Hierarchical model-based inference for forest inventory utilizing three sources of information

Svetlana Saarela, Sören Holm, Anton Grafström, Sebastian Schnell, Erik Næsset, Timothy G. Gregoire, Ross F. Nelson, Göran Ståhl

Annals of Forest Science (2016) 73:895–910
DOI 10.1007/s13595-016-0590-1

ORIGINAL PAPER

Hierarchical model-based inference for forest inventory utilizing three sources of information

Svetlana Saarela[1] · Sören Holm[1] · Anton Grafström[1] · Sebastian Schnell[1] · Erik Næsset[2] · Timothy G. Gregoire[3] · Ross F. Nelson[4] · Göran Ståhl[1]

Received: 26 February 2016 / Accepted: 13 October 2016 / Published online: 16 November 2016

Abstract

• ***Key message*** **The study presents novel model-based estimators for growing stock volume and its uncertainty estimation, combining a sparse sample of field plots, a sample of laser data, and wall-to-wall Landsat data. On the basis of our detailed simulation, we show that when the uncertainty of estimating mean growing stock volume on the basis of an intermediate ALS model is not accounted for, the estimated variance of the estimator can be biased by as much as a factor of three or more, depending on the sample size at the various stages of the design.**

• ***Context*** This study concerns model-based inference for estimating growing stock volume in large-area forest inventories, combining wall-to-wall Landsat data, a sample of laser data, and a sparse subsample of field data.

• ***Aims*** We develop and evaluate novel estimators and variance estimators for the population mean volume, taking into account the uncertainty in two model steps.

• ***Methods*** Estimators and variance estimators were derived for two main methodological approaches and evaluated through Monte Carlo simulation. The first approach is known as two-stage least squares regression, where Landsat

Handling Editor: Jean-Michel Leban

✉ Svetlana Saarela
svetlana.saarela@slu.se

Sören Holm
soren.holm@gronstenen.se

Anton Grafström
anton.grafstrom@slu.se

Sebastian Schnell
sebastian.schnell@slu.se

Erik Næsset
erik.naesset@nmbu.no

Timothy G. Gregoire
timothy.gregoire@yale.edu

Ross F. Nelson
rfn104@gmail.com

Göran Ståhl
goran.stahl@slu.se

[1] Department of Forest Resource Management, Swedish University of Agricultural Sciences, SLU Skogsmarksgränd, SE-90183 Umeå, Sweden

[2] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway

[3] School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA

[4] NASA/Goddard Space Flight Center, Greenbelt, Marylands 20771, USA

data were used to predict laser predictor variables, thus emulating the use of wall-to-wall laser data. In the second approach laser data were used to predict field-recorded volumes, which were subsequently used as response variables in modeling the relationship between Landsat and field data.
• ***Results*** The estimators and variance estimators are shown to be at least approximately unbiased. Under certain assumptions the two methods provide identical results with regard to estimators and similar results with regard to estimated variances.
• ***Conclusion*** We show that ignoring the uncertainty due to one of the models leads to substantial underestimation of the variance, when two models are involved in the estimation procedure.

Keywords Landsat · Large-scale forest inventory · Monte Carlo simulation · Two-stage least squares regression

1 Introduction

During the past decades, the interest in utilizing multiple sources of remotely sensed (RS) data in addition to field data has increased considerably in order to make forest inventories cost efficient (e.g., Wulder et al. 2012). When conducting a forest inventory, RS data can be incorporated at two different stages: the design stage and the estimation stage. In the design stage, RS data are used for stratification (e.g., McRoberts et al. 2002) and unequal probability sampling (e.g., Saarela et al. 2015a), they may be used for balanced sampling (Grafström et al. 2014) aiming at improving estimates of population parameters. To utilize RS data at the estimation stage, either model-assisted estimation (Särndal et al. 1992) or model-based inference (Matérn 1960) can be applied. While model-assisted estimators describe a set of estimation techniques within the design-based framework of statistical inference, model-based inference constitutes is a different inferential framework (Gregoire 1998). When applying model-assisted estimation, probability samples are required and relationships between auxiliary and target variables are used to improve the precision of population parameter estimates. In contrast, the accuracy of estimation when assessed in a model-based framework relies largely on the correctness of the model(s) applied in the estimators (Chambers and Clark 2012). While this dependence on the aptness of the model may be regarded as a drawback, this mode of inference also has advantages over the design-based approach. For example, in some cases, smaller sample sizes might be needed for attaining a certain level of accuracy, and in addition, probability samples are not necessary, which is advantageous for remote areas with limited access to the field.

While several sources of auxiliary information can be applied straightforwardly in the case of model-assisted estimation following established sampling theory (e.g., Gregoire et al. 2011; Massey et al. 2014; Saarela et al. 2015a), this issue has been less well explored for model-based inference for the case when the different auxiliary variables are not available for the entire population. However, recent studies by Ståhl et al. (2011) and Ståhl et al. (2014) and Corona et al. (2014) demonstrated how probability samples of auxiliary data can be combined with model-based inference. This approach was termed "hybrid inference" by Corona et al. (2014) to clarify that auxiliary data were collected within a probability framework.

A large number of studies have shown how several sources of RS data can be combined through hierarchical modeling for mapping and estimation of forest attributes such as growing stock volume (GSV) or biomass over large areas. For example, Boudreau et al. (2008) and Nelson et al. (2009) used a combination of the Portable Airborne Laser System (PALS) and the Ice, Cloud, and land Elevation/Geoscience Laser Altimeter System (ICESat/GLAS) data for estimating aboveground biomass for a 1.3 Mkm^2 forested area in the Canadian province of Québec. A Landsat 7 Enhanced Thematic Mapper Plus (ETM+) land cover map was used to delineate forest areas from non-forest and as a stratification tool. These authors used the PALS data acquired on 207 ground plots to develop stratified regression models linking the biomass response variable to PALS metrics. They then used these ground-PALS models to predict biomass on 1325 ICESat/GLAS pulses that have been overflown with PALS, ultimately developing a regression model linking the biomass response variable to ICESat/GLAS waveform parameters as predictor variables. The latter model was used to predict biomass across the entire Province based on 104044 filtered GLAS shots. A similar approach was applied in a later study by Neigh et al. (2013) for assessment of forest carbon stock in boreal forests across 12.5± 1.5 Mkm^2 for five circumpolar regions – Alaska, western Canada, eastern Canada, western Eurasia, and eastern Eurasia. The latest study of this kind is from Margolis et al. (2015), where the authors applied the approach for assessment of aboveground biomass in boreal forests of Canada (3,326,658 km^2) and Alaska (370,074 km^2). The cited studies have in common that they ignore parts of the models' contribution to the overall uncertainty of the biomass (forest carbon stock) estimators, i.e., they can be expected to underestimate the variance of the estimators.

With non-nested models, the assessment of uncertainty is straightforward. McRoberts (2006) and McRoberts (2010) used model-based inference for estimating forest area using Landsat data as auxiliary information. The studies were performed in northern Minnesota, USA. Ståhl et al. (2011)

presented model-based estimation for aboveground biomass in a survey where airborne laser scanning (ALS) and airborne profiler data were available as a probability sample. The study was performed in Hedmark County, Norway. Saarela et al. (2015b) analysed the effects of model form and sample size on the accuracy of model-based estimators through Monte Carlo simulation for a study area in Finland. However, model-based approaches that account correctly for hierarchical model structures in forest surveys still appear to be lacking.

In this study, we present a model-based estimation framework that can be applied in surveys that use three data sources, in our case Landsat, ALS and field measurements, and hierarchically nested models. Estimators of population means, their variances and corresponding variance estimators are developed and evaluated for different cases, e.g., when the model random errors are homoskedastic and heteroskedastic and when the uncertainty due to one of the model stages is ignored. The study was conducted using a simulated population resembling the boreal forest conditions in the Kuortane region, Finland. The population was created using a multivariate probability distribution copula technique (Nelsen 2006). This allowed us to apply Monte Carlo simulations of repeated sample draws from the simulated population (e.g., Gregoire 2008) in order to analyse the performance of different population mean estimators and the corresponding variance estimators.

2 Simulated population

The multivariate probability distribution copula technique is a popular tool for multivariate modelling. Ene et al. (2012) pioneered the use of this technique to generate simulated populations which mimic real-world, large-area forest characteristics and associated ALS metrics. Copulas are mathematical functions used to model dependencies in complex multivariate distributions. They can be interpreted as d-dimensional variables on $[0, 1]^d$ with uniform margins and are based on Sklar's theorem (Nelsen 2006), which establishes a link between multivariate distributions and their univariate margins. For arbitrary dimensions, multivariate probability densities are often decomposed into smaller building blocks using the pair-copula technique (Aas et al. 2009). In this study, we applied C-vine copulas modeled with the package "VineCopula" (Schepsmeier et al. 2015) of the statistical software R (Core Team 2015). As reference data for the C-vine copulas modeling, a dataset from the Kuortane region was employed. The reference set consisted of four ALS metrics: maximum height (h_{max}), the 80th percentile of the distribution of height values (h_{80}), the canopy relief ratio (CRR), and the number of returns above 2 m divided by the total number of returns as a measure for canopy cover (p_{veg}), digital numbers of three Landsat spectral bands: green (B20), red (B30) and shortwave infra-red (B50), and GSV values per hectare from field measurements using the technique of Finnish national forest inventory (NFI) (Tomppo 2006). For details about the reference data, see Appendix A.

A copula population of 3×10^6 observations was created, based on which GSV was distributed over the study area using nearest neighbour imputation with the Landsat and ALS variables as a link, and a sample of 818,016 observations corresponding to the 818,016 grid cells of 16m × 16m size, belonging to the land-use category forest. The selected sample of 818,016 elements is our simulated population with simulated Landsat spectral values, ALS metrics and GSV values (Saarela et al. 2015b). An overview of the study population is presented in Fig. 1:

3 Methods

3.1 Statistical approach

The model-based approach is based on the concept of a superpopulation model. Any finite population of interest is seen as a sample drawn from a larger universe defined by the superpopulation model (Cassel et al. 1977). For large populations, the model has fixed parameters, whose values are unknown, and random elements with assigned attributes. The model-based survey for a finite population mean approximately corresponds to estimating the expected value of the superpopulation mean (e.g., Ståhl et al. 2016). Thus, in this study, our goal was to estimate the expected value of the superpopulation mean, $E(\mu)$, for a large finite population U with N grid cells as the population elements. Our first source of information is Landsat auxiliary data, which are available for each population element (grid cell). The second information source is a sample of M grid cells, denoted S_a. Each grid cell in S_a has two sets of RS auxiliary data available: Landsat and ALS. The third source of information is a subsample S of m grid cells, selected from S_a. For each element in S, Landsat, ALS, and GSV values are available. For simplicity, simple random sampling without replacement was assumed to be performed in both phases of sampling. The size of S was 10 % of S_a, and S_a ranged from $M = 500$ to $M = 10,000$ grid cells, i.e., S ranged from $m = 50$ to $m = 1000$. We applied ordinary least square (OLS) estimators for estimating the regression model parameters and their covariance matrices for models that relate a response variable in one phase of sampling to the auxiliary data. One such example is ALS metrics regressed against GSV in the sample S. The OLS estimator

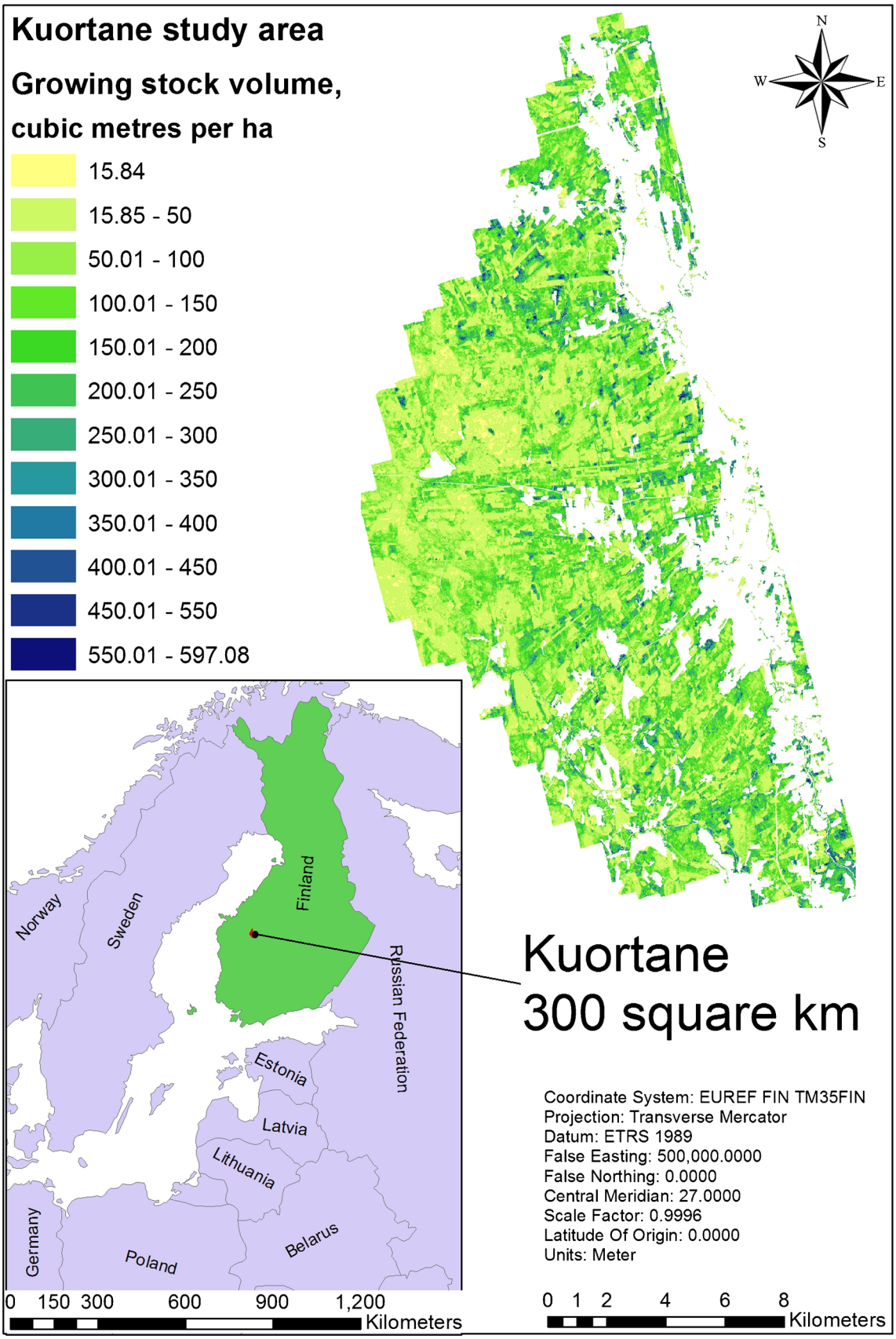

Fig. 1 The Kuortane study area. The image was shown at the SilviLaser 2015 - ISPRS Geospatial Week where the study's preliminary results were presented (Saarela et al. 2015c)

was applied under the usual assumptions, i.e., (i) *independence*, assuming that the observations are identically and independently distributed (i.i.d.); this assumption is guaranteed by simple random sampling; (ii) *exogeneity*, assuming that the (normally distributed) errors are uncorrelated with the predictor variables, and (iii) *identifiability*, assuming that there is one unique solution for the estimated model parameters, i.e., $(\boldsymbol{X}^T\boldsymbol{X})$ has full column rank.

Our study focused on the following cases:

Case A: Model-based estimation, where Landsat data are available wall-to-wall and GSV values are available for the population elements in the sample S. In the following sections, the case is also referred to as *standard model-based inference*.

Case B: Two-phase model-based estimation, where ALS data are available for S_a and GSV values for the subsample S. This case is also referred to as *hybrid inference* (Ståhl et al. 2016), since it utilizes both model-based inference and design-based inference.

Case C: Model-based estimation based on hierarchical modeling, with wall-to-wall Landsat data as the first source of information, ALS data from the sample S_a as the second information source, and GSV data from the subsample S as the third source of information. The case is referred to as *model-based inference with hierarchical modeling*.

Case C was separated into three sub-cases. The difference between the first two concerns the manner in which the three sources of data were utilized in the estimators and the corresponding variance and variance estimators. The third sub-case was introduced since it reflects how this type of nested regression models have been used in previous studies by simply ignoring the model step from GSV to GSV predictions based on ALS data, i.e., by treating the GSV predictions as if they were true values (e.g., Nelson et al. 2009; Neigh et al. 2011, 2013).

C.1: *Predicting ALS predictor variables from Landsat data – two-stage least squares regression.* – In this case information from the subsample S was used to estimate regression model parameters linking GSV values as responses with ALS variables as predictors. Information from S_a was then used to estimate a system of regression models linking ALS predictor metrics as response variables to Landsat variables as predictors. Based on Landsat data ALS predictor variables were then predicted for each population element and utilized for predicting GSV values with the first model. The reason for this rather complicated approach was that variances and variance estimators could be straightforwardly derived based on two-stage least squares regression theory (e.g., Davidson and MacKinnon 1993).

C.2: *Predicting GSV values from ALS data – hierarchical model-based estimation.* – In this case a model based on ALS data was used to predict GSV values for all elements in S_a. The predicted GSV values were then used for estimating a regression model linking the predicted GSV as a response variable with Landsat variables as predictors. This model was then applied to all population elements in order to estimate the GSV population mean.

C.3: *Ignoring the uncertainty due to predicting GSV based on ALS data—simplified hierarchical model-based estimation.* In this case, the estimation procedure was the same as in C.2, but in the variance estimation we ignored the uncertainty due to predicting GSV values from ALS data. As mentioned previously, the reason for including this case is that this procedure has been applied in several studies.

3.1.1 Case A: Standard model-based inference

This case follows well-established theory for model-based inference (e.g., Matérn et al. 1960; McRoberts 2006; Chambers & Clark 2012). For estimating the expected value of the superpopulation mean $E(\mu)$ (Ståhl et al. 2016), we utilise a regression model linking GSV values as responses with Landsat variables as predictors using information from the subsample S. We assume a linear model to be appropriate, i.e.,

$$\boldsymbol{y}_S = \boldsymbol{Z}_S\boldsymbol{\alpha} + \boldsymbol{w}_S \tag{1}$$

where $\boldsymbol{y}_S$ is a column vector of length m of GSV values, $\boldsymbol{Z}_S$ is a $m \times (q+1)$ matrix of Landsat predictors (with a first column of unit values and q is the number of Landsat predictors), $\boldsymbol{\alpha}$ is a column vector of model parameters with length $(q+1)$, and $\boldsymbol{w}_S$ is a column vector of random errors with zero expectation, of length m. Under assumptions of independence, exogeneity, and identifiability (e.g., Davidson and MacKinnon 1993), the OLS estimator of the model parameters is

$$\widehat{\boldsymbol{\alpha}}_S = (\boldsymbol{Z}_S^T\boldsymbol{Z}_S)^{-1}\boldsymbol{Z}_S^T\boldsymbol{y}_S \tag{2}$$

where $\widehat{\boldsymbol{\alpha}}_S$ is a $(q+1)$-length column vector of estimated model parameters.

The estimated model parameters $\widehat{\boldsymbol{\alpha}}_S$ are then used for estimating the expected value of the population mean, $\widehat{E(\mu)}$, Ståhl et al. (2016):

$$\widehat{E(\mu)}_A = \boldsymbol{\iota}_U^T\boldsymbol{Z}_U\widehat{\boldsymbol{\alpha}}_S \tag{3}$$

where $\boldsymbol{\iota}_U$ is a N-length column vector, where each element equals $1/N$, $\boldsymbol{Z}_U$ is a $N \times (q+1)$ matrix of Landsat auxiliary variables, i.e., for the entire population.

The variance of the estimator $\widehat{E(\mu)}_A$ is (Ståhl et al. 2016):

$$V\left[\widehat{E(\mu)}_A\right] = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U \boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_S) \boldsymbol{Z}_U^T \boldsymbol{\iota}_U \tag{4}$$

where $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_S)$ is the covariance matrix of the model parameters $\widehat{\boldsymbol{\alpha}}_S$. To obtain a variance estimator, the covariance matrix in Eq. 4 is replaced by an estimated covariance matrix.

When the errors, $\boldsymbol{w}_S$, in Eq. 1, are homoskedastic, the OLS estimator for the covariance matrix is (e.g., Davidson and MacKinnon 1993):

$$\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\alpha}}_S) = \frac{\widehat{\boldsymbol{w}}_S^T \widehat{\boldsymbol{w}}_S}{m-q-1} (\boldsymbol{Z}_S^T \boldsymbol{Z}_S)^{-1} \tag{5}$$

where $\widehat{\boldsymbol{w}}_S = \boldsymbol{y}_S - \boldsymbol{Z}_S \widehat{\boldsymbol{\alpha}}_S$ is a m-length column vector of residuals over the sample S, using Landsat auxiliary information.

When the errors, $\boldsymbol{w}_S$, in Eq. 1 are heteroskedastic, the covariance matrix can be estimated consistently (HC) with the estimator proposed by White (1980), namely

$$\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\alpha}}_S) = (\boldsymbol{Z}_S^T \boldsymbol{Z}_S)^{-1} \Big[\sum_{i=1}^{m} \hat{w}_i^2 \boldsymbol{z}_i^T \boldsymbol{z}_i \Big] (\boldsymbol{Z}_S^T \boldsymbol{Z}_S)^{-1} \tag{6}$$

where $\hat{w}_i$ is a residual and $\boldsymbol{z}_i$ is a $(q+1)$-length row vector of Landsat predictors for the i^{th} observation from the subsample S. To overcome an issue of the squared residuals $\hat{w}_i^2$ being biased estimators of the squared errors w_i^2, we applied the correction $\frac{m}{m-q-1}\hat{w}_i^2$ (Davidson and MacKinnon 1993), i.e., all the $\hat{w}_i^2$-terms in Eq. 6 were multiplied with $\frac{m}{m-q-1}$.

3.1.2 Case B: Hybrid inference

In the case of hybrid inference, expected values and variances were estimated by considering both the sampling design by which auxiliary data were collected and the model used for predicting values of population elements based on the auxiliary data (e.g., Ståhl et al. 2016). For this case, a linear model linking ALS predictor variables and the GSV response variable were fitted using information from the subsample S

$$\boldsymbol{y}_S = \boldsymbol{X}_S \boldsymbol{\beta} + \boldsymbol{e}_S \tag{7}$$

where $\boldsymbol{X}_S$ is the $m \times (p+1)$ matrix of ALS predictors over sample S, $\boldsymbol{\beta}$ is a $(p+1)$-length column vector of model parameters, and $\boldsymbol{e}_S$ is an m-length column vector of random errors with zero expectation. Under assumptions of independence, exogeneity and identifiability the OLS estimator of the model parameters is (e.g., Davidson & MacKinnon 1993):

$$\widehat{\boldsymbol{\beta}}_S = (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S \tag{8}$$

where $\widehat{\boldsymbol{\beta}}_S$ is a $(p+1)$-length column vector of estimated model parameters.

Assuming simple random sampling without replacement in the first phase, a general estimator of the expected value of the superpopulation mean $\widehat{E(\mu)}$ is (e.g., Ståhl et al. 2014):

$$\widehat{E(\mu)}_B = \boldsymbol{\iota}_{S_a}^T \boldsymbol{X}_{S_a} \widehat{\boldsymbol{\beta}}_S \tag{9}$$

where $\boldsymbol{\iota}_{S_a}$ is a M-length column vector of entities $1/M$ and $\boldsymbol{X}_{S_a}$ is a $M \times (p+1)$ matrix of ALS predictor variables.

The variance of the estimator $\widehat{E(\mu)}_B$ is presented by Ståhl et al. (2014, Eq.5, p.5.), ignoring the finite population correction factor:

$$V\left[\widehat{E(\mu)}_B\right] = \frac{1}{M}\omega^2 + \boldsymbol{\iota}_{S_a}^T \boldsymbol{X}_{S_a} \boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}_S) \boldsymbol{X}_{S_a}^T \boldsymbol{\iota}_{S_a} \tag{10}$$

where ω^2 is the sample-based population variance from the M-length column vector of $\widehat{\boldsymbol{y}}_{S_a}$-values and $\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}_S)$ is the covariance matrix of estimated model parameters $\widehat{\boldsymbol{\beta}}_S$. The $\widehat{\boldsymbol{y}}_{S_a}$ values were estimated as

$$\widehat{\boldsymbol{y}}_{S_a} = \boldsymbol{X}_{S_a} \widehat{\boldsymbol{\beta}}_S \tag{11}$$

By replacing ω^2 and $\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}_S)$ with the corresponding estimator, we obtain the variance estimator. The sample-based population variance ω^2 is estimated by $\widehat{\omega^2} = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{y}_i - \bar{\hat{y}})^2$ (cf. Gregoire 2008), and the OLS estimator for $\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}_S)$ is (e.g., Davidson & MacKinnon 1993):

$$\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\beta}}_S) = \widehat{\sigma_e^2} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \tag{12}$$

where $\widehat{\sigma_e^2} = \frac{\widehat{\boldsymbol{e}}_S^T \widehat{\boldsymbol{e}}_S}{m-p-1}$ is the estimated residual variance and $\widehat{\boldsymbol{e}}_S = \boldsymbol{y}_S - \boldsymbol{X}_S \widehat{\boldsymbol{\beta}}_S$ is an m-length column vector of residuals over sample S, using ALS auxiliary information.

In the case of heteroscedasticity, the OLS estimator (Eq. 8) can still be used for estimating the model parameters $\widehat{\boldsymbol{\beta}}_S$ but the covariance matrix is estimated by the HC estimator (White 1980)

$$\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\beta}}_S) = (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \Big[\sum_{i=1}^{m} \hat{e}_i^2 \boldsymbol{x}_i^T \boldsymbol{x}_i \Big] (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \tag{13}$$

where $\hat{e}_i$ is a residual and $\boldsymbol{x}_i$ is the $(p+1)$-length row vector of ALS predictors for the i^{th} observation from the subsample S. Like for the **Case A**, we corrected the the squared residuals $\hat{e}_i^2$ by a factor $\frac{m}{m-p-1}$ (Davidson and MacKinnon 1993).

3.1.3 Case C: model-based inference with hierarchical modelling

We begin with introducing the hierarchical model-based estimator for the expected value of the superpopulation mean, $E(\mu)$:

$$\widehat{E(\mu)}_C = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S \tag{14}$$

where in addition to the already introduced notation, $\boldsymbol{Z}_{S_a}$ is a $M \times (q+1)$ matrix of Landsat predictors for the sample S_a. In the following, it is shown that the hierarchical model-based estimators for **Case C.1** and **Case C.2** turn out to be identical under OLS regression assumptions. In the case of weighted least squares (WLS) regression, the estimators differ (see Appendix B).

C.1: ***Predicting ALS predictor variables from Landsat data – two-stage least squares regression.***

In this case, we applied a two-stage modeling approach (e.g., Davidson & MacKinnon 1993). Using the sample S_a, we developed a multivariate regression model linking ALS variables as responses and Landsat variables as predictors, i.e.

$$\boldsymbol{x}_{S_a j} = \boldsymbol{Z}_{S_a} \boldsymbol{\gamma}_j + \boldsymbol{d}_j, [j{=}1, 2, ..., (p+1)] \tag{15}$$

where $\boldsymbol{x}_{S_a j}$ is a M-length column vector of ALS variable j, $\boldsymbol{\gamma}_j$ is an $(q+1)$-length column vector of model parameters for predicted ALS variable j, and $\boldsymbol{d}_j$ is an M-length column vector of random errors with zero expectation. We assumed that "all" Landsat predictors $\boldsymbol{Z}$ are used so $\boldsymbol{Z}_{S_a}$ is the same for all variables $\boldsymbol{x}_{S_a j}$.

There are $(p+1)\times(q+1)$ parameters γ_{ij} in $\boldsymbol{\Gamma}$, an $(q+1)\times(p+1)$ matrix of model parameters, to be estimated. If we assume simultaneous normality the simultaneous least squares estimator can be used as:

$$\widehat{\boldsymbol{\gamma}}_j = (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{x}_{S_a j} \tag{16}$$

We denote $\widehat{\boldsymbol{\Gamma}}$ as a $(q+1)\times(p+1)$ matrix of estimated model parameters, where the first column of $\widehat{\boldsymbol{\Gamma}}$ is the column vector $(\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{1}_M$, which equals $\begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}^T_{1\times(q+1)}$, where $\boldsymbol{1}_M$ is an M-length column vector of unit values. Thus, we can predict ALS variables for all population elements using Landsat variables, i.e.:

$$\widehat{\boldsymbol{X}}_U = \boldsymbol{Z}_U \widehat{\boldsymbol{\Gamma}} \tag{17}$$

where $\widehat{\boldsymbol{X}}_U$ is a $N \times (p+1)$ matrix of predicted ALS variables over the entire population U.

Then, the predicted ALS variables $\widehat{\boldsymbol{X}}_U$ were coupled with the estimated model parameters $\widehat{\boldsymbol{\beta}}_S$ from Eq. 8 to estimate the expected value of the mean GSV:

$$\widehat{E(\mu)}_{C.1} = \boldsymbol{\iota}_U^T \widehat{\boldsymbol{X}}_U \widehat{\boldsymbol{\beta}}_S \tag{18}$$

To show that this equals Eq. 14, we can rewrite Eq. 18, using Eq. 8, as

$$\widehat{E(\mu)}_{C.1} = \boldsymbol{\iota}_U^T \widehat{\boldsymbol{X}}_U (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S$$

which evidently is equivalent to

$$\widehat{E(\mu)}_{C.1} = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U \widehat{\boldsymbol{\Gamma}} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S \tag{19}$$

Finally, using the estimator for $\widehat{\boldsymbol{\Gamma}}$ (Eq. 16), we can rewrite Eq. 19 as

$$\widehat{E(\mu)}_{C.1} = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S$$

which coincides with Eq. 14 proposed at the start of this section.

Since Eq. 18 can be rewritten as $\widehat{E(\mu)}_{C.1} = \sum_{i=1}^{p+1} \boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_i} \hat{\beta}_{S_i}$, the variance $V\left[\widehat{E(\mu)}_{C.1}\right]$ of the estimator in Eq. 18 can be expressed as

$$V\left[\widehat{E(\mu)}_{C.1}\right] = \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} Cov(\hat{\beta}_{S_i}[\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_i}], \hat{\beta}_{S_j}[\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_j}]) \tag{20}$$

Since $\widehat{\boldsymbol{\beta}}_S$ is based on the subsample S and $\widehat{\boldsymbol{X}}_U$ is based on the sample S_a, $\boldsymbol{e}_S$ and $\boldsymbol{d}_j$ are considered to be independent, and as a consequence we have

$$\begin{aligned} &Cov(\hat{\beta}_{S_i}[\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_i}], \hat{\beta}_{S_j}[\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_j}]) = \beta_i \beta_j Cov([\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_i}], [\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_j}]) \\ &+[\boldsymbol{\iota}_U^T \boldsymbol{x}_{U_i}][\boldsymbol{\iota}_U^T \boldsymbol{x}_{U_j}] Cov(\hat{\beta}_{S_i}, \hat{\beta}_{S_j}) \\ &+Cov(\hat{\beta}_{S_i}, \hat{\beta}_{S_j}) Cov([\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_i}], [\boldsymbol{\iota}_U^T \hat{\boldsymbol{x}}_{U_j}]) \end{aligned} \tag{21}$$

The covariances $Cov(\hat{\beta}_{S_i}, \hat{\beta}_{S_j})$ are given by the elements of the matrix $\sigma_e^2 (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1}$, where σ_e^2 is the variance of the residuals $\widehat{\boldsymbol{e}}_S$, estimated as $\widehat{\sigma_e^2} = \frac{\widehat{\boldsymbol{e}}_S^T \widehat{\boldsymbol{e}}_S}{m-p-1}$ (same as in Section 3.1.2). Thus, we estimate $Cov(\hat{\beta}_{S_i}, \hat{\beta}_{S_j})$ as

$$\widehat{Cov}(\hat{\beta}_{S_i}, \hat{\beta}_{S_j}) = \widehat{\sigma_e^2} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1}_{ij} \tag{22}$$

Further, Eq. 17 gives

$$Cov([\boldsymbol{\iota}_U^T\hat{\boldsymbol{x}}_{U_i}],[\boldsymbol{\iota}_U^T\hat{\boldsymbol{x}}_{U_j}]) = \sum_{k=1}^{q+1}\sum_{l=1}^{q+1}[\boldsymbol{\iota}_U^T\boldsymbol{z}_{U_k}][\boldsymbol{\iota}_U^T\boldsymbol{z}_{U_l}]Cov(\hat{\gamma}_{ki},\hat{\gamma}_{lj}) \tag{23}$$

The covariance of the estimated model parameters $\widehat{\boldsymbol{\Gamma}}$, assuming homoskedasticity,

$$Cov(\hat{\gamma}_{ki},\hat{\gamma}_{lj}) = Cov(\widehat{\boldsymbol{\Gamma}}) = \boldsymbol{\Lambda}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1} \tag{24}$$

where $\boldsymbol{\Lambda}$ is a $(p+1)\times(p+1)$ matrix of covariances of the $M\times(p+1)$ matrix of residuals $\boldsymbol{D}$, which are estimated as $\widehat{\boldsymbol{D}} = \boldsymbol{X}_{S_a} - \boldsymbol{Z}_{S_a}\widehat{\boldsymbol{\Gamma}}$, hence the covariance matrix $\boldsymbol{\Lambda}$ is estimated as:

$$\widehat{\boldsymbol{\Lambda}} = \frac{\widehat{\boldsymbol{D}}^T\widehat{\boldsymbol{D}}}{M-q-1} \tag{25}$$

Combining Eqs. 20–24, we can derive the least squares (LS) variance $V\left[\widehat{E(\mu)}_{C.1}\right]$:

$$\begin{aligned} V_{LS}\left[\widehat{E(\mu)}_{C.1}\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(z_i(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{z}_j^T\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} \\ &\quad +\sigma_e^2 z_i\widehat{\boldsymbol{\Gamma}}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\widehat{\boldsymbol{\Gamma}}^T\boldsymbol{z}_j^T \\ &\quad +\sigma_e^2 z_i(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{z}_j^T\sum_{k=1}^{p+1}\sum_{l=1}^{p+1}\lambda_{kl}(\boldsymbol{X}_S^T\boldsymbol{X}_S)_{kl}^{-1}\Big) \\ &= \boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} \\ &\quad +\boldsymbol{\iota}_U^T\boldsymbol{Z}_U\widehat{\boldsymbol{\Gamma}}\boldsymbol{Cov}_{OLS}(\widehat{\boldsymbol{\beta}}_S)\widehat{\boldsymbol{\Gamma}}^T\boldsymbol{Z}_U^T\boldsymbol{\iota}_U \\ &\quad +\sigma_e^2\boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\sum_{k=1}^{p+1}\sum_{l=1}^{p+1}\lambda_{kl}(\boldsymbol{X}_S^T\boldsymbol{X}_S)_{kl}^{-1} \end{aligned} \tag{26}$$

Here, λ_{kl} is the $[k,l]^{th}$ element of he matrix Λ.

To derive an estimator $\widehat{V}_{LS}\left[\widehat{E(\mu)}_{C.1}\right]$ for the variance Eq. 26, we replace $\boldsymbol{\beta}$ with estimated $\widehat{\boldsymbol{\beta}}_S$, the covariance matrix $\boldsymbol{\Lambda}$ with $\widehat{\boldsymbol{\Lambda}}$ from Eq. 25, and σ_e^2 with the estimated $\widehat{\sigma_e^2}$. Knowing that $E(\hat{\beta}_{S_i}\hat{\beta}_{S_j}) = \beta_i\beta_j + Cov(\hat{\beta}_{S_i}\hat{\beta}_{S_j})$ we have a "minus" sign between the second and third terms of Eq. 26 due to subtracting a product of the estimated covariances. Hence, our estimator for the variance $V_{LS}\left[\widehat{E(\mu)}_{C.1}\right]$ is

$$\begin{aligned} \widehat{V}_{LS}\left[\widehat{E(\mu)}_{C.1}\right] &= \boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\widehat{\boldsymbol{\beta}}_S^T\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\beta}}_S \\ &\quad +\boldsymbol{\iota}_U^T\boldsymbol{Z}_U\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\beta}}_S)\widehat{\boldsymbol{\Gamma}}^T\boldsymbol{Z}_U^T\boldsymbol{\iota}_U \\ &\quad -\widehat{\sigma_e^2}\boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\sum_{k=1}^{p+1}\sum_{l=1}^{p+1}\hat{\lambda}_{kl}(\boldsymbol{X}_S^T\boldsymbol{X}_S)_{kl}^{-1} \end{aligned} \tag{27}$$

where $\hat{\lambda}_{kl}$ is a $[k,l]^{th}$ element of the estimated covariance matrix $\widehat{\boldsymbol{\Lambda}}$ of residuals $\widehat{\boldsymbol{D}}$.

In the special case when any potential heteroskedasticiy is limited to the GSV function of ALS predictor variables over the sample S, the heteroskedasticity-consistent variance estimator is:

$$\begin{aligned} \widehat{V}_{HC}\left[\widehat{E(\mu)}_{C.1}\right] &= \boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\widehat{\boldsymbol{\beta}}_S^T\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\beta}}_S \\ &\quad +\boldsymbol{\iota}_U^T\boldsymbol{Z}_U\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\beta}}_S)\widehat{\boldsymbol{\Gamma}}^T\boldsymbol{Z}_U^T\boldsymbol{\iota}_U \\ &\quad -\boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_U^T\boldsymbol{\iota}_U\sum_{k=1}^{p+1}\sum_{l=1}^{p+1}\hat{\lambda}_{kl}\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\beta}}_S)_{kl} \end{aligned} \tag{28}$$

C.2: ***Predicting GSV values from ALS data – hierarchical model-based estimation.***

In this case, the predicted GSV variable $\widehat{\boldsymbol{y}}_{S_a}$ is used as a response variable for estimating model parameters linking GSV and Landsat-based predictors over the sample S_a, i.e., our assumed model is

$$\boldsymbol{X}_{S_a}\boldsymbol{\beta} = \boldsymbol{Z}_{S_a}\boldsymbol{\alpha} + \boldsymbol{w}_{S_a} \tag{29}$$

where $\boldsymbol{X}_{S_a}\boldsymbol{\beta}$ is an M-length column vector of expected values of predicted GSV values $\widehat{\boldsymbol{y}}_{S_a} = \boldsymbol{X}_{S_a}\widehat{\boldsymbol{\beta}}_S$ using ALS data, $\boldsymbol{\alpha}$ is a $(q+1)$-length column vector of model parameters linking estimated GSV values and Landsat predictor variables, and $\boldsymbol{w}_{S_a}$ is an M-length column vector of random errors with zero expectation.

In case the $\boldsymbol{X}_{S_a}\boldsymbol{\beta}$ values were observable, the OLS estimator of $\boldsymbol{\alpha}$ would be

$$\widetilde{\boldsymbol{\alpha}}_{S_a} = (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}\boldsymbol{\beta} \tag{30}$$

However, we use the $\boldsymbol{X}_{S_a}\widehat{\boldsymbol{\beta}}_S$ values and thus our OLS estimator of $\boldsymbol{\alpha}$ is

$$\widehat{\boldsymbol{\alpha}}_{S_a} = (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}\widehat{\boldsymbol{\beta}}_S \tag{31}$$

Thus, using the estimator $\widehat{\boldsymbol{\beta}}_S$ (Eq. 8), we obtain:

$$\widehat{\boldsymbol{\alpha}}_{S_a} = (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S^T\boldsymbol{y}_S \tag{32}$$

Then the estimated model parameters $\widehat{\boldsymbol{\alpha}}_{S_a}$ were employed for estimating the expected value of superpopulation mean $E(\mu)$:

$$\begin{aligned} \widehat{E(\mu)}_{C.2} &= \boldsymbol{\iota}_U^T\boldsymbol{Z}_U\widehat{\boldsymbol{\alpha}}_{S_a} \\ &= \boldsymbol{\iota}_U^T\boldsymbol{Z}_U(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S^T\boldsymbol{y}_S \end{aligned}$$

which coincides with Eq. 14. Thus, for models with homogeneous random errors, the estimators of the expected mean are the same for **Cases C.1** and **C.2**.

Based on the estimator $\widehat{E(\mu)}_{C.2} = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U \widehat{\boldsymbol{\alpha}}_{S_a}$, the variance is Ståhl et al. (2016)

$$V\left[\widehat{E(\mu)}_{C.2}\right] = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U \boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a}) \boldsymbol{Z}_U^T \boldsymbol{\iota}_U \tag{33}$$

where $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a})$ is the covariance matrix of $\widehat{\boldsymbol{\alpha}}_{S_a}$. By replacing the covariance $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a})$ with estimated covariance $\widehat{\boldsymbol{Cov}}(\widehat{\boldsymbol{\alpha}}_{S_a})$ in the expression Eq. 33, we obtain a variance estimator.

Under OLS assumptions $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a})$ is estimated as

$$\begin{aligned}\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= \frac{\widehat{\boldsymbol{w}}_{S_a}^T \widehat{\boldsymbol{w}}_{S_a}}{M-q-1} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \\ &+ (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \left[\boldsymbol{X}_{S_a} \widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\beta}}_S) \boldsymbol{X}_{S_a}^T\right] \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1}\end{aligned} \tag{34}$$

where, $\widehat{\boldsymbol{w}}_{S_a} = \boldsymbol{X}_{S_a} \widehat{\boldsymbol{\beta}}_S - \boldsymbol{Z}_{S_a} \widehat{\boldsymbol{\alpha}}_{S_a}$ is an M-length vector of residuals.

For the derivation of the estimator in Eq. 34, see Appendix C.

In case of heteroskedasticy of the random errors in the sample S_a and the sample S, the HC covariance matrix estimator (White 1980) of the estimated model parameters $\widehat{\boldsymbol{\alpha}}_{S_a}$ was applied (like before, the OLS estimator for $\widehat{\boldsymbol{\alpha}}_{S_a}$ was used):

$$\begin{aligned}\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \Big[\sum_{i=1}^{M} \hat{w}_i^2 \boldsymbol{z}_i^T \boldsymbol{z}_i\Big] (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \\ &+ (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \left[\boldsymbol{X}_{S_a} \widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\beta}}_S) \boldsymbol{X}_{S_a}^T\right] \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1}\end{aligned} \tag{35}$$

where $\hat{w}_i^2$ is a squared residual for the i^{th} observation in the sample S_a. As in **Cases A** and **B**, we applied the correction $\frac{M}{M-q-1}\hat{w}_i^2$ (Davidson and MacKinnon 1993).

A derivation of the estimator (Eq. 35) is given in see Appendix C.

C.3: ***Ignoring the uncertainty due to predicting GSV values based on ALS data – simplified hierarchical model-based estimation.***

This case is included since several studies have used predicted values $\widehat{\boldsymbol{y}}_{S_a}$, using ALS models, as if they were true values, and hence, the uncertainty of their estimation has been ignored. In this case, the same estimator (Eq. 14) for the expected value of mean was used, but for the variance estimator, Eqs. 33 and 34 were applied. Under OLS assumption, the matrix $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a})$ was estimated as

$$\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\alpha}}_{S_a})_{C.3} = \frac{\widehat{\boldsymbol{w}}_{S_a}^T \widehat{\boldsymbol{w}}_{S_a}}{M-q-1} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} + 0 \tag{36}$$

In the case of heteroskedasticity, it was estimated as

$$\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\alpha}}_{S_a})_{C.3} = (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \Big[\sum_{i=1}^{M} \hat{w}_i^2 \boldsymbol{z}_i^T \boldsymbol{z}_i\Big] (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} + 0 \tag{37}$$

Thus, in these estimators, we ignored the uncertainty due to the regression model based on information from the sample S.

3.2 Sampling simulation

Monte Carlo sampling simulation with $R = 3 \times 10^4$ repetitions was applied. At each repetition, new regression model parameter estimates for the pre-selected variables and their corresponding covariance matrix estimates were computed. Based on the computed model parameters, the expected value of the population mean and its variance were estimated for each case. Averages of estimated values $\widehat{E(\mu)}$ and their estimated variances $\widehat{V}\left[\widehat{E(\mu)}\right]$ were recorded. Empirical variances $V\left[\widehat{E(\mu)}\right]_{\text{emp}}$ were computed based on the outcomes from the R repetitions as

$$V\left[\widehat{E(\mu)}\right]_{\text{emp}} = \frac{1}{R-1} \sum_{k=1}^{R} \left[\widehat{E(\mu)}_k - \overline{\widehat{E(\mu)}}\right]^2 \tag{38}$$

Further, the empirical mean square error (MSE) was estimated based on the R repetitions

$$MSE\left[\widehat{E(\mu)}\right]_{\text{emp}} = \frac{1}{R} \sum_{k=1}^{R} \left[\widehat{E(\mu)}_k - E(\mu)\right]^2 \tag{39}$$

with the known expected value of the superpopulation mean $E(\mu) = 104.27$ m^{3}ha^{-1}, as the simulated finite population mean.

For all cases, we calculated the relative bias of the estimated variance

$$\widehat{RelBIAS} = 100\% \times \frac{\overline{\widehat{V}\left[\widehat{E(\mu)}\right]} - V\left[\widehat{E(\mu)}\right]_{emp}}{V\left[\widehat{E(\mu)}\right]_{emp}} \tag{40}$$

In order to make sure that the number of repetitions in the Monte Carlo simulations was sufficient, we graphed the average value of the target parameter estimates versus the number of simulation repetitions. For all our cases and

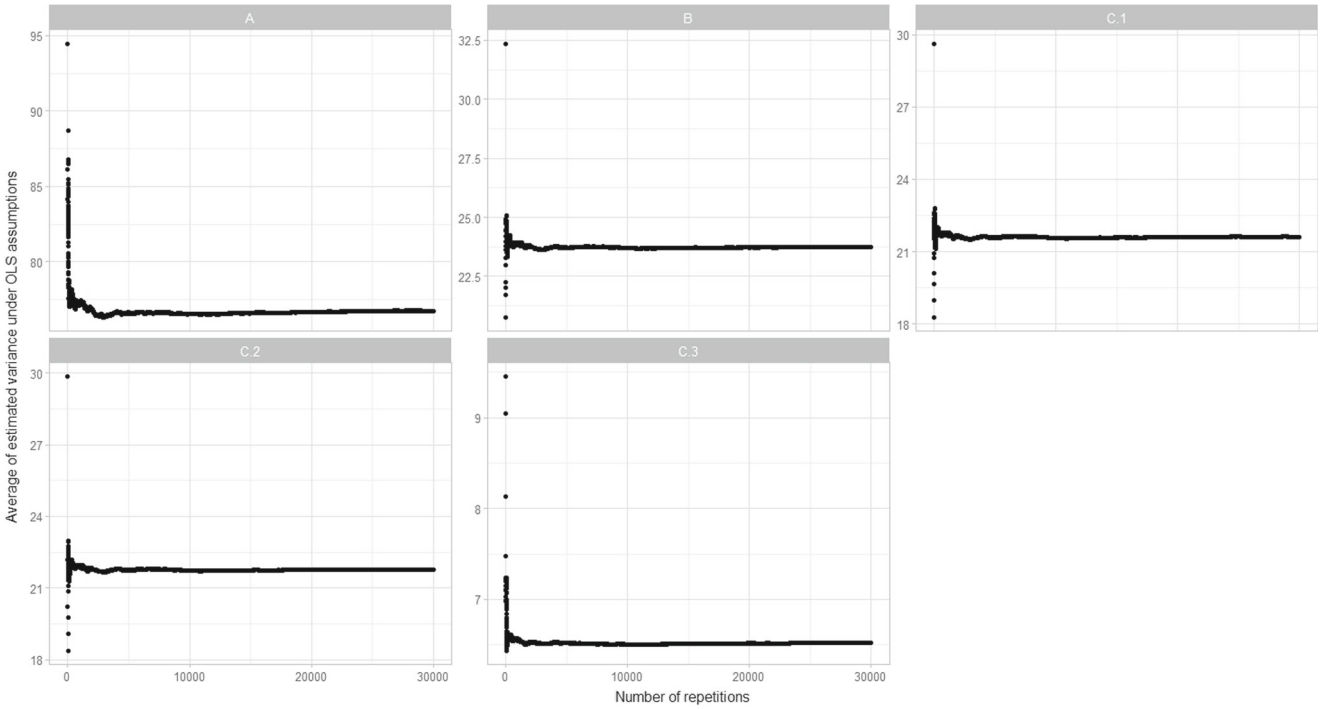

Fig. 2 The convergence of the estimated model-based variance under OLS assumptions $\widehat{\overline{V}}_{OLS}\left[\widehat{E(\mu)}\right]$ over the number of repetitions in the Monte Carlo simulations, sample sizes: m = 100 grid cells, M = 1000 grid cells.

estimators, the graphs showed that the estimators stabilized fairly rapidly and that our 3×10^4 repetitions were sufficient. An example is shown in Fig. 2, for the case of estimating $\widehat{\overline{V}}_{OLS}\left[\widehat{E(\mu)}\right]$:

4 Results

As expected, the accuracy of the model-based estimator with hierarchical modeling (**Case C**) increased as sample sizes in the two phases increased. The estimator is at least approximately unbiased, because for every group of sample sizes $MSE_{\text{emp}}\left[\widehat{E(\mu)}\right] \approx V_{\text{emp}}\left[\widehat{E(\mu)}\right]$. The Landsat variables had less predictive power than ALS metrics in prediction GSV; hence, the accuracy of the **Case B** estimator is higher than the **Case A** estimator. However, including wall-to-wall Landsat auxiliary information improved the accuracy compared to using ALS sample data alone, i.e., the MSE of **Case C** is lower than the MSE of **Case B** (Table 1).

Comparing the performances of the **Case C.3** variance estimator and the hierarchical model-based variance estimator of **Case C.2**, we observed that ignoring the uncertainty due to the GSV-ALS model leads to underestimation of the variance by about 70 % (Table 1).

In Table 2, we present examples of the goodness of fit of the models used in the **Case C.2** (and **C.3**). From Table 2, it can be observed that the goodness of fit was substantially better for the ALS models compared to the Landsat models.

5 Discussion

In this study, we have presented and evaluated novel estimators and their corresponding variance estimators for model-based inference using three sources of information and hierarchically nested models, for applications in forest inventory combining RS and field data. The estimators were evaluated through Monte Carlo simulation, for the case of estimating the population mean GSV. The estimators and the variance estimators were found to be at least approximately unbiased, unless in the **Case C.3** where the uncertainty of one of the models was ignored. The precision of the estimators depended on the number of observations used for developing the models involved; the uncertainties due to both model steps involved were found to substantially contribute to the overall uncertainty of the estimators.

Our first main methodological approach (**Case C.1**) uses wall-to-wall Landsat data to predict the ALS predictor variables involved when regressing field-measured GSV as a response variable on ALS data. In this way, we emulated wall-to-wall ALS data, which were used for estimating the population mean across the study area. The method is

Table 1 Averages of estimated expected values of the superpopulation mean $\overline{\widehat{E(\mu)}}$ and their estimated analytical variances $\overline{\widehat{V}}\left[\widehat{E(\mu)}\right]$, corresponding MSE $MSE\left[\widehat{E(\mu)}\right]$, empirical variances $V_{emp}\left[\widehat{E(\mu)}\right]$, and estimated relative bias $\widehat{RelBIAS}$: **Case A** – standard model-based inference, **Case B** – hybrid inference, **Case C** −− model-based inference with hierarchical modelling: **C.1** – two-stage lest squares regression, **C.2** – hierarchical model-based estimation, **C.3** – simplified hierarchical model-based estimation

Case		$\overline{\widehat{E(\mu)}}$, $[m^3 ha^{-1}]$	$\overline{\widehat{V}}_{OLS}\left[\widehat{E(\mu)}\right]$	$\overline{\widehat{V}}_{HC}\left[\widehat{E(\mu)}\right]$	$MSE_{emp}\left[\widehat{E(\mu)}\right]$	$V_{emp}\left[\widehat{E(\mu)}\right]$	$\widehat{RelBIAS}_{(OLS)}$, [%]	$\widehat{RelBIAS}_{(HC)}$, [%]
$m = 50\,grid\,cells,\ M = 500\,grid\,cells$								
A		102.96	159.05	156.97	157.43	155.72	2.14	0.80
B		104.74	47.74	49.05	53.12	52.90	-9.77	-7.28
C	**C.1**	104.65	43.48	44.56	48.37	48.23	-9.84	-7.60
	C.2		43.73	44.99			-9.33	-6.72
	C.3		13.46	13.46			-72.08	-72.08
$m = 100\,grid\,cells,\ M = 1000\,grid\,cells$								
A		103.68	76.78	76.27	75.67	75.33	1.93	1.25
B		104.54	23.82	24.37	25.15	25.07	-5.00	-2.80
C	**C.1**	104.48	21.69	22.16	22.69	22.64	-4.19	-2.15
	C.2		21.75	22.27			-3.92	-1.61
	C.3		6.54	6.54			-71.12	-71.12
$m = 500\,grid\,cells,\ M = 5000\,grid\,cells$								
A		104.13	15.01	14.99	14.98	14.96	0.32	0.20
B		104.30	4.75	4.78	4.80	4.80	-1.12	-0.46
C	**C.1**	104.29	4.32	4.35	4.31	4.31	0.28	0.92
	C.2		4.33	4.36			0.33	1.05
	C.3		1.27	1.27			-70.43	-70.43
$m = 1000\,grid\,cells,\ M = 10000\,grid\,cells$								
A		104.21	7.48	7.48	7.44	7.44	0.60	0.53
B		104.30	2.38	2.39	2.41	2.41	-1.36	-0.95
C	**C.1**	104.29	2.17	2.17	2.17	2.17	-0.12	0.29
	C.2		2.17	2.18			-0.09	0.36
	C.3		0.64	0.64			-70.69	-70.69

Table 2 Averages of adjusted coefficients of determination R_a^2 and estimated residual standard errors $\widehat{\sigma_e}$ and $\widehat{\sigma_w}$ for the ALS- and Landsat-based models developed in **Case C.2**

ALS			Landsat		
Number of grid cells, m	R_a^2	$\widehat{\sigma_e}$	Number of grid cells, M	R_a^2	$\widehat{\sigma_w}$
50	0.87	35.61	500	0.25	81.09
100	0.86	37.38	1000	0.25	80.40
500	0.85	38.74	5000	0.25	79.74
1000	0.85	38.96	10,000	0.25	79.65

straightforward but rather cumbersome to apply when the ALS models involve several predictor variables. Our second main methodological approach (**Case C.2**) is more intuitive, since it proceeds by first estimating a model between field GSV and ALS data; subsequently, GSV is predicted for all sample units with ALS data and these predictions are used as responses in modeling GSV based on Landsat data. Finally, wall-to-wall Landsat data are used for making predictions across the entire study area and for estimating the population mean GSV. Compared to the first method, this method is simpler to apply for ALS models with a large number of predictor variables. For models with homogeneous residual variances, fitted using OLS, the estimators obtained from the two different methods are identical, but the variances and variance estimators differ. However, the variance estimates obtained in the simulation study were similar for the two methods.

Several previous studies have combined two sources of RS data and field data in connection with hierarchical model-based estimation of forest resources. Boudreau et al. (2008), Nelson et al. (2009), Neigh et al. (2013), and Margolis et al. (2015) applied estimators of the kind denoted **C.3** in this study, i.e., they accounted for only one model step in the assessment of uncertainties. This is pointed out by Margolis et al. (2015), and Neigh et al. (2013) concluded that this would lead to a substantial underestimation of the variance. In our study, with the new set of estimators to specifically address this issue, we found that the underestimation of the variance may be as high as 70 % if the model step linking field and ALS data is ignored in the assessment of uncertainties. However, the magnitude of the underestimation depends on the properties of the models involved and the sample sizes applied for developing the models. Our findings also are important for studies (e.g., Rana et al. 2013; Ota et al. 2014) where ALS data are taken as true values in developing models where other types of RS data are used for stand or plot level predictions of forest attributes such as GSV, biomass, or canopy height.

Compared to hybrid inference using only the ALS sample and field data (**Case B**), using any of the two main methodological approaches of this study (**Cases C.1** and **C.2**) improved the precision of the estimated mean GSV. Compared to using only wall-to-wall Landsat and field data (**Case A**), the improvement in precision was very large.

An advantage of model-based inference and thus the estimators we propose is that they do not require probability samples of field or ALS data. Purposive sampling can be applied in all phases. This property makes the proposed inference technique attractive for forest surveys in remote areas, such as Siberia in the Russian Federation and Alaska in the USA, where field plots cannot easily be established in all parts of the target area due to the poor road infra-structure. However, in this study, we applied simple random sampling as a means to provide an objective description of the data collection; further, one of the methods evaluated, i.e., hybrid estimation, requires a probability sample of auxiliary data. Note that simple random sampling was applied in both phases, which to some extent limits the generality of the results since ALS samples are typically acquired as clusters of grid cells (e.g., Gobakken et al. 2012). Ongoing studies are addressing this issue in order to make the proposed type of estimators more general.

The new estimators are derived for both homoskedasticity and heteroscedasticity conditions, regarding the random errors variance. In case of heteroscedasticity, typically the OLS estimator of the covariance matrix of estimated model parameters overestimates the actual variances the model parameters (White 1980; Davidson and MacKinnon 1993). Thus, a heteroskedasticity-consistent estimator should be applied in such cases. In our simulation study, we applied a modified HC estimator; however, our results do not indicate any major difference between using different types of covariance matrix estimators. Another technical detail regards whether or not linear models can always be successfully applied for modelling GSV, as assumed in this study where OLS regression and linear models were applied. With nonlinear models or other parameter estimation techniques the proposed theory would need to be slightly modified.

Although some simplifying assumptions were made, we suggest that the proposed set of estimators (**Cases C.1 and C.2**) has a potential to substantially contribute to the development of new techniques for large-area forest surveys, utilizing several sources of auxiliary information in connection with model-based inference.

Appendix A: Reference data

A.1 Study site

To demonstrate the validity of our estimators, we chose the Kuortane area in the southern Ostrobothnia region of western Finland as study site. The main reason for this was the availability of data from earlier studies that have been conducted in the same region (e.g., Saarela et al. 2015b, 2016). The area has a size of approximately 30,000 ha of which 20,941 ha are covered by forests with *Pinus sylvestris* being the main tree species. *Picea abies* and *Betula spp.* usually occur as mixtures. The remaining parts of the landscape are formed by peat lands and open mires on higher elevations, and agricultural fields and water bodies at lower elevations and terrain depressions, respectively.

A.2 Field data

Field data were collected in 2006 using a systematic sample of circular field plots that were arranged in clusters. Each cluster consisted of 18 plots with a radius of 9 m, and the sample covered all land use types. For this study, however, only plots in forest areas were considered for further analysis. The distance between plots in a cluster was 200 m and the distance between clusters was 3500 m. In total, measurements from 441 forest field plots were available. GSV values per hectare were calculated for each field plot following the Finnish National Forest Inventory (NFI) procedure (Tomppo et al. 2008). Plots with GSV values of zero were omitted.

At all trees with a diameter at breast height (*dbh*) larger than 5 cm the following variables were observed: *dbh*, tree story class, and tree species. Tree height was measured for one sample tree per plot and species, while height for the remaining trees was predicted using models from Veltheim (1987). For calculating GSV, which is our variable of interest, individual tree models from Laasasenaho (1982) were applied. Individual tree volumes were then aggregated on the plot level and expanded to per hectare values.

A.3 ALS data

ALS of the study area was conducted in July 2006 using an Optech 3100 laser scanning system. The average flying altitude above terrain was 2000 m. The mean footprint diameter was 60 cm and the average point density was 0.64 echoes m^{-2}. Altogether, 19 north-south oriented flight lines were flown using a side overlap of about 20 %. The point cloud was normalized to terrain height using a digital terrain model generated with the Orientation and Processing of Airborne Laser Scanning data (OPALS) software (Pfeifer et al. 2014) from the same data, and divided along a grid of 16x16 m large cells. For each cell and field plot the height values of laser echoes were used to calculate several metrics related to observed values of GSV. Four metrics were calculated with the FUSION software (McGaughey 2012), and used for this study: maximum height observation (h_{max}); the 80th percentile of the distribution of height values (h_{80}); the canopy relief ratio (CRR); and the number of returns above 2 m divided by the total number of returns as a measure for canopy cover (p_{veg}). For details about the ALS data, see Saarela et al. (2015a).

A.4 Landsat data

Landsat 7 ETM+ orthorectified (L1T) multi-spectral imagery data were downloaded from U.S. Geological Survey (2014). The images were acquired in June 2006. For each field plot and grid cell, digital numbers of spectral values from the green (B20), red (B30), and shortwave infrared (B50) bands were extracted using the nearest neighbour re-sampling method in ArcGIS software (ESRI 2011).

Appendix B: Weighted least squares regression estimator for the model-based inference in Case C.1

In the case of applying weighted least squares estimator for heteroskedasticity removal in **Case C.1**, the estimator (Eq.(8)) will have the following form

$$\widehat{\boldsymbol{\beta}}_S = (\boldsymbol{X}_S^T \boldsymbol{G}_S^{-1} \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{G}_S^{-1} \boldsymbol{y}_S \tag{41}$$

where $\boldsymbol{G_S}$ is a $m \times m$ diagonal matrix with weight elements in the diagonal and zeros outside of the diagonal.

The estimator [corresponding to Eq. 16] will be

$$\widehat{\boldsymbol{\gamma}}_j = (\boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{aj}}^{-1} \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{aj}}^{-1} \boldsymbol{x}_{S_{aj}} \tag{42}$$

where $\boldsymbol{H}_{S_{aj}}$ is a $M \times M$ diagonal matrix with weight elements in the diagonal and zeros outside of the diagonal for

j^{th} ALS variable over sample S_a, and $\boldsymbol{x}_{S_{aj}}$ is an M-length column vector of j^{th} ALS variable.

Thus, estimator for the expected value of the superpopulation mean estimation in **Case C.1** will be

$$\widehat{E[\mu_{C.1}]}_{WLS} = \boldsymbol{\iota}_U^T \boldsymbol{Z}_U \begin{pmatrix} \left(\widehat{\boldsymbol{\gamma}}_{(p+1)} = (\boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a(p+1)}}^{-1} \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a(p+1)}}^{-1} \boldsymbol{x}_{S_{a(p+1)}}\right)^T \\ \left(\widehat{\boldsymbol{\gamma}}_p = (\boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_ap}^{-1} \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_ap}^{-1} \boldsymbol{x}_{S_ap}\right)^T \\ \vdots \\ \left(\widehat{\boldsymbol{\gamma}}_2 = (\boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a2}}^{-1} \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a2}}^{-1} \boldsymbol{x}_{S_{a2}}\right)^T \\ \left(\widehat{\boldsymbol{\gamma}}_1 = (\boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a1}}^{-1} \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{H}_{S_{a1}}^{-1} \boldsymbol{1}_M\right)^T \end{pmatrix}^T \times (\boldsymbol{X}_S^T \boldsymbol{G}_S^{-1} \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{G}_S^{-1} \boldsymbol{y}_S \quad (43)$$

Appendix C Proof of the hierarchical model-based variance estimators in Case C.2.

C.1 General derivation

For each element in the sample S_a, there are two models: $y_i = \boldsymbol{x}_i\boldsymbol{\beta} + e_i$ with $\boldsymbol{e} \sim N(\sigma_e^2, 0)$ and $\boldsymbol{x}_i\boldsymbol{\beta} = \boldsymbol{z}_i\boldsymbol{\alpha} + w_i$ with $\boldsymbol{w} \sim N(\sigma_w^2, 0)$, where y_i is GSV value, $\boldsymbol{x}_i$ is an $(p+1)$-length row vector of ALS predictor variables, $\boldsymbol{z}_i$ is a $(q+1)$-length row vector of Landsat predictor variables, and e_i and w_i are independent and identically distributed (i.i.d.) random errors for the i^{th} observation. Combining these two models, we can develop a composite model:

$$\begin{aligned} \boldsymbol{x}_i\boldsymbol{\beta} &= \boldsymbol{z}_i\boldsymbol{\alpha} + w_i \\ y_i - e_i &= \boldsymbol{z}_i\boldsymbol{\alpha} + w_i \\ y_i &= \boldsymbol{z}_i\boldsymbol{\alpha} + w_i + e_i \end{aligned} \quad (44)$$

That is, in vector notation the regression model applied in **Case C.2** is

$$\boldsymbol{y}_{S_a} = \boldsymbol{Z}_{S_a}\boldsymbol{\alpha} + \boldsymbol{w}_{S_a} + \boldsymbol{e}_{S_a} \quad (45)$$

For deriving an estimator for the covariance matrix of the estimated model parameters $\widehat{\boldsymbol{\alpha}}_{S_a}$, we modify Eq. 32 as:

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{S_a} &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{y}_S \\ &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T (\boldsymbol{X}_S\boldsymbol{\beta} + \boldsymbol{e}_S) \\ &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{X}_S\boldsymbol{\beta} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \end{aligned}$$

Knowing that $(\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{X}_S = \boldsymbol{I}_{(m\times m)}$ in the first term of the expression, we obtain:

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{S_a} &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a}\boldsymbol{\beta} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \end{aligned}$$

Recalling from Eq. 44 that $\boldsymbol{X}_{S_a}\boldsymbol{\beta} = \boldsymbol{Z}_{S_a}\boldsymbol{\alpha} + \boldsymbol{w}_{S_a}$, we modify further to obtain:

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{S_a} &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T (\boldsymbol{Z}_{S_a}\boldsymbol{\alpha} + \boldsymbol{w}_{S_a}) \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \\ &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a}\boldsymbol{\alpha} + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \end{aligned}$$

Knowing that $(\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a} = \boldsymbol{I}_{(M\times M)}$, we obtain:

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{S_a} &= \boldsymbol{\alpha} + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \end{aligned}$$

Moving $\boldsymbol{\alpha}$ to the left side of the expression, we get

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{S_a} - \boldsymbol{\alpha} &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \end{aligned} \quad (46)$$

Now, we derive the estimator for the covariance of $\widehat{\boldsymbol{\alpha}}_{S_a}$:

$$\begin{aligned} \boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= E\left[(\widehat{\boldsymbol{\alpha}}_{S_a} - \boldsymbol{\alpha})(\widehat{\boldsymbol{\alpha}}_{S_a} - \boldsymbol{\alpha})^T\right] = E\Big[\big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S\big) \\ &\quad \times \big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S\big)^T\Big] \\ &= E\Big[(\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} \big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a}\big)^T \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a} \big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S\big)^T \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{w}_{S_a}\big)^T \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S \big((\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T \boldsymbol{e}_S\big)^T\Big] \\ &= (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T E\left[\boldsymbol{w}_{S_a} \boldsymbol{w}_{S_a}^T\right] \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T E\left[\boldsymbol{w}_{S_a} \boldsymbol{e}_S^T\right] \boldsymbol{X}_S (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_{S_a}^T \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T E\left[\boldsymbol{e}_S \boldsymbol{w}_{S_a}^T\right] \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \\ &\quad + (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \boldsymbol{Z}_{S_a}^T \boldsymbol{X}_{S_a} (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_S^T E\left[\boldsymbol{e}_S \boldsymbol{e}_S^T\right] \boldsymbol{X}_S (\boldsymbol{X}_S^T \boldsymbol{X}_S)^{-1} \boldsymbol{X}_{S_a}^T \boldsymbol{Z}_{S_a} (\boldsymbol{Z}_{S_a}^T \boldsymbol{Z}_{S_a})^{-1} \end{aligned}$$

Assuming that $\boldsymbol{w}_{S_a}$ and $\boldsymbol{e}_S$ are independent and uncorrelated, and knowing that $E[\boldsymbol{w}_{S_a}] = 0$ and $E[\boldsymbol{e}_{S_a}] = 0$, we have $E\left[\boldsymbol{w}_{S_a}\boldsymbol{e}_S^T\right] = E\left[\boldsymbol{e}_S\boldsymbol{w}_{S_a}^T\right] = E\left[\boldsymbol{w}_{S_a}\right]E\left[\boldsymbol{e}_S\right] = 0$. Thus,

$$\begin{aligned}\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T E\left[\boldsymbol{w}_{S_a}\boldsymbol{w}_{S_a}^T\right]\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &\quad+(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S^T E\left[\boldsymbol{e}_S\boldsymbol{e}_S^T\right]\boldsymbol{X}_S(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_{S_a}^T\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &= (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{\Sigma}\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &\quad+(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S^T\boldsymbol{\Omega}\boldsymbol{X}_S(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_{S_a}^T\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\end{aligned} \tag{47}$$

where $\boldsymbol{\Sigma}$ is a covariance matrix of errors $\boldsymbol{w}_{S_a}$ and $\boldsymbol{\Omega}$ is a covariance matrix of errors $\boldsymbol{e}_S$.

C.2 Under homogeneous random errors

Under the general OLS assumptions $\boldsymbol{\Sigma} = \sigma_w^2\boldsymbol{I}_{(M\times M)}$, where σ_w^2 is estimated as $\widehat{\sigma_w^2} = \frac{\widehat{\boldsymbol{w}}_{S_a}\widehat{\boldsymbol{w}}_{S_a}^T}{M-q-1}$, where $\widehat{\boldsymbol{w}}_{S_a} = \boldsymbol{X}_{S_a}\widehat{\boldsymbol{\beta}}_S - \boldsymbol{Z}_{S_a}\widehat{\boldsymbol{\alpha}}_{S_a}$ is an M-length column vector of residuals over sample S_a. Thus, first part of Eq. 47, i.e., $(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{\Sigma}\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}$, can be estimated as

$$\begin{aligned}&(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\left[\widehat{\sigma_w}^2\boldsymbol{I}_{(M\times M)}\right]\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &\quad= \frac{\widehat{\boldsymbol{w}}_{S_a}\widehat{\boldsymbol{w}}_{S_a}^T}{M-q-1}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\end{aligned}$$

In the second term of Eq. 47, $\boldsymbol{\Omega}$ is estimated as $\widehat{\sigma_e^2}\boldsymbol{I}_{(m\times m)} = \frac{\widehat{\boldsymbol{e}}_S\widehat{\boldsymbol{e}}_S^T}{m-q-1}\boldsymbol{I}_{(m\times m)}$, where $\widehat{\boldsymbol{e}}_S = \boldsymbol{y}_S - \boldsymbol{X}_S\widehat{\boldsymbol{\beta}}_S$ is an m-length column vector of residuals over the sample S. Thus, the second term can be estimated as

$$\begin{aligned}&(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_S^T\left[\frac{\widehat{\boldsymbol{e}}_S\widehat{\boldsymbol{e}}_S^T}{m-q-1}\boldsymbol{I}_{(m\times m)}\right]\boldsymbol{X}_S(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_{S_a}^T\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &= (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}\left[\frac{\widehat{\boldsymbol{e}}_S\widehat{\boldsymbol{e}}_S^T}{m-q-1}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\right]\boldsymbol{X}_{S_a}^T\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\end{aligned}$$

We can see that the expression $\left[\frac{\widehat{\boldsymbol{e}}_S\widehat{\boldsymbol{e}}_S^T}{m-q-1}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\right]$ is in fact the estimator of the covariance matrix of the estimated model parameters $\widehat{\boldsymbol{\beta}}_S$ (Eq. 12). Therefore, we can write the estimator of $\boldsymbol{Cov}(\widehat{\boldsymbol{\alpha}}_{S_a})$ as

$$\begin{aligned}\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= \frac{\widehat{\boldsymbol{w}}_{S_a}\widehat{\boldsymbol{w}}_{S_a}^T}{M-q-1}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &\quad+(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\left[\boldsymbol{X}_{S_a}\widehat{\boldsymbol{Cov}}_{OLS}(\widehat{\boldsymbol{\beta}}_S)\boldsymbol{X}_{S_a}^T\right]\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\end{aligned} \tag{48}$$

C.3 Under heteroskedasticity

In the case of heteroskedasticity, we followed the theoretical framework developed by White (1980). Thus, the expression $\boldsymbol{Z}_{S_a}^T\boldsymbol{\Sigma}\boldsymbol{Z}_{S_a}$ in the first term of Eq. 47 can be estimated as $\sum_{i=1}^{M}\hat{w}_i^2\boldsymbol{z}_i^T\boldsymbol{z}_i$; correspondingly, in the second term the expression $\boldsymbol{X}_S^T\boldsymbol{\Omega}\boldsymbol{X}_S$ can be estimated as $\sum_{i=1}^{m}\hat{e}_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i$. Further, the second term of Eq. 47 can be estimated as

$$(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\boldsymbol{X}_{S_a}(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\left[\sum_{i=1}^{m}\hat{e}_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i\right](\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\boldsymbol{X}_{S_a}^T\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}$$

We can see that $(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}\left[\sum_{i=1}^{m}\hat{e}_i^2\boldsymbol{x}_i^T\boldsymbol{x}_i\right](\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}$ is in fact the heteroskedasticity-consistent estimator of the covariance matrix of the estimated model parameters $\widehat{\boldsymbol{\beta}}_S$ (Eq. 13). Therefore, the heteroskedasticity-consistent covariance matrix estimator for the estimated model parameters $\widehat{\boldsymbol{\alpha}}_{S_a}$ is

$$\begin{aligned}\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\alpha}}_{S_a}) &= (\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\left[\sum_{i=1}^{M}\hat{w}_i^2\boldsymbol{z}_i^T\boldsymbol{z}_i\right](\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\\ &\quad+(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\boldsymbol{Z}_{S_a}^T\left[\boldsymbol{X}_{S_a}\widehat{\boldsymbol{Cov}}_{HC}(\widehat{\boldsymbol{\beta}}_S)\boldsymbol{X}_{S_a}^T\right]\boldsymbol{Z}_{S_a}(\boldsymbol{Z}_{S_a}^T\boldsymbol{Z}_{S_a})^{-1}\end{aligned} \tag{49}$$

References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence. Insurance: Math Econ 44:182–198

Boudreau J, Nelson RF, Margolis HA, Beaudoin A, Guindon L, Kimes DS (2008) Regional aboveground forest biomass using airborne and spaceborne LiDAR in Qébec. Rem Sens of Envir 112:3876–3890

Cassel C-M, Särndal CE, Wretman JH (1977) Foundations of inference in survey sampling. (Book) Wiley

Chambers R, Clark R (2012) An introduction to model-based survey sampling with applications. (Book) OUP

Core Team R (2015) R: A language and environment for statistical computing r foundation for statistical computing, Vienna, Austria
Corona P, Fattorini L, Franceschi S, Scrinzi G, Torresan C (2014) Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. Can J of For Res 44:1303–1311
Davidson R, MacKinnon JG (1993) Estimation and inference in econometrics. OUP
Ene LT, Næsset E, Gobakken T, Gregoire TG, Ståhl G, Nelson R (2012) Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach. Rem Sens of Env 123:579–592
ESRI (2011) ArcGIS Desktop: Release 10 Redlands, SA: Environmental Systems Research Institute
Gobakken T, Næsset E, Nelson RF, Bollandsås OM, Gregoire TG, Ståhl G, Holm S, Ørka HO, Astrup R (2012) Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Rem Sens of Env 123:443–456
Grafström A, Saarela S, Ene LT (2014) Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can J of For Res 44:1156–1164
Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. Can J of For Res 28:1429–1447
Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. (Book) CRC Press
Gregoire TG, Ståhl G, Næsset E, Gobakken T, Nelson R, Holm S (2011) Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. Can J of For Res 41:83–95
Laasasenaho J (1982) Taper curve and volume functions for pine, spruce and birch [Pinus sylvestris Picea abies, Betula pendula, Betula pubescens]. Communicationes Instituti Forestalis Fenniae (Finland)
Margolis HA, Nelson RF, Montesano PM, Beaudoin A, Sun G, Andersen H-E, Wulder M (2015) Combining satellite LiDAR, airborne lidar and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of north America. Can J of For Res 45:838–855
Massey A, Mandallaz D, Lanz A (2014) Integrating remote sensing and past inventory data under the new annual design of the Swiss National Forest Inventory using three-phase design-based regression estimation. Can J of For Res 44:1177–1186
Matérn B (1960) Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Survey and Other Sampling Investigations. (Book) Esselte
McGaughey RJ (2012) FUSION/LDV: Software for LIDAR data analysis and visualization. Version 3.10. USDA Forest Service. Pacific Northwest Research Station. Seattle, WA. http://www.fs.fed.us/eng/rsac/fusion/. Accessed: 24 August 2012
McRoberts RE, Nelson MD, Wendt DG (2002) Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. Rem Sens of Env 82:457–468
McRoberts RE (2006) A model-based approach to estimating forest area. Rem Sens of Env 103:56—66
McRoberts RE (2010) Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. Rem Sens of Env 114:1017—1025
Neigh CS, Nelson RF, Sun G, Ranson J, Montesano PM, Margolis HA (2011) Moving Toward a Biomass Map of Boreal Eurasia based on ICESat GLAS, ASTER GDEM, and field measurements: Amount, Spatial distribution, and Statistical Uncertainties. In: AGU Fall Meeting Abstracts 2011 Dec (Vol. 1, p. 07)
Neigh CS, Nelson RF, Ranson KJ, Margolis HA, Montesano PM, Sun G, Kharuk V, Næsset E, Wulder MA, Andersen H-E (2013) Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne lidar. Rem Sens of Env 137:274—287
Nelsen RB (2006) An introduction to copulas. (Book) Springer
Nelson RF, Boudreau J, Gregoire TG, Margolis H, Næsset E, Gobakken T, Ståhl G (2009) Estimating Quebec provincial forest resources using ICESat/GLAS. Can J of For Res 39:862—881
Ota T, Ahmed OS, Franklin SE, Wulder MA, Kajisa T, Mizoue N, Yoshida S, Takao G, Hirata Y, Furuya N, Sano T (2014) Estimation of airborne lidar-derived tropical forest canopy height using landsat time series in Cambodia. Rem Sens 6:10750–10772
Pfeifer N, Mandlburge G, Otepka J, Karel W (2014) OPALS—a framework for airborne laser scanning data analysis. Computers, Env and Urban Syst 45:125–136
Rana P, Tokola T, Korhone L, Xu Q, Kumpula T, Vihervaara P, Mononen L (2013) Training area concept in a two-phase biomass inventory using airborne laser scanning and RapidEye satellite data. Rem Sens 6:285–309
Saarela S, Grafström A, Ståhl G, Kangas A, Holopainen M, Tuominen S, Nordkvist K, Hyyppä J (2015a) Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. Rem Sens of Env 158:431–440
Saarela S, Schnell S, Grafström A, Tuominen S, Nordkvist K, Hyyppä J, Kangas A, Ståhl G (2015b) Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. Can J of For Res 45:1524–1534
Saarela S, Grafström A, Ståhl G (2015c). Three-phase model-based estimation of growing stock volume utilizing Landsat, LiDAR and field data in large-scale surveys. Full Proceedings, SilviLaser 2015 - ISPRS Geospatial Week: invited session Estimation, inference, and uncertainty, Sept. 27-30, 2015, La Grande-Motte, France.
Saarela S, Schnell S, Tuominen S, Balazs A, Hyyppä J, Grafström A, Ståhl G (2016) Effects of positional errors in model-assisted and model-based estimation of growing stock volume. Rem Sens of Env 172:101–108
Särndal CE, Swensson B, Wretman J (1992) Model Assisted Survey Sampling. (Book) Springer
Schepsmeier U, Stoeber J, Brechmann EC, Graeler B, Nagler MT, Suggests TSP (2015) Package xVineCopula
Ståhl G, Holm S, Gregoire TG, Gobakken T, Næsset E, Nelson R (2011) Model-based inference for biomass estimation in a liDAR sample survey in Hedmark County, Norway. Can J of For Res 41:96–107
Ståhl G, Heikkinen J, Petersson H, Repola J, Holm S (2014) For Sc 60:3–13
Ståhl G, Saarela S, Schnell S, Holm S, Breidenbach J, Healey SP, Patterson PL, Magnussen S, Næsset E, McRoberts RE, Gregoire TG (2016) Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. For Ecosyst 3:1—11
Tomppo E (2006) The Finnish national forest inventory. In: Forest Inventory (Book) 179–464, 194. Springer
Tomppo E, Haakana M, Katila M, Peräsaari J (2008) Multi-source national forest inventory—methods and applications. (Book) Managing Forest Ecosystems 18
U.S. Geological Survey (2014). Landsat Missions. http://landsat.usgs.gov/index.php/. Accessed: 28 March 2011
Veltheim T (1987) Pituusmallit männylle, kuuselle ja koivulle. [Height models for pine, spruce and birch]. Master's thesis Department of Forest Resources Management. University of Helsinki, Finland
Wulder MA, White J, Nelson RF, Næsset E, Ørka HO, Coops NC, Hilker T, Bater CW, Gobakken T (2012) Lidar sampling for large-area forest characterization: a review. Rem Sens of Env 121:196–209
White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: J of the Econometric Society 817–838