SCANDINAVIAN JOURNAL OF FOREST RESEARCH, 2017
VOL. 32, NO. 6, 488–500
http://dx.doi.org/10.1080/02827581.2016.1259425

Effects of field plot configurations on the uncertainties of ALS-assisted forest resource estimates

Erkki Tomppo [a,b], Nea Kuusinen[a], Kai Mäkisara [a], Matti Katila [a] and Ronald E. McRoberts[c]

[a]Forest Planning and Forest Resources, Natural Resources Institute Finland (Luke), Helsinki, Finland; [b]Natural Resources Institute Finland (Luke), Helsinki, Finland; [c]Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA

ABSTRACT
The effects of field plot configurations on the uncertainties of plot-level forest resource estimates were analyzed using airborne laser scanner data, aerial photographs and field measurements. The aim was to select a field sample plot configuration that can be used for both large area and management inventories. Error estimates were evaluated at the plot level using six different training plot configurations. Additionally, separate plots with two different sizes were used for evaluation. Stem volume and five other forest resource characteristics were considered. The field measurement costs of the different plot configurations were also studied. RMSEs and mean deviations for airborne laser scanning ALS-assisted estimates were practically the same for the fixed radius plot, the two concentric plots and the angle count plot with a basal area factor of $q = 1$ for all three evaluation plot sizes. Angle count plots with basal area factors of $q = 1.5$ and 2 increased the RMSEs. For the former plot configurations, the RMSEs for the ALS-assisted estimates could be attributed to inaccuracy in the predicted relationships between the field data and ALS data, not to the training plot configuration. Tree measurements and costs can, therefore, be reduced from those of the Finnish management inventories without increasing RMSEs.

ARTICLE HISTORY
Received 4 June 2016
Accepted 4 November 2016

KEYWORDS
Forest inventory; ALS; aerial photographs; plot configuration; genetic algorithm

Introduction

The primary sources of forest resource information are forest inventories of which there are two primary types: national forest inventories (NFI) and forest management inventories (FMI). NFIs have a strategic focus and provide data for national and regional policy-making, forest management planning, and forest industry investment planning. NFIs use statistically rigorous sample-based methods and have been conducted in the Nordic countries and the United States of America (USA) for nearly 100 years (Tomppo et al. 2010). FMIs are used for specific purposes such as stand-level and forest holding-level management planning for individual forests (e.g. Loetsch & Haller 1973). Because precise stand-level estimates are more important for FMIs than accurate large area estimates based on probability sampling strategies, greater sampling design flexibility, even with selection bias, is permitted for FMIs than for NFIs.

Remotely sensed data have been used to enhance forest inventory resource estimates since the 1940s (Barrett et al. 2016). Much of the recent emphasis has focused on the use of airborne laser scanning (ALS) data. For NFI applications, Næsset et al. (2011) used model-assisted estimators with ALS data as auxiliary information and found that the variance of the estimator of mean growing stock volume per unit area for a Norwegian study area was reduced by a factor of more than 6.0 relative to simple random sampling estimators that used no auxiliary information. For a similar study in Washington, USA, Strunk et al. (2012) reported a variance reduction of 5.0. For a second Norwegian study area, McRoberts et al. (2013) compared simple random sampling, stratified, model-assisted and model-based estimators for estimating mean growing stock volume per unit area using ALS data as auxiliary information. The stratified estimators reduced the variance by a factor of approximately 3.0, whereas the model-assisted estimators reduced the variance by factors greater than 5.5. Thus, there is ample evidence that ALS data can contribute substantially to increasing the precision of large area forest attribute estimators (Næsset et al. 2013).

For FMIs, Næsset (1997a, 1997b) developed the key methods. The first operational ALS-assisted FMIs were conducted in Norway in 2002 using an area-based method that has been continuously modified and improved (Næsset 2002, 2004). In Finland, the operational use of ALS data for FMIs began in 2010. The plan is to complete the first round throughout the entire country by 2019 (Maltamo et al. 2011; Maltamo & Packalén 2014; Næsset 2014). Predictions for stand-level forest characteristics are made using ALS data, aerial photographs and data from field sample plots established and measured solely for this inventory. Predictions are calculated for 16-m × 16-m cells that tessellate the area of interest and are aggregated for micro-stands and stands resulting from a segmentation algorithm. The primary purpose is to minimize expensive field work.

The key features of any forest inventory are the sampling design, the sampling intensity, and the plot configuration which refers to the plot's size, shape, and components (e.g. Ware & Cunia 1962; Mandallaz 2008; Gregoire and Valentine 2008). Multiple factors affect the efficiency of plot

CONTACT Erkki Tomppo erkki.tomppo@luke.fi Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland

configurations including the expected number of plot trees to be measured; the tree, plot, and stand variables of interest; the diversity of the forest; and remote sensing considerations. Efficient use of remotely sensed data places different demands on field plots than an inventory that uses field data only. Because the variation in the remote sensing variables should relate to variation in the field variables, the field plot data should describe the areal unit for which the remotely sensed data are obtained. In addition, when using ALS data, the plot should be large enough that the number of pulses per plot is sufficient to produce accurate ALS metrics (Vauhkonen et al. 2014b). Thus, the optimal field plot size for use with remotely sensed data may be greater than the field plot size for employing field data only. In addition, more detailed plot measurements may be required. Overall, optimization of a plot configuration requires careful consideration of both field and remote sensing factors.

ALS-assisted inventories have generally used fixed area plot configurations. However, multiple studies comparing the effects of size for fixed area and angle plot configurations have also been reported. Maltamo et al. (2007) compared estimation errors for both fixed area and truncated angle count plot configurations and concluded that the angle count plot performed well compared to the fixed area plot. Hollaus et al. (2007) reported root-mean-square errors (RMSEs) for ALS-assisted regression model predictions of stem volume when using angle count plots. Among radii in the range 18–26 m, the RMSE was smallest for a plot with a radius of 24 m. Scrinzi et al. (2015) reported that ALS-assisted estimates obtained using angle count plot configurations with three maximum distances compared well with estimates obtained using fixed area plots. Thus, despite a general inventory preference for fixed area plots, angle count plot configurations should not be too readily dismissed, either for large area inventories such as NFIs or for FMIs.

When both NFIs and ALS-assisted FMIs are conducted in the same country, as is the case in the Nordic countries, the financiers often ask whether the two systems or components of the two systems can be merged. Because of their different goals, rotation systems, and data requirements, complete merger of NFIs and FMIs is not a realistic option. Nevertheless, because field measurements account for a substantial proportion of the costs for both NFIs and FMIs, the degree to which a common set of field measurements obtained from a common plot configuration could be used for the two inventories merits consideration.

Use of the same field data for both NFIs and ALS-assisted FMIs is challenging because of the different inventory objectives and the different cost efficiency requirements for the sampling designs and field plot configurations. A well-known fact is that the number of trees measured per plot for large area forest inventories such as NFIs is quite small, often less than 10 trees per plot in boreal forests for which trees in close physical proximity are typically very similar (Tomppo et al. 2010). Conversely, for ALS-assisted FMIs, measurement of all trees, or at least all trees with diameters greater than a specified threshold, is often strongly advised. Another challenge is that many NFIs use panel designs for which the field plots in a particular inventory region may be measured in different years. Thus, data for field plots measured before or after the year the remotely sensed data are acquired might need to be updated or transformed to correspond to the year of the remotely sensed data acquisition. Despite the challenges, the potential efficiencies that could be realized from a common NFI and FMI plot configuration justify an investigation of the possibilities.

The overall objective was to determine if the same field plot configuration can be used for both NFIs and ALS-assisted FMIs. The main technical objective was to compare the effects of typical NFI plot configurations and typical ALS-assisted FMI plot configurations on the uncertainty of the ALS-assisted forest resource estimates. A second technical objective was to compare different field plot configurations with respect to the errors in field data-based estimates. The error criteria are the mean deviations and RMSEs when the estimates are compared to the field data-based results on the plot configuration currently used in the ALS-assisted inventories by the Finnish Forest Centre (Section 'The plot alternatives'). These results are used for comparisons and assumed for this study to have only negligible errors. The key FMI characteristics for which estimates were obtained were stem volume, stem volume by tree species group, basal area, mean diameter and mean height of trees and the number of trees per hectare. The analyses included assessments of the costs of obtaining tree measurements for the different plot configurations. Special emphasis was placed on estimation for mixed and dense forests for which ALS-assisted estimation errors are known to be large (Packalén & Maltamo 2007; Maltamo et al. 2015). In the following, ALS-assisted forest inventory means an inventory method that uses ALS data and aerial photographs, in addition to field measurements (Sections 'Materials and methods' and 'ik-nn estimation').

The novel features of the study are investigation of the problems in an operational ALS-assisted setting with advanced multivariate estimation as well as multifaceted error analyses with a special emphasis on forests in which ALS-assisted approaches often fail to produce precise estimates.

Materials and methods

The field data

Study site and plot measurements

A total of 2468 field plots was measured in 2013 in a study site in Central Finland. The land area was 570,000 hectares, of which 431,000 hectares were forestry land. Forestry land here includes forest land, poorly productive forest land and unproductive land as defined in the Finnish national forest inventory and management inventories. The combined area of the first two categories (423,000 hectares) corresponds to the area based on FAO forest definition (Tomppo et al. 2011). The area of unproductive land was thus 8000 hectares.

Systematic sampling with L-shaped clusters and eight plots per cluster was used. The distance between the clusters was 4.3 km and the distance between the plots in a cluster was 250 m. Of the plots, 1867 were either on forest land, poorly productive forest land or unproductive land. Of these, 1165 plots belonged to the development classes (maturity

Table 1. The average values and standard deviations (SD) of the forest variables calculated with (a) trees dbh > 0 on the plots in the maturity classes young thinning stand – mature stand, n = 1150, (b) as (a), but only trees with dbh ≥ 4.5 cm included.

	All trees		dbh ≥ 4.5 cm	
	Average	SD	Average	SD
Volume, all species, m^3/ha	156.20	86.72	154.36	87.00
Pine vol, m^3/ha	98.00	70.57	97.78	70.67
Spruce vol, m^3/ha	37.02	73.73	36.54	73.68
Broadl. vol, m^3/ha	21.16	34.15	20.04	33.48
Mean diameter, cm	18.35	5.35	18.74	5.24
Mean height, dm	151.15	40.74	153.99	39.90
Basal area, m^2/ha	20.03	7.96	19.49	7.91
Number of stems/ha	3867.80	4373.83	1157.71	626.33

classes) young thinning stand, advanced thinning stand or mature stand (denoted here by 4, 5 and 6, respectively) and were entirely within a single stand, that is, did not intersect two or more stands. Fifteen of these 1165 plots included errors in the field data and were removed. Thus, 1150 plots remained for the final analyses. One plot from this data set is used in turn as validation data when using leave-one-out cross validation. All these plots are used as the training data when a separate validation data set is used (Section 'Additional validation data measured in 2014'). The average values of the forest variables studied on the 1150 plots are given in Table 1.

All trees with diameter at breast height (dbh, in our case, diameter at the height of 1.3 m) ≥ 4.5 cm were calipered for a fixed radius plot of 9 m. An additional angle count sample plot (Bitterlich plot) with a basal area factor of q = 1.5 was measured on each plot. The maximum distance of the angle count plot was also 9 m.

Tree species, dbh, tree class, distance from the center point of a plot to the tree as well as the azimuth of the tree were recorded for all tally trees. The distance and azimuth were measured simultaneously with the Masser Sonar caliper. These measurements made it possible to select several types of samples from the recorded plot data and to construct several plot configurations after the measurements were taken. Additionally, the age and height of the trees were recorded for the basal area median trees by tree species groups. The age is determined either using increment borer or by counting the knot ringlets when possible. The age at the breast height must be added (from tables) when using boring. These trees were selected by a program in the data logger after tree calipering. Additionally, a small set of stand-level characteristics were measured from each stand intersecting a plot.

The time needed for different measurement phases was recorded by the data loggers. The phases were (a) the initial preparations, including installation of GPS equipments, (b) tally tree measurements, (c) sample tree measurements and (d) stand-level measurements. These recordings made it possible to assess the measurement costs for the different plot configurations. Only time consumption from tally tree measurements was taken from each plot. For the other work on a plot, the averages of the time consumptions from the Finnish NFI12 field plots were used. The plots on forest land including tally trees were used for averages.

The simple assumptions made after the time data checks were, in addition to tally tree measurements by one person (Table 7): (1) the initial preparations, including installation of GPS equipments, took five minutes for two persons, (2) sample tree measurements took 15 minutes for one person and (3) plot and stand-level measurements took 17.5 minutes for one person. The time needed to complete all measurements was summed up and divided by two as the crew consisted of two persons. It was assumed that the crew members could flexibly change their tasks during the measurement of the plot.

The total number of the measured trees on 1150 plots was 33,879 and the number of the median height trees 2685.

In the ALS-assisted management inventory, the heights are estimated first for each tally tree after which the volumes are estimated for each tree using taper curve models by Laasasenaho (1982). The heights of the sample trees are used for height model calibration (Eerikäinen 2009). This method was used in the volume estimation of this study.

The plot alternatives

A fixed radius plot with a radius of 9 m is used in the Finnish ALS-assisted management inventory. All trees with dbh ≥ 4.5 cm are measured. In the Finnish NFI, a new plot configuration was introduced in 2014 to replace the earlier truncated angle count plot (Tomppo et al. 2011). The new plot consists of two concentric circles: trees with dbh 4.5–9.4 cm are measured on a circle with a radius of 5.64 m, whereas the radius for the thicker trees is 9 m. The new plot also includes angle count sampling of trees with dbh < 4.5 cm and with a basal area factor q = 1.5.

The plot configurations and their abbreviations used in this study are:

SMK: All trees with dbh ≥ 4.5 cm are measured on a 9 m fixed radius circular plot. The plot is used by the Finnish Forest Centre (SMK).

k75: Trees with dbh 4.5–7.4 cm are measured on a fixed radius plot of 5.64 m and trees with dbh ≥ 7.5 cm on a 9 m radius plot.

k95: Trees with dbh 4.5–9.4 cm are measured on a fixed radius plot of 5.64 m and trees with dbh ≥ 9.5 cm on a 9 m radius plot.

q1: Trees are selected on an angle count plot with a basal area factor of q = 1.0 and with a maximum distance from the plot center of 9 m.

q1.5: Trees are measured on an angle count plot with a basal area factor of q = 1.5 and a maximum distance from the plot center of 9 m.

q2: Trees are measured on an angle count plot with basal area factor of q = 2.0 and a maximum distance from the plot center of 9 m.

In the resulting calculations, the angle count plots ignored trees with dbh < 4.5 cm so that inclusion of smaller trees would not introduce bias into the estimators in the comparisons.

Results were not computed for basal area factors q larger than 2.0 because these factors are not feasible in the Finnish NFI, and even less for the ALS-assisted FMI.

Additional validation data measured in 2014

An additional data set was measured in 2014 and used as validation data. The data consist of observations and measurements for 30 field plots with a size of 32 m × 32 m. These plots are subdivided into four subplots each of size 16 m × 16 m. The locations of these plots were selected subjectively from forests, in which estimation with ALS usually leads to large RMSEs. Large RMSEs have been noticed in practice, e.g. in the Finnish ALS-assisted FMI, in dense forests, possibly with several tree layers, see also Næsset (2014), Maltamo et al. (2014) and Maltamo et al. (2015). Each 32 m × 32 m plot was entirely within one forest stand and belonged to one of the three development classes young thinning stand, advanced thinning stand or mature stand.

The plot locations were selected to align with the grid used in the management inventory of the Finnish Forest Centre (SMK). A two-phase procedure using post-processed Global Navigation Satellite System (GNSS) observations was used to position the plots as close as possible to the planned locations. The dbh, as well as the distance and direction from the center of the subplot to the tree, was measured for every tree with dbh $\geq$ 2.5 cm. Tree height and age were additionally measured from sample trees. On average, 37 sample trees per 32 m × 32 m plot were chosen based on the cumulative dbh. Additional sample trees were selected from basal area median trees by tree species groups as for the reference data (Section 'Study site and plot measurements'). The purpose of the large number of the sample trees was to estimate the plot-level forest variables accurately.

Calculation of the plot-level estimates based on field data

The variables were estimated in all cases using only trees with dbh $\geq$ 4.5 cm following the practice of the Finnish ALS-aided management inventory of the SMK. Note that in the Finnish NFI, all trees with heights greater than 1.3 meters are included in the volume, basal area and diameter estimates.

The inclusion probability of a tree for each plot configuration depended on the radius of the circle from which the tree was measured and thus on the dbh of the tree. This was taken into account in the estimators of forest variables for each plot configuration, see Tomppo et al. (2011). The estimators are thus unbiased, also for the truncated angle count plots. The mean diameter and the mean height of trees were defined as the basal area weighted means of the tree-level variables.

The ALS data

The ALS data used were a part of the operative management inventory data by the Finnish Forest Centre. The data were acquired by Blom Kartta Oy, Finland between 28 June and 27 August 2013 using the Piper Navajo airplane and the Optech Gemini ALTM scanner. The main parameters were: the flight altitude 1730 m, strip overlap 20%, pulse frequency, 70,000 Hz, scanning frequency 37 Hz, half scan angle 20 degrees, pulse density 0.89/m^2, the maximum number of observed pulse returns 4.

The digital elevation model (DEM) and the digital surface model (DSM) were derived from the ALS data. Pulse returns with heights greater than 1.3 m above the ground were classified as canopy.

The ALS feature variables were calculated for the squares of 16 m × 16 m around the center of each plot following the practice of the management inventory. The DEM and DSM were calculated for the squares of 8 m × 8 m and the values from the square closest to the plot centre were used.

The features were:

1	The height of the canopy, i.e. the difference between surface model and elevation model
2	Minimum height of the first returns
3	Maximum height of the first returns
4	Mean height of the first returns
5	Standard deviation of heights of the first returns
6	Skewness of heights of the first returns
7	Kurtosis of heights of the first returns
8	Width of range of heights of the first returns
9–15	The features similar to 2–8 for the last returns
16	Proportion of canopy returns, all pulse returns
17–27	Percentiles (5, 10, 20, … , 90, 95%) for the first returns
28–38	The same features as 17–27 for the last returns
39–49	The cumulative proportions of the foliage returns 0–5, 5–10, 10–20, … , 90–95% in the range of the height of the first returns
50–60	The same features as 39–49 for the last returns
61–62	Mean intensity (first and last returns)

The aerial photographs

Aerial imaging data were obtained from three flights on 12 July, 27 July, and 25 August in 2015 with the Microsoft UltraCam Eagle camera. The imaging altitude was 4700 m and the pixel size 0.3 m. The along track overlap was 80% and across track overlap was 35%. The sun elevation varied between 26 and 46 degrees. Three spectral bands, green, red and near infrared (NIR) were used for computing features NIR, Red and Green means as well as their standard deviations in a 16 m × 16 m square (53 × 53 pixels). Furthermore, the Haralick texture features 1–13 based on the co-occurrence matrix and the lag of 19 pixels were calculated separately for the NIR, red and green bands of the aerial photographs, in total 39 features (Haralick et al. 1973).

The total number of the feature variables, the potential explanatory variables to be used in estimation was thus 107 (Section ik-nn estimation).

General design of the study

The objective was to compare the RMSEs and mean deviations of the field plot alternatives in a realistic setting of an ALS-assisted inventory. One principle in selecting the estimation methods was to use a method that produces error estimates that are typical in operational ALS-assisted management inventories. We followed therefore the procedure used in the Finnish ALS-assisted FMI. The forest variables were measured on the circular plots with a maximum radius of 9 m and the ALS features were calculated for the squares of 16 m × 16 m, the center points coinciding with the center points of the field sample plots. These squares are the prediction units in the Finnish FMI. This procedure may have introduced noise but we have treated the plot alternatives in the same way, and in line with the objectives of this study. Furthermore, the plots that intersected several forest stands or land classes were removed following again the practice of ALS-assisted FMI.

The plot-level results are presented in two different ways: (1) the plot-level estimates for different plot configurations are compared against the estimates based on the field data on the SMK plots using RMSEs, and mean deviations as criteria (2) the errors, RMSEs and mean deviations, in the ALS-assisted estimates obtained with each plot configuration as a source of training data are compared against the similar errors obtained when using the SMK data as the training data. The estimates in the second-type analyses are always compared to the field-data-based estimates obtained with the SMK plot configuration. The first type of comparisons are often used to justify a certain plot configuration, and to compare the plot-level errors, arising from the plot configuration, to the errors arising from ALS-assisted estimation. The second kind of comparison is more relevant to the ALS data users. If another field plot configuration yields ALS-assisted estimates consistently similar to those obtained with plot configuration SMK and ALS data, both plot configurations are equally usable.

Eight forest inventory variables were selected for the analyses: (1) stem volume per hectare, (2)–(4) stem volumes per hectare of the three tree species groups, pine (*Pinus sylvestris* L.), spruce (*Picea abies* (L.) Karst.) and broad-leaved trees, mainly birch (*Betula* spp.) with some aspen (*Populus tremula* L.) and alder (*Alnus* spp.), (5) mean diameter of trees (the basal area of a tree as the weight), (6) mean height (the basal area of a tree as the weight), (7) basal area and (8) the number of stems per hectare.

The following questions were addressed:

(1) What are the errors of the ALS-assisted estimates? The mean deviations and RMSEs were used as the criteria (Equations (5) and (6)). We do not consider specifically the effect of the model prediction errors or measurement errors on the estimates. Both types of errors are indirectly included in the error estimates, but do not affect the comparisons of the RMSEs and mean deviations of the estimates with different plot configurations.

    The error quantities were calculated using two different procedures,
    (a) leave-one-out method using the data measured in 2013;
    (b) using the data set measured in 2014 as the validation data; The data measured in 2013 were used as the training data.
(2) What are the differences in the errors of the field data-based estimates caused by the different plot configurations compared to the differences in the errors caused by the ALS-assisted estimation?
(3) What are the measurement times and costs for the different plot configurations?

The estimates to be validated were computed using the selected field plot configuration as the training data in the ALS-assisted estimation and the field data-based estimates on the plot configuration when using field data alone. The results are always compared to the field data-based estimates on the SMK plots. The resulting statistics were computed for all plots (n = 1150) and for several subgroups: plots with the volume of 0–49 m^3/ha (n = 88), ≥ 200 m^3/ha (n = 285) and plots with mean dbh ≤ 11 cm (n = 70). The last subgroup was included to test specifically whether the decrease in the number of thin trees in the training data, caused by either the use of the concentric plots or angle count plots, increased the ALS-assisted estimation errors. The validation of the estimates from practical ALS-assisted FMIs has shown that the errors of the estimates and particularly the estimates by tree species are larger in young, dense, mixed forests, than in older and more homogenous forests, cf. Packalén and Maltamo (2007) and Maltamo et al. (2015). Young, dense, mixed forests were intentionally included in the validation data set measured in 2014. Note that the training data from year 2013 represent "average forests" in the region in question.

ik-nn estimation

The well-known k-nn estimation method was employed in ALS-estimation. The weights for the features selected were calculated using genetic algorithm and leave-one-out cross validation (Tomppo and Halme 2004; Tomppo et al. 2014). This k-nn method is called the ik-nn method here.

Let's recall the main features of the ik-nn estimation with the genetic algorithm in the feature weighting. Denote the k nearest feasible field plots by $i_1(p), \ldots, i_k(p)$. The weight $w_{i,p}$ of field plot i to pixel p is defined as

$$w_{i,p} = \frac{1}{d^t_{p_i,p}} \Big/ \sum_{j \in \{i_1(p),\ldots,i_k(p)\}} \frac{1}{d^t_{p_j,p}}, \quad \text{if and only if } i \in \{i_1(p), \ldots, i_k(p)\} \qquad (1)$$
$$= 0 \quad \text{otherwise.}$$

The value of k was fixed to be five after preliminary tests using the RMSEs and mean deviations as the criteria. The distance weighting power t is a real number, usually $t \in [0, 2]$. The value t = 1 was used here. A small quantity, greater than zero, is added to d when d = 0 and $i \in \{i_1(p), \ldots, i_k(p)\}$.

The distance metric d employed was

$$d^2_{p_j,p} = \sum_{l=1}^{n_f+n_g+n_h} \omega_l{}^2 (f_{l,p_j} - f_{l,p})^2, \qquad (2)$$

where $f_{l,p}$ is the lth feature variable for pixel p, n_f the number of aerial image feature variables, n_g the number of the ALS feature variables, n_h the number of the Haralick feature variables and ω_l the weight for the lth feature variable.

The values of the elements ω_l of the weight vector ω were selected with a genetic algorithm (Tomppo and Halme 2004; Tomppo et al. 2008).

The fitness function to be minimized for the optimization of the distance metric and calculating the values of the ω

vector was

$$f(\omega, \gamma_\delta, \gamma_e, \hat{\delta}, \hat{\bar{e}}) = \sum_{j=1}^{n_e} \gamma_{\delta,j} \hat{\delta}_j(\omega) + \sum_{j=1}^{n_e} \gamma_{e,j} \hat{\bar{e}}_j(\omega), \qquad (3)$$

where $\hat{\delta}_j(\omega)$ is the standard error of the estimate of the forest variable i, $\hat{\bar{e}}_j(\omega)$ is the mean deviation of the forest variable i, n_e the number of the forest variables used in the algorithm (here eight) and γ_δ and γ_e are fixed constant vectors.

The values of the vector ω were calculated using the plot configuration SMK. The same values were used with all the other plot configurations to investigate the effect of the plot configuration only on the error estimates and to keep the comparability of the estimates based on different plot configurations. We also tested the weights optimized by the plot configurations. With the procedure selected, the purpose was to be conservative in the sense that any of the plot configurations tested (k75, k95, q1, q1.5, q2) were not favored over the fixed radius plot.

After some tests with the principle of trial and error and heuristic tradeoffs between standard errors and mean deviations of different forest variables, the vectors γ_δ and γ_e were fixed to be both $\gamma_\delta = \gamma_e = (4, 4, 4, 2, 1, 1, 1, 0)$. The plot-level estimate of the variable "number of the trees per hectare" varied with each plot configuration so much that a value of zero was given to the weights of that variable in the optimization to avoid its adverse effect on the optimization. Otherwise, the parameters of the genetic algorithm itself were the same as in Tomppo and Halme (2004).

A pixel-level estimate of variable Y for pixel p is defined as

$$\tilde{y}_p = \sum_{i \in I_p} w_{i,p} y_i, \qquad (4)$$

where y_i is the value of the variable Y on plot i and I_p the set of the field plots used in calculating the estimates for pixel p. I_p may be the same for all pixels or depend on a possible stratification. Note that the set of the positive values of $w_{i,p}$ in I_p vary by pixel p.

The ik-nn estimates ($\hat{y}_{\text{ik-nn},a}$) based on each plot configuration a were compared to the plot-level values of the variables based on the field data on the plot configuration SMK, denoted by $\hat{y}_{f,\text{SMK}}$.

The mean deviation

$$\widehat{\text{meandev}}_a = \sum_{i=1}^{n} (\hat{y}_{i,\text{ik-nn},a} - \hat{y}_{i,f,\text{SMK}})/n$$

and RMSE

$$\widehat{\text{RMSE}}_a = \sqrt{\sum_{i=1}^{n} (\hat{y}_{i,\text{ik-nn},a} - \hat{y}_{i,f,\text{SMK}})^2/n}$$

were employed in assessing the goodness of the ALS-estimates, n being the number of plots. Similar quantities were employed when evaluating the field data-based estimates. In this case, $\hat{y}_{i,\text{ik-nn},a}$ is replaced by $\hat{y}_{i,f,a}$ where f refers to field data. Recall that the SMK, k75, k95, q1, q1.5 and q2 plot configurations were studied.

Selection of the features and validation of the estimates

It turned out that it was difficult to find a stable solution near the global optimum with the genetic algorithm and 107 feature variables. Reduction of the feature variables was necessary. The Lasso method (Tibshirani 1996) was first tested, as well as a stepwise regression analysis, but the final selection was based on a simple correlation analysis. Note that stepwise methods are known to perform poorly when the feature variables are highly correlated (Harrel 2001). The purpose was only to eliminate the irrelevant feature variables (the variables without any explanation power). The final weighting of the feature variables was determined with the genetic algorithm. As a heuristic, non-parametric method, it does not presume any distributional properties from the variables.

Three aerial photograph feature variables, 27 ALS feature variables, including the difference between the surface model and elevation model, and six Haralick feature variables based on aerial photographs were selected for genetic algorithm optimization. The feature variables were as follows:

(1) Aerial photograph feature variables NIR, red and green means, calculated from 53 × 53 pixels in a 16 m × 16 m square.
(2) ALS features, the height of the canopy, i.e. difference between surface model and elevation model, maximum height of first returns, mean height of first returns, maximum height of last returns, mean height of last returns, standard deviation of heights of last returns, proportion of canopy returns, all pulse returns, 20, 30, …, 90, 95% percentile for the first return, 20, 30, …, 90, 95% percentile for the last return, the proportions of the returns in 30–40% of the range of the heights of the first return pulses, the proportions of the returns in 40–50% of the range of the heights of the first return pulses,
(3) Haralick features, Sum Average, Haralick features 6, for NIR, red and green, as well as Sum Entropy, Haralick features 8, for NIR, red and green.

Results

The effect of the plot configuration on the estimates

The field data-based estimates for forest variables of the tested plot configurations ($\hat{y}_{f,a}$) deviated from the field data-based estimates of SMK plot ($\hat{y}_{f,\text{SMK}}$) least on the k75 plots, slightly more on the k95 plots and clearly more on the angle count plots (RMSE$_f$ and mean deviation$_f$, Tables 2–5). For the angle count plots, the differences increased when the basal area factor increased. For example, the field data-based estimate of the stem volume of all species of the k75 plots deviated 1.9–3.4 m^3/ha from that on the SMK plots depending on the group, whereas the difference was 3.3–6.7 m^3/ha with the k95 plots, 8.6–12.1 m^3/ha with the q1 plots, 10.8–21 m^3/ha with the q1.5 plots and 13.7–28.1 m^3/ha with the q2 plots. The differences between the field data-based estimates of the k75 and k95

Table 2. Comparisons of the field data-based forest variable estimates to those calculated with the SMK plots ($RMSE_f$, and mean deviation$_f$), as well as comparisons of the ALS-assisted forest variable estimates calculated with leave-one-out cross validation to the field-based estimates on the SMK plots ($RMSE_{ik-nn}$ and mean deviation$_{ik-nn}$). The upper row for each variable shows the RMSE and the lower row the mean deviation. All plots, n = 1150.

	$RMSE_f$ and mean deviation$_f$					$RMSE_{ik-nn}$ and mean deviation$_{ik-nn}$					
	k75	k95	q1	q1.5	q2	SMK	k75	k95	q1	q1.5	q2
Volume, m^3/ha	1.9	3.6	11.4	18.8	25.6	40.1	40.1	40.1	40.3	41.2	41.8
	−0.1	−0.0	0.3	0.3	0.3	0.2	0.2	0.2	0.5	0.9	1.4
Pine, vol m^3/ha	0.9	1.9	7.4	13.3	19.9	47.3	47.3	47.3	47.3	47.6	48.0
	0.0	0.0	0.3	0.7	0.8	1.5	1.5	1.5	1.8	2.6	2.9
Spruce, vol m^3/ha	1.0	1.8	5.5	9.1	12.9	44.5	44.6	44.6	44.7	44.9	45.0
	−0.0	−0.0	−0.3	−0.6	−0.8	−0.3	−0.3	−0.3	−0.6	−0.9	−0.9
Broadl, vol m^3/ha	1.2	2.4	6.5	9.9	11.9	20.7	20.7	20.8	20.9	21.1	21.4
	−0.1	−0.0	0.3	0.2	0.2	−1.0	−1.0	−1.0	−0.7	−0.8	−0.7
Mean dbh, cm	0.3	0.4	0.9	1.2	1.4	2.7	2.7	2.7	2.7	2.7	2.7
	0.0	0.0	0.1	0.2	0.3	−0.2	−0.2	−0.2	−0.1	−0.0	0.0
Mean height, dm	1.6	2.3	5.0	6.5	7.9	13.1	13.1	13.1	13.2	13.3	13.4
	0.1	0.2	0.7	1.0	1.4	−1.2	−1.1	−1.1	−0.6	−0.3	−0.0
Basal area, m^2/ha	0.5	0.8	2.0	3.0	3.9	4.7	4.7	4.7	4.7	4.9	4.9
	−0.0	−0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.3	0.3
Number of trees/ha	157.6	196.6	422.8	541.3	621.3	479.2	488.7	492.2	517.8	540.4	556.7
	−5.4	−4.2	2.5	−3.0	−13.5	10.5	8.7	9.0	14.6	13.7	8.2

plots from the SMK plot were the largest for the group with mean dbh ≤ 11 cm.

Contrary to the differences between field plot configurations in the field data-based estimates, the errors in the ALS-assisted estimates were similar for all tested field plot configurations ($RMSE_{ik-nn}$ and mean deviation$_{ik-nn}$, Tables 2–5). The RMSEs and mean deviations for the SMK plot and the k75, k95 and q1 plots were practically same. The errors increased slightly when the basal area factor for the angle count plot was increased from q = 1. This applies to the results for all subgroups.

The volume estimates for the plots with mean dbh ≤ 11 cm were studied in detail for tree species pine and spruce. The ALS-assisted estimates of pine and spruce stem volume with the SMK and k75 plots were very similar, but the differences between these estimates and the field data-based estimates on SMK plots were much larger. Each point in Figure 1 represents the estimates for both pine and spruce. The points corresponding to the two ALS-assisted estimates (training field plot configurations SMK and k75) and the field data-based estimate for the plot configuration SMK are connected with lines. Figure 2 shows that the differences between the ALS-assisted estimates are larger when plot configuration k95 is used as training data instead of plot configuration k75. The differences between the estimates are still much smaller than the estimation errors. The error patterns by tree species are almost identical.

The ALS-assisted estimates of the mean volume (all tree species together) based on the k95 plot and on the SMK plot in the group with mean dbh ≤ 11 cm are also similar (Figure 3). The estimates on one plot are connected by a line. The two ALS-assisted estimates deviated somewhat more when using the SMK plot and q1 plot as the training data (Figure 4).

The ik-nn method underestimated the large values and overestimated the small values of the forest variables. This can be seen from the positive conditional mean deviations of the estimates on the plots with small volumes (Tables 3 and 5) and negative conditional mean deviations in the large volume group (Table 4).

Table 3. Comparisons of the field data-based forest variable estimates to those calculated with the SMK plots ($RMSE_f$, and mean deviation$_f$), as well as comparisons of the ALS-assisted forest variable estimates calculated with leave-one-out cross validation to the field-based estimates on the SMK plots ($RMSE_{ik-nn}$ and mean deviation$_{ik-nn}$).

	$RMSE_f$ and mean deviation$_f$					$RMSE_{ik-nn}$ and mean deviation$_{ik-nn}$					
	k75	k95	q1	q1.5	q2	SMK	k75	k95	q1	q1.5	q2
Volume, m^3/ha	2.1	3.4	8.6	10.8	13.7	21.9	22.0	22.2	23.5	24.4	25.5
	−0.0	−0.2	−1.0	−1.2	−1.8	12.6	12.7	12.6	13.4	13.7	14.0
Pine, vol m^3/ha	1.5	2.8	7.3	9.6	12.4	21.4	21.7	21.8	22.5	23.7	24.0
	−0.0	−0.4	−0.5	−0.6	−0.9	11.8	12.0	11.8	12.4	12.7	13.1
Spruce, vol m^3/ha	0.8	1.3	3.5	4.1	4.1	10.5	10.6	10.5	10.7	10.6	11.0
	0.1	0.2	−0.3	−0.5	−0.7	0.1	0.1	0.2	0.1	−0.2	−0.1
Broadl. vol m^3/ha	0.9	1.9	4.4	5.2	6.0	8.2	8.2	8.3	9.1	8.9	10.0
	−0.1	−0.0	−0.2	−0.2	−0.1	0.8	0.6	0.6	0.9	1.1	1.0
Mean dbh, cm	0.4	0.6	1.4	1.7	2.1	1.9	1.9	1.9	2.1	2.2	2.3
	0.1	0.1	0.3	0.6	0.8	0.3	0.3	0.3	0.5	0.7	0.8
Mean height, dm	2.2	3.0	6.6	8.5	10.7	11.8	11.7	11.6	12.4	12.9	12.9
	0.2	0.2	1.0	2.4	3.5	2.3	2.2	2.2	3.1	4.2	4.7
Basal area, m^2/ha	0.6	0.9	1.9	2.4	3.0	3.7	3.8	3.8	4.1	4.3	4.5
	0.0	−0.1	−0.2	−0.3	−0.5	2.1	2.1	2.1	2.3	2.3	2.3
Number of trees/ha	211.0	230.5	440.0	595.0	654.0	513.9	518.3	530.0	625.1	668.6	676.7
	2.8	−7.5	−18.9	−70.3	−119.0	194.7	215.3	213.0	250.2	219.2	199.3

Notes: The upper row for each variable shows the RMSE and the lower row the mean deviation. Plots with total volume 0–49, n = 88.

Table 4. Comparisons of the field data-based forest variable estimates to those calculated with the SMK plots ($RMSE_f$, and mean $deviation_f$), as well as comparisons of the ALS-assisted forest variable estimates calculated with leave-one-out cross validation to the field-based estimates on the SMK plots ($RMSE_{ik-nn}$ and mean $deviation_{ik-nn}$).

	$RMSE_f$ and mean $deviation_f$					$RMSE_{ik-nn}$ and mean $deviation_{ik-nn}$					
	k75	k95	q1	q1.5	q2	SMK	k75	k95	q1	q1.5	q2
Volume, m³/ha	2.0	3.3	11.7	21.0	28.1	59.7	59.8	59.7	59.9	60.6	61.3
	−0.3	−0.1	−0.3	−1.3	−1.4	−26.6	−26.8	−26.7	−27.1	−27.9	−25.7
Pine vol. m³/ha	0.4	1.0	5.8	11.9	18.4	71.2	71.2	71.2	71.2	71.2	71.3
	0.0	0.0	−0.0	0.1	−0.5	−15.6	−15.6	−15.6	−15.5	−14.6	−14.0
Spruce vol m³/ha	1.3	2.2	7.1	13.0	18.2	76.4	76.4	76.5	76.7	77.0	77.2
	−0.1	−0.0	−0.1	−1.1	−0.6	−11.4	−11.5	−11.4	−11.9	−13.5	−12.5
Broadl. vol m³/ha	1.3	2.2	6.1	10.0	12.4	30.0	30.0	30.0	29.9	30.5	30.5
	−0.2	−0.1	−0.1	−0.3	−0.2	0.4	0.3	0.3	0.3	0.2	0.8
Mean dbh, cm	0.2	0.3	0.8	1.0	1.2	3.5	3.5	3.5	3.4	3.5	3.5
	0.0	0.0	0.1	0.2	0.2	−0.6	−0.6	−0.6	−0.5	−0.4	−0.4
Mean height, dm	1.5	2.1	4.9	6.4	7.7	15.4	15.5	15.5	15.3	15.2	15.1
	0.2	0.1	0.7	0.7	0.9	−3.9	−3.7	−3.7	−3.0	−2.9	−2.5
Basal area, m²/ha	0.4	0.7	1.9	3.0	3.7	6.3	6.3	6.3	6.4	6.4	6.5
	−0.1	−0.0	−0.1	−0.2	−0.2	−2.7	−2.7	−2.7	−2.8	−2.9	−2.7
Number of trees/ha	155.0	179.0	391.5	502.2	569.7	503.5	515.4	516.7	526.3	536.4	545.0
	−24.6	−21.2	−31.8	−35.0	−39.5	−116.7	−139.6	−138.2	−163.5	−177.5	−178.0

Notes: The upper row for each variable shows the RMSE and the lower row the mean deviation. Plots with total volume ≥ 200, n = 285.

Table 5. Comparisons of the field data-based forest variable estimates to those calculated with the SMK plots ($RMSE_f$, and mean $deviation_f$), as well as comparisons of the ALS-assisted forest variable estimates calculated with leave-one-out cross validation to the field-based estimates on the SMK plots ($RMSE_{ik-nn}$ and mean $deviation_{ik-nn}$).

	$RMSE_f$ and mean $deviation_f$				$RMSE_{ik-nn}$ and mean $deviation_{ik-nn}$						
	k75	k95	q1	q1.5	q2	SMK	k75	k95	q1	q1.5	q2
Volume, m³/ha	3.4	6.7	12.1	17.3	23.1	16.0	16.0	16.4	16.9	18.2	18.7
	0.7	0.9	2.4	2.8	2.1	7.0	7.3	7.0	8.5	9.4	9.5
Pine, vol m³/ha	2.0	3.8	7.5	10.9	12.8	21.8	21.9	22.0	22.1	22.6	23.0
	0.2	0.1	1.2	1.1	0.2	6.4	6.6	6.4	6.8	7.5	7.9
Spruce, vol m³/ha	1.3	2.1	3.3	5.5	7.2	13.0	13.0	13.1	13.3	13.7	13.5
	0.2	0.2	0.4	0.6	0.4	0.5	0.6	0.8	0.7	0.7	0.1
Broadl. vol m³/ha	2.4	5.0	10.1	13.1	16.3	11.7	11.6	11.2	11.7	12.1	12.3
	0.3	0.6	0.8	1.1	1.5	0.1	0.0	−0.1	1.1	1.2	1.4
Mean dbh, cm	0.3	0.4	0.9	1.2	1.5	2.2	2.1	2.2	2.4	2.5	2.6
	−0.0	−0.0	0.0	0.2	0.2	1.7	1.7	1.7	1.8	1.8	2.0
Mean height, dm	1.8	2.2	5.2	6.9	9.4	12.4	12.3	12.4	13.3	14.1	14.6
	−0.3	−0.4	−0.5	0.6	0.7	8.3	8.1	8.1	8.5	9.1	9.6
Basal area, m²/ha	0.9	1.5	2.6	3.7	4.6	2.9	2.9	3.0	3.1	3.3	3.4
	0.2	0.2	0.5	0.6	0.4	0.6	0.7	0.6	0.9	1.1	1.1
Number of trees/ha	290.9	369.0	682.6	928.7	1075.6	749.2	746.3	760.4	773.0	808.1	817.7
	64.2	72.1	160.3	148.0	109.3	−306.8	−272.9	−281.4	−197.0	−191.3	−207.4

Notes: The upper row for each variable shows the RMSE and the lower row the mean deviation. Plots with mean dbh ≤ 11 cm, n = 70.

The errors for volume groups closer to the overall mean volume of the 1150 plots (156.2 m³) were smaller than those presented and discussed here.

The errors of estimates with the validation data from 2014

Using the plots measured in 2014, it was possible to assess how the RMSEs decrease, that is, how the variance part of the RMSEs decrease when the area of the target unit increases. Using data from the 2013 plots as the training data with the configurations described in Section 'The plot alternatives' and the 32 m × 32 m plots measured in 2014 as the target units, reduced the RMSEs of the ALS-assisted estimates compared to the case where 16 m × 16 m plots were tested. The RMSE of the mean volume of all species decreased from about 37–39 m³/ha to 25–26 m³/ha (15–30%) (Table 6). The different training plot alternatives produced similar estimates.

The mean deviations (the other part of the RMSEs) were large when the 2013 field plot data were used as the training data to estimate forest variables for the 16 m × 16 m and 32 m × 32 m plots measured in 2014. For example, the mean deviations for mean volumes varied between −8 and 10 m³/ha for species together depending on the training plot configuration, between 16–17 m³/ha for pine and around 20 m³/ha for spruce for the both evaluation plots. Note that the mean deviation does not decrease when the plot size increases when calculating over the same area.

The effect of the plot configurations on the costs

The average time needed for the tally tree measurements on one plot for the plot configurations SMK, k75, k95 and q1 were 23.0, 19.3, 17.4 and 11.6 minutes, respectively (Table 7). Although the time savings in tree measurements are significant, the total savings are not that big due to the time needed for assessing and measuring plot-level and stand-level variables. The average measurement costs of one SMK plot including tree and plot-level variables were estimated to be 35.72 euros and time needed for two persons 32.8 minutes.

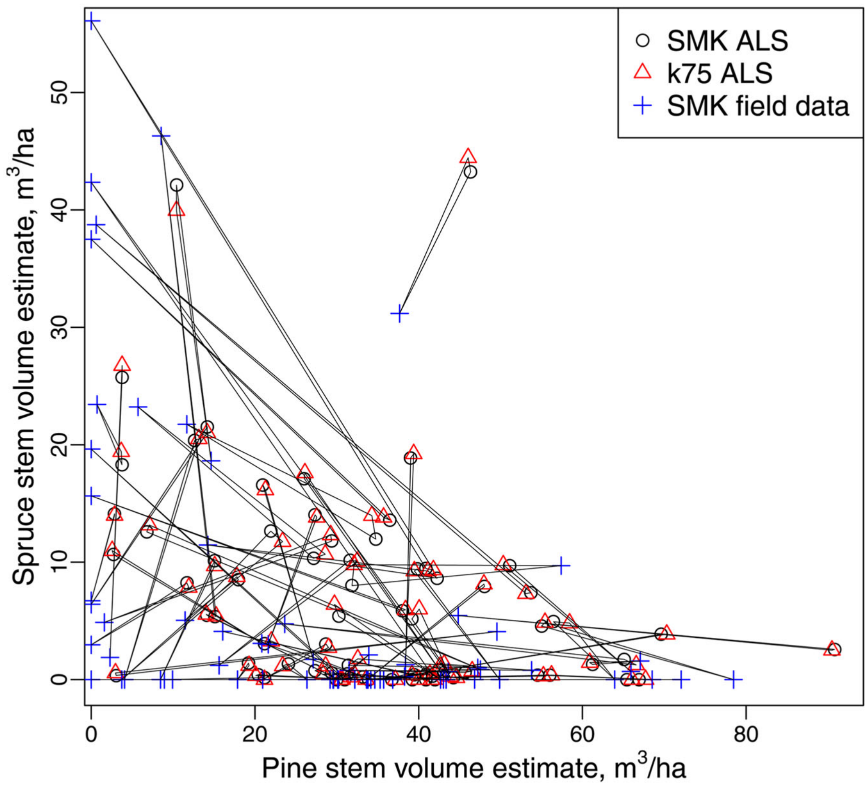

Figure 1. Spruce stem volume estimates versus pine stem volume estimates on the plots with mean dbh ≤ 11 cm ($n = 70$). The three configurations of estimates (six estimates) on one plot are (1) field data-based estimates using SMK plots (SMK field data), (2) ALS-assisted estimates using SMK plots (SMK ALS) and (3) ALS-assisted estimates using k75 plots (k75 ALS) and are connected by a polygon.

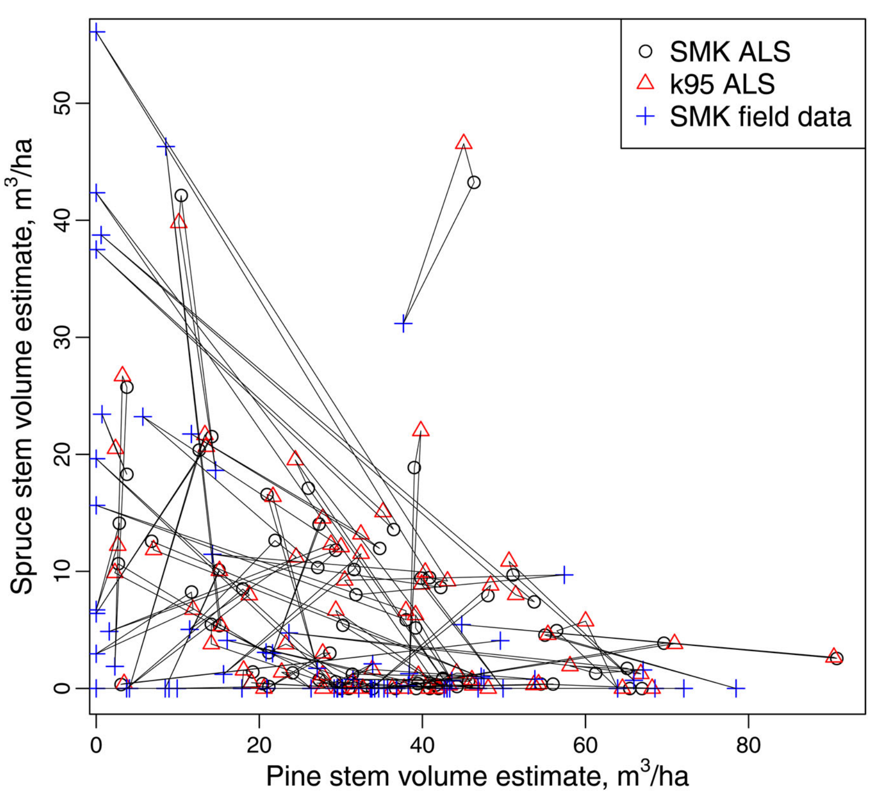

Figure 2. Spruce stem volume estimates versus pine stem volume estimates on the plots with mean dbh ≤ 11 cm ($n = 70$). The three configurations of estimates (six estimates) on one plot are (1) field data-based estimates using SMK plots (SMK field data), (2) ALS-assisted estimates using SMK plots (SMK ALS) and (3) ALS-assisted estimates using k95 plots (k95 ALS) are connected by a polygon.

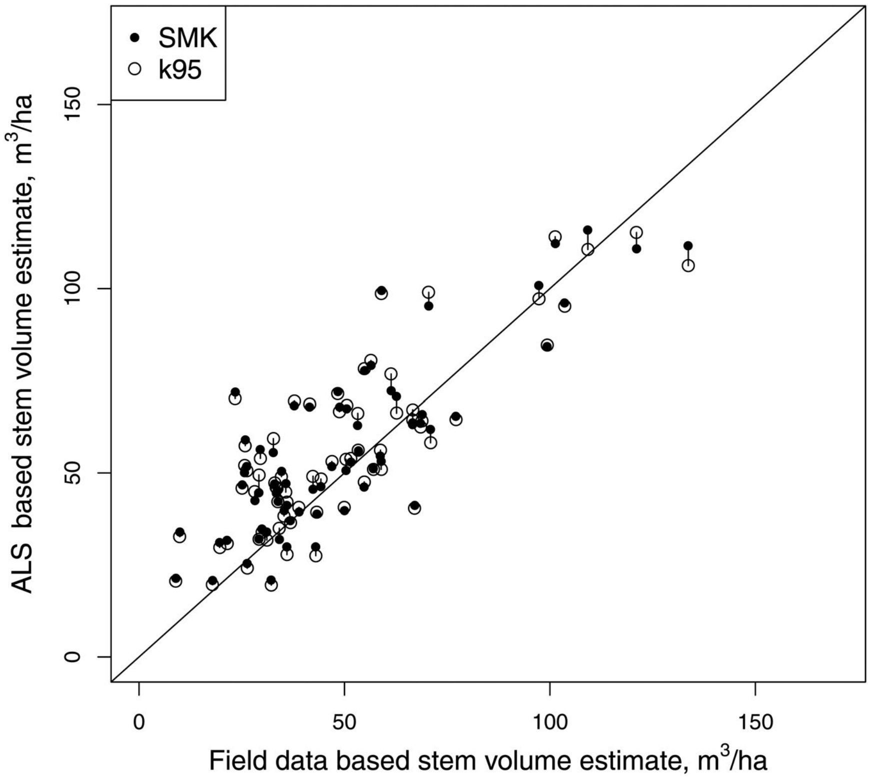

Figure 3. Two ALS-assisted stem volume estimates (a) using SMK plots (SMK) and (b) using k95 plots (k95) versus field data-based stem volume estimates using SMK plots on the plots with mean dbh ≤ 11 cm (n = 70). The estimates on one plot are connected by a line.

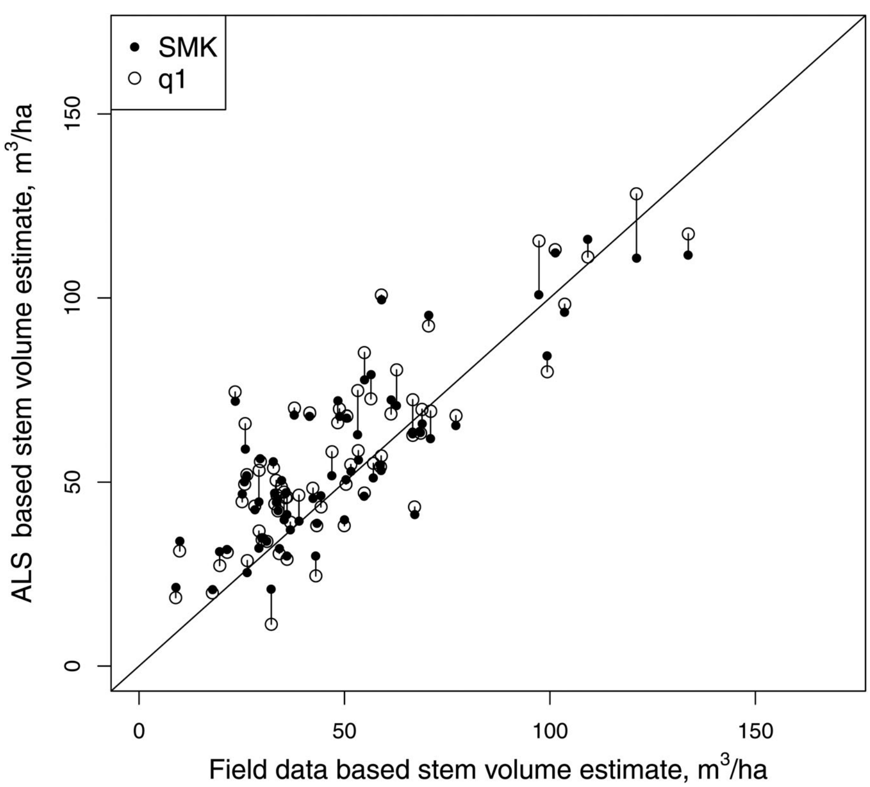

Figure 4. Two ALS-assisted stem volume estimates (a) using SMK plots (SMK) and (b) using q1 plots (q1) versus field data-based stem volume estimates using SMK plots on the plots with mean dbh ≤ 11 cm (n = 70). The estimates on one plot are connected by a line.

Table 6. Comparisons of the ALS-assisted estimates to the field data-based estimates on the 16 m × 16 m plots and on the 32 m × 32 m plots measured in 2014. The ALS-assisted estimates and RMSEs are calculated with the SMK, k75, k95, q1, q1.5, and q2 plots and 2013 data (n = 1150) as the training data and the 2014 data (n = 116 and n = 29) as the validation data.

	RMSE											
	16 m × 16 m plots						32 m × 32 m plots					
	SMK	k75	k95	q1	q1.5	q2	SMK	k75	k95	q1	q1.5	q2
Volume, m^3/ha	37.3	37.3	37.5	37.8	38.4	39.4	25.6	25.4	25.6	25.8	26.2	26.7
Pine, vol m^3/ha	47.1	47.1	47.0	47.1	47.6	49.2	39.3	39.3	39.2	39.4	39.4	40.4
Spruce, vol m^3/ha	44.2	44.1	44.0	44.8	45.7	45.3	35.4	35.3	35.2	35.9	36.7	36.0
Broadl. vol m^3/ha	21.0	21.0	20.9	21.2	22.3	22.6	17.0	17.2	17.0	16.9	17.8	18.0
Mean dbh, cm	2.3	2.3	2.4	2.4	2.4	2.5	2.0	2.0	2.0	2.0	2.1	2.1
Mean height, dm	11.4	11.4	11.3	10.9	10.8	10.9	9.6	9.5	9.5	9.1	9.0	9.1
Basal area, m^2/ha	5.2	5.2	5.2	5.3	5.4	5.5	4.2	4.2	4.2	4.3	4.3	4.4
Number of trees/ha	869.9	868.4	877.2	911.7	961.7	987.2	808.3	802.3	810.3	832.0	870.7	887.1

Table 7. Tally tree measurements, time, costs and savings compared to SMK plot measurements.

	Tally tree measurements		Savings, tally tree	Savings in total cost
Plot configuration	Mean, min.	Cost, euro	euro	%
SMK	23.0	12.6	0.00	0.0
k75	19.3	10.5	2.07	5.8
k95	17.4	9.5	3.07	8.6
q1	11.6	6.3	6.25	17.5

The reduction in total measurement costs was almost 9% when using the k95 plot and almost 18% when using the q1 plot instead of the SMK plot (Table 7).

Discussion

The effect of the field plot configuration on the errors of the ALS-assisted forest resource estimates were analysed. The estimation method, an improved non-parametric k-NN method, was purposively same for all plot configurations. The reason for using the same method was to focus on the effect of the plot configuration type only, and to be conservative in the sense that configurations with less tree data were not favored. The field measurement costs of the different plot configurations were also assessed.

The plot-level errors for the ALS-assisted estimates (Table 2) were well in line with previous results in estimating forest variables (Næsset 2002, 2004, 2007; Næsset et al. 2004; Maltamo et al. 2007; Nilsson et al. 2015). The plot-level RMSEs by tree species groups were also at the same level as in the other studies, and larger than the RMSEs of the estimates for all species together (Packalén and Maltamo 2007; Maltamo et al. 2015). Our aim was not to minimize the errors but to use in the tests a method that is comparable to those currently used in operative ALS-assisted applications.

The sampling designs with an inclusion probability of a tree that depends on the size of the tree is known to be efficient when estimating averages or totals. Concentric plots or angle count plots are often used (Tomppo et al. 2011). Our results confirm that, in forests similar to Finnish forests, even the angle count plot with the basal area factor $q = 1$ is a feasible option for ALS-assisted inventory because the ALS-derived estimates were not significantly different from those derived from the fixed area plots. cf. Maltamo et al. (2007), Hollaus et al. (2007) and Scrinzi et al. (2015). The basal area factors 1.5 and 2 increased the errors to some extent.

The error estimates were compared to those in given in a few other ALS-assisted forest inventory studies. The error estimates for the 32 m × 32 m plots were similar to those in Nilsson et al. (2015). Our smaller areal unit in validation than in Nilsson et al. (2015) may explain the differences. The size of the validation units, 1000 m^2 in Næsset (2007), is near our larger square. The RMSEs for the common variables, volume, mean diameter, basal area and mean height are comparable, though a bit larger in our case.

The errors of the estimates by tree species groups were larger than the errors of the estimates with tree species together, a common problem in any ALS-assisted approach (Packalén and Maltamo 2007; Maltamo et al. 2015).

The increased work load and costs, caused by the increased tree measurements on the SMK plots compared to the alternative plots, were assessed also using temporary field plot data from the Finnish eleventh NFI from years 2009–2013 and twelfth NFI from year 2014 in Southern part of the country and the plot-level time recordings. A truncated angle count plot with a basal area factor of $q = 2$ and a maximum distance of 12.52 m was used in the eleventh NFI while k95 plot configuration is used in the twelfth NFI. Additionally to the k95 configuration, the trees with diameter < 4.5 cm are measured using angle count sampling and a basal area factor of $q = 1.5$. The average plot measurement time was 44% larger in the twelfth NFI than in the eleventh NFI. The plots on forest land that did not intersect multiple forest stands were included in the comparisons. If the plots on the maturity classes 4–6 were used (Section ' Study site and plot measurements'), as in this study, the average plot measurement time increased by 38%.

The RMSE differences between field plot-based estimates of forest variables at the plot level were negligible when comparing the estimates for the k75 and k95 plots to those on the SMK plots. The RMSE differences increased to some extent on the angle count plots, slightly with the basal area factor of q = 1 and somewhat more with the factors of 1.5 and 2, as expected.

The slightly increased RMSEs based on the field data (conclusion 1) did not increase the RMSEs or the absolute values of the mean deviations of the ALS-assisted estimates when using the plot-level estimates of the k75 or k95 plots as the estimation training data instead of the SMK plots. Only the use of the angle count plots with the basal area factor of $q = 2$ as the source of training data in estimation increased the error quantities noticeably, and the use of the factor $q = 1.5$ slightly.

The conclusions remained unchanged when looking at the plots with the mean diameter of the trees not more than 11 cm and the plots with a large number of trees. These plots are often in mixed species forests and also with several tree layers, that is, in forests in which the errors of the ALS-assisted estimates are large.

Using concentric circular plots or an angle count plot with basal area factor $q = 1$ instead of the SMK plot could reduce costs about 10–20% without noticeably increasing the errors of the estimates.

The overall conclusion is that it is possible to select a field plot configuration that fulfills the efficiency needs of NFIs and plot-level accuracy requirements of ALS-assisted inventories. Of the tested alternatives, the angle count plot with a basal area factor of $q = 1$ could be recommended. It saves costs noticeably compared to a fixed area plot. Also the use of the concentric plot k95 would save costs. The practical Finnish NFI in 2015 and 2016 with the plot k95 has shown that our costs saving estimates are conservative. One technical conclusion is that the differences of the errors of the estimates of forest variables based on field data are not directly related to the differences of the errors in the ALS-assisted estimates. The reason is a slight increase in the RMSEs of the field data-based estimates does not directly contribute to the ALS-assisted estimates. Estimation with ALS and other auxiliary data used causes such large ik-nn prediction errors that they cover the smaller errors arising from the field data-based estimates. One should also note that the total volume, basal area and other characteristics of small trees ($d < 7.5$ cm or $d < 9.5$ cm) can be estimated with small absolute errors without measuring each tree on a fixed radius plot. The tree measurements play an important role in total time consumed wherefore thorough error and cost analyses are necessary when planning a cost-effective forest inventory. The use of the field plot for multiple inventories and purposes introduces additional considerations.

Acknowledgments

The Finnish Forest Centre provided the ALS data and the aerial image data. Ms. Megan McCormick gave technical help for the early versions of the manuscript. Mikael Strandström, Arto Ahola, Ville Pietilä and Juho Pitkänen as well as unnamed field crew members had a key role in the field data acquisition and processing. The Editor and an anonymous reviewer provided very useful and constructive comments that improved the manuscript noticeably.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The funding by the Finnish Ministry of Agriculture and Forestry, together with the funding by the Finnish Forest Research Institute, currently the Natural Resources Institute Finland, as well as the Finnish Forest Centre made this study possible.

ORCID

Erkki Tomppo http://orcid.org/0000-0002-5715-8912
Kai Mäkisara http://orcid.org/0000-0003-4949-5406
Matti Katila http://orcid.org/0000-0001-6946-5736

References

Barrett F, McRoberts RE, Tomppo E, Cienciala E, Waser LT. 2016. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. Remote Sens Environ. 174:279–289.

Eerikäinen K. 2009. A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the Finnish National Forest Inventory. For Sci. 55:480–493.

Gregoire TG, Valentine HT. 2008. Sampling strategies for natural resources and the environment. Applied environmental statistics. Boca Raton (FL): Chapman & Hall/CRC.

Haralikck RM, Shanmugam K, Dinstein I. 1973. Textural features for image classification. IEEE Trans Syst Man Cybern. SMC-3:610–621.

Harrell FE. 2001. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer.

Hollaus M, Wagner W, Maier B, Schadauer K. 2007. Airborne laser scanning of forest stem volume in a mountainous environment. Sensors. 7:1559–1577.

Laasasenaho J. 1982. Taper curve and volume functions for pine, spruce and birch. Commun Inst Fenn. 108:1–74.

Loetsch F, Haller KE. 1973. Forest inventory, Vol. I. München: BLV Verlagsgesellschaft mbH.

Maltamo M, Korhonen KT, Packalén P, Mehtätalo L, Suvanto A. 2007. Testing the usability of truncated angle count sample plots as ground truth in airborne laser scanning-based forest inventories. Forestry. 80:73–81.

Maltamo M., Næsset E., Vauhkonen J, editors. 2014. Forestry applications of airborne laser scanning – concepts and case studies. Managing forest ecosystems, Vol. 27. Heidelberg: Springer.

Maltamo M, Ørka OM, 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 Res. 4:336–345.

Maltamo M, Packalén P, Kallio E, Kangas J, Uuttera J, Heikkilä J. 2011. Airborne laser scanning based stand level management inventory in Finland. In: SilviLaser 2011. Hobart, Australia.

Maltamo M, Packalén P. 2014. Species-specific management inventory in Finland. In: Maltamo, M, Næsset, E, Vauhkonen, J, editors. Forestry applications of airborne laser scanning. Managing forest ecosystems, Vol. 27. Heidelberg: Springer; p. 241–252.

Mandallaz D. 2008. Sampling techniques for forest inventories. Boca Raton (FL): Chapman Hall/CRC.

McRoberts RE, Næsset E, Gobakken T. 2013. Inference for Lidar-assisted estimation of forest growing stock volume. Remote Sens Environ. 128:268–275.

Nilsson M, Nordkvist K, Jonzén J, Axensten P, Olsson H, Wallerman J, Egbert M, Lindgren N, Larsson S, Nilsson L, Eriksson J. 2015. A nationwide forest attribute map of Sweden derived using airborne laser scanning data and field data from the national forest inventory. In: Proceedings of SilviLaser 2015, September 28–30, 2015. La Grande Motte, France, p. 211–213.

Næsset E. 1997a. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J Photogramm Remote Sens. 52:49–56.

Naesset E. 1997b. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens Environ. 61:246–253.

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. 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scand J Res. 19:164–179.

Næsset E. 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. Scand J Res. 22:433–442.

Næsset E. 2014. Area-based inventory in Norway – from Innovation to an operational reality. In: Maltamo, M, Næsset, E, Vauhkonen, J, editors. Forestry application of airborne laser scanning. Managing forest ecosystems, Vol. 27. Heidelberg: Springer; p. 215–240.

Næsset E, Gobakken T, Bollandsås OM, Gregoire TG, Nelson R, Ståhl G. 2013. Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. Remote Sens Environ. 130:108–120.

Næsset E, Gobakken T, Holmgren J, Hyyppä H, Hyyppä J, Maltamo M, Nilsson M, Olsson H, Persson Å, Söderman U. 2004. Laser scanning of forest resources: the Nordic experience. Scand J Res. 19:482–499.

Næsset E, Gobakken T, Solberg S, Gregoire TG, Nelson R, Ståhl G, Weydahl D. 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: a case study from a boreal forest area. Remote Sens Environ. 115:3599–3614.

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.

Scrinzi G, Clementel F, Floris A. 2015. Angle count sampling reliability as ground truth for area-based LiDAR applications in forest inventories. Can J Res. 45:506–514.

Strunk JL, Reutebuch SE, Andersen H-E, Gould PJ, McGaughey RJ. 2012. Model-assisted forest yield estimation with light detection and ranging. West J Appl. 27:53–59.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol. 58:267–288.

Tomppo E, Gschwantner Th, Lawrence M, McRoberts RE, editors. 2010. National forest inventories – pathways for common reporting. Heidelberg: Springer.

Tomppo E, Haakana M, Katila M, Peräsaari J. 2008. Multi-source national forest inventory – methods and applications. Managing forest ecosystems, Vol. 18. Dordrecht: Springer.

Tomppo E, Halme M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sens Environ. 92:1–20.

Tomppo E, Heikkinen J, Henttonen HM, Ihalainen A, Katila M, Mäkelä H, Tuomainen T, Vainikainen N. 2011. Designing and conducting a forest inventory – case: 9th National Forest Inventory of Finland. Springer, Managing Forest Ecosystems, Vol. 22. Dordrecht: Springer.

Tomppo E, Katila M, Mäkisara K, Peräsaari J. 2014. The Multi-source National Forest Inventory of Finland – methods and results 2011. Working Papers of the Finnish Forest Research Institute 319. Available from: http://www.metla.fi/julkaisut/workingpapers/2014/mwp319.htm.

Vauhkonen J, Maltamo M, McRoberts RE, Næsset E. 2014. Introduction to forestry applications of airborne laser scanning. In: Maltamo, M, Næsset, E, Vauhkonen, J, editors. Forestry applications of airborne laser scanning. Managing forest ecosystems, Vol. 27. Dordrecht: Springer; p. 1–16.

Ware KD, Cunia T. 1962. Continuous forest inventory with partial replacement of samples. Forest Science Monograph, Vol. 3. Bethesda, MD: Society of American Foresters; p. 1–40.