Field calibration of merchantable and sawlog volumes in forest inventories based on airborne laser scanning

Karjalainen, Tomi

Canadian Science Publishing

Tieteelliset aikakauslehtiartikkelit
© The Author(s) 2020
All rights reserved
http://dx.doi.org/10.1139/cjfr-2020-0033

https://erepo.uef.fi/handle/123456789/24249
Downloaded from University of Eastern Finland's eRepository
Field calibration of merchantable and sawlog volumes in airborne laser scanning-based forest inventories

Tomi Karjalainen\textsuperscript{1*} (tomi.karjalainen@uef.fi)

Lauri Mehtätalo\textsuperscript{2} (lauri.mehtatalo@uef.fi)

Petteri Packalen\textsuperscript{1} (petteri.packalen@uef.fi)

Terje Gobakken\textsuperscript{3} (terje.gobakken@nmbu.no)

Erik Næsset\textsuperscript{3} (erik.naesset@nmbu.no)

Matti Maltamo\textsuperscript{1} (matti.maltamo@uef.fi)

\textsuperscript{1} School of Forest Sciences, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland.
\textsuperscript{*} Corresponding author: Tomi Karjalainen (e-mail: tomi.karjalainen@uef.fi)
\textsuperscript{2} School of Computing, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland.
\textsuperscript{3} Faculty of Environmental Science and Natural Resource Management, Norwegian University of Life Sciences, Box 5003, 1430 Ås, Norway.
ABSTRACT

In many countries, airborne laser scanning (ALS) inventories are implemented to produce predictions for stand-level forest attributes. Nevertheless, mature stands are usually field-visited prior to clear-cutting, so some measurements can be conducted on these stands to calibrate the ALS-based predictions. In this paper, we developed a seemingly unrelated multivariate mixed-effects model system that includes component models for basal area, merchantable volume and sawlog volume for 225 m² cells. We used ALS data and accurately positioned cut-to-length harvester observations from Norway spruce dominated clear-cut stands. Our aim was to study the effect of 1–10 local angle gauge basal area measurements on the accuracy of predicted merchantable and sawlog volumes. Seemingly unrelated mixed-effect model system was fitted to estimate cross-model correlations in residuals and random effects, which were then utilized to predict all the random effects of the system for calibrated stand-level predictions. The 10 angle gauge plots decreased the relative Root Mean Square Error (RMSE%) of the basal area and merchantable volume predictions from 16.8% to 10.5%, and from 15.8% to 11.9%, respectively. Cross-model correlations of the stand effects of sawlog volume with the other responses were low, therefore the initial RMSE% of ~22% was decreased only marginally by the calibration.

Keywords: LiDAR, area-based approach, quality, estimated best linear unbiased predictor, harvester data
INTRODUCTION

Many remote sensing technologies have proven to be useful in forest inventories around the world. In Norway, for example, almost all forest management inventories are based on airborne laser scanning (ALS), georeferenced field sample plots measured in the inventory area of interest (i.e. training data for ALS models), and visual interpretation of aerial images (Næset 2014). In Finland, aerial images are used in combination with ALS data and sample plots to predict species-specific stand attributes (Maltamo and Packalen 2014). The stand attributes of these inventories are suitable for forest management planning, but for some purposes, such as harvest planning, the obtained error level is not sufficient, or the inventory data are not sufficiently up to date for the intended purpose. For example, volumes of different timber assortments are very difficult to predict accurately (Korhonen et al. 2008).

In this study, we use term merchantable volume for such part of the stem that can be utilized in a way or another (sawlogs, pulpwood, energy wood) and sawlog volume for such parts of the stem that fulfill the requirements for minimum sawlog diameter, length, and quality (has no defects). What is considered as a defect, differs between species, but most of the defects are related to branches, crookedness, and rot. The requirements for sawlogs are quite similar across the Nordic countries, see SDC (2014) for the basic quality requirement for Scots pine (Pinus sylvestris L.) and Norway spruce (Picea abies [L.] Karst) in Sweden. Usually, the price of sawlogs is considerably higher than the price of pulpwood (Natural Resources Institute Finland 2019), so sawlog volume is an important attribute that describes tree quality from a commercial point of view. Various defects may downgrade sawlog-sized parts of the stems to the notably less valuable pulpwood during harvesting, making the estimation of the value of the growing stock problematic and prone to errors. Thus, more accurate predictions of sawlog volume would reduce the uncertainty in the decision making when harvesting operations
are planned and scheduled. In Finnish harvest planning practice, the theoretical sawlog volume is calculated using taper curves and requirements for dimensions, and the sawlog volume is then estimated from the theoretical sawlog volume with different sawlog reduction models that employ a range of predictors, such as site type and age of trees (e.g. Mehtätalo 2002). For Norway spruce, the defects reduce the sawlog volume by 18% on average (Mehtätalo 2002), but it is very hard to predict which trees and stands have these defects. Therefore, the predictions are not sufficiently accurate for the needs of stand-level harvest planning (Malinen et al. 2007) and it is common that the stands are visited in the field prior to clear-cutting.

In boreal forests in Nordic countries, the prediction of sawlog volumes with the area-based approach (ABA) (Næsset 1997) has typically yielded Root Mean Square Error (RMSE%) values of 20–30% at both the plot-(Bollandsås et al. 2011) and the stand-level (Korhonen et al. 2008). Diverse methods and datasets have been used. For example, harvester data based on actual sawlog volumes have been used in some studies. Bollandsås et al. (2011) fitted a model for sawlog volume with harvester data and ALS metrics, and reported a RMSE% value of 24% for 50 m × 50 m cells. Korhonen et al. (2008) also used harvester data, but only to validate their results. The actual sawlog volume estimates were produced by first bucking field measured trees using a taper curve and then employing a sawlog reduction model. The resulting RMSE% value for sawlog volume was 18% at the stand-level. Karjalainen et al. (2019) used visually bucked Scots pines in an ABA, which resulted in a RMSE% value of 21% for 30 m × 30 m cells.

Sawlog volume have been predicted by means of ALS also in other parts of the world, but due to greatly differing forest conditions, practices, and methods the comparison of the results to those obtained in boreal forests should be done with caution. In Wisconsin, USA, Hawbaker et al. (2010) predicted the sawlog volume for circular plots with a radius of 15.25 m by using ALS data. The resulting $R^2$ value for different sawlog volume models was about 0.65. For
Brazilian loblolly pine (*Pinus taeda* L.) plantations, on the other hand, Silva et al. (2017) predicted the sawlog volume for 20 m × 30 m plots with the Random Forest method. The resulting RMSE% for the predicted sawlog volume was 7.7%. The notably better accuracy in South American plantations compared to boreal conditions can probably be explained by the greater homogeneity of the trees.

In addition to area-based approach, ALS-based sawlog volume predictions can be made with the Individual Tree Detection (ITD) approach as well. Terrestrial Laser Scanning (TLS) has also potential for the purpose as it can be used to detect tree dimensions and defects below canopy. Kankare et al. (2014) used 1) ALS with ITD, 2) TLS, and 3) combination of ALS and TLS to predict sawlog volume for individual Scots pines in southern Finland. The corresponding RMSE% values for predicted sawlog volumes after dropping some extreme outliers with respect to tree quality were 34.7%, 17.5%, and 16.8%.

Many stand attributes, such as basal area and volume, are correlated, meaning that the measurements of one or more responses can be used to improve the predictions of the other responses as well. In general, such a cross-calibration procedure can be implemented with a multivariate mixed-effects modelling approach in which the information obtained from local observations of one or more responses is used to predict the random parts for all responses (Mehtätalo and Lappi 2020). If the correlation between responses is non-zero, such calibration improves the prediction also for variables without calibration measurements, compared to prediction based on the fixed part of the model only. In a forestry context, this approach was proposed by Lappi (1991), who discussed the calibration of tree volumes with height measurements (see also Lappi (1986) for an application to taper curves). Other examples of a similar approach are the studies of Mehtätalo (2005), who calibrated tree diameter percentiles with measured sample quantiles, and de Souza Vismara et al. (2016) who calibrated volume models of Eucalyptus plantations across two rotations.
ALS-based multivariate mixed-effects models can also be calibrated. Maltamo et al. (2012) studied the calibration of various tree-level attributes of Scots pine trees predicted by ALS. After constructing seemingly unrelated mixed-effects models, they calibrated the models to validation stands using local field measurements of some of the tree attributes. In most cases, the accuracy of predictions increased as the number of sample trees (1–10) used in the calibrations increased. Korhonen et al. (2019) studied the transferability and calibration of tree-level mixed-effects models. They focused only on sawlog-sized Scots pine trees, and the attributes of interest were tree height, diameter at breast height (DBH) and crown base height. Their results showed that the accuracy of predictions decreased after transferring the models to another inventory area where data was collected using a different ALS instrument, but the predictions improved with a few local measurements.

In harvest planning, there is an interest to obtain merchantable volume data and sawlog volume data. A motivation for this study was to assess if the accuracy of ALS-based predictions of these two volume attributes could be improved using field-based basal area (m² ha⁻¹) measurements that are easy and fast to carry out in the field, especially with an angle gauge. To realistically study the effect of such calibration, validation at the stand-level is required. For this purpose, spatially accurate harvester data from clear-cuts is ideal (see Hauglin et al. 2018) as, in addition to the location, the merchantable and sawlog volumes can be accurately obtained for harvested trees from the stem data (hpr-file), produced by a modern cut-to-length harvester. In addition, accurately positioned tree-level data allows us to sample an infinite number of artificial sets of angle gauge plots within stands, and to assess the predictions at the stand-level, which is a rarely viable scale of validation.

Traditionally, the lack of accurate positioning has prevented the utilization of harvester data, as the spatial accuracy has typically been about 10 m (Holmgren et al. 2012). Knowing the accurate location of the harvester head has many potential applications (Lindroos et al. 2015), so
systems that can provide such data have also been developed (Hauglin et al. 2017). Hauglin et al. (2018) studied the utilization of accurately positioned harvester data in the modelling of forest volume with ALS; the smallest RMSE% value for predicted volume on a 400 m² grid cell level was 19.2 %, and they concluded that spatially accurate harvester data could be used in forest inventories. Furthermore, Maltamo et al. (2019) used the same dataset in the prediction of diameter distribution, which resulted in stand-level RMSE% values less than 10 % for forest volume. The same harvester and ALS data are used in the current study.

In this paper, our aim was to study the effects of calibrations, based on 1–10 basal area measurements, on the accuracy of predicted merchantable volumes (m³ ha⁻¹) and sawlog volumes (m³ ha⁻¹). The initial predictions were conducted with a seemingly unrelated multivariate mixed-effects model system that employs ALS metrics as predictors. The application of the method proposed in this study is based on two presumptions: 1) that an ALS inventory has been carried out for the area of interest, and 2) mature stands are visited in field (e.g. for harvest planning purposes) prior to clear-cut. If the stands are to be visited in any case, undertaking simple measurements on the stand would be realistic as they would not increase the total costs significantly. These measurements can then be used to calibrate predictions for all stand attributes by utilizing the correlations between attributes.

MATERIAL AND METHODS

Study area

The study area was located in a boreal forest area in the Romerike region in Akershus County, southeastern Norway (60°25′ N, 11°4′ E). The forests are dominated by Norway spruce and Scots pine. In our data, the proportions of Norway spruce and Scots pine were 87.0 % and 7.3 % of total measured merchantable volumes, respectively. Some deciduous tree species, such as birch (Betula spp.), also exist in the area (5.7 % of total merchantable volume).
Harvester data

The tree data were collected using a John Deere 1270E harvester between January and October 2017. The harvester was equipped with an integrated positioning system to enable accurate positioning of the harvester head. Therefore, the accurate position of each harvested tree could also be recorded. The positioning system used is presented in greater detail by Hauglin et al. (2018), who also reported a mean positioning error of 0.75 m for the harvester-based tree positions. In addition to geographical positions, the tree data included DBH, diameter measurements for the whole stem at 10 cm intervals, and the length of each log. The operator also manually recorded the tree species and the quality assortment class of each log. Merchantable volume was the total volume of logs that passed the harvester head, whereas sawlog volume was the volume of logs qualified as sawlogs.

Layout of the grids

As we intended to construct ABA models (see next section), we first had to tessellate the harvester tree areas (also known as stands) into grid cells. We used a cell size of 15 m × 15 m, because this size would be similar to, or close to the size that is commonly used in operational ALS inventories in Nordic countries (e.g. Maltamo & Packalen 2014, Næsset 2014). Instead of laying one comprehensive grid with a fixed origin atop the entire study area, we optimized the location of separate grids for each individual clear-cut stand by maximizing the total number of cells that would fit within each individual stand. The process was conducted in the R software (R Core Team 2017) as follows. We had neither stand border geometries nor any control of the status of the surrounding stands, so we had to first generate realistic borders for the clear-cut stands. To begin, we aggregated our point-wise tree data to polygons by creating two-dimensional alpha shapes using the alphahull R package (Pateiro–Lopez & Rodríguez–Casal 2016). An alpha shape is a generalization of the convex hull (Edelsbrunner 1983). We used an alpha value of 10 m to obtain useful stand borders. The same alpha value was used...
with the same data for the same purpose also by Hauglin et al. (2018) and Maltamo et al. (2019).

Next, we used the minimum x- and y-coordinates of each stand to lay out an initial 15 m × 15 m grid. Furthermore, each of the grid cells were divided into 225 1-m$^2$ regular sub-cells, which were all inspected to determine whether they intersected the boundaries of the stand in question. For the 225 m$^2$ cells, only those that contained at least 215 1-m$^2$ sub-cells within the stand were accepted and included in the stand. This threshold of 215/225 was defined subjectively by us. The decision to include cells that did not fall entirely within the original stand borders was supported by the fact that the borders created by the alpha shape followed the positions of the outermost tree stems. This means that a large proportion of the crowns of these border trees fell outside the stand. Therefore, allowing a cell to expand slightly outside the stand border would better reflect the stand association of a tree as it is observed in the ALS data. Finally, the origin (xy) of a stand level grid for which the number of cells was maximized was found iteratively in an incremental search by 1-m steps in the xy-directions. The principle of the process is illustrated in Fig. 1. Later in the validation, the resulting stand-specific formations of 15 m × 15 m cells will be considered as the true stands, i.e. the excluded areas near stand borders will be ignored from the analysis as illustrated in Fig. 1 D. Moreover, those cells that slightly intersected the stand borders but were included in the stand will be considered as full cells, i.e. they will not be weighted differently when stand-level values are aggregated.

The harvested trees were assigned to grid cells according to their known positions. The following stand attributes were calculated for each cell: basal area (m$^2$ ha$^{-1}$) (the sum of cross-sectional areas at the height of 1.3 m of all harvested trees within the cell scaled to hectare level), merchantable volume (m$^3$ ha$^{-1}$), and sawlog volume (m$^3$ ha$^{-1}$). After the tessellation process, the entire dataset consisted of 2292 cells from 48 separate clear-cut stands.
We had prior knowledge of the spatial correlation of stem volume up to about 200 m in the study area (Hauglin et al. 2018). Therefore, to avoid problems later in the validation process, we divided our data into separate training and validation datasets by stands. The largest stand and the 3rd to the 16th largest stands were chosen for the validation (15 stands, 1654 cells), whereas the remainder of the stands (including the 2nd largest stand) comprised the training data (33 stands, 638 cells). The 2nd largest stand (285 cells) was included in the training data to provide more observations (also from one very large stand) for model fitting. We divided the data in this way, because the calibrations are more reasonable in larger stands than in stands composed of only a few 15 m × 15 m cells. In addition, placing 10 non-overlapping plots (see section on calibration plots) within smaller stands would have been a challenge, if not impossible. A summary of the attributes of the training and validation datasets is displayed in Table 1. The total area of training stands created by alpha shape was 22.7 ha of which the eventual 638 training cells covered 14.4 ha (63.3%). The total area of validation stands, on the other hand, was 47.6 ha of which the validation cells covered 37.2 ha (78.2%).

ALS data

The ALS data were collected in the summer 2013 with a Leica ALS70 instrument. The average flying altitude was 3000 m above ground level, pulse repetition frequency was 104.6 kHz, and the average point density was 0.7 points m⁻² at ground level. The ALS echoes were classified into ground hits and vegetation hits as proposed in Axelsson (1999). The aboveground height of the echoes was then defined as the vertical distance from a triangulated irregular network created from the ground echoes. The Leica ALS70 is able to record multiple echoes for each pulse. We created two different sets of echoes: the first (first of many + only) and the last (last of many + only) echoes. We extracted the ALS echoes for each grid cell and calculated the ALS metrics separately from the first and last echoes using the LASmetrics function.
in rLiDAR package (Silva et al. 2017) in R software (R Core Team 2017). The derived ALS
metrics are described in Table 2.

Seemingly unrelated linear mixed-effects models
Our data had a spatially grouped structure as the clear-cut stands consisted of many 15 m × 15
m grid cells (Fig. 1 D). Therefore, we used a mixed-effects modelling approach. First, we
constructed linear mixed-effects models for basal area, merchantable volume and sawlog vol-
ume separately using the training data. The optimal composition, i.e. the predictors and their
transforms, the structure of the random part of the model, the optimal variance function to
model heteroscedasticity, and the correlation structure to model the dependence among the
within-group errors, was examined separately for each response using function lme of the
nlme package (Pinheiro et al. 2019) in R software (R Core Team 2017). Models were fitted
using the Restricted Maximum Likelihood approach (e.g. Fahrmeir et al. 2013). The various
models were compared using Akaike Information Criterion values.

These univariate models were then merged into one multivariate (seemingly unrelated) linear
mixed-effect model system so as to model all three responses simultaneously. The motivation
for it was the estimation of cross-model correlation, as simultaneous fitting of well-formu-
lated seemingly unrelated system does not provide meaningful benefits in the estimation of
the fixed effects of the model (see Chapter 9 of Mehtätalo and Lappi (2020) for discussion).
Without local (i.e. stand, in this study) information, it is only possible to use the fixed part of
the model system in prediction. However, if local observations are available, it also enables
simultaneous prediction and utilization of the random parts of all responses. These calibra-
tions were carried out by employing the estimated best linear unbiased predictor (EBLUP).
See Appendix A for more details and for an example of the calibration procedure.
Calibration plots

Calibrations were tested in the validation stands using 1–10 simulated sample plots. First, stand borders were buffered 5 m inward. Furthermore, the fixed radius plots \((r = 8.46 \text{ m})\) were not allowed to intersect the 5 m buffered area, so the total buffer was 13.46 m inwards. Such buffering was justified as angle gauge plots would not be measured very close to stand borders due to edge effects in practice either. The first plot center of a stand was then placed randomly inside the buffered stand. In operational application the plots would not be measured very close to each other. To model such hard-core sampling of plot locations, the remaining nine calibration plots, which were sampled one at a time, were not allowed to be located too close to the previous plots. The radius of the hard-core area \((m)\) for the new plot center was defined separately for each stand according to the stand area in \(m^2\) (Eq. 1). The layout of the sample plots is illustrated in Fig. 2.

\[
r_{hc}(m) = 8.46 \text{ m} + \sqrt{\frac{\text{Area in } m^2}{10\pi}} \]

We computed similar ALS metrics for each plot with a radius of 8.46 m as for the grid cells (Table 2). The plot placement process (always 10 plots) was repeated 500 times for each validation stand to avoid any extreme effects caused by the randomness in the process. Let us consider this dataset as an array with 500 rows and 10 columns. Our aim was to study the effects of using 1–10 plots in the calibration. Instead of sampling unique sets of plots for each of the ten scenarios with a different number of plots, we always dropped the extra columns from the previously described array. For example, when we used six plots in the calibration, we dropped the last four columns. As the placement of 10 plots was repeated 500 times for each of the 15 validation stands, the total number of simulated plots was 75,000.

In practice, the simulated field measurements are used only to provide local measurements of basal area. Therefore, we calculated the basal area for each plot assuming: 1) a fixed radius
plot, and 2) an angle gauge plot. In the fixed radius plots, the radius was 8.46 m (225 m$^2$), which corresponds to the cell size used in this study. In the angle gauge plots, the distance how far a tree is counted in depends on the DBH of the tree in question and the applied basal area factor (BAF), so that the inclusion zone (in m) of each tree is defined by DBH / (2*sqrt(BAF)), where DBH is in cm. Here we used a BAF of 1 m$^2$ha$^{-1}$, i.e. for example a tree with a DBH of 25 cm was counted in if the distance to the plot center was less than 12.5 m and the tree count on the plot is directly the basal area in m$^2$ha$^{-1}$. BAF of 1 is most often being used also on mature stands in Nordic countries. Even though the plot centers had at least 13.46 m distance to stand border, it was still possible that some large trees could have been counted in from beyond the stand borders and the basal area measurements would therefore by slightly downward biased. To correct for such edge-bias, we generated a 30 m buffer around each stand. The locations of the trees in the buffer zone were generated by assuming a complete spatial randomness of the locations and using stand density of the stand as point density. The diameters were generated by sampling with replacement from among the observed trees of the stand. These generated trees are also illustrated in Fig. 2. Note that some inaccuracy is unavoidable when the angle gauge plots are linked to the corresponding ALS metrics because the plot radius depends on tree diameter. Here, we opted to use the same radius as employed in the fixed area plots in computing the ALS metrics.

Accuracy assessment

Cell-level predictions were aggregated to the stand-level where the performance was assessed. We report the results in the training and validation stands without calibration and with a different number of calibration plots. Results for basal area are also provided. For comparison, we provide also the results for such estimates that were derived as the mean observation from ten fixed area field plots, i.e. excluding ALS inventory entirely. We assessed the accuracy using empirical RMSE% (Eq. 2) and relative mean difference (MD%) (Eq. 3).
\begin{equation}
\text{RMSE}\% = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \times \frac{100}{\bar{y}}
\end{equation}

\begin{equation}
\text{MD}\% = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n} \times \frac{100}{\bar{y}}
\end{equation}

where \( n \) is the number of stands, \( y_i \) is the observed value for stand \( i \) (basal area, merchantable volume, or sawlog volume), \( \hat{y}_i \) is the predicted value for stand \( i \) and \( \bar{y} \) is the mean of the observed values (basal area, merchantable volume, or sawlog volume). For the calibrations, \( \hat{y}_i \) is the mean of the 500 predicted values that were obtained using different sets of either angle gauge or fixed radius plots in the calibration. Thus, \( \hat{y}_i \) varied for each scenario with a different number of plots. The estimates derived as the mean observation from ten plots were correspondingly the means of the 500 replicates.

**RESULTS**

**Linear mixed-effects models**

The response-specific fixed parts of the multivariate model system are shown in Table 3. A strong predictor for all the responses was the 60th percentile of height of first echoes. In the basal area and merchantable volume models, the standard deviation of height (first echoes) also performed well. The squared 60th percentile of height of first echoes was included in all models to model the nonlinear trends.

Variances, covariances, and correlations of the random parts of the multivariate model system are shown in Table 4. A random intercept was sufficient for the basal area and merchantable volume models, whereas a random slope for predictor \( f_{H60}^2 \) was also needed for the sawlog volume. The power-type variance function (Pinheiro and Bates 2000, Pinheiro et al. 2019) using \( f_{H60} \) as the covariate was used for all responses in order to model heteroscedasticity, and a constant correlation between the residual errors of each model pair was used to model the cross-model correlation of residuals (Table 5) (see Appendix A for formal definitions).
The random intercept of the basal area strongly correlated with the random intercept of merchantable volume (Table 4, Fig. 3), whereas for the random effects of the sawlog volume model the correlation with the random intercept of the basal area were small. The corresponding correlations were rather small between merchantable volume and sawlog volume as well. Similarly, the relationship of the Pearson residuals was clearly weaker between basal area and sawlog volume, than between basal area and merchantable volume (Fig. 4). Neither of the plots in Fig. 4 indicated a departure from a linear trend. Overall, these correlations indicate that when using basal area information in calibrations, the accuracy of merchantable volume predictions will improve, whereas significant improvement in the sawlog volume will not happen.

**Basal areas of angle gauge plots**

In the 75,000 simulated plots, the RMSE% and MD% values of the basal area of the angle gauge plots (compared to the fixed radius plots) were 23.2 % and 0.01 %, respectively. In general, the angle gauge plots slightly overestimated the basal area at small observed values and underestimated the basal area at large observed values (Fig. 5). These differences were mostly caused by the variable radius of the angle gauge plots. It appeared that altogether 10,785 generated trees on 7,676 (of 75,000) sample plots were counted in from outside the stands, so the edge-correction by means of generated trees was needed despite the used buffering. The importance was naturally the greater the smaller the stand in question was.

**Effects of calibrations**

The effects of angle gauge calibrations in the validation stands are illustrated in Fig. 6. The numerical values are also provided in Tables B1 and B2 in the Appendix B. There were some clear differences between the training and validation stands, especially in the case of sawlog volume, when only fixed effects were used in the prediction. Nevertheless, as expected, the most accurate calibrated predictions were obtained with 10 plots. However, as the number of
plots increased, the slopes of curves slowly approached zero, i.e. the benefit of each additional plot got smaller as the number of measured plots increased. The accuracies of basal area and merchantable volume predictions were increased notably, whereas the increase in accuracy of sawlog volume prediction was only marginal. For the basal area and merchantable volume predictions, the results of the calibrations were only slightly more accurate with the fixed radius (Tables B1 and B2 in Appendix B) than the angle gauge plots. For sawlog volume, on the other hand, the use of basal area derived from fixed radius plots even decreased the accuracies of predictions compared to the use of fixed effects only. For comparison, we calculated the RMSE% and MD% values also by assuming that the estimates for each stand were derived directly as the mean observation from ten plots, i.e. without ALS data (Tables B1 and B2 in appendix B). As expected, the accuracies increased notably. The results were not severely biased either.

The distributions of the calibration effects when merchantable volume was calibrated with the angle gauge plots are illustrated in Fig. 7. When one calibration plot was used instead of only the fixed effects of the model, the relative error of predictions decreased by approximately 0.5 percent points (pp), on average. However, with one calibration plot per stand, the accuracy of merchantable volume predictions decreased due to calibration in about 33% of replicates. With an increasing number of plots, the accuracy of predictions increased on average, but the variation also increased. With 10 plots, the relative error of predictions decreased by 2.9 pp on average, but the accuracy decreased in 23.6% of the repetitions.

DISCUSSION

Overview

The aim of this paper was to study the effects of calibrations (based on basal area measurements) on the accuracy of ALS-based predictions of stand-level merchantable volume and
sawlog volume. Our results showed that the accuracy of merchantable volume predictions can be expected to increase with such calibrations. On the other hand, obtaining an increase in accuracy of sawlog volume predictions is improbable because the correlation between the basal area and sawlog volume was only weak. Due to this weak correlation, the calibrated sawlog volume predictions were close to those that were obtained by using only fixed effects. Anyway, these results have practical implications for harvest planning and scheduling, assuming that an ALS inventory has been carried out for the area of interest, and that the mature stands will be visited in the field prior to clear-cutting. In such cases, a few simple and fast measurements would not increase the total costs of the inventory substantially.

The same local dataset was used both as training and validation data, which may result in overly optimistic model predictions. In our case, we had to use the same data since, to our knowledge, similar datasets that include sawlog volumes do not yet exist. On the other hand, we separated the data into training and validation stands based on the areas of the stands. The RMSE% and MD% values of predictions that were based on only fixed effects clearly differed between the training and validation stands (B1 and B2 in Appendix B), which indicates that there were differences also in forest structures (and/or ALS data) between the training and validation stands. On cell-level the mean attributes were quite similar (Table 1), so the differences can probably be explained by the fact that the training stands were clearly smaller than the validation stands. In addition, there occurred also some temporal and spatial mismatches between the ALS and field data (see Hauglin et al. 2018 for more details), thereby decreasing the correlation between ALS echoes and the forest attributes. Nevertheless, there are no reasons to doubt the generalizability of the method and results, even though the numeric results with other data could differ from what was obtained here.

In the calibrations, we assumed that the angle gauge measurements had the same variance as the 15 m × 15 m cells. However, as seen in Fig. 5, errors are unavoidable when the angle
gauge measurements are merged to fixed radius plots: the estimates based on angle gauge plots are affected also by trees that are outside the fixed radius plot. Therefore, the sum of random effect and residual of each measured plot also includes the merging error. On the other hand, the angle-gauge plots may have smaller variance than the fixed area plots. We could have made more specific calibrations by adjusting diagonal of $R_o$ matrix to describe the within-stand variance of the measurements in different situations (see Appendix A) (Lappi 1986, Mehtätalo and Kangas 2005). However, such process would have complicated the study even more and we believe that the results would not have changed much. Therefore, we were content with the simpler calibration procedure in this study.

When the plots were sampled, we used such buffering that plot centres were located at least 13.46 m (radius 8.46 m + 5 m) from the stand borders. The plots were not allowed to locate too close to each other either. Similar assumptions are likely to be applied in practice as well when angle gauge plots are being measured and are therefore justified. The use of additional 5 m buffer inward caused that areas very near stand edges were not presented in the samples. However, as seen in Table B2 in Appendix B, the results based only on field data were not severely biased, so the effect of not having plots very close to stand borders was only minor. In addition, the ALS-based calibration process is based on the observed residuals from sample plots, so it is not so straightforward to conclude that the use of 5 m buffer systematically changed the results. The hard-core selection of plots, on the other hand, does not lead to bias as all locations are still equally probable (excluding edges). In fact, the variance of estimates from the current procedure is lower than what it would be in independent plot placement.

Moreover, the samples should be collected from the population of interest, i.e. in this study from within the area covered by the accepted 15 m × 15 m cells. As seen in Fig. 1 D, many cells near stand borders were excluded from the analysis. Without using the 5 m additional
buffer inward, the area covered by the plots but not by the accepted cells would have been not-
ably larger. One may ask why the fixed radius plots were not forced within accepted cells, but then also angle gauge plots would have expanded outside the cells, and the areas near stand edges would have been represented even less. The issue that some sample plots partly covered small areas that were not covered by the accepted cells can be assumed to have only minor effect on the observed results. This is because the stands were homogenous and cells were spatially autocorrelated, i.e. differences in trees between adjacent cells were likely to be only small.

All in all, the main aim of this paper was not to study the edge-effects but to evaluate whether basal area information can be used to increase the accuracy of predicted merchantable and sawlog volumes. For this reason, the irregular areas next to stand edges were ignored and the stands were buffered inwards in the plot sampling process. Nevertheless, we think that the chosen methodology serves adequately for the aim of this study so that reliable conclusions can be made.

**Angle gauge calibrations**

Usually the relationship between basal area and merchantable volume is strong. In our study, the random parts of the models and the residuals of the predictions between these two responses showed a strong correlation (Fig. 3 subplot 1; Fig. 4 subplot 1). Therefore, our angle gauge calibrations also improved the accuracy of these two responses. The most accurate post-calibration RMSE% values for the stand-level basal area and merchantable volume were small, 10.5 % and 11.9 %, respectively.

For the stand-level basal area and total volume, RMSE% values of 10–15 % have typically been obtained with ALS in boreal forest conditions (e.g. Næsset 2007). Our results with only fixed effects are almost within this range, although instead of total volume, we focused on
merchantable volume which excluded treetops and the above-ground stumps that did not pass
through the harvester head. Moreover, Maltamo et al. (2019) used the same dataset and their
smallest RMSE% value was 8.4 % for the stand merchantable volume using the k-nearest
neighbor approach. The slightly greater accuracy of merchantable volume predictions in
Maltamo et al. (2019) compared to the current study, can probably be explained by the fact
that they only used stands larger than 0.5 ha in their validation.

There are many plausible reasons why the sawlog volume model was, in general, notably less
accurate than the merchantable volume model. Most importantly, sawlog volume is affected
by tree quality. Due to defects, two different trees with the same dimensions can yield very
different sawlog volumes. It is reasonable to assume that most of the defects, such as butt rot
or crooks in the stem, correlate neither with the crown attributes of the tree nor with basal
area. Therefore, they are difficult to detect in ALS data or in the field measurements of basal
area. In addition, the sawlogs in our data could only be bucked from spruce and pine trees –
this is the common policy in Norwegian forestry. Therefore, deciduous trees, which account
for 5.7 % of total merchantable volume, complicate the prediction of sawlog volume as they
could not be detected and separated from the conifers. In upcoming studies, aerial images that
were not available for this study could be utilized to detect deciduous trees. Among other
things, the operator of the harvester and the bucking of the logs also affect the sawlog vol-
umes.

A closer examination of the calibration of the sawlog volume predictions showed that the suc-
cess of calibration was stand-specific. In some stands, the accuracy of sawlog volume predic-
tions slightly increased by calibration, but in some other stands the decrease of the same mag-
nitude was observed, practically canceling the improvement in RMSE%. There were no ap-
parent differences in forest conditions between the successful and non-successful stands, and
some of the differences were illogical. For example, the mean of correlation between the grid
cell-level basal area and sawlog volume was actually stronger in those stands where the saw-
log volume predictions could not be improved with basal area information. Thus, the success
of calibration appears to depend on multiple aspects simultaneously, including the properties
of the plots. For example, a large basal area does not automatically mean large sawlog vol-
umes, as a large basal area may occur even in stands with a greater number of small trees with
small sawlog volumes. In mature stands, this probably represents an uncommon stand struc-
ture, but even a single plot with such properties may influence the calibration. Nevertheless,
the correlation between basal area and sawlog volume was not sufficiently strong to notably
improve the accuracy of sawlog volume predictions with angle gauge calibrations. The ob-
served RMSE% value of approximately 22 % is somewhat expected with reference to previ-
ous studies (see details in the introduction).

Conditions for successful calibration and distribution of calibration effects

The success of one implemented calibration depends on two things: 1) how accurate is the
stand-level prediction that is based only on fixed effects, and 2) how large is the residual error
of the measured plot. If the original prediction happens to be very close to the true value, it is
very likely that the calibration will decrease the accuracy. On the other hand, the absolute value
and the sign of the residual error of a plot compared to the random effect determine which
direction and how much the calibrated predictions will be shifted. It should be remembered that
if a plot has a different forest structure to the rest of the stand, the ALS metrics of the plot should
also indicate divergence. Thus, there are no methods to evaluate the representativeness of a plot
beforehand. If information about the representativeness of plots would be available, it would
be included in the model as a fixed predictor. Moreover, using specific requirements for plots
(such as a minimum stem number) would add a systematic error to the estimates, so the only
viable way to actually minimize the effect of unrepresentative plots is to increase the number of calibration plots.

In practice, the calibration plots would be measured only once. However, we simulated a large number of repeats and the results of the merchantable volume calibrations were generally as expected, i.e. the mean accuracy increased as more plots were used in the calibration (Fig. 7).

Also, the more plots that are measured, in general, the more plots with large residual error with respect to the model could be expected. This explains why the minimum of the boxes (Fig. 7), i.e. the largest reduction in the accuracy of predictions, steadily decreased as the number of plots increased. Nevertheless, it seems probable that one carefully measured set of angle gauge plots will increase the accuracy of merchantable volume predictions.

**Potential adoption in practice**

In this study, we fitted the models using observations from a cut-to-length harvester, which provided us with the sawlog volumes for each tree. However, harvester data are not essential, as the models could be fitted with regular, manually measured sample plots as well, as long as the same attributes are measured. In fact, as our results indicate that a successful calibration of sawlog volume predictions is unlikely, it can be questioned whether training data for sawlog volume is even needed. Then, laborious and time-consuming visual bucking to estimate tree level sawlog volumes (e.g. Karjalainen et al. 2019) would not be needed during field work.

Moreover, if the stands are visited, it would also be easy to visually evaluate the external quality of the trees. That is, in addition to measuring the angle gauge plots, the field worker could also visually assess a sawlog reduction factor for the stand. The stand sawlog volume could then be evaluated by applying the estimated sawlog reduction factor to the calibrated merchantable volume estimates. This method would be highly subjective, but on the other
hand it would not require any initial predictions of sawlog volume. The expected level of ac-
curacy of such sawlog volume predictions should be investigated and compared to what was
obtained in the current study with harvester data. However, if the acquisition and processing
of spatially accurate harvester data becomes a highly automatized and inexpensive by-product
of clear-cuts, then harvester-collected data would be an excellent choice for training data, as-
suming that the ALS data for the inventory area of interest had also been collected shortly be-
fore cutting. Such data would provide an easy avenue for the prediction of stand-level sawlog
volume for other (un-cut) stands on the inventory area.

Two alternatives to measure the basal area of a plot were also compared. In practice, basal
area can be measured with an angle gauge in less than one minute. On the other hand, the use
of an angle gauge includes some subjectivity and additional errors arise when the estimates
are merged to fixed-sized plots, from which the ALS metrics are computed. However, these
are probably just minor drawbacks. Angle gauge plots have been successfully combined with
fixed area ALS plots also by e.g. Hayashi et al. (2015). The calibration plots must also be po-
positioned. Sub-meter accuracy is required in positioning, as in normal ABA inventories (Næs-
devices and post-processing, the positioning may take several minutes under forest canopies
(Valbuena 2014; McGaughey et al. 2017). Therefore, to observe basal areas more accurately,
it might be possible to use fixed radius plots and measure the DBH of all trees while the posi-
tioning data are collected. Nevertheless, using angle gauge plots for calibration is an efficient
solution that performs almost as well as fixed radius plots (Tables B1 and B2 in Appendix B).
In addition, if the measurement of a fixed radius plot takes longer than the plot positioning,
then more angle gauge plots than fixed radius plots can be measured within the time available.
For example, the time consumption of two fixed radius plots (including plot positioning)
could turn out to be comparable to three angle gauge plots. Additional benefit of angle gauge plots is also that the measurements can be easily conducted by one person.

CONCLUSIONS

The prediction of sawlog volume yielded an RMSE% value of approximately 22 % after aggregating the 15 m × 15 m cell-level predictions to the stand-level, but only small improvements were obtained with calibrations based on angle gauge measurements. The RMSE% value of merchantable volume (and basal area) predictions improved considerably with the angle gauge plot calibrations, indicating possibilities for practical applications.

ACKNOWLEDGEMENTS

We would like to thank John Deere Forestry Norway AS, Viken Skog SA, Røsåsen Skogsmaskiner AS, Steinar Sætha, Steinar Åmål, and the Norwegian University of Life Sciences for collecting and providing the data. We also thank the two anonymous reviewers for their constructive comments on the manuscript. Petteri Packalen was supported by the Strategic Research Council of the Academy of Finland for the FORBIO project (Decision Number 314224).
REFERENCES


Natural Resources Institute Finland. (2019). Statistics Database: Stumpage prices of roundwood by Year, Region, Felling method and Roundwood assortment. Available from https://statdb.luke.fi/pxweb/pxweb/en/LUKE/LUKE_04%20Metsa_04%20Talous_02%20Teollisuuspuun_02%20kauppa_04%20Vuositilastot/02_Kantohinnat_v_maakunnittain.px/table tableViewLayout1/?rxid=876f7fa7-4bdf-4f29-b6ff-574f096266dc


**Table 1.** Attributes of the 15 × 15 m cells used in the training and validation datasets. Minimum / mean (standard deviation) / maximum.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Training (n₁=33, n₂=638)</th>
<th>Validation (n₁ = 15, n₂=1654)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal area (m² ha⁻¹)</td>
<td>0.9 / 31.4 (13.7) / 83.2</td>
<td>0.8 / 30.3 (12.6) / 76.0</td>
</tr>
<tr>
<td>Merchantable volume (m³ ha⁻¹)</td>
<td>6.9 / 294.5 (148.9) / 927.9</td>
<td>4.6 / 284.6 (132.5) / 845.0</td>
</tr>
<tr>
<td>Sawlog volume (m³ ha⁻¹)</td>
<td>0 / 98.6 (75.2) / 495.2</td>
<td>0 / 124.4 (85.9) / 544.9</td>
</tr>
<tr>
<td>Stand area (ha)</td>
<td>0.02 / 0.4 (1.1) / 6.4</td>
<td>0.9 / 2.5 (2.3) / 10.2</td>
</tr>
<tr>
<td>Mean diameter (cm)</td>
<td>11.9 / 22.2 (3.6) / 38.6</td>
<td>13.5 / 22.4 (3.8) / 48.5</td>
</tr>
<tr>
<td>Lorey’s mean height* (m)</td>
<td>11.5 / 20.8 (2.7) / 27.8</td>
<td>13.3 / 20.8 (2.3) / 29.0</td>
</tr>
</tbody>
</table>

**Note:** n₁ = number of stands, n₂ = number of cells, *Tree heights were estimated with Norwegian taper curves (see Hauglin et al. 2018).
Table 2. Description of the airborne laser scanning (ALS) metrics used as candidate predictors in the model system.

<table>
<thead>
<tr>
<th>ALS metric</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMAX/IMAX</td>
<td>Maximum H/I</td>
</tr>
<tr>
<td>HMIN/IMIN</td>
<td>Minimum H/I</td>
</tr>
<tr>
<td>HMEDIAN/IMEDIAN</td>
<td>Median H/I</td>
</tr>
<tr>
<td>HMEAN/IMEAN</td>
<td>Mean H/I</td>
</tr>
<tr>
<td>HSD/ISD</td>
<td>Standard deviation of H/I</td>
</tr>
<tr>
<td>HCV/ICV</td>
<td>Coefficient of variation of H/I</td>
</tr>
<tr>
<td>HMODE/IMODE</td>
<td>Mode of H/I</td>
</tr>
<tr>
<td>HVAR/IVAR</td>
<td>Variance of H/I</td>
</tr>
<tr>
<td>HKUR/IKUR</td>
<td>Kurtosis of H/I</td>
</tr>
<tr>
<td>HSKE/ISKE</td>
<td>Skewness of H/I</td>
</tr>
<tr>
<td>CRR</td>
<td>Canopy relief ratio</td>
</tr>
<tr>
<td>HiTH/iTH</td>
<td>$i^{th}$ percentile of H/I</td>
</tr>
</tbody>
</table>

Note: H=height, I=Intensity. In the percentiles $i = 1, 5, 10, 15...80, 90, 95$ and 99. The metrics were computed separately for both the first and the last echoes.
Table 3. Estimates of the fixed effects of the constructed multivariate model system.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Basal area (m² ha⁻¹)</th>
<th>Merchantable volume (m³ ha⁻¹)</th>
<th>Sawlog volume (m³ ha⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.024 (3.525)</td>
<td>74.313 (22.503)</td>
<td>30.087 (13.415)</td>
</tr>
<tr>
<td>f_H60</td>
<td>1.186 (0.512)</td>
<td>-2.296 (3.724)</td>
<td>-8.488 (2.304)</td>
</tr>
<tr>
<td>f_H60²</td>
<td>0.054 (0.018)</td>
<td>1.172 (0.141)</td>
<td>0.807 (0.100)</td>
</tr>
<tr>
<td>f_HSD</td>
<td>-2.115 (0.243)</td>
<td>-12.182 (1.823)</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Standard errors of the parameters are shown in parentheses. “f_” denotes that the variable has been derived from the first ALS echoes only. H60 = 60th percentile of height, HSD = standard deviation of height.
Table 4. The estimated variance-covariance matrix of random effects of the model system.

<table>
<thead>
<tr>
<th></th>
<th>Intercept of Basal area</th>
<th>Intercept of Merchantable volume</th>
<th>Intercept of Sawlog volume</th>
<th>H60² of Sawlog volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept of Basal area</td>
<td>3.322²</td>
<td>66.834</td>
<td>-5.542</td>
<td>0.139</td>
</tr>
<tr>
<td>Intercept of Merchantable volume</td>
<td>0.936</td>
<td>21.487²</td>
<td>-49.788</td>
<td>1.237</td>
</tr>
<tr>
<td>Intercept of Sawlog volume</td>
<td>-0.226</td>
<td>-0.314</td>
<td>7.390²</td>
<td>-1.325</td>
</tr>
<tr>
<td>H60² of Sawlog volume</td>
<td>0.233</td>
<td>0.321</td>
<td>-1.000</td>
<td>0.179²</td>
</tr>
</tbody>
</table>

Note: Variances on the diagonal (italics), covariances in upper right triangle, and correlations in the lower left triangle.
Table 5. The estimated cross-model correlations of residual errors (top two rows) and parameters for the applied power-type variance function (bottom two rows) (see Appendix A for definitions).

<table>
<thead>
<tr>
<th></th>
<th>Basal area</th>
<th>Merchantable volume</th>
<th>Sawlog volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal area</td>
<td>1</td>
<td>0.963</td>
<td>0.583</td>
</tr>
<tr>
<td>Merchantable volume</td>
<td>0.963</td>
<td>1</td>
<td>0.695</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.152$^2$</td>
<td>2.079$^2$</td>
<td>1.316$^2$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.772</td>
<td>1.363</td>
<td>1.303</td>
</tr>
</tbody>
</table>
**Figure 1.** An example run of a grid layout. A: Harvester positioned trees and corresponding stand borders created with alpha shape. B: A 15 m × 15 m grid laid atop a stand. C: Example area of 30 m × 30 m (shown in B as well) illustrating the grid cell-specific number of 1 m × 1 m cells that intersect the stand. Only grid cells with at least 215 intersecting 1 m × 1 m cells were included in the analysis. D: Included cells and the boundary of the original clear-cut stand. The map was created in the R software using the harvester-based tree positions (UTM zone 32N coordinate reference system).
**Figure 2.** An example run of a sample plot layout, where 10 plots are arranged randomly within the clear-cut stand. The dashed line illustrates the 5 m inner buffer zone which plots were not allowed to intersect. Outer circles around plot numbers 1–9 show the banned area for the next plot centers, and inner circles around all plots describe the actual plots with $r = 8.46$ m. Grey dots up to 30 m outside the stand describe the randomly generated trees. The map was created in the R software using the harvester-based tree positions (UTM zone 32N coordinate reference system).
Figure 3. Different random effects (n=33) of the multivariate model system plotted against each other. Correlations are also shown.
Figure 4. Pearson residuals (n=638). 1) Basal area vs. Merchantable volume, 2) Basal area vs Sawlog volume.
Figure 5. Basal areas (m$^2$ ha$^{-1}$) of fixed radius (8.46 m) vs. angle gauge measurements for all 75,000 simulated plots including generated trees. The darker the area in the plot, the more observations.
Figure 6. Root mean square error (RMSE%) and relative mean difference (MD%) values for all responses for the different numbers of plots used in the angle gauge calibrations.
Figure 7. Change in relative error (i.e. \( \frac{\text{observed}-\text{predicted}}{\text{observed}} \times 100 \)) of predicted merchantable volume of a stand when 1–10 angle gauge plots are used instead of the fixed effects of the model only (7,500 observations for each box). Above the \( y = 0 \) line the calibrated prediction is more accurate than the prediction based only on the fixed part of the model. In addition, variances are provided numerically above the boxes.
APPENDIX A

Here we provide an example of the calibration procedure. The same principle is described with more theory in Appendix A in Maltamo et al. (2012). An R-script example based on the study of Maltamo et al. (2012) is also provided in Mehtätalo and Lappi (2020).

Let us consider that we have a seemingly unrelated system of mixed-effects models including component models for basal area, merchantable volume, and sawlog volume. Each of the models was of form

\[ y_{ij}^{(k)} = x_{ij}^{(k)} \beta^{(k)} + z_{ij}^{(k)} b_i^{(k)} + \epsilon_{ij}^{(k)} \]

where \(i\) indicates the group, \(j\) indicates the member \(j\) of the group \(i\), \(k\) indicates the response \((k=1\) refers to basal area, \(k=2\) to merchantable volume and \(k=3\) to sawlog volume\), \(y\) is the value of the response, \(x\) is the vector including the values of the predictors for the \(j\)th observation, \(\beta\) is the vector including the regression coefficients, \(z\) is the vector of predictors in the random part, \(b\) is the vector of random effects, and \(\epsilon\) is the residual for the observation \(j\). If the model includes only the random intercept, then \(z = 1\) and the length of \(b\) is 1.

Furthermore, the model for group \(i\) can be written as

\[ y_i^{(k)} = X_i^{(k)} \beta^{(k)} + Z_i^{(k)} b_i^{(k)} + \epsilon_i \]

where the matrices are as illustrated in Mehtätalo and Lappi (2020). The predictors of both basal area and merchantable volume models are \(H60\) (60th percentile of height of airborne laser scanning [ALS] echoes), \(H60^2\) and \(HSD\) (standard deviation of heights of ALS echoes), and the predictors of sawlog volume model are \(H60\) and \(H60^2\). Basal area and merchantable volume models have only a random intercept, whereas the sawlog volume model also includes a random slope for the predictor \(H60^2\), in addition to the random intercept. A power-type variance function is also applied using the predictor \(H60\) of all component models:
\[ \text{Var}(\epsilon_{ij}) = \sigma^2 |H60|^{2\delta} \]

where \( \epsilon_{ij} \) is the within-group error, \( \sigma^2 \) is a scaling factor for the residual variance, H60 is the selected variance covariate, and \( \delta \) is the unrestricted variance parameter indicating the power (Pinheiro and Bates 2000, Pinheiro et al. 2019). Therefore, the possible difference between variances of the two different errors depends only on the variance covariate H60 of the observations. A general correlation structure, where

\[ \text{cor}(\epsilon_{ij}^{(k)}, \epsilon_{ij}^{(l)}) \]

was assumed separately for each pair \((k, l)\) of responses, was used to model the cross-model correlation of the components. Functions “varPower” and “corSymm” were used to implement these model components, see Mehtätalo and Lappi (2020) for an example.

Let us now consider the matrices and vectors needed for the model system. For all data, the model system can be written in the following form:

\[ \mathbf{y} = \mathbf{X}\beta + \mathbf{Zb} + \mathbf{e} \]

by defining:

\[ \mathbf{y} = [ \mathbf{y}^{(1)} \quad \mathbf{y}^{(2)} \quad \mathbf{y}^{(3)} ]' \]

\[ \mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} & 0 & 0 \\ 0 & \mathbf{X}^{(2)} & 0 \\ 0 & 0 & \mathbf{X}^{(3)} \end{bmatrix} \]

\[ \mathbf{\beta} = [ \mathbf{\beta}^{(1)} \quad \mathbf{\beta}^{(2)} \quad \mathbf{\beta}^{(3)} ]' \]

\[ \mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(1)} & 0 & 0 \\ 0 & \mathbf{Z}^{(2)} & 0 \\ 0 & 0 & \mathbf{Z}^{(3)} \end{bmatrix} \]

\[ \mathbf{b} = [ \mathbf{b}^{(1)} \quad \mathbf{b}^{(2)} \quad \mathbf{b}^{(3)} ]' \]
and:

\[ \mathbf{e} = [ e^{(1)\prime} e^{(2)\prime} e^{(3)\prime} ]' \]

where \( \mathbf{y} \) includes the vectors of responses for each component model, \( \mathbf{X} \) is a block-diagonal matrix including the predictors of component models in different columns, \( \mathbf{\beta} \) includes the vectors of regression coefficients for each component model, \( \mathbf{Z} \) is a block-diagonal matrix where each \( \mathbf{Z}_i \) is the model matrix of the random parts of the component model in question, \( \mathbf{b} \) includes the random effects vectors for each three component model, and \( \mathbf{e} \) includes the vectors of residual errors of each three component model.

Moreover, let us assume that only the basal area information has been observed from \( p \) angle gauge plots from the stand of interest. The aim is to predict the random stand effects vector \( \mathbf{b} \) by employing estimated best linear predictor (EBLUP) (Mehtätalo and Lappi 2020):

\[ \text{EBLUP}(\mathbf{b}) = \mathbf{C}_o \mathbf{Z}_o \mathbf{D}_o \mathbf{Z}_o' + \mathbf{R}_o)^{-1} (\mathbf{y}_o - \mathbf{X}_o \mathbf{\beta}_o) \]

Structures of these vectors and matrices in our case are described and illustrated next. \( \mathbf{C} \) and \( \mathbf{D}_o \) are based on the random effects variance-covariance matrix \( \mathbf{D} \) (see Pinheiro and Bates 1996 for parametrization of a variance-covariance matrix), which is of dimension 4 × 4, as follows:

\[
\begin{bmatrix}
\text{var}(C_1) & \text{cov}(C_2, C_1) & \text{cov}(C_3, C_1) & \text{cov}(C_4, C_1) \\
\text{cov}(C_1, C_2) & \text{var}(C_2) & \text{cov}(C_3, C_2) & \text{cov}(C_4, C_2) \\
\text{cov}(C_1, C_3) & \text{cov}(C_2, C_3) & \text{var}(C_3) & \text{cov}(C_4, C_3) \\
\text{cov}(C_1, C_4) & \text{cov}(C_2, C_4) & \text{cov}(C_3, C_4) & \text{var}(C_4)
\end{bmatrix}
\]

where \( C_1 \) is the random intercept of basal area, \( C_2 \) is the random intercept of merchantable volume, \( C_3 \) is the random intercept of sawlog volume, and \( C_4 \) is the random slope for H60² predictor of sawlog volume. \( \mathbf{C} \) is obtained by dropping all columns of \( \mathbf{D} \) that correspond to the unobserved response variable and \( \mathbf{D}_o \) is furthermore obtained by dropping all rows of \( \mathbf{C} \) that include ...
any information of unobserved response variable. Because only basal area has been observed, 
\( C \) and \( D_o \) have the following structures:

\[
C = \begin{bmatrix} \text{var}(C_1) \\ \text{cov}(C_1, C_2) \\ \text{cov}(C_1, C_3) \\ \text{cov}(C_1, C_4) \end{bmatrix}, 
D_o = [\text{var}(C_1)]
\]

\( Z_o \) is the random part that corresponds to the observed variables. The basal area model included 
only the random intercept, so in our case \( Z_o \) has only one column with \( p \) rows of ones:

\[
Z_o = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{p \times 1}
\]

\( R_o \) is the estimated variance-covariance matrix of residuals of observed response. Because only 
the basal area has been observed from \( p \) plots, \( R_o \) is of dimension \( p \times p \) and the off-diagonal 
elements are zeroes. The diagonal elements depend on the observed ALS predictor \( H60 \) for the 
plot in question, and on the parameters \( \sigma^2 \) and \( \delta \) that are estimated in accordance to the applied 
power-type variance function when the model system is fitted.

\[
R_o = \sigma^2 \begin{bmatrix} \lvert H60 \rvert^{2\delta} & 0 & \cdots & 0 \\ 0 & \lvert H60_2 \rvert^{2\delta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lvert H60_p \rvert^{2\delta} \end{bmatrix}
\]

Vector \( y_o \) includes the measured basal areas from the \( p \) plots, \( X_o \) includes the fixed predictors 
from the observed \( p \) plots for basal area model, and \( \beta_o \) includes the estimated fixed regression 
coefficients of the observed basal area model (see Table 3).

\[
y_o = \begin{bmatrix} y_{1}^{(1)} \\ y_{2}^{(1)} \\ \vdots \\ y_{p}^{(1)} \end{bmatrix}
\]
\[
\mathbf{X}_o = \begin{bmatrix}
1 & H60_1 & H60^2_1 & HSD_1 \\
1 & H60_2 & H60^2_2 & HSD_2 \\
\vdots & \vdots & \vdots & \vdots \\
1 & H60_p & H60^2_p & HSD_p \\
\end{bmatrix}
\]

\[
\mathbf{\beta}_o = \begin{bmatrix}
\beta_1^{(1)} \\
\beta_2^{(1)} \\
\beta_3^{(1)} \\
\beta_4^{(1)}
\end{bmatrix}
\]

A vector of predicted stand effects \( \mathbf{b} \) (length = 4) results when the above described vectors and matrices are inserted in the EBLUP equation. Locally calibrated predictions are provided for the stand in question when these predicted stand effects are used together with the fixed part of the model system.

REFERENCES


**APPENDIX B**

**Table B1.** Stand-level root mean square error (RMSE%) values in the different datasets and in the different calibration cases with varying number of plots. For the results of ALS-based calibrations (F + n plots): Angle gauge / fixed radius plot. On the lowest row, the RMSE% values when the estimates were derived as the mean observation from ten fixed radius plots (also mean of 500 repeats).

<table>
<thead>
<tr>
<th>Dataset/scenario</th>
<th>Basal area</th>
<th>Merchantable volume</th>
<th>Sawlog volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training stands (F)</td>
<td>14.75</td>
<td>14.17</td>
<td>41.08</td>
</tr>
<tr>
<td>Training stands (F+R)</td>
<td>10.40</td>
<td>12.27</td>
<td>15.62</td>
</tr>
<tr>
<td>Validation stands (F)</td>
<td>16.80</td>
<td>15.80</td>
<td>22.17</td>
</tr>
<tr>
<td>V: F + 1 plot</td>
<td>15.58 / 15.36</td>
<td>15.04 / 14.92</td>
<td>22.06 / 22.17</td>
</tr>
<tr>
<td>V: F + 2 plots</td>
<td>14.54 / 14.16</td>
<td>14.40 / 14.19</td>
<td>21.93 / 22.15</td>
</tr>
<tr>
<td>V: F + 5 plots</td>
<td>12.22 / 11.49</td>
<td>12.97 / 12.57</td>
<td>21.78 / 22.22</td>
</tr>
<tr>
<td>V: F + 6 plots</td>
<td>11.62 / 10.78</td>
<td>12.60 / 12.14</td>
<td>21.75 / 22.28</td>
</tr>
<tr>
<td>V: F + 7 plots</td>
<td>11.17 / 10.19</td>
<td>12.32 / 11.79</td>
<td>21.70 / 22.34</td>
</tr>
<tr>
<td>V: F + 8 plots</td>
<td>11.01 / 9.98</td>
<td>12.22 / 11.66</td>
<td>21.67 / 22.36</td>
</tr>
<tr>
<td>V: F + 9 plots</td>
<td>10.76 / 9.62</td>
<td>12.05 / 11.43</td>
<td>21.62 / 22.38</td>
</tr>
<tr>
<td>V: F + 10 plots</td>
<td>10.48 / 9.26</td>
<td>11.88 / 11.22</td>
<td>21.61 / 22.41</td>
</tr>
<tr>
<td>Field data from 10 plots</td>
<td>2.57</td>
<td>2.73</td>
<td>4.24</td>
</tr>
</tbody>
</table>

**Note:** F = fixed effects, R = random effects.
Table B2. Stand level relative mean difference (MD%) values in the different datasets and in the different calibration cases with varying number of plots. For the results of ALS-based calibrations (F + n plots): Angle gauge / fixed radius plot. On the lowest row, the MD% values when the estimates were derived as the mean observation from ten fixed radius plots (also mean of 500 repeats).

<table>
<thead>
<tr>
<th>Dataset/scenario</th>
<th>Basal area</th>
<th>Merchantable volume</th>
<th>Sawlog volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training stands (F)</td>
<td>2.65</td>
<td>2.95</td>
<td>-0.91</td>
</tr>
<tr>
<td>Training stands (F+R)</td>
<td>2.65</td>
<td>2.95</td>
<td>2.47</td>
</tr>
<tr>
<td>Validation stands (F)</td>
<td>-10.90</td>
<td>-9.10</td>
<td>9.33</td>
</tr>
<tr>
<td>V: F + 3 plots</td>
<td>-9.00 / -8.32</td>
<td>-7.90 / -7.47</td>
<td>10.87 / 11.33</td>
</tr>
<tr>
<td>V: F + 4 plots</td>
<td>-8.54 / -7.71</td>
<td>-7.61 / -7.09</td>
<td>11.23 / 11.80</td>
</tr>
<tr>
<td>V: F + 6 plots</td>
<td>-7.84 / -6.66</td>
<td>-7.17 / -6.42</td>
<td>11.80 / 12.62</td>
</tr>
<tr>
<td>V: F + 7 plots</td>
<td>-7.61 / -6.25</td>
<td>-7.02 / -6.16</td>
<td>12.00 / 12.94</td>
</tr>
<tr>
<td>V: F + 8 plots</td>
<td>-7.50 / -6.06</td>
<td>-6.95 / -6.04</td>
<td>12.10 / 13.10</td>
</tr>
<tr>
<td>V: F + 9 plots</td>
<td>-7.38 / -5.81</td>
<td>-6.87 / -5.89</td>
<td>12.21 / 13.30</td>
</tr>
<tr>
<td>V: F + 10 plots</td>
<td>-7.24 / -5.57</td>
<td>-6.79 / -5.73</td>
<td>12.33 / 13.50</td>
</tr>
<tr>
<td>Field data from 10 plots</td>
<td>0.63</td>
<td>-0.09</td>
<td>-1.44</td>
</tr>
</tbody>
</table>

Note: F = fixed effects, R = random effects.