RESEARCH　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

On the potential to predetermine dominant tree species based on sparse-density airborne laser scanning data for improving subsequent predictions of species-specific timber volumes

Janne Räty, Jari Vauhkonen[*], Matti Maltamo and Timo Tokola

Abstract

Background: Tree species recognition is the main bottleneck in remote sensing based inventories aiming to produce an input for species-specific growth and yield models. We hypothesized that a stratification of the target data according to the dominant species could improve the subsequent predictions of species-specific attributes in particular in study areas strongly dominated by certain species.

Methods: We tested this hypothesis and an operational potential to improve the predictions of timber volumes, stratified to Scots pine, Norway spruce and deciduous trees, in a conifer forest dominated by the pine species. We derived predictor features from airborne laser scanning (ALS) data and used Most Similar Neighbor (MSN) and Seemingly Unrelated Regression (SUR) as examples of non-parametric and parametric prediction methods, respectively.

Results: The relationships between the ALS features and the volumes of the aforementioned species were considerably different depending on the dominant species. Incorporating the observed dominant species in the predictions improved the root mean squared errors by 13.3–16.4 % and 12.6–28.9 % based on MSN and SUR, respectively, depending on the species. Predicting the dominant species based on a linear discriminant analysis had an overall accuracy of only 76 % at best, which degraded the accuracies of the predicted volumes. Consequently, the predictions that did not consider the dominant species were more accurate than those refined with the predicted species. The MSN method gave slightly better results than models fitted with SUR.

Conclusions: According to our results, incorporating information on the dominant species has a clear potential to improve the subsequent predictions of species-specific forest attributes. Determining the dominant species based solely on ALS data is deemed challenging, but important in particular in areas where the species composition is otherwise seemingly homogeneous except being dominated by certain species.

Keywords: Forest inventory, Light Detection and Ranging (LiDAR), Area-based approach, Nearest neighbor estimation, Crown base height, Intensity, Volume model

* Correspondence: jari.vauhkonen@uef.fi
University of Eastern Finland, School of Forest Sciences, Yliopistokatu 7, P.O. Box 111 FI-80101 Joensuu, Finland

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 2 of 17

Background

Forest ecosystem modelling requires inventory estimates, which are traditionally acquired using stand-level (compartmentwise) forest inventories based on field assessments or visual interpretation of aerial images (e.g. Eid et al. 2004; Koivuniemi and Korhonen 2006; Ståhl et al. 2011). Due to species-specific growth and yield modeling, the inventories are required to provide species-specific predictions (e.g. Maltamo et al. 2011). The conventional inventories to provide stand-level estimates are currently being replaced in Scandinavia, in particular, by discrete-return Light Detection and Ranging (LiDAR) data recorded by small-footprint airborne laser scanning (ALS; for an overview, see Maltamo et al. 2014) incorporated with spectral data from aerial (Packalén and Maltamo 2006, 2007, 2008) or satellite images (Wallerman and Holmgren 2007) for species recognition. Extracting species information has also been tested in North America (Hudak et al. 2008; van Ewijk et al. 2014) and Central Europe (Latifi et al. 2010; Heinzel and Koch 2012; Torabzadeh et al. 2014), and a detailed review on the topic can be found from Vauhkonen et al. (2014a).

High species recognition accuracy is crucial for forest management planning systems that involve different treatment schedules depending on species and also important towards accurate growth and yield estimates. According to the simulations Korpela and Tokola (2006) carried out in forest conditions closely corresponding to our study area, predictions of the total stand volume based on tree-level, species-specific allometric dependencies had Root Mean Squared Errors (RMSEs) of 30 % and about 15 %, when the species of the individual trees were recognized at accuracies of 75 % and 80–90 %, respectively, and the other measurements were error-free. A similar result is reported by Tompalski et al. (2014) in Canada, who nevertheless found predictions based on species-specific equations more accurate than generic ones.

Using ALS data, high species recognition rates are generally based on detecting individual trees (e.g., Holmgren and Persson 2004; Kim et al. 2009; Ørka et al. 2009; Suratno et al. 2009), which requires acquiring data in a higher density than what is currently feasible from an operational viewpoint (e.g., Maltamo and Packalen 2014; Næsset 2014). However, several studies have reported successful predictions of the total (Woods et al. 2011; Nord-Larsen and Schumacher 2012; Villikka et al. 2012) and even species-specific forest attributes (Vauhkonen et al. 2012; Ørka et al. 2013) based on ALS data with pulse densities < 1 m^{-2} and other scanning parameters not permitting individual tree detection.

The ALS inventories employing the sparse-density data are most often implemented using so-called area-based approaches (Næsset 2002), in which (i) models to predict the forest attributes of interest for the individual areas-of-interest (AOIs) are fit based on a set of training field plots; and (ii) the resulting models are applied to all the AOIs of the entire inventory area to produce wall-to-wall predictions. Operational implementations are elaborated by White et al. (2013), Maltamo and Packalen (2014), and Næsset (2014). In particular the modeling of a multivariate response such as the species-specific attributes is generally built upon non-parametric nearest neighbor (NN) approaches, in which the predictions of the considered forest attributes are simultaneously obtained as (weighted) averages of the k most similar reference observations in terms of the considered distance metric applied in the predictor space.

NN predictions require a considerably large database of the reference observations (see Maltamo et al. 2009a), although some studies have indicated that accurate species-specific forest attribute estimates may be provided with a limited number of plots (Kotamaa et al. 2010; Villikka et al. 2012; Pippuri et al. 2013). Further, an adequately representative reference data with respect to the species and size distribution of the area may be difficult to obtain using systematic sampling designs (Maltamo et al. 2009b). The predictions could be improved by a complementary inventory according to the deficiencies of the initial estimation, as demonstrated by Vauhkonen et al. (2012) complementing the data of Maltamo et al. (2009b).

Due to the practical difficulties to obtain adequately extensive and representative field reference data for the NN predictions, parametric models such as those constructed by Seemingly Unrelated Regression (SUR) approaches could be seen as alternative methods (e.g., Lindberg et al. 2010; see also Maltamo et al. 2009c, 2012). Even if fitted with similarly limited data, the ability to linearly interpolate in between the observations could be a practical benefit compared to the NN predictions, which are, to some degree, always based on the discrete data points. Beside ALS studies, the SUR and other methods for fitting regression models based on systems of equations are presented by Siipilehto et al. (2007).

From a practical point of view, it is well-reasoned to seek alternative implementations for ALS inventories relying on the availability of both the ALS and image data. Even though aerial images are usually available for the purpose of visual forest stand delineation, using them as additional data complicates the inventory system due to the required co-registrations and calibrations of the radiometric differences of multiple images. Plot-level species-specific predictions based solely on ALS data have also been tested (Ørka et al. 2013; Vauhkonen et al. 2012, 2014b). The predictions related to the dominant species in particular have been accurate based on

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 3 of 17

ALS data (Ørka et al. 2013), but the availability of the spectral data has generally improved the predictions (Vauhkonen et al. 2012; Ørka et al. 2013).

Even if the main tree species were estimated correctly, large errors may be related to the predictions of other forest attributes, especially those of the non-dominant tree species (e.g., Maltamo et al. 2009b). For example, Packalén et al. (2009) proposed excluding species representing <10 % of the total volume from the accuracy measures due to the insignificance of such species in the compartmentwise inventory. Yet, even such "near to zero" predictions may distort species proportions and cause further problems in inventory areas with an unbalanced species distribution such as strongly pine-dominated areas typical to the boreal region (e.g., Maltamo et al. 2009b; Vauhkonen et al. 2012, 2014b). However, whether known beforehand that a subject stand was dominated by certain species with a proportion of, say, >75 % or >95 %, the maximum error level expected for the predictions of the minor species could be confined. Based on this reasoning and the encouraging results of successfully predicting the dominant species based on ALS data alone (Ørka et al. 2013; Vauhkonen et al. 2014b) and improving the results of NN methods by pre-classifying the inventory area (Maltamo et al. 2015), a test of using dominant species information for the species-specific predictions was motivated.

The purpose of the study is thus to predict dominant species and species-specific timber volumes in a strongly pine-dominated test area. Predictions of the dominant species based on ALS features are evaluated. Prediction models based on NN and SUR are formulated and compared with respect to accounting for the a priori information on the dominant species.

Methods
Study area and field data
The data studied were originally collected for crown base height assessments (Korhonen 2012). Two test areas within a geographical distance of 30 km were established in Kuhmo, northeastern Finland. The area is very homogenous and strongly dominated by Scots pine (*Pinus sylvestris* L.) trees. The other species to be distinguished are Norway spruce (*Picea abies* [L.] H. Karst.) and a group of deciduous trees consisting of mainly birches (*Betula* spp. L.) and aspen (*Populus tremula* L.), which form minor proportions and typically occur below the dominant canopy. Altogether 265 field sample plots with co-located ALS and field data were studied.

Circular sample plots with radii of 9 m were used in the field data collection. Every tree with a diameter at breast height (D_{bh}) >5 cm was measured for the D_{bh} and crown base height (CBH). Trees with a D_{bh} corresponding

to the basal area-weighted median tree of each species occurring on a plot were determined in the field and measured for tree height. The D_{bh} and height of these trees were used as the median tree diameter and height (D_{gM} and H_{gM}, respectively) of the corresponding species per plot, and the maxima of the values were used as the D_{gM} and H_{gM} of the entire plot. Plot basal area (G) was calculated by summing from the tree-level basal areas, determined as $\frac{\pi}{4}D_{bh}^2$. The missing tree heights were predicted by calibrating the prediction models for the parameters of Näslund's (1936) height curve presented by Siipilehto (1999) using the species-specific D_{gM} and H_{gM} estimates. The volumes of the individual trees were predicted by models of Laasasenaho (1982), employing the D_{bh}, height, and tree species as predictors. The models for birch were used for all deciduous trees. Central characteristics of the field measurements aggregated for the field plots are shown in Table 1.

ALS data and the extracted features
ALS data
The ALS data were acquired on September 4–7, 2011, under a leaf-on period of the deciduous vegetation. Leica ALS50-II scanner was operated from an altitude of 2000 m using a field-of-view of 30°, a scanning rate of 52 Hz, and a pulse frequency of 58.9 Hz. These scanning parameters resulted in a nominal measurement density of 0.52 observations m^{-2}. The analyses were focused only on the first echoes (i.e., "only" and "first of many" echoes per pulse), aiming to obtain the main information from the data, while retaining most generalization abilities over sensors that record a different number of echo categories (e.g., Næsset 2014). The ALS data were acquired, preprocessed and co-located with the field data as a part of an operational data acquisition campaign by Arbonaut, Ltd., and the accuracies and error sources of this process are expected to correspond with those reported in the literature (see, e.g., Maltamo and Packalen 2014; Næsset 2014).

The predictor features extracted for the study (Table 2) were selected based on the earlier studies (e.g., Vauhkonen et al. 2014b). However, since a prediction of the CBH of a tree has been found a useful indicator of its species based on tree-level studies (Holmgren and Persson 2004; Holmgren et al. 2008) and we had rich field data for examining the accuracies of these predictions, we focused a particular attention on examining whether the field measurement or an area-based estimate of the CBH could distinguish plots dominated by various species. The CBH was predicted by extracting connected alpha shape components from the lowest parts of the point cloud according to the method of Maltamo et al. (2010), which is a

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 4 of 17

Table 1 Species-specific volume characteristics of the 265 sample plots. Min: minimum, Max: maximum, Sd: standard deviation

Attribute	Population	Mean	Min	Max	Sd
Volume (m^3 ha^{-1})	Total	131.5	6.3	434.9	85.3
	Scots pine	87.2	0.0	295.6	66.4
	Norway spruce	28.6	0.0	401.6	53.7
	Deciduous trees	15.8	0.0	178.1	24.3
Basal area (m^2 ha^{-1})	Total	17.5	0.1	45.8	9.1
	Scots pine	11.1	0.0	38.1	7.6
	Norway spruce	4.0	0.0	39.5	6.6
	Deciduous trees	2.4	0.0	18.7	3.2
Basal-area weighted mean diameter (cm)	Scots pine	18.4	0.0	42.3	9.3
	Norway spruce	9.8	0.0	38.0	8.6
	Deciduous trees	4.0	0.0	51.4	6.9
Basal-area weighted mean height (m)	Scots pine	13.6	0.0	28.3	6.3
	Norway spruce	7.7	0.0	26.3	6.6
	Deciduous trees	3.7	0.0	28.7	5.9

variant of a tree-level method described by Vauhkonen (2010).

The other ALS features considered were the mean and standard deviation of the intensity values and the proportion of the different echoes (Vauhkonen et al. 2014b). Following Ørka et al. (2012) and Vauhkonen et al. (2014b), the intensity features were calculated separately based on all, only, or first-of-many echoes. The most common ALS-based predictor variables (Magnussen and Boudewyn 1998; Næsset 2002), i.e., the maximum, the mean and standard deviation of the height values; proportion of echoes above 2 m vegetation threshold; the 5th, 10th, 20th, …, 90th, and 95th percentiles and the corresponding proportional densities of the ALS-based canopy height distribution were calculated according to Korhonen et al. (2008, p. 502–503). The ALS features are listed in Table 2.

Predicting the dominant species using ALS

The species proportions were determined as the percentages of each species from the total plot basal area (*G*). The dominant species were subsequently determined based on these proportions. Several alternatives to determine the exact percentage values for the dominant species were tested (Table 2) to analyze whether predicting this information was feasible using ALS. First, the species with the highest percentage were set as the dominant species of the plot, yielding three dominant species classes (pine, spruce, and birch dominated). Second, the dominant species were determined using a threshold of 75 %: whether a species had a proportion higher or equal to this level, it was set as the dominant species of the plot. Whether no species reached this threshold, the plot was labeled as 'mixed', i.e. this classification yielded the dominant species classes of pine, spruce, birch, and mixed. Finally, since the area was known to be strongly dominated by the pine species, it was tested whether information on pure pine plots improved the results. Such plots were selected using a threshold of 95 % and tested along the aforementioned alternatives. The definition alternatives for the dominant species are listed in Table 3.

Table 2 The response and predictor variables and the principles of relating these. The modelling principles (MSN: Most Similar Neighbor, SUR: Seemingly Unrelated Regression) are detailed in Section 2.4

Response variables	Predictor variables[1]	Modelling principle	
		MSN	SUR
- Total plot volume - Species-specific volumes	- ALS-based CBH estimate - Maximum, mean, standard deviation and proportion - Percentiles 5, 10, 20, …, 90, 95 - Densities 5, 10, 20, …, 90, 95 - Mean and standard deviation of intensity values[2]	NN search based on canonical correlation analysis between all response and predictor variables. The dominant species are included as restrictions to the NN search.	System of linear regression equations based on 1–2 ALS features and a categorical predictor indicating the dominant species on the plot.

[1]Computed using height values > 2 m ground threshold
[2] Computed separately based on only, first-of-many and both the return types

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 5 of 17

Table 3 The different definitions used for the dominant tree species in this study

Abbreviation	Definition for the dominant species	Classes[1]
Sp_{max}	Highest species-specific proportion of G per plot.	P, S, D
Sp_{max+95}	Highest species-specific proportion of G per plot + separately labeled plots with $G \geq 95$ % of pine.	P_{95}, P, S, D
Sp_{75}	Species-specific proportion of $G \geq 75$ %; plots with a lower dominant proportion pooled in a separate class.	P_{75}, S_{75}, D_{75}, M
Sp_{75+95}	Species-specific proportion of $G \geq 75$ %; plots with a lower dominant proportion pooled in a separate class + separately labeled plots with $G \geq 95$ % of pine.	P_{95}, P_{75}, S_{75}, D_{75}, M

[1] Dominated by pine (P), spruce (S), deciduous trees (D), or the aforementioned species with the proportion given in the subscript; or mixed (M)

The scatter plots of the ALS-based predictors were first assessed with respect to their abilities to discriminate between species and invariance with respect to tree size, quantified in terms of the D_{gM} and H_{gM} characteristics. A linear discriminant analysis (LDA) implemented in the MASS package (Venables and Ripley 2002) of R (R Core Team 2013) was used to classify the data by tree species. The principle of LDA is to form linear combinations which maximize the ratio of the between-class to within-class variance based on the data of the original feature vectors (see, e.g. Venables and Ripley 2002). LDA was run with a leave-one-out cross validation, in which the priors were adjusted to give an equal probability for each species. The predictors used were selected manually according to the graphical assessments. First, the discriminant functions were fitted with one predictor variable at the time. The variables resulting to best accuracies were added with a second variable and the accuracies of these combinations were further ordered. The procedure was repeated until the number of predictors was 4, which was considered as an adequate upper limit given the number of classes considered.

Modelling the species-specific volumes
Prior to the modeling, the predictors based on the ALS data were evaluated with respect to their relationships with the species-specific volumes in a similar way than described in the previous section. Two modeling strategies, namely a non-parametric nearest neighbor and a parametric regression based approach, were tested for obtaining the prediction models. The methods are described in the sub-sections below and their main differences are presented by Table 2.

k-Most Similar Neighbor (k-MSN)
In the NN approach, the predictions of the forest attributes were based on an average of k-NN observations in terms of the ALS features. The NNs were determined according to the Most Similar Neighbor (MSN) distance metric (Moeur and Stage 1995), in which a canonical correlation analysis is used to produce a weighting matrix for selecting the NNs from the training data. The total and species-specific volumes and all the ALS features were employed in the correlation analysis.

The dominant species information (Table 3) was taken into account in the prediction step. Instead of using the k-NNs solely based on the predictor feature space, those NNs which were of a different dominant species than the target plot were not considered in the predictions. In practice, up to 1–10 NNs meeting the dominant species condition were selected from an initial neighborhood consisting of all the reference plots. The total and species-specific volumes were predicted simultaneously as arithmetic averages of the restricted k-NNs. The MSN imputation was implemented using the *yaImpute* package (Crookston and Finley 2007) of R (R Core Team 2013).

Seemingly Unrelated Regression (SUR)
Alternatively, the species-specific volumes were predicted as a simultaneously fitted system of equations based on the Seemingly Unrelated Regression (SUR) modeling implemented using the *systemfit* package (Henningsen and Hamann 2007) of R (R Core Team 2013). The main idea of SUR (Zellner 1962) is to account for the interactions between residual structures of different linear regression equations such that every regression model will be affected (Henningsen and Hamann 2007). The coefficients of the SUR model were based on generalized least squares (GLS) estimation. A presumption for the GLS method is that the matrices which are constructed from the regression models should be correlated but unequal (Henningsen and Hamann 2007).

In the SUR modelling, the dominant tree species (Table 3) were accounted for by introducing a categorical predictor variable with levels corresponding to the tree species considered. ALS features were added as further predictors of the model based on the coefficient of determination (R^2) values. Individual predictors were added attempting to maximize the R^2. However, a new predictor was included only if it affected the model significantly according to the p-value of a Student's t-test.

Accuracy assessment
The accuracies of the predictions were assessed separately at the model fitting and prediction stage. In the latter, the dominant species predicted according to the

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 6 of 17

Section (Predicting the dominant species using ALS) were used to replace those observed in the field and used for model fitting (Section Modelling the species-specific volumes).

The accuracy of the species classification was assessed by means of the overall accuracy and kappa (κ) scores. The overall accuracy gives the number of correctly classified cases as a proportion of all observations. The κ coefficient (Eq. 1) can be interpreted as a proportion of chance-expected disagreements which do not occur (Cohen 1960). In our case it presents how much better an LDA-classification is compared to the material which is classified by chance. The κ coefficient was obtained as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{1}$$

where p_o is proportion of correctly classified observations and p_e is probability of correct classification by chance.

The accuracy of the species-specific volume predictions was assessed by means of the root mean squared error (RMSE, Eq. 2) and mean difference (BIAS, Eq. 3) between the observed and estimated values,

$$RMSE = \sqrt{\frac{\sum (p-r)^2}{n}} \text{ and} \tag{2}$$

$$BIAS = \frac{\sum (p-r)}{n}, \tag{3}$$

where p is the observed value based on field measurements, r is the predicted value, and n is the number sample plots. The relative RMSE and bias were calculated by dividing the absolute RMSE and bias values by the mean value of the reference attribute.

Results

In this section, we first present the results of explanatory analyses on the relationships between the ALS features and species-specific attributes (the Section of relationships between ALS features and species-specific attributes) and the development of the SUR models based on these analyses together with the performance of the SUR and k-MSN predictions using the field-observed dominant species (the Section of models for species-specific volumes). The results of predicting the dominant species and the prediction accuracies when combining this information with the models developed with the field data (the Section of models for species-specific volumes) are presented in the Sections of Classification of the dominant species and Prediction accuracies, respectively.

Relationships between ALS features and species-specific attributes

The CBH predicted by ALS had RMSEs of 1.58 and 1.47 m and biases of −0.93 and 0.07 m, when evaluated against the arithmetic and basal-area weighted means of the field measurements, respectively. These accuracies suggest that the area-based prediction of the CBH is a reliable estimate of this measure particularly with respect to the largest trees. The results are on the same accuracy level as in the earlier studies (see Maltamo et al. 2012).

The CBH was however not an appropriate indicator of the tree species proportion (Fig. 1). Instead, other ALS features produced a better discrimination between the dominant species considered. For example, the features based on the proportions and intensities of the different echoes (Fig. 1) indicated a difference in the leveling between pine and spruce dominated plots. This difference was also invariant to the size according to the D_{gM} measure. For deciduous dominated plots it was difficult to find ALS features which could separate them from the other species groups.

The height and density metrics had a quasi-linear relationship between the total and main species volumes, as illustrated in Fig. 2 using a product of a height percentile and the ratio of echoes reflected above ground to all echoes, i.e. the canopy cover. However, the volumes of the other-than-dominant species were not favorably related to these metrics (Fig. 2). Because the proportions of the minor species and the related ALS metrics considerably vary depending on the main species (Fig. 2), incorporating this information in the species-specific prediction appears strongly justified.

Models for species-specific volumes

To analyze the goodness-of-fit of the species-specific volume models, the predictor variables were inserted systematically based on the earlier analyses (e.g., Fig. 2). Although the final composition of the predictor variables slightly varied depending on the species, the ratio of the echoes reflected above ground to all echoes combined with a height percentile were the most frequent predictors included in the models. This is reasonable, since their product (density × height) forms an approximation of the growing stock volume. However, for sample plots dominated by the deciduous trees, other variables performed better as predictors.

The SUR models were typically composed of two ALS features and a group of dummy variables for the dominant species (Tables 4 and 5). In the latter groups, one of the coefficients using the Sp_{max+95} (Table 4) or 1–3 coefficients per model using the Sp_{75+95} strategy (Table 5) were not significant according to the t-tests for the model coefficients. The models presented in Tables 4

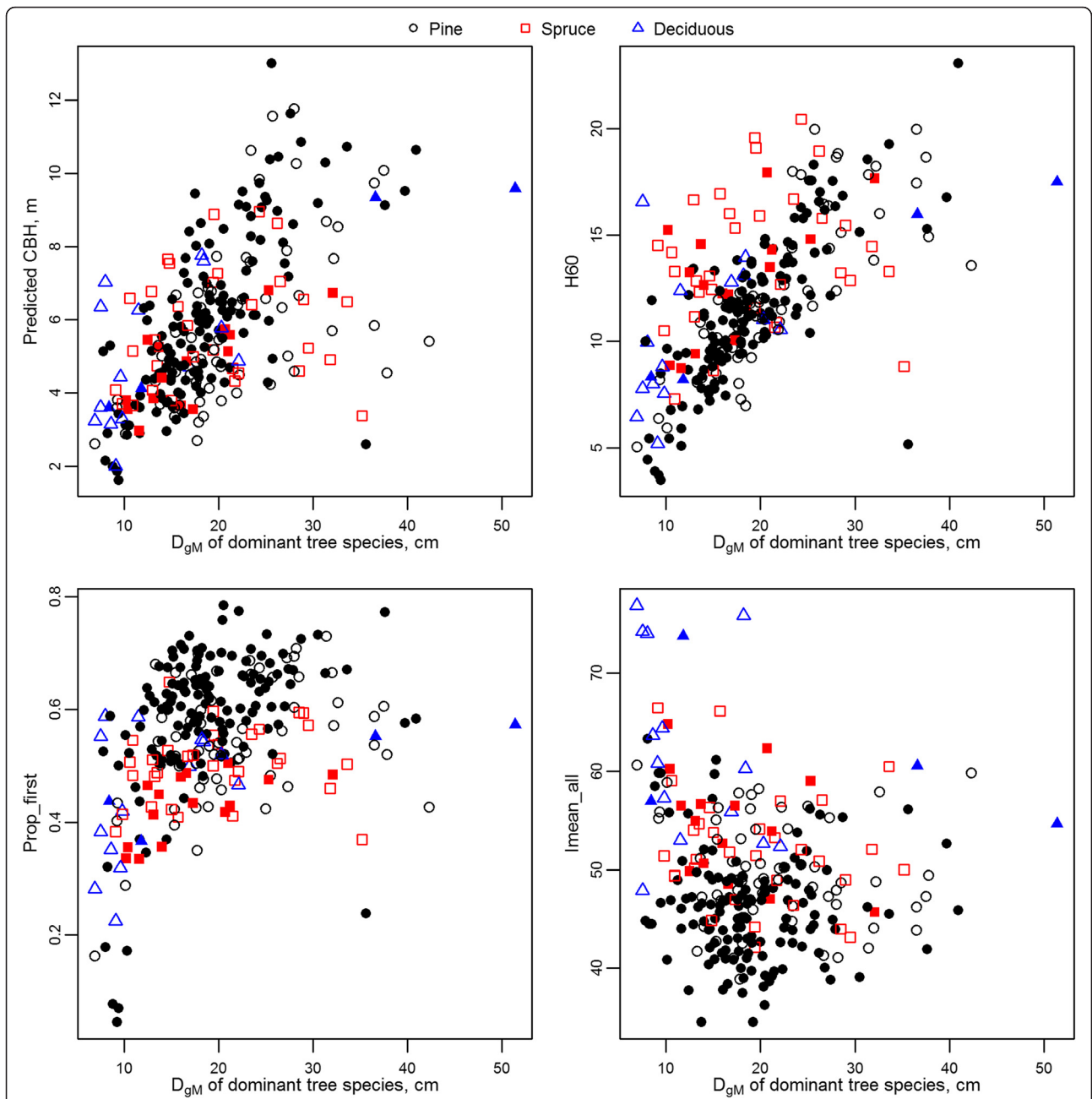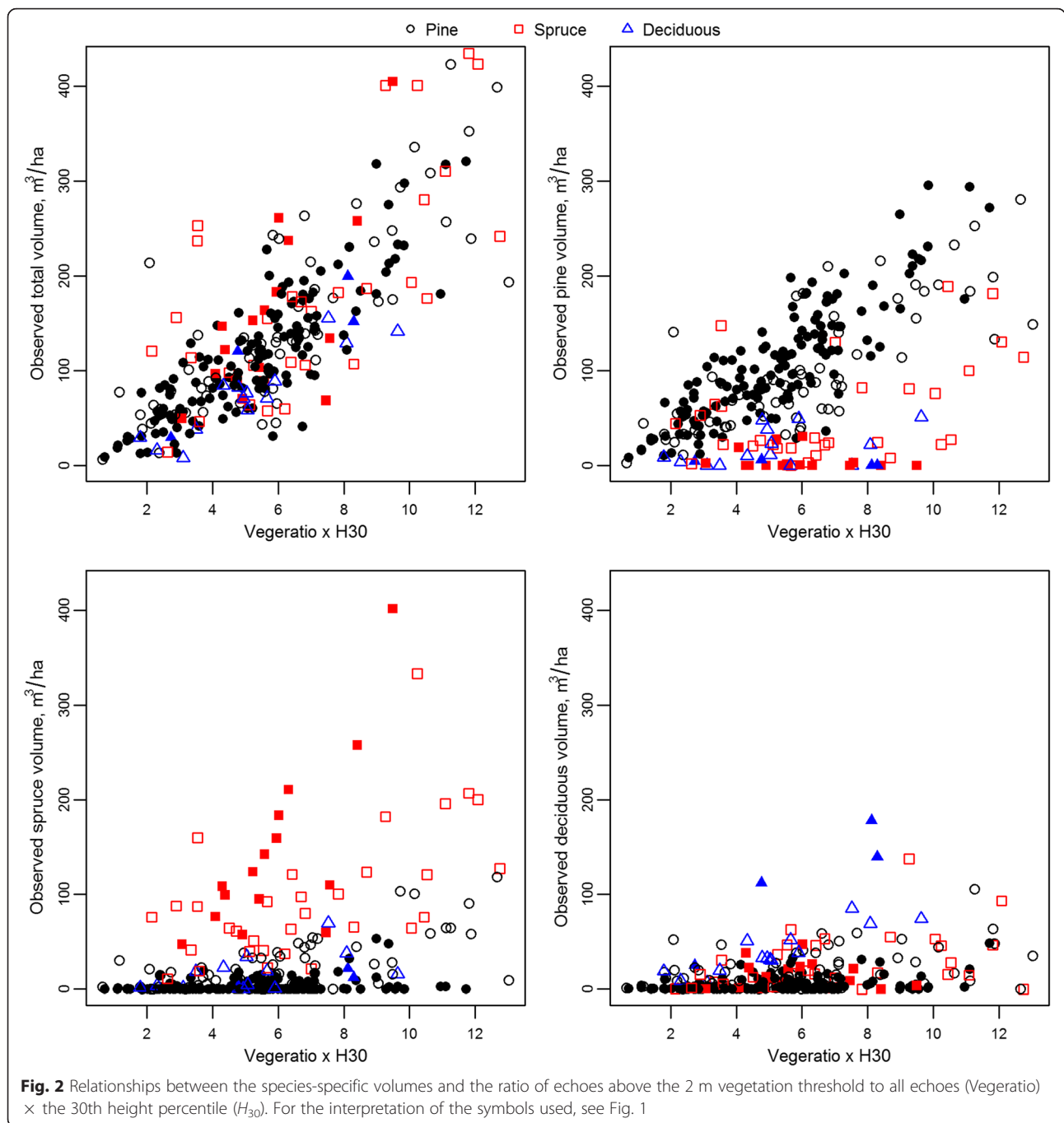Räty *et al. Forest Ecosystems* (2016) 3:1

Page 7 of 17

Fig. 1 Species-specific differences in selected ALS features, when the field-measured D_{gM} is used in the x-axes to assess the invariance of the features to the size. CBH: crown base height, H_{90}: the 90th height percentile, Prop_first: the proportion of the first-of-many returns to all returns above 2 m vegetation threshold, I_{mean_all}: mean intensity value of all returns above the vegetation threshold. Plots dominated by Scots pine, Norway spruce and deciduous trees are illustrated using circles, squares and triangles, respectively, and open or solid symbols refer to a proportion of either ≥50 % or ≥75 % of the dominant species

and 5 were fitted using the plots dominated by pine as the reference level, i.e. applying the models without the species-specific coefficients predicts the species-specific volumes of a pine-dominated plot. Similar to the results obtained previously in this study, the combination of the predictor features differed depending on the dominant species in question. The plots dominated by other than

pine species in particular affected the total volume and species-specific volumes of spruce and deciduous trees in these data. According to the statistical significance of the *t*-test on including the respective coefficient in the models (Tables 3 and 4), separating the plots with $G \geq$ 95 % of pine significantly affected the model predictions based on Sp_{max+95} (Table 4), whereas this information

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 8 of 17

Fig. 2 Relationships between the species-specific volumes and the ratio of echoes above the 2 m vegetation threshold to all echoes (Vegeratio) × the 30th height percentile (H_{30}). For the interpretation of the symbols used, see Fig. 1

was insignificant using the Sp_{75+95} strategy to stratify the species (Table 5).

The performance of the SUR models and the MSN imputation in the training data are compared in Table 6. Further, Figs. 3 and 4 show the predicted versus observed values of the total and species-specific volumes based on the SUR and MSN predictions, respectively. The MSN predictions were generally more accurate than those obtained by SUR except when including the dominant species information from the deciduous dominated

plots. However, the comparison is based on the MSN applied with $k = 5$, which produced the most accurate predictions with this method.

Using both the methods, the predictions regarding the total volume and the volume of pine on pine-dominant plots were well in line with the observed values (Figs. 3 and 4). However, the predictions of the minor species had lower accuracies with both the methods. Due to the coefficient structure of the SUR model (Table 4), the predictions could not show values between 50 and

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 9 of 17

Table 4 The SUR model for the plot volume based on the Sp$_{max+95}$ strategy to stratify the dominant species[1]

Predictor[a]	V_{total}	V_{pine}	V_{spruce}	$V_{deciduous}$
Intercept	−83.1778 ***	−72.1165 ***	−11.8327	−11.8591 *
Species				
P$_{95}$	−17.0412 *	12.47493 *	−13.6919 **	−14.5252 ***
S	21.61634 *	−89.1431 ***	99.7402 ***	7.486791 *
D	−41.1525 **	−87.2537 ***	−0.87776	44.85359 ***
ALS				
Vegeratio	167.1603 ***	113.0326 ***	54.32393 ***	–
H$_{30}$	12.73679 ***	–	–	–
H$_{40}$	–	10.46561 ***	–	–
H$_{95}$	–	–	−0.1767	–
H$_{mean}$	–	–	–	2.173142 ***
I$_{mean, first}$	–	–	–	0.175805 .

[1] Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
[a] Species: pine with $G \geq 95~\%$ (P_{95}), spruce (S) or deciduous trees (D). The *I*, *D*, and *H* refer to intensity, density, and height metrics; Vegeratio is the ratio of echoes above 2 m height to all echoes; and Prop_first is the proportion of first echoes to all echoes. The subscript indicates which descriptive statistic or percentile value was used and whether it was applied to a proportion of the echoes (sd = standard deviation, first = first-of-many echoes)

100 m^3·ha^{-1} of the spruce volume (Fig. 3). The predictions also saturated at certain values (150 m^3·ha^{-1} for spruce), whereas the true observed volumes were considerably higher (e.g. 400 m^3·ha^{-1} for spruce). Also the predictions using *k*-MSN were more inaccurate for the groups of spruce and deciduous plots than for total and pine plots (Fig. 4), but considerably more in line with the observed values compared to the SUR models.

The inclusion of the main species improved both the prediction types considerably. Using Sp$_{max+95}$ as the information on the dominant species, the RMSEs of the pine, spruce, deciduous, and total volumes improved by

Table 6 RMSEs (m^3 ha^{-1}) of the MSN/SUR predictions with different strategies to stratify the main species when evaluated in the training data. With the MSN method, $k = 5$ was applied

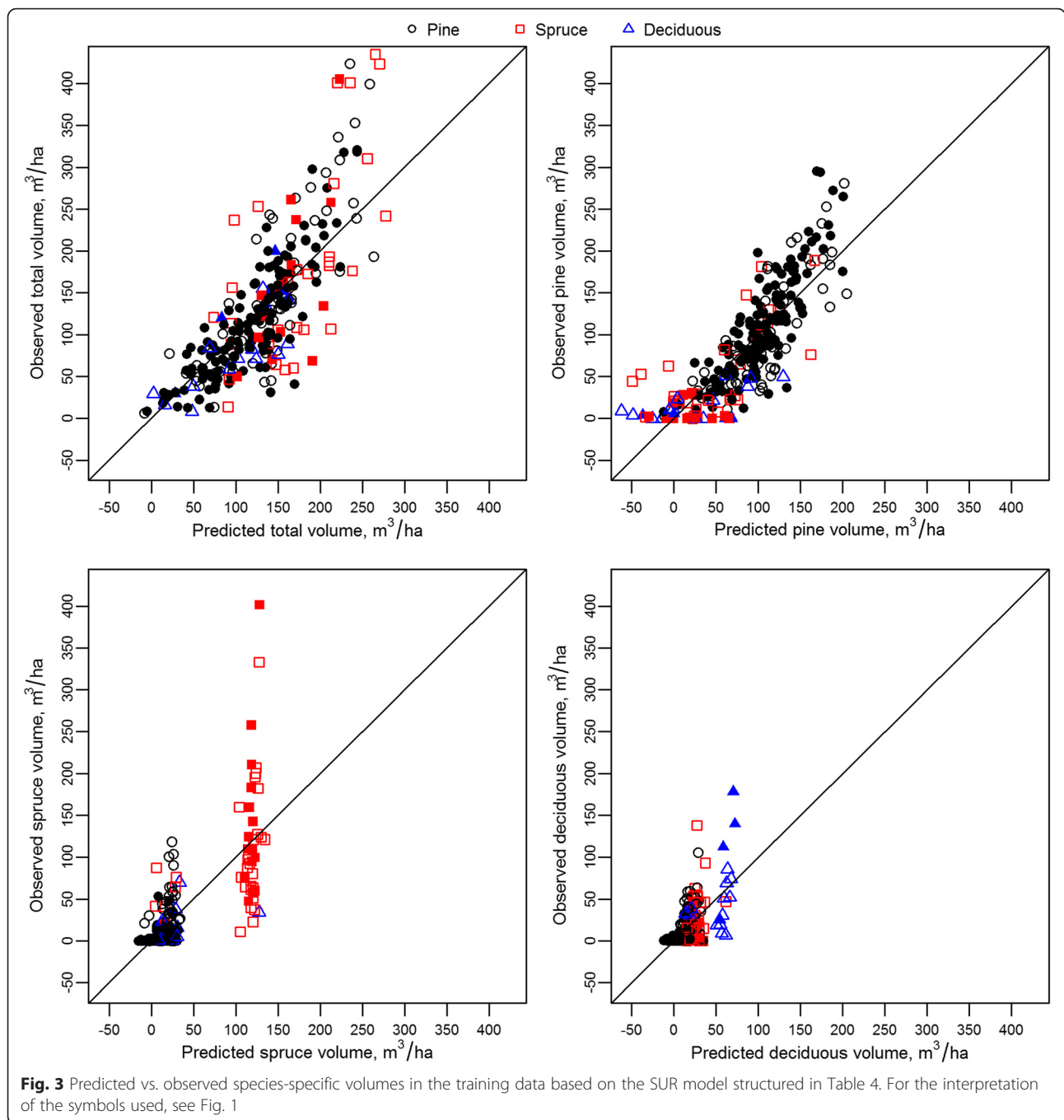| Tree species	Dominant species information		
	–	Sp$_{max+95}$	Sp$_{75+95}$
Pine	42.6/52.9	35.6/37.6	39.0/41.9
Spruce	36.0/47.6	31.2/35.5	33.2/39.3
Deciduous trees	21.3/21.5	18.9/18.8	18.4/17.1
Total	50.2/53.5	51.4/52.5	52.2/52.2

28.9, 25.4, 12.6, and 1.9 %, respectively, using SUR, whereas the corresponding species-specific figures for *k*-MSN were 16.4, 13.3, and 13.6 %, respectively. However, using the *k*-MSN method with the species restriction degraded the accuracy of the total volume by 2.4 %. In the case of *k*-MSN, the species-specific improvement was particularly due to removing close-to-zero observations from the plots dominated by certain species employing the dominant species restriction for the neighborhood. This restriction however reduced the number of potential nearest neighbors for some plots and therefore had a degrading effect on certain accuracy levels.

Classification of the dominant species

The accuracies to predict the dominant species (Table 3) using LDA are summarized in Table 7 and the full confusion matrices of the classifications are presented as Appendix I. The species stratification with only three classes (Sp$_{max}$) was most simple to predict and these predictions also yielded the highest overall accuracies of 73.6 and 76.2 % and kappa coefficients 0.40 and 0.48 using 3 and 4 predictors, respectively. When the plots were classified according to the ≥75 % species proportion, the most problematic case was the 'mixed' class including plots of lower dominance of the various species.

Table 5 The SUR model for the plot volume based on the Sp$_{75+95}$ strategy to stratify the dominant species. For the abbreviations used, please refer to Table 4
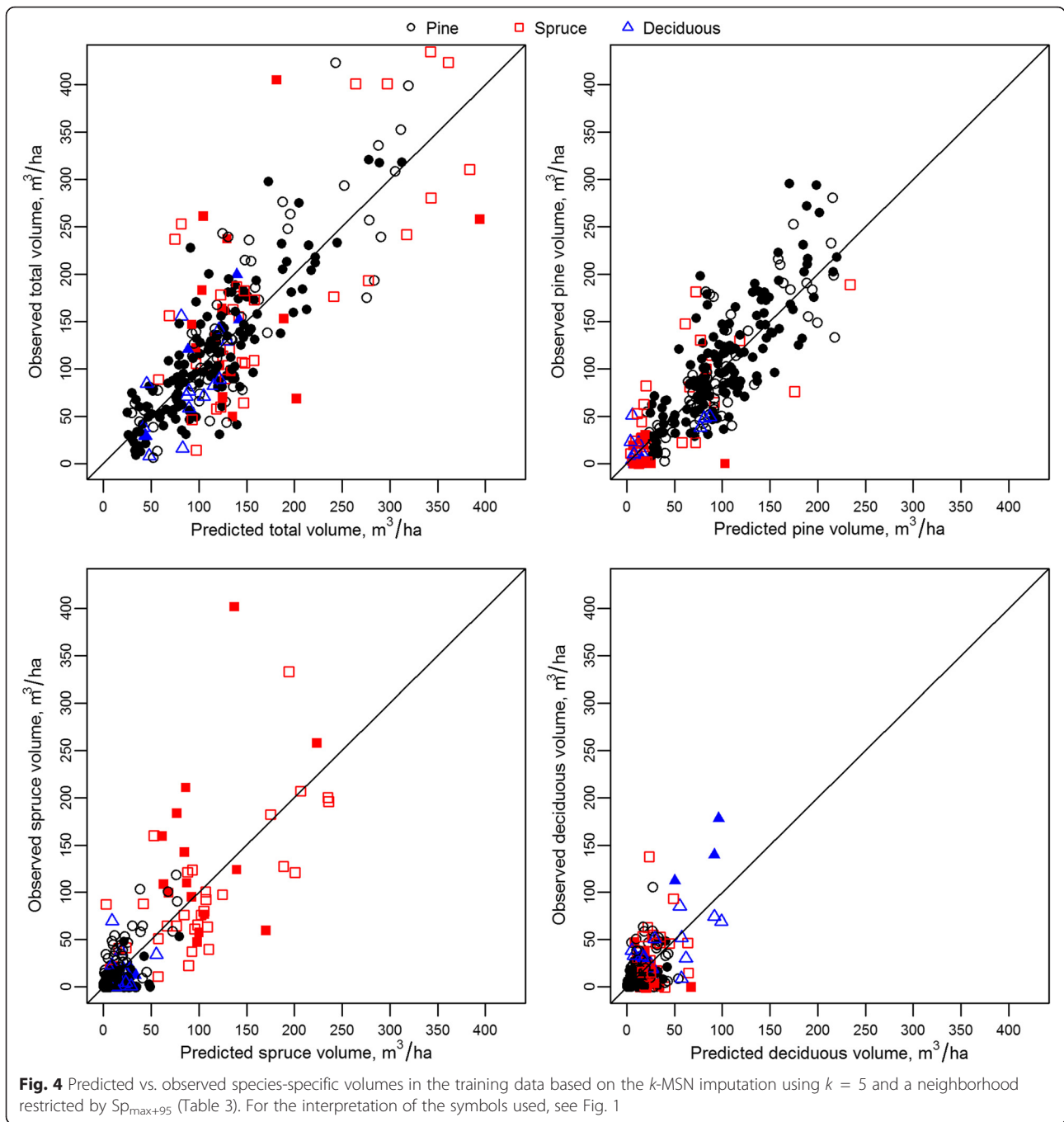
Predictor	V_{total}	V_{pine}	V_{spruce}	$V_{deciduous}$
Intercept	−99.86995 ***	−63.060416 ***	−32.585827 **	−15.382497 **
Species				
P$_{95}$	−12.04051	0.247694	−3.370746	−8.13935 **
S	30.36624 *	−108.00624 ***	129.469296 ***	5.270839
D	−14.26998	−117.48788 ***	−0.555964	100.272105 ***
M	9.99804	−42.385164 ***	35.570611 ***	15.234861 ***
ALS				
Vegeratio	147.73339 ***	87.519629 ***	56.8353 ***	–
H$_{30}$	15.05592 ***	13.061476 ***	–	–
H$_{sd}$	–	–	2.170459 *	–
H$_{mean}$	–	–	–	1.632179 ***
I$_{sd, first}$	–	–	–	0.413249 *

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 10 of 17

Fig. 3 Predicted vs. observed species-specific volumes in the training data based on the SUR model structured in Table 4. For the interpretation of the symbols used, see Fig. 1

Without this class, i.e. using the plots that had a dominant species with ≥ 75 % proportion (*n* = 158), the overall accuracy and the kappa of the classifier were 87.3 % and 0.56, respectively, using 3 predictors.

The inclusion of the pine plots with ≥ 95 % species proportion also complicated the classification and lowered the success rates. Instead of increasing the number of classes in LDA, however, it was found equally accurate to distinguish the plots with ≥ 95 % species

proportion separately based on thresholding of the predictor variables and adding the result manually to the LDA solution. Selecting the plots with ≥ 95 % species proportion manually was implemented and tested using the classification of Sp_{max} and Sp_{75} provided by LDA. In both cases, selecting the plots which had a standard deviation of the intensity values of all pulses < 30, a proportion of first pulses < 0.6 and a density in the 10th height percentile < 0.2 increased the overall accuracy by 4–6 %

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 11 of 17

Fig. 4 Predicted vs. observed species-specific volumes in the training data based on the *k*-MSN imputation using *k* = 5 and a neighborhood restricted by Sp$_{max+95}$ (Table 3). For the interpretation of the symbols used, see Fig. 1

compared to including a class with the plots with ≥ 95 % species proportion in the LDA. Applying these rules mainly resulted in confusion between plots with less pine, but of pine-dominance anyhow. The result could be related to the priors applied with LDA, which should however yield balanced predictors of each class considered. For these reasons, the manually composed classification of Sp$_{max+95}$ and Sp$_{75+95}$ is presented in Table 7 and used later in this study.

Prediction accuracies

To obtain indications of accuracies obtainable in a practical prediction, the dominant species predicted by LDA were combined with the fits of MSN and SUR. The dominant species predictions which included 24–59 % of errors (Table 7) degraded the accuracies obtained earlier (Fig. 5). The *k*-MSN method provided more accurate results than the SUR models. Nevertheless, when the dominant species had to be predicted with the

Table 7 The classification accuracies for the dominant species. For the abbreviations used, please refer to Tables 3 and 4. The confusion matrices of the classifications are presented as Appendix I

Classifier	Number of explanatory variables	Explanatory variables	Overall accuracy (%)	Kappa coefficient
Sp_{max}	3	$I_{mean,\ all}$ + Prop_first + D_{40}	73.6	0.40
	4	$I_{mean,\ all}$ + Prop_first + D_{40} + H_{60}	76.2	0.48
Sp_{max+95}	3	$I_{mean,\ all}$ + Prop_first + D_{40}	55.5	0.34
	4	$I_{mean,\ all}$ + Prop_first + D_{40} + H_{60}	57.7	0.39
Sp_{75}	3	$I_{mean,\ all}$ + Prop_first + D_{30}	58.5	0.34
	4	$I_{mean,\ all}$ + Prop_first + D_{40} + H_{70}	59.6	0.35
Sp_{75+95}	3	$I_{mean,\ all}$ + Prop_first + D_{30}	46.8	0.30
	4	$I_{mean,\ all}$ + Prop_first + D_{40} + H_{70}	45.7	0.28

aforementioned error levels, the predictions that did not consider the dominant species were generally clearly more accurate than those refined with the predicted species.

The most accurate results based on SUR were obtained using the model structured in Table 4 and a LDA model with four explanatory variables to predict Sp_{max} and Sp_{max+95} (Table 8). Considering the RMSEs of the total volumes there were no considerable differences whether the dominant species was either predicted or observed (cf. Table 6, Table 8). However, the RMSE values of spruce were in particular considerably poorer and the predictions of pine plots in particular were biased (Table 8).

The most accurate *k*-MSN predictions for the species-specific volumes were obtained using four explanatory variables to predict Sp_{max+95} (Table 9). Compared to the *k*-MSN predictions using the field-observed dominant tree information, predicting the dominant tree species degraded the RMSEs 25, 13, 15 and 0.1 % for pine, spruce, deciduous and total volumes, respectively. The definition of the dominant species had generally less importance in the *k*-MSN predictions than those based on the SUR models.

Discussion

The obtained results supported the initial hypothesis on the importance of being able to stratify the target plots according to the dominant species. The ALS features were found to be considerably different relative to the volumes of the aforementioned species. The proportion of the minor species in the data in particular varied according to the dominant species, which is realistic with respect to the composition of certain species according to site types, for example. Incorporating the observed dominant species in the MSN and SUR predictions showed potential to improve the accuracies by 13.3–16.4 % and 12.6–28.9 %, respectively, depending on the species.

The proposition to stratify the study area per species or to use the dominant species information overall is not unique to this study. Earlier, Maltamo et al. (2015) stratified the reference data of a species-specific prediction based on *k*-NN according to canopy height and spectral data acquired by ALS and aerial photography, respectively, aiming at a stratification imitating main tree species and stand development stages. The obtained stratification improved the accuracies of the species-specific inventory attributes. Pippuri et al. (2013), on the other hand, used proportions of tree species as a predictor in plot-level basal area predictions based on both *k*-NN and regression methods. A potential of using tree species proportions as a substitute to aerial photographs was noted (Pippuri et al., 2013).

Despite similarities to the earlier studies, our approach has considerable differences in terms of the studied species composition, the source of the dominant species information for the stratification, and the methods to utilize this information. For example, Pippuri et al. (2013) studied hardwood species that are uncommon in Finland and difficult to distinguish even from aerial images, whereas our problem was related to a more frequently occurring forest stand structure in the boreal forest. The coniferous study area had a clearly skewed species distribution, which was strongly dominated by the pine species. However, as noted above, minor species occurred in the area, distinguishing of which had a particular effect on the accuracies.

Even though the theoretical potential to improve the species predictions is clearly shown above, it could not be realized in the practical predictions combining the predicted dominant species to the models formulated. The main reason was the low success rate of classifying the dominant species based on ALS data alone. Maltamo et al. (2015) already cautioned on a limiting inaccuracy originating due to a visual photo-interpretation. A similar effect was observed here due to the lack of discriminative power in the ALS features. In the presence of a fundamentally simple species composition, it was assumed that the area could be inventoried based on ALS data as the sole remotely sensed data source. However,
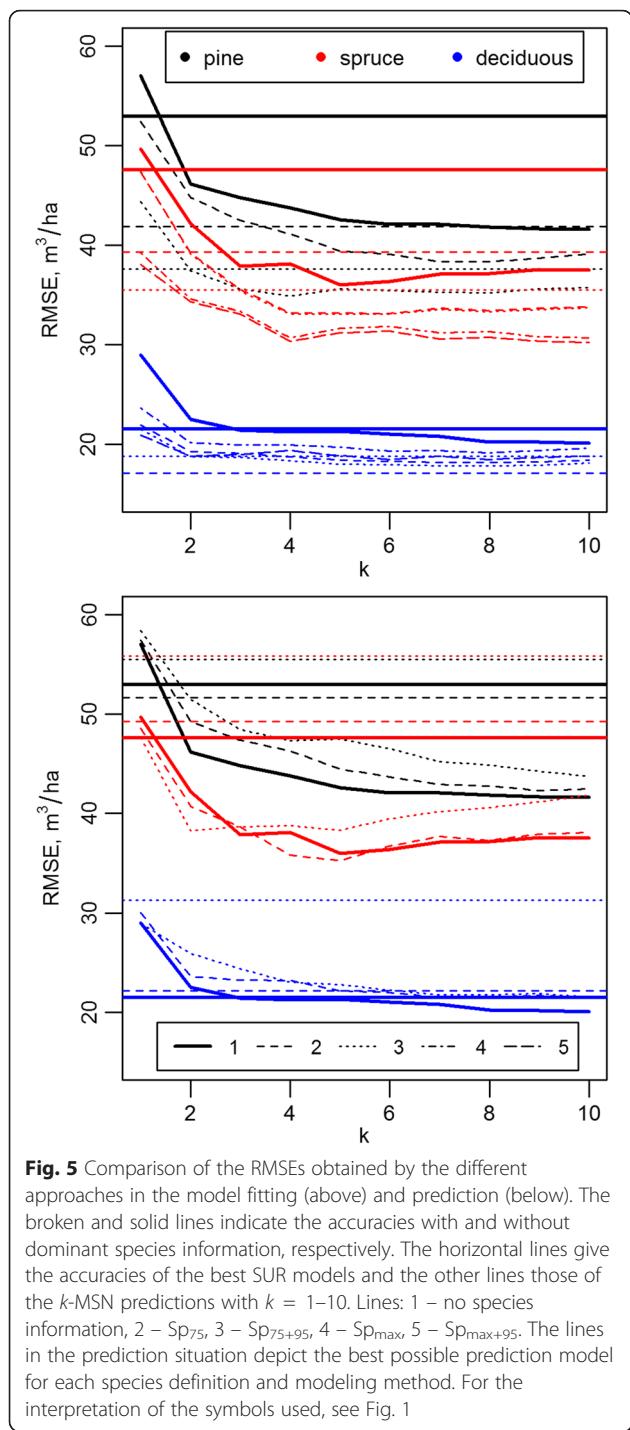
Räty *et al. Forest Ecosystems* (2016) 3:1

Page 13 of 17

Fig. 5 Comparison of the RMSEs obtained by the different approaches in the model fitting (above) and prediction (below). The broken and solid lines indicate the accuracies with and without dominant species information, respectively. The horizontal lines give the accuracies of the best SUR models and the other lines those of the k-MSN predictions with k = 1–10. Lines: 1 – no species information, 2 – Sp_{75}, 3 – Sp_{75+95}, 4 – Sp_{max}, 5 – Sp_{max+95}. The lines in the prediction situation depict the best possible prediction model for each species definition and modeling method. For the interpretation of the symbols used, see Fig. 1

the absence of spectral image data or other information on the minor species clearly degraded the accuracies.

First, it was assumed that the Scots pine and Norway spruce could be better discriminated by the CBH due to the structural differences between the canopies of these species as observed in individual trees (Holmgren and Persson 2004; Holmgren et al. 2008). In the studied plot-level data, this difference was not discriminative,

however. It was particularly challenging to distinguish young forest plots dominated by the coniferous trees based on the CBH. A slightly better discrimination was observed among more mature plots, but with respect to them, the data were overly scarce to draw any more justified conclusions. However, it was verified that the CBH itself could be predicted accurately also at the area-level, which is well in line with the findings of the previous studies (e.g., Maltamo et al. 2010; see also Maltamo et al. 2012).

Although the CBH was an inadequate feature for separating the species, some other ALS features had more discriminative power. The features describing the proportion of the first echoes and the intensity recordings of the echoes were among the best features for the species discrimination. Even though a leveling difference between the species studied was observed in the data, this difference was not as strong as observed earlier. For example, based on Fig. 1 in Vauhkonen et al. (2014b), the plots with a varying degree of dominance of Scots pine were more distinct from the other species in a boreal forest closely resembling the conditions studied here. The difference could be related to the calibration of the intensity recordings: in Vauhkonen et al. (2014b), the data had been range-corrected (e.g., Korpela et al. 2010), whereas our data were not calibrated nor did we have an access to the trajectory data to perform the calibration.

Even though an obvious difference in the ALS features between the sample plots dominated by varying proportions of pine (i.e. ≥ 75 % or 95 %) was not observed, distinguishing these plots was attempted due to the potential to increase the information for the later species-specific predictions. It was observed that classifying these plots successfully was difficult with LDA, whereas the classification accuracy could be slightly improved by first excluding the ≥ 95 % class and later manually sub-selecting these plots based on the determined threshold values. These values were not even optimized, but selected based on a visual assessment of the ALS features. The result could correspond to the observation made by Heinzel and Koch (2011), who obtained higher classification rates by reducing a so-called classification depth, i.e. reducing the number of classes by combining species. The difficulty of the classification task is increased by an increasing number of classes, but including such decision rules could improve the rates obtainable.

The results of the species classification based on the linear discriminant analysis were found comparable with the corresponding results reported by the previous studies. For example, Ørka et al. (2013) evaluated various remote sensing inventory approaches and data sources for the prediction of main species and species proportions. Less differences were found between area and tree based

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 14 of 17

Table 8 RMSEs and (BIASes) of the species-specific volumes based on the SUR models, when the dominant species were predicted by LDA. For the abbreviations used, please refer to Table 3

Species information	Number of explanatory variables	Volume (m³ ha⁻¹)			
		Pine	Spruce	Deciduous	Total
Sp$_{max}$	3	52.2 (11.5)	60.0 (−8.3)	22.8 (−3.4)	56.8 (−0.5)
Sp$_{max}$	4	53.0 (14.2)	49.3 (−11.4)	23.2 (−3.8)	54.1 (−1.3)
Sp$_{max+95}$	3	51.7 (11.3)	61.1 (−8.1)	22.2 (−2.9)	58.2 (−0.2)
Sp$_{max+95}$	4	53.1 (13.9)	49.8 (−11.1)	22.6 (−3.1)	55.4 (−0.8)
Sp$_{75}$	3	56.7 (13.9)	61.3 (−8.1)	31.9 (−6.0)	54.9 (−0.6)
Sp$_{75}$	4	54.4 (11.5)	55.8 (−5.1)	31.8 (−5.8)	54.4 (0.2)
Sp$_{75+95}$	3	55.5 (12.7)	60.4 (−7.1)	31.8 (−5.1)	55.2 (0.2)
Sp$_{75+95}$	4	55.7 (12.8)	59.9 (−7.8)	31.3 (−4.6)	54.9 (0.0)

inventory approaches than data sources. Using an area-based approach, an average accuracy of 89.1 % (κ = 0.78) of predicting the dominant species was reported. However, this result was based on separating the plots dominated by certain species, while an inclusion of a mixed class reduced the average performance to 80.4 % (κ = 0.70). The results were always improved by an availability of spectral data, and no separate figures with respect to the dominant species classification based solely on ALS are reported. However, the performance reduction observed by including a mixed class corresponds well to our findings and highlights the operational challenges that are emerging already from the definitions of the dominant species. A better strategy could be to predict the probability of a plot to consist of certain species, which somewhat resembles fuzzy classification of species tested already by Packalén and Maltamo (2006) using remotely sensed data.

The previous studies on predicting species-specific timber volumes by ALS have generally considered *k*-NN estimation approaches. We also produced the corresponding predictions by a Seemingly Unrelated Regression. The SUR and the *k*-MSN approach differ between each other on how the dominant species information

was used for the predictions. There is also a difference to the previous studies, in that we did not consider the dominant species as additional predictors added to the methods similar to the ALS features, for example. Rather, the dominant species information was added to each prediction method as an information that was fundamentally missing, considering the natural properties of a method. In the SUR modeling, the dominant species information was accounted for as a categorical variable, therefore introducing separate coefficients to predict the species proportions under a dominance of each dominant species. In the *k*-MSN modeling, the neighborhood was restricted in the prediction stage such that only the neighbor candidates matching the dominant species condition were considered.

Whether no or limited dominant species information is available, the proposed prediction methods can still be operated in a population-specific mode. The presence of improved information, such as identifying the plots with ≥ 95 % pine proportion, improves the accuracies. In turn, false predictions of the dominant species have an opposite effect. However, these modifications of the modeling strategy do not compensate for the defects in the modeling data used for training the prediction

Table 9 RMSEs and (BIASes) of the species-specific volumes based on MSN, when the dominant species were predicted by LDA. For the abbreviations used, please refer to Table 3

Species information	Number of explanatory variables	Volume (m³ ha⁻¹)			
		Pine	Spruce	Deciduous	Total
Sp$_{max}$	3	44.9 (0.9)	38.1 (−1.9)	22.2 (−0.6)	53.8 (−1.7)
Sp$_{max}$	4	44.8 (0.5)	35.3 (−0.9)	22.0 (−0.4)	52.0 (−0.8)
Sp$_{max+95}$	3	45.3 (1.6)	38.1 (−1.9)	22.2 (−0.8)	51.5 (−1.1)
Sp$_{max+95}$	4	44.5 (1.2)	35.2 (−1.3)	21.8 (−0.7)	51.5 (−0.8)
Sp$_{75}$	3	47.1 (−1.4)	38.4 (−2.8)	22.8 (−0.6)	54.6 (−4.8)
Sp$_{75}$	4	46.8 (−0.5)	38.2 (−2.2)	22.5 (−0.6)	51.9 (−3.3)
Sp$_{75+95}$	3	45.9 (−2.4)	39.6 (−3.0)	22.6 (−0.7)	53.5 (−6.0)
Sp$_{75+95}$	4	47.1 (−1.6)	37.7 (−2.4)	22.9 (−0.3)	53.0 (−4.3)

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 15 of 17

methods. For example, the saturating and incorrect predictions of both the methods in particular for the species occurring moderately in the area but dominating some plots (e.g. spruce) could be partly explained by the absence of spruce-dominated, very highly stocked plots. The only way to mitigate for such effects is to acquire an adequately representative reference data, which is already noted by Vauhkonen et al. (2012).

Overall, taken together with the earlier results, improvements in the scale of 9–47 and 33–50 percentage points are obtainable by first balancing the field reference data with respect to the inadequately represented species and then including spectral information as predictors in addition to those extracted by sole ALS data, respectively (Vauhkonen et al. 2012). According to our results, an additional 13.3 %–28.9 % increase in the species-specific accuracies may be obtainable by correctly predicting the dominant species and incorporating this information in the estimation. The results thus suggest the importance of investing in the data sources to improve the quality of the information. Yet, corresponding accuracy improvements are also reported based on optimizing the distance metric or the feature space considered by the NN methods (Latifi et al. 2010; Packalén et al. 2012).

Conclusions

The relationships between the predictor features derived from the ALS data and the volumes of Scots pine, Norway spruce, and deciduous species were considerably different depending on the dominant species. Incorporating the observed dominant species in the predictions based on MSN and SUR showed a potential improve the prediction accuracies by 13.3–16.4 % and 12.6–28.9 %, respectively, depending on the species. However, the overall accuracy of classifying the dominant species based solely on ALS data (76 % at best) was not adequate for reaching the aforementioned improvements. Rather, the predictions that did not consider the dominant species were more accurate than those refined with the predicted species. The MSN method gave slightly better results than models fitted with SUR. Determining the dominant species based solely on ALS data is deemed challenging, but important in areas where the species composition is otherwise seemingly homogeneous except being dominated by certain species. Provided an increase in the accuracy to determine the dominant species based on other data sources, for example, considerable improvements in the species-specific accuracies are obtainable by accounting for this information following a strategy proposed here.

Appendix I

Confusion matrices of the species classifications summarized in Table 7.

Table 10 Sp_{max}, 3 explanatory variables

		Predicted species		
		Pine	Spruce	Decid.
Observed species	Pine	163	33	13
	Spruce	10	24	7
	Decid.	2	5	8

Table 11 Sp_{max}, 4 explanatory variables

		Predicted species		
		Pine	Spruce	Decid.
	Pine	160	34	15
Observed species	Spruce	4	33	4
	Decid.	3	3	9

Table 12 Sp_{max+95}, 3 explanatory variables

		Predicted species			
		Pine, >95 %	Pine	Spruce	Decid.
	Pine, >95 %	41	24	6	1
	Pine	25	74	27	11
Observed species	Spruce	1	9	24	7
	Decid.	0	2	5	8

Table 13 Sp_{max+95}, 4 explanatory variables

		Predicted species			
		Pine, >95 %	Pine	Spruce	Decid.
	Pine, >95 %	41	25	5	1
	Pine	25	70	29	13
Observed species	Spruce	1	3	33	4
	Decid.	0	3	3	9

Table 14 Sp_{75}, 3 explanatory variables

		Predicted species			
		Pine	Spruce	Decid.	Mixed
	Pine	105	8	7	19
	Spruce	0	10	1	4
Observed species	Decid.	0	1	2	1
	Mixed	29	25	15	38

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 16 of 17

Table 15 Sp$_{75}$, 4 explanatory variables

		Predicted species			
		Pine	Spruce	Decid.	Mixed
	Pine	109	5	7	18
	Spruce	0	9	2	4
Observed species	Decid.	0	0	2	2
	Mixed	31	24	14	38

Table 16 Sp$_{75+95}$, 3 explanatory variables

		Predicted species				
		Pine, >95 %	Pine	Spruce	Decid.	Mixed
	Pine, >95 %	41	15	4	1	11
	Pine	17	33	4	5	8
Observed species	Spruce	0	0	10	1	4
	Decid.	0	0	1	2	1
	Mixed	9	20	25	15	38

Table 17 Sp$_{75+95}$, 4 explanatory variables

		Predicted species				
		Pine, >95 %	Pine	Spruce	Decid.	Mixed
	Pine, >95 %	41	22	2	1	6
	Pine	17	32	3	5	10
Observed species	Spruce	0	0	9	2	4
	Decid.	0	0	0	2	2
	Mixed	9	23	24	14	37

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JR participated in the computations of the field plot data, carried out the analyses related to dominant species and SUR models and drafted these parts of the manuscript under the supervision of JV. JV extracted the ALS features, carried out the analyses related to the MSN method, and conceived and organized the computations and reporting. MM participated in the computations of the field plot data, conceived some parts of the study and participated in its design. TT managed the funding of the study and participated in its design and coordination. All authors read and approved the final manuscript.

References

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 96:37–46

Crookston NL, Finley AO (2007) yalmpute: An R package for k-NN imputation. J Stat Softw 23:1–16

Eid T, Gobakken T, Næsset E (2004) Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scand J For Res 19:512–523

Heinzel J, Koch B (2011) Exploring full-waveform LiDAR parameters for tree species classification. Int J Appl Earth Obs Geoinfo 13:152–160

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 Geoinfo 18:101–110

Henningsen A, Hamann JD (2007) System fit: A package for estimating systems of simultaneous equations in R. J Stat Softw 23:1–40

Holmgren J, Persson Å (2004) Identifying species of individual trees using airborne laser scanner. Remote Sens Environ 90:415–423

Holmgren J, Persson Å, Söderman U (2008) Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. Int J Remote Sens 29:1537–1552

Hudak AT, Crookston NL, Evans JS, Hall DE, Falkowski MJ (2008) Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sens Environ 112:2232–2245

Kim S, McGaughey RJ, Andersen HE, Schreuder G (2009) Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sens Environ 113:1575–1586

Koivuniemi J, Korhonen KT (2006) Inventory by compartments. In: Kangas A, Maltamo M (eds) Forest inventory – Methodology and applications. Managing Forest Ecosystems, vol 10. Springer, Dordrecht, pp 271–278

Korhonen M (2012) Puuston latvusrajan ennustaminen harvapulssisesta laserkeilausaineistosta mäntyvaltaisella alueella ja latvusrajan mittauksen tehostaminen (In Finnish for" Predicting crown base height of the tree stock using sparse airborne laser scanning data in a pine-dominated area and streamlining the reference measurements of the crown base height"). In: M. Sc. thesis. University of Eastern Finland, Joensuu

Korhonen L, Peuhkurinen J, Malinen J, Suvanto A, Maltamo M, Packalén P, Kangas J (2008) The use of airborne laser scanning to estimate sawlog volumes. Forestry 81:499–510

Korpela I, Ørka HO, Hyyppä J, Heikkinen V, Tokola T (2010) Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. ISPRS J Photogramm Remote Sens 65:369–379

Korpela I, Tokola T (2006) Potential of aerial image-based monoscopic and multiview single-tree forest inventory: A simulation approach. For Sci 52:136–147

Kotamaa E, Tokola T, Maltamo M, Packalén P, Kurttila M, Mäkinen A (2010) Integration of remote sensing-based bioenergy inventory data and optimal bucking for stand-level decision making. Eur J For Res 129:875–886

Laasasenaho J (1982) Taper curve volume functions for pine, spruce and birch. Comm Inst For Fenn 108:74

Latifi H, Nothdurft A, Koch B (2010) Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. Forestry 83:395–407

Lindberg E, Holmgren J, Olofsson K, Wallerman J, Olsson H (2010) Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. Int J Remote Sens 31:1175–1192

Magnussen S, Boudewyn P (1998) Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can J For Res 28:1016–1031

Maltamo M, Bollandsås OM, Vauhkonen J, Breidenbach J, Gobakken T, Næsset E (2010) Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. Forestry 83:257–268

Maltamo M, Mehtätalo L, Vauhkonen J, Packalén P (2012) Predicting and calibrating tree attributes by means of airborne laser scanning and field measurements. Can J For Res 42:1896–1907

Maltamo M, Ørka HO, Bollandsås OM, Gobakken T, Næsset E (2015) Using pre-classification to improve the accuracy of species-specific forest attribute estimates from airborne laser scanner data and aerial images. Scand J For Res 30:336–345

Maltamo M, Packalen P (2014) Species-specific management inventory in Finland. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry applications of airborne

Räty *et al. Forest Ecosystems* (2016) 3:1

Page 17 of 17

laser scanning - concepts and case studies. Managing Forest Ecosystems, vol 27. Springer, Dordrecht, pp 241–252

Maltamo M, Packalén P, Kallio E, Kangas J, Uuttera J, Heikkilä J (2011) Airborne laser scanning based stand level management inventory in Finland. In: Paper presented at the Silvi Laser 2011 – 11th International Conference on LiDAR Applications for Assessing Forest Ecosystems, 16–20 October 2011, Hobart, Australia., http://www.iufro.org/download/file/8239/5065/40205-silvilaser2011_pdf/ Accessed 2 Nov 2015

Maltamo M, Næsset E, Bollandsås OM, Gobakken T, Packalén P (2009a) Non-parametric prediction of diameter distributions using airborne laser scanner data. Scand J For Res 24:541–553

Maltamo M, Packalén P, Suvanto A, Korhonen KT, Mehtätalo L, Hyvönen P (2009b) Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. Eur J For Res 128:305–317

Maltamo M, Peuhkurinen J, Malinen J, Vauhkonen J, Packalén P, Tokola T (2009c) Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. Silva Fenn 43:507–521

Maltamo M, Næsset E, Vauhkonen J (2014) Forestry applications of airborne laser scanning - concepts and case studies. Managing Forest Ecosystems, vol 27. Springer, Dordrecht

Moeur M, Stage AR (1995) Most similar neighbor: An improved sampling inference procedure for natural resource planning. For Sci 41:337–359

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

Næsset E (2014) Area-based inventory in Norway – From innovation to an operational reality. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry applications of airborne laser scanning - concepts and case studies. Managing Forest Ecosystems, vol 27. Springer, Dordrecht, pp 215–240

Nord-Larsen T, Schumacher J (2012) Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. Remote Sens Environ 119:148–157

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

Ørka HO, Gobakken T, Næsset E, Ene L, Lien V (2012) Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. Can J Remote Sens 38:125–138

Ørka HO, Næsset E, Bollandsås OM (2009) Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sens Environ 113:1163–1174

Packalén P, Maltamo M (2006) Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. For Sci 52:611–622

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 109:328–341

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

Packalén P, Suvanto A, Maltamo M (2009) A two stage method to estimate species-specific growing stock. Photogramm Eng Remote Sens 75:1451–1460

Packalén P, Temesgen H, Maltamo M (2012) Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Can J Remote Sens 38:557–569

Pippuri I, Maltamo M, Packalen P, Mäkitalo J (2013) Predicting species-specific basal areas in urban forests using airborne laser scanning and existing stand register data. Eur J For Res 132:999–1012

Core Team R (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org Accessed 2 Nov 2015

Siipilehto J (1999) Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fenn 33:281–301

Siipilehto J, Sarkkola S, Mehtätalo L (2007) Comparing regression estimation techniques when predicting diameter distributions of Scots pine on drained peatlands. Silva Fenn 41:333–349.

Ståhl G, Allard A, Esseen P-A, Glimskär A, Ringvall A, SvenssonJ SS, Christensen P, Gallegos Torell Å, Högström M, Lagerqvist K, Marklund L, Nilsson B, Inghe O (2011) National Inventory of Landscapes in Sweden (NILS) – Scope, design, and experiences from establishing a multiscale biodiversity monitoring system. Environ Monit Assess 173:579–595

Suratno A, Seielstad C, Queen L (2009) Tree species identification in mixed coniferous forest using airborne laser scanning. ISPRS J Photogramm Remote Sens 64:683–693

Tompalski P, Coops NC, White JC, Wulder MA (2014) Simulating the impacts of error in species and height upon tree volume derived from airborne laser scanning data. For Ecol Manage 327:167–177

Torabzadeh H, Morsdorf F, Leiterer R, Schaepman ME (2014) Fusing imaging spectrometry and airborne laser scanning data for tree species discrimination. IEEE Int Geosci Remote Sens Symp 2014:1253–1256

van Ewijk KY, Randin CF, Treitz PM, Scott NA (2014) Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. Remote Sens Environ 150:120–131

Vauhkonen J (2010) Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. Int J Remote Sens 31:1213–1226

Vauhkonen J, Ørka HO, Holmgren J, Dalponte M, Heinzel J, Koch B (2014a) Tree species recognition based on airborne laser scanning and complementary data sources. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry applications of airborne laser scanning - concepts and case studies. Managing Forest Ecosystems, vol 27. Springer, Dordrecht, pp 135–156

Vauhkonen J, Packalen P, Malinen J, Pitkänen J, Maltamo M (2014b) Airborne laser scanning-based decision support for wood procurement planning. Scand J For Res 29(sup1):132–143

Vauhkonen J, Seppänen A, Packalén P, Tokola T (2012) Improving species-specific plot volume estimates based on airborne laser scanning and image data using alpha shape metrics and balanced field data. Remote Sens Environ 124:534–541

Venables WN, Ripley BD (2002) Modern applied statistics with S. Springer, Dordrecht

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

Wallerman J, Holmgren J (2007) Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. Remote Sens Environ 110:501–508

White JC, Wulder MA, Varhola A, Vastaranta M, Coops NC, Cook BD, Pitt D, Woods M (2013) A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach (Version 2.0). Canadian Forest Service, Canadian Wood Fibre Centre, Information report FI-X-010.http://www.cfs.nrcan.gc.ca/pubwarehouse/pdfs/34887.pdf Accessed 2 Nov 2015

Woods M, Pitt D, Penner M, Lim K, Nesbitt D, Etheridge D, Treitz P (2011) Operational implementation of a LiDAR inventory in Boreal Ontario. For Chron 87:512–528.

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J Amer Stat Ass 57:348–368.