

Article

Integrating Remotely Sensed Data to Reconcile Gaps in Growing Stock Volume Accounting for National Forest Inventory

Temitope Olaoluwa Omoniyi ^{1,*} , Allan Sims ¹ , Ronald E. McRoberts ² and Mercy Ajayi-Ebenezer ³

¹ Chair of Forest Management and Forest Industry, Institute of Forestry and Engineering, Estonian University of Life Sciences, Fr.R. Kreutzwaldi 5, 51006 Tartu, Estonia; allan.sims@emu.ee

² Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue North, Saint Paul, MN 55108, USA; mcrob001@umn.edu

³ Chair of Plant Biology and Agriculture, Institute of Agricultural and Environmental, Estonian University of Life Sciences, Fr.R. Kreutzwaldi 5, 51006 Tartu, Estonia; mercy.ajayi-ebenezer@emu.ee

* Correspondence: temitope.omoniyi@emu.ee

Abstract

National forest inventory (NFI) data are often collected over a 5-year or 10-year period, meaning some are already outdated by the time the complete results are available. This study assesses changes in growing stock volume (GSV, m³/ha) using hybrid estimation supported by Sentinel-2 metrics. It focuses on constructing a model for estimating the change in GSV using NFI plot data and bitemporal remotely sensed auxiliary data, where such data are available for both points in time (t_1 and t_2), and unitemporal data for which remotely sensed auxiliary data are available only for t_2 . A machine-learning approach based on the random forests (RFs) algorithm was used to predict plot-level GSV change. The original data for t_2 and t_3 were first used to evaluate the accuracy of the change prediction at the plot level, after which the predicted changes were applied to update the plot-level GSV to predict plot-level GSV at t_3 , which was then assessed against the observed plot-level GSV at t_3 . Predicted change was assessed with the Mean Average Annual Volume Change (MAAVC) method, representing the average annual change in GSV over a given period. The results indicate that at the plot level, the bitemporal model produced GSV change estimates with low accuracy ($R^2 = 0.26$, RMSE = 4.06 m³/ha, and MAE = 3.26 m³/ha), while the unitemporal model achieved $R^2 = 0.40$, RMSE = 3.64 m³/ha, and MAE = 2.65 m³/ha when predicting the $t_1 - t_2$ GSV change. Using the predicted change to predict plot-level GSV at t_3 , the MAAVC based on field data yielded $R^2 = 0.91$ and RMSE = 45.11 m³/ha, while the RS unitemporal yielded $R^2 = 0.73$ and RMSE = 83.79 m³/ha, and the bitemporal yielded $R^2 = 0.72$ and RMSE = 83.61 m³/ha. Mean population GSV at t_3 , estimated from the RF models, was 254.61 and 255.19 m³/ha for the unitemporal and bitemporal models, respectively. Monte Carlo simulations with a novel stopping criterion were then used to estimate total standard errors, which were 10.48 and 10.40 m³/ha for the unitemporal and bitemporal models, respectively, incorporating both model prediction uncertainty and sampling variability. A test of significance revealed a significant effect of the proposed method on the estimated mean population GSV at t_3 ($p < 0.001$). Conclusively, MAAVC and spatiotemporal RS methods provide a robust framework for predicting GSV at t_3 using Estonian NFI and Sentinel-2 data.



Academic Editors: Guojie Wang and Zengxin Zhang

Received: 22 January 2026

Revised: 15 February 2026

Accepted: 15 February 2026

Published: 18 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: growing stock volume; national forest inventory; change detection; sentinel-2; random forests; uncertainty

1. Introduction

The advent of national forest inventories (NFIs) has led to a comprehensive shift in forest growth measurements and analytical practices, thus supporting the implementation, monitoring, and promotion of sustainable forest management practices globally. Advanced inventory systems, such as NFIs, have been developed in many countries across Europe and beyond, evolving over decades and sometimes for more than a century [1,2]. Although an NFI is described in the Food and Agriculture Organization of the United Nations (FAO)'s Voluntary Guidelines on National Forest Monitoring (VGNFM) as a technical process for compiling data and analyzing forest resources at the national level [3], its content, concept, and definition are continually refined to align with users' needs [2]. A country's inventory capacity development is customized to its unique needs, covering activities such as planning, data collection, analysis, integration with remotely sensed data, quality assurance and control, data archiving, documentation, dissemination, and reporting [4]. Initially, NFIs focused on assessing timber resources and supporting sustainable forestry. They have been expanded over the years to encompass forest health monitoring, management practices, carbon sequestration, biodiversity, ecosystem services, and other variables, alongside greater diversity in sampling protocols and a more comprehensive, holistic approach [5].

Forest biophysical variables, such as diameter at breast height (dbh) and tree height (h), used to be the primary variables measured directly in the field. However, there has been a growing interest in a wider range of variables to provide comprehensive insights into forest ecosystems. Metrics such as growing stock volume (GSV, m³/ha), aboveground biomass (AGB), and belowground biomass are classified as predicted variables. Among these, GSV and AGB are particularly essential in many NFIs, especially when focusing on forest productivity, policy development, carbon storage, sustainability, and enhanced decision-making processes [6,7]. To monitor forest sustainability, track management practices, and evaluate forests' contribution to the carbon cycle, organizations, such as the FAO, United Nations Framework Convention on Climate Change (UNFCCC), and the Kyoto Protocol, rely on estimates of forest conditions and changes [1]. NFIs are instrumental in providing these essential estimates, which are critical for effective forest management planning. Having accurate and current data is crucial [8]; therefore, member countries are required to report the status of their forests every five years for transparency and accountability [9,10].

The availability of remotely sensed auxiliary data has helped in the development of techniques that can be used to predict forest variables and enhance precision [10]. Remotely sensed data, such as optical imagery (e.g., Landsat, Sentinel-2 MSI), lidar data, radar data (e.g., SAR), and hyperspectral imagery, are widely used for predicting variables including GSV, forest structure, biomass, carbon stocks, tree species composition, forest health, and canopy cover [11–13]. For example, Lee and Lee [14] predicted forest height using discrete-return lidar data, SRTM, satellite L-band SAR data, and optical data, achieving improved results through the application of small baseline subset (SBAS) algorithms and linear regression. Condés and McRoberts [15] combined auxiliary site data, such as topographic slope, mean annual temperature, annual precipitation, and the Martonne aridity index, with remotely sensed data such as Landsat imagery, to update GSV estimates for the Spanish National Forest Inventory (SNFI) through hybrid inference. One of their findings was that predictions using field and satellite data could be leveraged to update NFI estimates in years for which field observations were absent but spectral data were available. White et al. [13] concluded that remotely sensed data enhance NFIs in four main ways: (1) enabling quicker and cheaper methods for estimating forest attributes; (2) improving the accuracy of large-area inventory estimates, often through stratification or weighted estimation; (3) offering inventory estimates with acceptable error and precision for small areas lacking

sufficient field data; and (4) producing forest thematic maps that support timber production, procurement, and ecological studies.

Despite the issue of data saturation with optical data [7,16,17], such data have been extensively used for forest GSV estimation due to their diverse spatial, spectral, radiometric, and temporal resolutions, advanced processing technologies, abundant data sources, and extensive coverage [18–23]. Landsat has been a longstanding auxiliary data source in support of forest estimation, but the emergence of Sentinel-2 data has introduced competition and provided a compelling alternative. Zhou and Feng [24] evaluated the accuracy of GSV estimates based on Sentinel-2 and Landsat 8 OLI as auxiliary datasets, finding that models based on Sentinel-2 delivered greater precision. In 2020, Clark [25] conducted a comparison of multi-season Landsat 8, Sentinel-2, and hyperspectral imagery for classifying forest alliances in northern California and highlighted Sentinel-2's greater overall prediction accuracy than Landsat 8. Mura et al. [26] assessed the ability of Sentinel-2's Multi-Spectral Instrument (S2-MSI) to predict GSV in Italy. They tested its performance against data from Landsat 8 OLI and the RapidEye scanner using a consistent experimental protocol and found that Sentinel-2 delivered more accurate outcomes.

Change estimation from remotely sensed data can be approached either indirectly or directly. Indirect estimation involves constructing models of the relationship between the response variable and remotely sensed auxiliary variables separately for two temporal points, with change estimated as the difference between the predictions. Direct estimation involves constructing models of change directly using observations of change in the response variable and the corresponding changes in remotely sensed auxiliary variables over the same two temporal points. Mean change is then estimated from the average change predictions, with the model's prediction errors incorporated in the uncertainty estimation process [6,27]. While achieving temporal alignment between remotely sensed data and field inventory measurements is recognized as optimal for enhancing forest inferences [27], the utility of auxiliary satellite data remains significant even when temporal gaps exist [15,27]. In cases where models are employed to predict forest attributes due to the lack of direct observations for sample plot locations at the time of interest, uncertainties in the predictions must be incorporated to maintain the unbiasedness of variance estimators [15]. Despite the active use of remotely sensed data and machine learning (ML) methods in forest mapping, the estimation of the effects of prediction uncertainty is relatively uncommon [28]. Therefore, this study aims to create a model for estimating changes in GSV using NFI data and predicting GSV values for permanent sample plots before the next revisit. It also investigates Sentinel-2 optical data as a cost-effective, independent auxiliary data source for improving forest management and the associated uncertainties.

2. Materials and Methods

2.1. Study Area

The study area is situated in Estonia, a northern European country located along the Baltic Sea, between latitudes 57°30'34" N and 59°49'12" N and longitudes 21°45'49" E and 28°12'44" E. Estonia is bordered by the Baltic Sea to the west and the Gulf of Finland to the north (Figure 1).

Forests dominate approximately 53.6% of the country's land area, covering about 23,308 km² of Estonia's total 45,339 km² [29,30]. The primary tree species contributing substantially to Estonia's GSV include Scots pine (*Pinus sylvestris* L.), which covers approximately 34% of the forested area; downy birch (*Betula pubescens* Ehrh.) and silver birch (*Betula pendula* Roth), jointly occupying approximately 31%; and Norway spruce (*Picea abies* (L.) Karst), representing 16%. Other notable species include European aspen (*Populus tremula* L.), gray alder (*Alnus incana* (L.) Moench), black alder (*Alnus glutinosa* (L.) Gaertn.),

and common ash (*Fraxinus excelsior* L.) [31]. For this study, all permanent sample plots with repeated data available from the 2013, 2018, and 2023 NFI years were used. The species observed in this selection are detailed in Table 1.

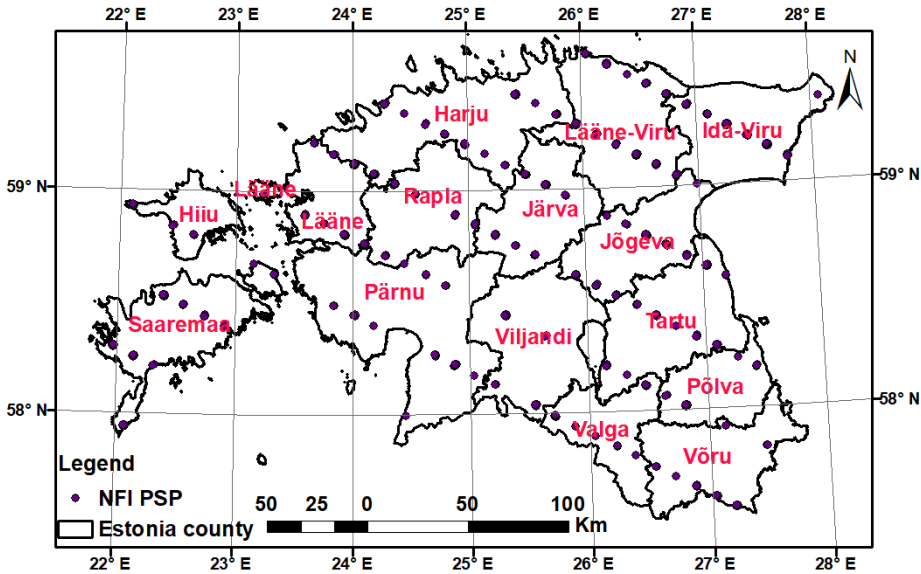


Figure 1. Estonian network of clusters of permanent NFI sample plots (PSP) for 2023.

Table 1. Species characteristics of the sample plots.

Species *	Frequency	Proportion	Cumulative Frequency
HB	17	5.5	17
KS	83	26.86	100
KU	49	15.86	149
LM	13	4.21	162
LV	32	10.36	194
MA	110	35.6	305
Others	5	1.62	309

* HB is European aspen (*Populus tremula* L.), KS is birch (silver birch (*Betula pendula* Roth), and downy birch (*Betula pubescens* Ehrh.)), KU is Norway spruce (*Picea abies* (L.) Karst.), LM is common alder (*Alnus glutinosa*), LV is gray alder (*A. incana* (L.) Moench, MA is Scots pine (*Pinus sylvestris* L.).

2.2. General Overview of Methods

For this study, Estonian NFI data for the years 2013, 2018, and 2023 were used to estimate changes in plot-level forest GSV and to estimate population-level GSV using MAAVC and unitemporal and bitemporal remote sensing metrics from Sentinel-2, with RF algorithms. The MAAVC method refers to average annual GSV change over a 5-year period. The predicted changes were then compared with in situ inventory data using standard performance metrics. Data for a total of 219 plots were used, with model development based on observed plot-level changes between 2013 and 2018 (t_1 and t_2). We organized our data into two categories, Dataset A and Dataset B. Dataset A included permanent plot-level GSV observations for 2013 (t_1) and 2018 (t_2), while Dataset B comprised data for 2018 (t_2) and 2023 (t_3), featuring plot-level GSV observations. For the random forests (RF) model development, changes between 2013 and 2018 were used with data for 80% of the total plots as training datasets, and the remaining 20% as a test set; the developed models and the MAAVC approach were used to predict plot-level GSV at 2023 (t_3), which was then compared with observed plot-level GSV at t_3 . The methodology flowchart is shown in Figure 2.

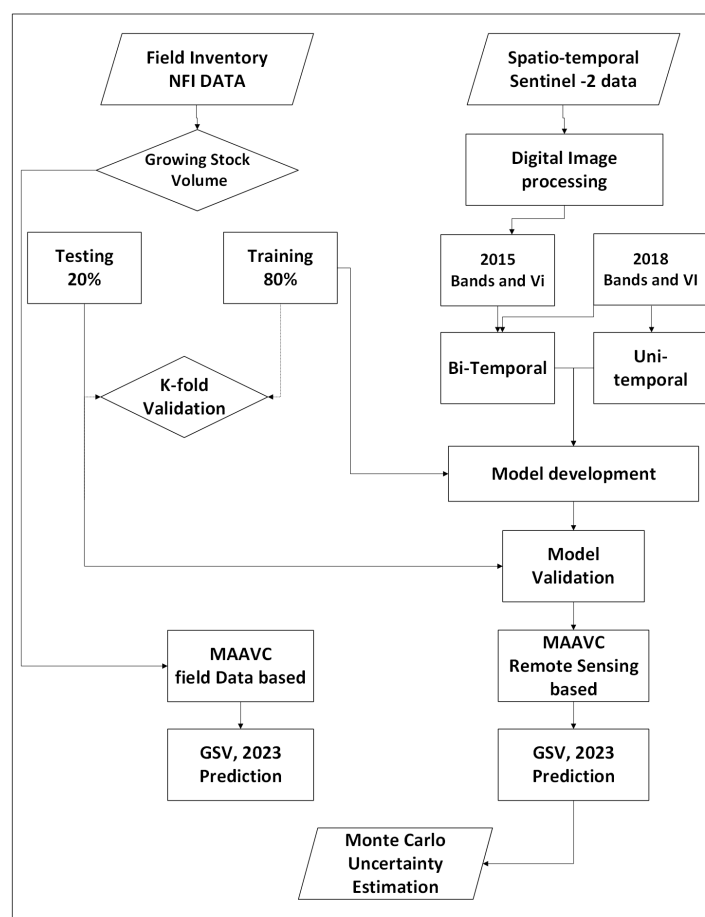


Figure 2. Methodology flowchart for this study.

2.3. National Forest Inventory Data

The Estonian NFI follows a structured interpenetrating panel sampling approach whereby a systematic 20% of the plots are measured each year [7], a standard methodology adopted by many NFIs globally. In 1999, Estonia initiated its first nationwide NFI [32], aimed at evaluating the country's forest resources. The NFI uses both temporary and permanent sample plots. The permanent sample plots play a crucial role by enabling consistent monitoring and estimation of changes over time. These plots provide reliable data for plot-wise analysis, facilitating effective detection of changes in forest structure and composition. The sample plots are systematically distributed across Estonia using a 5 km × 5 km grid, based on the L-EST coordinate system (EPSG:3301) to ensure uniform sampling intensity. Three types of circular sample plots with fixed radii are included: (i) volume plots, (ii) site category plots, and (iii) regeneration/felling plots. Plots were established and grouped into clusters spanning 800 × 800 m for efficient fieldwork and data collection. For this research, volume plots were used, and data were obtained from permanent plots and recorded during three (3rd, 4th, and 5th) consecutive Estonian NFI cycles, which correspond to years 2009–2013, years 2014–2018, and years 2018–2023, with specific data from years 2013, 2018, and 2023 selected to represent consistent five-year intervals.

2.4. GSV Estimation from the Estonian NFI Field Plots

The principal sample plots in the Estonian NFI are circular with radii of 10 m for volume sampling and 7 m for site category [16] which correspond to areas of approximately 315 m² and 154 m², respectively. Within each plot, trees with a dbh of 1.3 m and greater were measured for height according to Estonian NFI criteria, designed to minimize errors

in dbh-to-height conversions [33]. The dbh measurements were obtained using a caliper with ± 1.0 mm accuracy, applying a consistent single-measurement approach. Heights were measured using a Vertex hypsometer instrument, which uses ultrasonic signals and trigonometric principles to derive height from distance and angle data. The center of each sample plot was located using a Global Navigation Satellite System (GNSS) receiver with an accuracy as great as 5 m, thereby ensuring accurate alignment with satellite-based data. The longitude and latitude of each plot center were converted into spatial objects and represented as $20 \text{ m} \times 20 \text{ m}$ vector polygons using a bounding-box approach implemented in R. A regional species-specific allometric model—Equation (1)—was used to predict tree-level GSV.

$$GSV = (\beta_1 + \beta_2 \cdot h + \beta_3 \cdot f) \cdot \left(\frac{t}{t+5} \right)^{(\beta_5 + \beta_6 \cdot h)} + \varepsilon, \quad (1)$$

where GSV is the growing stock volume of each tree; β_1 , β_2 , β_3 , β_4 , and β_5 are species-specific parameters whose estimates were obtained from unpublished local sources; h is tree height; t denotes tree age; and f is a dummy variable assigned a value of 1 when the sample plot is located on an island in Estonia and zero (0) if the plot is on the mainland. Individual tree predictions from Equation (1) were then added to predict the sample plot GSV and scaled to a per-unit area basis (m^3/ha). The uncertainty in these tree-level predictions was considered negligible and ignored.

2.5. Remotely Sensed Data

This study harnessed Sentinel-2 MSI data, an advanced optical satellite product from the European Space Agency (ESA) mission known for fine-resolution, wide-swath multispectral imaging with a 5-day revisit at the Equator [26]. Sentinel-2 features a multispectral sensor capturing data across 13 bands, including visible, NIR, and SWIR regions, with three distinct resolutions of 10 m (visible/NIR), 20 m (SWIR), and 60 m for Coastal Aerosol, Water Vapor, and Cirrus bands [28]. The imagery was downloaded from the Copernicus Open Access Hub (<https://dataspace.copernicus.eu>) on 24 November 2024, with bands 2–8, 8A, 11, and 12 selected for analysis (Table 2). Atmospheric corrections were applied using Sen2Cor version 2.11 within the Sentinel-2 Application Platform (SNAP), converting top-of-atmosphere (TOA) reflectance to bottom-of-atmosphere (BOA) surface reflectance, followed by resampling to a 10 m spatial resolution. Cloud and shadow masking were performed using SNAP tools. A single image covering the entire Estonian landmass requires approximately 14 Sentinel-2 scenes. These scenes were mosaicked to create a unified dataset. Corresponding band reflectance values were extracted using a weighted-mean approach, averaging the pixel values intersecting with the shapefiles. As noted by Puliti et al. [27], remotely sensed data should be temporally aligned with field reference data for optimal results. However, the ESA mission for Sentinel-2 began its operation in 2015. For this study, Sentinel-2 data for 2015 were used as an alternative for the 2013 data, when matching spectral data for the bitemporal analysis were unavailable. Such small temporal mismatches between plot measurements and remotely sensed imagery are common in forest-inventory applications and have been addressed similarly in previous studies [15,27,34]. Meanwhile, data for the years 2018 and 2023 were used for the unitemporal analysis, aligning with the timing of the 2018 and 2023 field inventories.

Table 2. Origin and explanation of variables.

Variable		Descriptions
Sentinel-2 band	B2	Blue (458–523 nm) 10 m
	B3	Green (543–578 nm) 10 m
	B4	Red (650–680 nm) 10 m
	B5	Red-edge 1 (698–713 nm) 20 m
	B6	Red-edge 2 (733–748 nm) 20 m
	B7	Red-edge 3 (773–793 nm) 20 m
	B8	NIR 1 (785–899 nm) 10 m
	B8A	NIR 2 (855–875 nm) 20 m
	B11	SWIR 1 (1565–1655 nm) 20 m
	B12	SWIR 2 (2100–2280 nm) 20 m
Veg. indices	NDVI	$(B8 - B4)/(B8 + B4)$
	ARVI	$(B8A - B4 - y \times (B4 - B2))/(B8A + B4 - y \times (B4 - B2))$
	NDVIre705	$(B8-B5)/(B8 + B5)$
	NDVIre740	$(B8-B6)/(B8 + B6)$
	NDVIre783	$(B8-B7)/(B8 + B7)$
	EVI	$2.5 \times ((B8-B4)/(B8 + 6 \times B4 - 7.5 \times B2 + 1))$
	SAVI	$(B8-B4)/(B8 + B4 + L) \times (1 + L)$
	chlre	$((B8)/(B5)) - 1$
	GNDVI	$(B8-B3)/(B8 + B3)$
	MSAVI	$0.5 \times (2 \times B8 + 1 - \sqrt{(2 \times B8 + 1)^2 - 8 \times (B8 - B4)})$

nm is the wavelength in nanometers, 10 m and 20 m are the spatial resolution, L is the soil brightness correction factor set at 0.5, and $y = 0.106$. The vegetation indices were selected based on their relevance to forest structure and condition, and their common use in forest remote sensing studies [24,35,36].

2.6. Datasets Description and GSV Change Prediction

We used a direct method for change estimation, which models change explicitly by using data from two points in time, where the response variable is the change in the variable of interest and the predictor variables are either the changes in auxiliary variables between the two time points or the auxiliary variables at the second point in time. The study explored three approaches based on the availability of data: (i) the bitemporal RS approach, for which data are available for the initial measurement and the remeasurement time (t_1 and t_2) for both ground observations and RS; (ii) the unitemporal approach, for which RS data are available only for the second measurement (t_2), whereas ground observations are available for both measurement times, and so we can connect plot-level GSV changes to the reflectance properties of the forest at the remeasurement time (t_2); and (iii) the mean annual average volume change approach (MAAVC) based on the average annual observed change in plot-level GSV over a given period.

For both unitemporal and bitemporal approaches, two datasets were constructed according to Condés and McRoberts [15]. We organized our data into two categories, Dataset A and Dataset B. Dataset A included permanent plot-level GSV observations for 2013 (t_1) and 2018 (t_2), along with Sentinel-2 derived variables for the latter year and their temporal differences. Dataset B comprised data for 2018 (t_2) and 2023 (t_3), featuring plot-level GSV observations for both years and the associated changes in Sentinel-2 derived variables. The unitemporal approach represents cases for which observed GSV at the plot-level were available for t_1 and t_2 , while RS data are only available for t_2 . In this case, we predict plot-level GSV changes based on the use of remotely sensed data at the end of the monitoring period, such as t_2 . For each dataset category, because tree-level allometric model prediction uncertainty is considered negligible, the average annual volume change (ΔGSV) for each sample plot j was considered as an observation without error and was calculated using Equations (2) and (3):

$$\Delta GSV_{t21j} = \frac{GSV_{t2j} - GSV_{t1j}}{t_2 - t_1}, \quad (2)$$

where ΔGSV_{t1j} is the average annual change in GSV for Dataset A, GSV_{t1j} and GSV_{t2j} are the GSV observations for individual plot j at the times t_1 and t_2 respectively, and $t_2 - t_1$ is the number of years between the revisited measurement in Dataset A. For clarity and specificity, ΔGSV for Dataset B is calculated using Equation (3).

$$\Delta GSV_{t32j} = \frac{GSV_{t3j} - GSV_{t2j}}{t_3 - t_2}, \quad (3)$$

where ΔGSV_{t32j} is the average annual change in GSV for Dataset B, GSV_{t2j} and GSV_{t3j} are the observed GSV values for individual plot j at the times t_2 and t_3 , (Dataset B) and $t_3 - t_2$ is the number of years between the repeated measurements in Dataset B.

2.7. Methods for Predicting Plot-Level GSV at t_3

To estimate plot-level GSV at t_3 , we predicted the total volume at the end of the monitoring period by accounting for the annual change in GSV between t_2 , and t_3 , (Dataset B) for each sample plot.

The corresponding plot level GSV at t_3 was then predicted using linear Equation (4).

$$\widehat{GSV}_{t3j} = GSV_{t2j} + \widehat{\Delta GSV}_{t23j} \times (t_3 - t_2), \quad (4)$$

where $(t_3 - t_2)$ represents the number of years between the times t_2 and t_3 for plot j . This approach uses sample plot-level GSV change predictions, $\widehat{\Delta GSV}_{t23j}$, which are reliable when changes between periods are consistent; major disturbances such as windthrow, fire, or harvesting may cause larger deviations, and in such cases, ALS data or gain or loss methods can be applied. Dataset B was then used to assess the results by comparing the predicted plot-level GSV at t_3 with the corresponding observed plot-level GSV at t_3 . Three methods were used to predict plot-level change, $\widehat{\Delta GSV}_{t23j}$, MAACV, unitemporal RF, and bitemporal RF.

2.7.1. Mean Average Annual Volume Change Prediction from Plot Data

Having established the average annual change in plot-level GSV between the times t_1 and t_2 (Dataset A), we predicted the average annual volume change between t_2 and t_3 (Dataset B) for the NFI cycle using the Mean Average Annual Volume Change (MAACV). This method refers to how much GSV changes each year, on average, over a certain period. In this approach, MAACV volume changes over all plots; $\widehat{\Delta GSV}_{t32j}$ was estimated using Equation (5). This method is partially motivated by well-established knowledge in forestry that recent past growth is often one of the strongest predictors of subsequent tree or stand growth (e.g., previous growth has been shown to be a strong determinant of future growth [37]).

$$\widehat{\Delta GSV}_{t32j} = \Delta GSV_{t21} = \frac{1}{n} \sum_{j=1}^n \frac{GSV_{t2j} - GSV_{t1j}}{t_2 - t_1} \quad (5)$$

where ΔGSV_{t21} is the mean average annual change over all plots for the years between t_2 and t_1 (Dataset A), GSV_{t2j} and GSV_{t1j} are the observed GSV values for individual plot j at the times t_2 and t_1 , $t_2 - t_1$ is the number of years between the repeated measurements in Dataset B, and n is the total number of plots.

2.7.2. Mean Average Annual Volume Change Prediction from RS Data

For the RS approaches, two methods were adopted: the unitemporal and bitemporal. In both cases, an RF algorithm was calibrated using observed plot-level GSV change between t_1 and t_2 (Dataset A) as the dependent variable. In the bitemporal approach, the RS change metrics between t_1 and t_2 were used as the independent variables during

training, and the trained RF model was then applied using the RS change metrics between t_2 and t_3 (Dataset B) to predict plot-level GSV change for that interval, resulting in $\widehat{\Delta GSV}_{t32j}$; the prediction of plot-level GSV at time t_3 for plot j was obtained using Equation (4) with this predicted value. In the unitemporal approach, RS metrics at t_2 were used as the independent variables during training, and the trained RF model was then applied using RS metrics at t_3 to predict plot-level GSV change between t_2 and t_3 (Dataset B), which also produced $\widehat{\Delta GSV}_{t32j}$ for the unitemporal approach, as in the bitemporal case, the prediction of plot-level GSV at time t_3 for plot j was again obtained using Equation (4) with this predicted value.

2.8. Independent Variable Screening

A set of 20 predictor variables, encompassing Sentinel-2 reflectance values and vegetation-derived indices (Table 2), was analyzed. To handle multicollinearity among the independent variables, we applied the Boruta feature selection method and variance inflation factors (VIFs) to refine our variables [38]. Initially, Boruta was used to rank independent variables by their importance, as reflected in their Z-score values, and to further enhance the selection process, we calculated VIFs for the remaining predictor variables. A stepwise VIF diagnostic was performed using the *car* package in R (version 3.1.2). An iterative procedure was applied in which the variable with the greatest VIF was sequentially removed until all remaining predictors exhibited VIF values lower than the threshold of 10. This threshold followed the common criterion for identifying potentially problematic multicollinearity. The final subset of predictors thus retained only variables demonstrating acceptable levels of collinearity.

2.9. Random Forests Prediction

RF is a robust and widely used machine learning algorithm that combines multiple decision trees to enhance predictive performance. It was developed by Breiman [39] and operates through two key steps: (1) bootstrapping, where subsets of the training data are randomly selected with replacement to create diverse training datasets for each decision tree; and (2) feature randomness, where a random subset of features is considered at each split. RF predicts by averaging outputs from all decision trees, while for classification, it uses majority voting [40]. This ensemble approach handles multicollinearity and produces high accuracies, even with large datasets and many-dimensional features. While computationally intensive, RF's ability to predict complex relationships makes it a versatile choice for classification and regression tasks across various domains [41]. The RF prediction in this study was calibrated using the R package RandomForest [41]. It was initially set to *n*tree value of 100, then adjusted by steps of 100. The number of features at each split (*m*try) was fine-tuned using the *tuneRF* function, which selects the optimal *m*try value that minimizes the out-of-bag (OOB) error estimate.

2.10. Prediction Evaluation

The dataset structure and analysis necessitate that we perform accuracy tests for our predictions at two levels, according to Condés and McRoberts [15]: (i) the plot-level assessment and (ii) the population inference level. In the plot-level assessment, we calculate residuals for each plot and summarize them over all plots using mean absolute error (MAE) and coefficient of determination (R^2) based on Dataset B and the predicted GSV at t_3 ($\widehat{GSV}_{t3j}$) using Equations (6) and (7).

$$MAE = \frac{\sum_{j=1}^n |\widehat{GSV}_{t3j} - GSV_{t3j}|}{n}, \quad (6)$$

$$R^2 = 1 - \frac{\sum_{j=1}^n \left(GSV_{t3j} - \widehat{GSV}_{t3j} \right)^2}{\sum_{j=1}^n \left(GSV_{t3j} - \overline{GSV}_{t3} \right)^2} \quad (7)$$

2.11. Hybrid Inference Framework for Mean GSV

Hybrid inference, introduced by Corona et al. [42], which they initially called “hybrid perspective”, is a statistically robust framework for estimating population parameters when only predictions of the response variable based on auxiliary information, rather than direct observations, are available for sample plots. It deals with situations for which models predict a response variable using auxiliary data for a probability sample, and probability-based estimators rely on these predictions to estimate population parameters. It is based on this four-step framework: (1) a probability sample of the population for which auxiliary data are available; (2) a predictive model using the auxiliary data to estimate the response variable for the sample plots; (3) a probability design-based estimator of the population mean using the model predictions for the sample plots; and (4) estimating the variance of the population mean estimator through both model-based and probability-based methods [15,43].

For our study, the probability sample consisted of the sample plots unit locations for t_3 while the auxiliary information included Sentinel-2 derived variables (Table 2) for both the unitemporal and bitemporal methods. In our variance estimation, we accounted for both RF-based prediction uncertainties and design-based sampling uncertainties, ensuring a comprehensive representation of uncertainty from the prediction model and the sampling design, as described below.

To have an unbiased estimator of the variance, we estimated the variance of the estimated population mean $\widehat{Var}(\hat{\mu}_{GSVt3})$ using hybrid inference implemented with a replicated Monte Carlo simulation technique according to Condés and McRoberts [15]. This helps us incorporate the two uncertainty components: (i) the uncertainty arising from using model predictions ($\widehat{\Delta GSV}_{t3j}$) from Equation (5) instead of direct observations of volume change between the t_3 and t_2 , as in Equation (10); and (ii) the sampling uncertainty due to relying on a probability sample rather than a complete population census at time t_3 , as in Equation (9).

For this method, replicate bootstrap resamples of the same size as the original dataset were randomly drawn with replacement from Dataset A, and the models were recalibrated using these resampled data. These recalibrated models were then applied to a second bootstrap resample, drawn from Dataset B, which matched the size of the original sample size in t_3 . This process ensured that each replication used a different dataset for model calibration and prediction. For every replication, r , an estimate of the parameter of interest $\hat{\mu}_{GSVt3}^r$ was obtained. The population mean ($\hat{\mu}_{GSVt3}$) and its variance ($\widehat{Var}(\hat{\mu}_{GSVt3})$) were then estimated using Equations (8) and (11). The criterion used to stop replications follows the criterion proposed by McRoberts et al. [44], which states that $\widehat{SE}_B(\hat{\mu})$ for the final replication B does not deviate by more than 0.5% from bootstrap SE estimates for any of the previous 50% replications. In our design, we ensure the robustness of the stopping rule, as we require three consecutive passes of the stoppage criterion, as seen in Equation (12).

$$\hat{\mu}_{GSVt3}^{\text{avg}} = \frac{1}{B} \sum_{r=1}^B \hat{\mu}_{GSVt3}^{(r)} \quad (8)$$

$$W_1 = \frac{1}{R-1} \sum_{r=1}^R \left(\hat{\mu}_{GSVt3}^{(r)} - \hat{\mu}_{GSVt3}^{\text{avg}} \right)^2, \quad (9)$$

$$W_2 = \frac{1}{R} \sum_{r=1}^R \widehat{\text{Var}}_{\text{within}}^{(r)} \quad (10)$$

$$\widehat{\text{Var}}(\hat{\mu}_{\text{GSV}t3}) = W_1 + W_2 \quad (11)$$

where W_1 and W_2 are the model and sampling error component, and the standard error (SE) is calculated as $SE = \sqrt{\widehat{\text{Var}}(\hat{\mu}_{\text{GSV}t3})}$

$$\left| \frac{\widehat{SE}_b(\hat{\mu}) - \widehat{SE}_j(\hat{\mu})}{\widehat{SE}_j(\hat{\mu})} \right| \leq 0.5\% \text{ for all } j \in \left\{ \left\lceil \frac{b}{2} \right\rceil, \dots, b-1 \right\} \quad (12)$$

where $b = B, B-1$, and $B-2$ are all denoting three consecutive passes.

2.12. Statistical Analysis

A one-way repeated-measures ANOVA was conducted to examine the effect of the proposed method on GSV estimates. The assumption of sphericity, which requires that the variance of the differences between all pairs of conditions be equal, was assessed using Mauchly's test. In cases where sphericity was violated, Greenhouse–Geisser and Huynh–Feldt corrections were applied to adjust the degrees of freedom. Effect sizes were reported using generalized eta squared (η^2). All analyses were performed in R, with a significance threshold of $p < 0.05$.

3. Results

3.1. Variable Screening for Modeling

The variables described in Section 2.4 (Table 2) include 20 remote sensing metrics derived from Sentinel-2 data, covering spectral indices and reflectance bands commonly used in vegetation analysis. The Boruta feature selection procedure (Figure 3a,b) identified the most relevant predictors by comparing each variable's importance with the importance of randomized “shadow” features [38]; however, NDVIre783, band2, and NDVIre740 and band3 in the unitemporal dataset and NDVIre783, band 3, and band5, in the bitemporal dataset did not meet this threshold and were therefore marked for exclusion. The predictors were then evaluated using a stepwise VIF diagnostic.

3.2. Model Performance

3.2.1. Model Performance on Test Data at Plot-Level

On the Dataset A ($t_1 - t_2$) testing data, using the initial Boruta feature selection setup, the bitemporal RF model achieved $R^2 = 0.26$, RMSE of $4.06 \text{ m}^3/\text{ha}$, and MAE of $3.26 \text{ m}^3/\text{ha}$ when predicting average annual plot-level changes between $t_2 - t_1$ (Dataset B) (Figure 4a). In contrast, the unitemporal model yielded $R^2 = 0.40$, RMSE of $3.64 \text{ m}^3/\text{ha}$, and MAE of $2.65 \text{ m}^3/\text{ha}$ (Figure 4b), indicating improved explanatory power and reduced prediction error relative to the bitemporal approach.

3.2.2. Temporal Dynamics of Model Prediction at Plot Level

At the GSV plot-level, the unitemporal model (Figure 5b) showed greater predictive accuracy than the bitemporal model (Figure 5c) for GSV change $t_2 - t_3$. Relative to the bitemporal model, it explains 105% more variance and reduces prediction error by approximately 45%, indicating substantially stronger explanatory power and precision. The plot-level Dataset B GSV change predictions between t_2 and t_3 were added to the t_2 plot-level GSV to obtain updated t_3 GSV predictions (Equation (4)). Predictive performance at t_3 was assessed for both plot-level GSV change and plot-level GSV.

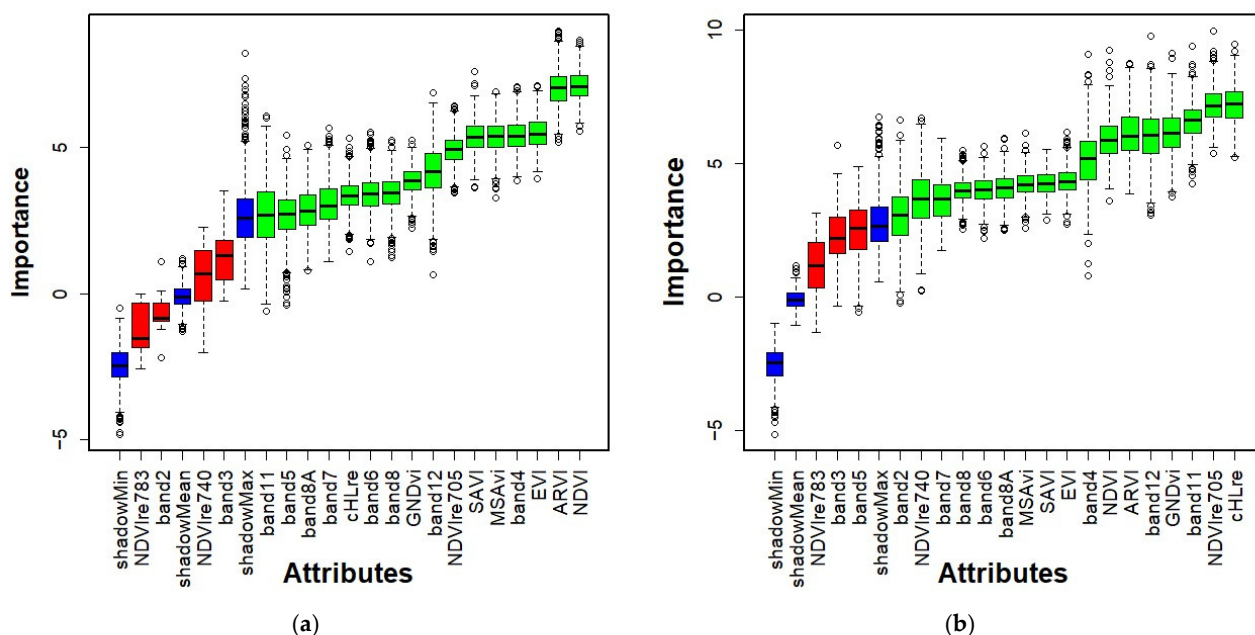


Figure 3. Boruta Feature Selection for unitemporal (a) and bitemporal (b). Green indicates confirmed important variables, red indicates rejected variables, and blue represents shadow attributes used as a reference for importance.

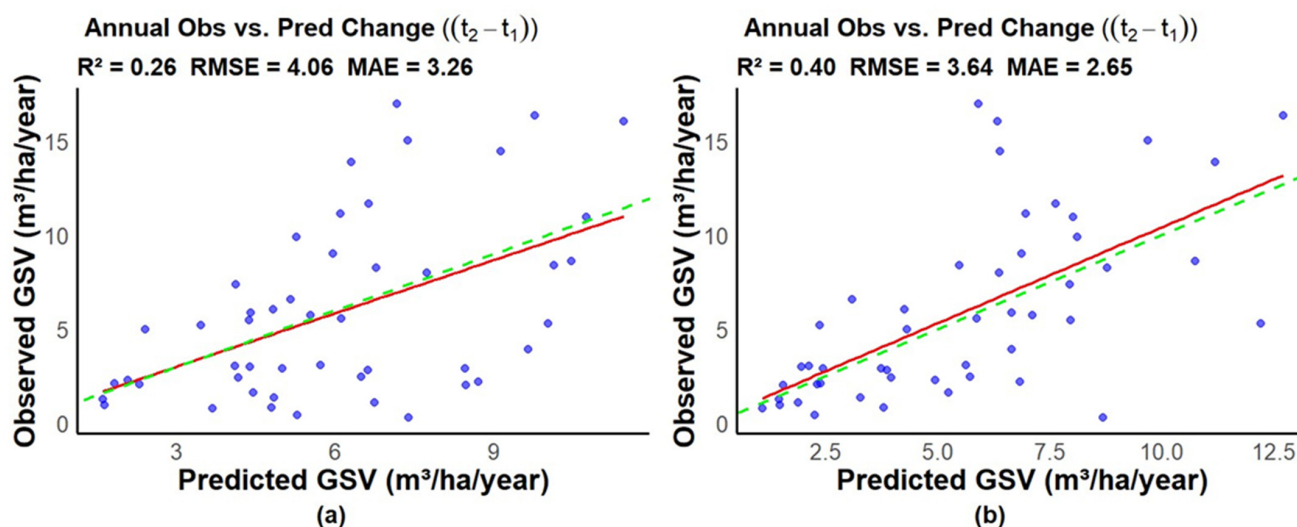


Figure 4. Scatter plot of observed versus predicted Dataset A, t_1 to t_2 , plot-level change for bitemporal (a) and unitemporal (b). Observed versus predicted GSV change per year, shown as blue points, with the fitted linear regression indicated by a red line and the 1:1 reference relationship shown as a green dashed line.

For the GSV change prediction assessment, R^2 , RMSE, and scatter plots were used. The scatter plots were constructed by graphing the observed plot-level GSV change from t_2 to t_3 versus observed plot-level GSV from t_1 to t_2 , consistent with the MAACV approach, versus unitemporal RF GSV t_2 to t_3 change predictions, and versus the bitemporal RF GSV t_2 to t_3 change predictions (Figure 5). The change prediction was performed with the split rule fixed to “variance” to minimize residual error. Algorithm performance during tuning was evaluated using 10-fold cross-validation, with hyperparameters selected to minimize RMSE and maximize R^2 . As shown in Figure 5a, the MAACV comparison yielded $R^2 = 0.49$ and RMSE = 19.76 m³/ha, indicating moderate temporal consistency in growth dynamics. The results for predicting GSV change in Dataset B show the RF unitemporal and

bitemporal models (Figure 5b,c) achieved $R^2 = 0.82$, $RMSE = 10.87 \text{ m}^3/\text{ha}$, and $R^2 = 0.40$, $RMSE = 19.65 \text{ m}^3/\text{ha}$, respectively.

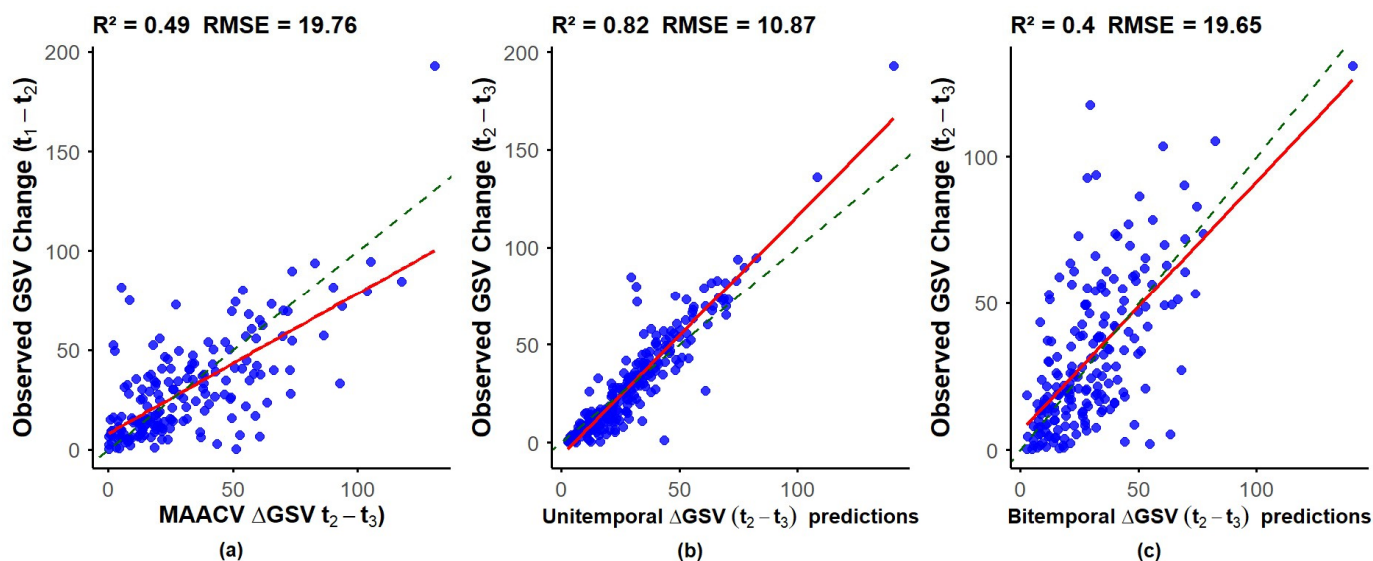


Figure 5. Plot-level comparison of observed and predicted GSV changes: (a) MAAVC, (b) unitemporal RF prediction, and (c) bitemporal RF prediction. Observed versus predicted GSV annual change, shown as blue points, with the fitted linear regression indicated by a red line and the 1:1 reference relationship shown as a green dashed line.

For the GSV prediction assessment, model performance was evaluated using R^2 and RMSE, and scatter plots of observed versus predicted plot-level GSV at t_3 were constructed for each method (Figure 6). The MAAVC predictions had $R^2 = 0.91$ and $RMSE = 45.11 \text{ m}^3/\text{ha}$, while the RF unitemporal and bitemporal predictions had $R^2 = 0.73$, $RMSE = 83.79 \text{ m}^3/\text{ha}$ and $R^2 = 0.72$, $RMSE = 83.61 \text{ m}^3/\text{ha}$, respectively. All predictions were based on Equation (4). Although the MAAVC shows superior accuracy at this population prediction, this comparison is presented to assess the feasibility of RS-based methods as an alternative when field measurements are not available.

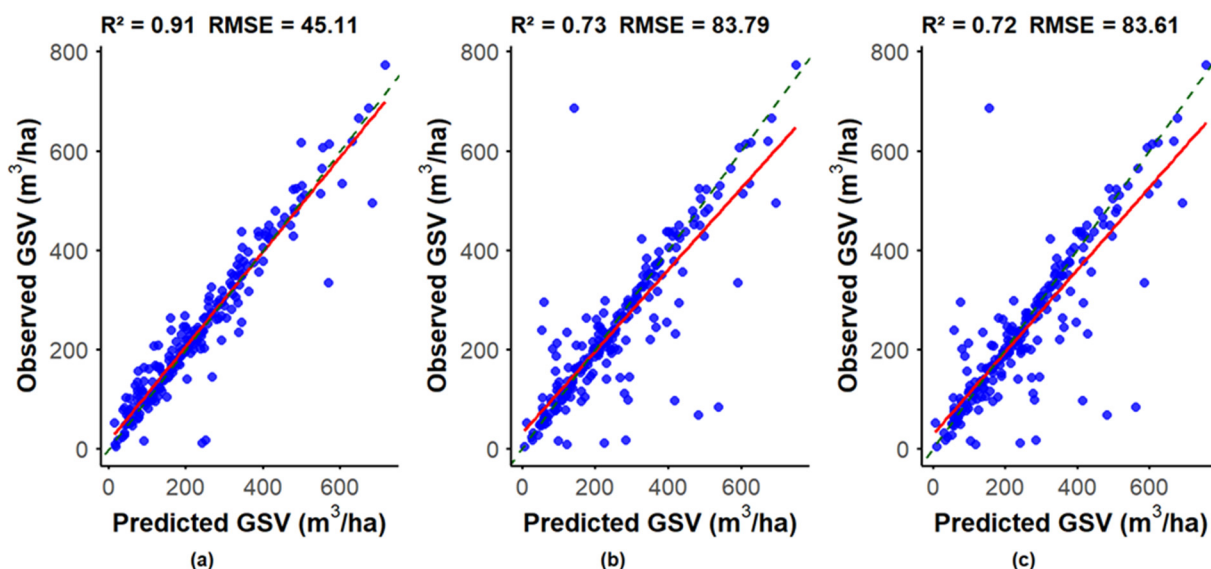


Figure 6. Plot-level comparison of observed and predicted GSV at t_3 : (a) MAAVC, (b) unitemporal RF prediction, and (c) bitemporal RF prediction. Observed versus predicted GSV at t_3 with the fitted linear regression indicated by a red line and the 1:1 reference relationship shown as a green dashed line.

3.3. Pairwise Difference Matrix for the Estimate of Mean Population GSV at t_3

Comparative performance of the prediction methods relative to the estimate of the population's mean GSV at t_3 is summarized in Table 3. The table presents the estimated population means for each method, alongside pairwise differences expressed in both absolute (m^3/ha) and relative (%) terms, with the observed data used as the control.

Table 3. Estimate of the population's mean GSV at t_3 with pairwise differences among the methods.

Reference\Estimate	GSV2023	MAAVC	Unitemporal	Bitemporal
Population mean $\pm$ SE (m^3/ha)	240.56 $\pm$ 10.08	237.01 $\pm$ 10.07	254.61 $\pm$ 10.48	255.19 $\pm$ 10.40
GSV2023	—	−3.55 (−1.48%)	+14.05 (+5.84%)	+14.63 (+6.08%)
MAAVC	+3.55 (+1.50%)	—	+17.60 (+7.43%)	+18.18 (+7.67%)
Unitemporal	−14.05 (−5.52%)	−17.60 (−6.91%)	—	+0.58 (+0.23%)
Bitemporal	−14.63 (−5.73%)	−18.18 (−7.12%)	−0.58 (−0.23%)	—

Estimated population means are descriptive (mean $\pm$ SE) and summarize overall GSV for each method. Matrix values represent model-predicted pairwise differences in GSV (m^3/ha), with percentage differences shown in parentheses. Positive values indicate greater GSV for the column method relative to the row method. The statistical significance of differences was evaluated using a linear mixed-effects model with plot included as a random intercept. The model showed a significant effect of the method, $F(3, 654) = 8.66$, $p < 0.001$ (Tables 4 and A1). None of the MAAVC, unitemporal, and bitemporal estimates of mean population GSV are statistically significantly different than the estimate based on t_3 plot observations (GSV2023), thereby supporting an assertion of unbiasedness of the MAAVC, unitemporal, and bitemporal approaches.

Table 4. Overall model test (linear mixed-effects model).

	Sum Sq	Mean Sq	NumDF	DenDF	F	Pr (>F)
Method	58,246	19,415	3	654	8.6562	1.222×10^{-5} ***

Signif. codes: *** $p < 0.001$.

Holm-adjusted post hoc [45] comparisons indicated that unitemporal and bitemporal methods produced significantly greater GSV estimates at t_3 than both MAAVC ($ps < 0.001$) and GSV2023 ($ps < 0.01$), whereas no significant differences were observed between unitemporal and bitemporal or between MAAVC and observed GSV at t_3 . Effect sizes for significant contrasts were small; paired Cohen's $ds = 0.23$ – 0.24 (Table A2).

3.4. Bootstrap-Based Hybrid Uncertainty Analysis RF Models

Bootstrap resampling combined with Monte Carlo simulation was used to estimate the uncertainty associated with the RF-based population mean GSV t_3 estimates. Convergence of the estimated standard error was assessed using the relative change in standard error as a function of iteration number. Relative changes in the estimated standard error decreased rapidly with increasing number of iterations and stabilized below the predefined tolerance threshold for both unitemporal and bitemporal RF-based models (Figure 7a,b).

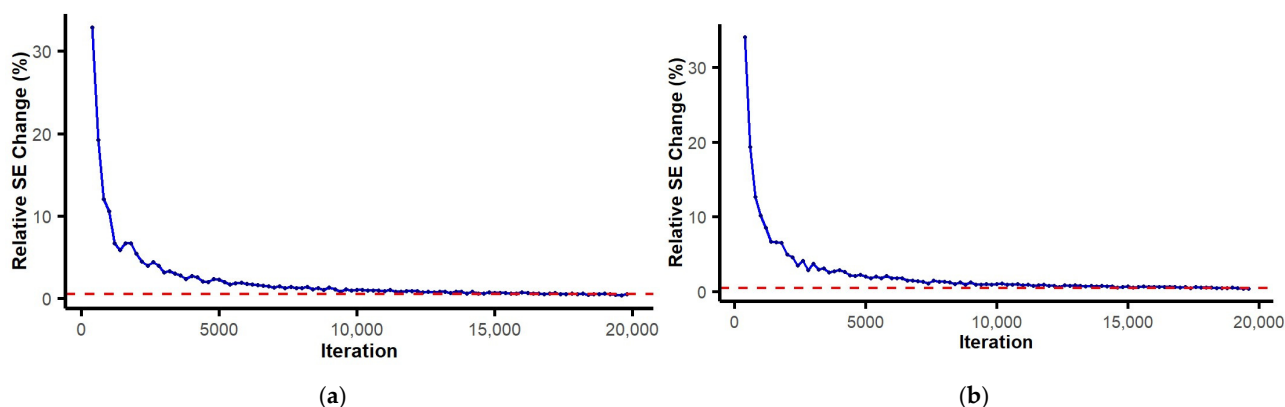


Figure 7. Bootstrap convergence diagnostics for the (a) unitemporal, and (b) bitemporal. The dashed horizontal line indicates the convergence tolerance threshold.

Stabilization was achieved after 19,600 iterations for the unitemporal and 19,800 iterations for the bitemporal model. The associated standard errors is presented in Table 5.

Table 5. Bootstrap uncertainty decomposition for random forest models.

Model (RF)	MC Iterations	Bootstrap Mean (m ³ /ha)	Model SE (m ³ /ha)	Sampling SE (m ³ /ha)	Total SE (m ³ /ha)
Unitemporal	19,600	255.09	1.22	10.41	10.48
Bitemporal	19,800	255.04	1.30	10.33	10.40

4. Discussion

Traditional and RS approaches to predicting forest variables are complementary in most cases [7]. Field methods (e.g., plot sampling, harvest records) provide accurate values for forest variables through direct observation [46] but are labor-intensive and costly, with limited temporal continuity in large areas [47,48]. By contrast, RS provides spatially explicit, wall-to-wall assessments of forest variables [26]. This study explored the potential of readily available Sentinel-2-derived unitemporal and bitemporal remote sensing metrics, coupled with RF prediction and a MAAVC approach to estimate plot and population-level GSV (m³/ha) changes in Estonian forests over a 5-year period. The findings highlighted a promising outcome for GSV change assessment and modeling using satellite-based predictors, especially with multitemporal observation when field inventories are not available or difficult to implement. The methodological choice was necessitated by the Sentinel-2 mission's operational timeline, as systematic image acquisition began only in 2015, precluding direct correspondence with the 2013 NFI cycle. This approach aligns with established practices: Puliti et al. [27] used 2015 Sentinel-2 imagery as a surrogate for 2014 field data, while Condés and McRoberts demonstrated the feasibility of using imagery acquired approximately 2–4 years before or after NFI cycles.

A more notable result from the Boruta feature selection is the exclusion of NDVIre783, band2, NDVIre740, and band3 in the unitemporal dataset, and NDVIre783, band3, and band5 in the bitemporal dataset, as their importance values did not exceed the maximum importance of the corresponding randomized shadow features (Figure 3a,b). According to Zhu and Liu [49], the use of optical satellite data acquired during the peak growing season may lead to spectral saturation, resulting in reduced accuracy in forest variable estimation. Previous studies have shown that inclusion of red-edge bands and NDVIre-based vegetation indices improves the magnitude of the relationship between forest variables and optical remote sensing metrics, particularly Sentinel-2 metrics [24,50]. For example, Zhou and Feng [24] estimated forest stock volume (FSV) using Sentinel-2 and Landsat

8 OLI imagery in combination with forest inventory data, and found models based on Sentinel-2 achieved higher accuracy than those using Landsat 8. They further reported that the red-edge bands of Sentinel-2 showed stronger correlations with FSV and had the potential to reduce model error. In contrast, the exclusion of several red-edge and spectral variables in the present study suggests that these predictors contributed limited or unstable information, particularly when temporal information was incorporated, indicating possible redundancy or saturation effects.

In predicting average annual changes, the unitemporal RF model explained more variance than the bitemporal model ($R^2 = 0.40$ versus $R^2 = 0.26$), representing approximately a 54% increase in explained variance in the Dataset A ($t_2 - t_1$) test data. This difference may be related to the timing of satellite acquisitions relative to the field measurement interval, as remotely sensed and field data are rarely temporally coincident, and forests may change between acquisition dates, which can influence the correspondence between spectral change signals and field-observed growth [51,52]. Fayad et al. [34] noted that, although there were minor temporal and spatial mismatches between field plots and remote sensing data, such discrepancies are unlikely to bias large-scale regression estimates because field measurements capture broad environmental trends in AGB rather than short-term forest dynamics. Notably, several studies show that, despite temporal mismatches, considerable accuracy can still be achieved [15,27,34]. Interestingly, among the results for annual change prediction, change prediction, and GSV estimation at t_3 (Figures 4–6), consistent patterns emerged. For both the annual and interval GSV change, the unitemporal RF model demonstrated greater performance than MAACV and the bitemporal RF configuration, which showed comparable error levels. In contrast, for GSV estimation at t_3 , the two RF models performed similarly, and MAACV achieved the strongest overall performance. Unitemporal Sentinel-2 metrics performed best for predicting GSV change, whereas the three approaches showed similar performance for predicting GSV at t_3 . Puliti et al. [27] assessed the aboveground biomass change using NFI data with Sentinel-2 and Landsat metrics in a boreal forest in south-eastern Norway. In their study, model-assisted estimation using bitemporal Sentinel-2 data produced the most precise estimate with a standard error (SE) = 1.7 Gg when compared to the unitemporal approach, which produced SE = 1.8 Gg. They found that using bitemporal data resulted in only a slight increase in precision compared to unitemporal data; however, they concluded that bitemporal data are the most precise overall and noted that Δ AGB can also be estimated when remotely sensed data are available at the end of the monitoring period. Similar results were also reported by McRoberts et al. [6]. In their study, using indirect and direct estimation of forest biomass change based on forest inventory and airborne laser scanning data, they discovered that the use of the ALS auxiliary information greatly increased the precision of change estimates, regardless of whether indirect or direct methods were used.

For the plot-level GSV change between t_2 and t_3 (Figure 5), the unitemporal RF model demonstrated greater performance ($R^2 = 0.82$, RMSE = 10.87) than both MAACV ($R^2 = 0.49$, RMSE = 19.76) and the bitemporal RF configuration ($R^2 = 0.40$, RMSE = 19.65). The latter two approaches produced nearly identical RMSE values, although MAACV explained more variance modestly. These results demonstrate that Sentinel-2 metrics are effective for predicting GSV change. A linear model (Equation 4), which uses input from MAACV and RS unitemporal and bitemporal approaches, was applied to estimate population- and plot-level GSV (m^3/ha) at t_3 within the study window. This model combined the field-derived MAACV of Equation (4) with field data from t_2 . The results were then compared with the predictions obtained using RS-based approaches (Figure 6). In the Dataset B analysis, the MAACV model showed greater precision than the RS-based models, as indicated by its R^2 of 0.91 and RMSE of $45.11 \text{ m}^3/\text{ha}$, compared to the RS-unitemporal model with R^2 of 0.73

and RMSE of 83.79 m³/ha, and bitemporal with R² of 0.72 and RMSE of 83.61 m³/ha. These results indicate that although MAAVC provides the most precise predictions, Sentinel-2 auxiliary metrics can still be effectively used to update GSV at the plot level using the temporal approach. At the population level, the temporal model produced comparable mean GSV estimates at t_3 ; a linear mixed-effects model revealed a significant effect of the proposed method on the GSV population mean at t_3 ($F(3, 654) = 8.66, p < 0.001$). Post hoc comparisons showed that MAAVC and GSV were significantly less than uni- and bitemporal population-level estimates at t_3 , with small effect sizes (Table A3).

Various concerted research efforts have been geared towards estimating forest variable changes using repeated data or single data [27,53]. For example, Noordermeer et al. [53] showed that bitemporal data acquired as part of repeated ALS-based forest inventories can be used to classify various changes in forest structure reliably. However, because change is derived from the difference between two error-prone values, it is inherently difficult to predict, and predictions of changes in a response variable are often less precise than predictions of the variable itself [6]. According to FAO [3], integrating uncertainty estimation requires further emphasis in NFI workflows.

Adherence to the IPCC Good Practice Guidance requires that estimates minimize bias and reduce uncertainty to the degree possible [43]. Such uncertainty stems from the inherent variability associated with random sampling and limited sample sizes. These variations lead to sampling uncertainty, which can be estimated through repeated simulations, such as bootstrap resampling. However, McRoberts [44] noted that a key concern with bootstrapping is the number of replications (B) required for $\hat{SE}_B(\hat{\mu})$ to approximate $SE(\hat{\mu})$ closely enough, and, more importantly, the criterion used to determine B and therefore proposed a statistically robust stopping criterion for Monte Carlo iterative procedures based on the stabilization of the SE estimate. In this study, we adopted this criterion to ensure convergence of the uncertainty of the mean GSV estimate. We further enhanced the stopping rule by introducing an additional requirement that the convergence threshold be met in three consecutive checks prior to terminating the simulation (Equation (12)). Applying this rule increased the number of iterations required for convergence but provided more reliable variance estimates. Stabilization was achieved after 19,600 iterations (Figure 7a) for the unitemporal and 19,800 iterations (Figure 7b) for the bitemporal RF models, ensuring stable variance estimates (Table 5). The Monte Carlo estimated mean for the unitemporal model was 255.09 m³/ha (Total SE = 10.48 m³/ha), while the bitemporal model produced a very similar estimated mean of 255.04 m³/ha (Total SE = 10.40 m³/ha). As shown in Table 5, sampling variability accounts for a larger portion of the total standard error than model prediction uncertainty. This occurs even in cases where model prediction accuracy is not large, a trend that has been noted in previous studies [54]. Melo et al. [55] demonstrated, using a hybrid bootstrap framework, that sampling variability represents the main source of uncertainty in short-term predictions, exceeding the contribution of model-related variance.

Overall, the enhanced Monte Carlo stopping rule provided stable uncertainty estimates and showed that, even with a slight difference in deviation at the plot level, the temporal RS approach still produces population-level GSV estimates that are fully consistent with both the design-based reference mean and the MAAVC model.

5. Conclusions

This study demonstrates the effectiveness of integrating temporal remote sensing metrics with RF algorithms for monitoring changes and updating forest GSV using NFI data. Four main conclusions were drawn. First, the unitemporal RF configuration demonstrated greater performance in predicting GSV change, whereas MAACV and the bitemporal RF

model exhibited comparable error levels. Second, for the GSV estimation at t_3 , the two RF configurations performed similarly, and MAACV demonstrated greater performance. Third, the consistent predictive performance of temporal RS with GSV estimation at t_3 indicates that Sentinel-2–derived metrics provide a reliable basis for updating GSV and enabling spatially explicit forest monitoring. Lastly, in estimating uncertainty, the use of objective Monte Carlo stoppage procedures is more defensible than arbitrarily selecting the number of iterations. The proposed methodology offers a reliable framework for bridging temporal gaps in national forest inventory data collection and supporting reporting obligations. These conclusions should be interpreted in the context of data availability and acquisition timing, as differences between satellite acquisition dates and field measurement intervals may influence model performance.

Author Contributions: Conceptualization, T.O.O., A.S. and R.E.M.; methodology, T.O.O., A.S. and R.E.M.; formal analysis, T.O.O., A.S., R.E.M. and M.A.-E.; writing—original draft preparation, T.O.O.; writing—review and editing, T.O.O., R.E.M. and M.A.-E.; supervision, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Estonian Research Council, grant number PRG2214.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We extend our sincere thanks to Copernicus open-access data for making the satellite data available. Additionally, we appreciate the efforts of all authors whose contributions were vital to the development and completion of this paper.

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

Appendix A. Tests of Significance for Differences in Volume Among Methods

Statistical tests were conducted using a linear mixed-effects model with plot included as a random intercept. Degrees of freedom for fixed effects were estimated using the Satterthwaite approximation (Table A1). Effect sizes are reported as paired Cohen’s d (Table A2), Pairwise comparisons were adjusted for multiple testing using the Holm method (Table A3).

Table A1. Estimated mean GSV by method based on linear mixed-effects model.

Method	Emmean	SE	df	Lower.CL	Upper.CL
GSV_2023	241	10.2	253	220	261
MAAVC	237	10.2	253	217	257
UNI_2023	255	10.2	253	234	275
BI_2023	255	10.2	253	235	275

Table A2. Effect sizes for significant contrasts at t_3 .

Contrast (Groups)	Cohen’s d	Magnitude
MAAVC–Unitemporal	−0.234	Small
MAAVC–Bitemporal	−0.244	Small

Table A3. Holm-adjusted pairwise comparisons.

Contrast	Estimate (m ³ /ha)	SE	df	t	p
GSV2023–MAAVC	3.55	4.53	654	0.78	0.866
GSV2023–Unitemporal	−14.05	4.53	654	−3.10	0.006
GSV2023–Bitemporal	−14.62	4.53	654	−3.23	0.005
MAAVC–Unitemporal	−17.60	4.53	654	−3.89	<0.001
MAAVC–Bitemporal	−18.17	4.53	654	−4.02	<0.001
Unitemporal–Bitemporal	−0.57	4.53	654	−0.13	0.899

Degrees-of-freedom method: Kenward–Roger *p*-value adjustment. Holm method for 6 tests.

References

- Vidal, C.; Alberdi, I.A.; Hernández Mateo, L.; Redmond, J.J. (Eds.) *National Forest Inventories*; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-44014-9.
- Tomppo, E.; Gschwantner, T.; Lawrence, M.; McRoberts, R.E. (Eds.) *National Forest Inventories: Pathways for Common Reporting*; Springer: Dordrecht, The Netherlands, 2010; ISBN 978-90-481-3232-4.
- FAO. *Voluntary Guidelines on National Forest Monitoring*; FAO: Rome, Italy, 2017; ISBN 978-92-5-109619-2.
- FAO. *Global Forest Resources Assessment 2020: Main Report*; FAO: Rome, Italy, 2020; ISBN 978-92-5-132974-0. [[CrossRef](#)]
- Fridman, J.; Holm, S.; Nilsson, M.; Nilsson, P.; Ringvall, A.; Ståhl, G. Adapting National Forest Inventories to Changing Requirements—The Case of the Swedish National Forest Inventory at the Turn of the 20th Century. *Silva Fenn.* **2014**, *48*. [[CrossRef](#)]
- McRoberts, R.E.; Næsset, E.; Gobakken, T.; Bollandsås, O.M. Indirect and Direct Estimation of Forest Biomass Change Using Forest Inventory and Airborne Laser Scanning Data. *Remote Sens. Environ.* **2015**, *164*, 36–42. [[CrossRef](#)]
- Omoniyi, T.O.; Sims, A. Enhancing the Precision of Forest Growing Stock Volume in the Estonian National Forest Inventory with Different Predictive Techniques and Remote Sensing Data. *Remote Sens.* **2024**, *16*, 3794. [[CrossRef](#)]
- Lindgren, N.; Wästlund, A.; Bohlin, I.; Nyström, K.; Nilsson, M.; Olsson, H. Updating of Forest Stand Data by Using Recent Digital Photogrammetry in Combination with Older Airborne Laser Scanning Data. *Scand. J. For. Res.* **2021**, *36*, 401–407. [[CrossRef](#)]
- FAO; UNEP. *The State of the World's Forests 2020*; FAO and UNEP: Rome, Italy, 2020; ISBN 978-92-5-132419-6.
- Gschwantner, T.; Alberdi, I.; Bauwens, S.; Bender, S.; Borota, D.; Bosela, M.; Bouriaud, O.; Breidenbach, J.; Donis, J.; Fischer, C.; et al. Growing Stock Monitoring by European National Forest Inventories: Historical Origins, Current Methods and Harmonisation. *For. Ecol. Manag.* **2022**, *505*, 119868. [[CrossRef](#)]
- Xie, B.; Cao, C.; Xu, M.; Bashir, B.; Singh, R.P.; Huang, Z.; Lin, X. Regional Forest Volume Estimation by Expanding LiDAR Samples Using Multi-Sensor Satellite Data. *Remote Sens.* **2020**, *12*, 360. [[CrossRef](#)]
- Takahashi, T.; Yamamoto, K.; Senda, Y.; Tsuzuku, M. Predicting Individual Stem Volumes of Sugi (*Cryptomeria japonica* D. Don) Plantations in Mountainous Areas Using Small-Footprint Airborne LiDAR. *J. For. Res.* **2005**, *10*, 305–312. [[CrossRef](#)]
- White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [[CrossRef](#)]
- Lee, W.-J.; Lee, C.-W. Forest Canopy Height Estimation Using Multiplatform Remote Sensing Dataset. *J. Sens.* **2018**, *2018*, 1–9. [[CrossRef](#)]
- Condés, S.; McRoberts, R.E. Updating National Forest Inventory Estimates of Growing Stock Volume Using Hybrid Inference. *For. Ecol. Manag.* **2017**, *400*, 48–57. [[CrossRef](#)]
- Lang, M.; Sims, A.; Pärna, K.; Kangro, R.; Möls, M.; Möistus, M.; Kiviste, A.; Tee, M.; Vajakas, T.; Rennel, M. Remote-Sensing Support for the Estonian National Forest Inventory, Facilitating the Construction of Maps for Forest Height, Standing-Wood Volume, and Tree Species Composition Kaugseirel Põhinev Lahendus Eesti Statistilise Metsainventuuri Jaoks Puistute Kõrguse, Tüvemahu Ja Liigilise Koosseisu Kaartide Koostamiseks. *For. Stud.* **2020**, *73*, 77–97. [[CrossRef](#)]
- Li, G.; Xie, Z.; Jiang, X.; Lu, D.; Chen, E. Integration of ZiYuan-3 Multispectral and Stereo Data for Modeling Aboveground Biomass of Larch Plantations in North China. *Remote Sens.* **2019**, *11*, 2328. [[CrossRef](#)]
- Du, C.; Fan, W.; Ma, Y.; Jin, H.I.; Zhen, Z. The Effect of Synergistic Approaches of Features and Ensemble Learning Algorithm on Aboveground Biomass Estimation of Natural Secondary Forests Based on Als and Landsat 8. *Sensors* **2021**, *21*, 5974. [[CrossRef](#)] [[PubMed](#)]

19. Asner, G.P.; Knapp, D.E.; Martin, R.E.; Tupayachi, R.; Anderson, C.B.; Mascaro, J.; Sinca, F.; Chadwick, K.D.; Higgins, M.; Farfan, W.; et al. Targeted Carbon Conservation at National Scales with High-Resolution Monitoring. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5016–E5022. [\[CrossRef\]](#)
20. Loveland, T.R.; Irons, J.R. Landsat 8: The Plans, the Reality, and the Legacy. *Remote Sens. Environ.* **2016**, *185*, 1–6. [\[CrossRef\]](#)
21. Chen, Q.; Gong, P.; Baldocchi, D.; Tian, Y.Q. Estimating Basal Area and Stem Volume for Individual Trees from Lidar Data. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1355–1365. [\[CrossRef\]](#)
22. Zheng, S.; Cao, C.; Dang, Y.; Xiang, H.; Zhao, J.; Zhang, Y.; Wang, X.; Guo, H. Retrieval of Forest Growing Stock Volume by Two Different Methods Using Landsat TM Images. *Int. J. Remote Sens.* **2014**, *35*, 29–43. [\[CrossRef\]](#)
23. Ni, W.; Zhang, Z.; Sun, G.; Liu, Q. Modeling the Stereoscopic Features of Mountainous Forest Landscapes for the Extraction of Forest Heights from Stereo Imagery. *Remote Sens.* **2019**, *11*, 1222. [\[CrossRef\]](#)
24. Zhou, Y.; Feng, Z. Estimation of Forest Stock Volume Using Sentinel-2 MSI, Landsat 8 OLI Imagery and Forest Inventory Data. *Forests* **2023**, *14*, 1345. [\[CrossRef\]](#)
25. Clark, M.L. Comparison of Multi-Seasonal Landsat 8, Sentinel-2 and Hyperspectral Images for Mapping Forest Alliances in Northern California. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 26–40. [\[CrossRef\]](#)
26. Mura, M.; Bottalico, F.; Giannetti, F.; Bertani, R.; Giannini, R.; Mancini, M.; Orlandini, S.; Travaglini, D.; Chirici, G. Exploiting the Capabilities of the Sentinel-2 Multi Spectral Instrument for Predicting Growing Stock Volume in Forest Ecosystems. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *66*, 126–134. [\[CrossRef\]](#)
27. Puliti, S.; Breidenbach, J.; Schumacher, J.; Hauglin, M.; Klingenberg, T.F.; Astrup, R. Above-Ground Biomass Change Estimation Using National Forest Inventory Data with Sentinel-2 and Landsat. *Remote Sens. Environ.* **2021**, *265*, 112644. [\[CrossRef\]](#)
28. Suleymanov, A.; Shagaliev, R.; Belan, L.; Bogdan, E.; Tuktarova, I.; Nagaev, E.; Muftakhina, D. Forest Growing Stock Volume Mapping with Accompanying Uncertainty in Heterogeneous Landscapes Using Remote Sensing Data. *Earth Sci. Inf.* **2024**, *17*, 5359–5369. [\[CrossRef\]](#)
29. Raukas, A. Briefly about Estonia. *Dyn. Environnement.* **2018**, *42*, 284–291. [\[CrossRef\]](#)
30. Keskkonnaagentuur. *Yearbook Forest 2018. (Aastaraamat Mets 2018)*; Raudsaar, M., Siimon, K.L., Valgepea, M., Eds.; Keskkonnaagentuur: Tallinn, Estonia, 2020.
31. Raudsaar, M.; Merenäkk, M.; Valgepea, M. *Yearbook Forest 2013*; Estonian Environment Agency: Tallinn, Estonia, 2014.
32. Kohava, P. *Forests in Estonia 1999 (Eesti Metsad 1999)*; Eesti Metsakorralduskeskus: Tartu maakond, Estonia, 2000; p. 44.
33. Sims, A. Principles of National Forest Inventory Methods. In *Definition and Uncertainty of Forests. Managing Forest Ecosystems*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 43.
34. Fayad, I.; Baghdadi, N.; Guitet, S.; Bailly, J.-S.; Hérault, B.; Gond, V.; El Hajj, M.; Tong Minh, D.H. Aboveground Biomass Mapping in French Guiana by Combining Remote Sensing, Forest Inventories and Environmental Data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 502–514. [\[CrossRef\]](#)
35. Blickensdörfer, L.; Oehmichen, K.; Pflugmacher, D.; Kleinschmit, B.; Hostert, P. National Tree Species Mapping Using Sentinel-1/2 Time Series and German National Forest Inventory Data. *Remote Sens. Environ.* **2024**, *304*, 114069. [\[CrossRef\]](#)
36. Chen, J.; Li, X.; Huang, Z.; Xuan, J.; Chen, C.; Hu, M.; Tan, C.; Zhou, Y.; Zhao, Y.; Yu, J.; et al. Forest Age Estimation Using UAV-LiDAR and Sentinel-2 Data with Machine Learning Algorithms—A Case Study of Masson Pine (*Pinus massoniana*). *Geo-Spat. Inf. Sci.* **2025**, *28*, 1051–1071. [\[CrossRef\]](#)
37. Brien, R.J.W.; Zuidema, P.A.; During, H.J. Autocorrelated Growth of Tropical Forest Trees: Unraveling Patterns and Quantifying Consequences. *For. Ecol. Manag.* **2006**, *237*, 179–190. [\[CrossRef\]](#)
38. Kursa, M.B.; Jankowski, A.; Rudnicki, W.R. Boruta—A System for Feature Selection. *Fundam. Informaticae* **2010**, *101*, 271–285. [\[CrossRef\]](#)
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
40. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [\[CrossRef\]](#)
41. Awad, M.; Fraihat, S. Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *J. Sens. Actuator Netw.* **2023**, *12*, 67. [\[CrossRef\]](#)
42. Corona, P.; Fattorini, L.; Franceschi, S.; Scrinzi, G.; Torresan, C. Estimation of Standing Wood Volume in Forest Compartments by Exploiting Airborne Laser Scanning Information: Model-Based, Design-Based, and Hybrid Perspectives. *Can. J. For. Res.* **2014**, *44*, 1303–1311. [\[CrossRef\]](#)
43. McRoberts, R.E.; Chen, Q.; Domke, G.M.; Ståhl, G.; Saarela, S.; Westfall, J.A. Hybrid Estimators for Mean Aboveground Carbon per Unit Area. *For. Ecol. Manag.* **2016**, *378*, 44–56. [\[CrossRef\]](#)
44. McRoberts, R.E.; Næsset, E.; Hou, Z.; Ståhl, G.; Saarela, S.; Esteban, J.; Travaglini, D.; Mohammadi, J.; Chirici, G. How Many Bootstrap Replications Are Necessary for Estimating Remote Sensing-Assisted, Model-Based Standard Errors? *Remote Sens. Environ.* **2023**, *288*, 113455. [\[CrossRef\]](#)
45. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.

46. Brown, S. Measuring Carbon in Forests: Current Status and Future Challenges. *Environ. Pollut.* **2002**, *116*, 363–372. [[CrossRef](#)] [[PubMed](#)]
47. Kangas, A.; Astrup, R.; Breidenbach, J.; Fridman, J.; Gobakken, T.; Korhonen, K.T.; Maltamo, M.; Nilsson, M.; Nord-Larsen, T.; Næsset, E.; et al. Remote Sensing and Forest Inventories in Nordic Countries—Roadmap for the Future. *Scand. J. For. Res.* **2018**, *33*, 397–412. [[CrossRef](#)]
48. Wang, X.; Zhang, C.; Qiang, Z.; Xu, W.; Fan, J. A New Forest Growing Stock Volume Estimation Model Based on AdaBoost and Random Forest Model. *Forests* **2024**, *15*, 260. [[CrossRef](#)]
49. Zhu, X.; Liu, D. Improving Forest Aboveground Biomass Estimation Using Seasonal Landsat NDVI Time-Series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231. [[CrossRef](#)]
50. Chrysafis, I.; Mallinis, G.; Siachalou, S.; Patias, P. Assessing the Relationships between Growing Stock Volume and Sentinel-2 Imagery in a Mediterranean Forest Ecosystem. *Remote Sens. Lett.* **2017**, *8*, 508–517. [[CrossRef](#)]
51. Oh, S.; Jung, J.; Shao, G.; Shao, G.; Gallion, J.; Fei, S. High-Resolution Canopy Height Model Generation and Validation Using USGS 3DEP LiDAR Data in Indiana, USA. *Remote Sens.* **2022**, *14*, 935. [[CrossRef](#)]
52. Knott, J.A.; Liknes, G.C.; Giebink, C.L.; Oh, S.; Domke, G.M.; McRoberts, R.E.; Quirino, V.F.; Walters, B.F. Effects of Outliers on Remote Sensing-assisted Forest Biomass Estimation: A Case Study from the United States National Forest Inventory. *Methods Ecol Evol.* **2023**, *14*, 1587–1602. [[CrossRef](#)]
53. Noordermeer, L.; Økseter, R.; Ørka, H.O.; Gobakken, T.; Næsset, E.; Bollandsås, O.M. Classifications of Forest Change by Using Bitemporal Airborne Laser Scanner Data. *Remote Sens.* **2019**, *11*, 2145. [[CrossRef](#)]
54. Li, C.; Yu, Z.; Dai, H.; Zhou, X.; Zhou, M. Effect of Sample Size on the Estimation of Forest Inventory Attributes Using Airborne LiDAR Data in Large-Scale Subtropical Areas. *Ann. For. Sci.* **2023**, *80*, 40. [[CrossRef](#)]
55. Melo, L.C.; Schneider, R.; Fortin, M. Estimating Model- and Sampling-Related Uncertainty in Large-Area Growth Predictions. *Ecol. Model.* **2018**, *390*, 62–69. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.