This is a post-peer-review, pre-copyedit version of an article published in Remote Sensing of Environment.
The final authenticated version is available online at: http://dx.doi.org/10.1016/j.rse.2018.04.028.

Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data

Johannes Breidenbach

Norwegian Institute of Bioeconomy Research, NIBIO, Postboks 115, 1431 Ås, Norway

Steen Magnussen

Natural Resources Canada, Canadian Forest Service, 506 West Burnside Road, Victoria British Columbia V8Z 1M5, Canada

Johannes Rahlf

NIBIO, Postboks 115, 1431 Ås, Norway

Rasmus Astrup

NIBIO, Postboks 115, 1431 Ås, Norway

Abstract

In many applications, estimates are required for small sub-populations with so few (or no) sample plots that direct estimators that do not utilize auxiliary variables (e.g. remotely sensed data) are not applicable or result in low precision. This problem is overcome in small area estimation (SAE) by linking the variable of interest to auxiliary variables using a model. Two types of models can be distinguished based on the scale on which they operate: i) Unit-level models are applied in the well-known area-based approach (ABA) and are commonly used in forest inventories supported by fine-resolution 3D remote sensing data such as airborne laser scanning (ALS) or digital aerial photogrammetry (AP); ii) Area-level models, where the response is a direct estimate based on a sample within the domain and the explanatory variables are aggregated auxiliary variables, are less frequently applied. Estimators associated with these two model types can make use of sample plots within domains if available and reduce to so-called synthetic estimators in domains where no sample plots are available. We used both model types and their associated model-based estimators in the same study area with AP data as auxiliary variables. Heteroscedasticity, i.e. for continuous dependent variables typically an increasing dispersion of residuals with increasing predictions, is often observed in models linking field- and remotely sensed data. This violates the model assumption that the distribution of the residual errors is constant. Complying with model assumptions is required for model-based methods to result in reliable estimates. Addressing heteroscedasticity in models had considerable impacts on standard errors. When complying

Email address: job@nibio.no (Johannes Breidenbach)

with model assumptions, the precision of estimates based on unit-level models was, on average, considerably greater (29%-31% smaller standard errors) than those based on area-level models. Area-level models may nonetheless be attractive because they allow the use of sampling designs that do not easily link to remotely sensed data, such as variable radius plots.

Keywords: **Keywords:** Forest inventory, model-based inference, synthetic estimator, variance estimation, image matching

1. Introduction

Fine-resolution remotely sensed data such as 3D point clouds acquired using airborne laser scanning (ALS) or digital aerial photogrammetry (AP) can be utilized for estimating forest parameters with individual tree crown approaches (ITC), area-based approaches (ABA), or domain-level approaches (DLA). The three approaches can be linked to two model types (unit-level and area-level) and their associated model-based estimators in small area estimation (SAE). The aim is typically to provide estimates of forest parameters on the level of stands or other domains. For example, main outcomes of an inventory using the ABA are stand-level mean timber volume estimates resulting from averaging the grid-cell predictions per stand. Today, ITC and ABA are commonly applied in operational forest management inventories (Næsset, 2014) while domain-level approaches (DLA) have mostly been used in research studies (van Aardt et al., 2006; Goerndt et al., 2011).

Unit-level models are applied in ITC and ABA because the variable of interest and the explanatory variables are available on the level of population units such as geo-located trees or field sample plots (Rao, 2003, Ch. 5.3). Area-level models are applied in DLA if variables of interest are estimates based on a sample within a domain (Rao, 2003, Ch. 5.2). The explanatory variables in area-level models are obtained by aggregating the auxiliary variables to the domain level, for example by calculating the mean of the ALS heights within each forest stand. The advantage of area-level models is that no exact geo-locations are required for sample plots which can be also of interest in cases where plot coordinates are confidential. Furthermore, it allows the use of plot designs that cannot be exactly matched to remotely sensed data. For example, it may be difficult to link a line transect, as used in distance sampling, to remotely sensed data and to tesselate the study area according to the transect as required for unit-level estimators (Bäuerle et al., 2009). Similar issues arise with variable radius plots although workarounds may be feasible (Kirchhoefer et al., 2017). In that way, area-level models may simplify the use of remote sensing techniques, such as mobile laser scanning (Saarela et al., 2017) or UAV-based data acquisitions (Nevalainen et al., 2017), as reference data in larger scale applications.

Inference in remote sensing-supported forest inventories, i.e. the estimation of a population parameter and the associated uncertainty, using unit-level estimators has recently received more attention (Mandallaz, 2013; McRoberts et al., 2014; Saarela et al., 2015, 2016; Gregoire et al., 2016; Chen et al., 2016; Mauro et al., 2016). Also, area-level estimators are becoming more studied (Goerndt et al., 2011, 2013; Boubeta et al., 2015; Magnussen et al., 2017). However, unit- and area-level estimators are seldom compared (Hidiroglou and You, 2016). Mauro et al. (2017b) compared unit- and area-level estimators in an ALS-supported forest inventory where the domains were forest stands aggregated to management units larger than 4 ha. They found that the root mean squared errors (RMSE) of area-level estimates were, on average, between 1.3 (Lorey's height) and 2.8 (timber volume) larger than RMSE of unit-level estimates.

For continuous dependent variables such as timber volume, heteroscedasticity typically manifests itself as an increasing dispersion of residuals with increasing predictions. It is frequently observed in linking models for remote-sensing supported forest inventories (e.g., Rahlf et al., 2014; Saarela et al., 2016). Heteroscedasticity may be caused by natural phenomena but can also indicate an omitted explanatory variable or a mis-specified model shape (e.g., linear instead of curvylinear relationship). How to handle heteroscedasticity in SAE is still an active field of research and did not receive much attention in area-level estimation. In unit-level estimation, Militino et al. (2006) used the number of sample units within a domain for considering heteroscedasticity. This choice, however, does not allow any synthetic estimates that are often required in forest inventories. Mauro et al. (2017b) chose one of three transformations of the most important explanatory variable based on goodness of fit criteria and visual inspections of the residuals to model the heteroscedasticity. An alternative was suggested by Jiang and Nguyen (2012) who assumed that the data come from different super-populations by categorizing the sample units into few groups. The mean squared error is then a function of the empirical residuals within each group. This has, however, the disadvantage that the continuous nature of the heteroscedasticity is categorized and the number of super-populations has to be selected.

The aim of this study was to compare area-level and unit-level models and associated model-based estimators in a case study with forest inventory data using AP as auxiliary variables. Specifically, we analyze the consequences of heteroscedasticity because realistic uncertainty assessment of estimators associated with area-level and unit-level models depends on compliance with model assumptions. We study the case where several field sample plots are available within stands which are used in linear models linking the variable of interest with auxiliary information. Synthetic estimators, i.e. aggregates of model predictions, are applied in stands without sample plots (Breidenbach et al., 2015; Magnussen et al., 2016).

2. Methods

2.1. The direct estimator

The aim is to estimate the population mean of some variable of interest (e.g., timber volume) for each of $i = 1, \ldots, M$ domains (small areas). Each domain is composed of $j = 1, \ldots, N_i$ population units, such that

$$\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij} \tag{1}$$

where y_{ij} is the variable of interest of the j-th population unit within the i-th domain and N_i is the known number of population units within the i-th domain.

The estimate is denoted $\hat{\bar{Y}}_i$ and can be calculated from the sample units (e.g., sample plots) for the $m \leq M$ domains (e.g., forest stands). This estimator is denoted direct (D) and is usually a Horvitz-Thompson estimate of the mean. Assuming simple random sampling (SRS) within a domain, the estimator is given by

$$\hat{\bar{Y}}_i^D = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \tag{2}$$

where y_{ij} is the j-th sample unit in domain i, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, and n_i is the number of sample units within domain i. Its variance is estimated by

$$\hat{\sigma}_i^2 = F_i \frac{s_i^2}{n_i} \tag{3}$$

where

$$F_i = \frac{N_i - n_i}{N_i} \tag{4}$$

is the finite population correction that is close to one and can thus be ignored for large populations where the sampling fraction $f_i = \frac{n_i}{N_i}$ is small, and

$$s_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2/(n_i - 1) \tag{5}$$

is the sample variance. Domain totals and their variances can be estimated by multiplying the mean estimate with the domain size N_i and the variance estimate with N_i^2.

2.2. Notation for the use of auxiliary variables

We assume that for each population unit (e.g., pixels or grid cells) a vector of p explanatory variables $\boldsymbol{x}_{ij}, j = 1, \ldots, N_i$ (auxiliary variables), obtained from remotely sensed data, is available (wall-to-wall). Consequently, the explanatory variables are available for each sample unit $\boldsymbol{x}_{ij}, i = 1, \ldots, n_i$, and can be used in unit-level models (next section).

The mean of the explanatory variables over all population units within a domain ($\boldsymbol{x}_{ij}, j = 1, \ldots, N_i$) will be denoted $\bar{\boldsymbol{x}}_{iP}$ and can be used for area-level or unit-level estimates on domain level. The mean of the explanatory variables over the $N_i - n_i$ population units not sampled within a domain ($\boldsymbol{x}_{ij}, j = 1, \ldots, N_i - n_i$) will be denoted $\bar{\boldsymbol{x}}_{iR}$. In forest inventories, it will often be the case that some domains (stands) do not contain any sample plots. Explanatory variables for those domains are still available and will be used for so-called synthetic estimates. In this case, synthetic estimates are aggregate statistics, such as mean or sum, of the model predictions over all population units in a domain.

Following Rao and Molina (2015), the term variance will be used for design-based estimators (i.e., the direct estimator, section 2.1), while the term mean squared error (MSE) will be used for model-based estimators which are not necessarily design-unbiased. The square root of variance or MSE estimates will be denoted standard error (SE) (Rao and Molina, 2015, p. 187). The software implementation of the methods described below is available as an `R`-package (Breidenbach, 2013).

2.3. Unit-level models and estimators

To estimate the population mean (Rao and Molina, 2015, Ch. 7), unit-level estimators use the nested-error linking model

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i + \epsilon_{ij}, \quad v_i \sim N(0, \sigma_v^2), \quad \epsilon_{ij} \sim N(0, k_{ij}^2 \sigma_e^2), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i \quad (6)$$

where $\boldsymbol{\beta}$ is a vector of coefficients, v_i is an independently identically distributed random intercept with variance σ_v^2 which is assumed to be independent of the residuals ϵ_{ij} which are assumed to be mutually independent with variance σ_e^2, and k_{ij} is a known constant used for capturing heteroscedasticity. For $k_{ij} = 1$, the model is suitable for homoscedastic residual variances and all further estimators simplify considerably (Prasad and Rao, 1990; Breidenbach and Astrup, 2012). Transformed residuals and standardized residuals (σ_v- and σ_e-residuals) can be used to test the model assumptions of homoscedasticity and normality (see AppendixA.1.1 for details).

For large populations and negligible sampling fractions (Battese et al., 1988), the residual error mean assumes its expected value (zero) and the EBLUP estimator (empirical best linear predictor) of the domain mean is

$$\hat{\mu}_i^{\mathrm{UE}} = \bar{\boldsymbol{x}}_{iP}^T \hat{\boldsymbol{\beta}} + \hat{v}_i = \hat{\mu}_i^{\mathrm{US}} + \hat{v}_i \tag{7}$$

where the superscript UE denotes the Unit-level EBLUP estimator and the superscript US denotes the Unit-level (EBLUP) Synthetic estimator. The synthetic estimator

$$\hat{\mu}_i^{\mathrm{US}} = \bar{\boldsymbol{x}}_{iP}^T \hat{\boldsymbol{\beta}} \tag{8}$$

is the mean over all predictions within a domain which makes the estimator applicable for domains without samples.

The weight $\hat{\gamma}_i$ is the ratio of the unexplained among-domain variability (the random-effect variance, $\hat{\sigma}_v$) and the total variability

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / a_{i\cdot}} \tag{9}$$

where $a_{i\cdot} = \sum_{j=1}^{n_i} a_{ij}$ and $a_{ij} = k_{ij}^{-2}$. Under homoscedasticity ($k_{ij} = 1$), $a_{i\cdot}$ reduces to n_i. The smaller the relative unexplained among-domain variability (i.e., the more variance is explained by the fixed part of model (6)), the smaller is $\hat{\gamma}_i$ and the more weight is given to the synthetic estimator (i.e., the smaller is the predicted random effect $\hat{v}_i$).

For domains with small populations (Rao and Molina, 2015, Ch. 7.1.3), sampling fractions (f_i) can be non-negligible and the EBLUP estimator is given by

$$\hat{\bar{Y}}_i^{\mathrm{UE}} = \hat{\mu}_i^{\mathrm{US}} + \omega_i (\hat{\bar{Y}}_i^D - \hat{\mu}_i^{\mathrm{US}}) \tag{10}$$

where $\omega_i = f_i + (1 - f_i)\hat{\gamma}_i$ and $\hat{\bar{Y}}_i^D$ is the sample mean (eq. 2). The symbol $\hat{\bar{Y}}_i^{\mathrm{UE}}$ is used here rather than $\hat{\mu}_i^{\mathrm{UE}}$ because the sample units are considered in the estimate.

The uncertainty (MSE) of unit-level estimators results from estimating the random effects, the fixed model parameters, and the residual error (see AppendixA.1.2 for details).

2.4. Area-level models and estimators

In the area-level estimators (Rao and Molina, 2015, Ch. 6), the model that links the auxiliary variables to the direct mean estimate is described as a mixed-effects model

$$\hat{\bar{Y}}_i^D = \bar{\boldsymbol{x}}_{iP}^T \boldsymbol{\beta} + b_i \upsilon_i + \epsilon_i, \quad \upsilon_i \sim N(0, \sigma_\upsilon^2), \quad \epsilon_i \sim N(0, \sigma_i^2), \quad i = 1, \ldots, m \tag{11}$$

where υ_i is a random intercept, b_i is a known constant for considering heteroscedasticity and σ_i^2 is the variance of the direct estimator. With $b_i = 1$, the model is suited for homoscedastic data. The direct estimator $\hat{\bar{Y}}_i^D$ is based on sample units within a domain such as eq. (2) but can be based on any probability sampling design. This allows the use of plot designs that are difficult to link to remotely sensed data on the population unit level. Because of the missing repetitions of observations on domain-level, the estimation of the fixed-effects parameters $\boldsymbol{\beta}$ is non-standard (see AppendixA.2.1 for details on estimating fixed and random effects).

The area-level EBLUP estimator (superscript AE) (Fay and Herriot, 1979)

$$\tilde{\mu}_i^{\text{AE}} = \hat{\mu}_i^{\text{AS}} + b_i \upsilon_i = \bar{\boldsymbol{x}}_{iP}^T \hat{\boldsymbol{\beta}} + b_i \upsilon_i \tag{12}$$

is the weighted average of the synthetic and a direct estimate such as eq. (2).

$$\hat{\mu}_i^{\text{AE}} = \hat{\gamma}_i \hat{\bar{Y}}_i^D + (1 - \hat{\gamma}_i) \hat{\mu}_i^{\text{AS}} \tag{13}$$

with

$$\hat{\gamma}_i = \frac{\hat{\sigma}_\upsilon^2 b_i^2}{\hat{\sigma}_\upsilon^2 b_i^2 + \hat{\sigma}_i^2}. \tag{14}$$

Here, $\hat{\gamma}_i$ gives more weight to the direct estimate if its variance is relatively small and vice versa. The more variance is explained by the fixed part of model (11), the more weight is given to the synthetic estimator. The area-level synthetic estimator (superscript AS) can be used for domains without observations and is given by

$$\hat{\mu}_i^{\text{AS}} = \bar{\boldsymbol{x}}_{iP}^T \hat{\boldsymbol{\beta}}. \tag{15}$$

The uncertainty (MSE) of area-level estimators results from estimating the random effect and the fixed model parameters, and the uncertainty of the direct estimate (see AppendixA.2.2 for details).

3. Case study

3.1. Overview

The study area consisted of parts of Vestfold county in south-eastern Norway with a full coverage of digital aerial photogrammetry (AP) data. Our aim was to estimate the mean timber volume scaled to per-hectare values for each of $i = 1, \ldots, m = 64$ stands (domains) from which between $n_i = 4$ and $n_i = 7$ population units were sampled using fixed-area sample plots. The sample plots had an area of 250 m^2 and trees were recorded according to the protocol of the Norwegian National Forest Inventory (NFI) for temporary plots (Landsskogtakseringen, 2008). Single tree volume was predicted from diameter at breast height and tree height recordings applying the standard models used in the NFI. Volume per hectare on plot-level was obtained by aggregating and expanding the single tree predictions. Uncertainties arising from volume models were ignored in this study. Volume per hectare, y_{ij}, ranged between 0.0 and 947.8 m^3/ha on the $n = \sum_{i=1}^{i=64} n_i = 382$ sample plots. The stands were sampled from areas with available forest management inventories (FMI) in the municipalities of Holmestrand, Lardal, and Stokke (Fig. 1).

A total of 30 stands was selected in Lardal with an aim to take a sample of 7 plots per stand (two stands had 6 plots). In Holmestrand and Stokke, 14 and 15 stands were selected with an aim to take a sample of 5 plots per stand (two stands had 4 plots). A constrained random sample of stand polygons that were delineated in the FMI was selected. The constraints were based on area (between 1 and 3 ha in size), and shape (avoidance of stands with a very complex outline) in order to simplify the field work. More information on the field data can be found in Breidenbach et al. (2015).

Potential explanatory variables were the mean and other metrics (McGaughey, 2014) of AP heights. The AP metrics were calculated for the sample plots and for square grid cells with 16 m side length tessellating the study area. The grid cell size was selected to correspond approximately with the sample plot size. Image matching was performed using Socet Set 5.5.0 on digital aerial images with a pixel size of 20 cm that were acquired using a Vexcel UltraCamX sensor. More information on the remotely sensed data can be found in Breidenbach and Astrup (2012).

The number of grid cells within a stand N_i ranged from 38 to 159 with a mean of 68 (1.7 ha). For 4-7 sample plots per stand, this resulted in sampling fractions of 4%-15% with a mean of 10% and corresponding finite population corrections (eq. (4)) of 0.84-0.96 with a mean of 0.90.

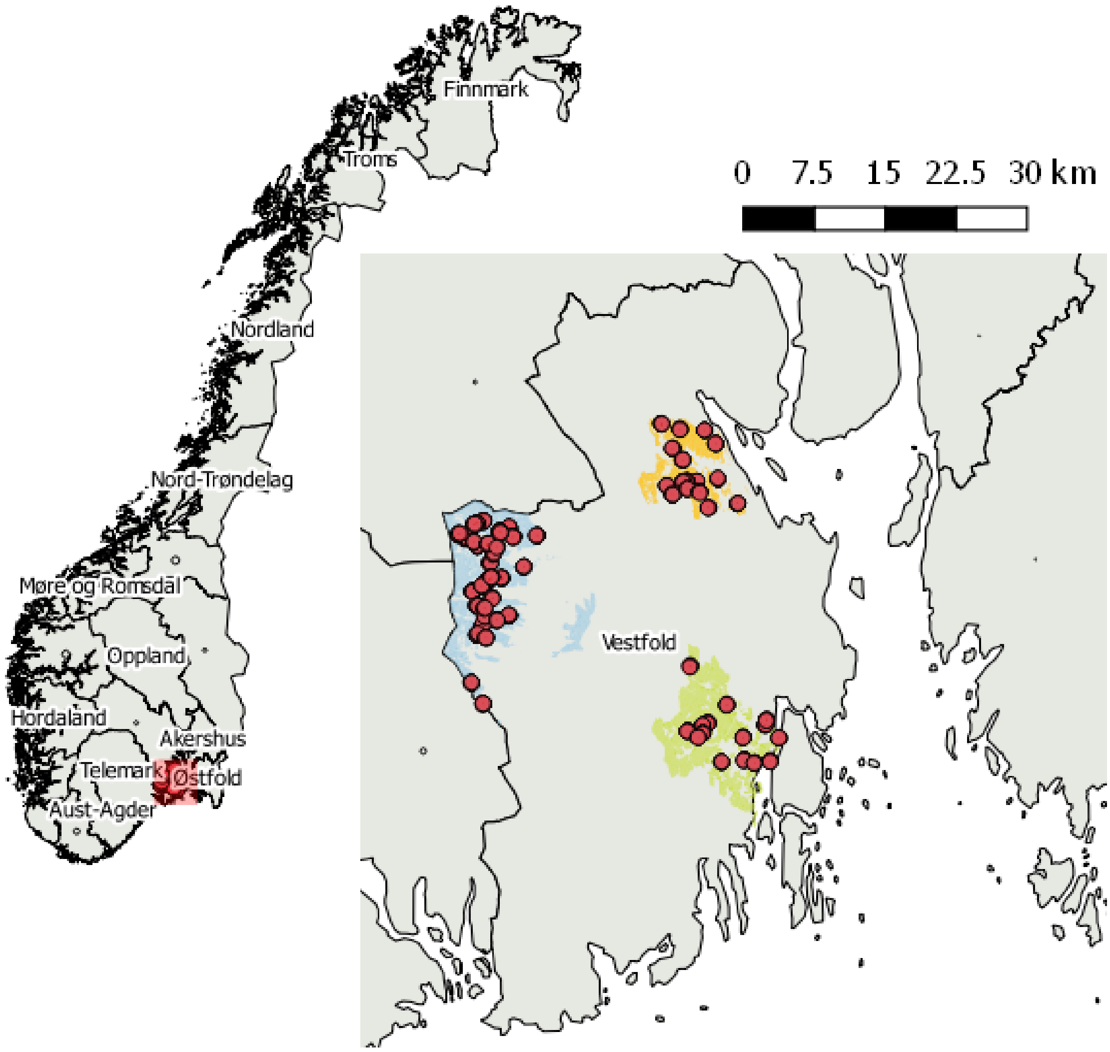

Figure 1: Left: Location of the study area within Norway (red square). Right: Location of the sample plots within the forest areas of Holmestrand (orange), Lardal (blue), and Stokke (green).

3.2. Modeling

3.2.1. Unit-level models

The unit-level variable of interest y_{ij} in the mixed effects model (6) was timber volume scaled to per-hectare values observed on the sample plots. Possible effects of unequal sampling fractions among the stands were ignored when estimating the model parameters. Based on previous experience, AP mean height observed on a sample plot (x_{2ij}) and its square (x_{3ij}) were the only explanatory variables in models with intercept ($x_{1ij} = 1$), resulting in $p = 3$. Restricted maximum likelihood was used to estimate the model parameters. The fixed part of the mixed models explained more than 74% of the total variation.

Models ignoring heteroscedasticity ($k_{ij} = 1$, eq. 9) showed clear patterns of increasing variance with increasing x values in the residuals (Fig. 2 A and B). The constant k_{ij} for considering heteroscedasticity affects standard errors of estimates, and to a smaller degree, also the estimates themselves. Because influential observations may have a strong influence of the decision, we used the following procedure to select k_{ij}:

- The Akaike's information criterion (AIC) was minimized by varying the parameter ξ in $k_{ij} = x_{2ij}^{\xi}$.
- Five influential observations were removed temporarily from the data set after inspecting the σ_v-residuals (eq. A.1) and σ_e-residuals (eq. A.2). The sample plot with the largest observed timber volume was among the influential observations.
- Minimizing the AIC with the reduced data set still left structures in the residuals. Therefore, the harmonic mean of the p-values of Breusch-Pagan tests (Breusch and Pagan, 1979) for the squared σ_v- and σ_e-residuals was maximized under the constraint that both p-values were > 0.05 by varying the parameter ξ in $k_{ij} = x_{2ij}^{\xi}$. With p-values > 0.05, the hypothesis of homoscedasticity is not rejected given a 95% significance level.

The optimized value was $\xi = 0.48$ resulting in Breusch-Pagan p-values of 0.14 and 0.10 for the σ_v- and σ_e-residuals, respectively. Breusch-Pagan p-values were smaller than 0.05 for the full data set with the influential observations. Larger values of ξ would have led to a stronger visibility of heteroscedasticity in the estimated standard errors but also to structures in the residuals. The values of k_{ij} ranged from 0.7 to 14.5 with a mean of 8.2. The value of k_{ij} is selected based on data although it is assumed to be known (Rao and Molina, 2015, Ch. 4.3). Possible influences on MSE estimators were ignored.

In some studies "outliers" are removed in order to meet the model assumptions (Battese et al., 1988). The four influential observations that were removed temporarily in order to select ξ were, however, not outliers because we could not find any evidence of data errors. We therefore kept all observations in the models but also report results for models without the influential observations.

Although the scaled residuals were approximately symmetrically distributed around zero (Fig. 2 C and D), it was not possible to select k_{ij} values that would have resulted in normally distributed residuals (test: Shapiro-Wilk given a 95% significance level), even after removing the influential observations.

For stand-level estimates, the stand-level mean of AP mean height observed on all grid cells ($\bar{x}_{2iP}$), and the mean of squared AP mean height ($\bar{x}_{3iP}$) were calculated. For considering small population sizes (eq. (10)), stand-level means need to be calculated also for the non-sampled population units ($\bar{\boldsymbol{x}}_{iR}$, eq. (A.7)). However, the sample plots are circular and not aligned with the grid cells. Therefore, for each sample plot the closest grid cell was omitted in order to calculate $\bar{x}_{2iR}$, and $\bar{x}_{3iR}$. Table 1 summarizes some characteristics of the response and explanatory variables. Synthetic estimates (eq. (8)) were generated for the 64 stands with plots by assuming that no field data were available for them.

	Mean	Min	Max	SD
y_{ij}	193.02	0.00	947.80	141.23
x_{2ij}	91.70	0.47	263.73	57.11
$\bar{x}_{2iP}$	91.73	8.31	201.92	45.75

Table 1: Characteristics of plot-level timber volume (y_{ij}, m^3/ha), plot-level AP mean height (x_{2ij}, dm), and stand-level means of AP mean height observed on all grid cells ($\bar{x}_{2iP}$, dm).

3.2.2. Area-level models

The response variable in the area-level model (eq. (11)) was the direct mean timber volume estimate of the sample plots ($\hat{\bar{Y}}_i^D$, eq. (2)). The stand-level mean of AP mean height observed on all grid cells ($\bar{x}_{2iP}$) was the only explanatory variable in models with intercept ($\bar{x}_{1iP} = 1$) such that $p = 2$. As opposed to the unit-level model, the parameter estimate of the square of the AP mean height was not significantly different from 0. Restricted maximum likelihood was used to estimate the model parameters.

The constant b_i for considering heteroscedasticity affects standard errors of estimates, and to a smaller degree, also the estimates themselves. Furthermore, influential observations may have a strong influence on the selection of b_i. Models ignoring heteroscedasticity ($b_i = 1$) showed clear patterns of increasing variability in the residuals with increasing $\bar{x}_{2iP}$ (Fig. 3 A). However, values

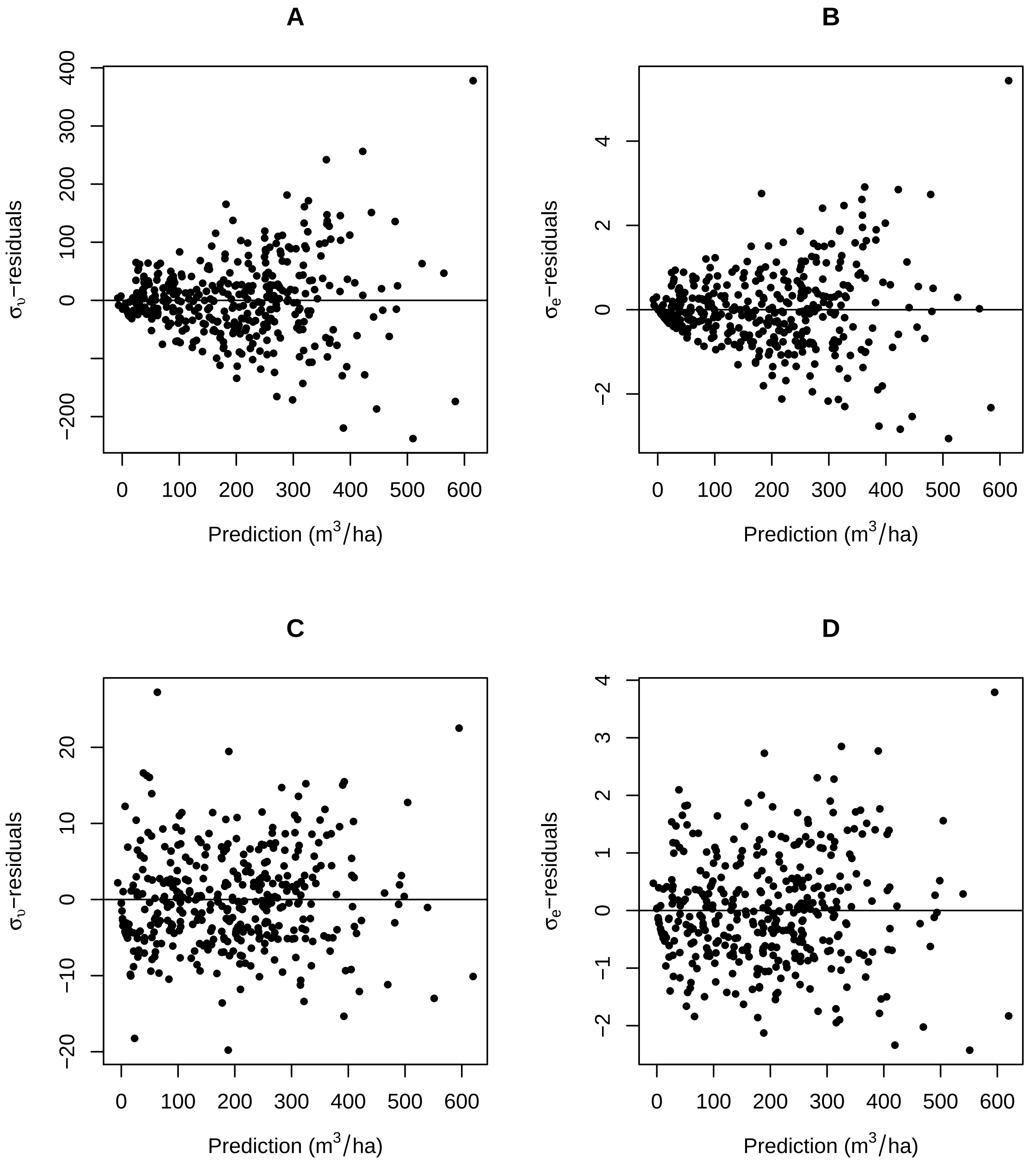

Figure 2: Transformed and scaled residuals (see AppendixA.1.1, including influential observations) versus predicted values for the unit-level model with $k_{ij} = 1$ (A and B), and $k_{ij} = x_{2ij}^{0.48}$ (C and D).

of b_i in the order of $\bar{x}_{2iP}$ led to numeric instability. Therefore, we used the following procedure to select b_i:

- The explanatory variable $\bar{x}_{2iP}$ was transformed ($\bar{x}^*_{2iP}$) to the range $0, \ldots, 1$ and the AIC was minimized by varying the parameter ζ in $b_i = \bar{x}^*_{2iP} + \zeta$.
- One influential observation was removed temporarily from the data set and scaled random effects (eq. A.15) were obtained. The influential observation was also among the influential observations in the unit-level model.
- Minimizing the AIC with the reduced data set still resulted in structures in the residuals. Therefore, the harmonic mean of the p-values of Breusch-Pagan and Shapiro-Wilk tests for the scaled random effects was maximized under the constrain that both p-values were $>$ 0.05 by varying the parameter ζ in $b_i = \bar{x}^*_{2iP} + \zeta$.

The selected value was $\zeta = 0.39$ resulting in Breusch-Pagan and Shapiro-Wilk p-values of 0.32 and 0.14, respectively. The Breusch-Pagan p-value was larger than 0.05 for the data set including the influential observation but the Shapiro-Wilk p-value was smaller than 0.05 (Fig. 3 B). As before, we did not have any indication of errors and therefore did not exclude the influential observation. The values of b_i ranged from 0.39 to 1.39 with a mean of 0.82. Possible influences on MSE estimators due to the selection of b_i were ignored.

3.3. Results

3.3.1. Comparison of unit- and area-level estimators

This section gives an overview of the results which are further described in the two following sections. Due to the small sample size within stands, results of the design-based direct estimator are presented with the area-level estimates.

Although unit-level and area-level EBLUP estimates at stand-level were similar (Fig. 4), their SE differed considerably (Fig. 4 and 5). When not considering heteroscedasticity, SE of unit-level EBLUP estimates varied unrealistically little, and were on average similar to SE of area-level EBLUP estimates where SE increased with increasing estimates. Furthermore, SE of synthetic estimates were similar for unit-level and area-level models. However, not considering heteroscedasticity strongly violated the model assumption of homogeneity of variance.

In models where heteroscedasticity was addressed, SE of unit-level EBLUP estimates increased in parallel with increasing mean estimates but were on average 31% smaller than SE of

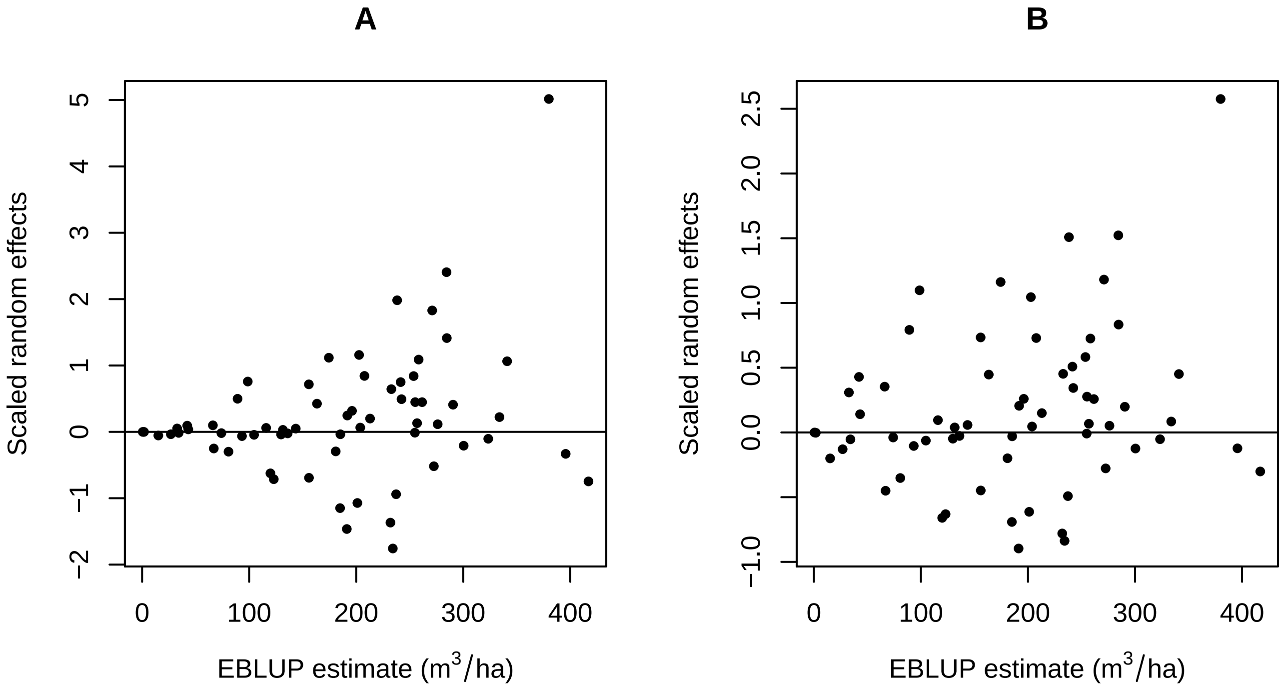

Figure 3: Scaled random effects (including one influential observation) versus area-level EBLUP estimates on stand-level with $b_i = 1$ (A), and $b_i = \bar{x}_{2iP} + 0.39$ (B).

area-level estimates. However, for 11 of the 64 stands, the SE of area-level EBLUP estimates was slightly smaller than the SE of unit-level EBLUP estimates. These were typically stands with relatively precise direct estimates. SE of unit-level synthetic estimates were consistently and, on average, 28% smaller than SE of area-level synthetic estimates.

The influence of heteroscedasticity was hardly visible in SE unit-level and area-level synthetic estimates. For unit-level estimates, this is because the contribution of the residual variance $(k_{ij}^2 \sigma_e^2)$ that models the heteroscedasticity is negligible as it is divided by a large number of grid cells (eq. A.11). Therefore, heteroscedasticity is only indirectly accounted for by the uncertainty of the fixed parameter estimates. Also for area-level synthetic estimates, heteroscedasticity is only accounted for by the uncertainty of the fixed parameter estimates (A.21).

3.3.2. Unit-level estimators

Stand-level synthetic estimates (assuming no sample plots within stands, eq. (8)) of mean timber volume were similar whether heteroscedasticity was considered ($k_{ij} = x_{2ij}^{0.48}$) or not ($k_{ij} = 1$) and whether influential observations were included in the model or not (Fig. B.6 A). In other words, the fixed effects estimates differed only slightly between the models. EBLUP estimates ($k_{ij} = x_{2ij}^{0.48}$, all observations) of stand-level timber volume ranged from 6.13 to 437.77 with a

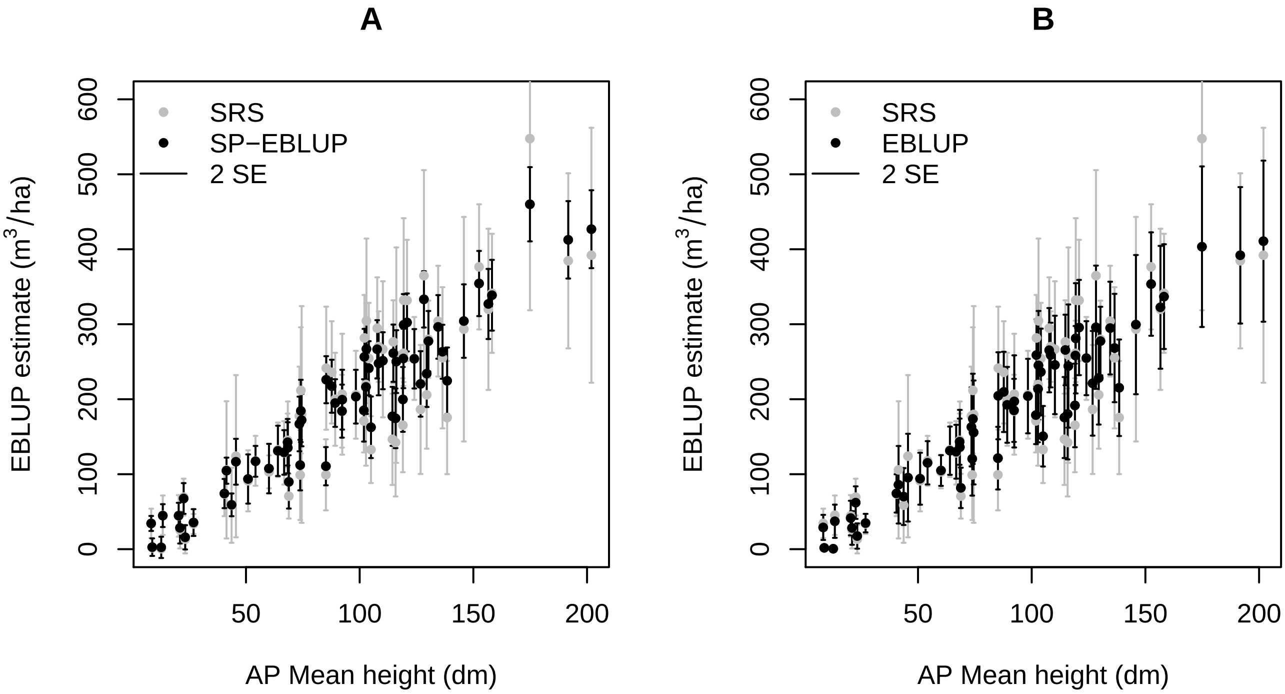

Figure 4: Stand-level estimates assuming small populations (SP-EBLUP) based on unit-level models (A) and area-level models (B). SRS = direct estimates.

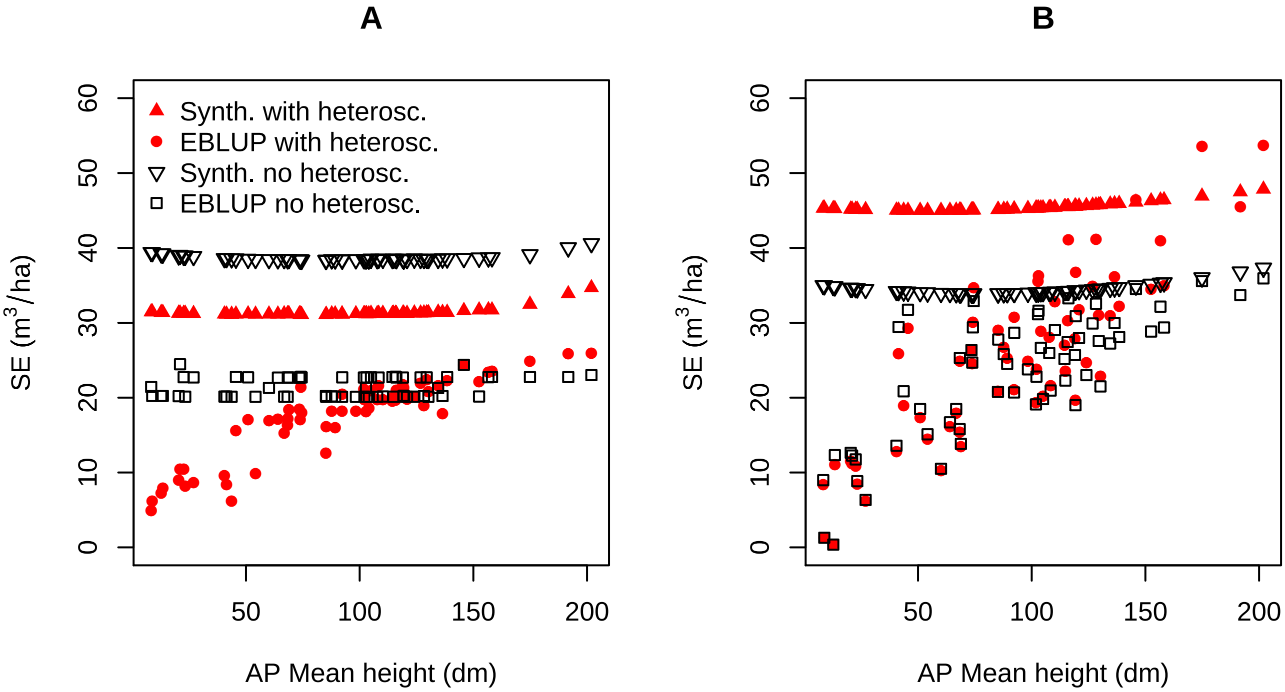

Figure 5: Standard errors versus AP mean height for unit-level (A) and area-level estimates (B).

mean of 190.83 m^3/ha.

SE increased slightly for models including influential observations (Tab. B.7). The fact that mean estimates remained similar while standard errors increased made us more confident in keeping the influential observations despite of the formal violation of the assumptions of homoscedasticity and normality of the residuals.

SE of mean estimates based on models considering heteroscedasticity or not differed drastically (Fig. 5 A and Tab. 3) which also resulted in differences in the EBLUP estimates (Fig. B.6 B). Random-effect variances of models assuming heteroscedasticity were smaller than of models assuming homoscedasticity (Tab. 2), which resulted in larger values of $\hat{\gamma}_i$ and thus more weight assigned to the synthetic estimator.

	R^2	σ_v^2	σ_e^2	mean(γ_i)	min(γ_i)	max(γ_i)
k=1 all obs.	0.75	1437.71	3744.15	0.69	0.61	0.73
k=f(x) all obs.	0.74	943.55	44.86	0.67	0.38	0.98
k=1 sel. obs.	0.75	1401.29	3331.46	0.71	0.63	0.75
k=f(x) sel. obs.	0.75	995.00	41.08	0.69	0.38	0.98

Table 2: Unit-level model characteristics. k=1: no heteroscedasticity; k=f(x): heteroscedasticity considered; all obs.: all observations; sel. obs.: selected observations (without influential observations).

SE of EBLUP estimates assuming homoscedasticity are almost exclusively dependent on the number of observations within the stands (eqs. (9) and (A.4)), which made them appear rather unrealistic for the given data set. Most of the SE follow two imaginary lines in Fig. 5 A (black hollow squares) for stands with 7 sample plots (SE approximately 20 m^3/ha) and 5 sample plots (SE approximately 23 m^3/ha). Deviations from the two lines are visible for four stands with 6 sample plots (SE between 20 and 23 m^3/ha) or 4 sample plots (SE greater than 23 m^3/ha). Some smaller deviations are caused by the secondary error terms (eqs. (A.5) and (A.6)). The influence of sample size on SE of EBLUP estimates was visible because the sampling fraction within stands was variable. SE of synthetic estimates (the same stands ignoring the sample plots) were, on average, almost twice as large as SE of EBLUP estimates (the same stands considering the sample plots) and, as expected, increased toward the extremes of the explanatory variable (Fig. 5 A, black hollow triangles).

SE of EBLUP estimates considering heteroscedasticity showed an increasing trend over the explanatory variable (Fig. 5 A, red filled dots) and were, on average, smaller than SE of EBLUP estimates assuming homoscedasticity. SE of synthetic estimates considering heteroscedasticity (Fig. 5 A, red filled triangles) exhibited a stronger increase toward the maximum than toward the minimum of the explanatory variable. SE of synthetic estimates were always larger than SE for stands with observations but the difference decreased with increasing estimates (Fig. 5 A, red filled symbols).

	mean(SE)	min(SE)	max(SE)	mean(SE%)	min(SE%)	max(SE%)
k=1 all obs.	21.48	20.11	24.45	22.30	5.15	219.86
k=f(x) all obs.	17.42	4.92	25.89	13.55	5.55	100.74
k=1 all obs. synth.	38.64	38.35	40.59	35.09	8.91	199.87
k=f(x) all obs. synth.	31.38	31.08	34.67	25.43	7.51	110.30

Table 3: Standard errors (SE, m^3/ha) and relative SE (%) of unit-level estimates. k=1: no heteroscedasticity; k=f(x): heteroscedasticity considered; all obs.: all observations; synth.: synthetic estimate.

Considering the small population size (eq. 10) had only minor effects for most stands (see AppendixB.2).

3.3.3. Area-level estimators

Direct (SRS) estimates of stand-level timber volume ($\hat{\bar{Y}}_i^D$) are a part of area-level EBLUP estimates (eq. 13), and ranged from 0.6 to 547.6 with a mean of 193.5 m^3/ha. Standard errors (SE) including finite population correction (fpc, eq. 4) were on average 2 percentage points (pp) smaller than SE not considering fpc (Tab. 4).

	mean(SE)	min(SE)	max(SE)	mean(SE%)	min(SE%)	max(SE%)
SRS	34.35	0.38	114.47	22.37	8.33	81.33
SRS fpc	32.67	0.37	107.61	21.24	7.85	74.58

Table 4: Standard errors (SE) in m^3/ha and SE relative to the estimate (%) of direct estimates (SRS) with and without finite population correction (fpc).

Differences between mean estimates and SE of area-level EBLUP estimates based on models including all observations or omitting one influential observation were less than 1%. Results for omitting one influential observation are therefore not reported. This suggests that a minor violation of model assumptions as indicated by test statistics (the null-hypothesis of normal distributed residuals was rejected when including one influential observation) was of little practical relevance. Some model characteristics can be found in Table 5.

Considering heteroscedasticity ($b_i = x_{2i}^* + 0.39$) or not ($b_i = 1$), had little effect on the fixed-effects parameter estimates $\hat{\boldsymbol{\beta}}$ (Fig. B.8 A), however EBLUP estimates (Fig. B.8 B) and SE changed considerably (Fig. B.9, Tab. 6). EBLUP estimates considering heteroscedasticity ranged from 0.6 m^3/ha to 411.0 m^3/ha with a mean of 186.8 m^3/ha.

Assuming heteroscedasticity in the model, resulted in almost a doubling of the random-effect variance (Fig. B.9, Tab. 5), and thus in more weight on the direct estimator (eq. (13)), and consequently a larger SE of EBLUP estimates. This effect was most noticable in stands with large SE (Fig. B.9 A). For synthetic estimates (assuming observations were not available), the SE increased on average by 12 pp (Fig. B.9, Tab. 6).

In tendency, $\hat{\gamma}_i$ decreased with increasing EBLUP estimates which is the reason for similar SE of area-level estimates and direct (SRS) estimates up to direct SE of approximately 30 m^3/ha (Fig. B.9 B). For the smallest mean estimate, $\hat{\gamma}_i$ was close to 1 such that the EBLUP estimate was approximately equal to the direct (SRS) estimate (Tab. 5).

	R^2	σ_v^2	mean(γ_i)	min(γ_i)	max(γ_i)
b=1	0.78	1116.54	0.57	0.09	1.00
b=f(x)	0.78	2002.78	0.61	0.19	1.00

Table 5: Area-level model characteristics for models including one influential observation. b=1: no heteroscedasticity; b=f(x): heteroscedasticity considered.

	mean(SE)	min(SE)	max(SE)	mean(SE%)	min(SE%)	max(SE%)
b=1 RV	23.10	0.37	35.92	17.31	7.39	74.05
b=f(x) RV	25.43	0.37	53.70	17.73	7.61	72.67
b=1 synth.	34.39	33.87	37.31	40.41	8.86	390.72
b=f(x) synth.	45.46	45.01	47.84	53.50	11.33	521.73

Table 6: Standard errors (SE, m^3/ha) and SE relative to the estimate (%) of area-level estimates. b=1: no heteroscedasticity; b=f(x): heteroscedasticity considered; RV: variance of the direct estimate considered; synth.: synthetic estimate.

4. Discussion

Individual tree crown approaches (ITC), area-based approaches (ABA), or domain-level approaches (DLA) can utilize explanatory variables obtained from fine-resolution 3D remotely sensed data such as airborne laser scanning (ALS) or digital aerial photogrammetry (AP). The three approaches can be linked to two model types (unit-level and area-level) and their associated model-based estimators. Unit-level and area-level EBLUP estimators are weighted averages of estimators that are exclusively dependent on the fitted linking model (synthetic estimators) and estimators that are based on the sample units within a domain. The weight depends on the model accuracy which again depends on the quality of the explanatory variables. In a case study, unit-level and area-level estimators were used for inference on mean timber volume within stands using linear linking models. Adjustments to the described methods would be required if non-linear models were to be applied (Rao and Molina, 2015, Ch. 4.6).

Satisfying all model assumptions (e.g., homogeneity of variance, normality of residual dispersion) using real data can be difficult as we learned from our case study. However, small deviations from model assumptions may be inconsequential. For important decisions that require large accuracy, investing more in field sample plots deserves consideration. Due to their robustness against model-misspecification, model-assisted (design-based) estimators (e.g. Mandallaz, 2013; McRoberts et al., 2014) may be an interesting alternative to the model-based estimators used here if sufficient sample sizes per domain are available.

As observed in other studies (Hidiroglou and You, 2016; Mauro et al., 2017b), mean estimates from unit-level and area-level estimators can be similar. Furthermore, when satisfying the model assumptions unit-level estimators, on average, result in smaller standard errors (SE) than area-level estimators and SE decrease with domain size which is an intuitive property (Breidenbach et al., 2015). Furthermore, unit-level models can be used to generate sub-domain maps of forest resources. For these reasons, unit-level estimators may be preferred over area-level estimators, if the data afford their use.

However, the field data acquisition for area-level estimators can be considerably cheaper than for unit-level estimators, because exact sample locations are not required and efficient plot designs such as variable radius plots can be used without compromising the link to remotely sensed data. The reduced costs per sampling unit can be used toward a larger field sample which, in turn, may improve the precision of area-level estimates relative to unit-level estimates under a given budget.

Heteroscedasticity has a strong influence on the SE of unit-level and area-level estimates. If observed, it should be considered in the models to avoid violated assumptions and unrealistic SE. However, selecting the constants k_{ij} (unit-level) and b_i (area-level) for considering heteroscedasticity is a delicate matter because they affect standard errors of estimates and, to a smaller degree, also the estimates themselves. Therefore, we based the selection of these constants on objective methods that aim at fulfilling the model assumptions. Based on a visual inspection of residuals, Mauro et al. (2017b) selected similar values for the constants k_{ij} of unit-level models for estimating timber volume.

While the challenge of the chosen method for considering heteroscedasticity in unit-level models is to select an adequate value of the constant k_{ij}, the challenges of an alternative method proposed by Jiang and Nguyen (2012) is that the continuous nature of the heteroscedasticity is categorized and the number of categories has to be selected. Transformation of the response variable is another method that can help meeting model assumptions (e.g. Næsset, 1997). Inference on domain means using transformed variables would, however, require modifications to the methods used here due to the required back-transformation (Rao and Molina, 2015, p. 140).

Considering heteroscedasticity in SAE deserves more attention. A reason for the limited attention to this topic in area-level estimators is that heteroscedasticity is naturally included because of their relation to the direct estimator. Furthermore, synthetic estimates using area-level estimators are often not considered when observations are available for each domain (Hidiroglou and You, 2016; Molina and Marhuenda, 2015). However, in forest inventories, the majority of stands may not contain any sample plots due to small sampling fractions and thus require synthetic estimates.

Remotely sensed data play a pivotal role in the discussed methods as they provide auxiliary information which are highly correlated to the variables of interest. Without the availability of auxiliary information, synthetic estimates (for stands without sample plots) would be of little practical relevance. With the availability of remotely sensed data that are less closely related to the variables of interest than the AP data we had available (for example Landsat images), synthetic estimates need to be used with care as they can have large systematic errors. Due to the advantages of coarser-resolution remotely sensed data, such as easier data handling, they may nonetheless be useful, especially for estimates on larger scales. With the availability of remotely sensed data that are more closely related to the variables of interest than the AP data we had available, issues of heteroscedasticity may be reduced.

To exploit remotely sensed data for estimation of volume and biomass we need auxiliary information that provide information proxies for stand structures related to densities of volume and biomass. Canopy height metrics and metrics related to canopy density, as provided by 3D remotely sensed data such as ALS and AP, are therefore key. But of course, environmental auxiliary information may also be useful, but typically more so the stronger the environmental gradients within the study region are expressed. For example, exposure and elevation can be useful predictors, but only when the vegetation is strongly influenced by these factors and if other auxiliary information (such as 3D remotely sensed data) have not already explained the variation in the vegetation. In sum, the scale of the study, and the study environment dictates the utility of the available auxiliaries.

5. Conclusions

The following conclusions can be drawn from this study. i) If present, including heteroscedasticity in models used for unit- and area-level SAE is important for obtaining realistic measures of precision. ii) SAE under heteroscedasticity should be studied more, especially for area-level estimation. The same is true for synthetic estimation in domains without samples, which is uncommon in area-level SAE. iii) On average, unit-level estimates can be expected to be more precise than area-level estimates. However, if the direct estimate (based on sample units only) has high precision (e.g., due to sufficient number of sample units and small variability such as in young stands), area-level estimates can have greater precision than unit-level estimates. iv) The use of digital aerial photogrammetry data considerably improved the precision of estimates.

6. Acknowledgments

This study was supported by the Norwegian Institute of Bioeconomy Research (NIBIO) and Nordic Forest Research (SNS) over the Centre of Advanced Research for the innovative use of 3D remote sensing in mapping of forest and landscape attributes based on national forest inventories (CARISMA). We would like to thank the editor and three anonymous reviewers for constructive comments.

7. References

Battese, G. E., Harter, R. M., Fuller, W. A., 1988. An error-components model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association 83 (401), 28–36.

Boubeta, M., Lombardía, M. J., Marey-Pérez, M. F., Morales, D., 2015. Prediction of forest fires occurrences with area-level poisson mixed models. Journal of environmental management 154, 151–158.

Breidenbach, J., 2013. JoSAE: Functions for some unit-level small area estimators and their variances. R package version 0.2.2.
URL http://CRAN.R-project.org/package=JoSAE

Breidenbach, J., Astrup, R., 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. European Journal of Forest Research 131 (4), 1255–1267.

Breidenbach, J., McRoberts, R. E., Astrup, R., 2015. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sensing of Environment 173, 274–281.

Breusch, T. S., Pagan, A. R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society, 1287–1294.

Bäuerle, H., Nothdurft, A., Kändler, G., Bauhus, J., 11 2009. Monitoring habitat trees and coarse woody debris based on sampling schemes. Allgemeine Forst und Jagdzeitung 180, 249–259.

Chen, Q., McRoberts, R. E., Wang, C., Radtke, P. J., 2016. Forest aboveground biomass mapping and estimation across multiple spatial scales using model-based inference. Remote Sensing of Environment 184, 350–360.

Fay, R. E., Herriot, R. A., 1979. Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association 74 (366a), 269–277.

Goerndt, M. E., Monleon, V. J., Temesgen, H., 2011. A comparison of small-area estimation techniques to estimate selected stand attributes using lidar-derived auxiliary variables. Canadian Journal of Forest Research 41 (6), 1189–1201.

Goerndt, M. E., Monleon, V. J., Temesgen, H., 2013. Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. Forest Science 59 (5), 536–548.

Gregoire, T. G., Næsset, E., McRoberts, R. E., Ståhl, G., Andersen, H.-E., Gobakken, T., Ene, L., Nelson, R., 2016. Statistical rigor in lidar-assisted estimation of aboveground forest biomass. Remote Sensing of Environment 173, 98–108.

Hidiroglou, M. A., You, Y., 2016. Comparison of unit level and area level small area estimators. Survey Methodology 42, 41–61.

Jiang, J., Nguyen, T., 2012. Small area estimation via heteroscedastic nested-error regression. Canadian Journal of Statistics 40 (3), 588–603.

Kirchhoefer, M., Schumacher, J., Adler, P., Kändler, G., Jul 2017. Considerations towards a novel approach for integrating angle-count sampling data in remote sensing based forest inventories. Forests 8 (7), 239.
URL http://dx.doi.org/10.3390/f8070239

Landsskogtakseringen, 2008. Landsskogtakseringens feltinstruks 2008, Håndbok fra Skog og landskap 05/08. Skog og landskap, Ås, Norway, 152 p.

Magnussen, S., Fransisco, M., Breidenbach, J., Lanz, A., Kändler, G., 2017. Area-level analysis of forest inventory variables. European Journal of Forest Research 136 (5), 839–855.

Magnussen, S., Frazer, G., Penner, M., 2016. Alternative mean-squared error estimators for synthetic estimators of domain means. Journal of Applied Statistics, 1–24.

Mandallaz, D., 2013. Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Canadian Journal of Forest Research 43 (5), 441–449.

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., Ayuga-Téllez, E., 2016. Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics 27 (4), 225–238.

Mauro, F., Monleon, V., Temesgen, H., Ruiz, L., 2017a. Analysis of spatial correlation in predictive models of forest variables that use lidar auxiliary information. Canadian Journal of Forest Research 47 (6), 788–799.
URL https://doi.org/10.1139/cjfr-2016-0296

Mauro, F., Monleon, V. J., Temesgen, H., Ford, K. R., 12 2017b. Analysis of area level and unit level models for small area estimation in forest inventories assisted with lidar auxiliary information. PLOS ONE 12 (12), 1–14.

McGaughey, R., 2014. Fusion. USDA forest service.

McRoberts, R., Andersen, H.-E., Næsset, E., 2014. Using airborne laser scanning data to support forest sample surveys. In: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning. Springer, pp. 269–292.

McRoberts, R. E., 2006. A model-based approach to estimating forest area. Remote Sensing of Environment 103 (1), 56–66.

Militino, A., Ugarte, M., Goicoa, T., 2007. A blup synthetic versus an eblup estimator: an empirical study of a small area estimation problem. Journal of Applied Statistics 34 (2), 153–165.

Militino, A., Ugarte, M., Goicoa, T., González-Audícana, M., 2006. Using small area models to estimate the total area occupied by olive trees. Journal of agricultural, biological, and environmental statistics 11 (4), 450–461.

Molina, I., Marhuenda, Y., jun 2015. sae: An R package for small area estimation. The R Journal 7 (1), 81–98.
URL `http://journal.r-project.org/archive/2015-1/molina-marhuenda.pdf`

Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment 61 (2), 246–253.

Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyyppä, J., Saari, H., Pölönen, I., Imai, N., et al., Feb 2017. Individual tree detection and classification with uav-based photogrammetric point clouds and hyperspectral imaging. Remote Sensing 9 (3), 185.
URL `http://dx.doi.org/10.3390/rs9030185`

Næsset, E., 2014. Area-based inventory in Norway - from innovation to an operational reality. In: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning. Springer, pp. 215–241.

Prasad, N., Rao, J., 1990. The estimation of the mean squared error of small-area estimators. Journal of the American statistical association 85 (409), 163–171.

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R., 2014. Comparison of four types of 3d data for timber volume estimation. Remote Sensing of Environment 155, 325–333.

Rao, J. N., Molina, I., 2015. Small area estimation, 2nd Edition. John Wiley & Sons.

Rao, J. N. K., 2003. Small area estimation. Wiley-Interscience.

Rivest, L.-P., Vandal, N., 2003. Mean squared error estimation for small areas when the small area variances are estimated. In: Proceedings of the International Conference on Recent Advances in Survey Sampling. Laboratory for Research in Statistics and Probability, Carleton University Ottawa, ON, Canada.

Saarela, S., Breidenbach, J., Raumonen, P., Grafström, A., Ståhl, G., Ducey, M. J., Astrup, R., 2017. Kriging prediction of stand-level forest information using mobile laser scanning data adjusted for nondetection. Canadian Journal of Forest Research 47 (9), 1257–1265. URL `https://doi.org/10.1139/cjfr-2017-0019`

Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., Nordkvist, K., Hyyppä, J., 2015. Model-assisted estimation of growing stock volume using different combinations of lidar and landsat data as auxiliary information. Remote Sensing of Environment 158, 431–440.

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T. G., Nelson, R. F., Ståhl, G., 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. Annals of Forest Science 73 (4), 895–910.

van Aardt, J., Wynne, R., Oderwald, R., 2006. Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segment basis. Forest Science 52 (6), 636–649.

Wang, J., Fuller, W. A., 2003. The mean squared error of small area predictors constructed with estimated area variances. Journal of the American Statistical Association 98, 716–723.

AppendixA. Methods

AppendixA.1. Unit-level EBLUP estimators

AppendixA.1.1. Transformed and scaled residuals

The transformed residuals (σ_v-residuals) are given by

$$u_{ij} = \frac{(y_{ij} - \hat{\tau}\hat{\bar{Y}}_i^D) - (\boldsymbol{x}_{ij} - \hat{\tau}\bar{\boldsymbol{x}}_i)^T \hat{\boldsymbol{\beta}}}{k_{ij}} \tag{A.1}$$

with $\hat{\tau} = 1 - \sqrt{1-\hat{\gamma}}$ where $\hat{\gamma} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2}$ and $\bar{\boldsymbol{x}}_i$ is the sample mean of the explanatory variables (Militino et al., 2006). The standardized residuals (σ_e-residuals) are given by

$$\varepsilon_{ij} = \frac{e_{ij}}{k_{ij}\hat{\sigma}_e} \tag{A.2}$$

where e_{ij} is an empirical residual.

AppendixA.1.2. Estimating uncertainty

For large populations (Rao and Molina, 2015, Ch. 7.2), the (unconditional) MSE is estimated as the sum of three terms

$$\widehat{\mathrm{MSE}}(\hat{\mu}_i^{\mathrm{UE}}) = g_{1i} + g_{2i} + 2g_{3i} \tag{A.3}$$

where

$$g_{1i} = (1 - \hat{\gamma}_i)\hat{\sigma}_v^2 \tag{A.4}$$

describes the influence of the random-effect variance (Militino et al., 2007). Because g_{1i} is the leading term (has the most influence on the MSE) and due to the structure of $\hat{\gamma}_i$ (eq. 9), the variability of the MSE among domains is almost exclusively affected by the number of samples within a domain (n_i) in the case of homoscedasticity ($k_{ij} = 1$).

The term g_{2i} describes the uncertainty due to the estimation of the fixed-effects parameters $\boldsymbol{\beta}$

$$g_{2i} = (\bar{\boldsymbol{x}}_{iP} - \hat{\gamma}_i\bar{\boldsymbol{x}}_{ia})^T \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}})(\bar{\boldsymbol{x}}_{iP} - \hat{\gamma}_i\bar{\boldsymbol{x}}_{ia}) \tag{A.5}$$

where $\bar{\boldsymbol{x}}_{ia} = \sum_{j=1}^{n_i} a_{ij}\boldsymbol{x}_{ij}/a_{i\cdot}$ is the weighted sample mean of the explanatory variables and $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}})$ is the covariance matrix of the estimated fixed-effect parameters $\hat{\boldsymbol{\beta}}$.

The term g_{3i} describes the uncertainty due to the estimation of σ_v^2 and σ_e^2

$$g_{3i} = \frac{\hat{\sigma}_e^4 \overline{V}_v + \hat{\sigma}_v^4 \overline{V}_e - 2\hat{\sigma}_e^2 \hat{\sigma}_v^2 \overline{V}_{ve}}{a_{i\cdot}^2 (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 a_{i\cdot})^3} \tag{A.6}$$

where $\overline{V}_v, \overline{V}_e$ and $\overline{V}_{ve}$ are the asymptotic variance and covariance estimates of $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$, respectively.

For small populations (Rao and Molina, 2015, Ch. 7.2.3), the MSE is estimated by

$$\widehat{\text{MSE}}(\hat{\bar{Y}}_i^{\text{UE}}) = (1 - f_i)^2 \widetilde{\text{MSE}}(\hat{\mu}_i^{\text{UE}}) + g_{4i} \tag{A.7}$$

where

$$\widetilde{\text{MSE}}(\hat{\mu}_i^{\text{UE}}) = g_{1i} + \tilde{g}_{2i} + 2g_{3i}. \tag{A.8}$$

$\tilde{g}_{2i}$ is obtained by substituting the population mean of the explanatory variables $\bar{\boldsymbol{x}}_{iP}$ with the mean of the the explanatory variables of the non-sampled population units $\bar{\boldsymbol{x}}_{iR}$ in g_{2i} (eq. A.5). The fourth component of the MSE estimate considers the residual error variance

$$g_{4i} = g_{4i}^* \hat{\sigma}_e^2 \tag{A.9}$$

where

$$g_{4i}^* = \frac{\boldsymbol{k}_{iR}^T \boldsymbol{k}_{iR}}{N_i^2} \tag{A.10}$$

is the proportion to which the residual error variance is incorporated into the MSE, with $\boldsymbol{k}_{iR}$ as the vector of constants for considering heteroscedasticity of the non-sampled population units. Under homoscedasticity, the term g_{4i}^* reduces to $g_{4i}^* = N_i^{-2}(N_i - n_i)$ which means that it is not necessary to know which population unit is part of the sample in that case. The term g_{4i} can also be used to consider spatial autocorrelation (Mauro et al., 2017a; Breidenbach et al., 2015), which is, however, outside the scope of this study.

The MSE of the synthetic estimator (eq. 8) results from

$$\widehat{\text{MSE}}(\hat{\mu}_i^{\text{US}}) = g_{1i} + g_{2i} + g_{4i} \tag{A.11}$$

by setting $\hat{\gamma}_i := 0$ and noting that $\boldsymbol{k}_{iR} = \boldsymbol{k}_{iP}$ and $n_i = 0$ in g_{4i} (eq. (A.9)). Under homoscedasticity, g_{4i} reduces to

$$g_{4i} = \frac{\hat{\sigma}_e^2}{N_i} \tag{A.12}$$

(McRoberts, 2006; Breidenbach et al., 2015). From this form of the component g_{4i} it becomes clear how quickly the influence of the residual error variance reduces with domain size and that g_{4i} can be ignored for large populations.

AppendixA.2. Area-level EBLUP estimators

AppendixA.2.1. Estimating model parameters

The fixed model parameters are estimated by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \frac{\bar{\boldsymbol{x}}_{iP} \bar{\boldsymbol{x}}_{iP}^T}{\hat{\vartheta}} \right)^{-1} \left(\sum_{i=1}^{m} \frac{\bar{\boldsymbol{x}}_{iP} \hat{\bar{Y}}_i^D}{\hat{\vartheta}} \right) \tag{A.13}$$

where $\hat{\vartheta} = \hat{\sigma}_i^2 + \hat{\sigma}_v^2 b_i^2$ is the estimated total model variance (Rao and Molina, 2015, Ch. 6.1.1).

The random effects are indirectly estimated by

$$\hat{\upsilon}_i = \frac{\hat{\bar{Y}}_i^D - \hat{\mu}_i^{\mathrm{AE}}}{b_i} \tag{A.14}$$

and scaled random effects result from

$$\hat{\varphi}_i = \frac{\hat{\upsilon}_i}{\hat{\sigma}_v^2}. \tag{A.15}$$

AppendixA.2.2. Estimating uncertainty

The MSE of area-level estimates as described by Fay and Herriot (1979) can be estimated by adding the terms

$$\widehat{\mathrm{MSE}}(\tilde{\mu}_i^{\mathrm{AE}}) = g_{1i} + g_{2i} + 2g_{3i} \tag{A.16}$$

where

$$g_{1i} = (1 - \hat{\gamma}_i)\hat{\sigma}_v^2 \tag{A.17}$$

reflects the uncertainty in the random effect,

$$g_{2i} = (1 - \hat{\gamma}_i)^2 \bar{\boldsymbol{x}}_{iP}^T \widehat{Cov}(\hat{\boldsymbol{\beta}}) \bar{\boldsymbol{x}}_{iP} \tag{A.18}$$

reflects the uncertainty in the fixed model parameter estimates $\hat{\boldsymbol{\beta}}$, and

$$g_{3i} = \hat{\sigma}_i^4 (\hat{\sigma}_i^2 + \hat{\sigma}_v)^{-3} \overline{V}_v \tag{A.19}$$

reflects the uncertainty due to estimating the random effect, where $\overline{V}_v$ is the asymptotic variance of the random effect.

The assumptions in the MSE estimator (A.16) include that the variance of the direct estimate is known without uncertainty. In practice, the variance $\hat{\sigma}_i^2$ is estimated (i.e., is not known without uncertainty) and is either directly plugged-in to (A.16) as described in the equations above or after smoothing. However, Wang and Fuller (2003) described general estimators for area-level models where the assumptions about known model parameters are relaxed. Similarly, Rivest and Vandal (2003) developed estimators for the special case where the direct estimate is based on unit-level samples as in our case study. To account for the uncertainty in the variance of the direct estimate $\hat{\sigma}_i^2$, an additional term is added to the MSE estimator (A.16)

$$\widehat{\mathrm{MSE}}_{RV}(\hat{\mu}_i^{\mathrm{AE}}) = \widehat{\mathrm{MSE}}(\hat{\mu}_i^{\mathrm{AE}}) + 2(\hat{\sigma}_v^2 + \hat{\sigma}_i^2)^{-3}\hat{\sigma}_v^4\hat{\delta} \tag{A.20}$$

where $\hat{\delta} = 2\hat{\sigma}_i^4/(n_i - 1)$.

For domains without samples, the variance of the synthetic estimate (Rao and Molina, 2015, Ch. 6.2.2) is the sum of the random effect variance estimate and the uncertainty in the model parameter estimates

$$\widehat{\mathrm{MSE}}(\tilde{\mu}_i^{\mathrm{AS}}) = g_{1i} + g_{2i} \tag{A.21}$$

by setting $\hat{\gamma}_i := 0$.

AppendixB. Results

AppendixB.1. Unit-level EBLUP estimates

AppendixB.2. Unit-level EBLUP estimates for small populations

For small populations, the samples have to be considered in the EBLUP estimates (eq. 10). The latter were therefore more variable (Fig. B.7 A) than estimates assuming large populations. EBLUP estimates of stand-level timber volume ranged from 3.35 to 440.89 with a mean of 190.35 m^3/ha. Furthermore, the residual error term has to be considered in the standard errors (eq. A.7) which had different effects for models considering heteroscedasticity or not (Fig. B.7 B and Tab. 3). For models assuming homoscedasticity, the standard errors were generally smaller when considering the sampling fraction. For models considering heteroscedasticity, the standard errors for small populations were often smaller for $x_{2ij} < 100$ dm, but in tendency larger otherwise. The reason for this difference is that the residual error term receives more weight for larger estimates

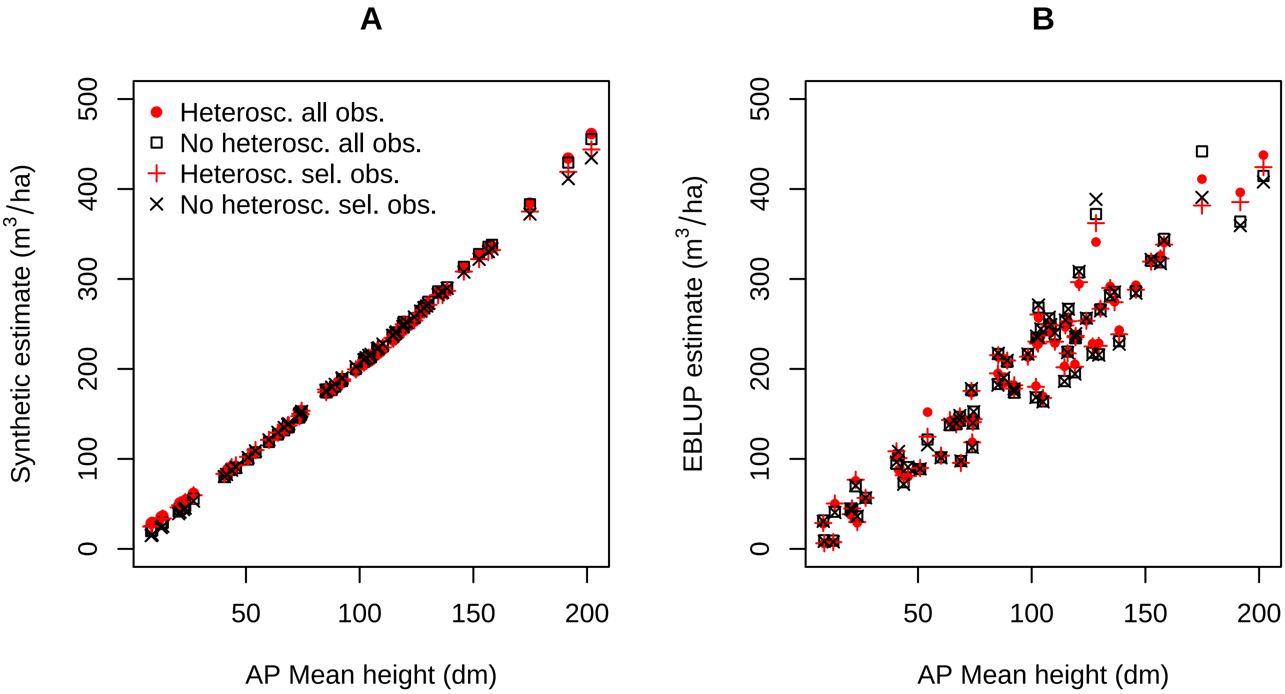

Figure B.6: Mean timber volume estimates using unit-level models for 64 stands. Synthetic estimates (A) using all observations or selected observations (omitting influential observations). EBLUP estimates (B).

640 due to heteroscedasticity which counter-acts the reduction of the SE by considering the sampling
641 fraction (eq. (A.7)).

	mean(SE)	min(SE)	max(SE)	mean(SE%)	min(SE%)	max(SE%)
k=1 all obs.	21.48	20.11	24.45	22.30	5.15	219.86
k=f(x) all obs.	17.42	4.92	25.89	13.55	5.55	100.74
k=1 sel. obs.	20.61	19.17	23.43	22.45	5.24	247.70
k=f(x) sel. obs.	17.29	4.72	25.84	13.28	5.55	95.35
k=1 all obs. synth.	38.64	38.35	40.59	35.09	8.91	199.87
k=f(x) all obs. synth.	31.38	31.08	34.67	25.43	7.51	110.30
SP k=1 all obs.	20.79	18.95	24.22	21.49	4.96	214.06
SP k=f(x) all obs.	17.18	5.01	25.96	13.30	5.50	95.74

Table B.7: Standard errors (SE, m^3/ha) and relative SE (%) of unit-level estimates. k=1: no heteroscedasticity; k=f(x): heteroscedasticity considered; all obs.: all observations; sel. obs.: selected observations (without influential observations); synth.: synthetic estimate; SP: considering small population size.

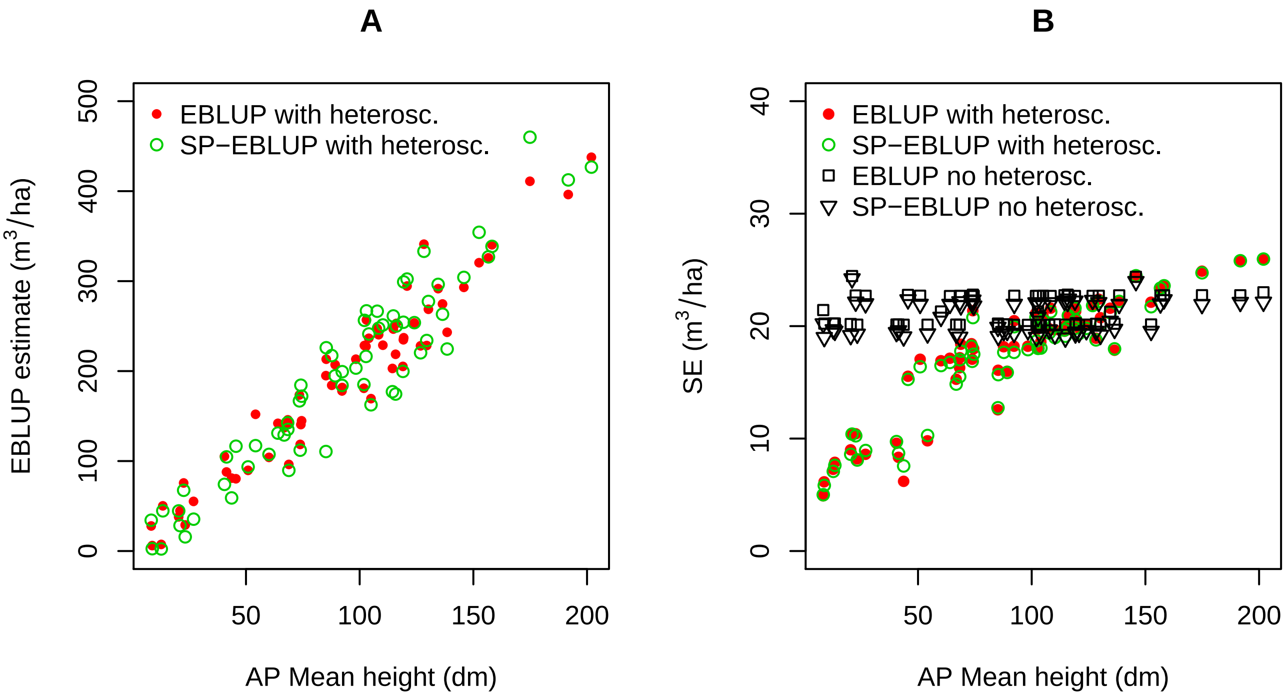

Figure B.7: EBLUP estimates assuming small populations (SP-EBLUP) and EBLUP estimates (A). Standard errors of EBLUP estimates (B).

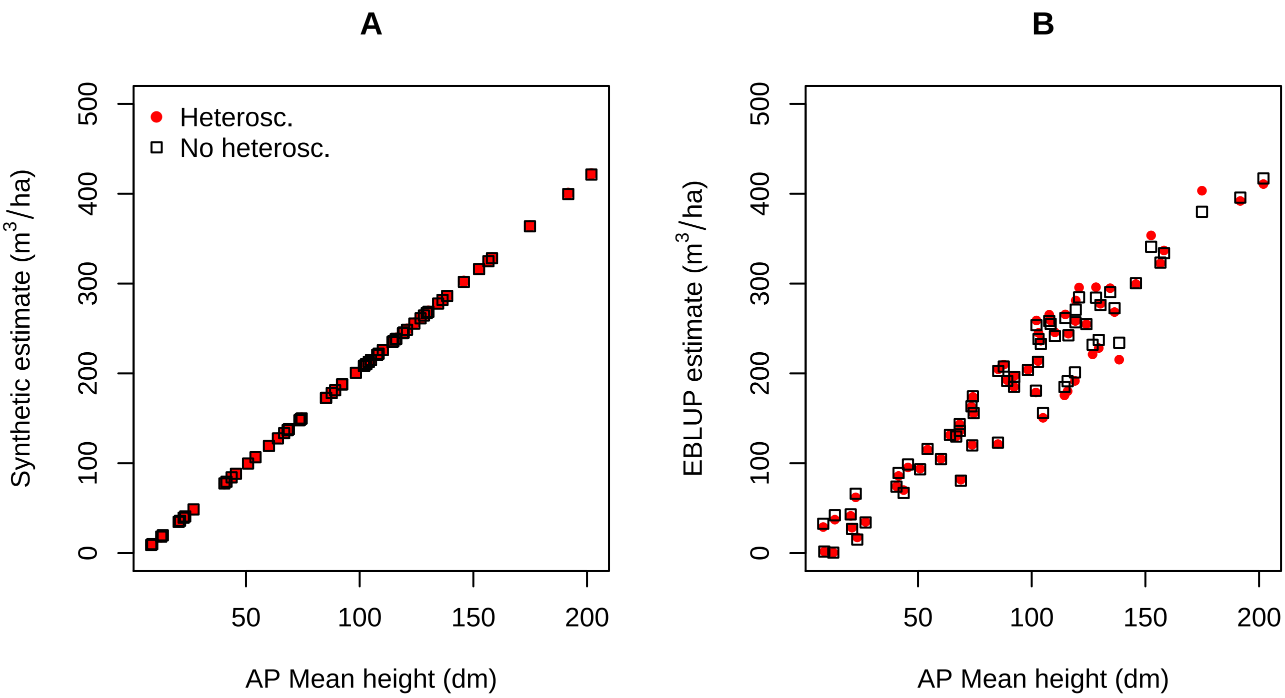

Figure B.8: Mean timber volume estimates using area-level models for 64 stands. Synthetic estimates (A), EBLUP estimates (B).

AppendixB.3. Area-level estimates

SE of estimates considering the variance of the direct estimator ($\widehat{\mathrm{MSE}}_{RV}(\hat{\mu}_i^{\mathrm{AE}})$) were similar but slightly larger than SE not considering the variance of the direct estimator ($\widehat{\mathrm{MSE}}_{FH}(\hat{\mu}_i^{\mathrm{AE}})$) (Tab. B.8).

	mean(SE)	min(SE)	max(SE)	mean(SE%)	min(SE%)	max(SE%)
b=1 FH	21.44	0.37	34.38	16.33	6.96	74.01
b=f(x) FH	23.45	0.37	51.48	16.62	7.28	72.65
b=1 RV	23.10	0.37	35.92	17.31	7.39	74.05
b=f(x) RV	25.43	0.37	53.70	17.73	7.61	72.67
b=1 synth.	34.39	33.87	37.31	40.41	8.86	390.72
b=f(x) synth.	45.46	45.01	47.84	53.50	11.33	521.73

Table B.8: Standard errors (SE, m^3/ha) and SE relative to the estimate (%) of area-level models. b=1: no heteroscedasticity; b=f(x): heteroscedasticity considered; FH: variance of the direct estimate not considered; RV: variance of the direct estimate considered; synth.: synthetic estimate.

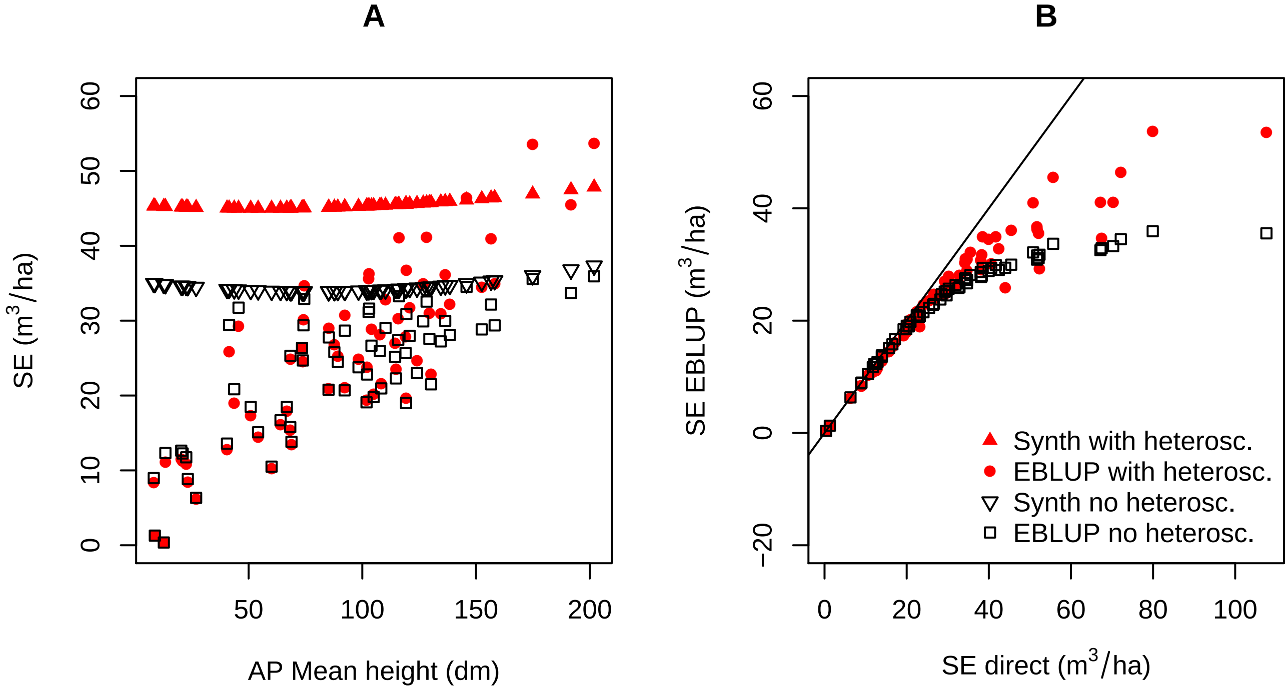

Figure B.9: Standard errors (SE) versus AP mean height for area-level estimates including all observations (A) and SE of EBLUP estimates vs SE of direct estimates (B).

LIST OF FIGURE CAPTIONS

Figure 1 Left: Location of the study area within Norway (red square). Right: Location of the sample plots within the forest areas of Holmestrand (orange), Lardal (blue), and Stokke (green).

Figure 2 Transformed and scaled residuals (see AppendixA.1.1, including influential observations) versus predicted values for the unit-level model with $k_{ij} = 1$ (A and B), and $k_{ij} = x_{2ij}^{0.48}$ (C and D).

Figure 3 Scaled random effects (including one influential observation) versus area-level EBLUP estimates on stand-level with $b_i = 1$ (A), and $b_i = \bar{x}_{2iP} + 0.39$ (B).

Figure 4 Stand-level estimates assuming small populations (SP-EBLUP) based on unit-level models (A) and area-level models (B). SRS = direct estimates.

Figure 5 Standard errors versus AP mean height for unit-level (A) and area-level estimates (B).

Figure B.6 Mean timber volume estimates using unit-level models for 64 stands. Synthetic estimates (A) using all observations or selected observations (omitting influential observations). EBLUP estimates (B).

Figure B.7 EBLUP estimates assuming small populations (SP-EBLUP) and EBLUP estimates (A). Standard errors of EBLUP estimates (B).

Figure B.8 Mean timber volume estimates using area-level models for 64 stands. Synthetic estimates (A), EBLUP estimates (B).

Figure B.9 Standard errors (SE) versus AP mean height for area-level estimates including all observations (A) and SE of EBLUP estimates vs SE of direct estimates (B).