Forest management inventories assisted by airborne laser scanning (ALS) can be used to predict different forest attributes. These predictions are utilized in practical forestry, but in the case of timber assortment-specific volumes, the ALS-based predictions can be inaccurate. This causes uncertainty in harvest planning. However, ALS-based predictions can be calibrated to achieve greater accuracy with local measurements. In this study, we used ALS data and accurately positioned cut-to-length harvester measurements from Norway spruce (Picea abies (L.) Karst.) dominated clear-cuts. We fitted linear mixed-effects (LME) models with exponential correlation structure for merchantable volume and sawlog volume for 225 m$^2$ cells. Our aim was to study the effect of local harvester measurements on the accuracy of stand level merchantable and sawlog volumes. LME-based predictions were calibrated repeatedly up to 40 times as the cutting progressed. ALS data and harvester measurements were used to predict both the random effects and residual errors for each validation unit. At best, relative root mean square error (RMSE%) of initial predictions of 15.4 per cent for merchantable volume and 22.1 per cent for sawlog volume were reduced to 4.1 and 5.3 per cent, respectively, when measurements from 40 harvested cells of size 225 m$^2$ were used. These results suggest that spatially accurate harvester data could be utilized during harvesting to increase the accuracy of volume and timber assortment predictions.

**Introduction**

In many countries, forest management inventories assisted by airborne laser scanning (ALS) are carried out to produce predictions for various stand level forest attributes such as the total volume (Næsset, 2014). These ALS-based predictions are utilized, for example, when timber assortment volumes are predicted for the marked stands. However, for individual stands, the accuracy of such predictions may be low (Holopainen et al., 2010). Therefore, different calibration procedures based on field measurements have been proposed (Korhonen et al., 2019; Karjalainen et al., 2020; Maltamo et al., 2012). These calibrations are usually based on linear mixed-effects (LME) models that are locally adjusted by predicting the stand effects with local observations. Usually, the aim of such calibrations has been to increase the accuracy of predictions for the standing stock. More accurate pre-harvest information about, for example, the tree quality or the sawlog volume would ease the scheduling and overall planning of silvicultural and harvesting operations. With spatially or temporally correlated data, the prediction can be further improved by also predicting the residual errors of the model (Mehtätalo and Lappi, 2020), but that possibility has not been previously used in this context.

In addition to obtaining more accurate pre-harvest information, calibrations can be done also during harvesting by using the data collected by cut-to-length (CTL) harvesters. That is, the predictions for the currently uncut part of a stand are successively re-calibrated with the harvester measurements as the harvesting is progressing through the stand. A similar application was tested by Uusitalo et al. (2006), who aimed to estimate the stand diameter distribution by using the measurements of the first 30, 90 and 150 harvested trees. More accurate information about the diameter distribution could potentially be useful, for example, in the tree bucking control. According to Uusitalo et al. (2006), the reliability of predicted diameter distributions seemed to increase if harvester measurements were combined with prior information. However, without prior information, the process resulted in inaccurate predictions.

© The Author(s) 2021. Published by Oxford University Press on behalf of Institute of Chartered Foresters. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Along with bucking optimization, accurately positioned CTL harvester data also have the potential to increase the efficiency of harvesting operations. In CTL harvesting operations, logs need to be cut into the correct dimensions during the harvesting operation. Consequently, there may be numerous different timber assortments (e.g., species- and quality-specific sawlogs, species-specific pulpwood, energy wood) that are being cut simultaneously from a stand. At the intermediate storage at the roadside, these different timber assortments are stacked in different piles because they are often transported to specific processing facilities. Sometimes, the amount of suitable space for the storage may be very limited (Metsätalo, 2010) which may require new temporary spots for log piles to be found or the logs at the roadside to be transported to a different location during harvesting. Therefore, predictions of accruals of different timber assortments during harvesting may improve the logistics.

In addition, in clear-cut stands where small diameter trees are rare and there exist some estimate of mean diameter, volume of trees is a key indicator for the productivity of a CTL harvester (Eriksson and Lindroos, 2014). Therefore, improved prediction of the total merchantable volume in a stand would also help to estimate the total time needed for harvesting. This, again, helps the harvesting organizations to plan and schedule consecutive operations in different stands more effectively so that unnecessary idle time of expensive machinery can be kept at a minimum. In smaller stands, the advantages of calibrations during the harvesting would be only minor, but in larger stands where the cutting takes multiple days, the work effectiveness could be improved by using calibrated predictions.

The proposed in-situ calibrations are made possible by the advanced technology that is mounted on modern CTL harvesters. The computer in a harvester records, among other things, the volume and timber assortment for each log. Therefore, harvesters provide an easy way to measure attributes such as merchantable volume and sawlog volume that are time-consuming to measure manually (see, e.g., Karjalainen et al., 2019 for visual bucking). However, a prerequisite for such calibrations that combine harvester and ALS data is that the trees can be positioned accurately, preferably with positioning errors less than 1 m. So far, the inaccuracy in the positioning of the harvested trees has limited the usability of harvester measurements together with ALS data. For example, the position for the harvested trees has sometimes been determined as the position of the harvester itself at the time of felling which means that the movement of the boom and the harvester head has been ignored (Bollandsdás et al., 2011). The large potential of spatially accurate harvester data (Lindroos et al., 2015) has sparked the development of such positioning systems that take the movement of the boom into account (Hauglin et al., 2017; Melkas and Riekki, 2017). For example, in the study by Hauglin et al. (2018), sub-metre accuracy was obtained for the harvested trees. However, systems providing sub-metre accuracies for positions of harvested trees are not commonly available yet, so more development is still needed.

The aim of this study was to test the effects of in-situ calibrations of LME models on the accuracy of stand level merchantable and sawlog volume (in cubic metres per hectare) predictions. The calibrations were carried out while the cutting was in progress in the stand in question, and they were based on ALS data and observations collected with a CTL harvester from different numbers of 15 m x 15 m (225 m²) grid cells tessellating the harvested stands. In the calibrations of LME models, both the random effects and residual errors were predicted for each validation unit. For comparison, the use of the mean values of the harvester measurements as stand level volume estimates was also evaluated.

**Methods**

In the current study, the pre-processed field data and ALS data as used by Karjalainen et al. (2020) were subject to analysis. The method used to acquire the data is summarized below; full details can be found in Karjalainen et al. (2020).

**Study area**

The study area is located in southeastern Norway (60°25’ N, 11°4’ E) in Romerike region (Figure 1), and the forests are dominated by Norway spruce (Picea abies (L.) Karst.) (87.0 per cent of total measured merchantable volume). Scots pine (Pinus sylvestris L.) (7.3 per cent) and deciduous species (5.7 per cent) are also abundant in the area.

**Field data and data procession**

The field data were collected with a John Deere 1270E CTL harvester in 2017. The harvester was equipped with a system that provided sub-metre accuracy for the position of each harvested tree. In accordance with the StanForD2010 forest data standardization (Skogforsk, 2018), the harvester recorded the volume of all logs (‘merchantable volume’) and the volume of those logs that were qualified as sawlogs (‘sawlog volume’) for each harvested tree. The deciduous trees were not classified as sawlogs which is consistent with common practice in Norway. The data consisted of 48 clear-cut stands.

The field data included only the tree positions, not any geometries for stand borders. As our intention was to use the ALS data with the area-based approach, and to produce wall-to-wall predictions, the stands needed to be tessellated into grid cells first. We opted to use a cell size of 15 m x 15 m, which is close to the size that is commonly used in operational ALS-assisted management inventories in Norway (Næsset, 2014) and Finland (Metsäkeskus, 2021). The R software (R Core Team, 2017) was used in the process. First, we generated realistic stand borders by aggregating the point-wise tree data to polygons by creating 2-D alpha shapes using the alphashape R package (Pateiro-López and Rodriguez-Casal, 2015) with an alpha value of 10. The same procedure with the same alpha value was used also in the studies of Hauglin et al. (2018) and Maltamo et al. (2019). Stand borders were necessary to determine which grid cells were ‘acceptably within’ the clear-cut area. A cell was ‘acceptably within’ if at least 215 (of total 225) 1-m² sub-cells intersected the underlying clear-cut area created with the alpha shape. All the accepted cells were considered equal, i.e., no weighting was applied (Karjalainen et al., 2020). We also wanted to maximize the number of cells in the data set, so in each stand, instead of accepting the first selected position of the grid, an optimal position for the grid was found through an iterative process where the number of accepted cells was maximized (Karjalainen et al., 2020). After
the optimal position for the grid was found, the harvested trees were assigned to grid cells by using their harvester head-based positions. For each cell, the merchantable volume (m$^3$ ha$^{-1}$) and the sawlog volume (m$^3$ ha$^{-1}$) were calculated.

The entire data set consisted of 2292 cells distributed on the 48 stands. The data set was split into training and validation stands so that the 16 largest stands according to area, excluding the second largest stand, were chosen as validation data and the rest of the stands (including the second largest stand) composed the training data. Such a split of the data was adopted because the calibrations implemented in Karjalainen et al. (2020) were more reasonable for large stands than in stands that consist of only a few 15 m × 15 m grid cells. The same justification applied in this study, too, so we did not make any changes to training and validation stands. The second largest stand (285 cells) was included in the training data to provide observations also from one of the large stands in the model fitting. The smallest validation stand consisted of 41 cells (0.92 ha), whereas the largest validation stand consisted of 453 cells (10.19 ha). Mean values for the training stands and stand-specific information for the validation stands are provided in Table 1. For detailed information about the cells within the training and validation stands, see Table 1 in Karjalainen et al. (2020).

ALS data

The ALS data were acquired in 2013 using a Leica ALS70 instrument operated from an average flying altitude of 3000 m above ground level. The pulse repetition frequency was 104.6 kHz, and the resulting point density was about 0.7 points m$^{-2}$ at ground level. The classification algorithm proposed by Axelsson (1999) was used to separate ground hits and vegetation hits. Moreover, as the used ALS instrument could record multiple echoes for each pulse, two echo categories were constructed and used in the analysis, namely first (first of many + only) and last (last of many + only) echoes. Finally, the ALS echoes were extracted for each cell and the ALS metrics were calculated separately for both echo categories using the LASmetrics function in rLiDAR package (Silva et al., 2017). The derived ALS metrics included all the common metrics such as the maximum, median, mean, standard deviation and the variance of heights of the ALS echoes. In addition, different height percentiles (1, 5, 10, …, 90, 95, 99) were calculated.

LME models and their calibration

LME models were used as they are able to take the hierarchical structure of the data (i.e. grid cells within stands) into account. Furthermore, LME models also allowed to calibrate the models for individual stands. The optimal structure (including the predictors, random part, variance function and correlation structure) of the models for both merchantable and sawlog volumes was carefully examined using the nlme package (Pinheiro et al., 2019) in the R software (R Core Team, 2017). See Appendix for formal definition of the model. For merchantable and sawlog volumes, the same univariate model forms as were constructed by Karjalainen et al. (2020) were used also in the current study, with the exception that for both models, the general correlation structure was changed to a better justified exponential spatial correlation structure, which also allows more efficient use of the calibration measurements through prediction of the residual errors.

Local observations make it possible to predict the random effects for all the responses, and thus, to calibrate the models in each stand. In addition, also the residual errors can be predicted by utilizing the spatial autocorrelation of the data (Mehtätalo and Lappi, 2020). If the residual errors are spatially correlated,
then their prediction should increase the accuracy of predictions for the cells that are located close to the calibration cells. Three different components can be summed up to obtain total predictions: (1) predictions based on the fixed part of the model, (2) the predicted random effects and (3) the predicted residual errors. When local observations are not available, only the fixed part of the model can be used.

In this study, calibrations were based on the harvester measurements from the grid cells. Predictions were always updated after a new 15 m × 15 m grid cell was clear-cut. By using the observed merchantable and sawlog volumes and the known ALS metrics, random effects for the entire group (i.e. all the cells within the same stands) could be predicted by using the Estimated Best Linear Unbiased Predictor (EBLUP). The observed residuals of the harvested cells were utilized to predict the residual errors for the uncut cells in the stand. The whole calibration procedure is described in detail in Appendix.

Selection of calibration cells

The sizes of the validation stands ranged from 41 to 453 cells. We wanted to have the same number of stands for each calibration with different number of calibration cells, so the maximum number of calibration cells was 41. We opted to test the calibration by using 1–40 cells in the calibration in each stand. For the smallest stand, this meant that almost the entire stand was already clear-cut after the 40th cell was harvested. In the largest stand, 40 cells cover only 8.8 per cent of the total area.

Two different strategies for choosing the calibration cells were tested: (1) using the actual cutting order based on the time stamp of each harvested tree and (2) using a strategy by which one main strip road is first harvested, and then subsequently, the rest of the area is harvested by creating smaller loops from the main strip road. In the current study, we refer to this latter strategy as the ‘backline method’. The main advantage of the backline method in practical forestry is to direct most of the forest machine traffic into main strip road placed on a sturdy ground, whereas more vulnerable and soft soils are less trafficked. This method is quite commonly used in operational harvesting, for example, in Finnish forestry (Uusitalo, 2003). Here, the cells for the backline method were subjectively chosen so that the cells composed a clear main strip road along the longest axis, and possibly some loops until the threshold of 40 cells was reached. The backline method was included in the study to provide a plausible comparison for the actual cutting order. It appeared that in most of the stands, the actual cutting order moved forward rather systematically from one corner of the stand to the remaining areas without any clearly observable main strip road. The two strategies for different cutting orders are illustrated in Figure 2. Note that the results of this study were obtained with only one set of calibration cells for both strategies.

Every time the model was re-calibrated, predictions for all the cells in the stand were updated, that is, also the predictions for the already harvested cells changed. Our interest was to improve the predictions of stand level merchantable and sawlog volumes as the clear-cutting progressed. Therefore, it was justified to fix all the predictions of the already harvested cells to the correct values that were observed with the harvester. Consequently, with this approach, the errors of predictions would have eventually become zero if the re-calibration would had continued until the stand had been completely harvested. For one of the small stands (stand #15), this was close to appear as only a single cell remained uncut.

In addition to the calibration that utilizes the ALS data, we calculated the results also by determining the stand level merchantable and sawlog volumes directly as the observed mean of the harvested cells. Similarly, the predictions were always updated after a new cell was completed harvested. For example,

| ID | n cells | Cell area (ha) | Stand area (ha) | Merch. V (m³ ha⁻¹) | Sawlog V (m³ ha⁻¹) |
|----|---------|---------------|----------------|-----------------|-------------------|
| Mean T | 19.3 | 0.4 | 0.7 | 290.3 | 104.2 |
| 1 | 453 | 10.2 | 11.4 | 294.7 | 82.1 |
| 2 | 166 | 3.7 | 5.0 | 230.0 | 129.9 |
| 3 | 136 | 3.1 | 4.2 | 248.4 | 116.9 |
| 4 | 117 | 2.6 | 3.3 | 275.8 | 147.6 |
| 5 | 107 | 2.4 | 3.0 | 291.9 | 95.6 |
| 6 | 100 | 2.3 | 3.0 | 272.1 | 152.3 |
| 7 | 81 | 1.8 | 2.5 | 333.8 | 188.2 |
| 8 | 78 | 1.8 | 2.6 | 227.5 | 146.9 |
| 9 | 76 | 1.7 | 2.2 | 210.8 | 118.2 |
| 10 | 74 | 1.7 | 2.3 | 301.4 | 159.4 |
| 11 | 63 | 1.4 | 1.8 | 354.7 | 162.0 |
| 12 | 59 | 1.3 | 1.8 | 332.9 | 151.6 |
| 13 | 54 | 1.2 | 1.8 | 389.3 | 165.3 |
| 14 | 49 | 1.1 | 1.6 | 286.9 | 119.1 |
| 15 | 41 | 0.9 | 1.2 | 349.3 | 169.4 |
| Mean | 110.3 | 2.5 | 3.2 | 293.3 | 140.3 |

n cells = number of the cells, Cell area (ha) = the total area covered by the cells, Stand area (ha) = the total area of the stand created with the alpha shape, Merch. = merchantable, V = volume, T = training stands.
if merchantable volumes of the three first harvested cells were 200, 220 and 150 m$^3$ ha$^{-1}$, then stand level merchantable volume prediction would have been 200, 210 and 190 m$^3$ ha$^{-1}$ after harvesting of the first, second and third cell, respectively.

### Accuracy assessment

Our interest was in stand level predictions as the clear-cutting progressed. Therefore, the cell level predictions were aggregated to stand level prior to accuracy assessment. The accuracy of stand level predictions was assessed using empirical relative root mean squared error (RMSE%) (Eq. 1) and relative mean difference (MD%) (Eq. 2).

\[
\text{RMSE\%} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \times \frac{100}{\bar{y}}
\]

(1)

\[
\text{MD\%} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n} \times \frac{100}{\bar{y}},
\]

(2)

where $n$ is the number of stands in the dataset, $y_i$ is the observed merchantable/sawlog volume for stand $i$, $\hat{y}_i$ is the predicted merchantable/sawlog volume for stand $i$ and $\bar{y}$ is the measured mean of the merchantable/sawlog volume in the dataset. In addition, plots showing the error of predictions on the y-axis and the relative proportion of the harvested stand on the x-axis were visually inspected.

### Results

#### LME models

The univariate models that were constructed for the multivariate model system in Karjalainen et al. (2020) were used in the current study with the exception that exponential spatial correlation structure was now applied. Fixed effects for the univariate models are provided in Table 2. Variances for the random parts of the models and the estimated parameters for the power-type variance function and the exponential spatial correlation structure are provided in Table 3.

#### Calibration of merchantable and sawlog volumes

The RMSE% and MD% values of predicted merchantable and sawlog volumes for actual cutting order and the backline method are illustrated in Figure 3. The corresponding RMSE% and MD% values for the cases where 1, 10, 20, 30 and 40 cells were used to calibrate the predictions are provided in Tables 1 and 2 in Supplementary data for merchantable volume and sawlog volume, respectively.

The RMSE% and MD% values of predictions clearly improved due to calibration. For the LME model, the calibration that also utilized the prediction of residual errors proved to be more accurate than the calibration based on the prediction of random effects only. At best, the RMSE% values of LME-based merchantable

---

**Figure 2** The selection of calibration cells with the actual cutting order and backline method for stand #9. Note that in the actual cutting order, the cells are in partly and seemingly illogical order because the cells were not picked before the last tree within the 15 m × 15 m cell was harvested.

**Table 2** Estimates of the fixed effects of the constructed univariate models.

| Variable | Merchantable volume (m$^3$ ha$^{-1}$) | Sawlog volume (m$^3$ ha$^{-1}$) |
|----------|--------------------------------------|---------------------------------|
| Intercept | 131.770 (29.557)                     | 54.704 (15.107)                 |
| $f_{H60}$ | -7.520 (5.156)                       | -12.819 (2.597)                 |
| $f_{H60}^2$ | 1.371 (0.196)                       | 0.984 (0.112)                  |
| $f_{HSD}$ | -17.634 (3.003)                      | -                            |

Note: Standard errors of the parameters are shown in parentheses. ‘$f_-$’ denotes that the variable has been derived from the first ALS echoes only. $H60 = 60$th percentile of height, $HSD = $ standard deviation of height.
volume and sawlog volume predictions decreased from 15.4 to 4.1 per cent and from 22.1 to 5.3 per cent, respectively.

In most of the cases, the calibration based on both ALS and harvester data resulted in smaller RMSE% values than the calibration based on the harvester data alone. The only exception was the prediction of merchantable volume with the backline method. In that case, the calibration based only on the harvester data was eventually slightly more accurate than the LME-based calibration that also utilized the prediction of residual errors. When less than 16 cells were used in the calibration, the LME-based calibration was clearly more accurate also in this case.

The lines illustrating the actual cutting order and backline method were somewhat similar in Figure 3, especially in the case when both the harvester and ALS data were used. If only harvester data were used, then there were some clear differences especially in the MD% values: the backline method overestimated both the merchantable and sawlog volumes, particularly in the beginning of the clear-cut. However, after 40 cells were harvested, the accuracies were mostly greater with the backline method than with the actual cutting order. Especially, when also the residuals were predicted, the backline method resulted in more accurate predictions in all the cases except when \( n = 1 \) and \( n = 17 \), for merchantable volume, and \( n = 1 \) for sawlog volume (not shown). With 40 harvested cells, the differences in RMSE% values were greater than 1 per cent point for both responses.

The stand-specific results for the calibration of merchantable volume are provided in Figure 4 for the actual cutting order and in Figure 5 for the backline method. For actual cutting order (Figure 4), the calibration of the LME-model with measurements from 40 grid cells improved the predictions in all stands, except for stands #1 and #4. Especially in the stand #1, the calibration of the LME model was clearly disturbed by a single extraordinary cell: the use of the cell #32 shifted the LME-based merchantable volume prediction from \( \sim 291 \) to \( 340 \text{ m}^3\text{ ha}^{-1} \) (observed value was \( 295 \text{ m}^3\text{ ha}^{-1} \)). However, also the curve showing the mean of the harvester measurements started to recede the zero line after about 10 harvested cells, indicating that thereafter merchantable volumes of harvested cells were clearly smaller than the actual mean of the stand. In the stand #4, on the other hand, the accuracy of the LME-based calibrations decreased rather constantly as the clear-cutting progressed. Assumingly, the used calibration cells were not representative for the entire stand #4.

For the backline method (Figure 5), extraordinary cells were avoided in stand #1 and the errors of the calibrated LME-based predictions were always close to zero. On the other hand, in stand #12, there were now some clearly visible fluctuations in the predictions when about 15 cells were used in the calibration. Otherwise, the curves of ALS-based predictions were similar between the cutting approaches. Overall, regardless of cutting approach, it appears that on stand level, the calibrations based on mean of

| Table 3 Information about the random parts of the models, applied power-type variance function, and applied exponential spatial correlation structure |
|-----------------------------------------------|
|                      | Merchantable volume | Sawlog volume |
| Variance of random intercept          | 17.171\(^2\)          | 8.761\(^2\)          |
| Variance of random \( f_{H60} \)        | –                | 0.164\(^2\)          |
| Correlation between random effects     | –                | –1                  |
| varPower: \( \sigma^2 \)                | 2.600\(^2\)          | 1.260\(^2\)          |
| varPower: \( \delta \)                  | 1.285              | 1.326               |
| corExp: Range                         | 30.356             | 27.007              |
| corExp: Nugget                        | 0.549              | 0.517               |

Figure 3 RMSE% and MD% values for the predicted merchantable and sawlog volumes with different numbers of cells used in the calibration. Black solid line = calibration of LME model with both predicted random effects and residual errors; black dashed line = calibration of LME model with predicted random effects only; dotted horizontal line = the level of accuracy with the fixed effects of the LME-model; grey dashed line = mean of harvester measurements.
the harvester measurements resulted in greater variation than the calibration of the LME model as the cutting progressed, as expected by the theory (see Mehtätalo and Lappi, 2020, section 5.4.4).

When predicted volumes were compared with observed volumes in individual stands instead of calculating RMSE% and MD% values over all 15 validation stands, it turned out that the differences between cutting approaches were not as clear.
Forestry

Figure 5 The change in stand-specific errors (observed – predicted volume at stand level) of merchantable volume predictions v. the proportion of the harvested area of the total stand area. Backline method. Solid black line = calibration of LME-model with both the predicted random effects and residual errors; dashed grey line = mean of harvester measurements.

anymore. In fact, with 40 harvested cells, it appeared that the backline method was more accurate in just 8 of 15 stands for both response variables (in stands #1, 4, 5, 9, 11, 12, 13 and 14 for merchantable volume and stands #1, 2, 3, 4, 5, 11, 13 and 15 for sawlog volume, respectively). There were no clear trends with respect to the stand size, that is, the actual cutting order resulted in greater accuracy in some large stands as well as in small stands. The average cell number of those stands in
which actual cutting order was more accurate than the backline method was about 97 and 74 cells for merchantable volume and sawlog volume, respectively.

Discussion

The aim of this study was to test whether calibrations based on ALS data and harvester data from 1 to 40 grid cells of size 15 m × 15 m could be used to improve the predictions of stand level merchantable and sawlog volumes, while the cutting is in progress in the stand in question. The results showed that for both response variables, notable increases in accuracy can be expected as the clear-cutting progresses. Therefore, the proposed calibration procedure could be useful in practice. However, positioning systems providing sub-metre accuracies for the positioning of harvested trees would be required to allow practical implementation of the application examined in this study.

There were some restrictions associated with the data and methods used in this study and which may influence on the extent to which the results can be generalized. First, the results were calculated with just a single set of grid cells in the calibration and for a restricted number of stands. Allocating other grid cells and in different order might alter the results. Likewise, stands with properties different from those we analyzed may yield different results. In addition, the used harvester data only provided information regarding categorization into sawlogs or non-sawable logs. These non-sawable logs included both pulpwood and energy wood, so we were not able to distinguish the energy wood and pulpwood logs from each other. Moreover, local aerial images were not available, so we were not able to do any species-specific predictions for each timber assortments. Thus, a more thorough analysis for the need of storage space for log piles would require species-specific and timber assortment-specific predictions. Nevertheless, based on the results of this study, it is likely that at least some level of increase in the accuracies of also species-specific and timber assortment-specific predictions would likely be obtained if such data were available.

In Karjalainen et al. (2020), it was noted that the residuals and the random effects of the merchantable and sawlog volume models are correlated in the used dataset. Thus, more information would have been obtained if the calibration utilized also the cross-model correlation. However, using cross-model correlation is expected to provide only minor improvement compared with the individual model-specific calibration when both responses have been observed for all calibration cells. In addition, fitting of a spatially autocorrelated multivariate mixed-effects model is a complicated task and to the best of our knowledge, such a model cannot be fitted straightforwardly with currently available software. Our available resources did not allow to fit such a model for this study. Therefore, in this study, we opted for keeping the two models separate instead of combining them to a spatially autocorrelated seemingly unrelated model system. Anyway, fitting and calibrating a spatially autocorrelated multivariate mixed-effects model is an interesting topic for a future study.

On general level, the utilization of ALS data in the calibration resulted in greater accuracies compared with the mean of harvester measurements (Figure 3). However, on stand level (Figures 4 and 5), the more accurate calibration method varied between stands. One explanation for the generally better performance of ALS-based calibrations could be that it is not that sensitive to extremely small or large measured volumes because the predictions are shrunk towards the marginal mean given by the fixed part of the model. If the volume of a cell is small, then also the initial LME-based prediction should be small, and the residual can be either positive or negative, shifting the calibrated predictions in either direction. In the case of harvester measurements, a cell with small volume automatically reduces the predicted stand level volume so occurrence has a greater effect. The difference in accuracies between calibration methods is especially pronounced in the beginning of a clear-cutting. As assumed, the performance of the calibrations based only on the mean of the harvester measurements improved the more cells were harvested. In some stands, it happened to result in greater accuracies than the calibrations of LME models.

We tested two different strategies to choose the 1–40 cells for the calibration. From a practical point of view, the cells derived by the actual cutting order were probably the more interesting ones as they described how the cuttings are carried out in practice in Norway. However, there may be differences in practices between countries and, for example, in Finland, the presented backline method could also be applied in stands where the soil bearing capacity varies within the stand steering the main forest machine traffic into the main strip road placed on a robust ground. Therefore, the comparison of the two strategies was also interesting. In the calibrations based only on the harvester data, especially the MD% values differed between the cutting orders. The result that the backline method was prone to overestimate both the merchantable and sawlog volume predictions in the beginning of clear-cut was logical because the cells for the backline method were chosen from the central part of the stand. That is, the cells near the stand borders with presumably different volumes were not presented in the sample. Consequently, stand level volumes were here overestimated.

The results with 40 harvested cells were mostly more accurate with the backline method. This was logical as well, because instead of observing 40 units from only one part of the stand, the entire stand was now better represented with the backline method as the strip road travelled through the stand along the longest axis. This point was important also in the case of ALS-based predictions with the predicted residuals: with the backline method, the harvested cells were distributed at greater distances from each other, and therefore, a greater portion of the predicted residuals for the uncut cells were non-zero. As the data were spatially autocorrelated, at least the signs of the predicted residuals were mostly correct, and therefore, the accuracies of total predictions increased. To our knowledge, this is the first study in which also the residual errors were predicted as a part of the calibration of an LME model. In general, the acquisition of such data that allows to model the spatial autocorrelation of residuals has been challenging, so there are not yet that many studies on the topic (Breidenbach et al., 2016; Mauro et al., 2017).

Another interesting result was that the sawlog volume predictions were clearly improved with the calibrations, resulting in an unforeseen level of accuracy. However, a pre-requisite for improved predictions was that numerous local measurements were carried out. With only the fixed part of the model, the
RMSE% was about 22 per cent, so local measurements are crucial. The obtained results are in line with Karjalainen (2020) who concluded that for RMSE% values notably smaller than 20 per cent some stand-specific auxiliary information should be collected, preferably underneath the canopy. This is because the correlation between ALS data and the defects causing sawlog reduction is negligible (see, e.g. Karjalainen, 2020, for requirements for sawlogs). On the other hand, for example due to properties of the site, the tree quality seems to correlate strongly within a stand which allows effective calibration of stand-level sawlog volume.

The proposed calibration approach could potentially be used to also improve the ALS-based diameter distribution prediction of a stand as the cutting progresses. By knowing the diameter distribution of the remaining trees more accurately, bucking of logs could possibly be optimized during the clear-cut to better match to the demands of sawmills (Jusitalo et al., 2006). Jusitalo et al. (2006) reported that when the first harvested trees were combined with prior information, the reliability of diameter distribution predictions was increased. In the proposed calibration approach, such prior information about stand composition can be derived from ALS data. In addition, the quality of trees that has great effect on the accruals of different timber assortments can be considered with the harvester measurements. Calibration of diameter distributions is another interesting topic for a future study. Overall, the results of this study indicate that spatially accurate harvester data have a great potential for improving the performance of harvest operations in practical forestry.

Conclusion
Harvester measurements allow to calibrate ALS-based sawlog and merchantable volume models, while the harvesting is in progress. In this paper, the harvester-based in-situ calibrations increased the accuracies of predictions notably, indicating potential usefulness of the developed methodology in harvest management of individual clear cuts. In many cases, the utilization of ALS data in the calibrations resulted in more accurate predictions than the mean of the harvester measurements only. In addition, in the calibration of LME models, the prediction of residual errors (allowed by spatial autocorrelation of the data) increased the accuracies further compared with prediction of random effects only. The chosen strategy on how the clear-cutting is started may affect how effectively the predicted residuals can be utilized.

Data availability statement
The data underlying this article were provided by Norwegian University of Life Sciences by permission. Data will be shared on reasonable request to authors from Norwegian University of Life Sciences.

Supplementary data
Supplementary data are available at Forestry online.

Acknowledgements
We express our gratitude to John Deere Forestry AS, Viken Skog SA, Råsdøn Skogsmaskiner AS, Steinar Satha, Steinar Åmå and Norwegian University of Life Sciences for collecting and providing the data.

Conflict of interest statement
None declared.

Funding
‘PRECISION project’ (project #281140) funded by the Research Council of Norway; the Finnish Flagship Programme of the Academy of Finland for the Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) -project (decisions #337127 and #337655); and the Academy of Finland through the projects Unmanned Aerial Vehicles in Forest Remote Sensing (the Research Council for Biosciences, Health and Environment) (grant #323484) and Models of Heterogeneity, Contextuality and Self-interaction in Ordered Spatial Point Patterns with Applications to Animal Movement and Forest Inventory (grant #310073).

References
Axelsson, P. 1999 Processing of laser scanner data – algorithms and applications. ISPRS J. Photogramm. Remote Sens. 54, 138–147.

Bollandsdss, O.M., Maltamo, M., Gobakken, T., Lien, V. and Næsset, E. 2011 Prediction of timber quality parameters of forest stands by means of small footprint airborne laser scanner data. Int. J. For. Eng. 22, 14–23.

Breidenbach, J., McRoberts, R.E. and Astrup, R. 2016 Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. 173, 274–281.

Eriksson, M. and Lindroos, O. 2014 Productivity of harvesters and foresters in CTL operations in northern Sweden based on large follow-up datasets. Int. J. For. Eng. 25, 179–200.

Haaglin, M., Hansen, E., Næsset, E., Bysterud, B., Gjevestad, J. and Gobakken, T. 2017 Accurate single-tree positions from a harvester: a test of two global satellite-based positioning systems. Scand. J. For. Res. 32, 774–781.

Haaglin, M., Hansen, E., Næsset, E. and Gobakken, T. 2018 Utilizing accurately positioned harvester data: modelling forest volume with airborne laser scanning. Can. J. For. Res. 48, 913–922.

Holappa, M., Vastaranta, M., Rasimäki, J., Kalliovirta, J., Mäkinen, A., Haaparanta, R., et al. 2010 Uncertainty in timber assortment estimates predicted from forest inventory data. Eur. J. For. Res. 129, 1131–1142.

Karjalainen, T. 2020 Predicting commercial tree quality by means of airborne laser scanning. Dissertationes Forestales 307, 1–63.

Karjalainen, T., Packalen, P., Raty, J. and Maltamo, M. 2019 Predicting factual sawlog volumes in Scots pine dominated forests using airborne laser scanning data. Silva Fennica 53, 1–17.

Karjalainen, T., Mehtatalo, L., Packalen, P., Gobakken, T., Næsset, E. and Maltamo, M. 2020a Field calibration of merchantable and sawlog volumes in forest inventories based on airborne laser scanning. Can. J. For. Res. 50, 1352–1364.
In-situ calibration of stand level merchantable and sawlog volumes

Korhonen, L., Repola, J., Karjalainen, T., Packalen, P. and Maltamo, M. 2019 Transferability and calibration of airborne laser scanning based mixed-effects models to estimate the attributes of sawlog-sized scots pines. Silva Fennica 53, 1–18.

Lindroos, O., Ringdahl, O., Herra, P.L., Hohnlaser, P. and Hellström, T. 2015 Estimating the position of the harvester head – a key step towards the precision forestry in the future? Croat. J. For. Eng. 36, 147–164.

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

Maltamo, M., Hauglin, M., Næsset, E. and Gobakken, T. 2019 Estimating stand level stem diameter distribution utilizing harvester data and airborne laser scanning. Silva Fennica 53, 1–19.

Mauro, F., Monleon, V.J., Temesgen, H. and Ruiz, L.A. 2017 Analysis of spatial correlation in predictive models of forest variables that use LiDAR auxiliary information. Can. J. For. Res. 47, 788–799.

Mehtätalo, L. and Lappi, J. 2020a Biometry for Forestry and Environmental Data: With Examples in R. Chapman and Hall/CRC, New York.

Mehtätalo, L. and Lappi, J. 2020b Biometry for Forestry and Environmental Data: With Examples in R. Chapman and Hall/CRC, New York.

Metsäkeskus [Finnish Forest Centre]. 2021 Metsävaratiedon laatuseloste. [Quality description for the Forest Resource Information of Finnish Forest Centre.] https://www.metsakeskus.fi/sites/default/files/document/metsavaratiedon-laatuseloste.pdf (accessed on 18 May 2021) (In Finnish).

Metsäteho. 2010. Korjuun suunnittelu ja toteutus. Guide book for harvest planning and implementation. http://www.metsateho.fi/wp-content/uploads/2015/03/Korjuun_suunnittelua_toteutus_ver02.pdf. (accessed on 18 May 2021) (In Finnish).

Næsset, E. 2014 Area-based inventory in Norway – from innovation to an operational reality. In Forestry Applications of Airborne Laser Scanning – Concepts and Case Studies. M., Maltamo, E., Næsset, J., Vauhkonen (eds.). Vol. 27. Springer. Managing Forest Ecosystems, pp. 215–240.

Pateiro-Lopez, B. and Rodríguez-Casal, A. 2015 R package alphahull: generalization of the convex hull of a sample of points in the plane. R package version 2, 1. https://CRAN.R-project.org/package=alphahull.

Pinheiro, J., Bates, D., Deb Roy S., Sarkar, D. and R Core Team. 2019a nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-140. https://CRAN.R-project.org/package=nlme.

R Core Team. 2017 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Skogforsk. 2018 StanForD/StanForD 2010-Standard for Forest Machine Data and Communication. https://www.skogforsk.se/english/projects/stanford/ (accessed on 18 May 2021).

Uusitalo, J. 2003 Metsäteknologian perusteet. Metsäkustannus Oy. ISBN: 952-5118-58-4.

Uusitalo, J., Puustelli, A., Kivinen, V.-P., Nummi, T. and Sinha, B.K. 2006 Bayesian estimation of diameter distribution during harvesting. Silva Fennica 40, 663–671.

Karjalainen, T., Mehtätalo, L., Packalen, P., Gobakken, T., Næsset, E. and Maltamo, M. 2020b Field calibration of merchantable and sawlog volumes in forest inventories based on airborne laser scanning. Can. J. For. Res. 50, 1352–1364.

Lappi, J. 1986 Mixed Linear Models for Analyzing and Predicting Stem Form Variation of Scots Pine. Communications Instituti Forestalis Fenniae 134, FFRI 69p.

Mehtätalo, L. and Lappi, J. 2020b Biometry for Forestry and Environmental Data: With Examples in R. Chapman and Hall/CRC, New York.

Pinheiro, J.C. and Bates, D.M. 2000 Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York, USA, p. 528p.

Pinheiro, J., Bates, D., Deb Roy S., Sarkar, D. and R Core Team. 2019b nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-140. https://CRAN.R-project.org/package=nlme.
Appendix

A formal definition for the used LME model is as follows (Eq. A1):

\[ y_{ij} = x_i \beta + z_{ij} b_i + e_{ij} \]  
(A1)

where \( i \) indexes the group, \( j \) indexes the member \( j \) of the group \( i \), \( y \) is the response variable, \( x \) is the vector of predictors, \( \beta \) is the vector of regression coefficients, \( z \) is the vector of predictors in the random part, \( b \) is the vector of random effects and \( e \) is the residual. For the model, it was assumed that \( e \) is normally distributed \( N(0, \sigma^2) \) (where \( \sigma \) indicates the power). Parameters \( \sigma^2 \) and \( \delta \) were estimated during the model fitting using separate modelling data. However, this part is regular linear regression and will not be discussed in more depth here. See, e.g. Mehtätalo and Lappi (2020) for more details.

\[ \text{var}(e_{ij}) = \sigma^2 |v_i|^{2\delta} \]  
(A2)

where \( \sigma^2 \) is a scaling factor for the residual variance, \( v \) is the variance covariate and \( \delta \) is the unrestricted variance parameter indicating the power. Parameters \( \sigma^2 \) and \( \delta \) are estimated when the model is fitted (we used Restricted Maximum Likelihood). The assumption about constant variance of residual errors between observations was relaxed due to spatial autocorrelation in the data. To allow the prediction of residual errors, exponential correlation structure was applied as follows (Eq. A3):

\[ \text{cor}(e_{ij}, e_{ik}) = \begin{cases} 
1 - \psi_1 \exp \left( -\frac{d}{\psi_2} \right) & 0 < \psi_1 < 1; \psi_2 > 0; d > 0, \\
1 & d = 0,
\end{cases} \]  
(A3)

where \( \psi_1 \) is the nugget effect, \( d \) is the Euclidean distance between observations \( j \) and \( k \) and \( \psi_2 \) is the range of the spatial correlation (Mehtätalo and Lappi, 2020). Again, parameters \( \psi_1 \) and \( \psi_2 \) are estimated during the model fitting. The correlation between observations \( j \) and \( k \) is smaller the longer they are spatially apart. The exponential correlation structure was included in the model by using the R function ‘corExp’ (Pinheiro et al., 2019).

Next, we illustrate how the calibration procedure was carried out. To avoid unnecessary repeat, let us consider only the model for the merchantable volume. It employs the following ALS metrics as the predictors: H60 (60th percentile of height of ALS echoes), H60\(^2\) and HSD (standard deviation of heights of ALS echoes). The random part of the model includes only the random intercept and power-type variance function is applied using the predictor H60 as the variance covariate. Predictor for correlated observation 0 in group \( i \) using a mixed model can be calculated as shown in Eq. A4 (Lappi, 1986; Mehtätalo and Lappi, 2020).

\[ \hat{y}_{i0} = x_{i0} \hat{\beta} + z_{i0} \hat{b}_i + w_i R_i^{-1} \left( y_i - \hat{X}_i \hat{\beta} - Z_i \hat{b}_i \right). \]  
(A4)

These symbols will be defined below. The first component of the sum is the fixed part of the model and the second component is the random effects. The third component takes the spatial within-group correlation into account. Let us next inspect all the three components of Eq. A4 in turn.

The first component \( x_i \hat{\beta} \) includes the observed predictors for the observation 0 in group \( i \), and the estimated regression coefficients \( \hat{\beta} \) as follows:

\[ x_{i0} = \left[ 1 \ H60 \ H60^2 \ \text{HSD} \right] \]

\[ \hat{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \]

\( \hat{\beta} \) were estimated during the model fitting using separate modelling data. However, this part is regular linear regression and will not be discussed in more depth here. See, e.g. Mehtätalo and Lappi (2020) for more details.

The second component \( z_{i0} \hat{b}_i \) includes the random predictors of the model, and the predicted random effects for group \( i \). The model for merchantable volume includes only a random intercept. Therefore, \( z_{i0} \) is 1. The random effect \( b \) for group \( i \), on the other hand, is predicted using information from all the observations from stand \( i \) and by employing Empirical Best Linear Unbiased Predictor (EBLUP) (Mehtätalo and Lappi, 2020) (Eq. A5):

\[ \text{EBLUP}(b_i) = D_i Z_i (Z_i D_i Z_i + R_i)^{-1} (\hat{y}_i - X_i \hat{\beta}). \]  
(A5)

In the following matrices, let us assume that the merchantable volume has been measured from \( p \) grid cells. \( D_i \) is the estimated variance–covariance matrix of the random effects. As there is only a random intercept in the model for merchantable volume, \( D_i \) is of size \( 1 \times 1 \) and includes the variance of the random intercept only. \( Z_i \) includes the random part of the model for the \( p \) calibration measurements.

\[ D_i = \begin{bmatrix} \text{var(} \text{Random intercept}) \end{bmatrix} \]

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

\( R_i \) is the estimated variance–covariance matrix of the residual errors. The size of \( R \) is \( p \times p \). The values in \( R \) are affected by the applied power-type variance function and the exponential correlation structure. However, all the values in \( R \) can be calculated using the equation \( \text{cor}(e_{ij}, e_{ik}) \sqrt{\text{var}(e_{ij}) \text{var}(e_{ik})} \). In the diagonal, \( j = k \), so \( \text{cor}(e_{ij}, e_{ik}) = 1 \). Consequently, the value is directly the variance of the cell (let us say ‘l’) in question, which is determined by the applied power-type variance function as \( \sigma^2 |H60|^\delta \). In off diagonals, the correlation between observations \( j \) and \( k \) is the outcome of the exponential correlation structure specified earlier. Furthermore, the covariance in row \( j \), column \( k \), can be calculated by multiplying the correlation by the standard deviations (that is, the square root of the variance) of the cells \( j \) and \( k \).
In-situ calibration of stand level merchantable and sawlog volumes

\( y_i \) includes the observed merchantable volumes and \( X_i \) the predictors for the \( p \) harvested grid cells as follows:

\[
y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} 1 & H60_1 & H60^2_1 & HSD_1 \\ 1 & H60_2 & H60^2_2 & HSD_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & H60_p & H60^2_p & HSD_p \end{bmatrix} \]

\( \hat{\beta} \) is the same as earlier in the first component of Eq. A4. Employing the previously described vectors and matrices to EBLUP equation results in prediction \( \hat{b}_i \) for the stand effect \( b_i \). The vectors and matrices in the third component \( w_i R_i^{-1} (y_i - X_i \hat{\beta} - Z_i \tilde{b}_i) \) are already described, expect for the \( w_i \), which includes the covariances between the target cell 0 and the other observations in the group \( i \) as follows:

\[
w_i = \begin{bmatrix} \text{cov} (\epsilon_{i0}, \epsilon_{i1}) & \text{cov} (\epsilon_{i0}, \epsilon_{i2}) & \cdots & \text{cov} (\epsilon_{i0}, \epsilon_{ip}) \end{bmatrix}.
\]

The covariances in \( w_i \) are calculated in the same way as in matrix \( R_i \). Predictions for each cell were always updated after a new cell was harvested. For example, when the third cell from a stand was clear-cut, the prediction for random effect \( \tilde{b}_i \) was recalculated using measurements from all three harvested cells instead of only the two previous ones. In addition, the predictions for the residual errors were also recalculated separately for each cell in the stand. The closer the cell located to the last harvested cell, the more the predicted residual error was affected.