remote sensing

MDPI

Article

Generalized Hierarchical Model-Based Estimation for Aboveground Biomass Assessment Using GEDI and Landsat Data

Svetlana Saarela[1,*], Sören Holm[1], Sean P. Healey[2], Hans-Erik Andersen[2], Hans Petersson[1], Wilmer Prentius[1], Paul L. Patterson[2], Erik Næsset[3], Timothy G. Gregoire[4] and Göran Ståhl[1]

1 Faculty of Forest Sciences, Swedish University of Agricultural Sciences, SLU Skogsmarksgränd 17, SE-90183 Umeå, Sweden; svetlana.saarela@slu.se; soren.holm@gronstenen.se; hans.petersson@slu.se; wilmer.prentius@slu.se; goran.stahl@slu.se;

2 Inventory and Monitoring, United States Department of Agriculture (USDA) Forest Service, 1400 Independence Ave, SW, Washington, DC 20250-1111, USA; seanhealey@fs.fed.us; handersen@fs.fed.us; plpatterson@fs.fed.us;

3 Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway; erik.naesset@nmbu.no;

4 School of Forestry and Environmental Studies, Yale University, 195 Prospect Street, New Haven, CT 06511, USA; timothy.gregoire@yale.edu.

* Correspondence: svetlana.saarela@slu.se

Received: 11 October 2018; Accepted: 14 November 2018; Published: 19 November 2018

Abstract: Recent developments in remote sensing (RS) technology have made several sources of auxiliary data available to support forest inventories. Thus, a pertinent question is how different sources of RS data should be combined with field data to make inventories cost-efficient. Hierarchical model-based estimation has been proposed as a promising way of combining: (i) wall-to-wall optical data that are only weakly correlated with forest structure; (ii) a discontinuous sample of active RS data that are more strongly correlated with structure; and (iii) a sparse sample of field data. Model predictions based on the strongly correlated RS data source are used for estimating a model linking the target quantity with weakly correlated wall-to-wall RS data. Basing the inference on the latter model, uncertainties due to both modeling steps must be accounted for to obtain reliable variance estimates of estimated population parameters, such as totals or means. Here, we generalize previously existing estimators for hierarchical model-based estimation to cases with non-homogeneous error variance and cases with correlated errors, for example due to clustered sample data. This is an important generalization to take into account data from practical surveys. We apply the new estimation framework to case studies that mimic the data that will be available from the Global Ecosystem Dynamics Investigation (GEDI) mission and compare the proposed estimation framework with alternative methods. Aboveground biomass was the variable of interest, Landsat data were available wall-to-wall, and sample RS data were obtained from an airborne LiDAR campaign that produced simulated GEDI waveforms. The results show that generalized hierarchical model-based estimation has potential to yield more precise estimates than approaches utilizing only one source of RS data, such as conventional model-based and hybrid inferential approaches.

Keywords: Carbon monitoring; GEDI; Landsat 7 ETM+; Model-based inference; Superpopulation models; Variance estimation.

1. Introduction

For several reasons society's interest in forests and forestry is increasing. Forests play a key role in ambitions to mitigate climate change through moving from fossil-based economies to bio-based

economies. They provide a long list of ecosystem services, including human recreational and economic uses. As a result, there is an increasing worldwide need for information about the state and change of forest resources and forest environmental conditions. One important example is the United Nations' Framework Convention on Climate Change (UNFCCC), which has become an important driver of forest inventory development since it highlights that forests are potentially huge sources or sinks of carbon dioxide and, thus, the agreements under UNFCCC require recurrent reporting of forest carbon pool changes [1]. Another example is the Forest Europe process, which relies on forest information for promoting sustainable management of European forests [2].

A major challenge is to develop methods to collect forest information cost-efficiently, without sacrificing information quality. Typically, National Forest Inventories (NFIs) based on field sample plots provide forest information at regional and national level. However, the ongoing development of earth observation technologies is rapidly improving, and the scope, quality and availability of the data provided offer substantial new possibilities for forest inventories [e.g., 3]. Over the last few decades several studies have been conducted where remote sensing (RS) and field data have been combined in order to enhance the precision of large-scale field based inventories, or to make forest surveys feasible in remote areas where field sampling is very costly.

When conducting a forest inventory, RS data can be incorporated at either the design stage or the estimation stage. At the design stage, RS data can be used for stratification [e.g., 4], unequal probability sampling [e.g., 5], or balanced sampling [e.g., 6]. To utilize RS data in the estimation stage, either model-based inference [7–9], or model-assisted estimation including post-stratification [10], i.e. design-based inference, can be applied.

The vast availability of RS data also opens new possibilities for improving estimation efficiency by using combinations of several sources of RS data. While this can be achieved straightforwardly in the case of model-assisted estimation following established sampling theory [e.g., 5,11–13], this issue has been less explored for model-based inference for the case when auxiliary data are not available for the entire population. An important case is when wall-to-wall RS data (or a large sample) are complemented by a (sparse) sample of RS data that are strongly correlated with the forest attribute variable of interest. While several studies have utilized this type of combination of RS data for model-based inference [e.g., 14–16], first steps towards a rigid statistical framework for this type of surveys were taken by Saarela *et al.* [17] and Holm *et al.* [18]. The current study is a generalization and expansion of those studies. The proposed technique has been termed hierarchical model-based estimation (ibid.) since in a typical case the data sources are nested. Commonly, the ambition is to utilize wall-to-wall (or a large sample of) RS data for the inference about population characteristics within a large study area. Since field data are expensive or may not be possible to obtain from all parts of the area, a sample of RS data is selected, and a model linking the study variable (from field plots) with sampled RS data is established. A key issue in hierarchical model-based estimation is that this model must provide precise predictions of the variable of interest. Thus, the sampled RS data should provide more accurate model predictions than the RS data available wall-to-wall. For example, in a biomass survey the wall-to-wall RS data might be obtained from the Landsat satellite and the sampled RS data from airborne laser scanning [cf. 19]. Model predictions across the sampled set of RS data are used for estimating a second model, linking the variable of interest with wall-to-wall RS data. The latter model is used for the model-based inference, but in the uncertainty assessment both modeling steps must be accounted for [17,18,20].

Hierarchical model-based estimation can be used also in cases where data are not nested, as an approach to make use of a sparse network of existing field data. In this case a small (possibly purposive) sample of RS data is selected to link field data with a large sample of RS data, or RS data available wall-to-wall. This approach was pioneered by Boudreau *et al.* [14] and Nelson *et al.* [15], who used a combination of the Portable Airborne Laser System (PALS) and data from ICESat/GLAS for estimating aboveground biomass (AGB) in Québec, Canada. A similar approach was applied in a study by Neigh *et al.* [16] for assessment of forest carbon stock across the entire boreal forest

region. However, these studies ignore parts of the models' contribution to the overall uncertainty of the biomass estimators. In Saarela *et al.* [17] it was shown that the studies might have underestimated the variance by up to 70%. However, the study by Saarela *et al.* [17] was conducted under simplifying assumptions, such as assuming all samples being conducted through simple random sampling and assuming that all models involved had homogeneous residual variance.

The main objective of this study is to generalize the existing estimators for hierarchical model-based estimation so that this estimation method can be applied also in cases when the model errors have non-homogeneous variance and cases where the errors are correlated, for example due to clustered sampled data. A further objective is to compare the proposed estimation method with existing methods, using data from study sites in the USA that mimic the type of data that will be available from the Global Ecosystem Dynamics Investigation (GEDI) mission [21]. The proposed generalization of existing theory in Saarela *et al.* [17] is important, since both field and intermediate RS data in most practical cases are clustered [e.g., 22] and many models between RS data and biophysical features, such as biomass, have non-homogeneous variance [e.g., 18,23].

2. Methods

2.1. Overview

In this section we first derive and present the theory for generalised hierarchical model-based (GHMB) estimation. In the next section, we then present the data from six study sites in the USA, which were used for assessing the performance of the new estimators. A key issue was to assess what precision can be expected if this estimation framework is used with a combination of field, GEDI and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data. Because GEDI data are not available yet, we used GEDI-like waveforms simulated from airborne small-footprint LiDAR (see Section 3.1.1 for GEDI data description).

We compared the performance of GHMB estimation with: (i) two-stage model-based estimation described in Holm *et al.* [18], based on GEDI, Landsat 7 ETM+ and field data; (ii) hybrid estimation [24], which utilizes only the sampled GEDI data and field data; and (iii) conventional model-based estimation [e.g., 25], which utilizes only wall-to-wall Landsat 7 ETM+ data and field data.

2.2. Generalized Hierarchical Model-Based estimation (GHMB)

In model-based inference, the vector of N population values is seen as a realization from a vector random variable $\mathbf{y} = (y_1, ..., y_N)$ with a given joint probability distribution [26]. The expected value of y_i is μ_i and $\sum_{i=1}^{N} \frac{\mu_i}{N} = \mu$. If the size N of the target population, U, is large, we can assume that $\mu \approx \bar{y}$ (where $\bar{y}$ is the population mean of the given realization) is a good approximation, and instead of predicting the target population mean $\bar{y}$, we estimate its expectation μ [e.g., 27].

The joint distribution G of the vector random variable $\mathbf{y} = (y_1, ..., y_N)$ is denoted a *superpopulation model* [cf. 26]. The GHMB estimation approach relies on two superpopulation models of the class of multiple regression models. The first superpopulation model links target population element values $\mathbf{y}$ with predictor variables $\mathbf{X}$ (including a column of units) assumed to be strongly correlated with $\mathbf{y}$. The second superpopulation model links $\mathbf{y}$ with predictor variables $\mathbf{Z}$ (including a column of units) assumed to be weakly correlated with $\mathbf{y}$. In our example the variable of interest, $\mathbf{y}$, is AGB, and $\mathbf{X}$ and $\mathbf{Z}$ are GEDI and Landsat 7 ETM+ data, respectively (see Section 3.1 for the data description). On this basis we denote the first superpopulation model G_X and the second G_Z. Since, G_X and G_Z describe the same joint distribution of $\mathbf{y} = (y_1, ..., y_N)$, they relate as $\mu|G_X = \mu|G_Z$, which is an essential basis for GHMB estimation.

In our study, G_X and G_Z are assumed to have a similar form and follow similar distributional properties, and thus, for our target population U we have:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0},\ \mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathsf{T}}] = \omega^2\boldsymbol{\Omega}, \tag{1}$$

and

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{v},\ \mathrm{E}[\boldsymbol{v}] = \mathbf{0},\ \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}}] = \theta^2\boldsymbol{\Delta}; \tag{2}$$

here, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are vectors of model parameters to be estimated, $\boldsymbol{\epsilon}$ and $\boldsymbol{v}$ are random error vectors, and $\omega^2\boldsymbol{\Omega}$ and $\theta^2\boldsymbol{\Delta}$ are the variance-covariance matrices with diagonal elements corresponding to individual error variances and off-diagonal elements to covariances between individual errors.

Three dataset are required for GHMB estimations.

- The first dataset, denoted S, contains a sample of field data for which sampled RS data are also available. The dataset is used for estimating the model parameters $\boldsymbol{\beta}$. Each estimator based on this set is given the subscript S; the data S comprise n field observations.
- The second dataset, denoted Sa, contains the enlarged sample of RS data and the corresponding RS data from the wall-to-wall dataset. It is used for estimating the model parameters $\boldsymbol{\alpha}$, and any estimator based on this set has the subscript Sa; the data Sa comprise M sampled RS observations.
- The third dataset contains the wall-to-wall RS data for the entire target population, U. The target population U comprises N population elements.

With a nested structure of data, Sa is a sample of U, and S is a sample of Sa. However, as we stated in the introduction, the datasets do not have to be nested, and the sample S may be selected independently from Sa. Figure 1 provides an overview of GHMB estimation.

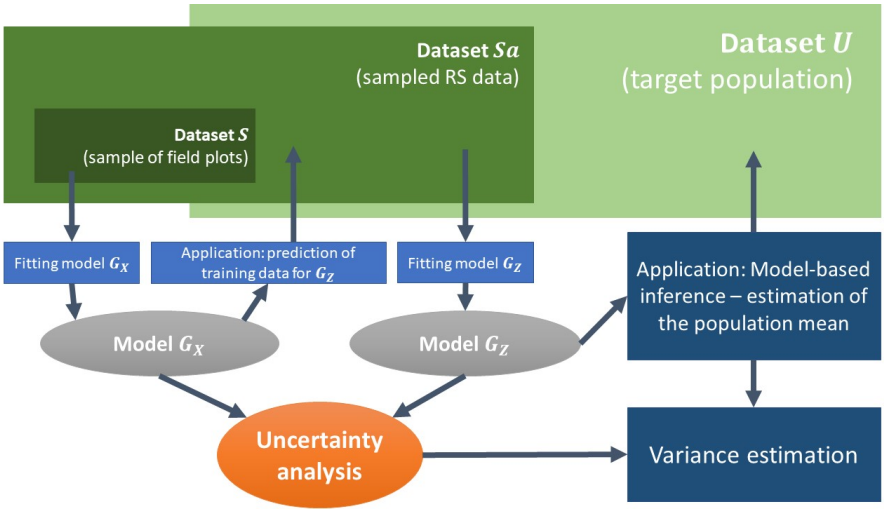


Figure 1. Overview of Generalized Hierarchical Model-Based estimation.

We estimated $\boldsymbol{\beta}$ by generalized least squares (GLS) estimators usinf dataset S [e.g., 28]:

$$\widehat{\boldsymbol{\beta}}_S = \left(\mathbf{X}_S^{\mathsf{T}}\widehat{\boldsymbol{\Omega}}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\mathsf{T}}\widehat{\boldsymbol{\Omega}}_S^{-1}\mathbf{y}_S, \tag{3}$$

Conditions $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{v}] = \mathbf{0}$ in G_X and G_Z, respectively, imply that for the superpopulation model G_X, $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$ and for the superpopulation model G_Z, $\mathrm{E}[\mathbf{y}|\mathbf{Z}] = \mathbf{Z}\boldsymbol{\alpha}$. This does not imply that $\mathrm{E}[\mathbf{y}|\mathbf{X}]$ equals $\mathrm{E}[\mathbf{y}|\mathbf{Z}]$ for a given realization of $\mathbf{X}$ and $\mathbf{Z}$. For example, assume tree biomass is modeled based on either tree diameter ($\mathbf{X}$) or tree height ($\mathbf{Z}$) these two models will be very different, and in formal terms this is because the models provide the expected values conditional on the different predictor variables. This difference between the two models may be more subtle when the models are based on different types of RS data, but the principle remains the same: the two models will be different. Thus, for a given realization of $\mathbf{X}$ and $\mathbf{Z}$, $\mathbf{y}|\mathbf{X} = \mathbf{y}|\mathbf{Z}$, but $\mathrm{E}[\mathbf{y}|\mathbf{X}] \neq \mathrm{E}[\mathbf{y}|\mathbf{Z}]$. Using the models (1) and (2) we obtain $\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{v}$, and thus,

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u},\ \mathrm{E}[\mathbf{u}] = \mathbf{0},\ \mathrm{E}[\mathbf{u}\mathbf{u}^{\intercal}] = \sigma^2\boldsymbol{\Sigma}, \tag{4}$$

where $\mathbf{u} = \boldsymbol{v} - \boldsymbol{\epsilon}$ is a random variable. It can be seen that $\bar{\mathbf{x}}\boldsymbol{\beta} = \bar{\mathbf{z}}\boldsymbol{\alpha}$ ($\bar{\mathbf{x}}$ and $\bar{\mathbf{z}}$ are vectors of average values) and thus, the relationship agrees with $\mu|G_X = \mu|G_Z$. The relationship (4) is employed to estimate the model parameters $\boldsymbol{\alpha}$ using the dataset Sa and $\widehat{\mathbf{y}}_{Sa} = \mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S$, i.e.

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{Sa} &= \left(\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\widehat{\mathbf{y}}_{Sa} \\ &= \left(\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S \\ &= \left(\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\widehat{\boldsymbol{\Sigma}}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\widehat{\boldsymbol{\Omega}}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\widehat{\boldsymbol{\Omega}}_S^{-1}\mathbf{y}_S. \end{aligned} \tag{5}$$

The variance-covariance matrices $\omega^2\boldsymbol{\Omega}$ and $\sigma^2\boldsymbol{\Sigma}$ are among the essential parameters to be estimated. Under assumptions of homoskedasticity and independence, $\omega^2\boldsymbol{\Omega}$ and $\sigma^2\boldsymbol{\Sigma}$ become $\omega^2\mathbf{I}$ and $\sigma^2\mathbf{I}$.

However, under heteroskedasticity and dependent observations the form of $\omega^2\boldsymbol{\Omega}$ and $\sigma^2\boldsymbol{\Sigma}$ is different. Under heteroskedasticity, the matrices' diagonal elements, corresponding to individual error variances, are unequal. Dependent errors may arise due to several reasons, important examples are clustered sample data, spatially autocorrelated errors and nested samples:

- with clustered data structure, $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are block-diagonal matrices, where the blocks correspond to the clusters;
- under spatial autocorrelation, the matrices' off-diagonal elements, corresponding to covariances between errors, are non-zero;
- with nested samples, the dependency between datasets S and Sa results in a non-zero covariance between errors of model (1) and relationship (4).

In the present study we didn't account for the uncertainty due to the estimation of $\omega^2\boldsymbol{\Omega}$ and $\sigma^2\boldsymbol{\Sigma}$, and assumed that they are known to a constant, i.e. $\widehat{\omega^2\boldsymbol{\Omega}}_S = \widehat{\omega^2}\boldsymbol{\Omega}_S$ and $\widehat{\sigma^2\boldsymbol{\Sigma}}_{Sa} = \widehat{\sigma^2}\boldsymbol{\Sigma}_{Sa}$. This is a common practice when GLS estimators are applied [cf. 28,29]. Therefore, since the $\widehat{\omega^2}$ and $\widehat{\sigma^2}$ terms will cancel, estimators (3) and (5) can be rewritten as

$$\widehat{\boldsymbol{\beta}}_S = \left(\mathbf{X}_S^{\intercal}\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\boldsymbol{\Omega}_S^{-1}\mathbf{y}_S, \tag{6}$$

and

$$\widehat{\boldsymbol{\alpha}}_{Sa} = \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\boldsymbol{\Omega}_S^{-1}\mathbf{y}_S. \tag{7}$$

The difference between (3) and (6), and (5) and (7) is that the covariance matrices are not estimated in (6) and(7) but assumed known [cf. 28,29].

The expected value of the finite population mean is then estimated as

$$\widehat{\mu}_{GHMB} = \boldsymbol{\iota}_U^{\intercal}\mathbf{Z}_U\widehat{\boldsymbol{\alpha}}_{Sa}, \tag{8}$$

where the subscript GHMB denotes "Generalized Hierarchical Model-Based" and $\boldsymbol{\iota}_U$ is an N-length vector, each element of which is $1/N$.

The variance of $\widehat{\mu}_{GHMB}$ is

$$\mathrm{V}(\widehat{\mu})_{GHMB} = \boldsymbol{\iota}_U^{\intercal}\mathbf{Z}_U\,\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)\mathbf{Z}_U^{\intercal}\boldsymbol{\iota}_U, \tag{9}$$

where $\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)$ (see Appendix A for details) is the core expression for the GHMB estimator variance. As shown in the Appendix A the covariance composed of four terms.

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right) = \mathrm{E}[\mathbf{b}\mathbf{b}^{\intercal}] + \mathrm{E}[\mathbf{c}\mathbf{c}^{\intercal}] + \mathrm{E}[\mathbf{b}\mathbf{c}^{\intercal}] + \mathrm{E}[\mathbf{c}\mathbf{b}^{\intercal}], \tag{10}$$

where

$$\mathrm{E}\left[\mathbf{b}\mathbf{b}^{\intercal}\right] = \omega^2 \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{X}_{S}\right)^{-1}\mathbf{X}_{Sa}^{\intercal}\right]\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1},$$

$$\mathrm{E}\left[\mathbf{c}\mathbf{c}^{\intercal}\right] = \sigma^2 \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1},$$

$$\mathrm{E}\left[\mathbf{b}\mathbf{c}^{\intercal}\right] = \mathrm{E}\left[\mathbf{c}\mathbf{b}^{\intercal}\right]^{\intercal} = \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{X}_{S}\right)^{-1}\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\,\mathrm{Cov}\left(\boldsymbol{\epsilon}_{S}, \mathbf{u}_{Sa}\right)\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}.$$

In the general case all these terms would be non-zero and contribute to the variance of the GHMB estimator. However, in case the S and Sa datasets are independent, then $\mathrm{E}\left[\mathbf{b}\mathbf{c}^{\intercal}\right] = \mathrm{E}\left[\mathbf{c}\mathbf{b}^{\intercal}\right]^{\intercal} = \mathbf{0}$. By replacing ω^2 and σ^2 with the corresponding estimators $\widehat{\omega^2}$ and $\widehat{\sigma^2}$ we obtain the covariance estimator $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)$, i.e.

$$\begin{aligned}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right) &= \widehat{\sigma^2}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} \\ &+ \widehat{\omega^2}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{X}_{S}\right)^{-1}\mathbf{X}_{Sa}^{\intercal}\right]\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}.\end{aligned} \tag{11}$$

where ω^2 in model (1) is estimated as [e.g., 28]

$$\widehat{\omega^2} = \frac{SSR(\boldsymbol{\beta}|\boldsymbol{\Omega}_S)}{df_S} = \frac{(\mathbf{y}_S - \mathbf{X}_S\widehat{\boldsymbol{\beta}}_S)^{\intercal}\boldsymbol{\Omega}_S^{-1}(\mathbf{y}_S - \mathbf{X}_S\widehat{\boldsymbol{\beta}}_S)}{n-(p+1)}, \tag{12}$$

whereas σ^2 due to relationship (4), was estimated as

$$\widehat{\sigma^2} = \frac{SSR(\boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa}) - \mathrm{Tr}\left[\mathbf{X}_{Sa}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right]}{M-(q+1)}, \tag{13}$$

where $SSR(\boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa}) = (\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})^{\intercal}\boldsymbol{\Sigma}^{-1}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})$, $\mathbf{I}_M$ is an identity matrix of dimension $M \times M$ and $\mathbf{H} = \mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}$. The correction $\frac{\mathrm{Tr}\left[\mathbf{X}_{Sa}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right]}{M-(q+1)}$ is needed because the response variable in relationship (4) is not measured but estimated using estimator (6).

It can be seen that the estimated covariance of $\widehat{\boldsymbol{\beta}}_S$ is a part of $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)$ by substituting in estimator (11) $\widehat{\omega^2}\left(\mathbf{X}_S^{\intercal}\boldsymbol{\Omega}_S\mathbf{X}_S\right)^{-1}$ to $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)$ [e.g., 28]. The derivation of estimators (11) and (13) are presented in Appendix A. Our GHMB variance estimator $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)$ is

$$\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB} = \boldsymbol{\iota}_U^{\intercal}\mathbf{Z}_U\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)\mathbf{Z}_U^{\intercal}\boldsymbol{\iota}_U. \tag{14}$$

The R package "HMB" by Saarela *et al.* [30] has function `ghmb()`, which provides estimates based on estimators (8) and (14). The package is based on a C++ library for linear algebra developed by [31].

2.3. Reference methods for comparison

We compared the performance of the GHMB estimator with the two-stage model-based estimation procedure presented in [18], which utilizes the same datasets. Additionally, we compared the GHMB estimator with estimators utilizing only a single source of RS data, i.e. an estimator for hybrid inference [24], and an estimator for conventional model-based inference [e.g., 17]. Some details of these reference methods are provided below.

2.3.1. Generalized Two-Stage Model-Based estimation (GTSMB)

In this estimation procedure, the regressors $\mathbf{X}$ in the model G_X, are regressands and dependent on $\mathbf{Z}$, i.e.

$$\mathbf{x}_k = \mathbf{Z}\boldsymbol{\gamma}_k + \mathbf{d}_k, \mathrm{E}\left[\mathbf{d}_k\right] = \mathbf{0},\ \mathrm{E}\left[\mathbf{d}_k\mathbf{d}_l^{\intercal}\right] = \delta_{kl}\boldsymbol{\Phi}_{kl}, \text{ in case } l{=}k\text{: } \delta_{kk}\boldsymbol{\Phi}_{kk} = \delta_k^2\boldsymbol{\Phi}_k \tag{15}$$

for the k^{th} variable out of $(p+1)$ variables in $\mathbf{X}$. The same set of $\mathbf{Z}$ is used to predict each variable in $\mathbf{X}$. Model parameters γ_k are estimated using information from the *Sa* dataset employing the GLS estimator, i.e. $\widehat{\gamma}_k = \left(\mathbf{Z}_{Sa}^\intercal \widehat{\mathbf{\Phi}}_k^{-1} \mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_{Sa}^\intercal \widehat{\mathbf{\Phi}}_k^{-1} \mathbf{x}_{S_{a_k}}$ [cf. 28]. Similarly to GHMB estimation, we do not account for the uncertainty due to the $\mathbf{\Phi}_k$ estimation, and, thus, assume $\delta_k^2 \mathbf{\Phi}_k$ to be known to a constant δ_k^2, i.e. $\widehat{\delta_k^2 \mathbf{\Phi}}_k = \widehat{\delta_k^2} \mathbf{\Phi}_k$. Thus,

$$\widehat{\gamma}_k = \left(\mathbf{Z}_{Sa}^\intercal \mathbf{\Phi}_k^{-1} \mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_{Sa}^\intercal \mathbf{\Phi}_k^{-1} \mathbf{x}_{S_{a_k}}. \tag{16}$$

In this approach the expected value of the finite population mean, μ, is estimated as

$$\widehat{\mu}_{GTSMB} = \boldsymbol{\iota}_U^\intercal \widehat{\mathbf{X}}_U \widehat{\boldsymbol{\beta}}_S, \tag{17}$$

where $\widehat{\mathbf{x}}_{U_k} = \mathbf{Z}_U \widehat{\gamma}_k$ and the subscript GTSMB denotes "Generalized Two-Stage Model-Based". The variance of $\widehat{\mu}_{GTSMB}$ is estimated as

$$\begin{aligned} \widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB} &= \boldsymbol{\iota}_U^\intercal \widehat{\mathbf{X}}_U \widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right) \widehat{\mathbf{X}}_U^\intercal \boldsymbol{\iota}_U \\ &+ \widehat{\boldsymbol{\beta}}_S^\intercal \widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^\intercal \widehat{\mathbf{X}}_U\right) \widehat{\boldsymbol{\beta}}_S \\ &- \textstyle\sum_{k=1}^{(p+1)} \sum_{l=1}^{(p+1)} \widehat{\mathrm{Cov}}\left(\widehat{\beta}_{S_k}, \widehat{\beta}_{S_l}\right) \widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^\intercal \widehat{\mathbf{x}}_{U_k}, \boldsymbol{\iota}_U^\intercal \widehat{\mathbf{x}}_{U_l}\right), \end{aligned} \tag{18}$$

where $\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^\intercal \widehat{\mathbf{x}}_{U_k}, \boldsymbol{\iota}_U^\intercal \widehat{\mathbf{x}}_{U_l}\right) = \widehat{\delta_{kl}} \boldsymbol{\iota}_U^\intercal \mathbf{Z}_U \left(\mathbf{Z}_{Sa}^\intercal \mathbf{\Phi}_k^{-1} \mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_{Sa}^\intercal \mathbf{\Phi}_k^{-1} \mathbf{\Phi}_{kl} \mathbf{\Phi}_l^{-1} \mathbf{Z}_{Sa} \left(\mathbf{Z}_{Sa}^\intercal \mathbf{\Phi}_l^{-1} \mathbf{Z}_{Sa}\right)^{-1} \mathbf{Z}_U^\intercal \boldsymbol{\iota}_U$ and $\widehat{\delta_{kl}} = \frac{(\mathbf{x}_{S_{a_k}} - \mathbf{Z}_{Sa}\widehat{\gamma}_k)^\intercal \mathbf{\Phi}_{kl}^{-1} (\mathbf{x}_{S_{a_l}} - \mathbf{Z}_{Sa}\widehat{\gamma}_l)}{M-(q+1)}$.

Details of the estimator (18) are presented in [18]. The R package "`HMB`" by Saarela *et al.* [30] was used to obtain estimates based on (17) and (18).

The correspondence between GHMB and GTSMB estimators were further evaluated as a part of this study. In Appendix B we show that under certain rather general conditions the estimators and variance estimators will provide approximately the same results for given datasets.

2.3.2. Hybrid estimation

We employed the hybrid estimators for clustered data described in [24]. The estimator utilizes sampled $\mathbf{X}$ data from the *Sa* dataset, and accounts for sampling uncertainty due to the *Sa* sample and modelling uncertainty due to the model (1) (i.e., $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$) using the dataset *S* [24].

Within hybrid inference, the population mean estimator is [24, Equation 11, p. 101]

$$\widehat{\mu}_{Hybrid} = \frac{\sum_{i=1}^{m} \widehat{F}_i}{\sum_{i=1}^{m} A_i}, \tag{19}$$

where m is the number of clusters in *Sa*, $\widehat{F}_i = \sum_{t=1}^{A_i} \widehat{y}_t = \sum_{t=1}^{A_i} \mathbf{x}_{it} \widehat{\boldsymbol{\beta}}_S$ is the i^{th} cluster total, and A_i is the number of population elements in the i^{th} cluster.

The corresponding variance estimator is [24, Equation 15, p. 101]

$$\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{Hybrid} = \frac{1}{\bar{A}^2} \left(\frac{\sum_{i=1}^{m} (\widehat{F}_i - \widehat{\mu}_{Hybrid} A_i)^2}{m(m-1)} \right) + \boldsymbol{\iota}_{Sa}^\intercal \mathbf{X}_{Sa} \widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right) \mathbf{X}_{Sa}^\intercal \boldsymbol{\iota}_{Sa}, \tag{20}$$

where $\bar{A} = \frac{1}{m} \sum_{i=1}^{m} A_i$.

2.3.3. Conventional model-based inference (MB)

This estimation procedure uses only one source of auxiliary information available wall-to-wall, in our case the Landsat 7 ETM+ data. The population mean estimator is [e.g., 17, Equation 3, p. 899]

$$\widehat{\mu}_{MB} = \boldsymbol{\iota}_U^\intercal \mathbf{Z}_U \widehat{\boldsymbol{\alpha}}_S, \tag{21}$$

where $\widehat{\boldsymbol{\alpha}}_S = \left(\mathbf{Z}_S^\intercal \boldsymbol{\Delta}_S^{-1} \mathbf{Z}_S\right)^{-1} \mathbf{Z}_S^\intercal \boldsymbol{\Delta}_S^{-1} \mathbf{y}_S$ is a vector of estimated model parameters using the dataset S following superpopulation model G_Z in equation (2). As in the previous approaches, in this case we also assume that the covariance $\theta^2 \boldsymbol{\Delta}_S$ is known to a constant θ^2, i.e. $\widehat{\theta^2 \boldsymbol{\Delta}_S} = \widehat{\theta^2} \boldsymbol{\Delta}_S$.

The model-based variance estimator is

$$\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{MB} = \boldsymbol{\iota}_U^\intercal \mathbf{Z}_U \widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_S\right) \mathbf{Z}_U^\intercal \boldsymbol{\iota}_U, \tag{22}$$

where $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_S\right) = \widehat{\theta^2}\left(\mathbf{Z}_S^\intercal \boldsymbol{\Delta}_S^{-1} \mathbf{Z}_S\right)^{-1}$ and $\widehat{\theta^2} = \frac{SSR(\boldsymbol{\alpha}|\boldsymbol{\Delta}_S)}{df_S} = \frac{(\mathbf{y}_S - \mathbf{Z}_S \widehat{\boldsymbol{\alpha}}_S)^\intercal \boldsymbol{\Delta}_S^{-1} (\mathbf{y}_S - \mathbf{Z}_S \widehat{\boldsymbol{\alpha}}_S)}{n-(q+1)}$ [cf. 28].

3. Material

For this study we simulated six populations mimicking forest conditions in study areas across the USA (Figure 2). For each site reference data were available from field measurements, Landsat 7 ETM+, and laser scanning using a method that mimics future measurements with the GEDI space LiDAR. In this section we first describe the datasets and, secondly, how these reference data were used for simulating AGB and RS data for each study site.

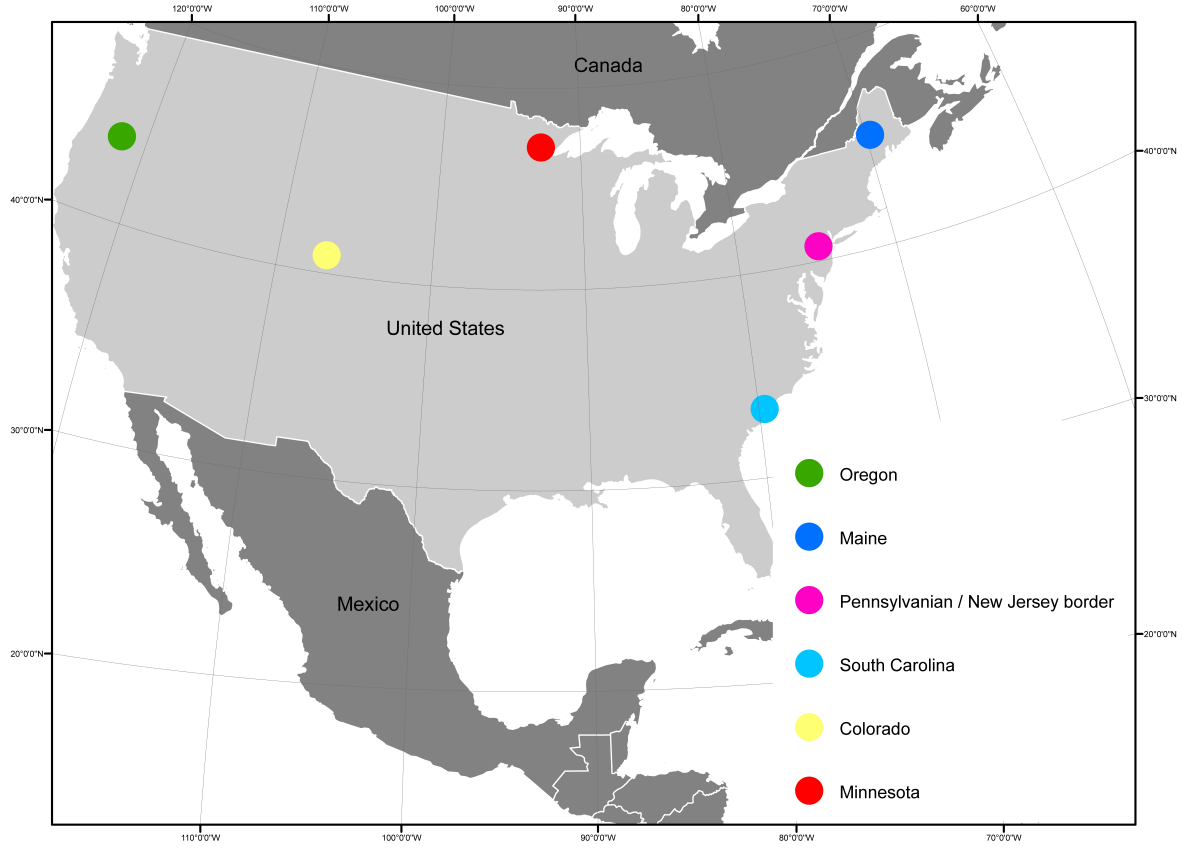


Figure 2. Study sites across the USA.

3.1. Reference data

3.1.1. Simulated GEDI data

GEDI's nominal footprint diameter will be approximately 25 meters, with approximately 60-meter along-track spacing [21]. After two years of operation aboard the International Space Station (ISS), the cross-track spacing of GEDI's sample is expected to be 500 meters, producing a semi-regular

lattice of sample lines from 51° South to 51° North (the area of the Earth traversed by the ISS orbit). Waveform properties used in the modeling process were quantified by relative heights (rh) from below which different amounts of energy were reflected. So, 'rh90' was the height below which 90% of the waveform energy returned, which was higher by an amount dictated by the vertical distribution of canopy material than 'rh30'.

The LiDAR data used in this study were derived by the GEDI Science Team from airborne discrete-return LiDAR (DRL) acquisitions, transformed to resemble the return waveforms to be collected by the GEDI mission. Further details related to LiDAR and field data acquisition are given in Appendix C. The transformation process is one where individual photon returns from across a given surface area are grouped into height quantiles, which are then integrated to derive a waveform describing the return height function for that area [32]. The GEDI Waveform Simulator realistically accounts for topographic and canopy penetration issues, while effecting the above physical transformation of return energy information from discrete to continuous functions.

The DRL acquisitions forming the basis of this study were collected with a RIEGL LMS-Q680i airborne laser scanner across the six study sites (Maine, Pennsylvania, South Carolina, Colorado, Minnesota) between June and August, 2014, the Oregon site was scanned in June, 2015. Data were collected in North-South lines that were 300-1000 meters in width and spaced 5 km apart. Pulse density was at least 4 pulses per square meter. The DRL sample lines were transformed to simulate a surface of contiguous GEDI footprints along the sample lines. Ten strips of GEDI data, each about 510 m long, were available from each site [33].

3.1.2. Landsat 7 ETM+ data

Landsat values were derived for each field plot from surface reflectance values (multiplied by 10^4) generated from the LEDAPS algorithm [34]. Pre-collection surface reflectance imagery from June to September, 2015 for band 3 (red; B3) and band 4 (near-infrared; B4) were composited using a medoid method (multi-dimensional median; [35]), screening out clouds and cloud shadows using the F-mask algorithm [36]. This processing was conducted on the Google Earth Engine platform [37].

3.1.3. Field data

Between 46 and 50 field plots were established in each of the six study sites from June-August, 2014 (Maine, Pennsylvania, South Carolina, Colorado, Minnesota) and from June - August 2015 (Oregon, see Appendix C for more details). The sample was designed to cover the range of vegetation conditions in each area, making use of 15 strata covering the bivariate distribution of LiDAR-based estimates of height (broken into 5 classes) and vegetation cover (3 classes). The sampled distribution excluded locations with less than 10% canopy cover (according to LiDAR metrics) and areas of non-forest according to the National Land Cover Dataset [38]. Approximately 25 random locations were chosen within each stratum, from which an approximately equal number (3-4) were chosen for field measurement based upon accessibility.

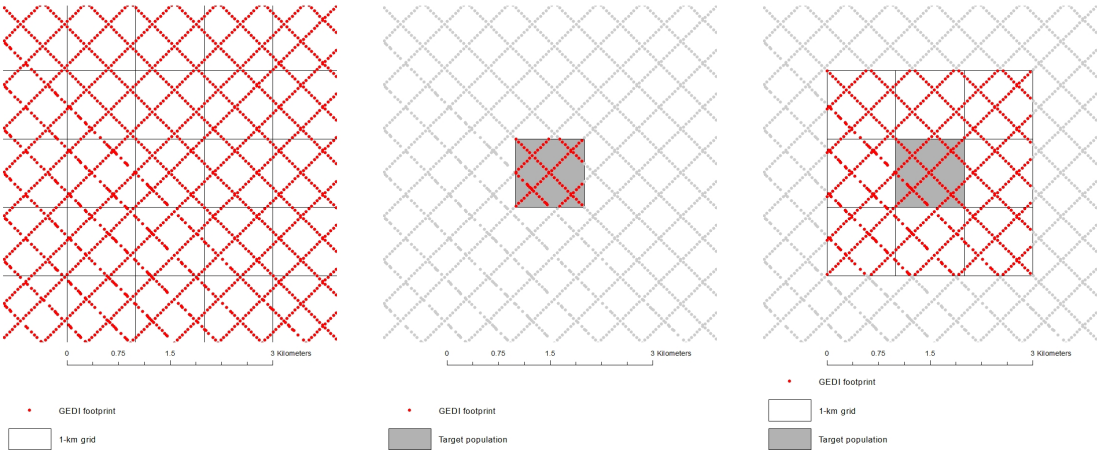
Standard field protocols used by the US National Forest Inventory (managed by the US Forest Service Forest Inventory and Analysis program) were used in measuring trees on 16.15 m radius plots (yielding approximately the same area plot as a Landsat pixel) [39]. For details see Appendix C. AGB was calculated from the tree list for each plot as described in [39]. Plot numbers, forest types according to [40,41] and statistics are summarized in Table 1.

Table 1. Overview of field data.

Study site	Forest type	Number of field plots	AGB, [Mg/ha] min	mean	max	sd
OR – Oregon	Douglas Fir (24%), Ponderosa Pine (33%), Fir Spruce/Mountain Hemlock (22%), Lodgepole Pine (20%)	49	4.7	227.5	775.1	205.7
ME – Maine	Spruce/Fir (49%), Maple/Beech/Birch (50%)	48	0.2	64.1	264.8	69.7
PANJ — Pennsylvanian / New Jersey border	Loblolly/Shortleaf Pine (25%), Oak/Hickory (62%), Maple/Beech/Birch (6%)	49	0.0	89.1	378.8	106.7
SC — South Carolina	Loblolly/Shortleaf Pine (66%), Oak/Gum/Cypress (28%)	50	0.0	72.7	373.3	80.1
CO – Colorado	Pinyon/Juniper (17%), Fir/Spruce/Mountain Hemlock (27%), Aspen/Birch (36%), Western Oak (12%)	46	0.0	133.8	353.5	96.2
MN – Minnesota	Spruce/Fir (28%), Aspen/Birch (68%)	47	0.0	48.1	202.0	48.5

3.2. Simulated populations

Each of the six areas depicted in Figure 2 were simulated independently; each contained approximately 50 field plots, with corresponding Landsat 7 ETM+ and GEDI data (simulated from DRL). In general, the target for the GEDI mission is to report AGB estimates by 1-km squares. Thus, the setup will be according to Figure 3**a**, i.e. the area is tessellated into 1-km grid-cells, each of which will have a certain number of GEDI footprints (the points) and potentially a wall-to-wall cover of Landsat data.



a) GEDI footprint simulated spatial pattern **b)** Case A **c)** Case B

Figure 3. a) Simulated spatial pattern of GEDI footprints and the 1-km grid delineation; **b)** Case A: data for a single 1-km grid-cell were used; **c)** Case B: data from neighboring grid-cells were used as well, for the GHMB and GTSMB estimators.

Our case study examples are of two kinds. In Case A, only data (GEDI and Landsat 7 ETM+) from within a given 1-km grid-cell are used. In Case B, data from eight neighboring grid-cells were used in addition to the center (target) grid, since model development data from a larger similar area may improve GHMB estimation. Note, that only the GHMB and GTSMB approaches can benefit from data from surrounding grid-cells, since such data can be used to improve the model used in the final step of the GHMB (and GTSMB) procedure, that is, developing a model linking predicted biomass values (based on GEDI data) with wall-to-wall Landsat data, and in the case of GTSMB procedure developing a multivariate model linking GEDI variables with Landsat data. The conventional MB and the hybrid methods cannot benefit from including a support area of this kind. The two cases are shown in Figure 3**b,c**.

Simulation were made to provide the data needed for Case B; Case A data were obtained as the corresponding data from the center 1-km grid-cell. Each 1-km grid-cell was tessellated into 2500 square plots of a size corresponding to the future GEDI LiDAR footprint size.

3.2.1. Correlated multinomial random variables

To create our simulated populations mimicking forest conditions in the study sites we need correlation matrices between the AGB, GEDI and Landsat 7 ETM+ variables. The matrices were calculated using reference data and are given in Table 2.

Table 2. Correlations between the study variables for each study site.

Study site	AGB	GEDI rh60	GEDI rh90	Landsat B3	Landsat B4	
	1	0.79	0.85	-0.49	-0.19	AGB
	–	1	0.78	-0.55	-0.03	GEDI rh60
OR	–	–	1	-0.43	-0.12	GEDI rh90
	–	–	–	1	-0.17	Landsat B3
	–	–	–	–	1	Landsat B4
	1	0.90	0.87	-0.26	-0.44	AGB
	–	1	0.83	-0.41	-0.35	GEDI rh60
ME	–	–	1	-0.28	-0.41	GEDI rh90
	–	–	–	1	0.14	Landsat B3
	–	–	–	–	1	Landsat B4
	1	0.82	0.70	-0.54	0.06	AGB
	–	1	0.87	-0.62	0.24	GEDI rh60
PANJ	–	–	1	-0.63	0.45	GEDI rh90
	–	–	–	1	-0.37	Landsat B3
	–	–	–	–	1	Landsat B4
	1	0.89	0.83	-0.36	0.07	AGB
	–	1	0.85	-0.37	0.03	GEDI rh60
SC	–	–	1	-0.34	0.05	GEDI rh90
	–	–	–	1	-0.26	Landsat B3
	–	–	–	–	1	Landsat B4
	1	0.82	0.78	-0.38	-0.36	AGB
	–	1	0.82	-0.42	-0.22	GEDI rh60
CO	–	–	1	-0.46	-0.35	GEDI rh90
	–	–	–	1	0.02	Landsat B3
	–	–	–	–	1	Landsat B4
	1	0.83	0.89	-0.24	0.04	AGB
	–	1	0.80	-0.33	0.06	GEDI rh60
MN	–	–	1	-0.27	0.12	GEDI rh90
	–	–	–	1	-0.14	Landsat B3
	–	–	–	–	1	Landsat B4

We applied a classical method for generating correlated multivariate normal random variables for a given covariance matrix, i.e.

$$\mathbf{a}^{\mathsf{T}} = \mathbf{L}\mathbf{f}^{\mathsf{T}} + \boldsymbol{\mu}^{\mathsf{T}} \tag{23}$$

where, $\mathbf{a} = (a_1, ..., a_t)$ is the desired multivariate normal vector of t variables (in our case study the variables are AGB, rh60, rh90, B3 and B4), $\mathbf{f} = (f_1, ..., f_t)$ is a row vector of independent standard normals, $\boldsymbol{\mu} = (\mu_1, ..., \mu_t)$ is the vector of corresponding means, and $\mathbf{L}$ is a triangular matrix such that $\mathbf{L}\mathbf{L}^{\mathsf{T}} = \mathrm{Cov}(a_k, a_l)$. We applied Cholesky decomposition to derive the matrix $\mathbf{L}$. The covariance matrix $\mathrm{Cov}(a_k, a_l) = \mathrm{Corr}(a_k, a_l)\,\mathrm{sd}(a_k)\,\mathrm{sd}(a_l)$ and the expected values $\boldsymbol{\mu}$ were estimated using reference data for each study site. Table 3 provides $\mathrm{sd}(a_k)$ and expected values $\boldsymbol{\mu}$ for each study site.

Table 3. Mean values and standard deviations of the study variables for each site.

Study site	AGB	GEDI rh60	GEDI rh90	Landsat B3	Landsat B4	
OR	227.5	10.3	22.1	378.3	1990.3	μ
	205.7	11.1	13.7	205.5	540.7	sd
ME	64.1	5.2	10.2	278.2	3074.5	μ
	69.7	4.6	5.4	94.5	562.7	sd
PANJ	89.1	8.2	14.1	279.6	3505.6	μ
	106.7	9.2	9.7	89.4	93.6	sd
SC	72.7	7.5	13.2	289.3	2825.7	μ
	80.1	7.9	8.3	128.8	417.4	sd
CO	133.8	5.9	12.8	388.6	2369.6	μ
	96.2	4.9	6.4	155.0	834.4	sd
MN	48.1	4.9	10.2	280.0	3243.0	μ
	48.5	4.7	6.4	89.0	761.3	sd

The independent vector of standard normal random variables $\mathbf{f}$ was generated randomly for a given simulated spatial autocorrelation of exponential form as a function of spatial distances between square plots, i.e. $\rho_{i,j} = e^{-a(distance_{i,j})}$. The spatial autocorrelation values are presented in Table 4.

Table 4. Simulated spatial autocorrelation between two square plots center-points at 20 m distance for AGB variable for each study site.

Study site	$\rho_{i,j}$	a	Study site	$\rho_{i,j}$	a	Study site	$\rho_{i,j}$	a
OR	0.72	1.64×10^{-2}	PANL	0.82	0.99×10^{-2}	CO	0.44	4.10×10^{-2}
ME	0.59	2.64×10^{-2}	SC	0.59	1.64×10^{-2}	MN	0.77	1.31×10^{-2}

We emphasis that for purposes of this study we assumed AGB autocorrelation as an average of GEDI rh60 and rh90 autocorrelation values (estimated using GEDI simulated strip data), given strong correlation between AGB and GEDI variables. This is an important point that the AGB autocorrelation assessment was not a part of the computation process.

3.3. Regression modeling

For each study site five regression models were fitted. Table 5 gives the model description and Table 6 provides information on the degrees of freedom, which were employed in the regression models involved for the different estimation methods. Dataset S was randomly selected from the square plots belonging to the target population U (1-km grid-cell) and its support area (eight neighboring 1-km grid-cells) corresponding to Case B, i.e. from the 3×3 km area (see Figure 3**c**).

Table 5. The regression models employed in the case study sites.

Notation	Description	Application	Dataset
AGB-GEDI	A model is linking AGB as a response variable with GEDI variables as predictor variables	GHMB (first level of modeling hierarchy), GTSMB and Hybrid	S
AGB-Landsat (Sa)	A model linking predicted AGB as a response variable with Landsat predictor variables	GHMB (second level of modeling hierarchy)	*Sa*
GEDIrh60-Landsat GEDIrh90-Landsat	Models linking GEDI response variables with Landsat predictor variables	GTSMB	*Sa*
AGB-Landsat (S)	A model linking AGB as a response variable with Landsat predictor variables	MB	*S*

Table 6. Degree of freedom (*df*) for the models involved.

Case	Model	*df*
A, B	AGB-GEDI and AGB-Landsat (S)	47
A	AGB-Landsat (Sa), GEDI(rh60)-Landsat and GEDI(rh90)-Landsat	91
B	AGB-Landsat (Sa), GEDI(rh60)-Landsat and GEDI(rh90)-Landsat	724

3.4. Evaluation criteria

We used the relative standard error (rSE) to assess the precision of the estimators

$$\widehat{rSE} = 100\frac{\sqrt{\widehat{V}(\widehat{\mu})}}{\mu} \tag{24}$$

where μ is the expected value of population mean and in our study cases is the mean value of AGB over field plots (see Table 3).

We also analyzed the uncertainty contribution of different sources in GHMB and hybrid estimation, i.e. for GHMB uncertainty contribution (i) due to the AGB-GEDI model, and (ii) due to the AGB-Landsat (Sa) model; for hybrid uncertainty contribution (i) due to the modeling, and (ii) due to the sampling. We estimated proportions of each uncertainty contribution, i.e. for GHMB estimation the proportions are estimated as (estimator (11) with substituting $\widehat{\omega^2}\left(\mathbf{X}_S^\intercal\mathbf{\Omega}_S\mathbf{X}_S\right)^{-1}$ to $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)$, and estimator (14)):

$$\mathrm{Prop}_{\text{AGB-GEDI}} = \frac{\boldsymbol{\iota}_U^\intercal\mathbf{Z}_U\left(\mathbf{Z}_{Sa}^\intercal\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^\intercal\mathbf{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^\intercal\right]\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^\intercal\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_U^\intercal\boldsymbol{\iota}_U}{\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB}}, \tag{25}$$

$$\mathrm{Prop}_{\text{AGB-Landsat(Sa)}} = \frac{\widehat{\sigma^2}\boldsymbol{\iota}_U^\intercal\mathbf{Z}_U\left(\mathbf{Z}_{Sa}^\intercal\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_U^\intercal\boldsymbol{\iota}_U}{\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB}}. \tag{26}$$

for hybrid estimation the proportions are estimated as (estimator (20)):

$$\text{Prop}_{\text{Modelling}} = \frac{\boldsymbol{\iota}_{Sa}^{\mathsf{T}} \mathbf{X}_{Sa} \widehat{\text{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right) \mathbf{X}_{Sa}^{\mathsf{T}} \boldsymbol{\iota}_{Sa}}{\widehat{\text{V}}\left(\widehat{\mu}\right)_{Hybrid}}, \tag{27}$$

$$\text{Prop}_{\text{Sampling}} = \frac{\left(\frac{\sum_{i=1}^{m}(\widehat{F}_i - \widehat{\mu}_{Hybrid} A_i)^2}{m(m-1)}\right)}{\overline{A}^2 \widehat{\text{V}}\left(\widehat{\mu}\right)_{Hybrid}}. \tag{28}$$

Ideally the performance of estimators and variance estimators should have been evaluated through Monte Carlo simulations with repeated simulation of both the populations and selection of datasets S and Sa, and the estimations. However, as seen from the formulas (11) and (14) the variance estimator is conditional on the sample outcomes $\mathbf{X}_S$ and $\mathbf{Z}_{Sa}$. Thus different simulated samples will result in different variances. It is the random deviations $\boldsymbol{\epsilon}$ and $\mathbf{u}$ that causes uncertainty according to the formulas. Therefor simulations that result in different $\mathbf{X}_S$ and $\mathbf{Z}_{Sa}$ cannot strictly be used to compare the empirical variance with the mean estimated. The simulated variances should be seen as examples of the size. For this reason we have chosen a restricted number of simulations. Simulations of only $\boldsymbol{\epsilon}$ and $\mathbf{u}$ could be used for empirical studies, but then (for a given set of simulations) only for a single fixed sample $\mathbf{X}_S$ and $\mathbf{Z}_{Sa}$.

4. Results

In Table 7, the expected population mean AGB values, μ, the corresponding simulated mean values, $\overline{y}_U$, and the estimates for the GHMB and GTSMB (Cases A and B), hybrid, and MB methods are presented. It can be seen that the AGB varied quite substantially between the different sites with the highest value in Oregon and the lowest in Minnesota. The average of estimated values over 250 repetitions were always fairly close to the average of simulated true means, $\overline{y}_U$.

Table 7. Estimated expected value of population mean, $\widehat{\mu}$, [Mg/ha].

Study site	μ	$\overline{y}_U$	Two sources of RS data				One source of RS data	
			Case A		Case B		Hybrid	MB
			GHMB	GTSMB	GHMB	GTSMB		
OR	227.5	226.8	227.4	227.4	226.5	226.5	226.3	225.6
ME	64.1	64.4	64.6	64.6	64.0	64.0	64.5	64.5
PANJ	89.1	88.1	88.0	88.0	88.3	88.3	88.0	88.2
SC	72.7	72.6	73.4	73.4	72.7	72.7	73.3	73.6
CO	133.8	133.1	133.7	133.7	133.6	133.6	133.7	133.2
MN	48.1	48.9	49.0	49.0	48.7	48.7	48.9	49.2

Our assessment of the performance of the different methods is based on estimated variances, recalculated and expressed as relative standard errors, for each method in each of the study areas as well as on average across the different study areas.

In Figure 4, the average relative standard error for the different methods across the study sites is presented. In this case a relative variance was first estimated for each site; then an average for the sites was calculated from which the square root was computed. On average, the GHMB and GTSMB methods performed about equally well (Appendix B), and they were superior to the other methods in case data from neighboring grid cells were used for improving the models (case B). When data from the target grid cell, only, were used for the model building (case A), the GHMB and GTSMB methods were outperformed by the hybrid and conventional MB estimation methods.

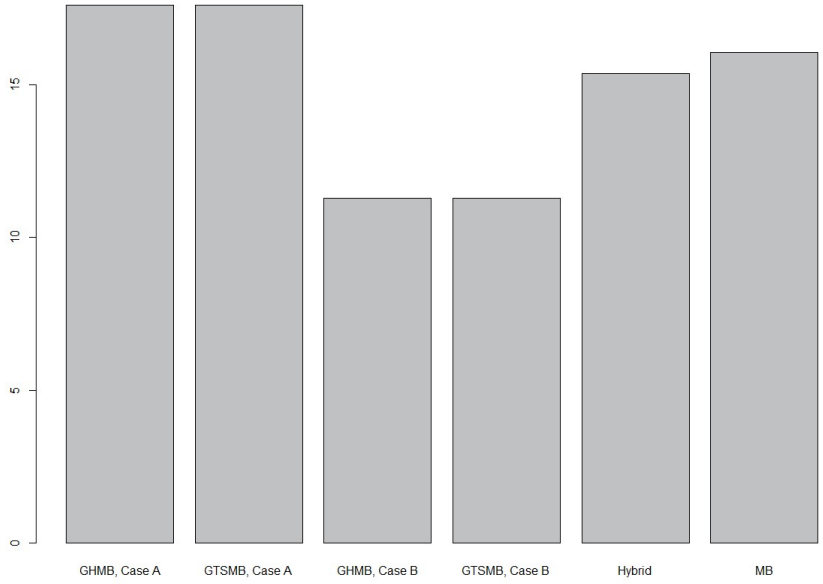


Figure 4. The average relative standard error for the different methods, [%].

Relative standard error for each of the methods in each site is detailed in Table 8. It can be seen that the performance of the methods, in terms of relative standard error, varied substantially between the sites. The hybrid estimation method showed fairly consistent results, in terms of precision, across the different sites with the smallest relative standard error in Colorado and the largest in Pennsylvanian and New Jersey border. In this simulation study we were not able to indicate any specific relation between estimators' performance and forest types. To conduct such analysis, real life data would be required rather than simulated. The MB method performed almost as well as hybrid estimation with similar patterns across study sites, and GHMB and GTSMB Case A estimations. The GHMB and GTSMB methods typically decreased their relative standard errors by about 40% when data from surrounding grid-cells were applied in the model building.

Table 8. Relative standard error in the study sites, $\widehat{rSE}$, [%]

Study site	Two sources of RS data				One source of RS data	
	Case A		Case B		Hybrid	MB
	GHMB	GTSMB	GHMB	GTSMB		
OR	14.8	14.8	9.4	9.4	13.1	13.3
ME	14.0	14.0	8.0	8.0	13.8	15.2
PANJ	25.4	25.4	17.5	17.5	20.3	21.2
SC	15.2	15.2	9.0	9.0	14.2	16.3
CO	8.6	8.6	6.3	6.3	8.8	9.1
MN	19.5	19.5	10.8	10.8	15.7	18.4

4.1. Sources of uncertainty for the GHMB and hybrid estimation methods

Table 9 shows the contribution of different sources of uncertainty to the variance GHMB and hybrid estimation methods. It can be seen that the contribution of the AGB-GEDI estimated model uncertainty is substantial for both estimation methods; for GHMB Case A it varies between 19% (MN) and 47% (CO), in Case B between 54% (CO) and 81% (MN), and in hybrid estimation between 24% (ME)

and 60% (PANJ). Comparing Cases A and B in GHMB estimation, we can see that with an increased number of GEDI footprints (from 94 to 727), the uncertainty contribution of the AGB-Landsat (Sa) estimated model decreased.

Table 9. Proportion (percentage) of the variance due to different sources for the GHMB and hybrid estimation methods, [%]

Study site	GHMB				Hybrid	
	Due to AGB-GEDI		Due to AGB-Landsat (Sa)		Due to Modeling	Due to Sampling
	Case A	Case B	Case A	Case B		
OR	30.4	76.1	69.6	23.9	39.3	60.7
ME	22.7	69.0	77.3	31.0	23.5	76.5
PANJ	37.9	80.3	62.1	19.7	60.5	39.5
SC	26.2	73.1	73.8	26.9	30.2	69.8
CO	46.5	87.0	53.5	13.0	44.8	55.2
MN	19.1	62.6	80.9	37.4	29.7	70.3

5. Discussion

In this article the hierarchical model-based estimation method presented by Saarela *et al.* [17] has been extended to cases when the model errors have non-homogeneous variance and cases where the errors are correlated, for example due to clustered sample data and/or spatial autocorrelation. A main component of this development is that generalized least squares theory is applied for the parameter estimation in the regression analysis, in which case the parameter estimation and the corresponding estimation of the covariance matrix for the parameter estimates accommodate such data structures. Also, the similarities between the GHMB and the GTSMB methods [18] are further explored and a proof is given (Appendix B) that the two methods will provide identical estimates and variances, and almost identical variance estimates, under certain conditions. However, with the expansion of the GHMB theory presented in this article, the GHMB method has a potential to be applied under a wider range of conditions than the GTSMB method, which among other things assumes independence between the two datasets used for the model building. The GHMB method might also be considered more intuitive to apply since it directly predicts the reference data (AGB values) that are used for the model building in the final step. The GTSMB method, on the other hand, uses the wall-to-wall RS data for predicting the metrics that would have been obtained from the sampled RS data source.

The results from the case studies indicate that the GHMB and GTSMB methods lead to more precise results, when a large support area is used for developing the model linking the wall-to-wall dataset (simulated Landsat data in our case) with the model predictions from the sample RS data (simulated GEDI data in our case). Restricting the data for the model building to a smaller area decreases the precision of the two methods. The hybrid estimation method and the conventional MB method were superior to the GHMB and GTSMB methods when only a 1-km support area was used. With the hybrid method, the AGB predictions from the GEDI sample, only, are used for the inference in our case. The conventional MB method makes predictions for all units in the wall-to-wall dataset (simulated Landsat data). Note, that Case B is only relevant for GHMB and GTSMB, since the dataset S (the field sample plots) was assumed to have a fixed size regardless of the size of the target area, and thus, the models used for the MB estimation would be the same in Case A and Case B. Thus "borrowing strength" for the model development from surrounding grid-cells does not change anything in the cases of conventional MB estimation. This holds true also for hybrid estimation, based on the same argument.

Overall, the performance of the different methods depends on many factors. A first requirement is that the study area is large enough, so that the assumption that the superpopulation mean value

is approximately the same as the study area mean holds. Further, the goodness-of-fit of the models involved is another core issue. With a very good model linking wall-to-wall data with AGB values, there is no need for complicating the estimation with hybrid, GHMB or GTSMB methods. However, most wall-to-wall RS datasets are weakly correlated with AGB which is the reason why samples of RS data that are strongly correlated with AGB are of interest as additional sources of auxiliary data in more advanced estimation procedures. Lastly, the sample sizes of the *S* and *Sa* datasets are important for the precision of the estimators. In general, increasing the sample sizes will increase their precision. Table 9 supports this statement, as it shows that 47 degree of freedom for fitting the AGB-GEDI model resulted in low precision of estimated $\boldsymbol{\beta}$ and, hence, the uncertainty contribution to the estimated variance in GHMB estimation for Case A was between 38% and 19%, and for Case B between 62% and 87%. The low precision of estimated $\boldsymbol{\beta}$ lead to the situation that the conventional MB method with only one source of weakly correlated wall-to-wall RS data (simulated Landsat data in our example) slightly outperformed GHMB (and GTSMB) in Case A, where only 94 GEDI footprints were available (see Figure 4).

The niche for GHMB and GTSMB estimation appears to be the important cases where the following requirements are fulfilled: (i) field data are sparse and expensive to acquire, (ii) wall-to-wall RS data (or a large sample of RS data) are available, which can be fairly well fit to the target variable, and (iii) RS data which can be very well fit to the target variable are available, but only samples of such data can be acquired. Since AGB models based on laser data can be expected to be more generally applicable across large regions than AGB models from Landsat data, it should make sense to train local models of the latter kind using pseudo-field data from predictions with the former kind of models, and thus apply GHMB or GTSMB estimation.

Several methodological issues may be brought up in relation to this study:

- One important issue is that all the case study data were simulated, based on sparse samples of reference data. The simulations are simplified generalizations of the real world, that leave out many important issues that must be handled in practical surveys, such as delineating forests from non-forest land [42]. In practical applications, land-use maps need to be applied to delineate forests before the GHMB method is employed.
- Another important restriction of the present study is that the results are based on estimates from a small number of iterations. A future expanded study should be based on Monte Carlo simulations of both the populations and the sampling from these populations, as a basis for empirical evaluation of the proposed estimators. Another method for simulating the multivariate variable **a** in equation (23) might be applied to demonstrate explicitly abilities of the proposed estimators.
- Further, future studies on this subject should also deal with the details of how to estimate the correlation matrices of model errors that are required for the GLS regression; in this article these details are only briefly addressed. One of the solutions to this problem could be iterative re-weighted least squares regression methods, such methods are often applied in geostatistical approaches.
- The current GHMB estimator is derived under the assumption that the target population is large, i.e. so that the population mean is at least approximately equal to the superpopulation mean. Modifications of the GHMB estimators for small-area estimation should be addressed in a potential future study.
- In the current study we assumed that the regression models involved in the AGB assessment by means of GEDI and Landsat data are linear. However, in reality the relationship between AGB (or growing stock volume) and height-like measures tends to be nonlinear [e.g., 23]. A further elaboration of the GHMB method for nonlinear models would be needed to handle such cases.
- Lastly, the performance of GHMB method in comparison to other methods should be analyzed as a basis for making recommendations on what method is appropriate under different conditions.

The above issues identify areas of needed research in the context of our findings. Including those issues into the current study would have increased the scope and length of the article considerably, and, thus, we chose to put those issues on hold for future studies.

6. Conclusions

Design-based inventory methods based on field sampling are limited by the availability of plot data. GHMB (and GTSMB) allows estimation of biomass in areas where there may be none or very limited field data, such as the 1-km grid-cells of interest to the GEDI mission, by taking advantage of multiple levels of RS data. Specifically, field data and high-quality RS data from similar areas outside the domain of interest can be used to calibrate wall-to-wall predictions within the domain of interest using synoptically collected RS data more weakly related to biomass. This paper augments previous hierarchical estimation methods by presenting methods that can be applied in cases where model errors have non-homogenous variance and where the model errors are correlated. Both cases are relevant for use of data from the upcoming GEDI LiDAR mission.

Author Contributions: Conceptualization, Svetlana Saarela, Sören Holm and Göran Ståhl; Data curation, Svetlana Saarela, Sean P. Healey, Hans-Erik Andersen and Paul L. Patterson; Funding acquisition, Sean P. Healey and Erik Næsset; Investigation, Svetlana Saarela, Sören Holm, Sean P. Healey, Hans-Erik Andersen, Hans Petersson, Wilmer Prentius, Paul L. Patterson, Erik Næsset, Timothy G. Gregoire and Göran Ståhl; Methodology, Svetlana Saarela, Sören Holm and Göran Ståhl; Project administration, Svetlana Saarela, Sean P. Healey and Erik Næsset; Software, Svetlana Saarela and Sören Holm; Writing – original draft, Svetlana Saarela, Sören Holm and Göran Ståhl; Writing – review & editing, Svetlana Saarela, Sören Holm, Sean P. Healey, Hans-Erik Andersen, Hans Petersson, Wilmer Prentius, Paul L. Patterson, Erik Næsset, Timothy G. Gregoire and Göran Ståhl.

Funding: Funding was provided by grants from NASA's Carbon Monitoring System (Healey CMS2016, Cohen CMS2013) and ERA-GAS FORCLIMIT (grant number FR-2017/0006).

Acknowledgments: The authors gratefully acknowledge Steven Hancock of the University of Maryland, who simulated the GEDI waveforms, Warren Cohen (US Forest Service), who helped to direct much of the field and LiDAR data collection, Zhiqiang Yang (US Forest Service), who helped in the development of "HMB" R package, and GEDI Science Team. The authors are thankful to the four anonymous Reviewers, whose comments helped to improve the clarity of the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AGB	AboveGround Biomass	LiDAR	Light Detection And Ranging
GEDI	Global Ecosystem Dynamics Investigation	MB	Model-Based
GHMB	Generalized Hierarchical Model-Based	NFI	National Forest Inventory
GLS	Generalized Lest Squares	PALS	Portable Airborne Laser System
GTSMB	Generalized Two-Stage Model-Based	RS	Remote Sensing
Landsat 7 ETM+	Landsat 7 Enhanced Thematic Mapper Plus	UNFCCC	United Nations' Framework Convention on Climate Change

Appendix A. Generalized hierarchical model-based estimators

For a realization, i.e. our target population, we have superpopulation models:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}, \mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\intercal}] = \omega^2\boldsymbol{\Omega}, \tag{A.1}$$

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{v}, \mathrm{E}[\boldsymbol{v}] = \mathbf{0}, \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^{\intercal}] = \theta^2\boldsymbol{\Delta}. \tag{A.2}$$

We also defined the following relationship between the two models for the given realization:

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}, \mathrm{E}[\mathbf{u}] = \mathbf{0}, \mathrm{E}[\mathbf{u}\mathbf{u}^{\intercal}] = \sigma^2\boldsymbol{\Sigma}. \tag{A.3}$$

For given datasets S, Sa, and U, and matrices $\mathbf{\Sigma}_{Sa}$ and $\mathbf{\Omega}_S$:

$$\widehat{\boldsymbol{\alpha}}_{Sa} = \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\widehat{\mathbf{y}}_{Sa}. \tag{A.4}$$

Knowing that $\widehat{\mathbf{y}}_{Sa} = \mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S$ and $\widehat{\boldsymbol{\beta}}_S = \left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{y}_S$, we have

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{Sa} &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S,\\ &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{y}_S.\end{aligned}$$

Due to $\mathbf{y}_S = \mathbf{X}_S\boldsymbol{\beta} + \boldsymbol{\epsilon}_S$:

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{Sa} &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\left(\mathbf{X}_S\boldsymbol{\beta}+\boldsymbol{\epsilon}_S\right),\\ &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\boldsymbol{\beta} + \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\boldsymbol{\epsilon}_S.\end{aligned}$$

Due to $\mathbf{X}_{Sa}\boldsymbol{\beta} = \mathbf{Z}_{Sa}\boldsymbol{\alpha} + \mathbf{u}_{Sa}$:

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{Sa} &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\left(\mathbf{Z}_{Sa}\boldsymbol{\alpha}+\mathbf{u}_{Sa}\right) + \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\boldsymbol{\epsilon}_S.\\ \widehat{\boldsymbol{\alpha}}_{Sa}-\boldsymbol{\alpha} &= \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{u}_{Sa} + \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\boldsymbol{\epsilon}_S.\end{aligned}$$

Denoting
$\mathbf{b} = \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\boldsymbol{\epsilon}_S$, $\mathbf{c} = \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{u}_{Sa}$ and given that $\mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right] = \boldsymbol{\alpha}$, we have

$$\widehat{\boldsymbol{\alpha}} - \mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right] = \mathbf{b} + \mathbf{c}. \tag{A.5}$$

This gives $\left(\widehat{\boldsymbol{\alpha}} - \mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right]\right)\left(\widehat{\boldsymbol{\alpha}} - \mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right]\right)^{\intercal} = \mathbf{b}\mathbf{b}^{\intercal} + \mathbf{c}\mathbf{c}^{\intercal} + \mathbf{b}\mathbf{c}^{\intercal} + \mathbf{c}\mathbf{b}^{\intercal}$.

Thus, covariance is

$$\begin{aligned}\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) &= \mathrm{E}\left[\left(\widehat{\boldsymbol{\alpha}} - \mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right]\right)\left(\widehat{\boldsymbol{\alpha}} - \mathrm{E}\left[\widehat{\boldsymbol{\alpha}}\right]\right)^{\intercal}\right]\\ &= \mathrm{E}\left[\mathbf{b}\mathbf{b}^{\intercal}\right] + \mathrm{E}\left[\mathbf{c}\mathbf{c}^{\intercal}\right] + \mathrm{E}\left[\mathbf{b}\mathbf{c}^{\intercal}\right] + \mathrm{E}\left[\mathbf{c}\mathbf{b}^{\intercal}\right],\end{aligned} \tag{A.6}$$

where
$\mathrm{E}\left[\mathbf{b}\mathbf{b}^{\intercal}\right] = \omega^2\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^{\intercal}\right]\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}$,
$\mathrm{E}\left[\mathbf{c}\mathbf{c}^{\intercal}\right] = \sigma^2\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}$,
$\mathrm{E}\left[\mathbf{b}\mathbf{c}^{\intercal}\right] = \mathrm{E}\left[\mathbf{c}\mathbf{b}^{\intercal}\right]^{\intercal} = \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathrm{Cov}\left(\boldsymbol{\epsilon}_S, \mathbf{u}_{Sa}\right)\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}$.

Under the $S \perp Sa$ assumption, $\mathrm{E}\left[\mathbf{b}\mathbf{c}^{\intercal}\right] = \mathrm{E}\left[\mathbf{c}\mathbf{b}^{\intercal}\right]^{\intercal} = \mathbf{0}$:

$$\begin{aligned}\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) &= \mathrm{E}\left[\mathbf{b}\mathbf{b}^{\intercal}\right] + \mathrm{E}\left[\mathbf{c}\mathbf{c}^{\intercal}\right]\\ &= \sigma^2\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} + \omega^2\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^{\intercal}\right]\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\end{aligned} \tag{A.7}$$

Knowing that $\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_S\right) = \omega^2\left(\mathbf{X}_S^{\intercal}\mathbf{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}$ [e.g., 28] we can rewrite

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\alpha}}\right) = \sigma^2\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} + \left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\,\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^{\intercal}\right]\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} \tag{A.8}$$

Here we show that the trivial estimator for σ^2 through the sum of squared residuals divided by the degree of freedom, leads to a biased estimation. We also derive an unbiased $\widehat{\sigma^2}$ estimator.

We denote $\frac{SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})}{df_{Sa}} = \widetilde{\sigma^2}$. The expected value of $\widetilde{\sigma^2}$ is:

$$\begin{aligned}\mathrm{E}\left[\widetilde{\sigma^2}\right] =& \mathrm{E}\left[\frac{SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})}{df_{Sa}}\right] \\ =& \frac{\mathrm{E}\left[(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})\right]}{M-(q+1)}.\end{aligned}$$

Denoting $\mathbf{H} = \mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^\intercal \boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}$ and $\mathbf{A} = \left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}$ we have

$$\begin{aligned}\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa} =& \mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{y}_S \\ & -\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{y}_S \\ =& \mathbf{X}_{Sa}\mathbf{A}\mathbf{y}_S - \mathbf{H}\mathbf{X}_{Sa}\mathbf{A}\mathbf{y}_S \\ =& (\mathbf{I}_M - \mathbf{H})\mathbf{X}_{Sa}\mathbf{A}\mathbf{y}_S \\ =& (\mathbf{I}_M - \mathbf{H})\mathbf{X}_{Sa}\mathbf{A}(\mathbf{X}_S\boldsymbol{\beta} + \boldsymbol{\epsilon}_S) \\ =& (\mathbf{I}_M - \mathbf{H})(\mathbf{X}_{Sa}\boldsymbol{\beta} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S).\end{aligned}$$

Due to $\mathbf{X}_{Sa}\boldsymbol{\beta} = \mathbf{Z}_{Sa}\boldsymbol{\alpha} + \mathbf{u}_{Sa}$ (for the given realization of $\mathbf{X}_{Sa}$ and $\mathbf{Z}_{Sa}$):

$$\begin{aligned}\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa} =& (\mathbf{I}_M - \mathbf{H})(\mathbf{Z}_{Sa}\boldsymbol{\alpha} + \mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S) \\ =& (\mathbf{I}_M - \mathbf{H})(\mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S).\end{aligned}$$

Therefore,

$$\begin{aligned}SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa}) =& (\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})^\intercal\boldsymbol{\Sigma}^{-1}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S - \mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa}) \\ =& (\mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S)^\intercal(\mathbf{I}_M - \mathbf{H})^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})(\mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S).\end{aligned}$$

Given $(\mathbf{I}_M - \mathbf{H})^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H}) = \boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})$ we have

$$\begin{aligned}SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa}) &= (\mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S)^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})(\mathbf{u}_{Sa} + \mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S) \\ &= \mathbf{u}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\mathbf{u}_{Sa} + \boldsymbol{\epsilon}_S^\intercal\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S \\ &+ \mathbf{u}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\epsilon}_S + \boldsymbol{\epsilon}_S^\intercal\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\mathbf{u}_{Sa}.\end{aligned} \tag{A.9}$$

And thus, the expected value of the sum of squared residuals is (note: here we used that the quadratic form $\mathbf{t}^\intercal\mathbf{J}\mathbf{t} = \mathrm{Tr}\left[\mathbf{t}\mathbf{t}^\intercal\mathbf{J}\right] = \mathrm{Tr}\left[\mathbf{J}\mathbf{t}\mathbf{t}^\intercal\right]$ for any column vector $\mathbf{t}$ and matrix $\mathbf{J}$),

$$\begin{aligned}\mathrm{E}\left[SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})\right] =& \mathrm{Tr}\left[\mathrm{E}\left[\mathbf{u}_{Sa}\mathbf{u}_{Sa}^\intercal\right]\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right] + \mathrm{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\,\mathrm{E}\left[\boldsymbol{\epsilon}_S\boldsymbol{\epsilon}_S^\intercal\right]\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right] \\ &+ \mathrm{Tr}\left[\mathrm{E}\left[\mathbf{u}_{Sa}\boldsymbol{\epsilon}_S^\intercal\right]\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right] + \mathrm{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\,\mathrm{E}\left[\boldsymbol{\epsilon}_S\mathbf{u}_{Sa}^\intercal\right]\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M - \mathbf{H})\right].\end{aligned}$$

Knowing that $\mathrm{E}\left[\mathbf{u}_{Sa}\mathbf{u}_{Sa}^\intercal\right] = \sigma^2\boldsymbol{\Sigma}_{Sa}$, $\mathrm{E}\left[\boldsymbol{\epsilon}_S\boldsymbol{\epsilon}_S^\intercal\right] = \omega^2\boldsymbol{\Omega}_S$ and $\mathrm{E}\left[\boldsymbol{\epsilon}_S\mathbf{u}_S^\intercal\right] = \mathrm{E}\left[\mathbf{u}_{Sa}\boldsymbol{\epsilon}_S^\intercal\right]^\intercal = \mathrm{Cov}\left(\boldsymbol{\epsilon}_S, \mathbf{u}_{Sa}\right)$, we have

$$\begin{aligned}\mathrm{E}\left[SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})\right] =& \sigma^2 \operatorname{Tr}\left[(\mathbf{I}_M-\mathbf{H})\right]+\omega^2 \operatorname{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\Omega}_S\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\\ &+2\operatorname{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\operatorname{Cov}\left(\boldsymbol{\epsilon}_S,\mathbf{u}_{Sa}\right)\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right].\end{aligned}$$

Given $\operatorname{Tr}\left[(\mathbf{I}_M-\mathbf{H})\right]=M-(q+1)$ and $\mathbf{X}_{Sa}\mathbf{A}\boldsymbol{\Omega}_S\mathbf{A}^\intercal\mathbf{X}_{Sa}^\intercal=\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^\intercal$ we have

$$\begin{aligned}\mathrm{E}\left[SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})\right] =& \sigma^2(M-(q+1))+\omega^2 \operatorname{Tr}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\\ &+2\operatorname{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\operatorname{Cov}\left(\boldsymbol{\epsilon}_S,\mathbf{u}_{Sa}\right)\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right].\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{E}\left[\widetilde{\sigma^2}\right] &= \mathrm{E}\left[\tfrac{SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})}{df_{Sa}}\right]\\ &= \sigma^2+\tfrac{\omega^2}{M-(q+1)}\operatorname{Tr}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\\ &+ \tfrac{2}{M-(q+1)}\operatorname{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\operatorname{Cov}\left(\boldsymbol{\epsilon}_S,\mathbf{u}_{Sa}\right)\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\end{aligned}\tag{A.10}$$

It can be seen that $\widetilde{\sigma^2}$ estimator is biased, and thus, to derive an unbiased estimator $\widehat{\sigma^2}$ we have to correct $\widetilde{\sigma^2}$ for the estimated *BIAS*, i.e.

$$\begin{aligned}\widehat{BIAS} &= \tfrac{\widehat{\omega^2}}{M-(q+1)}\operatorname{Tr}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\\ &+ \tfrac{2}{M-(q+1)}\operatorname{Tr}\left[\mathbf{X}_{Sa}\mathbf{A}\widehat{\operatorname{Cov}}\left(\boldsymbol{\epsilon}_S,\mathbf{u}_{Sa}\right)\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right].\end{aligned}\tag{A.11}$$

And thus,

$$\widehat{\sigma^2}=\widetilde{\sigma^2}-\widehat{BIAS}.$$

Under the independence assumption between datasets S and Sa, $\operatorname{Cov}\left(\boldsymbol{\epsilon}_S,\mathbf{u}_{Sa}\right)=\mathbf{0}$, we have

$$\begin{aligned}\widehat{\sigma^2} &= \tfrac{1}{M-(q+1)}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S-\mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})^\intercal\boldsymbol{\Sigma}^{-1}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S-\mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})\\ &- \tfrac{\widehat{\omega^2}}{M-(q+1)}\operatorname{Tr}\left[\mathbf{X}_{Sa}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right].\end{aligned}\tag{A.12}$$

Knowing that $\widehat{\operatorname{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)=\widehat{\omega^2}\left(\mathbf{X}_S^\intercal\boldsymbol{\Omega}_S^{-1}\mathbf{X}_S\right)^{-1}$ we can rewrite estimator $\widehat{\sigma^2}$ as

$$\begin{aligned}\widehat{\sigma^2} &= \tfrac{1}{M-(q+1)}\left((\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S-\mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})^\intercal\boldsymbol{\Sigma}^{-1}(\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_S-\mathbf{Z}_{Sa}\widehat{\boldsymbol{\alpha}}_{Sa})-\operatorname{Tr}\left[\mathbf{X}_{Sa}\widehat{\operatorname{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\right)\\ &= \tfrac{1}{df_{Sa}}\left(SSR(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{\Sigma}_{Sa})-\operatorname{Tr}\left[\mathbf{X}_{Sa}\widehat{\operatorname{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^\intercal\boldsymbol{\Sigma}_{Sa}^{-1}(\mathbf{I}_M-\mathbf{H})\right]\right).\end{aligned}\tag{A.13}$$

Appendix B. A comparison between expectations and variance estimators of the two methods: GHMB and GTSMB

Below it is shown that the two methods GHMB and GTSMB lead to identical results under certain conditions. Throughout below we neglect the effects of estimating covariance matrices and use notations as if the matrices were known to a constant.

We start with the estimators: for GHMB we have,

$$\widehat{\mu}_{GHMB}=\boldsymbol{\iota}_U^\intercal\mathbf{Z}_U\widehat{\boldsymbol{\alpha}}_{Sa},\tag{B.1}$$

where, according to estimator (7)

$$\begin{array}{rcl} \widehat{\boldsymbol{\alpha}}_{Sa} & = & \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\widehat{\boldsymbol{\beta}}_{S}, \\ & = & \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}\left(\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{X}_{S}\right)^{-1}\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{y}_{S}. \end{array} \tag{B.2}$$

For GTSMB we have the same $\widehat{\boldsymbol{\beta}}_{S} = \left(\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{X}_{S}\right)^{-1}\mathbf{X}_{S}^{\intercal}\boldsymbol{\Omega}_{S}^{-1}\mathbf{y}_{S}$. The link between GEDI and Landsat data is given by

$$\mathbf{x}_{k} = \mathbf{Z}\boldsymbol{\gamma}_{k} + \mathbf{d}_{k}, \mathrm{E}\left[\mathbf{d}_{k}\right] = \mathbf{0}, \mathrm{E}\left[\mathbf{d}_{k}\mathbf{d}_{l}^{\intercal}\right] = \delta_{kl}\boldsymbol{\Phi}_{kl}, \text{ in case } l{=}k\text{: } \delta_{kk}\boldsymbol{\Phi}_{kk} = \delta_{k}^{2}\boldsymbol{\Phi}_{k}, \text{ for } k,l = 2,...,(p+1). \tag{B.3}$$

The parameters $\boldsymbol{\gamma}_{k}$ are estimated from the *Sa* sample and we have the GLS estimators

$$\widehat{\boldsymbol{\gamma}}_{k} = \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Phi}_{k}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Phi}_{k}^{-1}\mathbf{x}_{Sa,k}, \tag{B.4}$$

where $\boldsymbol{\Phi}_{k}^{-1}$ should be understood as $\boldsymbol{\Phi}_{k,Sa}^{-1}$. From (B.4) we obtain the vector of estimated GEDI values, given the Landsat $\mathbf{Z}$ ones as the columns of $\widehat{\mathbf{X}} = \mathbf{Z}\widehat{\boldsymbol{\Gamma}}_{Sa}$, where $\widehat{\boldsymbol{\Gamma}}_{Sa}$ is the matrix with the vectors $\widehat{\boldsymbol{\gamma}}_{k}$ as columns.

Thus, for GTSMB estimation we have the estimator

$$\widehat{\mu}_{GTSMB} = \boldsymbol{\iota}_{U}^{\intercal}\mathbf{Z}_{U}\widehat{\boldsymbol{\Gamma}}_{Sa}\widehat{\boldsymbol{\beta}}_{S}. \tag{B.5}$$

Comparing (B.1) and (B.5), we can see that $\widehat{\mu}_{GHMB} = \widehat{\mu}_{GTSMB}$, if

$$\boldsymbol{\Phi}_{k,Sa} = \lambda_{k}\boldsymbol{\Sigma}_{Sa} \text{ for some constants } \lambda_{k}, \text{ for } k = 2,...,(p+1) \tag{B.6}$$

and thus

$$\widehat{\boldsymbol{\Gamma}}_{Sa} = \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{X}_{Sa}. \tag{B.7}$$

This is so because the λ_{k} cancel in (B.4), and (B.7) multiplied by $\widehat{\boldsymbol{\beta}}_{S}$ is equal to $\widehat{\boldsymbol{\alpha}}_{Sa}$.

The relation (B.6) seems very restrictive. It means that the correlation matrices for the k error terms $\mathbf{d}_{k}$ vectors are identical. However, this is not unrealistic as the GEDI variables $\mathbf{x}_{k}$ are strongly correlated. Also, this common correlation matrix is assumed to be equal to the correlation matrix of the error terms of the GHMB model. This is not unrealistic either, since the GEDI variables should show a high correlation with field data y and so should the GHMB predictions $\widehat{\mathbf{y}}$. At least, the assumption (B.6) is an acceptable approximation. In the homoskedastic GEDI-Landsat case (when $\boldsymbol{\Sigma}$ and the $\boldsymbol{\Phi}_{k}$ are normed unit matrices) the relation (B.6) holds automatically, whether the Field-GEDI relation is homoskedastic or not.

Thus, $\widehat{\mu}_{GHMB} = \widehat{\mu}_{GTSMB}$ under certain conditions, exactly or approximately. Hence, the variances are so too. Still, it remains to see whether or not the two variance estimators are (almost) identical under some conditions. We restrict this study to the case where the S and Sa samples are selected independently. The GHMB variance estimator is then (estimators (11) and (14))

$$\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB} = \boldsymbol{\iota}_{U}^{\intercal}\mathbf{Z}_{U}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}_{Sa}\right)\mathbf{Z}_{U}^{\intercal}\boldsymbol{\iota}_{U}, \tag{B.8}$$

where

$$\begin{array}{rcl} \widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\alpha}}\right) & = & \widehat{\sigma^{2}}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1} \\ & + & \left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\left[\mathbf{X}_{Sa}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_{S}\right)\mathbf{X}_{Sa}^{\intercal}\right]\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}. \end{array} \tag{B.9}$$

The GTSMB variance estimator is (estimator (18))

$$\begin{array}{rcl} \widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB} & = & \boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\widehat{\mathbf{X}}_U^{\mathsf{T}}\boldsymbol{\iota}_U \\ & + & \widehat{\boldsymbol{\beta}}_S^{\mathsf{T}}\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U\right)\widehat{\boldsymbol{\beta}}_S \\ & - & \sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\widehat{\mathrm{Cov}}\left(\widehat{\beta}_{S_k},\widehat{\beta}_{S_l}\right)\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U_k},\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U_l}\right). \end{array} \tag{B.10}$$

We introduce the matrix $\mathbf{P}$ that is the same for all k and that, in accordance with the condition (B.6) is assumed to fulfill

$$\boldsymbol{\Phi}_k = \varphi_k\mathbf{P}, \tag{B.11}$$

for some constants $\varphi_k, k, l = 1, ..., (p+1)$ ($\mathbf{P} = \mathbf{P}_{Sa}$ throughout). Then $\widehat{\boldsymbol{\gamma}}_k = (\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{x}_{Sa,k}$ since φ_k cancels (see (B.4)), and

$$\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U = \boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{X}_{Sa}. \tag{B.12}$$

Thus, the first term in the (B.10) equals

$$\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\widehat{\mathbf{X}}_U^{\mathsf{T}}\boldsymbol{\iota}_U = \boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\left[\mathbf{X}_{Sa}\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\beta}}_S\right)\mathbf{X}_{Sa}^{\mathsf{T}}\right]\mathbf{P}^{-1}\mathbf{Z}_{Sa}(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_U^{\mathsf{T}}\boldsymbol{\iota}_U. \tag{B.13}$$

Due to the assumptions (B.6) and (B.11) we also have $\boldsymbol{\Sigma}_{Sa} = h^2\mathbf{P}$ for some constant h^2. Since h^2 cancels in the second term of estimator (B.9), we see that the contribution of this term to $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB}$ equals the first term in $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB}$.

Next, we will show that he first term of $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB}$ equals to the second one in $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB}$. For this we need to add that condition (B.11) also holds for the cross-covariances, i.e., that

$$\boldsymbol{\Phi}_{kl} = \varphi_{kl}\mathbf{P} \tag{B.14}$$

We have, since φ_k cancels,

$$\begin{array}{rcl} \boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_{U,k} & = & \boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\boldsymbol{\Phi}_{kk}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_{Sa}^{\mathsf{T}}\boldsymbol{\Phi}_{kk}^{-1}\mathbf{X}_{Sa,k} \\ & = & \boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{X}_{Sa,k} \end{array}$$

and from this and (B.14) we obtain

$$\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U,k},\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U,l}\right) = \widehat{\delta}_{kl}\varphi_{kl}\boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_U^{\mathsf{T}}\boldsymbol{\iota}_U. \tag{B.15}$$

Hence, by summation, we obtain for the second term of $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB}$

$$\begin{array}{rcl} \widehat{\boldsymbol{\beta}}_S^{\mathsf{T}}\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U\right)\widehat{\boldsymbol{\beta}}_S & = & \sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\widehat{\beta}_k\widehat{\beta}_l\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U,k},\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{x}}_{U,l}\right) \\ & = & \sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\widehat{\beta}_k\widehat{\beta}_l\widehat{\delta}_{kl}\varphi_{kl}\boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_U^{\mathsf{T}}\boldsymbol{\iota}_U. \end{array} \tag{B.16}$$

Next we will show that $\sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\widehat{\beta}_k\widehat{\beta}_l\widehat{\delta}_{kl}\varphi_{kl} = \widehat{\sigma^2}h^2$ (where $\boldsymbol{\Sigma}_{Sa} = h^2\mathbf{P}$), exactly or approximately, and then

$$\begin{array}{rcl} \widehat{\boldsymbol{\beta}}_S^{\mathsf{T}}\widehat{\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\mathsf{T}}\widehat{\mathbf{X}}_U\right)\widehat{\boldsymbol{\beta}}_S & = & \sigma^2\boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_U^{\mathsf{T}}\boldsymbol{\iota}_U \\ & = & \sigma^2h^2\boldsymbol{\iota}_U^{\mathsf{T}}\mathbf{Z}_U(\mathbf{Z}_{Sa}^{\mathsf{T}}\mathbf{P}^{-1}\mathbf{Z}_{Sa})^{-1}\mathbf{Z}_U^{\mathsf{T}}\boldsymbol{\iota}_U. \end{array} \tag{B.17}$$

In the hierarchical approach for the given realization of $\mathbf{X}_{Sa}$ and $\mathbf{Z}_{Sa}$ we have

$$\mathbf{X}_{Sa}\boldsymbol{\beta} = \mathbf{Z}_{Sa}\boldsymbol{\alpha} + \mathbf{u}_{Sa}. \tag{B.18}$$

We insert the second stage model (B.3) of the two-stage approach,

$$\mathbf{X}_{Sa} = \mathbf{Z}_{Sa}\boldsymbol{\Gamma} + \mathbf{D}_{Sa} \tag{B.19}$$

to get

$$\mathbf{X}_{Sa}\boldsymbol{\beta} = \mathbf{Z}_{Sa}\boldsymbol{\Gamma}\boldsymbol{\beta} + \mathbf{D}_{Sa}\boldsymbol{\beta}. \tag{B.20}$$

We identify the fixed and random parts of expression (B.20) and see that

$$\mathbf{D}_{Sa}\boldsymbol{\beta} = \mathbf{u}_{Sa}. \tag{B.21}$$

Written in a clear way, we have (omitting the index *Sa*)

$$\beta_1\mathbf{d}_1 + \beta_2\mathbf{d}_2 + ... + \beta_{(p+1)}\mathbf{d}_{(p+1)} = \mathbf{u}_{Sa}. \tag{B.22}$$

By taking the expectation of the product of the two sides of (B.22) with their transposes we obtain

$$\begin{aligned}\sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\beta_k\beta_l\,\mathrm{Cov}\left(\mathbf{d}_k,\mathbf{d}_l\right) &= \mathrm{E}\left[\mathbf{u}_{Sa}\mathbf{u}_{Sa}^{\intercal}\right]\\ \sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\beta_k\beta_l\delta_{kl}\boldsymbol{\Phi}_{kl} &= \sigma^2\boldsymbol{\Sigma}_{Sa}\end{aligned}$$

Recalling that $\boldsymbol{\Sigma}_{Sa} = h^2\mathbf{P}$ and $\boldsymbol{\Phi}_{kl} = \varphi_{kl}\mathbf{P}$, we have

$$\sum_{k=1}^{(p+1)}\sum_{l=1}^{(p+1)}\beta_k\beta_l\delta_{kl}\varphi_{kl}\mathbf{P} = \sigma^2h^2\mathbf{P}.$$

and, thus, by replacing expected values with their estimators,

$$\begin{aligned}\widehat{\boldsymbol{\beta}_S^{\intercal}\mathrm{Cov}}\left(\boldsymbol{\iota}_U^{\intercal}\widehat{\mathbf{X}}_U\right)\widehat{\boldsymbol{\beta}}_S &= \widehat{\sigma^2}h^2\mathbf{Z}_U\left(\mathbf{Z}_{Sa}^{\intercal}\mathbf{P}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_U^{\intercal}\boldsymbol{\iota}_U\\ &= \widehat{\sigma^2}\mathbf{Z}_U\left(\mathbf{Z}_{Sa}^{\intercal}\boldsymbol{\Sigma}_{Sa}^{-1}\mathbf{Z}_{Sa}\right)^{-1}\mathbf{Z}_U^{\intercal}\boldsymbol{\iota}_U\end{aligned} \tag{B.23}$$

what is exactly the first term of the $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GHMB}$.

Numerical examples (with simulated data) have shown that the two first terms of the variances give identical sums. Further, the size of the third term of $\widehat{\mathrm{V}}\left(\widehat{\mu}\right)_{GTSMB}$ has been shown to be much smaller that the other two (it contributed with about 0.5% of the variance). The third term could be seen as a second order correction of the first two.

Appendix C. Field and LiDAR Data Collection Methods

LiDAR data were collected using a Riegl LMS-Q680i system at five sites (SC, PANJ, ME, MN, CO) in the summer of 2014 and a sixth site (OR) in the summer of 2015 (see Figure 2 for locations and Table 1 for site descriptions). All data were collected during snow-free time periods when vegetation was fully leaf-on and non-senescent. The data was collected in north-south oriented flight lines spaced 5 km apart. Some crossing flights were flown for calibration purposes. Technical specifications for the LiDAR acquisition included: point density of 4 pulses per square meter; altitude of 732 m; field of view of 60 degrees; pulse rate of 330 kHz; nominal swath width of 812 m; horizontal and vertical accuracy (root mean square error in z) of 50 cm [33]. The LiDAR data were processed in Fusion [43] to produce rasters of various metrics with a 30 meter cell size to support the plot selection process described below.

Approximately fifty field inventory plots were established at each area in the summer of 2015 [44]. To ensure that field plots represented the full range of biomass levels within forested areas of each scene, the LiDAR data was used to inform the selection of field plot locations via a stratified sampling procedure. After masking out non-forest areas within the LiDAR coverage using a LiDAR-based

percent cover threshold (10%) and the National Land Cover Dataset [38], fifteen strata were delineated based upon three vegetation cover and five height classes within the random cells. A minimum of 23-24 candidate plot locations were generated for each of the 15 identified strata. Final field plots were selected from these candidate plots on the basis of accessibility and valid forest land cover status. Once in the field, crews aimed to visit 50 sample plots at each study area, with 3-4 plots in each stratum.

At each plot location, live and dead trees with dbh > 12.7 cm were measured on a 16.2-meter radius plot, and trees with 2.54 cm < dbh < 12.7 cm were measured on a 4.57-m radius circular plot. Field protocols consistent with those used by the US Forest Service were used in measuring trees [39]. In addition, survey-grade GPS coordinates (< 1 meter error) were acquired for each plot center.

Large-footprint LiDAR waveforms similar to those expected from GEDI were simulated from the acquired lidar data using the method presented in [32], in which waveforms are modeled as the sum of individual returns from surfaces at different heights, accounting for instrument-specific properties. Realistic noise was added to the simulated waveforms following [45] and [46]. The expected signal to noise ratio (SNR) of GEDI signals has been predicted through link margin analysis. For a given SNR, the probability of the ground elevation being correctly identified can be calculated and used to quantify the expected measurement accuracy. The simulator has been validated against real LVIS data (airborne large-footprint, full-waveform liDAR similar to GEDI; [47]) in terms of waveform metrics and metric accuracy in the presence of noise.

References

1. UNFCCC. United Nations Framework Convention on Climate Change. Available online: http://unfccc.int/resource/convkp/kpeng.html (accessed on 16 November 2018).
2. Europe, F.; Unece, F. State of Europe's forests 2011. Available online: https://library.wmo.int/index.php?lvl=notice_display&id=5268#.W-6ixehKiUl (accessed on 16 November 2018).
3. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment* **2012**, *121*, 196–209. doi:10.1016/j.rse.2012.02.001.
4. McRoberts, R.E.; Wendt, D.G.; Nelson, M.D.; Hansen, M.H. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment* **2002**, *81*, 36–44. doi:10.1016/S0034-4257(01)00330-3.
5. Saarela, S.; Grafström, A.; Ståhl, G.; Kangas, A.; Holopainen, M.; Tuominen, S.; Nordkvist, K.; Hyyppä, J. Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. *Remote Sensing of Environment* **2015**, *158*, 431–440. doi:10.1016/j.rse.2014.11.020.
6. Grafström, A.; Schnell, S.; Saarela, S.; Hubbell, S.; Condit, R. The continuous population approach to forest inventories and use of information in the design. *Environmetrics* **2017**, *28*. doi:10.1002/env.2480.
7. Matérn, B. Spatial Variation: Stochastic models and their applictation to some problems in forest surveys and other sampling investiagtions. *Meddelanden från Statens skogsforskningsinstitut* **1960**, *49*, 144.
8. Gregoire, T.G. Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research* **1998**, *28*, 1429–1447. doi:10.1139/x98-166.
9. McRoberts, R.E.; Magnussen, S.; Tomppo, E.O.; Chirici, G. Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. *Remote Sensing of Environment* **2011**, *115*, 3165–3174. doi:10.1016/j.rse.2011.07.002.
10. Särndal, C.E.; Swensson, B.; Wretman, J.H. *Model Assisted Survey Sampling*; Springer, New York, NY, USA, 1992; p. 716.
11. Gregoire, T.G.; Ståhl, G.; Næsset, E.; Gobakken, T.; Nelson, R.; Holm, S. Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research* **2011**, *41*, 83–95. doi:10.1139/X10-195.
12. Massey, A.; Mandallaz, D.; Lanz, A. 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. *Canadian Journal of Forest Research* **2014**, *44*, 1177–1186. doi:10.1139/cjfr-2014-0152.

13. Chirici, G.; McRoberts, R.E.; Fattorini, L.; Mura, M.; Marchetti, M. Comparing echo-based and canopy height model-based metrics for enhancing estimation of forest aboveground biomass in a model-assisted framework. *Remote Sensing of Environment* **2016**, *174*, 1–9. doi:10.1016/j.rse.2015.11.010.
14. Boudreau, J.; Nelson, R.F.; Margolis, H.A.; Beaudoin, A.; Guindon, L.; Kimes, D.S. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment* **2008**, *112*, 3876–3890. doi:10.1016/j.rse.2008.06.003.
15. Nelson, R.; Boudreau, J.; Gregoire, T.G.; Margolis, H.; Næsset, E.; Gobakken, T.; Ståhl, G. Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research* **2009**, *39*, 862–881. doi:10.1139/X09-002.
16. Neigh, C.S.; Nelson, R.F.; Ranson, K.J.; Margolis, H.A.; Montesano, P.M.; Sun, G.; Kharuk, V.; Næsset, E.; Wulder, M.A.; Andersen, H.E. Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR. *Remote Sensing of Environment* **2013**, *137*, 274–287. doi:10.1016/j.rse.2013.06.019.
17. Saarela, S.; Holm, S.; Grafström, A.; Schnell, S.; Næsset, E.; Gregoire, T.G.; Nelson, R.F.; Ståhl, G. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science* **2016**, *73*, 895–910. doi:10.1007/s13595-016-0590-1.
18. Holm, S.; Nelson, R.; Ståhl, G. Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne lidar, and space lidar. *Remote Sensing of Environment* **2017**, *197*, 85–97. doi:10.1016/j.rse.2017.04.004.
19. Gobakken, T.; Næsset, E.; Nelson, R.; Bollandsås, O.M.; Gregoire, T.G.; Ståhl, G.; Holm, S.; Ørka, H.O.; Astrup, R. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment* **2012**, *123*, 443 – 456. doi:10.1016/j.rse.2012.01.025.
20. Puliti, S.; Saarela, S.; Gobakken, T.; Ståhl, G.; Næsset, E. Combining UAV and Sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. *Remote Sensing of Environment* **2018**, *204*, 485–497. doi:10.1016/j.rse.2017.10.007.
21. Dubayah, R.; Goetz, S.; Blair, J.B.; Fatoyinbo, T.; Hansen, M.; Healey, S.P.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; others. The Global Ecosystem Dynamics Investigation. http://adsabs.harvard.edu/abs/2014AGUFM.U14A..07D, 2014.
22. Tomppo, E.; Gschwantner, T.; Lawrence, M.; McRoberts, R.; Gabler, K.; Schadauer, K.; Vidal, C.; Lanz, A.; Ståhl, G.; Cienciala, E.; others. *National Forest Inventories. Pathways for Common Reporting*; Springer, Berlin/Heidelberg, Germany. 2010; p. 614.
23. Saarela, S.; Schnell, S.; Grafström, A.; Tuominen, S.; Nordkvist, K.; Hyyppä, J.; Kangas, A.; Ståhl, G. Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research* **2015**, *45*, 1524–1534. doi:10.1139/cjfr-2015-0077.
24. Ståhl, G.; Holm, S.; Gregoire, T.G.; Gobakken, T.; Næsset, E.; Nelson, R. Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research* **2011**, *41*, 96–107. doi:10.1139/X10-161.
25. McRoberts, R.E. A model-based approach to estimating forest area. *Remote Sensing of Environment* **2006**, *103*, 56–66. doi:10.1016/j.rse.2006.03.005.
26. Cassel, C.M.; Särndal, C.E.; Wretman, J.H. *Foundations of inference in survey sampling*; Wiley: Hoboken, NJ, USA. 1977; p. 192.
27. Ståhl, G.; Saarela, S.; Schnell, S.; Holm, S.; Breidenbach, J.; Healey, S.P.; Patterson, P.L.; Magnussen, S.; Næsset, E.; McRoberts, R.E.; Gregoire, T.G. Use of models for improved estimation in sample-based large-area forest surveys: a review. *Forest Ecosystems* **2016**, *3(5)*, 1–11. doi:10.1186/s40663-016-0064-9.
28. Davidson, R.; MacKinnon, J.G. *Estimation and inference in econometrics*; Oxford University Press: Oxford, UK. 1993; p. 896.
29. Melville, G.; Welsh, A.; Stone, C. Improving the efficiency and precision of tree counts in pine plantations using airborne LiDAR data and flexible-radius plots: model-based and design-based approaches. *Journal of Agricultural, Biological, and Environmental Statistics* **2015**, *20*, 229–257. doi:10.1007/s13253-015-0205-6.
30. Saarela, S.; Holm, S.; Yang, Z. HMB: Hierarchical Model-Based estimation approach. Available online: https://CRAN.R-project.org/package=HMB (accessed on 16 November 2018). R package version 1.0.
31. Sanderson, C.; Curtin, R. Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software* **2016**, *1*, 1–26.

32. Blair, J.B.; Hofton, M.A. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical research letters* **1999**, *26*, 2509–2512. doi:10.1029/1999GL010484.
33. Andersen, H.E.; Cohen, W.B.; Yang, Z.; Healey, S.P.; others. Model-assisted estimation of carbon using Landsat and a designed sample of lidar data. *Environmental Research Letters* **2018**, *In Press*.
34. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters* **2006**, *3*, 68–72. doi:10.1109/LGRS.2005.857030.
35. Flood, N. Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). *Remote Sensing* **2013**, *5*, 6481–6500. doi:10.3390/rs5126481.
36. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment* **2015**, *159*, 269–277. doi:10.1016/j.rse.2014.12.014.
37. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27. doi:10.1016/j.rse.2017.06.031.
38. Homer, C.; Dewitz, J.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.; Coulston, J.; Herold, N.; Wickham, J.; Megown, K. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing* **2015**, *81*, 345–354.
39. CMS. Carbon Monitoring System (CMS) Field Guide 2015. *Pacific Northwest Research Station, USDA Forest Service,* Portland, OR, USA. **2015**, p. 82.
40. Ruefenacht, B.; Finco, M.; Nelson, M.; Czaplewski, R.; Helmer, E.; Blackard, J.; Holden, G.; Lister, A.; Salajanu, D.; Weyermann, D.; others. Conterminous US and Alaska forest type mapping using forest inventory and analysis data. *Photogrammetric Engineering & Remote Sensing* **2008**, *74*, 1379–1388. doi:10.14358/PERS.74.11.1379.
41. Cohen, W.B.; Healey, S.P.; Yang, Z.; Stehman, S.V.; Brewer, C.K.; Brooks, E.B.; Gorelick, N.; Huang, C.; Hughes, M.J.; Kennedy, R.E.; others. How similar are forest disturbance maps derived from different Landsat time series algorithms? *Forests* **2017**, *8*, 98. doi:10.3390/f8040098.
42. McRoberts, R.E. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment* **2010**, *114*, 1017–1025. doi:10.1016/j.rse.2009.12.013.
43. McGaughey, R. FUSION/LDV: Software for LIDAR Data Analysis and Visualization, Version 3.01. Available online: http://forsys.cfr.washington.edu/fusion/fusionlatest.html (accessed on 24 August 2012).
44. Legner, K.; Andersen, H.E.; Dobelbower, K.; Cooke, A.; Cohen, W.; Healey, S.P. A cost-effective field measurement protocol to support carbon monitoring – Implementing a prototype design at six different US sites (SC, NJ/PA, ME, MN, CO, OR) . *Gen. Tech. Rep. PNW-GTR-XXX. Portland, OR* **2018**, *In Press*.
45. Davidson, F.M.; Sun, X. Gaussian approximation versus nearly exact performance analysis of optical communication systems with PPM signaling and APD receivers. *IEEE Transactions on Communications* **1988**, *36*, 1185–1192. doi:10.1109/26.8924.
46. Hancock, S.; Disney, M.; Muller, J.P.; Lewis, P.; Foster, M. A threshold insensitive method for locating the forest canopy top with waveform lidar. *Remote Sensing of Environment* **2011**, *115*, 3286–3297. doi:10.1016/j.rse.2011.07.012.
47. Blair, J.B.; Rabine, D.L.; Hofton, M.A. The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing* **1999**, *54*, 115–122. doi:10.1016/S0924-2716(99)00002-7.