The effect of forest characteristics on ALS-based inventory results

R. Haapanen[1], S. Tuominen[2], M. Holopainen[3] and R. Viitala[4]

[1]Haapanen Forest Consulting, Kärjenkoskentie 38, 64810 Vanhakylä, Finland; +358-40-502 1571; reija.haapanen@haapanenforestconsulting.fi

[2]Finnish Forest Research Institute, PL 18, 01301 Vantaa, Finland; +358-10-211 2167; sakari.tuominen@metla.fi

[3]University of Helsinki, Department of Forest Resource Management, P.O. Box 27, 00014 University of Helsinki, Finland; +358- 9-191 58181; markus.holopainen@helsinki.fi

[4]Risto Viitala, HAMK University of Applied Sciences, Saarelantie 1, 16970 Evo, Finland; +358-3-646 5300; risto.viitala@hamk.fi

Introduction

The management planning system of Finnish private forests is currently changing from mainly field inventory–based into one that makes use of airborne laser scanning (ALS) data, aerial photographs and a sample of field plots. In the traditional system every stand of the inventoried area is visited and some relascope sample plots measured within the stand. This is time consuming, expensive, and prone to error. In a study by Haara and Korhonen (2004) the tree species-specific relative mean volume RMSEs varied from 29% (Scots pine) to 65% (deciduous species trees), whereas the relative RMSE of stand mean total volume was 25%. Currently, the new ALS-based method is being tested in pilot projects. However, more should be known of its suitability in different types of forests.

Our aim was to compare the success of ALS-based forest inventory in two areas of Finnish boreal forest, with different forest structure, located approximately 200 km from each other in N-S and 100 km in E-W direction. The southern study site, Evo, contains mostly mineral soil stands with a relatively high mean volume and quite an even mixture of Norway spruce, Scots pine and deciduous tree species, while in the northern study site, Kuortane, the amount of mires is larger, the mean volume lower, and forests are dominated by Scots pine. The estimation was carried out with the *k*-nearest neighbor (*k*-nn) algorithm and we operated at the field plot level. The forest variables estimated included the mean volume of growing stock (m^3/ha), basal area (m^2/ha), height (m), diameter at breast height (cm), and the volumes of Scots pine, Norway spruce, and deciduous species trees (m^3/ha).

Material and methods

From the Evo area 282 fixed-radius (9.77 m) field plots were available, while in Kuortane the data consisted of 335 relascope sample plots with a maximum radius of 12.52 m. In Evo, diameters and heights of all sampled trees had been measured, while in Kuortane only diameters were measured and heights were based on a model calibrated with measurements from visually selected median trees. See Table 1 and Figure 1 for the characteristics of field sample plots.

Table 1. Characteristics of field sample plots in Kuortane and Evo study areas.

	Kuortane		**Evo**	
	Mean	**Std**	**Mean**	**Std**
D, cm	16.4	5.8	21.1	9.4
• Scots pine	17.0	6.3	23.3	9.5
• Norway spruce	17.5	7.2	17.4	10.2
• Deciduous species	14.4	6.2	17.8	8.7
H, m	13.2	4.7	17.0	6.7
• Scots pine	13.2	4.9	17.9	5.9
• Norway spruce	15.0	6.0	13.9	7.5
• Deciduous species	14.0	4.0	16.3	5.6
Basal area, m^2/ha	16.2	8.9	19.9	10.3
• Scots pine	12.1	8.2	7.7	8.6
• Norway spruce	2.6	6.2	6.7	8.5
• Deciduous species	1.5	3.6	5.4	6.0
Volume, m^3/ha	116.4	91.7	179.0	115.4
• Scots pine	83.1	70.5	69.9	86.8
• Norway spruce	23.0	60.7	63.5	94.8
• Deciduous species	10.3	26.9	45.6	56.3

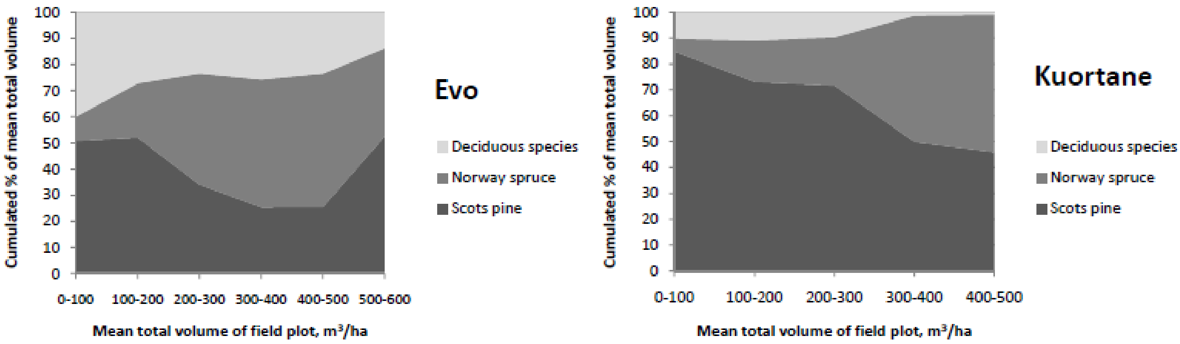

Figure 1. Tree species dominances in different volume classes in the field plot data of Evo and Kuortane.

From both study sites, low pulse-density (<2 hits/m^2) ALS data and orthorectified aerial photographs (near-infrared, red and green bands) were acquired. The ground resolution of aerial photographs was 0.5 m. ALS data were rasterized to a similar resolution for utilizing the textural features.

A large number (172) of statistical and textural features were extracted from the remotely sensed material. These included:

- Means, standard deviations and Haralick textural features (Haralick et al. 1973, Haralick 1979) of spectral values of aerial photographs, ALS height and intensity (first pulse only) from a window of 20 x 20 m.

- Height statistics for the first and last pulses as in Suvanto et al. (2005) of the points inside the field plot area.
- A number of std's extracted from a 32 x 32 pixel window using block sizes from 1 to 8 pixels.

All features were standardized to a mean of 0 and std of 1. The *k*-nearest neighbor (*k*-nn) method was used to estimate values for forest variables. Value of *k* was set to 5, Euclidean distances were used to measure closeness in the feature space and even weights were given to the 5 nearest neighbors. The RMSEs were estimated via leave-one-out cross-validation.

Automatic feature selection was carried out using a simple genetic algorithm presented by Goldberg (1989), and implemented in the GAlib C++ library (Wall 1996). The objective variable to be minimized during the process was a weighted combination of relative RMSEs of *k*-nn estimates for mean total volume, mean volumes of Scots pine, Norway spruce and deciduous species, mean diameter and mean height, with total volume having a weight of 50%, and the remaining variables 10% each. At this point of the study, we have not yet run the feature selection for Kuortane separately, but used features selected for Evo. There were 11 features selected into the final set, of which 7 were based on ALS height, one on ALS intensity and 3 on aerial photograph data (see Holopainen et al. 2008 for more details on estimation and feature selection processes and resulting features).

Results

The mean volume RMSE% was somewhat higher in Kuortane (34 vs. 29% corresponding to 40 and 53 m^3/ha). Tree species-specific estimation success followed the forest structure on each study area. In Evo, with relatively even proportions of tree species, the species-specific mean volume RMSEs were 80-90% of the means, whereas in Kuortane, with Scots pine dominating, the mean volume RMSE of Scots pine was 55%, while those for Norway spruce and deciduous species were 147 and 215%, respectively.

Figure 2 shows the relative mean volume error in different volume classes. Relative errors were largest in the 0-100 m^3/ha class as it may consist of a very heterogeneous selection of field plots: clear cut plots, plots with seed trees only, plots located in an opening in the forest, plots on treeless or sparsely forested mires, seedling stands and young stands. This was the case especially in Evo. Furthermore, in Evo this class had a high proportion of deciduous species trees (Fig. 1). In Kuortane the class was more homogeneous (mainly Scots pine, see Fig. 1), the amount of field plots was largest, and the RMSE% consequently far lower than in Evo. (However, the error did not follow the amount of field plots in each class.)

We set out to analyze the effect of within plot tree species distribution on the RMSEs. From table 2 it can be seen that the RMSE% first decreases when the share of dominant tree species of plot volume increases, but the largest errors are found in the class where the dominant species accounts for 90-100% of the plot volume. In both study areas, majority of plots belonged to this class, but this still did not mean a sufficient number of neighbors: in Evo, 2 Scots pine and 2 Norway spruce plots were separated from the main group by their exceptionally high volumes.

As a consequence, these plots had large estimation errors, significantly raising the RMSE of the whole class.

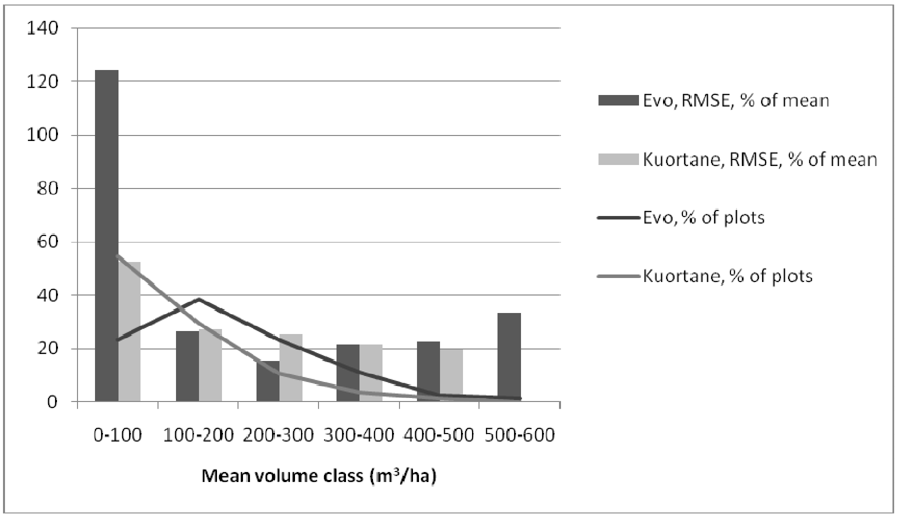

Figure 2. Relative mean volume errors and the share of field plots in different volume classes. Note that in Kuortane there are no plots in the 500-600 m^3/ha class.

Table 2. Effect of within-plot species distribution on the relative mean volume RMSE.

	Mean volume RMSE, %		**Proportion of field plots, %**	
Dominant tree species, % of volume	**Evo**	**Kuortane**	**Evo**	**Kuortane**
<50	29.7	35.3	7.9	3.0
50-60	28.8	27.6	17.7	7.2
60-70	19.5	25.4	14.0	6.6
70-80	23.0	36.0	12.8	4.8
80-90	23.7	33.3	14.3	10.7
90-100	32.0	36.8	33.2	67.8

Discussion

In this study we were interested in the accuracy of ALS-based forest variable estimation employing *k*-nn method at field plot level in two different areas of Finnish boreal forest. Both the tree species-specific errors and the errors by volume classes shoved variation between the areas. One can ask, whether the high relative errors are due to 1) lack of suitable neighbors (similar plots, i.e. amount of field plots is too small), 2) a mixture of species or a high amount of deciduous tree crowns on the plots, causing irregularity in the crown structure. Based on our results, we can say that both factors have a role in the generation of error. The species homogeneity of the stand decreases the error to some extent, but then the lack of suitable single tree species stands available as neighbors may limit the success of estimation. Both large variation in stand conditions (lack of suitable neighbors) and high proportion of deciduous species caused great errors in the 0-100 m^3/ha volume class in Evo.

Running a separate feature selection for Kuortane would probably have improved the results, as the distinctive capacity of remotely sensed material was now optimized for the forests of Evo with higher mean volumes and an even distribution of Scots pine, Norway spruce and deciduous species. Maltamo et al. (2009) obtained somewhat better results using the same data from Kuortane: they resulted in a mean volume RMSE of 30%, and species-specific RMSEs 49, 141 and 177% for Scots pine, Norway spruce and deciduous species, respectively. In addition to the optimization of features for this specific data, other major differences to our approach were the use of the *k* most similar neighbor method and spectrally calibrated aerial photographs.

Conclusions

We conclude that even with remotely sensed data having high correlations with forest variables, the characteristics of the inventoried forest area and the ability of the field sample to describe the locally exotic strata within the forest area have a great impact on the success of estimation. In our case, stratified sampling in the Scots pine dominated Kuortane area and a general increase in the number of field plots in the heterogeneous Evo area would probably have improved the accuracies of rarer strata. Furthermore, the applied sub-set of features should be optimized for each target area. Running the feature selection separately for Kuortane will be our next step.

References

Goldberg, D.E., 1989. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Publishing Company, Reading, Massachusetts, 412 p.

Haara, A. and K. Korhonen. 2004. Kuvioittaisen arvioinnin luotettavuus. Metsätieteen aikakauskirja, 4/2004: 489-508.

Haralick, R.M., K. Shanmugan and I. Dinstein. 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, 3(6): 610-621.

Haralick, R. 1979. Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5): 786-804.

Holopainen, M., R. Haapanen, S. Tuominen, and R. Viitala. 2008. Performance of airborne laser scanning- and aerial photograph-based statistical and textural features in forest variable estimation. In Hill, R., Rossette, J. and Suárez, J. 2008. Silvilaser 2008 proceedings: 105-112.

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

Suvanto, A., M. Maltamo, P. Packalén, and J. Kangas. 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. Metsätieteen aikakauskirja, 4/2005: 413-428.

Wall, M. 1996. GAlib: A C++ Library of Genetic Algorithm Components Version 2.4 Documentation, revision B. Massachusetts Institute of Technology. 101 pp.