

Article

Enhancing the Precision of Forest Growing Stock Volume in the Estonian National Forest Inventory with Different Predictive Techniques and Remote Sensing Data

Temitope Olaoluwa Omoniyi * and Allan Sims 

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

* Correspondence: temitope.omoniyi@emu.ee

Abstract: Estimating forest growing stock volume (GSV) is crucial for forest growth and resource management, as it reflects forest productivity. National measurements are laborious and costly; however, integrating satellite data such as optical, Synthetic Aperture Radar (SAR), and airborne laser scanning (ALS) with National Forest Inventory (NFI) data and machine learning (ML) methods has transformed forest management. In this study, random forest (RF), support vector regression (SVR), and Extreme Gradient Boosting (XGBoost) were used to predict GSV using Estonian NFI data, Sentinel-2 imagery, and ALS point cloud data. Four variable combinations were tested: CO1 (vegetation indices and LiDAR), CO2 (vegetation indices and individual band reflectance), CO3 (LiDAR and individual band reflectance), and CO4 (a combination of vegetation indices, individual band reflectance, and LiDAR). Across Estonia's geographical regions, RF consistently delivered the best performance. In the northwest (NW), the RF model achieved the best performance with the CO3 combination, having an R^2 of 0.63 and an RMSE of 125.39 $m^3/plot$. In the southwest (SW), the RF model also performed exceptionally well, achieving an R^2 of 0.73 and an RMSE of 128.86 $m^3/plot$ with the CO4 variable combination. In the northeast (NE), the RF model outperformed other ML models, achieving an R^2 of 0.64 and an RMSE of 133.77 $m^3/plot$ under the CO4 combination. Finally, in the southeast (SE) region, the best performance was achieved with the CO4 combination, yielding an R^2 of 0.70 and an RMSE of 21,120.72 $m^3/plot$. These results underscore RF's precision in predicting GSV across diverse environments, though refining variable selection and improving tree species data could further enhance accuracy.



Citation: 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. <https://doi.org/10.3390/rs16203794>

Academic Editor: Fernando
José Aguilar

Received: 19 September 2024

Revised: 10 October 2024

Accepted: 10 October 2024

Published: 12 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forests and their ecosystems play a vital role in the overall health and sustainability of the planet Earth, providing indispensable services such as carbon sequestration, biodiversity conservation, and water regulation [1,2]. These ecosystems support a vast array of life-forming processes for both plant and animal species, contributing to ecological functionality, stability, and resilience [3,4]. In addition, they offer economic benefits through the provision of resources such as timber and non-timber forest products (NTFPs), highlighting their multifaceted importance [5]. Forest growing stock volume (GSV) is a critical metric in forestry resource assessments (FRAs), representing the total volume of all living trees within a given area [6,7] and serving as an indicator for measuring the forest management capability, forest quality, and carbon sequestration potential [8]. The accurate estimation of this essential variable is pivotal for sustainable forest management, conservation practices, and the quantification of ecosystem services, and aids in providing detailed information to support decision-making mechanisms [9,10]. Advances in remote sensing data-capturing

technology and machine learning (ML) algorithms have significantly enhanced the precision and efficiency of GSV assessments, causing a fundamental change in estimation and thereby providing valuable insights for resource management [11,12].

Traditional and remote sensing (RS)-based estimation methods in assessing GSV offer distinct advantages and present challenges, shaping their complementary roles in forest inventory and management. Traditional methods, such as field sampling and harvesting data, provide accurate measurements at localized scales by directly observing tree dimensions and timber volume [6]. However, the fieldwork for acquiring data for forest growing stock volume at a larger scale is challenging [13] and, holistically, these approaches are labor-intensive, costly, and may lack continuous coverage over large, forested areas [14]. Conversely, the RS method offers significant advancements by providing spatially explicit and continuous assessments of forest-related variables. With a long history of remote sensing data availability, different data sources have been utilized to estimate forest variables, out of which optical satellite imagery such as Landsat has been widely used [15]. The synoptic coverage, moderate ground cover resolution, high repetitive coverage, and free access policy make some optical RS methods the most widely used [15–17]. Meanwhile, while RS methods enhance efficiency and coverage, it is essential to note that they cannot emphatically replace field sampling methods. They require careful calibration and validation with ground-truth data from traditional methods to ensure accuracy and precision [18]. The integration of both approaches through hybrid methodologies facilitates comprehensive GSV assessments [19], leveraging the strengths of RS for broad-scale monitoring and the precision of traditional methods for validation and localized accuracy.

Optical satellites are capable of capturing detailed horizontal information on forest vegetation using surface reflectance patterns, but information on vertical structures is usually not detailed since these satellites cannot penetrate the canopy or measure height variations effectively [20,21]. However, recent advancements in forestry inventory now enable the use of terrestrial optical cameras to capture the forest's vertical structure. Recent advancements, like multi-view stereopsis (MVS), combine computer vision and photogrammetry to enable 3D reconstruction using standard optical cameras with algorithms such as SIFT and SURF [22,23]. Juujärvi et al. [24] developed a digital image-based tree measurement system for forest inventory, while Panagiotidis et al. [23] compared Structure from Motion (SfM) and terrestrial laser scanning for measuring DBH and height. Their study found that SfM accuracy was lower on the tree stems due to fewer matching points, particularly at greater heights, resulting in errors often exceeding 11 cm. On the contrary, LiDAR remote sensing can provide both horizontal and vertical information, depending on the configuration of the LiDAR system used, and therefore it is well suited for describing forest structure [22,25,26]. The use of optical imagery with high resolution for estimating GSV has proven to be effective, for example, Mura et al. [27] explored the capabilities of Sentinel-2 for predicting growing stock volume in Italy. Zhou and Feng [8] compared the precision of Sentinel-2 and Landsat 8 OLI in estimating forest stock volume and concluded that the Sentinel-2-based model achieved higher accuracy. Hence, we chose to integrate LiDAR metrics and Sentinel-2 data in our current study. Research efforts geared towards the usage of airborne laser scanning (ALS) in the estimation of GSV has proven its resourcefulness as a valuable data source [13,28–31]. Lindgren et al. [32] emphasized that in addition to ALS, there are many other sources of remote sensing data that could be used for automated prediction of forest variables like GSV. However, a lack of good prediction accuracy has reduced their use in operational forest management.

In providing more accurate, non-destructive frequent updates, remotely sensed data can improve the quality of forest inventory database; thus, the synergy between traditional and RS methods in GSV estimation not only enhances forest management practices but also supports sustainable resource utilization and conservation efforts on a global scale, thereby improving the resource management activities [19,33–36]. Integrating LiDAR and optical remote sensing data for estimating forest biophysical parameters presents some associated challenges necessitating careful attention, robust methodologies, and often a hybrid method

that leverages the strengths of both LiDAR and optical sensors while mitigating their limitations. Researchers have addressed these challenges through various innovative approaches, for example, Hudak et al. [37] proved that the integration of LiDAR and optical remote sensing data significantly improves forest inventory accuracy, most especially in predicting forest structure attributes when LiDAR is combined with multispectral imagery. Asner et al. [38] stressed the importance of ground-based calibration and validation to achieve accurate carbon stock estimates with airborne LiDAR. Fernández-Landa et al. [39] analyzed the effectiveness of combining national wide LiDAR information and optical imagery to estimate forest variables and concluded that this integration may lead to an increase in estimation accuracy, most especially when the species are homogenized.

Several modeling techniques, either parametric (e.g., multiple linear regression, logistic regression, geographically weighted regression) or non-parametric (e.g., k-nearest neighbors (k-NNs), random forests, artificial neural networks, gradient nearest neighbors (GNNs), support vector machines, and decision trees), have been used to establish the relationship between the response variables (forest variables) and predictor variables (remote sensing-derived variables) [8,28,40–44]. Machine learning algorithms such as random forests, support vector machines, Extreme Gradient Boosting, and artificial neural networks (ANNs) are some of the commonest ML algorithms used for forest variable estimation and they have demonstrated superiority and proven efficient in estimating GSV, outperforming traditional methods and highlighting their robustness across diverse datasets and varying forest ecological conditions [8,11,41,42,45,46].

The aim of our research was to assess the predictability of forest growing stock volume across various Cartesian planes in Estonia, corresponding to the flight schedules of ALS based on forestry purposes, by utilizing three ML algorithms to analyze NFI plot data alongside multimodal remote sensing data. We integrated ALS height percentiles with multispectral data to create a comprehensive dataset for the analysis and incorporated ground-truth data from NFI plots to enhance the accuracy of remote sensing data interpretations and model training. We applied and compared the performance of three distinct machine learning algorithms, RF, SVR, and XGBoost, to predict forest growing stock volume, focusing on their accuracy and reliability at the plot level to ensure localized accuracy in forest growing stock volume. By achieving these objectives, this research aimed to advance the understanding of forest growing stock volume estimation using multimodal remote sensing data and machine learning techniques, ultimately contributing to more effective and sustainable forest management in Estonia.

2. Materials and Methods

2.1. Study Area

The study area, Estonia, covers approximately 45,339 km² of land area and is located in Northern Europe on the shore of the Baltic Sea. It borders the Baltic Sea and the Gulf of Finland. Forests in Estonia account for about 53.6%, approximately 23,308 km², of its land area [47]. It is situated between latitudes 57°30'34"N to 59°49'12"N and longitudes 21°45'49"E and 28°12'44"E, as shown in Figure 1a. [48]. Forests in Estonia are defined by its Forest Act [49] and cover 2.2 million ha of its total land area, which accounts for approximately half of the county's land area [50]. The main common tree species that contributes majorly to the growing stock volume quantification are Scots pine (*Pinus sylvestris* L.), which account for about 34% coverage of the forest area; downy birch (*Betula Pubescens* Ehrh) and silver birch (*B. Pendula* Roth) dominating about 31% of the forested land; Norway spruce (*Picea abies* (L.) Karst), with a share of 16%; and species like European aspen (*Populus tremula* L.), grey alder (*Alnus incana* (L.) Moench), black alder (*A. glutinosa* (L.) Gaertn, and common ash (*Fraxinus excelsior* L) [50]. Their coverage in space is either linear or clustered, resulting in heterogenous or homogenous stands with an average stand size of 1.25 ha [28,51]. The mean annual temperature is 4.7 °C; in February, it is –6.6 °C, and in July, it is 16.3 °C, while the mean annual precipitation is 500 to 700 mm [48]. Thirteen (13) soil taxa were identified and categorized into automorphic soils (leptosols, cambisols,

luvisols, planosols, podzoluvisols, podzols, arenosols), different gleysols, and histosols (mires and bogs), fluvisols (alluvial and saline littoral soils), and technogenic formations on reclaimed land [52].

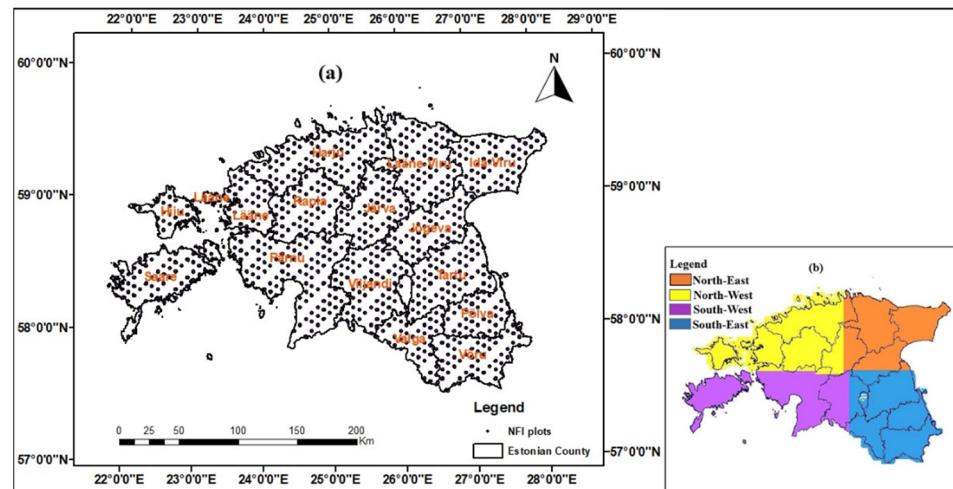


Figure 1. Cluster network (a) of the Estonia NFI permanent and temporary plot (2018–2022); Cartogram of the elevation model of the land cover (b).

2.2. General Overview of Methods

This study estimates the forest growing stock volume (GSV) by combining multi-sensor remote sensing and National Forest Inventory data using the RF, Extreme Gradient Boosting (XGBoost), and support vector regression (SVR) regression algorithms. The methodology flowchart is shown in Figure 2.

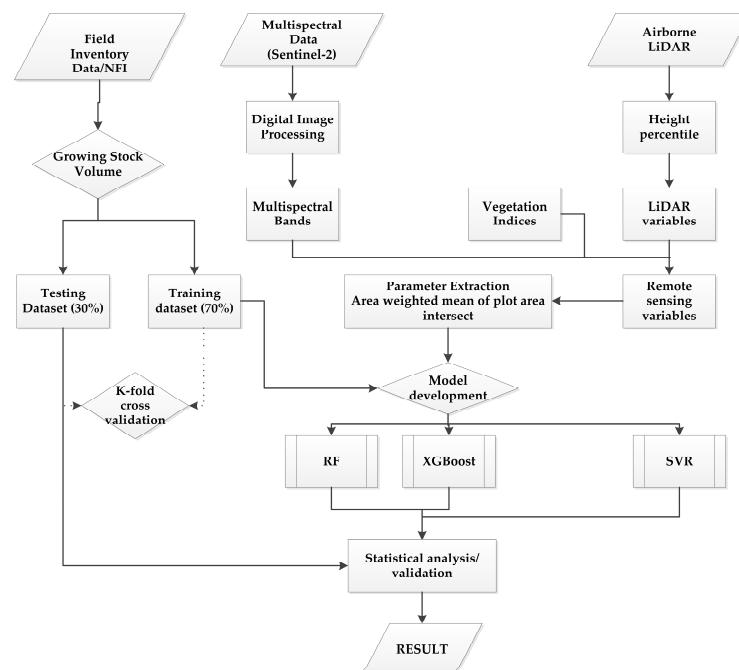


Figure 2. Methodology flowchart for this study.

2.3. Estonian National Forest Inventory

As part of the intensive forest management scheme, NFI is conducted in Estonia with a five-year repeat cycle. The first NFI covering the whole country dates back to 1999 [53] and was designed to give a description of Estonia's forests. Currently, the NFI gives more

robust information about the distribution of land by land-use classes, the afforestation and growing stock of non-forest land, etc. It is designed as an annual research effort, which, using optimal methods, enhances the continuous updating of its architectural database. A network of sample plots, covering the entire country with the same sampling intensity, has been designed for five years with 20% yearly coverage and a potential for increasing frequencies. This is designed in a way that allows one-fifth of the total sample plots to be inventory yearly and further enhance a revisit measurement of the permanent sample plots after each complete cycle (five years). The network of sample plots in the Estonian NFI is spatially distributed on a national grid of a $5\text{ km} \times 5\text{ km}$ quadrangle as determined by the L-EST coordinates system (EPSG:3301) with the same sampling intensity. Three types of circular sample plots with fixed radius are used in the NFI: (a) volume sample plots, (b) site category sample plots, and (c) regeneration and felling sample plots, but for the purpose of this study, volume-based sample plots are used. To enhance the efficiency of the survey, these circular sample plots are grouped into clusters, each covering an $800 \times 800\text{ m}$ square area. This clustering approach is used to optimize fieldwork and data collection. Detailed methodology for the NFI is described in the Estonia global forest resources assessment [54]. For this study, sample plots in the years 2018, 2019, 2020, 2021, and 2022 cycle were used.

2.4. Field Inventory Measurement

As an element of the thorough data acquisition strategies, a systematic sampling technique without pre-stratification was adopted in the Estonia NFI. Within each circular sample plot, the diameter at breast height (dbh) of trees with $\text{dbh} \geq 5\text{ cm}$ was recorded at 1.3 m height above the ground. Height was recorded for some trees according to the criteria that allow for dbh to height estimation with minimal prediction error in accordance with NFI criteria using Equation (1); information such as species, age, and sample plot center coordinates was also recorded within the 10 m and 7 m radii sample plot using a handheld GPS of up to 5 m accuracy. Trees' dbh was measured using a caliper with $\pm 1.0\text{ mm}$ accuracy on a single measurement method while height measurement was carried out using Vertex. Vertex measures precise distances using ultrasonic signals while heights are then determined trigonometrically based on these distance and angle measurements. The sample plots were further stratified into NE, SE, SW, and NW categories in order to align with the aerial laser scanning-based partitioning of Estonia (Figure 1b). A region-specific growing stock volume equation (Equation (2)) was used to estimate the volume of trees in each plot. The volume of the sample plots was then calculated by summing the volume of all the trees in the sample plot.

$$\text{Nc} = \frac{\text{BA}}{n} \quad (1)$$

$$\text{GSV} = (a_1 + a_2 \times h + a_3 \times f) \times (t/(t+5.0))^{(c_1+c_2 \times h)} \quad (2)$$

where Nc is the total tree count for height measurement; BA is the plot basal area in m^2/ha and n is the maximum number of trees in the plot; GSV is the growing stock volume of each plot; a_1 , a_2 , a_3 , c_1 , c_2 are species-based constants; h is the tree height; t is the age; and f is a dummy variable with a value of 1 if the sample plot is on any island in Estonia.

Outliers were detected and removed through the use of the PauTa criterion, where the outlier is statistically calculated and removed rather than through a manual specification of the outlier threshold. A threshold is determined according to a certain probability and any error exceeding this is taken to be a gross error rather than a random error. A numerical distribution interval of $(\mu - 3\sigma, \mu + 3\sigma)$ was used as this explained the largest proportion of any given data point [55]. Data points with values less than $\mu - 3\sigma$ or greater than $\mu + 3\sigma$ were deleted before the regression analysis to ensure that the data assumed a normal distribution. The interval $(\mu - 3\sigma, \mu + 3\sigma)$ represents the range in a normal distribution that includes about 99.7% of the data, where μ is the mean and σ is the standard deviation. Table 1 shows the geographical characterization of the GSV.

Table 1. Descriptive characteristics for GSV (m^3/plot).

Zone	No. of Plot	Min	Max	Mean	SD
N/W	3712	36.89	1007.01	531.42	162.36
S/W	3651	80.8	1038.24	573.22	153.58
N/E	3852	31.76	1101.03	569.59	175.65
S/E	3665	122.08	1101.03	625.40	147.57
All	14,880	31.76	1101.03	575.13	163.78

2.5. Remote Sensing Data

To establish the relationship between GSV and remote sensing data, this study utilized both optical satellite data (Sentinel-2) and ALS height percentile data. The Sentinel-2 images that cover each cartesian plane were downloaded from Copernicus Open Access Hub (<https://dataspace.copernicus.eu>) accessed on 23 February 2024. The images contained rich spectral information and were downloaded based on selection criteria such as maximum cloud cover <2% and sensor uniformity in cases where multiple sensors were available. A full representation of each plane contains a mosaic of a maximum of 4 or 3 scenes. The selected bands range from bands 2 through 8, 8A, 11, and 12 (Table 2). Each product was corrected from top-of-atmosphere (TOA) reflectance and processed into orthoimage bottom-of-atmosphere (BOA) surface reflectance using sen2cor version 2.11 of the Sentinel application platform (SNAP) and then resampled into 10 m resolution. Cloud and cloud shadow were determined and masked using ESA SNAP, and overlapping points on the pixel that were not properly classified were manually removed using R studio. Points were converted into shapefiles (20 m × 20 m vector box) using the bounding box technique and were used to extract the corresponding band reflectance values using a weighted mean approach which extracts the average value of pixels that interest with the shapefile.

Table 2. Source and description of variables.

Variable	Description
Sentinel-2 single band	
Blue	B2
Green	B3
Red	B4
Red edge 1	B5
Red edge 2	B6
Red edge 3	B7
NIR	B8
NIR narrow	B8A
SWIR	B11
SWIR	B12
Veg. indices	
NDVI	$(B08 - B04)/(B08 + B04)$
ARI1	$(1/B03)-(1/B05)$
MARI	$(1/B03)-(1/B05) \times B07$
ARVI	$(B8A - B04 - y \times (B04 - B02))/(B8A + B04 - y \times (B04 - B02))$
NDVIRE705	$(B08 - B05)/(B08 + B05)$
NDVIRE740	$(B08 - B06)/(B08 + B06)$
NDVIRE783	$(B08 - B07)/(B08 + B07)$
EVI	$2.5 \times (B08 - B04)/(B08 + 2.4 \times B04 + 1)$
SAVI	$(B08 - B04) \times (1 + L)/(B08 + B04 + L)$
MCARI	$((B05 - B04) - 0.2 \times (B05 - B03)) \times (B05/B04)$
Clgreen	$(B07)/(B03 - 1)$
ND11	$(B07 - B05)/(B07 + B05)$
chlre	$(B07)/(B05 - 1)$
GNDVI	$(B08 - B03)/(B08 + B03)$
MSAVI	$0.5 \times (2 \times B08 + 1 - \sqrt{(2 \times B08 + 1)^{-8} \times (B08 - B04)})$

Table 2. *Cont.*

Variable		Description
Airborne LiDAR	p25	Height percentile of 25%
	p50	Height percentile of 50%
	p60	Height percentile of 60%
	p70	Height percentile of 70%
	p75	Height percentile of 75%
	p80	Height percentile of 80%
	p90	Height percentile of 90%
	p95	Height percentile of 95%
	first_returns_above	number of first return laser pulses
	percentage_all	percentage of all laser pulses
	total_first_return	total number of first return laser pulses

2.6. LiDAR Data

In Estonia, airborne laser observation is carried out by the Estonian Land Board on a yearly basis during the spring and summer periods. For data optimization enhancement of the LiDAR data acquisition, the country is divided into four and a quarter is covered in the spring and a quarter in summer (Figure 1b). The summer flight program is designed to allow for forestry applications. The ALS data have a point density of 0.15–2 m², a pulse footprint diameter of ~0.5 m at the canopy level, and a scanning angle of up to 30° from the nadir using manned airplanes [28]. In this study, a total of eleven (11) height percentiles (Table 2) were used and these correspond to data collected during the summers of 2018, 2019, 2020, and 2021 using the Reigl VQ-1560i airborne laser scanning system. The system was configured to operate at altitudes of 1200 m (mapping of densely populated areas), 2000 m (mapping with imaging priority), and 2600 m during the spring and 3100 m in the summertime for forestry mapping.

2.7. Independent Variable Screening

A total of thirty-six (36) variables were pooled from the Sentinel -2 and LiDAR data (Table 2). In order to avoid multicollinearity among independent variables, the Boruta feature selection method and variance inflation factors (VIFs) were employed. First, we used the Boruta feature selection to rank all the independent variables according to their importance as a function of their Z-score value; secondly, in order to enrich the selection processes, VIFs were estimated among the remaining predictor variables and variables with VIFs lower than five (5) were considered an indicator of multicollinearity and filtered appropriately. We explored four different combinations of variables: (i) VIs combined with band reflectance values and LiDAR data. (ii) VIs combined with band reflectance values. (iii) VIs combined with LiDAR data. (iv) Band reflectance values combined with LiDAR data. For each year, a total of eleven (15) spectra vegetation indices (VIs) (Table 2) were estimated using R studio. The height percentile and band reflectance are as described under the sections (LiDAR) and (Sentinel-2), respectively.

2.8. Regression Analysis

Three (3) non-parametric models were constructed and compared using R statistical software version 4.3.3. These models encompass random forest (RF), support vector regression (SVR), and XGBoost.

2.8.1. Random Forest Regression

The random forest (RF) is a CART decision model [56] and one of the most commonly used non-parametric methods to establish the relationship between dependent and independent variables. It is a machine learning process where regression is based on a decision tree characterized by a selection of a subset from training data by a random sampling method with replacement. This bootstrapping process extracts a set of n data points repeat-

edly and randomly from the sample data as a training dataset (approximately two-thirds), and the remaining one-third which are left out-of-bag (OOB) are used to estimate the error and to assign a level of importance to the dependent variables. Handling redundancy in predictors when the number of predictors is very large, which tends to aggravate into multicollinearity, is properly minimized and overfitting is prevented [57]. For this study, the RF was constructed using the R package *RandomForest* [58]. It was initially calibrated with the default value of *ntrree* and subsequently varied with a step of 500. The number of feature variables at each split (*mtry*) was determined using the *tuneRF* function. This function facilitates the tuning of the hyperparameter *mtry* by searching for the optimal value that minimizes the out-of-bag (OOB) error estimate.

2.8.2. Support Vector Regression

Support vector regression (SVR) is a supervised machine learning algorithm that handles the prediction of continuous numeric values. It is a derivative of support vector machine (SVM), which was developed by Vapnik et al. [59], and it uses a kernel function to linearize a non-linear regression by mapping the input data into a high-dimensional function space [60]. This method has been proven to be one of the popular algorithms used for regression analysis to estimate forest-related variables using remotely sensed data [15,35,44]. It uses a set of functions to estimate the regression model and engage Structural Risk Minimization (SRM) to minimize model error and overfitting problems [35,61]. In this study, the radial basis function kernel (RBF) was used to train the model as it has a few tuning parameters and was implemented in R using the “e1071” R package.

2.8.3. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is one of the most effective advanced machine learning approaches whose iteratives are based on the gradient boosting framework. XGBoost was proposed by Chen et al. [62] to support both classification and regression models using *reg:logistics* and *reg:linear* as loss functions for its classification and regression algorithms, respectively. It applies the second-order Taylor expansion to the objective function and uses the combination of first- and second-order derivatives to approximate the objective function more accurately to arrive at a more precise definition that facilitates optimization [63]. It introduces a regularization term into the objective function to prevent overfitting and enhance the creation of simpler models. As an ensemble learning method, XGBoost combines multiple base learners to obtain a single prediction that is more robust towards achieving a better predictive accuracy.

2.9. Model Evaluation

The total sample plots in each quadrant (Table 1) were randomly portioned into 70% and 30% subsets; the first was used as training dataset and the latter was used as the testing dataset for the built model. In order to avoid the risk associated with overfitting and to obtain a more reliable estimate of machine learning-based models, a k-fold cross-validation technique is always a vital step. This method partitions the original sample into k equal numbers of subset and iteratively chooses different partitions for the training and validation datasets. Each partition is used as the testing set only once while the remaining k-1 subsets are used as the training sample in a leave-one-out cross-validation method (LOOCV). A parallel processing technique was employed to speed up the computational speed of the LOOCV using the *parallel* package in R. Performance metrics such as Root Mean Square Error (RMSE) and coefficient of determination (R^2) were calculated based on the predictions made on the testing dataset using 10-fold cross-validation.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\tilde{y}_i - y_i)^2}{n}} \quad (3)$$

$$RMSE\% = \frac{RMSE}{\bar{y}} \times 100 \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where n is the number of observations, y_i is the actual GSV of the i th plot, \tilde{y}_i is the predicted GSV of the i th plot, and \bar{y} is the mean of the actual GSV.

3. Results

A total of 14,880 sample plots were selected from the NFI data following the removal of outliers through the PauTa criteria. The share of each quadrant is represented in Table 1, where the largest share (3852) is on the NE zone followed by the NW with 3712 sample plots, while the SE and SW have shares of 3665 and 3651, respectively. Four (4) variable combinations as discussed in Section 2 were used in training the three predictive models, i.e., RF, SVR, and XGBoost. A maximum of a 10-fold cross-validation was considered sufficient enough to optimize the model according to the minimum RMSE and a considerable coefficient of determination. The sample plots in each share were used to generate a corresponding model and each model's performance metrics are explained as follows.

3.1. Performance Assessment Based on Random Forest Predictive Model

The performance metrics for the RF-based model for each combination of variables at the quadrant level are illustrated in Table 3. In the NW zone, RMSE values ranged from 125.39 m³/plot to 141.84 m³/plot for the CO3 and CO1 categories, respectively, with CO4 and CO3 showing relatively close performance with R² values of 0.64 (CO4) and 0.63 (CO3). The CO4 seems to have a relatively higher RMSE (126.74 m³/plot) and a higher RMSE% (13.09) as compared to the CO3 (Table 3); with this, preference was given to the combination with the lowest RMSE and RMSE%. The highest R² (0.78) also corresponds to a larger RMSE (141.84), as with the CO1 model, this necessitates the choice of a CO3 variable combination for the best performance at this zone. In the SW zone, RMSE varied from 126.85 m³/plot to 139.21 m³/plot for CO3 and CO1, respectively. In this zone, the CO4 variable combination exhibited a good fit with an RMSE of 129.22 m³/plot, R² of 0.74, and RMSE% of 14.04. Although the CO1 and CO2 models show higher R² values of 0.79 and 0.77, their associated RMSEs are larger as compared to the CO4 combination (Table 3).

Table 3. Support vector regression performance metrics.

Model	No. Variables		R ²	RMSE	RMSE%
CO1	17	NW	0.78	141.84	14.65
CO2	14		0.69	133.1	13.74
CO3	15		0.63	125.39	12.95
CO4	22		0.64	126.74	13.09
CO1	14	SW	0.79	139.21	15.12
CO2	13		0.77	138.57	15.05
CO3	20		0.7	126.85	13.78
CO4	30		0.74	129.22	14.04
CO1	18	NE	0.73	141.51	13.23
CO2	17		0.74	147.2	13.77
CO3	18		0.61	132.17	12.36
CO4	25		0.64	133.77	12.51
CO1	18	SE	0.81	135.49	13.84
CO2	23		0.74	128.41	13.12
CO3	17		0.68	119.35	12.19
CO4	29		0.7	120.56	12.32

For the NE zone, CO4 yielded the best results with an R^2 of 0.64 and an RMSE of $133.77 \text{ m}^3/\text{plot}$, followed by CO3 model with an R^2 of 0.61 and an RMSE of $132.17 \text{ m}^3/\text{plot}$. The highest R^2 and RMSE under this zone also correspond to CO2 and CO1, which is accompanied by a higher RMSE. Similarly, in the SE zone, the CO4 variables performed the best with an R^2 of 0.70 and an RMSE of $120.56 \text{ m}^3/\text{plot}$, followed by the CO3 variables with an R^2 0.68 and an RMSE of $119.35 \text{ m}^3/\text{plot}$. Although CO1 had the highest R^2 (0.81) and RMSE (135.49), the emphasis on minimizing the RMSE favored the CO4 variables as the variability explained under this combination is good on average.

3.2. Performance Assessment Based on Support Vector Regression

The results of the support vector regression model carried out at the quadrant level using each variable combination are presented in Table 4. In the NW zone, the best performance in terms of the lowest RMSE and RMSE% is the CO3-generated model having an R^2 of 0.52, an RMSE of $123.41 \text{ m}^3/\text{plot}$, and an RMSE% of 12.74%. However, this model captured less variability compared to its RF counterpart (Table 3). A notable trend of a consistent highest R^2 value across each quadrant is exhibited by the CO1 model, and also it is characterized by the highest RMSE. In the NW zone, for instance, the CO1 variable achieves an R^2 of 0.81 and an RMSE of $150.01 \text{ m}^3/\text{plot}$, indicating significant error when applied to the 3712 plots within the zone.

Table 4. Support vector regression performance metrics.

Model	No. Variables		R^2	RMSE	RMSE%
CO1	17	NW	0.81	150.01	15.49
CO2	14		0.53	130.53	13.48
CO3	15		0.52	123.41	12.74
CO4	22		0.51	125.05	12.91
CO1	14	SW	0.79	141.35	15.36
CO2	13		0.74	136.17	14.79
CO3	20		0.64	124.53	13.53
CO4	30		0.64	129.57	14.08
CO1	18	NE	0.65	153.53	14.06
CO2	17		0.65	149.16	13.95
CO3	18		0.43	129.83	12.14
CO4	25		0.6	138.98	13
CO1	18	SE	0.86	144.46	14.76
CO2	23		0.79	132.45	13.52
CO3	17		0.63	116.43	11.89
CO4	29		0.68	122.34	12.5

In the SW zone, the CO1 variable produced an R^2 value of 0.79 and an RMSE of $141.35 \text{ m}^3/\text{plot}$, while CO4 and CO3 both achieved an R^2 value of 0.64. However, the CO3 model has the lowest RMSE ($124.53 \text{ m}^3/\text{plot}$) compared to CO4 ($129.57 \text{ m}^3/\text{plot}$). Despite CO2's relatively average performance with an R^2 of 0.74 and an RMSE of $136.17 \text{ m}^3/\text{plot}$, the CO3 model is deemed more suitable with an RMSE of $124.53 \text{ m}^3/\text{plot}$. The range of the coefficient of determination in the NE zone is between 0.43 and 0.65, where CO1 and CO2 explain the same variability with an R^2 of 0.65. However, the RMSE in CO1 is $4.37 \text{ m}^3/\text{plot}$ higher than CO2. The CO3 model exhibits the lowest RMSE ($129.83 \text{ m}^3/\text{plot}$) and RMSE% (12.14), although its R^2 (0.43) is the lowest among all variable combinations. On average, CO4 performs notably better with an R^2 of 0.60 and an RMSE of $138.98 \text{ m}^3/\text{plot}$. In the SE zone, CO1 demonstrates the highest R^2 (0.86) and RMSE ($144.46 \text{ m}^3/\text{plot}$), while CO3 exhibits the lowest R^2 (0.63) and the lowest corresponding RMSE ($116.43 \text{ m}^3/\text{plot}$). Although the CO3 and CO4 models shows a relatively good fit, selecting CO4 ($R^2 = 0.68$, RMSE = $122.34 \text{ m}^3/\text{plot}$) is justifiable. In comparison with the RF model, a better precision is obtained with the RF using the CO4 model.

3.3. Performance Assessment Based on Extreme Gradient Boosting

XGBoost, Table 5, outperforms the SVR in the NW zone; meanwhile, a similar performance was achieved within the variable combination of the RF and XGBoost models. Notably, for the CO4 variable, both models (XGBoost and RF) had the same R^2 value of 0.64 and closely correlated RMSE values, differing by only $0.59\text{ m}^3/\text{plot}$. For the CO1 model, both XGBoost and RF achieved an R^2 of 0.78, though XGBoost had a higher RMSE of $1.8\text{ m}^3/\text{plot}$ compared to the RF model. An R^2 of 0.61 for XGBoost was found in the CO3 model, with an RMSE of $124.63\text{ m}^3/\text{plot}$, highlighting the robustness of CO3 variables for growing stock volume estimation.

Table 5. XGBoost performance metrics.

Model	No. Variables		R^2	RMSE	RMSE%
CO1	17	NW	0.78	143.67	14.84
CO2	14		0.68	133.77	13.81
CO3	15		0.61	124.46	12.85
CO4	22		0.64	126.15	13.03
CO1	14	SW	0.77	139.37	15.19
CO2	13		0.74	136.17	14.79
CO3	20		0.68	129.32	14.05
CO4	30		0.69	128.29	13.94
CO1	18	NE	0.69	139.63	13.09
CO2	17		0.72	148.16	13.86
CO3	18		0.6	132.63	12.4
CO4	25		0.59	134.25	12.56
CO1	18	SE	0.8	138.01	14.1
CO2	23		0.75	131.4	13.4
CO3	17		0.65	119.48	12.17
CO4	29		0.65	123.36	12.6

In the SW zone, we compared the unit differences in performance metrics and found that the CO4 model-based algorithm met our goals the best, with an R^2 of 0.69 and the lowest RMSE of $128.29\text{ m}^3/\text{plot}$. This was followed by the CO3 model with an R^2 of 0.68 and an RMSE of $129.32\text{ m}^3/\text{plot}$. As with the other models, the CO1 variable yielded the highest R^2 of 0.77, with an RMSE of $139.39\text{ m}^3/\text{plot}$. At the NE zone, the RMSEs are in the range of $132.58\text{ m}^3/\text{plot}$ to $148.16\text{ m}^3/\text{plot}$, which is similar to the range obtained with RF. The CO3 produced a more accurate estimate considering its R^2 and RMSE of 0.60 and $132.58\text{ m}^3/\text{plot}$, respectively. In the NE zone, the RMSE values range from $132.63\text{ m}^3/\text{plot}$ to $148.16\text{ m}^3/\text{plot}$, similar to the range obtained with RF. The CO3 model provided a more accurate estimate, with an R^2 of 0.60 and an RMSE of $132.63\text{ m}^3/\text{plot}$.

Comparing the results obtained in the SE zone based on the four variable combinations, we found the CO3 model to be the most precise, with an R^2 of 0.65 and an RMSE of $119.48\text{ m}^3/\text{plot}$. The CO4 model also achieved an R^2 of 0.65, but its associated RMSE is $123.36\text{ m}^3/\text{plot}$, which is $4.18\text{ m}^3/\text{plot}$ higher than CO3, making CO3 the preferred model. The CO2 model shows an increased R^2 of 0.75 but a decreased accuracy, with an RMSE of $131.40\text{ m}^3/\text{plot}$. Although the CO1 model has the highest R^2 of 0.80, its associated RMSE of $138.01\text{ m}^3/\text{plot}$ indicates a notable decrease in predictive power.

3.4. Model Predictive Power at Each Quadrant/Comparative Analysis of Predictive Model

In all quadrants, the RF algorithm proves to be the most effective. For instance, the RF algorithm with the CO3 variable combination proves to be the best predictive model in the NW zone, with the SVR-based model achieving the lowest RMSE of $123.41\text{ m}^3/\text{plot}$, followed by the CO3-XGBoost model with an RMSE of $124.63\text{ m}^3/\text{plot}$; however, considering the differences in R^2 values and associated RMSEs, the RF-based model with an R^2 value of 0.63 and an RMSE value of $125.39\text{ m}^3/\text{plot}$ is more preferable (Figure 3a), and the CO3 variable combination demonstrates robust performance, consistently yielding the

lowest RMSE values across all three predictive models. In the SW zone, the RF algorithm again outperforms other models in estimating GSV based on the CO4 variable combination (Figure 3b); although the SVR model achieved the lowest RMSE, it also showed more scattering with an R^2 of 0.6. Shifting focus to the NE zone, the RF algorithms using the CO4 variable combination yielded more preferable results (Figure 3c, followed by the XGBoost model using the CO3 variable combination, with the least effective results found for the SVR model (Table 6); turning to the SE zone, the use of RF-based algorithms with the CO4 variable combination also shows appreciable performance (Figure 3d), outperforming the SVR and XGBoost models (Table 6). As observed in other zones, the lowest RMSE does not necessarily correspond to the RF-based model; nonetheless, the minimal differences in performance metrics among these predictive models favor the RF model, and generally, the SVR yields the lowest RMSE across all quadrants and predictive models, but its CO1- and CO2-based models are characterized by relatively high RMSE values.

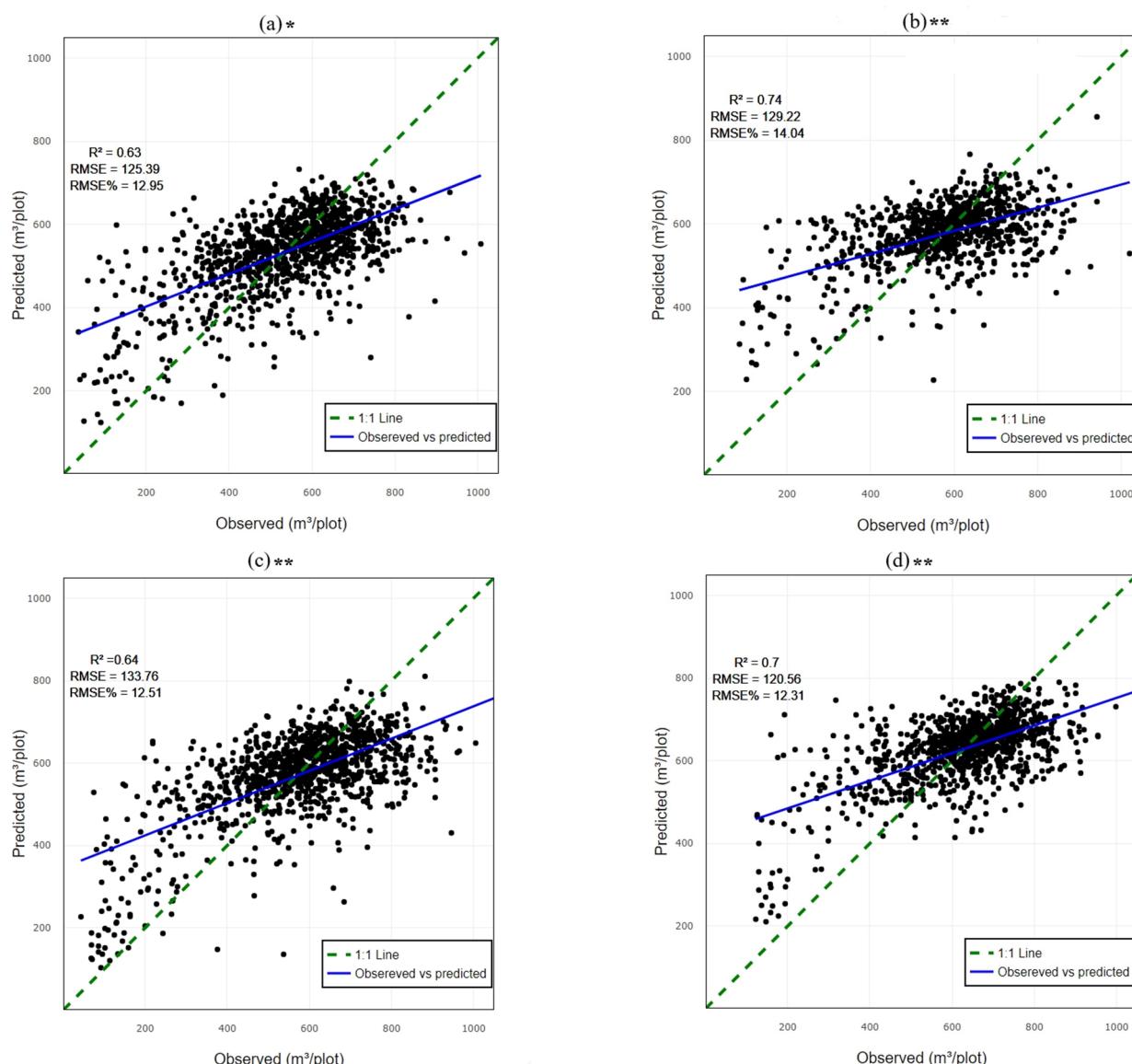


Figure 3. Scatter plot of observed vs. predicted GSV values for the validation plots using the best predictive model. The symbols * and ** represent the CO3 and CO4 combinations, respectively. (a), (b), (c), and (d) denote the random forest-based models for the northwest, southwest, northeast, and southeast regions, respectively.

Table 6. Comparative metrics.

Quadrant	RF R^2	RMSE	SVR R^2	RMSE	XGBoost R^2	RMSE
NW	0.64 *	125.39	0.52 *	123.41	0.64 **	126.15
SW	0.74 **	129.22	0.64 *	124.53	0.69 **	128.29
NE	0.64 **	133.77	0.60 **	138.98	0.60 *	132.63
SE	0.70 **	120.56	0.68 **	122.34	0.65 *	123.66

Here, * and ** are the CO3 and CO4 combinations, respectively.

4. Discussion

In this study, we demonstrated the prediction of forest tree growing stock volume using three predictive algorithms: RF, SVM, and XGBoost. In order to enrich our prediction, we made use of the Estonian National Forest Inventory data, Sentinel-2, and LiDAR height percentile data as the use of multi-source data has the potential to improve the prediction of growing stock volume [42,45,64]. Our study presents a pioneering model estimation of plot GSV using a complete-cycle NFI over Estonia forest, and as such provides an architectural framework for comparative study in the country. We used four variable combinations derived from the remotely sensed data as explained in Section 2 of this article. These include LiDAR and vegetation indices (CO1); LiDAR and band reflectance (CO2); vegetation indices and band reflectance value (CO3); and band reflectance value, vegetation indices, and LiDAR height percentile (CO4). In our characterization, we examined which of the machine learning algorithms performs the best and also delved into the variables' importance and their associated sources.

Notably, in all the quadrants, the RF-based algorithm outperformed the other predictive algorithms (Table 6). Although the lowest RMSE does not necessarily correspond to the RF model, the associated unit differences between the performance metric favor the RF model. For example, in the NW, the RF achieved an R^2 of 0.63 and an RMSE of $125.39 \text{ m}^3/\text{plot}$, followed by the XGBoost with an R^2 of 0.64 and an RMSE of $126.15 \text{ m}^3/\text{plot}$; the lowest RMSE corresponds to the SVR algorithm with an R^2 value of 0.52 and an RMSE value of $123.41 \text{ m}^3/\text{plot}$. The lowest RMSE of $124.53 \text{ m}^3/\text{plot}$ in the SW also corresponds to the SVR, whereas the best predictive model was found in the RF models with an R^2 of 0.74 and an RMSE of $129.22 \text{ m}^3/\text{plot}$. This further establishes the strength of RF algorithms in estimating forest variables.

The use of RF to estimate GSV has been demonstrated over time; in some cases, it was used exclusively [14,42,65], and in some, it was used comparatively with other ML algorithms [41,42,45]. For example, Chirici et al. [45] used two non-parametric and two parametric prediction methods (RF, K-NN, MLR, and GWR) to produce a wall–wall spatial prediction of growing stock volume using an Italian NFI plot and remotely sensed data and concluded that the RF algorithm produced the best accuracy. However, the variable used in their prediction was not the same as ours. In our study, we do not consider the environmental factors, nor distinguish the forest type and species at each quadrant level. This, of course, we believe would have helped to further homogenize our findings. For example, Estonia is characterized by both coniferous and deciduous tree species. Lang et al. [28], in their research to facilitate the construction of maps for forest height, standing wood volume, and tree species composition in Estonia, suggested that for a more reliable prediction, empirical data can be divided into subsets that account for variability before fitting the model.

The use of optical remote sensing data for the prediction of forest variables is usually accompanied by issues relating to saturation and this is believed to have an effect on the accuracy of the predicted variables. Areas within the forest with dense canopy closure are highly susceptible to saturation problems as the optical sensor only captures horizontal information about the forest. One of the ways to possibly reduce this impact is the inclusion of vegetation indices derived from red edge bands [66,67], and another is through the inclusion of satellite data that capture the vertical information about the forest structure.

Hence, we incorporated the height percentile data from the LiDAR metrics and vegetation indices derived from red edge bands (Table 2). As seen from our findings, the combination that incorporated vegetation indices, band reflectance values, and height percentile data consistently yielded the best results in each quadrant except in the NW quadrant, where the combination of band reflectance number and height percentile is considered to be more accurate (Table 3). However, the unit difference between the performance metrics of the CO3 and CO4 variable combinations at this quadrant is not well pronounced, having an RMSE of $125.39\text{ m}^3/\text{plot}$ and an R^2 value of 0.63 in the CO3 and an RMSE $126.7\text{ m}^3/\text{plot}$ and an R^2 value of 0.64 in the CO3 (Table 3).

When band reflectance and height percentiles are used, this approach yields the lowest RMSE across all quadrants and the three predictive models (Tables 3–5). Additionally, incorporating vegetation indices further improves the coefficient of determination (R^2). For example, in the SW zone, adding vegetation indices improves the R^2 by 4.11% and decreases the RMSE by 1.56%. Similarly, in the NE zone, there is a 4.7% improvement in R^2 and a 1.19% decrease in RMSE. In the SE zone, the R^2 increases by 2.86% and the RMSE decreases by 1.13%. The CO1 and CO2 combination consistently yields significantly higher R^2 values and erroneous RMSEs (Tables 3–5).

Lahssini et al. [64] were of the opinion that the availability of multimodal remote sensing data alone does not necessarily guarantee better performance; instead, how these data are integrated is crucial. Appendix A shows how the various variables were integrated and ranked in order of their importance. Based on the prediction in the NW quadrant, the LiDAR height percentile and band reflectance number show a good fit in GSV prediction with height percentiles at p95, p90, and p50 being the most important height percentile variables; this had been established by other studies [68–70]. Red edge bands 7 and 5 are the most important band reflectance numbers; this corresponds to other research where the red edge bands had been established to outperform other bands in forest variable prediction [8,71,72].

In the SW, NE, and SE quadrats where the combination of height percentile, individual band reflectance value, and vegetation indices proved to be the most efficient combination in GSV prediction, the height percentiles at p95, p90, p80, and p50 are crucial among the height percentile variables, followed by individual band reflectance values. This corresponds to [68,69,73], where these height percentiles were also reported to contribute effectively to forest variable prediction. Among the individual band reflectance numbers, the red edge and SWIR are the most commonly reported bands that correlate with GSV. Our findings further establish this, where the SWIR1, SWIR2, and red edge bands were part of the most important variables that contribute significantly to the GSV prediction, except in the NE quadrant where red edge band 6 was removed.

In our research, we only examined the effect of the inclusion of various band reflective values, vegetation indices, and height percentile data on the overall performance and the reliability of GSV prediction; meanwhile, the initial performance of the ML algorithm can be improved with a hybrid model [42,74]. For instance, in Hu et al. [42], the accuracies of three machine learning models, RF, SVR, and an artificial neural network (ANN), were improved by building hybrid models as an extension of these ML models. These later yielded an improvement by reducing the initial RMSE of RF by 11.45%, and there was an increase in the associated R^2 by 2.17% for the SVR, the RMSE was reduced by 20.37%, and the R^2 increased by 7%; these are called hybrid ordinary kriging models.

5. Conclusions

Estimating and quantifying forest-related variables such as aboveground biomass and growing stock volume are becoming increasingly important and there already exist considerable research findings on this; nonetheless, there exists a dearth of comprehensive research effort in Estonia. The primary aim of this study was to enhance the estimation of growing stock volume by integrating data from the National Forest Inventory, Sentinel-2, and LiDAR, employing three predictive algorithms: RF, SVR, and XGBoost. Our findings

reveal that the combined use of height percentile, band reflectance value, and vegetation indices significantly improves the accuracy of GSV predictions. The combination of band reflectance value and height percentile also exhibited a promising trend. Among the predictive algorithms tested, RF demonstrated the best performance, yielding the most considerable accuracy at each quadrant. The most commonly occurring height percentile variables are P95, p90, and p50, while the occurrences of the red edge and SWIR bands are significantly notable. Despite the promising outcomes of our study, certain research gaps exist that need addressing. For example, the variability in forest types and conditions across different geographical regions suggests that further testing and refinement of these methods are necessary. Efforts should attempt to homogenize the tree species in each quadrant into individual strata before modeling, and, additionally, integrating other data sources, such as ecological and climatological, and experimenting with emerging machine learning techniques could further enhance prediction accuracy. In conclusion, the enriched estimation of growing stock volume using NFI, Sentinel-2, and LiDAR data, combined with advanced predictive algorithms, offers a robust tool for forest management. This approach promises significant benefits for sustainable forestry practices and ecosystem monitoring, paving the way for more informed and effective climate smart decision-making in forest conservation and management.

Author Contributions: Conceptualization, T.O.O. and A.S.; methodology, T.O.O. and A.S.; formal analysis, T.O.O. and A.S.; writing—original draft preparation, T.O.O.; writing—review and editing, T.O.O. and A.S.; 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 the Estonian Land Board for supplying the LiDAR data. 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

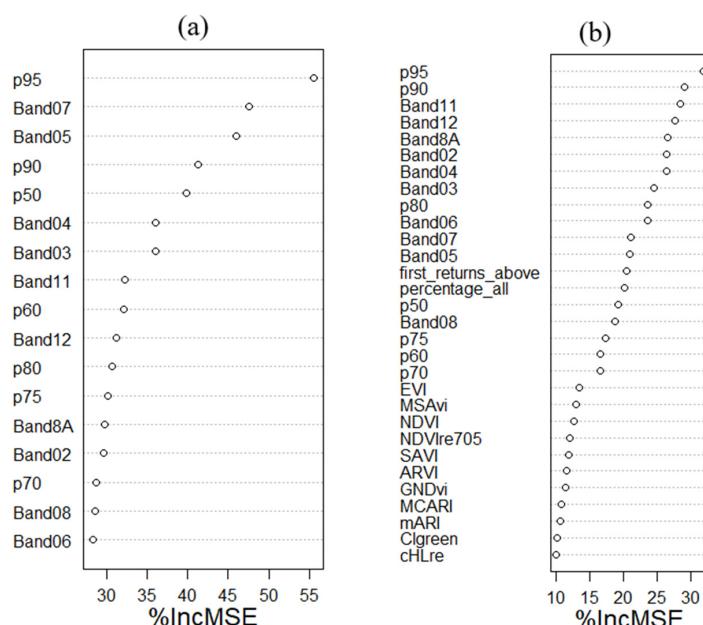


Figure A1. *Cont.*

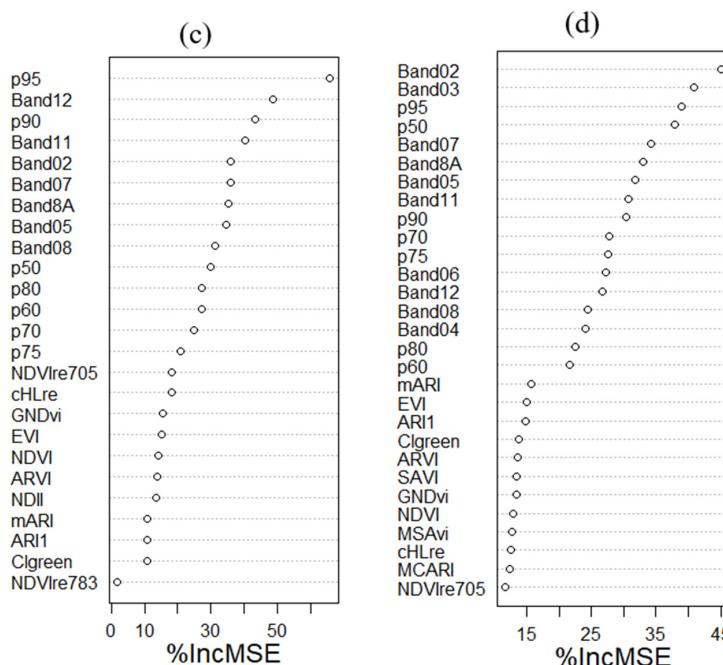


Figure A1. Variable important plot using the best predictive model. Where (a), (b), (c), and (d) denote the random forest-based models for the northwest, southwest, northeast, and southeast regions, respectively.

References

1. Bonan, G.B. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **2008**, *320*, 1444–1449. [[CrossRef](#)] [[PubMed](#)]
2. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A Large and Persistent Carbon Sink in the World’s Forests. *Science* **2011**, *333*, 988–993. [[CrossRef](#)] [[PubMed](#)]
3. Mori, A.S.; Furukawa, T.; Sasaki, T. Response Diversity Determines the Resilience of Ecosystems to Environmental Change. *Biol. Rev.* **2013**, *88*, 349–364. [[CrossRef](#)] [[PubMed](#)]
4. Sing, L.; Metzger, M.J.; Paterson, J.S.; Ray, D. A Review of the Effects of Forest Management Intensity on Ecosystem Services for Northern European Temperate Forests with a Focus on the UK. *For. Int. J. For. Res.* **2018**, *91*, 151–164. [[CrossRef](#)]
5. Belcher, B.; Ruíz-Pérez, M.; Achdiawan, R. Global Patterns and Trends in the Use and Management of Commercial NTFPs: Implications for Livelihoods and Conservation. *World Dev.* **2005**, *33*, 1435–1452. [[CrossRef](#)]
6. Brown, S. Measuring Carbon in Forests: Current Status and Future Challenges. *Environ. Pollut.* **2002**, *116*, 363–372. [[CrossRef](#)]
7. Chirici, G.; McRoberts, R.E.; Winter, S.; Bertini, R.; Brändli, U.-B.; Asensio, I.A.; Bastrup-Birk, A.; Rondeux, J.; Barsoum, N.; Marchetti, M. National Forest Inventory Contributions to Forest Biodiversity Monitoring. *For. Sci.* **2012**, *58*, 257–268. [[CrossRef](#)]
8. 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](#)]
9. FAO. Global Forest Resources Assessment Report, Estonia Global Forest Resources Assessment. Available online: <https://openknowledge.fao.org/server/api/core/bitstreams/83fe0535-10eb-4a48-b67d-6db7b6977e89/content> (accessed on 15 November 2022).
10. Næsset, E. Practical Large-Scale Forest Stand Inventory Using a Small-Footprint Airborne Scanning Laser. *Scand. J. For. Res.* **2004**, *19*, 164–179. [[CrossRef](#)]
11. Breidenbach, J.; Anton-Fernandez, C.; Petersson, H.; Mcroberts, R.E.; Astrup, R. Quantifying the Model-Related Variability of Biomass Stock and Change Estimates in the Norwegian National Forest Inventory. *For. Sci.* **2014**, *60*, 25–33. [[CrossRef](#)]
12. Wulder, M.A.; White, J.C.; Fournier, R.A.; Luther, J.E.; Magnussen, S. Spatially Explicit Large Area Biomass Estimation: Three Approaches Using Forest Inventory and Remotely Sensed Imagery in a GIS. *Sensors* **2008**, *8*, 529–560. [[CrossRef](#)] [[PubMed](#)]
13. 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](#)]
14. 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](#)]
15. 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](#)]

16. 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](#)]
17. Loveland, T.R.; Irons, J.R. Landsat 8: The Plans, the Reality, and the Legacy. *Remote Sens. Environ.* **2016**, *185*, 1–6. [[CrossRef](#)]
18. Gobakken, T.; Næsset, E. Assessing Effects of Laser Point Density, Ground Sampling Intensity, and Field Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data. *Can. J. For. Res.* **2008**, *38*, 1095–1109. [[CrossRef](#)]
19. Rosema, A.; Verhoef, W.; Noorbergen, H.; Borgesius, J.J. A New Forest Light Interaction Model in Support of Forest Monitoring. *Remote Sens. Environ.* **1992**, *42*, 23–41. [[CrossRef](#)]
20. Mulatu, K.A.; Decuyper, M.; Brede, B.; Kooistra, L.; Reiche, J.; Mora, B.; Herold, M. Linking Terrestrial LiDAR Scanner and Conventional Forest Structure Measurements with Multi-Modal Satellite Data. *Forests* **2019**, *10*, 291. [[CrossRef](#)]
21. Temitope Olaoluwa, O.; Akinyede Joseph, O.; Oyinloye, R.O. *Spatio-Temporal Analysis of Biomass Accumulation in Tectona Grandis and Gmelina Arboreal Plantation in Oluwa Afforestation*; Federal University of Technology: Akure, Nigeria, 2016.
22. Panagiotidis, D.; Abdollahnejad, A.; Slavík, M. 3D Point Cloud Fusion from UAV and TLS to Assess Temperate Managed Forest Structures. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102917. [[CrossRef](#)]
23. Panagiotidis, D.; Surový, P.; Kuželka, K. Accuracy of Structure from Motion Models in Comparison with Terrestrial Laser Scanner for the Analysis of DBH and Height Influence on Error Behaviour. *J. For. Sci.* **2016**, *62*, 357–365. [[CrossRef](#)]
24. Juujarvi, J.; Heikkonen, J.; Brandt, S.S.; Lampinen, J. *Digital-Image-Based Tree Measurement for Forest Inventory*; Casasent, D.P., Ed.; International Society for Optics and Photonics: Boston, MA, USA, 1998; pp. 114–123.
25. Lim, K.; Treitz, P.; Wulder, M.; St-Onge, B.; Flood, M. LiDAR Remote Sensing of Forest Structure. *Prog. Phys. Geogr. Earth Environ.* **2003**, *27*, 88–106. [[CrossRef](#)]
26. 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](#)]
27. 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](#)]
28. Lang, M.; Sims, A.; Pärna, K.; Kangro, R.; Möls, M.; Möistus, M.; Kivisté, 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 Pöhinev Lahendus Eesti Statistikilise Metsainventuuri Jaoks Puistute Kõrguse, Tüvemahu ja Liigilise Koosseisu Kaartide Koostamiseks. *For. Stud.* **2020**, *73*, 77–97. [[CrossRef](#)]
29. McRoberts, R.E.; Tomppo, E.O.; Næsset, E. Advances and Emerging Issues in National Forest Inventories. *Scand. J. For. Res.* **2010**, *25*, 368–381. [[CrossRef](#)]
30. Næsset, E. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [[CrossRef](#)]
31. Nilsson, M.; Nordkvist, K.; Jonzén, J.; Lindgren, N.; Axensten, P.; Wallerman, J.; Egberth, M.; Larsson, S.; Nilsson, L.; Eriksson, J.; et al. A Nationwide Forest Attribute Map of Sweden Predicted Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory. *Remote Sens. Environ.* **2017**, *194*, 447–454. [[CrossRef](#)]
32. Lindgren, N.; Olsson, H.; Nyström, K.; Nyström, M.; Ståhl, G. Data Assimilation of Growing Stock Volume Using a Sequence of Remote Sensing Data from Different Sensors. *Can. J. Remote Sens.* **2022**, *48*, 127–143. [[CrossRef](#)]
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. Chowdhury, T.A.; Thiel, C.; Schmullius, C. Growing Stock Volume Estimation from L-Band ALOS PALSAR Polarimetric Coherence in Siberian Forest. *Remote Sens. Environ.* **2014**, *155*, 129–144. [[CrossRef](#)]
35. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [[CrossRef](#)]
36. Zharko, V.O.; Bartalev, S.A.; Sidorenkov, V.M. Forest Growing Stock Volume Estimation Using Optical Remote Sensing over Snow-Covered Ground: A Case Study for Sentinel-2 Data and the Russian Southern Taiga Region. *Remote Sens. Lett.* **2020**, *11*, 677–686. [[CrossRef](#)]
37. Hudak, A.T.; Lefsky, M.A.; Cohen, W.B.; Berterretche, M. Integration of Lidar and Landsat ETM+ Data for Estimating and Mapping Forest Canopy Height. *Remote Sens. Environ.* **2002**, *82*, 397–416. [[CrossRef](#)]
38. Asner, G.P.; Mascaro, J.; Muller-Landau, H.C.; Vieilledent, G.; Vaudry, R.; Rasamoelina, M.; Hall, J.S.; van Breugel, M. A Universal Airborne LiDAR Approach for Tropical Forest Carbon Mapping. *Oecologia* **2012**, *168*, 1147–1160. [[CrossRef](#)] [[PubMed](#)]
39. Fernández-Landa, A.; Fernández-Moya, J.; Tomé, J.L.; Algeet-Abarquero, N.; Guillén-Climent, M.L.; Vallejo, R.; Sandoval, V.; Marchamalo, M. High Resolution Forest Inventory of Pure and Mixed Stands at Regional Level Combining National Forest Inventory Field Plots, Landsat, and Low Density Lidar. *Int. J. Remote Sens.* **2018**, *39*, 4830–4844. [[CrossRef](#)]
40. Chirici, G.; Mura, M.; McInerney, D.; Py, N.; Tomppo, E.O.; Waser, L.T.; Travaglini, D.; McRoberts, R.E. A Meta-Analysis and Review of the Literature on the k-Nearest Neighbors Technique for Forestry Applications That Use Remotely Sensed Data. *Remote Sens. Environ.* **2016**, *176*, 282–294. [[CrossRef](#)]

41. Hawryło, P.; Francini, S.; Chirici, G.; Giannetti, F.; Parkitna, K.; Krok, G.; Mitelsztedt, K.; Lisańczuk, M.; Stereńczak, K.; Ciesielski, M.; et al. The Use of Remotely Sensed Data and Polish NFI Plots for Prediction of Growing Stock Volume Using Different Predictive Methods. *Remote Sens.* **2020**, *12*, 3331. [[CrossRef](#)]
42. Hu, T.; Sun, Y.; Jia, W.; Li, D.; Zou, M.; Zhang, M. Study on the Estimation of Forest Volume Based on Multi-Source Data. *Sensors* **2021**, *21*, 7796. [[CrossRef](#)]
43. Myroniuk, V.; Bell, D.M.; Gregory, M.J.; Vasylyshyn, R.; Bilous, A. Uncovering Forest Dynamics Using Historical Forest Inventory Data and Landsat Time Series. *For. Ecol. Manag.* **2022**, *513*, 120184. [[CrossRef](#)]
44. Singh, C.; Karan, S.K.; Sardar, P.; Samadder, S.R. Remote Sensing-Based Biomass Estimation of Dry Deciduous Tropical Forest Using Machine Learning and Ensemble Analysis. *J. Environ. Manag.* **2022**, *308*, 114639. [[CrossRef](#)]
45. Chirici, G.; Giannetti, F.; McRoberts, R.E.; Travaglini, D.; Pecchi, M.; Maselli, F.; Chiesi, M.; Corona, P. 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.* **2020**, *84*, 101959. [[CrossRef](#)]
46. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Hall, D.E.; Falkowski, M.J. Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245. [[CrossRef](#)]
47. Forest. *Yearbook Forest 2018. (Aastaraamat Mets 2018)*; Raudsaar, M., Siimon, K.L., Valgepea, M., Eds.; Keskkonnaagentuur: Tallinn, Estonia, 2020.
48. Raukas, A. Briefly about Estonia. *Dyn. Environnementales* **2018**, *42*, 284–291. [[CrossRef](#)]
49. Riigikantselei. RT Forest Act. *Riigiteataja I* **2006**, *30*, 232.
50. Raudsaar, M.; Merenäkk, M.; Valgepea, M. *Yearbook Forest 2013*; Estonian Environment Agency: Tartu, Estonian, 2014.
51. Adermann, V. Eesti Metsad. Estonian Forests 2010. 2012. Available online: <http://www.digar.ee/id/nlib-digar:258209> (accessed on 4 February 2024).
52. Reintam, L. Soil Objectives. Estonian Environment: Past, Present and Future. In *Ministry of the Environment of Estonia, Environment Information Centre*; Raukas, A., Ed.; Ministry of the Envieonment of Estonia: Tallinn, Estonia, 1996; pp. 106–108.
53. Kohava, P. *Forests in Estonia 1999 (Eesti Metsad 1999)*; Eesti Metsakorralduskeskus: Tallinn, Estonia, 2000; p. 44.
54. FRA. Estonia Global Forest Resources Assessment 2015. Country Report. 2015. Available online: <https://openknowledge.fao.org/server/api/core/bitsstreams/5100a18e-1432-42b1-945e-398daac0176e/content> (accessed on 20 February 2024).
55. Zhang, C.; Gao, Z.; Null, S.Y.; Li, P. Edge Detectors Based on Pauta Criterion with Application to Hybrid Compact-WENO Finite Difference Scheme. *Adv. Appl. Math. Mech.* **2023**, *15*, 1379–1406. [[CrossRef](#)]
56. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
57. Breiman, L. *Bagging Predictors*; Kluwer Academic Publishers: New York, NY, USA, 1996; Volume 24, pp. 123–140.
58. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
59. Vapnik, V.; Golowich, S.; Smola, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 281–287.
60. Liu, K.; Wang, J.; Zeng, W.; Song, J. Comparison and Evaluation of Three Methods for Estimating Forest above Ground Biomass Using TM and GLAS Data. *Remote Sens.* **2017**, *9*, 341. [[CrossRef](#)]
61. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model Induction with Support Vector Machines: Introduction and Applications. *J. Comput. Civ. Eng.* **2001**, *15*, 208–216. [[CrossRef](#)]
62. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
63. Li, Y.; Li, C.; Li, M.; Liu, Z. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. *Forests* **2019**, *10*, 73. [[CrossRef](#)]
64. Lahssini, K.; Teste, F.; Dayal, K.R.; Durrieu, S.; Ienco, D.; Monnet, J.-M. Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4337–4348. [[CrossRef](#)]
65. Ma, T.; Hu, Y.; Wang, J.; Beckline, M.; Pang, D.; Chen, L.; Ni, X.; Li, X. A Novel Vegetation Index Approach Using Sentinel-2 Data and Random Forest Algorithm for Estimating Forest Stock Volume in the Helan Mountains, Ningxia, China. *Remote Sens.* **2023**, *15*, 1853. [[CrossRef](#)]
66. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the Capabilities of Sentinel-2 for Quantitative Estimation of Biophysical Variables in Vegetation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 83–92. [[CrossRef](#)]
67. Yu, T.; Pang, Y.; Liang, X.; Jia, W.; Bai, Y.; Fan, Y.; Chen, D.; Liu, X.; Deng, G.; Li, C.; et al. China’s Larch Stock Volume Estimation Using Sentinel-2 and LiDAR Data. *Geo-Spat. Inf. Sci.* **2023**, *26*, 392–405. [[CrossRef](#)]
68. Leite, R.V.; Amaral, C.H.d.; Pires, R.d.P.; Silva, C.A.; Soares, C.P.B.; Macedo, R.P.; Silva, A.A.L.d.; Broadbent, E.N.; Mohan, M.; Leite, H.G. Estimating Stem Volume in Eucalyptus Plantations Using Airborne LiDAR: A Comparison of Area- and Individual Tree-Based Approaches. *Remote Sens.* **2020**, *12*, 1513. [[CrossRef](#)]
69. Lin, W.; Lu, Y.; Li, G.; Jiang, X.; Lu, D. A Comparative Analysis of Modeling Approaches and Canopy Height-Based Data Sources for Mapping Forest Growing Stock Volume in a Northern Subtropical Ecosystem of China. *GISci. Remote Sens.* **2022**, *59*, 568–589. [[CrossRef](#)]

70. Saarela, S.; Grafström, A.; Ståhl, G.; Kangas, A.; Holopainen, M.; Tuominen, S.; Nordkvist, K.; Hyppä, J. Model-Assisted Estimation of Growing Stock Volume Using Different Combinations of LiDAR and Landsat Data as Auxiliary Information. *Remote Sens. Environ.* **2015**, *158*, 431–440. [[CrossRef](#)]
71. Fang, G.; He, X.; Weng, Y.; Fang, L. Texture Features Derived from Sentinel-2 Vegetation Indices for Estimating and Mapping Forest Growing Stock Volume. *Remote Sens.* **2023**, *15*, 2821. [[CrossRef](#)]
72. Korhonen, L.; Hadi; Packalen, P.; Rautiainen, M. Comparison of Sentinel-2 and Landsat 8 in the Estimation of Boreal Forest Canopy Cover and Leaf Area Index. *Remote Sens. Environ.* **2017**, *195*, 259–274. [[CrossRef](#)]
73. Saarela, S.; Wästlund, A.; Holmström, E.; Mensah, A.A.; Holm, S.; Nilsson, M.; Fridman, J.; Ståhl, G. Mapping Aboveground Biomass and Its Prediction Uncertainty Using LiDAR and Field Data, Accounting for Tree-Level Allometric and LiDAR Model Errors. *For. Ecosyst.* **2020**, *7*, 43. [[CrossRef](#)]
74. Chen, L.; Ren, C.; Zhang, B.; Wang, Z. Multi-Sensor Prediction of Stand Volume by a Hybrid Model of Support Vector Machine for Regression Kriging. *Forests* **2020**, *11*, 296. [[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.