (Lämås 2010). A summary of the field plots is given in Table 1.

Only field plots with at least seven trees, basal area at least 7 m^2/ha, and mean DBH ≥ 10 cm were included in the analysis, in total 186 plots (Table 1). The selection was done to make sure that all plots had closed forest and that young forest was excluded. Young forest differs spectrally from older forest and tree species classification from remotely sensed data is usually limited to middle-aged and mature forest (Peterson 1993). The field plots were divided into four groups: Scots pine-dominated forest, Norway spruce-dominated forest, deciduous-dominated forest (i.e. at least 80% of the basal area was Scots pine, Norway spruce or deciduous trees, respectively), and mixed species (Table 1). The threshold is a trade-off between having a pure species composition and having enough plots for training and validation.

2.3. ALS data acquisition and preprocessing

Multispectral ALS data were acquired on July 21, 2016 using the Optech Titan X system. The system uses three wavelengths: 1550 nm (C1; SWIR), 1064 nm (C2; NIR), and 532 nm (C3; green). The flying altitude was 400 m above ground and the return density was 30–40 per m^2 per channel. The flying altitude was chosen to obtain a sufficiently strong signal in the green channel.

A range-corrected intensity was calculated by multiplying with the squared distance from the reflecting surface to the scanner for each return (Kukkonen et al., 2019). The height above the ground of each return was calculated by subtracting the height of a digital terrain model (DTM) from Lantmäteriet (the Swedish Land Survey), which was derived from sparse ALS data with country-wide coverage. A normalized digital surface model (nDSM) with raster cell size 0.5 m was derived from the maximum height above the ground in each raster cell.

2.4. Data processing for mini raster cell method

The mini raster cell method was implemented as follows (Fig. 2): Raster images of the mean values of the intensity in 0.5 m raster cells were calculated separately for each channel of the multispectral ALS data. The raster cell size was selected to be large enough to always have data from each channel in each raster cell. Only first returns at least 2 m above the ground and less than 2 m from the top of the nDSM (i.e., in a 2 m thick layer at the top of the canopy) were included. From the resulting raster images, the mean and standard deviation were calculated in circles with 5 m radius centered on the field plots for each wavelength. All raster calculations were done using the raster package in R (Hijmans 2019).

The feature selection was done by comparing the overall accuracy (i.e., the share of correctly classified field plots) resulting from classification from each feature separately. The highest overall accuracy was achieved for classification from the mean values of the intensities of the three different wavelengths. Due to this, the mean intensities were selected for further analysis.

The field plots with forest dominated by one species class (i.e., Scots pine, Norway spruce, or deciduous species with basal area at least 80% of the total; Table 1) were used as training data for classification with linear discriminant analysis (LDA). As a comparison, we did the same analysis with Quadratic Discriminant Analysis (QDA). The response variable was $Speciesclass_{mini}$ with values Scots pine, Norway spruce or

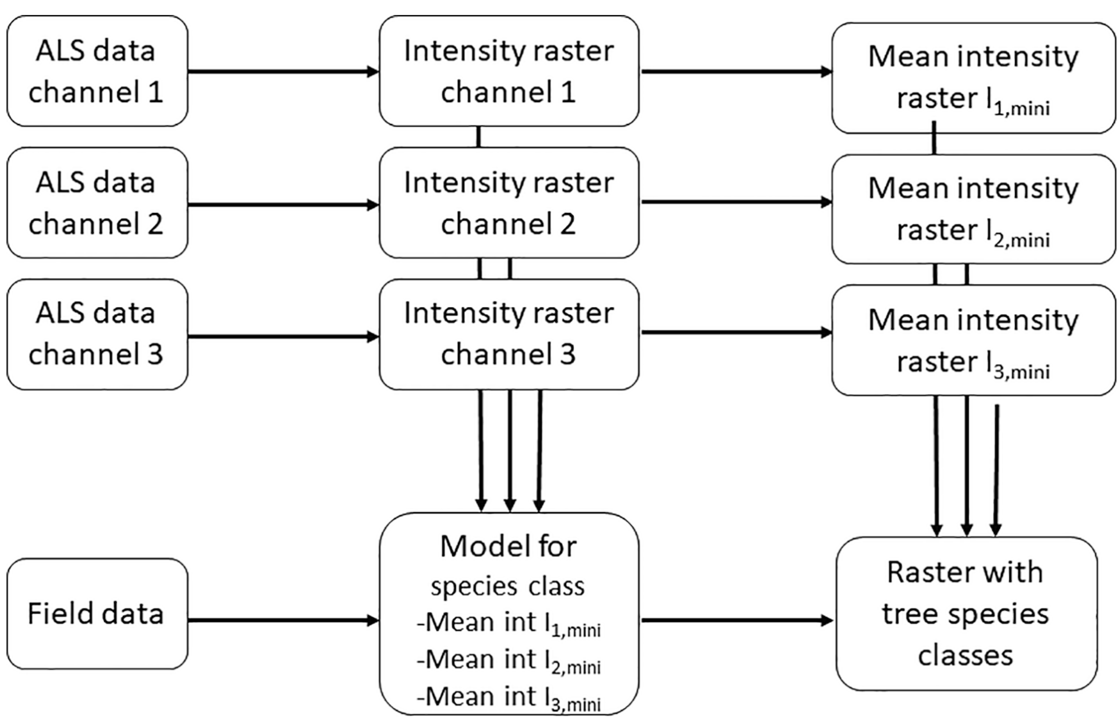

Fig. 2. Flowchart of the mini-raster method.

deciduous species and the explanatory variables were $I_{k,mini}$ which was the mean value of the raster cells in the original intensity raster in a circle with 5 m radius for the wavelengths k = 1,2, and 3. As a comparison, the same analysis was also done for circles with 3 m radius.

Based on the original intensity raster images, new raster images with 0.5 m raster cells were derived using a moving window average with a radius of 5 m (i.e., the mean value of the original intensity raster image within 5 m radius; Fig. 1) and 3 m. The idea was that the area of the moving window should contain only a few trees to make it likely that all were of the same species class. The circular moving window gave rise to circular patches in areas with very few ALS returns (Fig. 3).

The LDA model was applied to the raster images derived with moving window average. The result was a raster showing the tree species classes (i.e., Scots pine, Norway spruce or deciduous species) in each 0.5 m raster cell.

2.5. Data processing for plot level analysis

As a comparison, classification of the field plots was done with LDA from the mean intensity in three wavelengths of all raster cells in the original intensity raster images within each field plot (i.e., plot-level analysis). For this analysis, the training data were all field plots in four classes: Scots pine, Norway spruce, and deciduous-dominated plots (i.e., plots where the basal area of one tree species class made up at least 80% of the total) and mixed species, where no tree species class fulfilled the basal area condition (Table 1). The response variable was $Speciesclass_{ABA}$ with values Scots pine, Norway spruce, deciduous species or mixed species and the explanatory variables were $I_{k,mean}$ which was the mean value of ALS returns at least 2 m above the DTM in a circle with 10 m radius (i.e., corresponding to the field plot radius) of the intensity for k = 1,2, and 3.

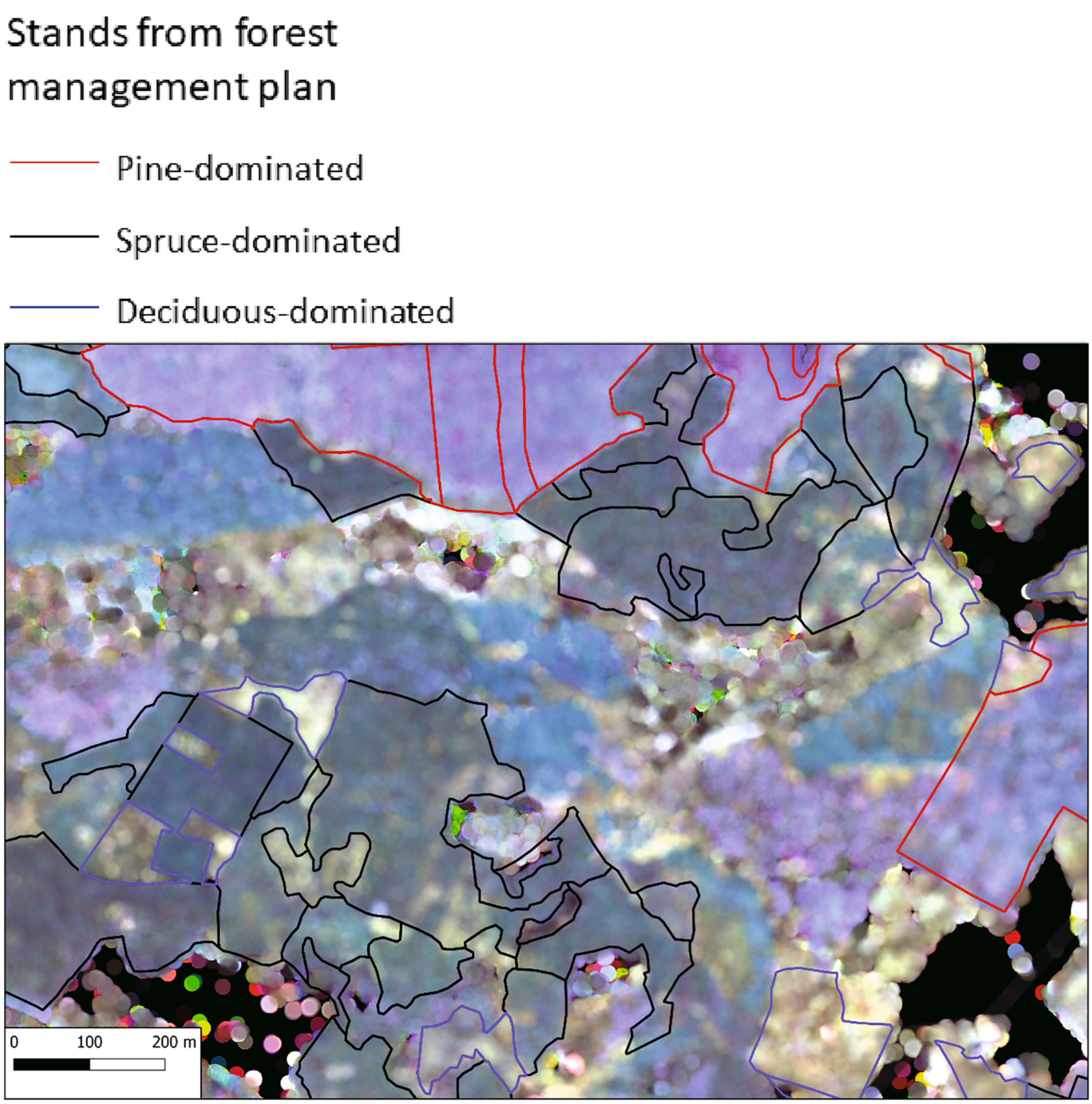

Fig. 3. Example of the image in Fig. 1 in higher resolution. The lines are borders of forest stands from the forest management plan with mean age at least 20 years and one dominant tree species (i.e., at least 80% of the stem volume). Areas outside those forest stands don't fulfill the criteria.

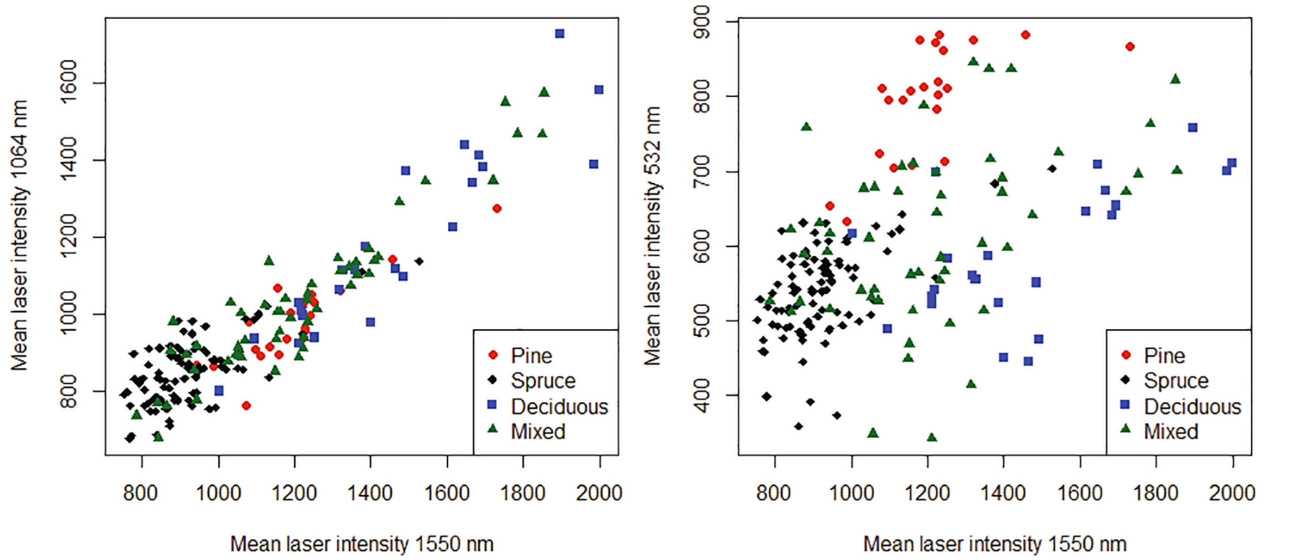

Fig. 4. Scatter plot for different tree species of the mean intensity of the 1550 nm channel versus the 1064 nm channel (left) and the 1550 nm channel versus the 532 nm channel (right).

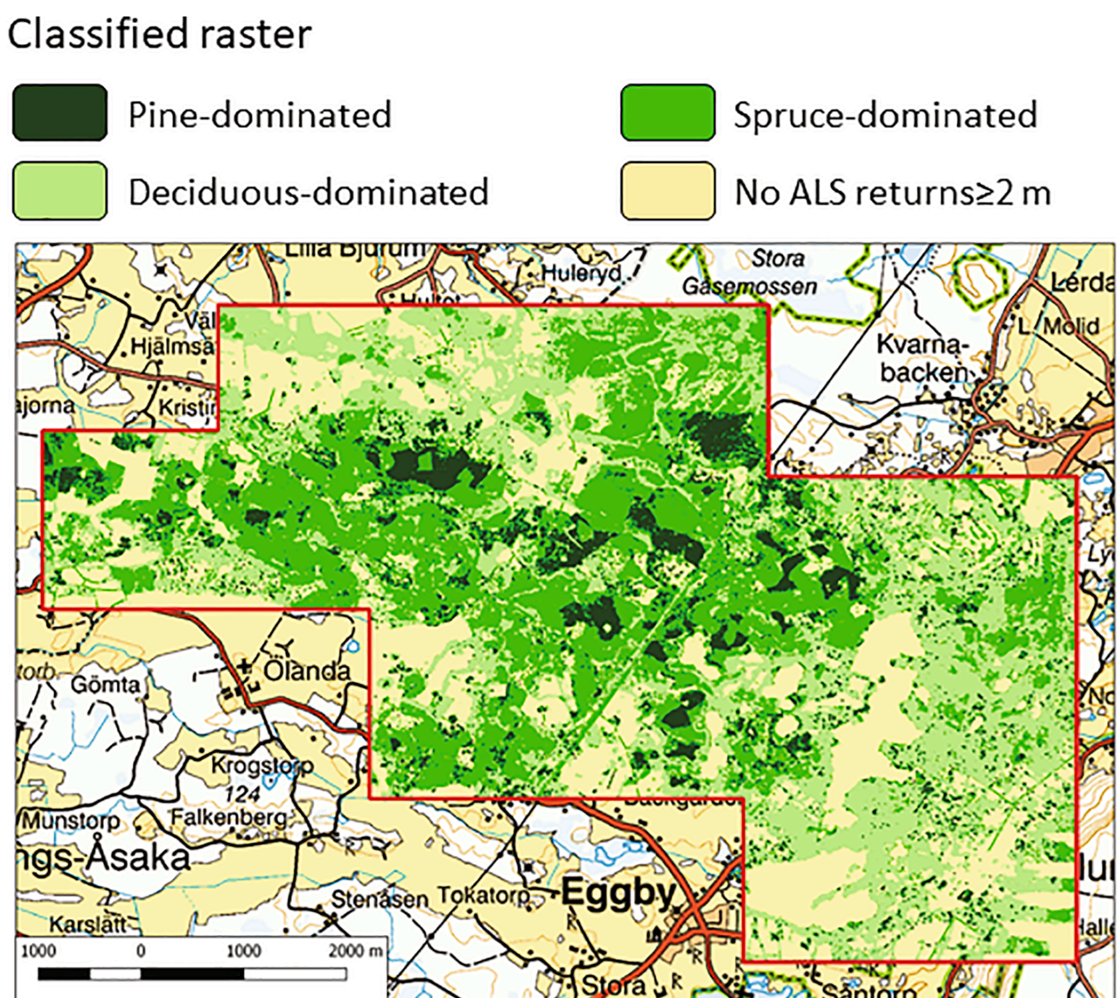

Fig. 5. The study area with the classified raster from the mini raster cell method inside the red polygon. The background (i.e., areas with no ALS returns at least 2 m above the DTM) inside the red polygon is light yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.6. Validation

For the mini raster cell method, the raster cells with their centers inside a radius of 10 m around each field plot center were summarized to calculate the tree species composition. The proportion of each tree species class in a field plot was calculated as the number of raster cells assigned to the tree species class divided by the total number of raster cells in the field plot. All 186 field plots with closed middle-aged and mature forest were used for validation of the resulting tree species raster image. Each field plot was assigned the class given by at least 80% of the raster cells within the field plot. If none of the tree species classes fulfilled that criterion, the plot was classified as mixed species. In other words, field plots where the raster cells were classified as different tree species classes and with no tree species class making up at least 80% of the total number of raster cells were considered as mixed species.

The validation was done using leave-one-out cross-validation for one field plot at a time, meaning that the classification was done for one field plot at a time while the field plot was excluded from the training data. This was done both for the classification results from the mini raster cell method and for the plot level analysis. A confusion matrix was calculated by comparing the classification result with the dominant tree species class derived from the field data, where a dominant tree species class was defined as the tree species class with at least 80% of the basal area in one field plot.

3. Results

The mean intensity of 1064 nm and 1550 nm were highly correlated while the intensities of 1550 nm and 532 nm were less correlated (Fig. 4).

The mini raster cell method resulted in a tree species map over the laser-scanned area (Fig. 5). Thanks to the small pixels used in the mini

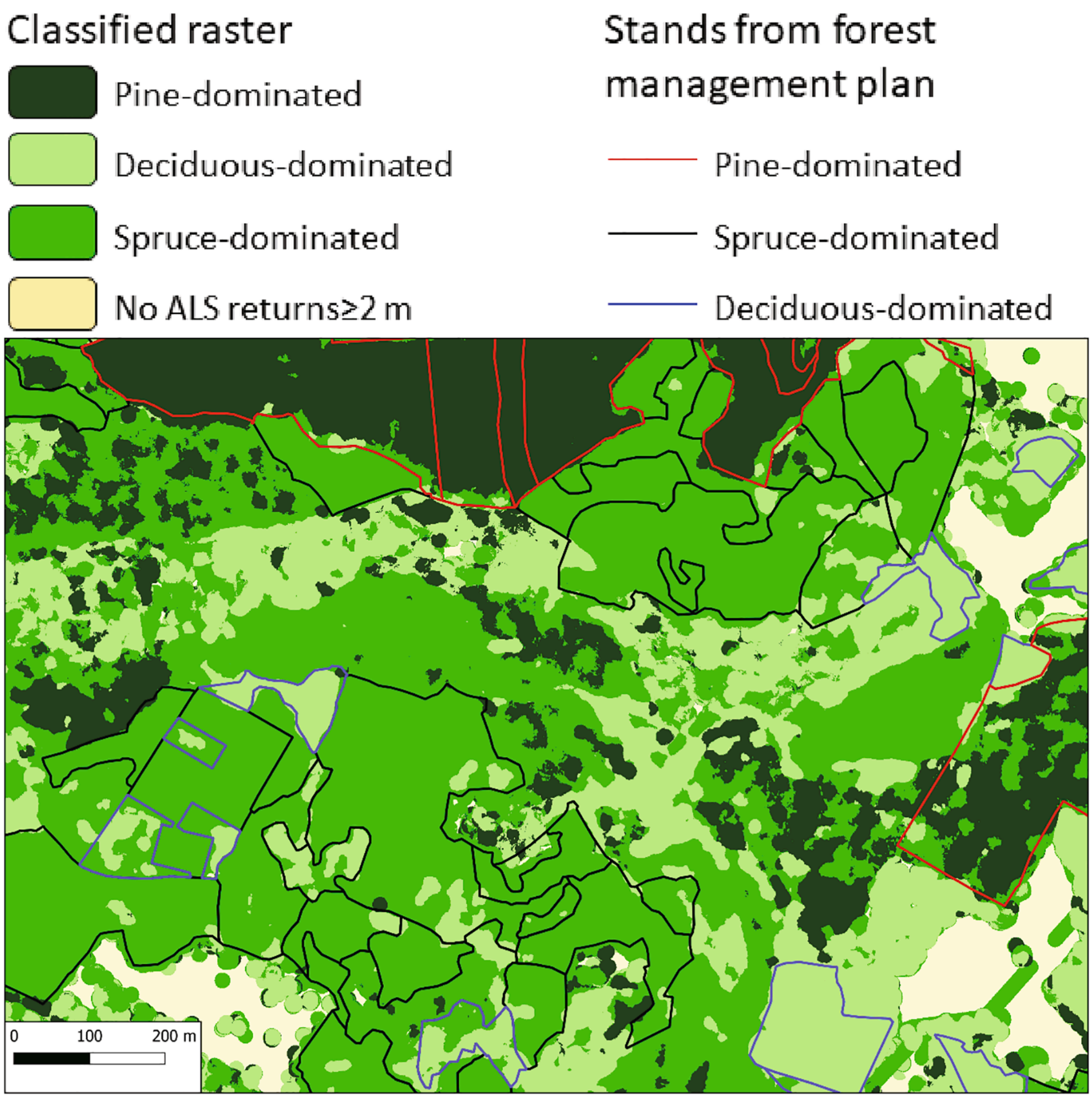

Fig. 6. Examples of the raster in Fig. 5 in higher resolution. The lines are borders of forest stands with mean age at least 20 years and one dominant tree species (i.e., at least 80% of the stem volume) from the forest management plan. Areas outside those forest stands don't fulfill the criteria.

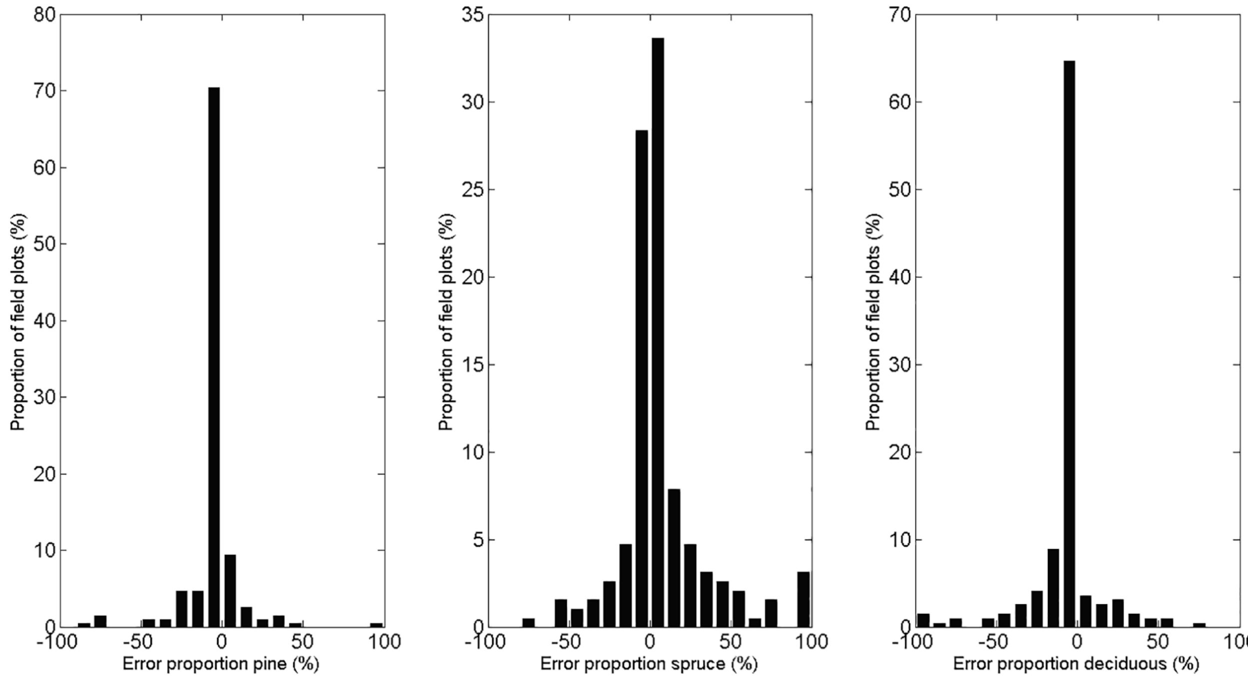

Fig. 7. Histograms of the errors of the proportion of different tree species in the field plots from the classified raster from the mini raster cell method relative to the proportions from the field data.

Table 2
Confusion matrix from classification based on average laser data intensities for field plots (i.e., plot-level analysis).

		Classified				Producer's accuracy
		Pine	Spruce	Deciduous	Mixed	
Field	Pine	19	2	0	1	86%
	Spruce	0	84	0	5	94%
	Deciduous	0	2	13	8	57%
	Mixed	5	25	11	11	21%
User's accuracy		79%	74%	54%	44%	68%

Table 3
Confusion matrix from classification based on raster cells corresponding to the field plot (i.e., mini raster cell method).

		Classified				Producer's accuracy
		Pine	Spruce	Deciduous	Mixed	
Field	Pine	17	2	0	3	77%
	Spruce	0	86	0	3	97%
	Deciduous	0	2	15	6	65%
	Mixed	3	16	11	22	42%
User's accuracy		85%	81%	58%	65%	75%

raster method, the tree species map could be visually compared with stand boundaries from the forest management plan (Fig. 6).

The confusion matrix showed high accuracy for Norway spruce and Scots pine and slightly lower accuracy for deciduous species and mixed species both for the classification based on plot level analysis (Table 2) and based on the mini raster cell method (Table 3). The results from the mini raster cell method were slightly more accurate with overall accuracy 75% with Cohen's kappa coefficient 0.62 compared to overall accuracy 68% with Cohen's kappa coefficient 0.50 for the plot level analysis. The difference in accuracy between the methods was tested with McNemar's Chi-squared test (Dietterich 1998), resulting in a p-value of 0.008829. The higher accuracy of the mini raster cell method was mainly due to higher accuracy for the mixed species class, both producer's and user's accuracy. The results from a 3 m radius (Table A1 and Fig. A1) and QDA (Table A2 and Fig. A2), respectively, resulted in similar accuracy for the classification (please see appendix).

For the mini raster cell method, a comparison of the proportion of tree species classes in the tree species map with the proportion from the field data showed that the difference was mostly close to zero for Scots pine and deciduous species, but slightly larger for Norway spruce (Fig. 7). The proportion of field plots where the difference was less than 20% was 87% for Scots pine, 75% for Norway spruce and 80% for deciduous species.

4. Discussion

4.1. Comparison between mini raster method and plot level analysis

The classification based on the mini raster cell method was slightly more accurate than classification based on plot level analysis with similar intensity metrics, in particular for the mixed species class. The explanation is probably that the mini raster cell method is successful at classifying patches of trees with the same tree species class. On the other hand, classification based on plot level analysis used metrics derived from the whole field plots, which increased the likelihood that the analyzed area contained a mix of tree species classes with a mixed signal as the result. If the classification is done based on plot level analysis, one or several classes of mixed species are needed, but those classes will be heterogeneous, making it difficult to separate them from other classes. To the best of our knowledge, the approach to classify tree species classes in small area units in a two-step processes and without delineation of individual tree crowns has not been used before. The closest previous study was by Packalen et al. (2009), who classified laser returns into tree species classes based on spectral data from aerial images and training data from forest stands with homogenous tree species. The laser returns were then used for area-based analysis of species-specific forest variables. This utilized information about the proportion of different tree species classes that can be derived from the laser returns, but the forest variables were still estimated for larger raster cells without the option of deriving a high-resolution tree species map or aggregating the result to arbitrary area units.

4.2. Properties of the tree species maps

The mini raster cell method used a subset of field plots with one dominant tree species class as training data to enable classification of homogenous patches of tree species classes from the ALS data. A potential source of error is that the intensity of some parts of the tree crowns differ from the major part of the tree crowns of that species class, which leads to small patches being classified as a different species class. This could be solved by applying morphological operators on the tree species map to remove patches smaller than a certain size. The aggregation to field plots defined the class of each field plot, which means that some field plots were classified as mixed species. The aggregation also resulted in a tree species composition in each field plot, which meant more information than for classification based on plot level analysis. The tree species composition can be combined with estimates of stem volume from ALS data to derive species-specific estimates of stem volume. Additionally, the mini raster cell method has the potential to provide a more detailed tree species maps showing the location of different tree species classes, which can be used to delineate micro-stands or treatment units without predefined stand borders (Hyvönen et al., 2005).

4.3. Classification results for different tree species classes and compared to other studies

The most common tree species classes were Norway spruce and Scots pine where the classification accuracy was the highest. Earlier studies using spectral data for tree species classification have shown slightly lower accuracy for separation between Scots pine and Norway spruce. Ørka et al. (2013) classified the dominant species classes pine, spruce, and deciduous species from ALS data and hyperspectral images with an average overall classification accuracy of 89.1%. Puliti et al. (2017) classified the same dominant species classes from ALS data and photogrammetric models with an overall accuracy of 79%. In this study, the overall accuracy for those three classes was 97% (i.e., excluding the mixed species class; Table 3). The same data as in this study have previously been used for classification of pine, spruce, birch, oak, and larch (i.e., splitting the deciduous class into several classes and excluding the mixed class; Persson et al., 2018) from four Sentinel-2 images with an overall classification accuracy of 88.2%. Comparisons between results at plot level and individual tree level are difficult, but some examples serve to illustrate the accuracy that can be achieved. Holmgren and Persson (2004) classified pine, spruce, and deciduous species trees for individual trees from structural features of ALS data with an overall accuracy of 95%. Yu et al. (2017) classified pine, spruce, and birch for individual trees from Optech Titan multispectral ALS data with an overall accuracy of 85.9% for the best combination of features. It has been suggested that texture metrics from high-resolution aerial images are more useful for discrimination of Scots pine and Norway spruce (Packalén and Maltamo 2006, Packalén and Maltamo 2007). The lowest accuracy in this study was achieved for the mixed species class where many field plots were classified as Norway spruce or deciduous species. The field plots with deciduous species also included a number of different tree species with slightly different spectral properties. The mix of deciduous tree species in that class is also likely an error source. Unfortunately none of the deciduous tree species was found in large enough abundance to be defined as a separate class.

4.4. Choice of metrics for classification

The classification was based only on mean intensity of the three available laser wavelengths, both for the mini raster cell method and the plot level analysis. The classification could potentially be improved by including more metrics derived from the intensity. Nevertheless, both methods in this study used the same kind of metrics. Regarding height and density metrics from ALS data that are commonly used for area-based estimation of forest variables, to the best of our knowledge no previous study has succeeded in using them for tree species classification since they are not distinctly different for different tree species when averaged over area units.

4.5. Choice of radius of local surrounding and raster cells size

The mini raster cell method provides classified tree species classes in small raster cells, 0.5 m. Each raster cell represents a circular surrounding with 5 m radius. The results were similar when using a 3 m radius. The reason for using small raster cells and surroundings rather than raster cells corresponding to the size of field plots is that the small raster cells are more likely to contain only one tree species class, which means that it is more likely that the intensity values from the ALS data are derived from only one tree species class. The size of a tree crown depends on the diameter and height of the tree, but since we selected field plots based on a minimum mean DBH, the variation was reduced, which makes it reasonable to use only one radius. The mini raster cells should be large enough to contain a few laser returns from each wavelength at the top of the canopy, but not larger since that would mean a lower resolution. With 30–40 returns per channel and m^2, each mini raster cell contained 7–10 returns per channel in total. Of these, 1–2 can be expected to be from the top 2 m of the canopy, considering that the 80th height percentile of the returns is around 2 m lower than the maximum height (e.g., Table 3 in Vastaranta et al. 2014).

4.6. Aggregation to area units of the mini raster cells

The small raster cells can be aggregated to arbitrary areas, for example forest stands or area units corresponding to field plots. In this study, the validation was done for raster cells aggregated within circular field plots. The validation was done by computing the proportion of area covered by raster cells classified as the different species classes in each field plot. This was based on the observation that the crown area of a tree is approximately proportional to its basal area, which is consistent with an approximately linear relationship between crown radius and DBH, at least for mean DBH $\geq$ 10 cm (Widlowski et al. 2003). The relationship differs slightly between tree species, but since it also depends on site conditions and stand history that are unknown, we assumed a simple linear relationship.

4.7. Choice of classification method

We choose LDA as the classification methods since it is a straightforward method. Quadratic Discriminant Analysis (QDA) is another classification method which differs from LDA in that LDA assumes that the variance is equal for all classes while QDA allows different variance for different classes. However, the dataset had a large proportion of Norway spruce-dominated field plots and much fewer observations for Scots pine and deciduous species. This would make it difficult to estimate the variance of each class separately. We still repeated the classification procedure with QDA instead of LDA and the results were similar. We chose to use a parametric method rather than a non-parametric method such as random forests since the number of observations in each class was smaller than what is usually used for non-parametric models.

4.8. Comparison to individual tree methods

Tree species classification using only ALS data has formerly been done using individual tree methods (e.g., Brandtberg et al., 2003, Holmgren and Persson 2004, Reitberger et al. 2008, Vauhkonen et al. 2009). Since the validation of the mini raster cell method was done at plot level, it is difficult to compare the accuracy to results published for individual tree methods. Delineation of individual trees requires more advanced algorithms than area-based methods. Additionally, adapting the parameters for the tree crown delineation to local forest conditions often requires information about the position of individual trees from the field data (e.g., Holmgren and Lindberg 2013), which can make the field inventories more complicated. The method presented here does not require tree positions. To collect training data in a cost-efficient way, one alternative would be to place field plots in homogenous forest stands with one dominant species. In that case the DBH of the trees would not be needed, only the position of the field plot center.

4.9. Practical applicability of multispectral ALS data

The ALS data in this study were collected from a flying altitude of 400 m, which is too low for operational use due to the comparatively high cost per area unit. Previous studies on tree species classification from Optech Titan data have used flying altitudes of 360 m and 800 m (Budei et al. 2017) and 400 m (Yu et al. 2017, Axelsson et al. 2018). One reason for using comparatively low flying altitudes is to get sufficient signal strength for all wavelengths. From our previous experience with data from Optech Titan (not published), a flying altitude of 1000 m resulted in a weak signal for the 532 nm channel, which had the consequence that in many cases no returns were registered from the tree canopy for the 532 nm channel although returns were registered for the other two channels. Additionally, the beam divergence of the 532 nm channel is approximately twice as large as for the two other channels (Optech 2015). For the data acquired from 1000 m altitude, the larger divergence caused averaging in the canopy for the 532 nm channel, providing less details than for the two other channels. A new dual-wave ALS system is currently available, RIEGL VQ-1560i-DW with wavelengths 1064 nm and 532 nm (RIEGL 2019). According to specifications, the maximum measuring range is equal for the two channels and the beam divergence of the 532 nm wavelength is smaller than for Optech Titan. The new system could potentially be operated at a higher flying altitude and still provide multispectral ALS data for tree species classification. If that is possible, operational use of the method presented here could be feasible. This could make it possible to derive maps of tree species composition automatically for large areas.

5. Conclusions

This study has presented a mini raster cell method for classification of tree species classes from multispectral ALS data. The method is an area-based approach where statistics are derived from ALS data in raster cells, but the raster cells are much smaller than the field sample plots commonly used for area-based methods. Hence, the resulting tree species map has a higher resolution.

The classification results from the mini raster cell method were more accurate than the results from a plot level area-based method. The most likely explanation is that the high resolution of the mini raster cell method enables classification of homogenous patches of tree species classes within a field plot. This is particularly useful in forest stands with mixed species.

The tree species map can be aggregated for arbitrary areas such as field sample plots or forest stands to estimate the proportion of the different tree species classes from the proportion of raster cells in different classes. This is another advantage compared to conventional area-based methods.

The mini raster method is a simple rasterization algorithm and requires less advanced algorithms than individual tree based-methods, another approach for forest estimations from ALS data. Individual tree delineation algorithms must be adapted or trained to local forest conditions while the mini raster method has few parameters: the size of the raster cells, the radius of the moving window, and the height limits above the ground and below the canopy for including laser returns. This means that the mini raster cell method could be more feasible for operational mapping of tree species.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Formas, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning through the project “Tree species classification from remote sensing - next generation forest maps for ecology and management” with the Formas DNR number 2018-01161, by the Swedish Forest Society Foundation through the project “New forest maps with tree species information from time series of satellite data” with the Id-number 2019-660-Steg 2 2018, and by the Hildur and Sven Wingquist Foundation for Forest Science Research through the projects "Tree species information from airborne and ground-based sensors" with the Id-number 15/16-4/101-6 and "New techniques for forest management planning" with the Id-number 16/17-5/104-5.

Appendix

Table A1
Confusion matrix from classification based on raster cells corresponding to the field plot (i.e., mini raster cell method) with 3 m radius.

		Classified				Producer's accuracy
		Pine	Spruce	Deciduous	Mixed	
Field	Pine	16	2	0	4	73%
	Spruce	0	86	0	3	97%
	Deciduous	0	3	14	6	61%
	Mixed	3	27	8	14	27%
User's accuracy		84%	73%	64%	52%	70%

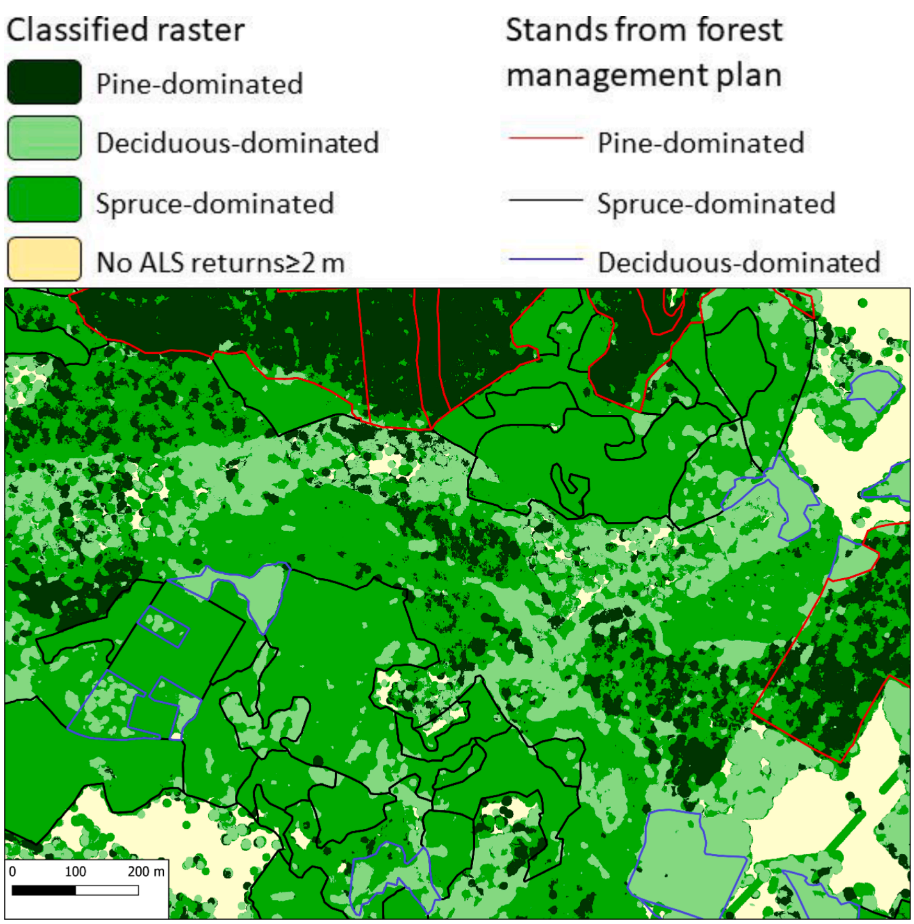

Fig. A1. Examples of the raster with 3 m radius in the same area as Fig. 6. The lines are borders of forest stands with mean age at least 20 years and one dominant tree species (i.e., at least 80% of the stem volume) from the forest management plan.

Table A2
Confusion matrix from classification based on raster cells corresponding to the field plot (i.e., mini raster cell method) with 5 m radius and QDA classification.

		Classified				Producer's accuracy
		Pine	Spruce	Deciduous	Mixed	
Field	Pine	18	2	0	2	82%
	Spruce	0	86	0	3	97%
	Deciduous	0	1	18	4	78%
	Mixed	5	19	13	15	29%
User's accuracy		78%	80%	58%	62%	74%

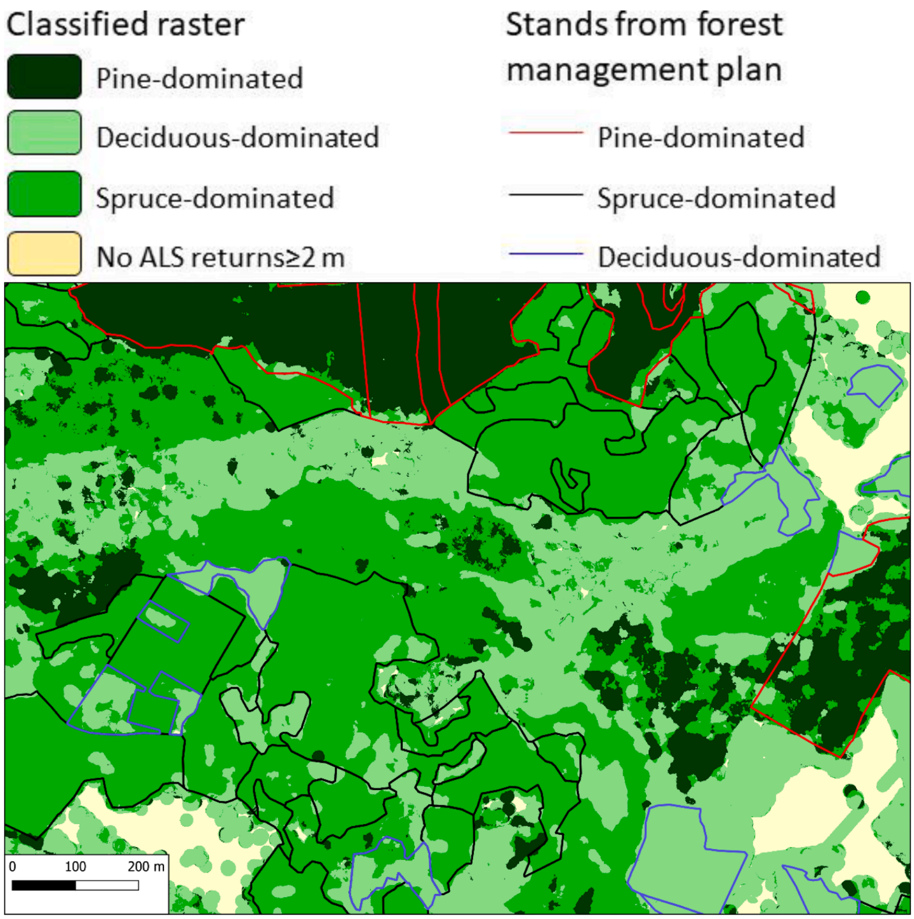

Fig. A2. Examples of the raster with 5 m radius and QDA classification in the same area as Fig. 6. The lines are borders of forest stands with mean age at least 20 years and one dominant tree species (i.e., at least 80% of the stem volume) from the forest management plan.

References

The Heureka project. Retrieved September 28, 2012, from http://www.slu.se/heureka.
Ahokas, E., Hyyppä, J., Yu, X., Liang, X., Matikainen, L., Karila, K., Litkey, P., Kukko, A., Jaakkola, A., Kaartinen, H., Holopainen, M., Vastaranta, M., 2016. Towards Automatic Single-Sensor Mapping by Multispectral Airborne Laser Scanning. Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, International Archives of the Photogrammetry.
Axelsson, A., Lindberg, E., Olsson, H., 2018. Exploring Multispectral ALS Data for Tree Species Classification. Remote Sens. 10 (2), 183.
Bakuła, K., Kupidura, P., Jełowicki, Ł., 2016. Testing of Land Cover Classifacation from Multispectral Airborne Laser Scanning Data. Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, International Archives of the Photogrammetry.
Brandtberg, T., 2002. Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. Fuzzy Sets Syst. 132 (3), 371–387.
Brandtberg, T., 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 Sens. Environ. 85 (3), 290–303.
Budei, B.C., St-Onge, B., Hopkinson, C., Audet, F.-A., 2017. Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. Remote Sens. Environ.
Dalponte, M., Ørka, H.O., Ene, L.T., Gobakken, T., Næsset, E., 2014. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sens. Environ. 140, 306–317.
Dietterich, T.G., 1998. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Comput. 10 (7), 1895–1923.
Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87.
Fassnacht, F.E., Mangold, D., Schaefer, J., Immitzer, M., Kattenborn, T., Koch, B., Latifi, H., 2017. Estimating stand density, biomass and tree species from very high resolution stereo-imagery - towards an all-in-one sensor for forestry applications? Forestry 90 (5), 613–631.
Franklin, S.E., Hall, R.J., Moskal, L.M., Maudie, A.J., Lavigne, M.B., 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. Int. J. Remote Sens. 21 (1), 61–79.
Heinzel, J., Koch, B., 2012. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. Int. J. Appl. Earth Obs. Geoinf. 18, 101–110.
Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling.
Holmgren, J., Lindberg, E., 2013. Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. Can. J. Remote Sens. 39, S86–S98.
Holmgren, J., Persson, Å., 2004. Identifying species of individual trees using airborne laser scanner. Remote Sens. Environ. 90 (4), 415–423.
Hopkinson, C., Chasmer, L., Gynan, C., Mahoney, C., Sitar, M., 2016. Multisensor and Multispectral LiDAR Characterization and Classification of a Forest Environment. Can. J. Remote Sens. 42 (5), 501–520.
Hyvönen, P., Pekkarinen, A., Tuominen, S., 2005. Segment-level stand inventory for forest management. Scand. J. For. Res. 20 (1), 75–84.
Hyyppä, J., Yu, X., Hyyppä, H., Maltamo, M., 2006. Methods of airborne laser scanning for forest information extraction. Workshop on 3D Remote Sensing in Forestry. University of Natural Resources and Applied Life Science, Vienna, Austria, Vienna.
Korpela, I., 2004. Individual tree measurements by means of digital aerial photogrammetry.
Korpela, I., Heikkinen, V., Honkavaara, E., Rohrbach, F., Tokola, T., 2011. Variation and directional anisotropy of reflectance at the crown scale - Implications for tree species classification in digital aerial images. Remote Sens. Environ. 115 (8), 2062–2074.
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.
Kukkonen, M., Maltamo, M., Korhonen, L., Packalen, P., 2019. Multispectral Airborne LiDAR Data in the Prediction of Boreal Tree Species Composition. IEEE Trans. Geosci. Remote Sens. 57 (6).
Li, X., Strahler, A.H., 1985. Geometric-Optical Modeling of a Conifer Forest Canopy. IEEE Trans. Geosci. Remote Sens. GE-23 (5), 705–721.
Lillesand, T., Kiefer, R.W., Chipman, J., 2007. Remote Sensing and Image Interpretation. John Wiley & Sons Inc., New York.
Lindberg, E., Briese, C., Doneus, M., Hollaus, M., Schroiff, A., Pfeifer, N., 2015. Multi-wavelength Airborne Laser Scanning for Characterization of Tree Species. SilviLaser 2015, La Grande Motte, France, IGN, LIF, Saint-Mandé Cedex, France.

Lindberg, E., Holmgren, J., 2017. Individual Tree Crown Methods for 3D Data from Remote Sensing. Curr. For. Reports 3, 19–31.

Lämås, T., 2010. The Haglöf PosTex ultrasound instrument for the positioning of objects on forest sample plots. Umeå, Sweden, Swedish University of Agricultural Sciences, Department of Forest Resource Management: 10.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., Kangas, J., 2006. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. Can. J. For. Res.-Revue Canadienne De Recherche Forestiere 36 (2), 426–436.

Maltamo, M., Naesset, E., Vauhkonen, J., 2014. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, Springer, Po Box 17, 3300 Aa Dordrecht, Netherlands.

Matikainen, L., Karila, K., Hyyppä, J., Litkey, P., Puttonen, E., Ahokas, E., 2017. Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating. ISPRS J. Photogramm. Remote Sens. 128, 298–313.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80 (1), 88–99.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., Olsson, H., 2015. A nationwide forest attribute map of Sweden derived using airborne laser scanning data and field data from the national forest inventory. SilviLaser 2015, La Grande Motte, France.

Olofsson, K., Wallerman, J., Holmgren, J., Olsson, H., 2006. Tree species discrimination using Z/I DMC imagery and template matching of single trees. Scand. J. For. Res. 21 (S7), 106–110.

Optech, T., 2015. Optech Titan Multispectral LiDAR System. Teledyne Optech, from https://www.teledyneoptech.com/en/products/airborne-survey/titan/.

Packalén, P., Maltamo, M., 2006. Predicting the Plot Volume by Tree Species Using Airborne Laser Scanning and Aerial Photographs. For. Sci. 52 (6), 611–622.

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 Sens. Environ. 109 (3), 328–341.

Packalén, P., Maltamo, M., 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Can. J. For. Res.-Revue Canadienne De Recherche Forestiere 38 (7), 1750–1760.

Packalen, P., Suvanto, A., Maltamo, M., 2009. A Two Stage Method to Estimate Species-specific Growing Stock. Photogramm. Eng. Remote Sens. 75 (12), 1451–1460.

Pellikka, P., King, D.J., Leblanc, S.G., 2000. Quantification and reduction of bidirectional effects in aerial cir imagery of deciduous forest using two reference land surface types. Remote Sens. Rev. 19 (1–4), 259–291.

Persson, M., Lindberg, E., Reese, H., 2018. Tree Species Classification with Multi-Temporal Sentinel-2 Data. Remote Sens. 10 (11), 1794.

Persson, Å., Holmgren, J., Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogramm. Eng. Remote Sens. 68 (9), 925–932.

Peterson, U., 1993. Successional reflectance trajectories in northern temperate forests. Int. J. Remote Sens. 14 (3), 609–613.

Puliti, S., Gobakken, T., Orka, H.O., Naesset, E., 2017. Assessing 3D point clouds from aerial photographs for species-specific forest inventories. Scand. J. For. Res. 32 (1), 68–79.

Reitberger, J., Krzystek, P., Stilla, U., 2008. Analysis of full waveform LiDAR data for the classification of deciduous and coniferous trees. Int. J. Remote Sens. 29 (5), 1407–1431.

RIEGL. 2019. RIEGL VQ-1560i-DW. RIEGL Laser Measurement Systems GmbH, from http://www.riegl.com/uploads/tx_pxpriegldownloads/RIEGL_VQ-1560i-DW_Datasheet_2017-11-27_Preliminary.pdf.

Teo, T.-A., Wu, H.-M., 2017. Analysis of Land Cover Classification Using Multi-Wavelength LiDAR System. Appl. Sci. 7 (7), 663.

Trimble, 2013. Trimble Geoexplorer 6000 Serie. Sunnyvale, California, US.

Wagner, W., 2010. Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts. ISPRS J. Photogramm. Remote Sens. 65 (6), 505–513.

Wagner, W., Hollaus, M., Briese, C., Ducic, V., 2008. 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. Int. J. Remote Sens. 29 (5), 1433–1452.

Vastaranta, M., Saarinen, N., Kankare, V., Holopainen, M., Kaartinen, H., Hyyppa, J., Hyyppa, H., 2014. Multisource Single-Tree Inventory in the Prediction of Tree Quality Variables and Logging Recoveries. Remote Sens. 6 (4), 3475–3491.

Vauhkonen, J., Hakala, T., Suomalainen, J., Kaasalainen, S., Nevalainen, O., Vastaranta, M., Holopainen, M., Hyyppa, J., 2013. Classification of Spruce and Pine Trees Using Active Hyperspectral LiDAR. IEEE Geosci. Remote Sens. Lett. 10 (5), 1138–1141.

Vauhkonen, J., Maltamo, M., Mcroberts, R.E., Næsset, E., 2014. Introduction to Forestry Applications of Airborne Laser Scanning. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. M. Maltamo, E. Næsset and J. Vauhkonen. Dordrecht, Springer: 1–16.

Vauhkonen, J., Tokola, T., Packalén, P., Maltamo, M., 2009. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. For. Sci. 55 (1), 37–47.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. Can. J. Remote Sens. 42 (5), 619–641.

Widlowski, J.-L., Verstraete, M., Pinty, B., Gobron, N., 2003. Allometric relationships of selected European tree species. Parametrizations of tree architecture for the purpose of 3-D canopy reflectance models used in the interpretation of remote sensing data: Betula pubescens, Fagus sylvatica, Larix decidua, Picea abies, Pinus sylvestris. I. f. E. a. Sustainability. Ispra, Italy, Office for Official Publications of the European Communities: 61.

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Orka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. LiDAR sampling for large-area forest characterization: a review. Remote Sens. Environ. 121, 196–209.

Yu, X., Hyyppä, J., Litkey, P., Kaartinen, H., Vastaranta, M., Holopainen, M., 2017. Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning. Remote Sens. 9 (2), 108.

Ørka, H.O., Dalponte, M., Gobakken, T., Næsset, E., Ene, L.T., 2013. Characterizing forest species composition using multiple remote sensing data sources and inventory approaches. Scand. J. For. Res. 28 (7), 677–688.