



Contents lists available at ScienceDirect

# International Journal of Applied Earth Observation and Geoinformation

journal homepage: [www.elsevier.com/locate/jag](http://www.elsevier.com/locate/jag)

## Leveraging remotely sensed non-wall-to-wall data for wall-to-wall upscaling in forest inventory

Fangting Chen<sup>a,b</sup>, Zhengyang Hou<sup>a,b,\*</sup>, Svetlana Saarela<sup>c</sup>, Ronald E. McRoberts<sup>d</sup>, Göran Ståhl<sup>e</sup>, Annika Kangas<sup>f</sup>, Petteri Packalen<sup>g</sup>, Bo Li<sup>h</sup>, Qing Xu<sup>i,1</sup>

<sup>a</sup> The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China

<sup>b</sup> Ecological Observation and Research Station of Heilongjiang Sanjiang Plain Wetlands, National Forestry and Grassland Administration, Shuangyashan 518000, China

<sup>c</sup> Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Postboks 5003, NMBU, 1432 Ås, Norway

<sup>d</sup> Department of Forest Resources, University of Minnesota, St. Paul, MN, United States

<sup>e</sup> Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

<sup>f</sup> Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Yliopistokatu 6, 80100 Joensuu, Finland

<sup>g</sup> Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Latokartanonkaari 9, FI-00790 Helsinki, Finland

<sup>h</sup> Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, United States

<sup>i</sup> Key Laboratory of National Forestry and Grassland Administration/Beijing for Bamboo & Rattan Science and Technology, International Centre for Bamboo and Rattan, Beijing 100102, China

### ARTICLE INFO

#### Keywords:

Non-wall-to-wall  
Wall-to-wall  
Model-based inference  
Two-stage modeling  
Survey sampling

### ABSTRACT

Remote sensing (RS) has enhanced forest inventory with model-based inference, that is, a family of statistical procedures rigorously estimates the parameter of a variable of interest (VOI) for a spatial population, e.g., the mean or total of forest carbon for a study area. Upscaling in earth observation, alias to this estimation, aggregates VOI from a finer spatial resolution to a coarser one with reduced uncertainty, serving decision making for natural resource management at larger scales. However, conventional model-based estimation (CMB) confronts a major challenge: it only supports RS wall-to-wall data, meaning that remotely sensed data must be available in panorama and non-wall-to-wall but quality data such as lidar or even cloud-masked satellite imagery are not supported due to incomplete coverage, impeding precise upscaling with cutting-edge instruments or for large scale applications. Consequently, this study aims to develop and demonstrate the use and usefulness of RS non-wall-to-wall data for upscaling with Hierarchical model-based estimation (HMB) which incorporates a two-stage model for bridging RS non- and wall-to-wall data; and for optimizing cost-efficiency, to evaluate the effects of non-wall-to-wall sample size on upscaling precision. Three main conclusions are relevant: (1) the HMB is a variant of the CMB estimator through trading in the uncertainty of the second-stage model to enable estimation using RS non-wall-to-wall data; (2) a quality first-stage model is key to exerting the advantage of HMB relative to the CMB estimator; (3) the variance of the HMB estimator is dominated by the first-stage model variance component, indicating that increasing the sample size in the first-stage is effective for increasing the overall precision. Overall, the HMB estimator balances tradeoffs between cost, efficiency and flexibility when devising a model-based upscaling in earth observation.

### 1. Introduction

Upscaling in earth observation serves natural resource management at larger scales through aggregation of variable of interest (VOI) from a finer spatial resolution to a much coarser one (Masek et al., 2015). Preservation of information integrity is essential for such upscaling

process, which in statistics corresponds to estimating the parameter of a VOI for a spatial population, e.g., the mean or total of forest carbon for a study area, in a way as precise as possible (Bazezew et al., 2018). While National Forest Inventory (NFI) programs in several countries provide precise estimates every five years for bio- and abiotic VOIs using upscaling procedures categorized as design-based inference, it is

\* Corresponding author at: The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China.  
E-mail addresses: [houshengyang@bjfu.edu.cn](mailto:houshengyang@bjfu.edu.cn) (Z. Hou), [qing.xu@icbr.ac.cn](mailto:qing.xu@icbr.ac.cn) (Q. Xu).

<sup>1</sup> Co-corresponding author.

<https://doi.org/10.1016/j.jag.2023.103314>

Received 15 February 2023; Received in revised form 1 April 2023; Accepted 16 April 2023

Available online 23 April 2023

1569-8432/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

difficult and expensive for programs of this sort to meet precision standards for annual reporting required by, for example, the United Nations Framework Convention on Climate Change (Hou et al., 2021; Vidal et al., 2016).

Alternative to the design-based inference, model-based inference, i.e., a family of statistical procedures capable of leveraging remote sensing (RS) for upscaling or estimation, has enhanced forest inventory with increased cost-efficiency (Bayat et al., 2021; Hou et al., 2018). Assisted with a variety of terrestrial, airborne, or spaceborne sources of RS auxiliary data, a conventional model-based estimator (CMB) is more precise than a design-based estimator using the same sample, or equivalently, CMB estimator achieves a similar precision but using a smaller sample (McRoberts et al., 2018). This is because CMB estimator works in a way by assuming a superpopulation model where RS auxiliaries are used as independent variables towards modeling a surrogate for the superpopulation model enroute to estimation (Chambers and Clark, 2012; Ståhl et al., 2016).

However, CMB confronts three major challenges. First, field campaigns to acquire reference data are nontrivial and expensive, but in most cases they are required for model construction. CMB relies on modeling a linear or nonlinear relationship between the sampled unit observations of the VOI and observed RS auxiliary variables (McRoberts et al., 2014). Second, a model constructed previously, i.e., an existing model, is usually overlooked for reusing (e.g., Nord-Larsen and Cao, 2006). When such a model is readily available, costs associated with field sampling and acquisition of RS data can be avoided, particularly for continuously monitored areas, e.g., forest compartments under forest management, Amazonian rain forests under ecological monitoring, and African tropical forests under biodiversity protection. These existing models are often constructed with terrestrial, airborne, or UAV lidar, or with very high-resolution satellite imagery (e.g., Heiskanen et al., 2019; Lu et al., 2020). Third, CMB requires RS coverage for the entire area, i.e., wall-to-wall auxiliary data, which is not always available or even feasible in large-scale inventories (Chirici et al., 2020). Although fine resolution satellite or lidar data are detailed, this type of existing models cannot be applied to CMB in large-scale inventories due to the cost and non-wall-to-wall coverage (Xu et al., 2023). However, non-wall-to-wall auxiliary data could be the only option in many applications, for example, upscaling the VOIs of understory vegetation with terrestrial lidar (Li et al., 2021). Hence, alternatives that overcome these challenges must be sought.

Hierarchical model-based estimation (HMB) is a viable option that harnesses existing model and non-wall-to-wall data in tandem. Proposed by Saarela et al. (2018, 2020), the HMB estimator is a model-based procedure that makes use of non-wall-to-wall data for upscaling forest biomass. Unlike the CMB estimator, which requires one source of wall-to-wall RS auxiliary data, the HMB estimator utilizes two sources of RS auxiliary data, one of greater quality but non-wall-to-wall, and one source of lesser quality but wall-to-wall. Two-stage modeling, which is a key component of the HMB estimator, combines them in a way that the first-stage model is constructed with field-observed VOI and non-wall-to-wall RS data, and then the resulting predicted values are used as reference data for fitting the second-stage model with the wall-to-wall RS data. The HMB estimator takes the uncertainty propagation from both the first- and the second-stage models into account. Because of the two-stage construction, the first-stage model can be replaced with an existing model to avoid the cost of field and RS data acquisition for model training. However, while the HMB estimator is more flexible than the CMB estimator, the inferential precision of the HMB estimator as the sample size of the non-wall-to-wall RS data increases, is still unclear.

Consequently, the objectives of this study are threefold: (1) to develop and demonstrate the use and usefulness of RS non-wall-to-wall data for upscaling with HMB; (2) to compare the wall-to-wall CMB estimator and the non-wall-to-wall HMB estimator; and (3) for optimizing cost-efficiency, to evaluate the effects of non-wall-to-wall sample size on upscaling precision.

## 2. Materials

### 2.1. Field data

The study area, i.e., the target population, has approximately 10,836 ha located in Kou, Burkina Faso with a fragmented landscape of dry savanna due to agricultural land uses (11°45'N, 1°57'W) (Fig. 1). The plot-level VOI in this study is firewood volume in m<sup>3</sup>/ha, aggregated from within-plot woody material usable as fuelwood. A sample of 160 circular plots selected by two-stage sampling was field surveyed during the dry season lasted from late November 2013 through early February 2014, with sample statistics listed in Table 1. The plots have a radius of 17.84 m with centers geo-referenced using Global Navigation Satellite System receivers that have a real-time precision of 60 cm based on free corrections from Satellite-Based Augmentation Systems supported by European Geostationary Navigation Overlay Service.

### 2.2. Remotely sensed auxiliaries

RapidEye and Landsat 8 provided wall-to-wall auxiliary data georeferenced to WGS84/UTM Zone 30 N with acquisition time matching the dry season when the field campaign was carried out. A single scene of respective sensors covered the target population. At a cost of roughly 1.3 USD/km<sup>2</sup>, RapidEye data were processed to Level 3A with radiometric, sensor and geometric corrections at a spatial resolution of 5 m. Landsat 8 data were the Provisional Surface Reflectance product that has a spatial resolution of 30 m, downloaded from the USGS at no cost. RS auxiliary variables such as the Enhanced Vegetation Index (EVI), Generalized Difference Vegetation Index (GDVI), the first principal component (PCA) and Haralick textures were calculated for respective datasets as detailed in Hou et al. (2018). The R-package “rgdal” was used in data processing (Bivand et al., 2013).

## 3. Methods

### 3.1. Overview

A flowchart for the study is provided in Fig. 2, including the following cases for CMB and HMB estimators:

Case A. The CMB estimator.

Case A.1. Using wall-to-wall Landsat 8 auxiliary data.

Case A.2. Using wall-to-wall RapidEye auxiliary data.

Case B. The HMB estimator using a sample of RapidEye data (non-wall-to-wall) and wall-to-wall Landsat 8 data.

All cases follow three fundamental assumptions: (i) the population,  $U$ , consists of  $N$  units with wall-to-wall Landsat 8 auxiliary data available where a unit is equivalent to a grid cell, i.e. a pixel, of size 30 by 30 m; (ii) a sample,  $S_a$ , selected from  $U$ , contains  $M$  units, for which each sample unit has both Landsat 8 and RapidEye auxiliary data available; (iii) a sample  $S$  containing  $n$  units for which each sample unit has the field observed firewood volume and the auxiliary data for both Landsat 8 and RapidEye. In this study, the population size is  $N = 120756$ , and the sample size for  $S$  is  $n = 160$ . For evaluating the effects of sample size for RS auxiliary data on the HMB estimator, samples,  $S_a$ , of size  $M \in (160, 320, 1600, 3200, 6400, 12800)$  was selected randomly without replacement from  $U$ . With the HMB estimator, RapidEye represents the greater-quality RS sample data, available non-wall-to-wall. Other examples of such data could be airborne, terrestrials lidar or UAV imagery.

For convenience, we itemize readily available tools for this section: (1) R-package “bootStepAIC” (Rizopoulos, 2022) for selecting independent variables; (2) R-package “nlme” (Pinheiro et al., 2016) for estimating model parameters; (3) R-package “nlme” (Pinheiro and Bates, 2000) for estimating parameters of residual variance; (4) R-package “HMB” (<https://CRAN.R-project.org/package=HMB>) for HMB estimator.

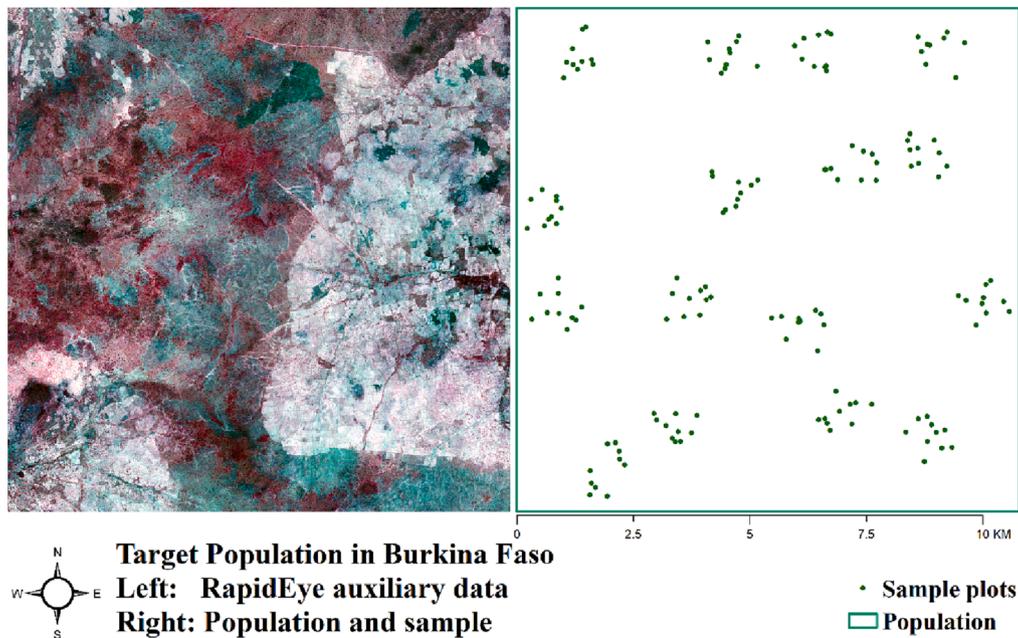


Fig. 1. Target population in Burkina Faso.

Table 1  
Sample statistics.

Forest attributes	Min	Max	Mean	SD
Tree density (stems/ha)	10	1935	494	401
Mean diameter (cm)	6.4	40	15.2	8.5
Basal area (m <sup>2</sup> /ha)	0.2	16.1	5.6	3.6
Firewood volume (m <sup>3</sup> /ha)	0	29.1	6.6	6.2

### 3.2. Conventional model-based (CMB) estimator

The CMB estimator supports upscaling with RS wall-to-wall auxiliary data. It uses the sample  $S$  for modeling and then plugs wall-to-wall auxiliary variables into the constructed model for population parameter prediction (i.e., upscaling). If there is an existing model, the CMB estimator directly applies given that the wall-to-wall auxiliary variables used by this model are available. However, for lidar instruments, for example, new acquisitions of timely wall-to-wall auxiliary variables are

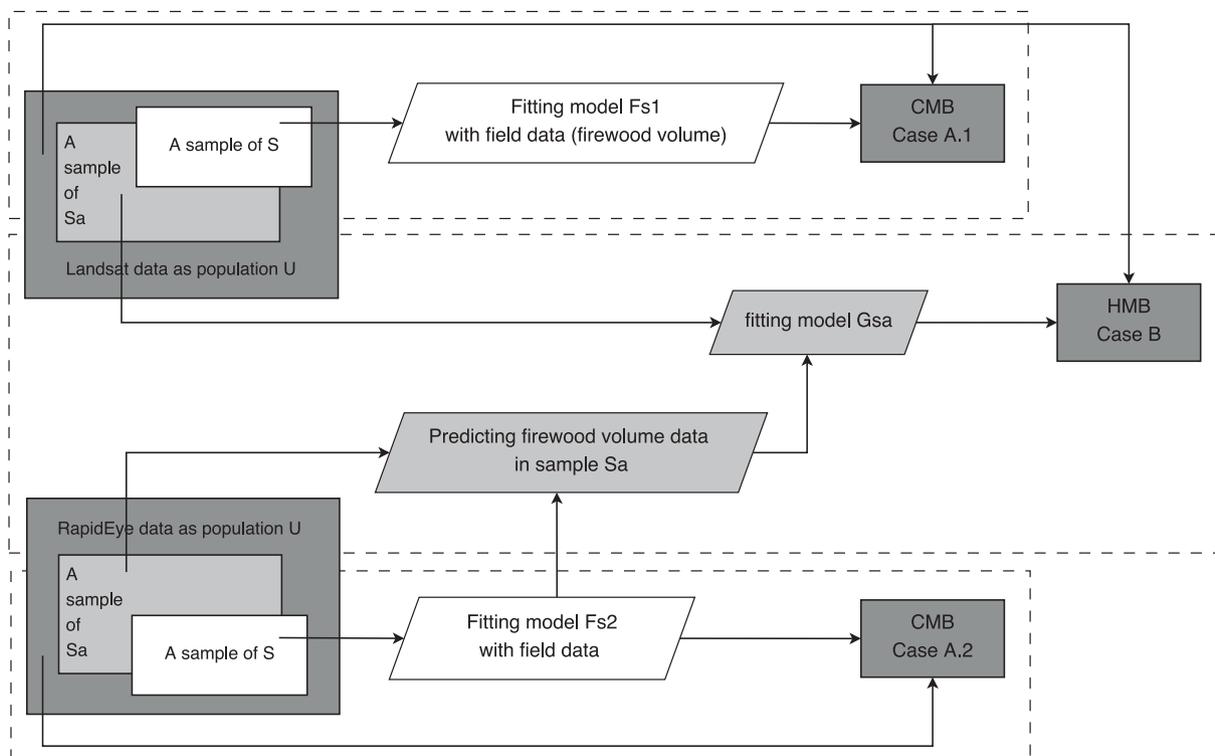


Fig. 2. Flowchart of the study cases.  $S$ , first-stage sample;  $S_a$ , second-stage sample;  $U$ , population; model  $F_{s1}$ , first-stage model using field data and Landsat 8 data; model  $F_{s2}$ , first-stage model using field data and RapidEye data; model  $G_{sa}$ , second-stage model using RapidEye-predicted data and Landsat 8 data.

hardly practical, thus restraining these cutting-edge instruments for large-scale inventories. Regardless, this is feasible with satellite data which, albeit coarser than lidar data, have timely and full coverages to the target population, and thus readily suit the CMB estimator.

The model is key to upscaling with the CMB estimator. For cases A.1 and A.2, the model is denoted as  $F_{s1}$  for Case A.1 and  $F_{s2}$  for Case A.2, and has the general form,  $y_i = f(x_i; \beta) + \varepsilon_i$ , which links the dependent variable  $y_i$  with a vector of RS auxiliary variables  $x_i$ , where  $i$  indexes population units;  $\beta$  is a vector of model parameters; and  $\varepsilon_i$  is a random error term following  $N(0, \sigma_\varepsilon^2)$ . The mean of the distribution of  $y$  for the  $i^{\text{th}}$  population unit is denoted  $y_{Fi}$ , estimated with  $\hat{y}_{Fi} = f(x_i; \hat{\beta}_S)$ , i.e.,  $\hat{y}_F = f(X_U; \hat{\beta}_S)$  in matrix form. Section 3.4 provides details about the estimation of model parameters.

With the fitted model, the population mean is estimated with a collection of estimates  $\hat{y}_{Fi}$  for every unit of the population (i.e., wall-to-wall),

$$\hat{\mu}_1 = \frac{1}{N} \sum_{i=1}^N \hat{y}_{Fi} \quad (1)$$

The variance estimator of  $\hat{\mu}_1$  takes the form (e.g., McRoberts et al., 2013)

$$\widehat{Var}(\hat{\mu}_1) = I_U^T \tilde{X}_U \widehat{Cov}(\hat{\beta}_S) \tilde{X}_U^T I_U \quad (2)$$

with

$$\widehat{Cov}(\hat{\beta}_S) = \left( \tilde{X}_S^T \hat{\Omega}_S^{-1} \tilde{X}_S \right)^{-1} \quad (3)$$

where  $\tilde{X}_U = \frac{\partial f(X_U; \hat{\beta}_S)}{\partial \beta_S}$  is a partial derivative matrix of  $f(X_U; \hat{\beta}_S)$  with respect to  $\hat{\beta}_S$  for population  $U$ , and  $I_U$  is an  $N$ -length vector with each element being  $1/N$ ;  $\widehat{Cov}(\hat{\beta}_S)$  is the estimated covariance matrix for  $\hat{\beta}_S$ , with  $\tilde{X}_S$  being a partial derivative matrix of  $f(X_U; \hat{\beta}_S)$  with respect to  $\hat{\beta}_S$  for the sample  $S$ ;  $\hat{\Omega}_S$  is a residual variance–covariance matrix for the model  $F_{s1}$  or  $F_{s2}$ , with diagonal elements estimated using an exponential variance function detailed in Section 3.4. Because  $\widehat{Var}(\hat{\mu}_1)$  was derived with the delta method, using first-order Taylor expansion, and thus is an approximation for nonlinear models (Gregoire et al., 2016). Note that  $\hat{\mu}_1$  and  $\widehat{Var}(\hat{\mu}_1)$  are general expressions for the CMB estimator, working for both A.1 and A.2 by using the corresponding set of  $\hat{y}_{Fi}$ ,  $\tilde{X}_U$ ,  $\tilde{X}_S$  and  $\hat{\Omega}_S$ .

### 3.3. Hierarchical model-based (HMB) estimator

The HMB estimator supports upscaling with RS non-wall-to-wall auxiliary data. It was derived from the general form of the CMB estimator, where the modeling is arranged in a hierarchical structure of two stages. The resulting HMB estimator takes the error-propagation across the hierarchy into account, and thus enables uncertainty assessment combining wall-to-wall and non-wall-to-wall data for large-scale inventories, greatly increasing the utility of cutting-edge remote sensing instruments. (e.g., Wang et al., 2020).

In case B, the first-stage model based on RapidEye,  $F_{s2}$ , is linked with the second-stage model based on Landsat 8,  $G_{sa}$ , in a way that model  $F_{s2}$  predicts, with a newly constructed or existing model, a set of non-wall-to-wall firewood volumes that are used for modeling  $G_{sa}$  with the sample  $Sa$ . The final upscaling is based on model  $G_{sa}$ , with uncertainties associated with model parameter estimation and sample size at each stage. The general form of model  $G_{sa}$  conforms to models  $F_{s1}$  and  $F_{s2}$  of the CMB estimator (Section 3.2), with modeling details described in Section 3.4.

The HMB estimator of the population mean is

$$\hat{\mu}_2 = \frac{1}{N} \sum_{i=1}^N \hat{y}_{Gi} \quad (4)$$

where  $\hat{y}_{Gi}$  is the predicted dependent variable based on the second-stage model  $G_{sa}$  for the  $i^{\text{th}}$  population unit. This is with respect to the population  $U$ , and thus requires RS wall-to-wall auxiliary variables, i.e., Landsat 8.

The variance estimator of  $\hat{\mu}_2$  takes the form

$$\widehat{Var}(\hat{\mu}_2) = I_U^T \tilde{Z}_U \widehat{Cov}(\hat{\alpha}_{Sa}) \tilde{Z}_U^T I_U \quad (5)$$

where  $\tilde{Z}_U = \frac{\partial f(Z_U; \hat{\alpha}_{Sa})}{\partial \alpha_{Sa}}$  is a partial derivative matrix of  $f(Z_U; \hat{\alpha}_{Sa})$  with respect to  $\hat{\alpha}_{Sa}$  in the population  $U$ ;  $\widehat{Cov}(\hat{\alpha}_{Sa})$  is the estimated variance–covariance matrix for  $\hat{\alpha}_{Sa}$ , and

$$\begin{aligned} \widehat{Cov}(\hat{\alpha}_{Sa}) = & \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \\ & + \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{X}_{Sa} \widehat{Cov}(\hat{\beta}_S) \tilde{X}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \end{aligned} \quad (6)$$

The law of total covariance was used for deriving Eq. (6),  $\widehat{Cov}(\hat{\alpha}_{Sa}) = E[\widehat{Cov}(\hat{\alpha}_{Sa} | \hat{y}_{Fsa})] + \widehat{Cov}(E[\hat{\alpha}_{Sa} | \hat{y}_{Fsa}])$  (Saarela et al., 2020). The first term can be expressed as the model-based covariance of estimated model parameters conditionally on the predicted firewood volume using model  $F_{s2}$ ,  $\hat{y}_{Fsa}$ , i.e.,  $E[\widehat{Cov}(\hat{\alpha}_{Sa} | \hat{y}_{Fsa})] = \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1}$ .  $\tilde{Z}_{Sa}$  is a partial derivative matrix of the fixed part of model  $G_{sa}$  with respect to  $\hat{\alpha}_{Sa}$  for sample  $Sa$ ; and  $\hat{\Sigma}_{Sa}^{-1}$  is the estimated variance–covariance matrix of model  $G_{sa}$  residuals detailed in Section 3.4. The second term estimates the propagated uncertainty stemming from  $\hat{y}_{Fsa}$ , i.e.,  $\widehat{Cov}(E[\hat{\alpha}_{Sa} | \hat{y}_{Fsa}]) = \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \widehat{Cov}(\hat{y}_{Fsa}) \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1}$ .

The variance–covariance matrix for  $\hat{\alpha}_{Sa}$  can be thus expressed as

$$\begin{aligned} \widehat{Cov}(\hat{\alpha}_{Sa}) = & \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \\ & + \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \widehat{Cov}(\hat{y}_{Fsa}) \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \end{aligned} \quad (7)$$

An approximately unbiased estimator of  $\widehat{Cov}(\hat{\alpha}_{Sa})$  can be obtained by substituting the estimated  $\widehat{Cov}(\hat{y}_{Fsa})$ , that is,

$$\begin{aligned} \widehat{Cov}(\hat{\alpha}_{Sa}) = & \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \\ & + \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \widehat{Cov}(\hat{y}_{Fsa}) \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \left( \tilde{Z}_{Sa}^T \hat{\Sigma}_{Sa}^{-1} \tilde{Z}_{Sa} \right)^{-1} \end{aligned} \quad (8)$$

where  $\widehat{Cov}(\hat{y}_{Fsa}) = \tilde{X}_{Sa} \widehat{Cov}(\hat{\beta}_S) \tilde{X}_{Sa}^T$ , with  $\tilde{X}_{Sa}$  being a partial derivative matrix of the fixed part of model  $F_{s2}$  with respect to  $\hat{\beta}_S$  for sample  $Sa$ ; and  $\widehat{Cov}(\hat{\beta}_S)$  being consistent with Eq. (3). Note the difference regarding subscripts  $S$ ,  $Sa$  and  $U$  that indicates non-wall-to-wall or wall-to-wall characteristics for a sample or population.

Therefore, the variance of the HMB estimator,  $\widehat{Var}(\hat{\mu}_2)$ , decomposes into two parts, with the first term estimating the variance associated with model  $G_{sa}$  at the second-stage, and the second term estimating variance associated with model  $F_{s2}$  at the first-stage and propagated to the second-stage.

### 3.4. Single- and two-stage modeling at the unit level

For modeling, the “bootstrap stepAIC” procedure was used for selecting independent variables parsimoniously from the large set of Landsat 8 or RapidEye auxiliary variables extracted, as described in

Section 2.3. This procedure integrates bootstrapping to assess the variability of stepwise model selections as per the Akaike information criterion, available in the R-package “bootStepAIC” (Rizopoulos, 2022).

As illustrated in Fig. 2, the single stage model  $F_{s1}$  describes the relationship between the VOI,  $y_S$  (i.e.  $y_S = [y_1, y_2, \dots, y_n]^T$ ), and the Landsat 8 independent variables,  $x_L$  (i.e.,  $x_L = [x'_1, \dots, x'_n]^T = [1, x_1, x_2, \dots, x_{p1}]$ ), at sample  $S$ , expressed as follows with  $\beta_L$  denoting a vector of model parameters to be estimated:

$$ModelF_{s1} : y_S = f(x_L; \beta_L) + e, e \sim N(0, \Omega_L) \tag{9}$$

The single stage model  $F_{s2}$  describes the relationship between the VOI,  $y_S$  (i.e.  $y_S = [y_1, y_2, \dots, y_n]^T$ ), and the RapidEye independent variables,  $x_R$  (i.e.,  $x_R = [x'_1, \dots, x'_n]^T = [1, x_1, x_2, \dots, x_{p2}]$ ) at sample  $S$ , expressed as follows with  $\beta_R$  denoting a vector of model parameters to be estimated:

$$ModelF_{s2} : y_S = f(x_R; \beta_R) + \epsilon, \epsilon \sim N(0, \Omega_R) \tag{10}$$

Following Davidson and MacKinnon (1993), the second-stage of the two-stage model  $G_{Sa}$  describes the relationship between the VOI,  $y_{Sa}$  (i.e.  $y_{Sa} = [y_1, y_2, \dots, y_M]^T$ ), and the Landsat 8 independent variables,  $z$  (i.e.  $z = [z'_1, \dots, z'_M]^T = [1, z_1, z_2, \dots, z_q]$ ) at sample  $Sa$ :

$$ModelG_{Sa} : y_{Sa} = g(z; \alpha) + v, v \sim N(0, \Sigma) \tag{11}$$

Where the  $y_{Sa}$ , is estimated using the first-stage model  $F_{s2}$ , using non-wall-to-wall RapidEye independent variables available in the sample  $Sa$ ; and  $\alpha$  is a vector of model parameters to be estimated. For generalizability, the role of model  $F_{s2}$  can be replaced with an existing model constructed previously, and the role of RapidEye data with airborne or UAV-borne lidar data which form the basis for tackling the challenges listed in the Introduction with HMB estimation.

In this study, models  $F_{s1}$ ,  $F_{s2}$  and  $G_{Sa}$  took nonlinear forms as  $y = \beta_{L0} \cdot x_{L1}^{\beta_{L1}} + e$ ,  $y = e^{(\beta_{R1} \cdot x_{R1} + \beta_{R2} \cdot x_{R2} + \beta_{R3} \cdot x_{R3})} + \epsilon$ , and  $y = \alpha_0 \cdot z_1^{\alpha_1} + v$  respectively, with one independent variable being selected for models  $F_{s1}$  and  $G_{Sa}$  and three independent variables for model  $F_{s2}$ .

Parameters of the three models were estimated with the restricted maximum likelihood estimation using “nlme” package in R (Pinheiro et al., 2016). The residual variance-covariance matrices for the models  $F_{s1}$ ,  $F_{s2}$  and  $G_{Sa}$ , i.e.,  $\Omega_L$  and  $\Omega_R$  corresponding to  $\Omega_S$  in Eq. (3) and  $\Sigma$  to  $\Sigma_{Sa}$  in Eq. (6), were estimated for on-diagonal elements using the exponential variance function structure conforming to the following form (e.g., Galecki and Burzykowski, 2013)

$$Var(\epsilon_i) = \sigma^2 \exp(2\delta v_i), \tag{12}$$

where  $\sigma^2$  and  $\delta$  are the parameters estimated using “varExp” function in R-package “nlme” (Pinheiro and Bates, 2000, p. 211); and  $v_i$  is the variance covariate defining the variance function for  $i^{th}$  population unit. In this study, we chose scalar-valued function of the independent variables of respective models as respective variance covariates  $v_i$  in that the results of which outperformed the most common choice i.e., the estimated  $E(y_i)$  (Mehtätalo and Lappi, 2020). Although spatial autocorrelation was not an issue in the present study, both CMB and HMB estimators in Sections 3.2 and 3.3 support its incorporation.

### 3.5. Evaluation criteria for modeling and upscaling

Root mean square error,  $RMSE$ , was used for evaluating the prediction accuracy of a fitted model.  $RMSE$  and its relative form on a percentage basis are respectively expressed as  $RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$ , and  $RMSE_{\%} = \frac{RMSE}{\bar{y}} \times 100$ , where  $n$  is the number of sample plots;  $y_i$  and  $\hat{y}_i$

refer to field measured and predicted firewood volume; and  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

As for upscaling or population prediction, we further evaluate the

precision using the sampling error, i.e.,  $SE_{\%} = 100 \times \frac{\sqrt{\widehat{Var}(\hat{\mu})}}{\hat{\mu}}$ . The smaller the  $SE_{\%}$ , the greater the precision. This measure jointly evaluates  $\hat{\mu}$  and  $\widehat{Var}(\hat{\mu})$  on a percentage basis, making it more intuitive to comparing estimators than using  $\widehat{Var}(\hat{\mu})$  alone.

## 4. Results and discussion

### 4.1. Models

The constructed models,  $F_{s1}$ ,  $F_{s2}$  and  $G_{Sa}$ , are summarized in Table 2. For Landsat 8, two first-stage models of different  $RMSEs$ , one using EVI and the other using GDVI, were constructed for evaluating the effects of different wall-to-wall auxiliaries on the HMB estimator. There was no model exhibiting a systematic lack of fit.

Three findings are relevant: (1) the  $RMSEs$  of the different models are generally similar, in the order of 45–66%, consistent with previous studies under tropical conditions using multispectral satellites (e.g., Hou et al., 2011, 2017; Næsset et al., 2016); (2) similar  $RMSEs$  indicate that wall-to-wall predictions for every unit in the population, i.e., mapping, are comparable for the CMB and HMB estimators; and (3) RapidEye outperforms Landsat 8 by 8% on average, i.e.,  $F_{s2}$  vs.  $F_{s1}$ , suggesting appreciable effects of using finer spatial resolution and within plot information, consistent with previous findings (e.g., Hou et al. 2018; Rahlf et al., 2014; Wittke et al., 2019).

In practice, any working existing model constructed with alternative RS auxiliary data can substitute for the role of model  $F_{s2}$  in Case B (Jayathunga et al., 2018; Chirici et al., 2020). Consequently, costs spent on the first-stage modeling including the field campaign and RS acquisitions at sample  $S$  would be circumvented. However, note that for selecting an existing nonlinear model that is highly wiggly, its approximated variance covariance matrix of model parameters may risk being underestimated, which is less problematic for linear models. Regardless, costs associated with applying the existing model with a sample of RS non-wall-to-wall auxiliary variables,  $Sa$ , are still required for predicting VOI values that will be used for fitting a second-stage model. Effects of  $Sa$  sample size are examined in Section 4.3.

### 4.2. Comparison of estimators

Comparison of the CMB and HMB estimator reveals the greater efficiency of the HMB estimator. The estimates resulting from the CMB

**Table 2**  
Summary of the models. A sample of 6400 observations predicted by the first-stage model  $F_{s2}$  were used to fit the second-stage model  $G_{Sa}$ , denoted with  $*6400$ .

Case	Model	RMSE	RMSE%	Independent variable	Estimate	SE
Case A1	Model $F_{s1}$	4.42	65.21	(Intercept)	4.79	0.36
				EVI	2.29	0.22
Case A1	Model $F_{s1}$	4.49	66.36	(Intercept)	17.96	1.62
				GDVI	5.87	0.60
Case A2	Model $F_{s2}$	4.09	60.35	PCA	0.31	0.03
				Textural mean of SR	13.14	0.80
				Textural variance of SR	-0.02	0.00
Case $B^{*6400}$	Model $G_{Sa}$	3.18	45.54	(Intercept)	6.01	0.06
Case $B^{*6400}$	Model $G_{Sa}$	3.25	46.68	EVI	1.40	0.03
				(Intercept)	12.42	0.12
				GDVI	3.07	0.05

**Table 3**

Summary of the estimation using CMB and HMB estimators. A sample of 6400 observations predicted by the first-stage model  $F_{s2}$  were used to fit the second-stage model  $G_{Sa}$ , denoted with  $*6400$ .

Case	Estimator	Model	$\hat{\mu}$	$\widehat{Var}(\hat{\mu})$	$SE_{\%}$
Case A1	CMB	Model $F_{s1}$ with EVI	6.584	0.119	5.24
Case A1	CMB	Model $F_{s1}$ with GDVI	6.624	0.125	5.34
Case A2	CMB	Model $F_{s2}$	6.955	0.098	4.50
Case B <sup>*6400</sup>	HMB	Model $G_{Sa}$ with EVI	6.923	0.099	4.54
Case B <sup>*6400</sup>	HMB	Model $G_{Sa}$ with GDVI	6.896	0.099	4.56

and HMB estimators are summarized in Table 3. From the upscaling perspective, these estimates are upscaled from the unit-level to the population-level. The advantage of using the CMB and HMB estimators for upscaling resides in its survey sampling nature that is statistically sound in terms of model-unbiasedness and uncertainty quantification in the form of precision (Chambers and Clark, 2012).

Three findings are relevant. First, the rate of error propagation is greater in inferring than modeling even though modeling underpins inference, resulting in numerous implications towards efficient inventory. From a comparison within Case A, the precision of Case A.2 is markedly greater than that of Case A.1, by 15% on average, while Case A.2 outperforms Case A.1 by 8% on average with respect to RMSE (Section 4.1), suggesting that (1) RMSE at the level of individual units cannot be used as a measure of the precision of the estimator of the population mean, where a  $\widehat{Var}(\hat{\mu})$  (or  $SE_{\%}$ ) should be used; and (2) adopt remotely sensed greater quality data as much as possible, which contributes to decreasing RMSE and  $\widehat{Var}(\hat{\mu})$  (or  $SE_{\%}$ ).

Second, the CMB estimator is more sensitive and reliant on correlation with wall-to-wall auxiliaries than the HMB estimator. Table 3 shows that the weaker the correlation between wall-to-wall auxiliary variables and VOI, the less the precision in the form of  $\widehat{Var}(\hat{\mu})$  or  $SE_{\%}$  for both estimators. However, with HMB, the loss in precision is negligible and embodied as almost the same  $\widehat{Var}(\hat{\mu})$  or  $SE_{\%}$  of EVI and GDVI models, the reason is associated with the share of uncertainty contributed from each of the stages as further elaborated in Section 4.4.

Third, the precision of the CMB and HMB estimators is comparable, with the latter outperforming the former due to the relatively more accurate first-stage model  $F_{s2}$  (Table 3). Taking NFI in China or the United States of America as reference, the  $SE_{\%}$  around 5% meets the official standard of the precision in respective countries (Tomppo et al., 2010; USDA, 2008), demonstrating the prospect of both estimators for improving annual reporting.

**Table 4**

Effects of the sample size of  $S_a$ , i.e., the data volume of remotely sensed non-wall-to-wall auxiliaries, on the precision of the HMB estimator with  $\widehat{Var}(\hat{\mu})$  decomposed for the first- and second-stage modeling.

Sample size of $S_a$	Sampling intensity of $S_a$ (%)	Wall-to-wall auxiliary variable	$\hat{\mu}$	$\widehat{Var}(\hat{\mu})$	$SE_{\%}$	Variance decomposition (%)	
						$F_{s2}$	$G_{Sa}$
160	0.13	EVI	6.793	0.142	5.55	70.42	29.58
		GDVI	6.842	0.146	5.58	67.81	32.19
320	0.26	EVI	6.891	0.132	5.27	72.27	27.73
		GDVI	6.907	0.134	5.30	70.90	29.10
1600	1.32	EVI	6.932	0.105	4.67	94.29	5.71
		GDVI	6.871	0.107	4.74	92.45	7.55
3200	2.65	EVI	6.923	0.101	4.59	97.03	2.97
		GDVI	6.862	0.101	4.62	95.83	4.17
6400	5.30	EVI	6.923	0.099	4.54	98.49	1.51
		GDVI	6.896	0.099	4.56	98.39	1.61
12,800	10.60	EVI	6.913	0.098	4.53	99.18	0.82
		GDVI	6.881	0.098	4.55	98.98	1.02

4.3. Effects of non-wall-to-wall sample size on HMB estimator

The non-wall-to-wall sample size ( $S_a$ ) determines the number of sample plots for which RS auxiliary data are collected for the dependent and the independent variables of model  $G_{Sa}$ . Apparently, with a given inferential precision, the smaller the  $S_a$  sample size, the fewer the non-wall-to-wall auxiliaries required, and thus the more cost-efficient the HMB estimator; or equivalently, with a given size of  $S_a$ , the precision is dependent on the sampling intensity, i.e., the ratio between the size of  $S_a$  and the size of population, useful for inventory planning and budgeting particularly when an existing model casts the role of model  $F_{s2}$ .

As the sample size for  $S_a$  increases, the predictions produced by model  $G_{Sa}$  are summarized in Table 4, and the pattern of convergence is graphed in Fig. 3. Two findings are relevant. First, as the sample size for  $S_a$  increases, the precision of the HMB estimator converges to that of CMB estimator for which model  $F_{s2}$  is used and with wall-to-wall auxiliaries available, suggesting that (1) HMB is a variant of CMB estimator and is a special case of CMB when sharing the same first-stage model (i.e., Case A.2 vs Case B); (2) by increasing the sample size for  $S_a$ , the margin of benefit resides in reducing the difference between  $\widehat{Var}(\hat{\mu}_1)$  of the CMB estimator and  $\widehat{Var}(\hat{\mu}_2)$  of the HMB estimator (Tables 3 and 4); and (3) when the sample size for  $S_a$  equals the population size, i.e., wall-to-wall, the greatest precision for the HMB estimator is at most  $\widehat{Var}(\hat{\mu}_1)$  (Fig. 3).

Second, the rate of convergence is nonlinear as the sample size for  $S_a$  increases (Table 4), with a reverse J-shaped pattern observed (Fig. 3B), indicating that (1) like design-based inference (Hou et al., 2022), model-based inference also follows the reverse-J pattern between sampling intensity and precision; and thus (2) the sampling intensity for  $S_a$  is cost-efficient around the turning point in the curvature of the reverse-J shape, i.e., the maxima of second-order derivative on its curve function, at empirically 1.32% in terms of sampling intensity. At this point, take EVI model for example, the precision of HMB estimator was only 4% less than that of the CMB estimator using RapidEye wall-to-wall auxiliaries, but 11% greater than that for Landsat 8 wall-to-wall auxiliaries (Tables 3 and 4). Nevertheless, when the sampling intensity is increased to 10.6%, the precision of the HMB estimator was 13.5% greater than that for the CMB estimator using Landsat 8 wall-to-wall auxiliaries, and only 0.9% less than that for RapidEye wall-to-wall auxiliaries. As expected, if the sample size of  $S_a$  increases further, the precision of the HMB estimator would converge to the precision of Case A.2 of the CMB estimator, reiterating the potential of the HMB estimator when used with a greater-quality model,  $F_{s2}$ , and provided with a reasonable size of the non-wall-to-wall auxiliaries for  $S_a$ .

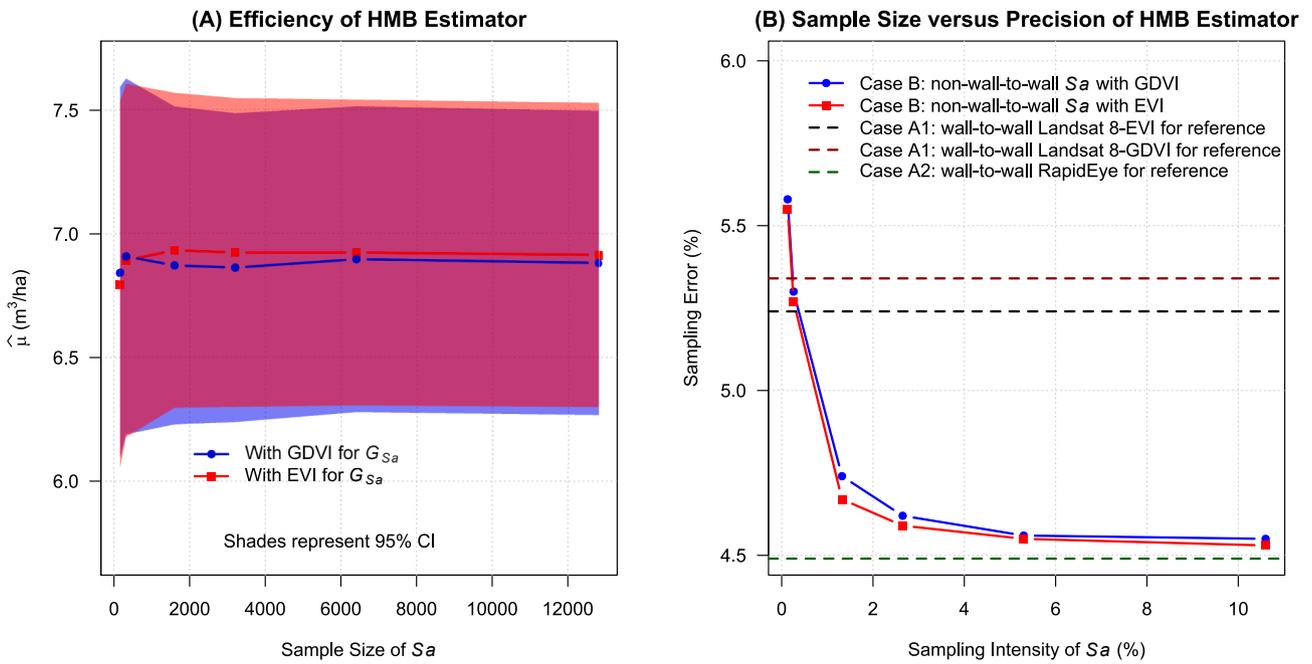


Fig. 3. Estimated population means with 95% confidence intervals (CI) plotted against the sample size of Sa (on the left), and the sampling error plotted against the sampling intensity of Sa (on the right).

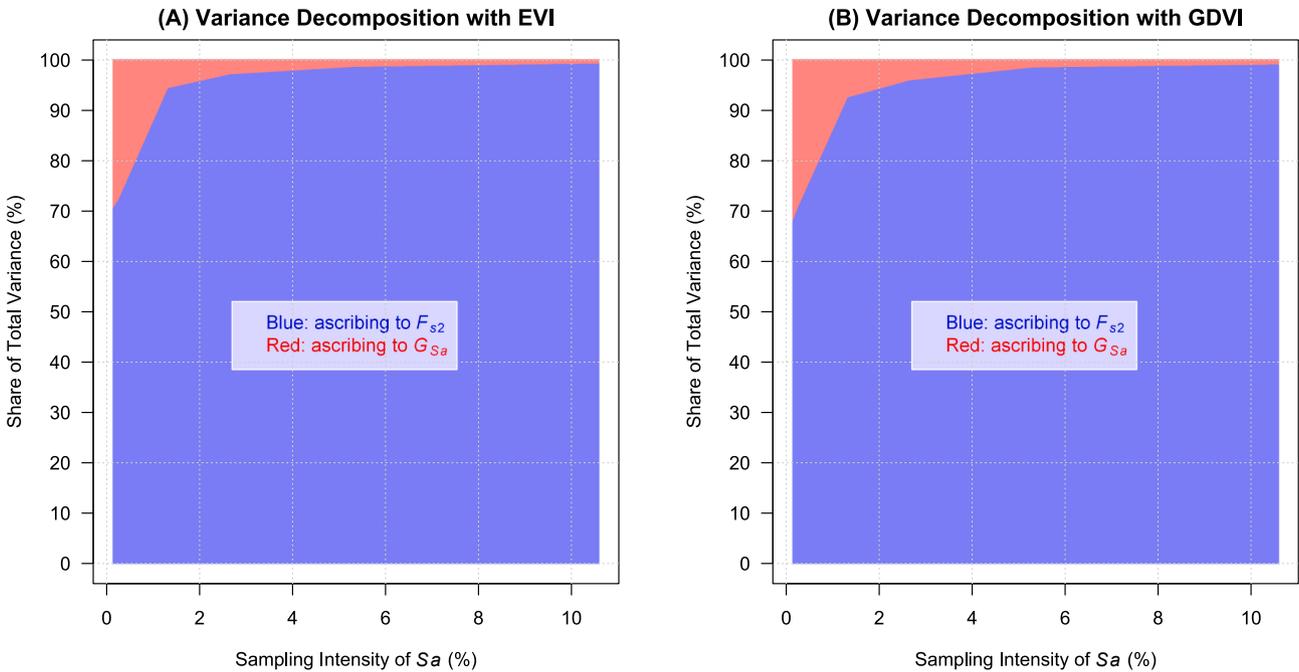


Fig. 4. The share of decomposed variance attributed to the first-, F<sub>S2</sub>, and second-stage, G<sub>Sa</sub>, modeling in the total variance as the sampling intensity of Sa increases with the auxiliary variable of model G<sub>Sa</sub> using EVI (on the left) or GDVI (on the right).

4.4. Variance decomposition for the HMB estimator

A closer look at the composition of the estimated total variance for the HMB estimator,  $\widehat{Var}(\hat{\mu}_2)$ , helps to identify the pattern of uncertainty propagation ascribed to the respective stages with a few towards practical guidance for effective variance reduction. The results of variance decomposition are summarized in Table 4 and graphed out in Fig. 4.

Five findings are relevant. First, the total variance of the HMB estimator is dominated by the first-stage variance component ascribed to model F<sub>S2</sub>. Second, the share of the first-stage variance is on average

3600% greater than that of the second-stage variance. Third, as the sampling intensity for Sa increases, the first-stage share of the total variance increases and would become 100% upon the full coverage of Sa, i.e., wall-to-wall. Fourth, likewise, as the sampling intensity for Sa increases, the share of the second-stage variance decreases and would become 0% upon the full coverage of Sa. Fifth, the difference in precision caused by different second-stage models in the HMB estimator decreases as the sampling intensity for Sa increases because of the dominance of first-stage variance component.

These findings suggest that (1) upon the full coverage for Sa, the first-

stage variance converges to the variance of the CMB estimator using model  $F_{s2}$  with wall-to-wall auxiliaries, and the second-stage variance converges to zero, reiterating that the HMB estimator is a special case of the CMB estimator when sharing the same first-stage model; (2) increasing the sample size for  $S$  for modeling  $F_{s2}$  is more effective for decreasing  $\widehat{Var}(\hat{\mu}_1)$  than increasing the sample size for  $S_a$  for modeling  $G_{sa}$ , because the first-stage modeling contributes to most of the total variance; (3) alternatively, employing a greater-quality model for the first-stage helps to reduce  $\widehat{Var}(\hat{\mu}_1)$  as well; (4) the existing model to be used as model  $F_{s2}$  is key to exerting the advantage of the HMB estimator relative to the CMB estimator; and (5) the HMB estimator balances tradeoffs between cost, efficiency and flexibility, and fulfils what the CMB estimator fulfils, but not vice-versa.

#### 4.5. Application scenarios for HMB estimator

The HMB estimator is not just comparable to the CMB estimator with respect to precision but also enables various practical applications that cannot be performed with the CMB estimator. These applications include but are not limited to scenarios as follows.

First, there exists a high-quality model, but not for its wall-to-wall auxiliaries. This scenario is common for populations under annual or periodic monitoring (Van Deusen, 2002). Reusing this existing model as the first-stage model of the HMB estimator exempts field campaign and RS acquisitions for model construction. Although for the purpose of calibrating the parameters of an existing model it may be necessary to observe the VOI with a few sample plots (Tompalski et al., 2019). Fieldwork of this sort is less burdensome than undertaking a full campaign and can be incorporated into the planning for acquiring second-stage sample,  $S_a$ , to reduce the overall cost. This scenario could be efficient and interesting for the provision of annual estimates required by the UNFCCC.

Second, mapping and upscaling are both desired. In model-based inference, there are few estimators enabling inference with non-wall-to-wall auxiliaries, except for the HMB and hybrid estimators. The hybrid estimator works for non-wall-to-wall auxiliaries (Ståhl et al., 2011). However, while the hybrid and HMB estimators are about equally efficient for upscaling, the hybrid estimator does not support mapping, unlike the HMB estimator. Comprehensive comparisons between the HMB and hybrid estimators are reported in Saarela et al. (2018, 2020).

Third, the acquired RS data cannot be wall-to-wall due to the obstruction of clouds, the malfunction of RS sensors, or the non-wall-to-wall nature of remote sensing instruments. Cloud cover is a major limiting factor due to the obstruction and shading effects in time-series auxiliaries acquired by optical spaceborne sensors including the freely available Landsat 8 and Sentinel-2 (Zhu et al., 2015). Although there are cloud masking algorithms that are increasingly effective (Joshi et al., 2019; Qiu et al., 2019), post-processing acquisitions are essentially non-wall-to-wall. Landsat 7 exemplifies the malfunction of RS sensors. In addition to orbit drifting, failure of the Landsat 7 scan line corrector causes the scanning pattern to exhibit wedge-shaped scan-to-scan gaps, leading to the missing about 22% of the normal scene area, i.e., non-wall-to-wall (Mueller-Warrant, 2019). The Global ecosystem dynamics investigation (GEDI) instrument is a spaceborne waveform lidar installed on the International Space Station for sampling land surface between 51.6° N and 51.6° S latitudes on the order of 4% with a nominal footprint of 25 m (Dubayah et al., 2020). Compared with 2D spectral data, GEDI lidar supplies 3D data that are appreciable for the sampling of surface topography, canopy height, canopy cover, and vertical canopy structure (Duncanson et al., 2022). Indubitably, particularity of these challenges resides in the non-wall-to-wall nature of RS auxiliaries, which is, albeit beyond the appliance of CMB estimator, within the appliance of HMB estimator, both for wall-to-wall mapping and upscaling.

## 5. Conclusions

In this study, the precisions of the CMB (Cases A.1 and A.2) and the HMB (Case B) estimators were compared to demonstrate the use and usefulness of RS non-wall-to-wall data for upscaling within the model-based inferential framework. Five conclusions are relevant: (1) HMB is a variant of CMB estimator and is a special case of CMB when sharing the same first-stage model; (2) an existing model of high quality is key to exerting the advantage of HMB estimator relative to the CMB estimator, e.g., high-quality non-wall-to-wall multi-spectral or 3D remote sensing data employed by the HMB estimator versus moderate wall-to-wall multi-spectral data employed by the CMB estimator; (3) the variance of the HMB estimator is dominated by the first-stage model variance component, indicating that increasing the sample size in the first-stage is effective for increasing the overall precision; (4) RMSE at the level of individual units cannot be used as a measure of the precision of the estimator of the population mean where a variance estimator or coefficient of variation should be used; and (5) overall, the HMB estimator balances tradeoffs between cost, efficiency and flexibility, fulfils what the CMB estimator fulfils, but not vice-versa, with three practical scenarios common to remote sensing-based wall-to-wall upscaling (i.e., estimation) summarized and discussed. Last but not the least, complement to the present study, further research outlined towards maximizing the application of RS non-wall-to-wall data is encouraged for upscaling scenarios where spatial correlation and estimation bias are present (Saarela et al., 2022; Fortin et al., 2023).

#### CRedit authorship contribution statement

**Fangting Chen:** Software, Writing – original draft, Writing – review & editing. **Zhengyang Hou:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Svetlana Saarela:** Methodology, Writing – review & editing. **Ronald E. McRoberts:** Validation, Writing – review & editing. **Göran Ståhl:** Validation, Writing – review & editing. **Annika Kangas:** Validation, Writing – review & editing. **Petteri Packalen:** Writing – review & editing. **Bo Li:** Validation, Writing – review & editing. **Qing Xu:** Investigation, Validation, Writing – original draft, Writing – review & editing, Funding acquisition.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

This work was supported by the National Social Science Fund of China (Grant No. 22BTJ005). Dr. Qing Xu was also supported by the International Center for Bamboo and Rattan (Grant No. 1632020029 and Grant No. 1632021024). We are grateful to the international project, Building Biocarbon and Rural Development in West Africa (BIO-DEV), and our Finnish colleague, Dr. Janne Heiskanen, for organizing the field campaign in Burkina Faso. Review comments provided by the editor and anonymous reviewers are sincerely acknowledged.

#### References

- Bayat, B., Camacho, F., Nickeson, J., Cosh, M., Bolten, J., Vereecken, H., Montzka, C., 2021. Toward operational validation systems for global satellite-based terrestrial essential climate variables. *Int. J. Appl. Earth Obs. Geoinf.* 95, 102240 <https://doi.org/10.1016/j.jag.2020.102240>.

- Bazewew, M.N., Hussin, Y.A., Kloosterman, E.H., 2018. Integrating airborne LiDAR and terrestrial laser scanner forest parameters for accurate above-ground biomass/carbon estimation in Ayer Hitam tropical forest, Malaysia. *Int. J. Appl. Earth Obs. Geoinf.* 73, 638–652. <https://doi.org/10.1016/j.jag.2018.07.026>.
- Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Baston, D., Rouault, E., Warmerdam, F., Ooms, J., Rundel, C., 2013. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.6-5. <https://CRAN.R-project.org/package=rgdal> (accessed 31 March 2023).
- Chambers, R.L., Clark, R.G., 2012. An introduction to model-based survey sampling with applications. Oxford University Press Inc., New York. <https://doi.org/10.1093/acprof:oso/9780198566625.001.0001>.
- Chirici, G., Giannetti, F., McRoberts, R.E., Travaglini, D., Pecchi, M., Maselli, F., Chiesi, M., Corona, P., 2020. Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data. *Int. J. Appl. Earth Obs. Geoinf.* 84, 101959 <https://doi.org/10.1016/J.JAG.2019.101959>.
- Davidson, R., MacKinnon, J.G., 1993. Estimation and inference in econometrics. Oxford University Press, New York.
- Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* 1, 100002 <https://doi.org/10.1016/j.srs.2020.100002>.
- Duncanson, L., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., Patterson, P.L., Saarela, S., Marselis, S., 2022. Aboveground biomass density models for NASA's global ecosystem dynamics investigation (GED) lidar mission. *Remote Sens. Environ.* 270, 112845 <https://doi.org/10.1016/j.rse.2021.112845>.
- Fortin, M., Lier, O.V., Côté, J.-F., 2023. Combining forest growth models and remotely sensed data through a hierarchical model-based inferential framework. *Can. J. For. Res.* 53, 1–13. <https://doi.org/10.1139/cjfr-2022-0168>.
- Galecki, A., Burzykowski, T., 2013. Linear Mixed-Effects Model. In: *Linear Mixed-Effects Models Using R*. Springer Texts in Statistics. Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-3900-4\\_13](https://doi.org/10.1007/978-1-4614-3900-4_13).
- Gregoire, T.G., Næsset, E., McRoberts, R.E., Ståhl, G., Andersen, H.-E., Gobakken, T., Ene, L., Nelson, R., 2016. Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sens. Environ.* 173, 98–108. <https://doi.org/10.1016/j.rse.2015.11.012>.
- Heiskanen, J., Adhikari, H., Piironen, R., Packalen, P., Pellikka, P.K.E., 2019. Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes? *Int. J. Appl. Earth Obs.* 81, 176–185. <https://doi.org/10.1016/j.jag.2019.05.017>.
- Hou, Z., Xu, Q., Tokola, T., 2011. Use of ALS, Airborne CIR and ALOS AVNIR-2 data for estimating tropical forest attributes in Lao PDR. *ISPRS-J. Photogramm. Remote Sens.* 66, 776–786. <https://doi.org/10.1016/j.isprsjprs.2011.09.005>.
- Hou, Z., Xu, Q., McRoberts, R.E., Greenberg, J.A., Liu, J., Heiskanen, J., Pitkänen, S., Packalen, P., 2017. Effects of temporally external auxiliary data on model-based inference. *Remote Sens. Environ.* 198, 150–159. <https://doi.org/10.1016/j.rse.2017.06.013>.
- Hou, Z., McRoberts, R.E., Ståhl, G., Packalen, P., Greenberg, J., Xu, Q., 2018. How much can natural resource inventory benefit from a finer resolution auxiliary data? *Remote Sens. Environ.* 209, 31–40. <https://doi.org/10.1016/j.rse.2018.02.039>.
- Hou, Z., Domke, G., Russell, M., Coulston, J., Nelson, M., Xu, Q., McRoberts, R.E., 2021. Updating annual state- and county-level forest inventory estimates with data assimilation and FIA data. *Forest Ecol. Manag.* 483, 118777 <https://doi.org/10.1016/j.foreco.2020.118777>.
- Hou, Z., McRoberts, R.E., Zhang, C., Ståhl, G., Zhao, X., Wang, X., Li, B., Xu, Q., 2022. Cross-classes domain inference with network sampling for natural resource inventory. *For. Ecosyst.* 9, 100029 <https://doi.org/10.1016/j.fecs.2022.100029>.
- Jayathunga, S., Owarib, T., Tsuyukia, S., 2018. The use of fixed-wing UAV photogrammetry with LiDAR DTM to estimate merchantable volume and carbon stock in living biomass over a mixed conifer–broadleaf forest. *Int. J. Appl. Earth Obs. Geoinf.* 73, 767–777. <https://doi.org/10.1016/j.jag.2018.08.017>.
- Joshi, P.P., Wynne, R.H., Thomas, V.A., 2019. Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8. *Int. J. Appl. Earth Obs. Geoinf.* 82, 101898 <https://doi.org/10.1016/j.jag.2019.101898>.
- Li, S., Wang, T., Hou, Z., Gong, Y., Feng, L., Ge, J., 2021. Harnessing terrestrial laser scanning to predict understory biomass in temperate mixed forests. *Ecol. Indic.* 121, 107011 <https://doi.org/10.1016/j.ecolind.2020.107011>.
- Lu, J.B., Wang, H., Qin, S.H., Cao, L., Pu, R.L., Li, G.L., Sun, J., 2020. Estimation of aboveground biomass of Robinia pseudoacacia forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds. *Int. J. Appl. Earth Obs.* 86, 102014 <https://doi.org/10.1016/j.jag.2019.102014>.
- Masek, J.G., Hayes, D.J., Joseph Hughes, M., Healey, S.P., Turner, D.P., 2015. The role of remote sensing in process-scaling studies of managed forest ecosystems. *For. Ecol. Manag.* 355, 109–123. <https://doi.org/10.1016/j.foreco.2015.05.032>.
- McRoberts, R.E., Næsset, E., Gobakken, T., 2013. Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sens. Environ.* 128, 268–275. <https://doi.org/10.1016/j.rse.2012.10.007>.
- McRoberts, R.E., Næsset, E., Gobakken, T., 2014. Estimation for inaccessible and non-sampled forest areas using model-based inference and remotely sensed auxiliary information. *Remote Sens. Environ.* 154, 226–233. <https://doi.org/10.1016/j.rse.2014.08.028>.
- McRoberts, R.E., Næsset, E., Gobakken, T., Chirici, G., Condés, S., Hou, Z., Saarela, S., Chen, Q., Ståhl, G., Walters, B.F., 2018. Assessing components of the model-based mean square error estimator for remote sensing assisted forest applications. *Can. J. For. Res.* 48, 642–649. <https://doi.org/10.1139/cjfr-2017-0396>.
- Mehtätalo, L., Lappi, J., 2020. *Biometry for Forestry and Environmental Data: With Examples in R*. CRC Press, p. 411.
- Mueller-Warrant, G., 2019. Multistep block mapping on principal component uniformity repairs Landsat 7 defects. *Int. J. Appl. Earth Obs. Geoinf.* 79, 12–23. <https://doi.org/10.1016/j.jag.2019.02.005>.
- Næsset, E., Ørka, H.O., Solberg, S., Bollandsås, O.M., Hansen, E.H., Mauya, E., Zahabu, E., Malimbwi, R., Chamuya, N., Olsson, H., Gobakken, T., 2016. Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: A comparison of estimated precision. *Remote Sens. Environ.* 175, 282–300. <https://doi.org/10.1016/j.rse.2016.01.006>.
- Nord-Larsen, T., Cao, Q.V., 2006. A diameter distribution model for even-aged beech in Denmark. *For. Ecol. Manag.* 231 (1–3), 218–225. <https://doi.org/10.1016/j.foreco.2006.05.054>.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-effects models in S and S-Plus*. Springer-Verlag New York Inc., New York. <https://doi.org/10.1007/b98882>.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2016. R Core Team nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-162. <http://CRAN.R-project.org/package=nlme/> (accessed 15 February 2023).
- Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sens. Environ.* 231, 111205 <https://doi.org/10.1016/j.rse.2019.05.024>.
- Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., Astrup, R., 2014. Comparison of four types of 3D data for timber volume estimation. *Remote Sens. Environ.* 155, 325–333. <https://doi.org/10.1016/j.rse.2014.08.036>.
- Rizopoulos, D., 2022. Model selection by bootstrapping the stepAIC procedure. <https://cran.r-project.org/web/packages/bootStepAIC/bootStepAIC.pdf> (accessed 15 February 2023).
- Saarela, S., Wästlund, A., Holmström, E., Mensah, A.A., Holm, S., Nilsson, M., Fridman, J., Ståhl, G., 2020. Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree level allometric and LiDAR model errors. *For. Ecosyst.* 7, 1–17. <https://doi.org/10.1186/s40663-020-00245-0>.
- Saarela, S., Holm, S., Healey, S.P., Andersen, H.E., Petersson, H., Prentiss, W., Patterson, P.L., Næsset, E., Gregoire, T.G., Ståhl, G., 2018. Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data. *Remote Sens.* 10, 1832. <https://doi.org/10.3390/rs10111832>.
- Saarela, S., Holm, S., Healey, S.P., Patterson, P.L., Yang, Z., Andersen, H., Dubayah, R.O., Qi, W., Duncanson, L.I., Armston, J.D., Gobakken, T., Næsset, E., Ekström, M., Ståhl, G., 2022. Comparing frameworks for biomass prediction for the Global Ecosystem Dynamics Investigation. *Remote Sens. Environ.* 278, 113074 <https://doi.org/10.1016/j.rse.2022.113074>.
- Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., Nelson, R., 2011. Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Can. J. For. Res.* 41, 96–107. <https://doi.org/10.1139/X10-161>.
- 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., 2016. Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *For. Ecosyst.* 3, 1–11. <https://doi.org/10.1186/s40663-016-0064-9>.
- Tompalski, P., White, J.C., Coops, N.C., Wulder, M.A., 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sens. Environ.* 227, 110–124. <https://doi.org/10.1016/j.rse.2019.04.006>.
- Tomppo, E., Gschwanter, T., Lawrence, M., McRoberts, R.E., 2010. *National Forest Inventories: Pathways for common reporting*. Springer, Berlin. <https://doi.org/10.1007/978-90-481-3233-1>.
- USDA, 2008. Accuracy standards. In: *Forest Survey Handbook*. FSH 4809.11, Chapter 10. Washington, DC: U.S. Department of Agriculture, Forest Service.
- Van Deusen, P.C., 2002. Comparison of some annual forest inventory estimators. *Can. J. For. Res.* 32, 1992–1995. <https://doi.org/10.1139/x02-115>.
- Vidal, C., Alberdi, I., Hernández, L., Redmond, J., 2016. *National Forest Inventories: Assessment of Wood Availability and Use*. Springer: Cham, Switzerland. <https://doi.org/10.1007/978-3-319-44015-6>.
- Wang, D., Wan, B., Liu, J., Su, Y., Guo, Q., Qiu, P., Wu, X., 2020. Estimating aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery. *Int. J. Appl. Earth Obs.* 85, 101986 <https://doi.org/10.1016/j.jag.2019.101986>.
- Wittke, S., Yu, X., Karjalainen, M., Hyypää, J., Puttonen, E., 2019. Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest. *Int. J. Appl. Earth Obs. Geoinf.* 76, 167–178. <https://doi.org/10.1016/j.jag.2018.11.009>.
- Xu, Q., Li, B., McRoberts, R.E., Li, Z., Hou, Z., 2023. Harnessing data assimilation and spatial autocorrelation for forest inventory. *Remote Sens. Environ.* 288, 113488 <https://doi.org/10.1016/j.rse.2023.113488>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.