



Article

High-Resolution Canopy Height Mapping: Integrating NASA's Global Ecosystem Dynamics Investigation (GEDI) with Multi-Source Remote Sensing Data

Cesar Alvites ^{1,*}, Hannah O'Sullivan ^{2,3,4}, Saverio Francini ^{5,6}, Marco Marchetti ¹, Giovanni Santopuoli ⁷, Gherardo Chirici ^{5,8}, Bruno Lasserre ¹, Michela Marignani ⁹ and Erika Bazzato ^{9,10}

- ¹ Department of Biosciences and Territory, University of Molise, Cda Fonte Lappone Snc, 86090 Pesche, Italy; marchettimarco@unimol.it (M.M.); lasserre@unimol.it (B.L.)
 - ² Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK; h.osullivan18@imperial.ac.uk
 - ³ Royal Botanic Gardens, Kew, Richmond TW9 3AB, UK
 - ⁴ Department of Forest Sciences, University of Helsinki, 00100 Helsinki, Finland
 - ⁵ Department of Agriculture, Food, Environment and Forestry, University of Florence, Via San Bonaventura 13, 50145 Florence, Italy; saverio.francini@unifi.it (S.F.); gherardo.chirici@unifi.it (G.C.)
 - ⁶ NBFC, National Biodiversity Future Center, 90133 Palermo, Italy
 - ⁷ Department of Agricultural, Environmental and Food Sciences, University of Molise, Via De Sanctis 1, 86100 Campobasso, Italy; giovanni.santopuoli@unimol.it
 - ⁸ Fondazione per il Futuro delle Città, 50133 Firenze, Italy
 - ⁹ Department of Life and Environmental Sciences, University of Cagliari, Via Sant'Ignazio da Laconi 13, 09123 Cagliari, Italy; marignani@unica.it (M.M.); erika.bazzato@hotmail.it (E.B.)
 - ¹⁰ Department of Agricultural Sciences, University of Sassari, Viale Italia 39, 07100 Sassari, Italy
- * Correspondence: cesar.alvites@unimol.it



Citation: Alvites, C.; O'Sullivan, H.; Francini, S.; Marchetti, M.; Santopuoli, G.; Chirici, G.; Lasserre, B.; Marignani, M.; Bazzato, E. High-Resolution Canopy Height Mapping: Integrating NASA's Global Ecosystem Dynamics Investigation (GEDI) with Multi-Source Remote Sensing Data. *Remote Sens.* **2024**, *16*, 1281. <https://doi.org/10.3390/rs16071281>

Academic Editor: Luis A. Ruiz

Received: 16 December 2023

Revised: 30 March 2024

Accepted: 2 April 2024

Published: 5 April 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/>).

Abstract: Accurate structural information about forests, including canopy heights and diameters, is crucial for quantifying tree volume, biomass, and carbon stocks, enabling effective forest ecosystem management, particularly in response to changing environmental conditions. Since late 2018, NASA's Global Ecosystem Dynamics Investigation (GEDI) mission has monitored global canopy structure using a satellite Light Detection and Ranging (LiDAR) instrument. While GEDI has collected billions of LiDAR shots across a near-global range (between 51.6°N and >51.6°S), their spatial distribution remains dispersed, posing challenges for achieving complete forest coverage. This study proposes and evaluates an approach that generates high-resolution canopy height maps by integrating GEDI data with Sentinel-1, Sentinel-2, and topographical ancillary data through three machine learning (ML) algorithms: random forests (RF), gradient tree boost (GB), and classification and regression trees (CART). To achieve this, the secondary aims included the following: (1) to assess the performance of three ML algorithms, RF, GB, and CART, in predicting canopy heights, (2) to evaluate the performance of our canopy height maps using reference canopy height from canopy height models (CHMs), and (3) to compare our canopy height maps with other two existing canopy height maps. RF and GB were the top-performing algorithms, achieving the best 13.32% and 16% root mean squared error for broadleaf and coniferous forests, respectively. Validation of the proposed approach revealed that the 100th and 98th percentile, followed by the average of the 75th, 90th, 95th, and 100th percentiles (AVG), were the most accurate GEDI metrics for predicting real canopy heights. Comparisons between predicted and reference CHMs demonstrated accurate predictions for coniferous stands (R-squared = 0.45, RMSE = 29.16%).

Keywords: Global Ecosystem Dynamics Investigation (GEDI); Sentinel data; NASA; downscaling; canopy heights; forest structure

1. Introduction

Forests play a vital role in regulating the carbon and water cycles, supporting biodiversity and providing economic benefits to society [1]. However, the stability of forests, and the ecosystem-services they provide, are increasingly threatened by anthropogenic change. Consequently, regular monitoring across large spatial scales is needed for effective forest conservation and management [2]. Forests have been recognised as an important nature-based solution to climate change, as they remove CO₂ from the atmosphere and store it as biomass. This has led to increased efforts to quantify global forest biomass and carbon stocks. In particular, remote sensing (RS) technologies have been used to assess canopy height, a key variable for estimating above-ground biomass and ultimately carbon stocks, as well as for identifying ecosystem services [3–6]. Satellite-based measurements of canopy height are now available on an unprecedented global scale [7–10]; however, there remain certain limitations that hinder the use of these data for localised forest studies.

LiDAR (Light Detection and Ranging) is one of the most useful remote sensing tools for quantifying forest structure. Different LiDAR instruments have different spatial resolutions depending on the distance from the sensor to the object, from terrestrial (millimetric) to airborne (centimetric) to satellite LiDAR systems (metric) [11,12]. Terrestrial LiDAR is by far the most accurate system for capturing forest structure, but the cost and time required to complete a survey means that these systems are rarely used [11]. However, airborne laser scanning (ALS) is a more cost-effective option, capable of surveying forest stands in a relatively short amount of time [11,13,14]. Consequently, ALS is one of the most frequently used LiDAR systems for field-based forest surveys in most EU countries. Finally, satellite LiDAR monitors changes in forest structure on a near-global scale. The Global Ecosystem Dynamics Investigation (GEDI) NASA mission characterises the structure of Earth's forests using a LiDAR instrument onboard the International Space Station (between 51.6°N and S) [15]. Since its launch, GEDI's satellite products have been widely adopted for forest-related research [16–18]. For instance, GEDI's products have been used to create forest composition maps [19], quantify carbon stocks [16], investigate the relationship between biodiversity and forest structure [20], and detect human logging activities, forest disturbance, and changes in forest structure [21–23].

The GEDI mission has improved our ability to understand and monitor changes in forest structure, but there are still a number of caveats to the data. The Level-2A Geolocated Elevation and Height Metrics GEDI Product (GEDI02_A) consists of a hundred relative height (Rh) metrics with a 25 m pixel spatial resolution (average footprint size) [15] (<https://gedi.umd.edu/data/products/>; accessed on 11 September 2023). This large pixel size can be problematic for stand-level forest monitoring, since a small stand might only have a few pixels covering the area. Furthermore, relative height averaged within a 25 m pixel loses information on structural heterogeneity. Another key caveat for using GEDI data is that they are discontinuous, resulting in large quantities of missing data. This is especially problematic for estimating above ground biomass and carbon stocks, since it leaves many areas unaccounted. The quality of GEDI's acquisition accuracy also depends on various factors, including footprint geolocation and footprint variability [22,24]. Consequently, although GEDI data are appropriate for investigating global patterns in forest structure, the large spatial resolution of individual pixels and discontinuous acquisitions pose challenges at smaller geographical scales. Trees are the fundamental building blocks of forests, and management decisions are ultimately conducted on a tree-by-tree basis. Therefore, downscaling GEDI data to enhance pixel resolution and extend coverage has become an area of active research in recent years (Table 1).

In order to achieve high-resolution, full-coverage canopy height maps, numerous studies have used machine learning (ML) for downscaling and spatializing satellite-LiDAR data. These studies often combine satellite LiDAR with satellite imagery to reduce the original footprint size and extend coverage for entire forest ecosystems [25–29] (Table 1). ML algorithms have proven effective in interpreting complex patterns within RS datasets, adjusting for overfitting biases, and efficiently handling large volumes of data [30]. Cur-

rently, the most widely used ML methods include random forests (RF), deep learning (DL), and gradient tree boost (GB) (Table 1). Despite the prevalence of RF and DL algorithms, there remains a gap in utilizing classification and regression trees (CART) algorithms for downscaling purposes (Table 1).

Table 1. Canopy height assessment studies using satellite LiDAR (light detection and ranging) data. The accuracy measurements from these studies are mean absolute error (MAE), coefficient of correlation (r) and determination (R-squared), bias, root mean square error (RMSE), overall accuracy (OA), and mean error (ME). Deep learning (DL), random forests (RF), linear model (LM), gradient tree boost (GB), ordinary least squares (OLS), convolutional neural network (CNN), NA for unknown methodology, Landsat (LDT), Sentinel-1 (S1), Sentinel-2 (S2), the ice, cloud, and land elevation satellite (ICESat), Global Ecosystem Dynamics Investigation (GEDI), and the National Terrestrial Ecosystem Monitoring System (NTEMS).

Site	Year	Methods	Dependent Variables	Independent Variables	Map Accuracy		Study
					Output Pixel-Based	Statistic Measurements	
Global map	2000–2017	NA	ICESat	LDT	30 m	MAE = 3.7 m; R-squared = 0.85–0.92	[31]
China's forest	2017–2019	DL and RF	ICESat-2	S1, S2 and LDT8	10 m–30 m–250 m–500 m–1000 m	R-squared = 0.68–0.78; bias = −1.46 m	[29]
USA	2019–2021	RF	GEDI	S1 and S2	30 m	$r = 0.58$; RMSE = 4.46 m	[32]
Canada	2019	LM (i.e., OLS)	ICESat-2	NTEMS (validation)	100 m segments	$r = 0.61$; mean difference = 0.55 m	[8]
Global map	April–October 2019	RF	GEDI	LDT	30 m	RMSE = 6.6 m; MAE = 4.45 m, R-squared = 0.62	[27]
China, France, and the United States	2019	RF	GEDI	S2	10 m	OA China = 0.89; OA France = 0.85; OA US = 0.91	[26]
Global map	2020	DL (i.e., CNN)	GEDI	S2	10 m	RMSE = 9.6 m; MAE = 7.4 m; ME = −4.8 m	[28]
Australia and the United States	2020	GB	GEDI	S1 and S2	100 m–200 m	R-squared of 0.66–0.74; RMSE of 41–77%	[25]

Google Earth Engine (GEE) is widely used for remote sensing analyses due to its cloud-based computing architecture and easy access to multi-temporal global satellite data, removing the computational limitations associated with local analysis [23,33–35]. This capability has empowered researchers to use a cloud-based platform to analyse petabytes of RS images and generate canopy height raster data, predominantly leveraging Sentinel-1 (S1), Sentinel-2 (S2), and Landsat (LDT) as data sources (Table 1). For instance, Potapov et al. [27] used a bagged regression trees ensemble method to merge GEDI data with multi-temporal Landsat images, to produce a global canopy height map with a spatial resolution of 30 m (hereafter CH-Potapov2019). However, they observed that the low resolution of Landsat prompted an overestimation in measurements of forest canopy height in temperate forests. A comparable map was produced by Lang et al. [28] using multi-temporal S2 images and a deep learning method (hereafter CH-Lang2020) without considering locally calibrated models.

However, none of the previous studies include topographic characteristics as independent variables for predicting canopy heights, despite the demonstrated signifi-

cance of topography in enhancing the accuracy of GEDI footprint measurements in forest canopies [25,26].

Although the downscaling method is quite similar across the previously mentioned studies, a significant difference lies in the metric used as a proxy to predict the top-of-canopies. For instance, the 90th, 95th, and 98th percentiles were most frequently used as Rh metrics (from 90th to Rh90) [27,28,36,37]. Interestingly, Potapov et al. [27] revealed that Rh90 tended to underestimate canopy height, whilst Rh100 tended to overestimate canopy height. Furthermore, to the best of our knowledge, most of the developed canopy height maps using locally calibrated models were focused on specific areas [8,25,32]. By contrast, global canopy height maps were year-specific (i.e., 2019 and 2020) and utilized GEDI footprints at the national or continental level [27,28]. Given these lacks (i.e., absence of topographic predictors, adoption of single ML models, and individual relative height without locally calibrations), there is a pressing need to investigate the effect of different Rh metrics on predicting canopy height, using locally calibrated ML models to investigate the influence of additional RS and topographic data on model outcomes.

In response to these gaps, this study proposes and evaluates an approach to generate high-resolution canopy height maps (10 m) by combining GEDI with S1, S2, and topographical data using different locally calibrated ML algorithms. To reach this aim, this study dealt with the following three specific research objectives:

- (1) To assess the performance of three ML algorithms, RF, GB, and CART, in predicting canopy heights from the most commonly used GEDI metrics.
- (2) To evaluate the performance of our canopy height maps using reference ALS-based CHMs.
- (3) To compare our canopy height maps with two existing canopy height maps, CH-Potapov2019 [27] and CH-Lang2020 [28].

The proposed approach is tested in two structurally contrasting Mediterranean coniferous and broadleaved forest sites, as they represent two of the most important European forest ecosystems with contrasting forest stand structures [2].

2. Study Area

Two Mediterranean forest test sites were selected, based on their contrasting structural and demographic profiles. Both sites are located in central Italy, known locally as Pennataro (41°44'5.97"N, 14°12'0.79"W) and Lago di Occhito (41°37'17"N, 14°58'19"W). Under the European Forest Type classification, Pennataro is an oak–hornbeam forest, and Lago di Occhito is a Mediterranean pine forest [38,39]. Lago di Occhito is characterized by approximately 997 ha of structurally homogenous forests, whilst Pennataro is covered by approximately 270 ha of structurally heterogeneous forests (Figure 1) [40,41]. Pennataro is a mixed broadleaved tree species, dominated by Turkey oak (*Quercus cerris*, 40%) and including European beech (*Fagus sylvatica*, 21%) and Italian maple (*Acer opalus*, 9.6%) [42]. In Lago di Occhito, forest plantation is characterized by Aleppo pine (*Pinus halepensis*, 61%), which is the dominant species, and Arizona cypress (*Cupressus arizonica*, 20%), plus a limited number of other coniferous tree species.

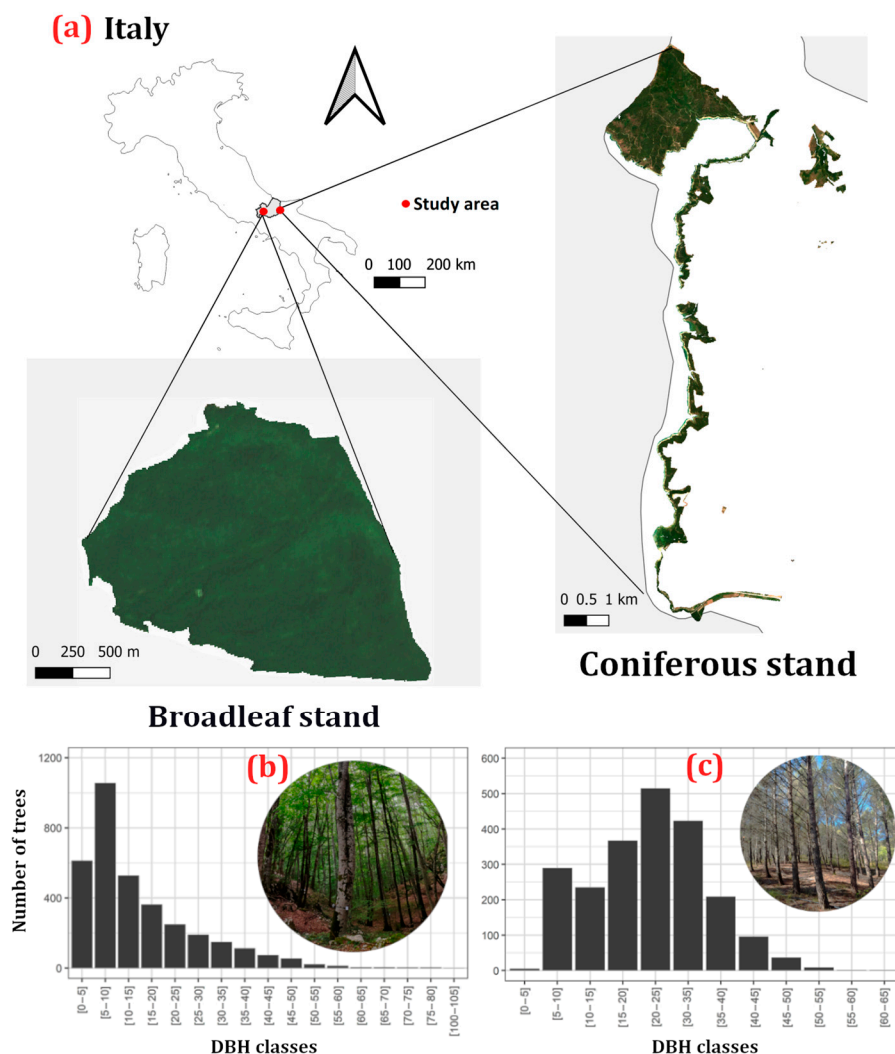


Figure 1. Study site locations and diametric class distribution of trees. (a) Locations visualised using a single RGB Sentinel – 2 image. (b) The diametric class distribution of trees in Pennataro (broadleaf stand). (c) The diametric class distribution of trees in Lago di Occhito (coniferous stand). The diametric class distribution of trees in Pennataro is positively skewed, indicating higher structural heterogeneity, compared to Lago di Occhito, which follows a normal distribution, indicating structural homogeneity.

3. Data

The following four main data sources were used in this study for model development and validation: (1) ALS data collection and processing for canopy height validation (Section 3.1); (2) GEDI relative height metrics to be downscaled (Section 3.2); (3) multi-spectral Sentinel-1 and Sentinel-2 satellite imagery for canopy height prediction (Section 3.3); and (4) topographical features such as elevation, slope, and aspect, as additional predictors for canopy height (Section 3.4). In addition, we used two existing GEDI-derived canopy height maps for further validation (Section 3.5).

3.1. Airborne Laser Scanning Data Collection and Processing

ALS data were acquired for both the broadleaf and coniferous forest sites. ALS data for the broadleaf forest site were collected in June 2016 during the leaf-on season, while data for the coniferous forest site were collected in July 2021. YellowScan Mapper+ sensors were used and mounted on a helicopter for flying over a broadleaf forest site and on a Matrice 300 RTK unmanned aerial vehicle (UAV) for flying over a coniferous forest site. In the broadleaf site, the sensor was configured with a maximum scan angle of $\pm 50^\circ$ and a

pulse frequency of 20 kHz, resulting in an average point density of 60/m². In the coniferous site, a most recent version of YellowScan Mapper+ sensor was used, mounted on a UAV flying at an altitude of 70 m above the canopy. This resulted in a higher scanning frequency of 10 Hz, generating a denser point cloud with an average density of 300 points/m² (<https://www.yellowscan-lidar.com/products/mapper-3/>, accessed on 3 March 2024). ALS data collection covered 90% of the broadleaf forest site and 16% of the coniferous forest site.

The following four main steps were used for generating a canopy height map from the ALS data: (1) point classification as ground/non-ground; (2) outlier removal; (3) Z-point normalization; and (4) CHM generation. The 'lidR' package [42] and 'rlas' package [43] (R version 4.3.0) were used to create CHMs [44]. The final CHMs (.tiff) with a spatial resolution of 1 m, for broadleaf and coniferous forest sites, were obtained. Hence, we standardized the spatial resolution of the CHMs of the study areas to 1 m, to ensure consistency in the validation approach and outputs.

3.2. Global Ecosystem Dynamics Investigation (GEDI) Level-2A Data

GEDI footprint and gridded datasets come in different spatial resolutions for 3D Earth observation [45]. Since we were focused on canopy height, we used the GEDI—L2A (Level-2A) top-of-canopy and relative height metrics. This dataset has high geolocation accuracy, equivalent to a horizontal error of 10.3 m [46], and comprises 25 m footprints, acquired at 60 m intervals along the track and 600 m intervals across the track [15].

Three quality control filters, "sensitivity", "quality", and "degrade flags", were used to filter the GEDI—L2A data. These flags indicate waveform reliability for measuring 3D surface structure [46,47]. Waveforms unsuitable for measuring potential top canopy height (quality flag = 1) were excluded, and only non-degraded samples (degrade flag = 0) were selected. Subsequently, four metrics, Rh90, Rh95, Rh98, and Rh100, were chosen due to their potential to accurately measure the top-of-canopy height [27,28,35,36]. Additionally, a metric based on the average of Rh75, Rh90, Rh95, and Rh100 (hereafter the AVG metric) was incorporated into this study. Rh75 is an effective metric for characterizing forest structure [48], but it is rarely used in deriving top-of-canopy heights. We hypothesize that integrating the Rh75 metric into an AVG metric will allow us to mitigate under- and over-estimations associated with Rh90 and Rh100 metrics [27].

3.3. Sentinel Mission Data

GEE was used to access harmonised Sentinel-2 Level-2A data and Sentinel-1 ground range detected (GRD) data between 1 July and 31 August 2023. To reduce leaf canopy occlusion, leaf-off canopy conditions were selected for the study period [49]. Sentinel-1 bands, with both ascending and descending orbits, were selected using the IW (interferometric wide) mode, including single ("VV" and "HH") and dual polarization ("VH") signals [50]. The Sentinel-1 bands were pre-processed using border noise correction, speckle filtering, and radiometric terrain normalization [51]. For the broadleaf forest site, 116 multi-temporal images were available, and 125 multi-temporal images for the broadleaf and coniferous forest sites, respectively, from S1 during the setting period. Each available image contained three channels, including the incident angle range ("Angle"), as well as single and dual polarizations ("VV"—"VH") bands. As preprocessing, a spatial filtering operation that computes the median value within a neighbourhood around each pixel in an image was applied to the previously acquired multi-temporal images. This process performed a composite band [51].

For the study period, 25 Sentinel-2 images were available for the broadleaf site and 13 for the coniferous site, after ensuring all images had less than 70% cloud cover. Images were masked using the QA60 band to limit the presence of opaque and cirrus clouds [23,52]. A final median composite image was then generated for each study site, excluding clouds and shadows [52–54].

3.4. Topographical Data

Three topographical variables were included (elevation, slope, and aspect) since they have been demonstrated to significantly impact forest structure [22,55–57]. The Global Multi-Resolution Terrain Elevation Data (GMTE)-2010 dataset, available from GEE, was used to compute slope and aspect in addition to elevation [58].

3.5. Existing Global Ecosystem Dynamics Investigation (GEDI)-Derived Canopy Height Maps

Two previously published global canopy height maps are readily accessible through GEE [28] and the Earth Map platform [59]. Both maps integrate 3D GEDI data with multi-spectral and multi-temporal satellite imagery for downscaling, making them ideal candidates for benchmarking future canopy height maps. The first canopy height map, referred to as CH-Lang2020 [28], predicts canopy height with a spatial resolution of 10 m. The second map, referred to as CH-Potapov2019 [27], predicts canopy height with a spatial resolution of 30 m. The CH-Potapov2019 map uses Rh95 to represent top-of-canopy, whilst CH-Lang2020 uses Rh98. Landsat satellite imagery and land surface elevation were used in CH-Potapov2019, whereas Sentinel-2 imagery and land surface elevation was in CH-Lang2020, resulting in a higher resolution of 10 m. For both studies, map accuracy was validated using ALS data, with CH-Potapov2019 achieving a RMSE of 9.07 m (Mean Absolute Error “MAE” = 6.36 m) and an RMSE of 6 m (MAE = 4 m) for CH-Lang2020 [28].

4. Methods

In this study we leveraged the readily available data and computational tools in GEE to develop canopy height maps for two distinct Mediterranean forest types [33]. The main steps of the workflow are as follows (Figure 2): (1) feature selection, model construction, and performance evaluation (Section 4.1); (2) map validation with ALS-based CHMs (Section 4.2); and (3) benchmarking with existing canopy height maps from CH-Potapov2019 and CH-Lang2020 (Section 4.3).

4.1. Canopy Height Map Prediction

GEE offers a number of built-in machine learning tools for supervised classification, unsupervised classification, and regression. To predict canopy heights from different GEDI Rh metrics, we used 18 predictors in total, including multi-spectral Sentinel-1 (three predictors) and Sentinel-2 (twelve predictors) imagery and ancillary topographical data (three predictors). The dataset was split into training (70% corresponding to 2586 and 2464 samples in broadleaf and coniferous forests, respectively) and testing (30% corresponding to 1108 and 1056 samples in broadleaf and coniferous forests, respectively) data [53,60–62], using a random sampling strategy within the masked forest sites. To downscale and spatialize GEDI data, three ML algorithms from the GEE Classifier were used: RF, GB, and classification and regression (CART). These algorithms were selected based on their demonstrated suitability for predicting canopy height and robustness to overfitting, due to their potential in applying decision tree analysis, enabling regression analysis for large datasets, assessing feature importance across variables (thus reducing overfitting), and replicating human reasoning in data processing [17,60,61]. The number of trees was set to 500 for the RF and GB models, with other hyperparameters set to default values (see Table 2) [63–65]. The CART models were run with default hyperparameters [66]. In total, 30 ML models were trained: five GEDI metrics (Rh90, Rh95, Rh98, Rh100, and AVG) were analysed using the RF, GB, and CART algorithms to obtain 15 canopy height maps, with 10 m pixel resolution, for each study area (i.e., the coniferous site and the broadleaf site). The testing data were used to validate the performance of the ML models through the coefficient of determination (R-squared, ranging from zero to one) and root mean squared error (RMSE, %). R-squared was used to gauge the extent to which ML models explain variability, whilst RMSE was used to assess the predictive accuracy of each model and map (lower RMSE values indicate more precise and reliable predicted outcomes).

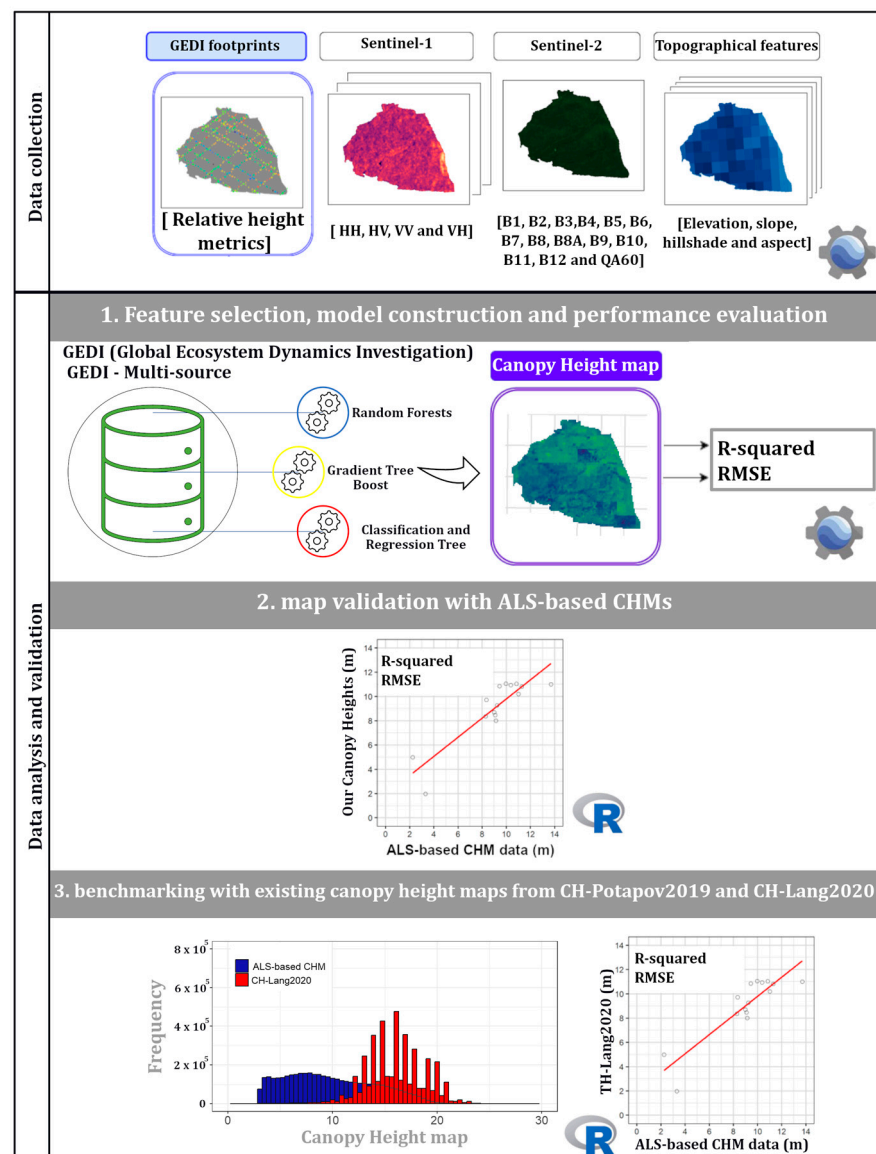


Figure 2. Canopy height map workflow. Random forests, gradient tree boost, and classification and regression trees were used to generate canopy height maps. The resulting canopy height map was validated using canopy height models from airborne laser scanning (ALS-based CHM) data through root mean square error (RMSE) and coefficient determination (R-squared).

4.2. Comparison of Predicted Canopy Height Maps with Reference Airborne Laser Scanning (ALS)-Based Canopy Height Models (CHMs)

To evaluate the performance of the resulting canopy height maps, we compared the predicted heights with a reference ALS-based CHM map. The maps were resampled and aligned using a nearest neighbour method, as we relied on the assumption that near point data tend to be more similar than far point data. As it is considered suitable for continuous data populations and forestry studies [67], the nearest neighbour method was applied to complete the following tasks: (1) to resample our canopy height maps (10 m) to the resolution of ALS-based CHM map (1 m); and (2) to check errors in raster alignment and adjust alignment using ALS-based CHM map as snap raster. Subsequently, we applied Cook's distance method to remove influential values (outliers) from pixel values from the predicted and reference maps [68–70]. Then, the predicted and reference data were compared using the linear regression model. Finally, we assessed map accuracy using

RMSE (%) and the variability explained by the regression models using R-squared (which ranged from zero to one) [71].

Table 2. Machine learning algorithms and their parameter settings to predict GEDI canopy height Rh metrics, based on Sentinel-1, Sentinel-2, and topographical data. Random Forests (RF), Gradient Tree Boost (GB), and Classification and Regression Trees (CART) are configured.

Machine Learning Algorithms	Parameter Name	Parameter Description	Parameter Setting
RF	numberOfTrees	Decision tree number	500
	variablesPerSplit	Number of variables per split (mtry)	4
	minLeafPopulation	Minimum number of training samples in each leaf node	1
	bagFraction	Input fraction to bag per tree	0.5
	maxNodes	Maximum number of leaf nodes in each tree	no limit
GB	numberOfTrees	Decision tree number	500
	shrinkage	Learning rate	0.005
	samplingRate	Sampling rate for stochastic tree boosting	0.7
	maxNodes	Maximum number of leaf nodes in each tree	no limit
	loss	Loss function for regression	LeastAbsoluteDeviation
CART	maxNodes	Maximum number of leaf nodes in each tree	no limit
	minLeafPopulation	Minimum number of training samples in each leaf node	1

4.3. Comparison of Predicted Canopy Height Maps with Other Existing Global Ecosystem Dynamics Investigation (GEDI)-Derived Canopy Height Maps

To further evaluate the robustness of the predicted canopy height maps, we compared map accuracy statistics with those of CH-Potapov2019 and CH-Lang2020 for the same ALS-based reference CHM. The existing CH-Potapov2019 and CH-Lang2020 maps were processed using the same method outlined in Section 4.2 to allow for fair comparisons. The overall evaluations were made using the R-squared (which ranged from zero to one), RMSE (%), and pixel frequency distribution [27,28].

5. Results

5.1. Canopy Heights Map Prediction

A single canopy height map was constructed for each of the study sites with 10 m pixel resolution. Contrasting results were obtained for the coniferous and broadleaf sites, with R-squared values ranging from 0.61 to 0.93 (RMSE % = 16.35–26.58) in the coniferous site and from 0.64 to 0.78 (RMSE % = 13.32–17.79) in the broadleaf site (Figure 3). Notably, the RF and GB models achieved better predictive performance than the CART models. In fact, the CART models exhibited low accuracy, with the highest RMSE reaching 17.79% in the broadleaf site and 26.58% in the coniferous site. Some variation was observed in the suitability of different GEDI Rh metrics for predicting canopy heights. We found that, in the study sites, the Rh98, Rh100, and AVG performed slightly better than Rh90 and Rh95, but no significant difference was found overall.

In the coniferous site, the RF and GB models produced the most accurate canopy height maps compared to the CART models (Figure 4). The maps derived from CART models exhibited pixelation, blurriness, and inconsistencies in brightness (see Figure 4).

In the broadleaf site, the RF and GB models also outperformed the CART models. However, the difference between the RF (RMSE = 2.61%) and GB (RMSE = 3.32%) models was minimal in terms of RMSE % (see Figures 3 and 5). CART-derived maps exhibited greater pixelation, blurriness, and brightness inconsistencies compared to RF and GB maps (Figure 5).

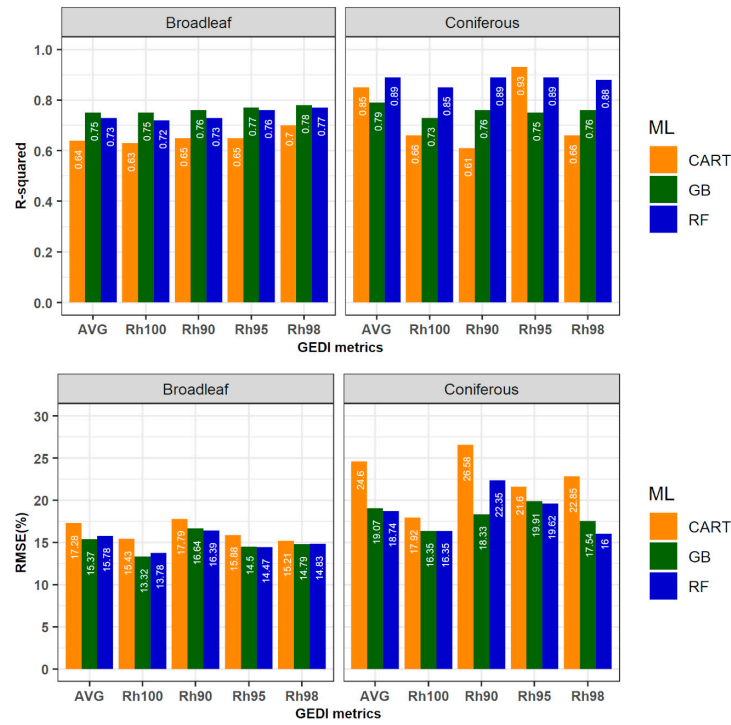


Figure 3. Map accuracy of downscaled canopy height maps. Five GEDI (Global Ecosystem Dynamics Investigation) metrics (Rh; Rh90, Rh95, Rh98, Rh100, and AVG of Rh75, Rh90, Rh95, Rh100) were predicted using three machine learning (ML) algorithms, namely RF, GB, and CART.

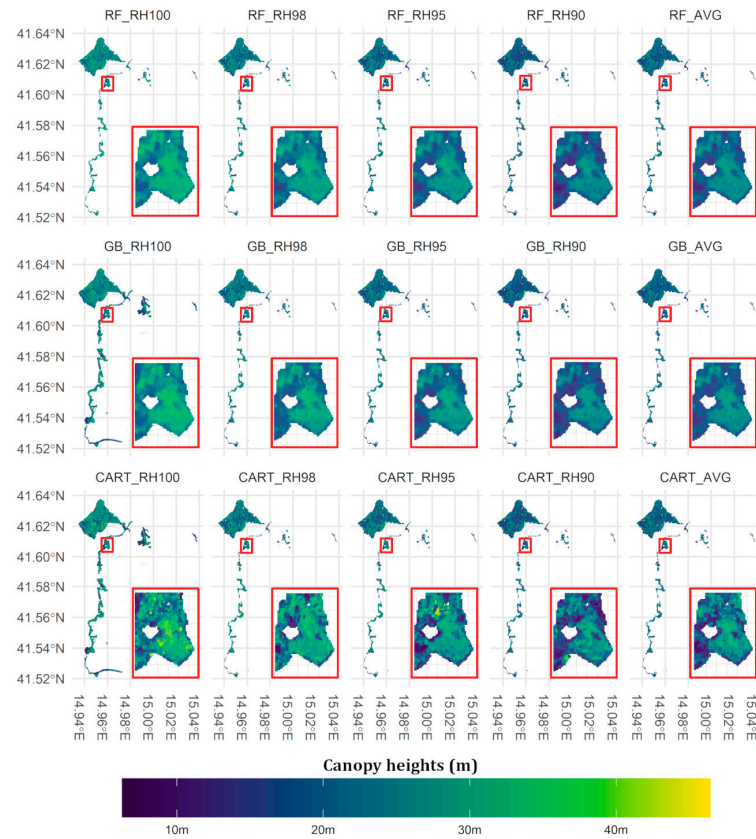


Figure 4. Predicted canopy height maps for each GEDI Rh metric and ML algorithm for the coniferous forest site. A subsection has been enlarged for interpretability.

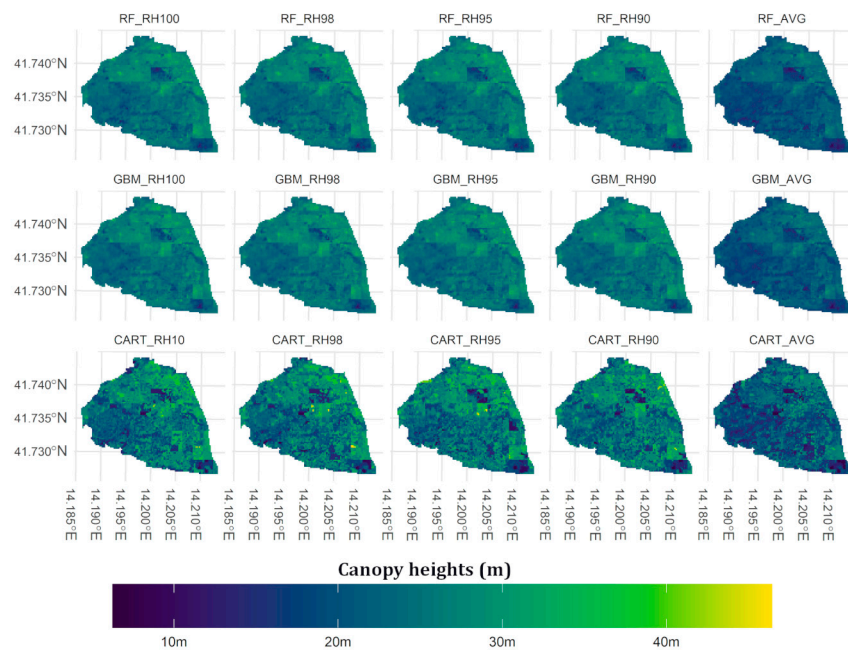


Figure 5. The predicted canopy height maps for all GEDI relative height metrics and ML algorithms for the broadleaf forest site.

The results revealed contrasting patterns in variable importance across the models and forest types. The most influential predictors in RF models developed for predicting canopy height in the broadleaf site were topographic data (slope, aspect, and DEM—digital elevation model) and Sentinel-2 band B1 (Figure S1). In contrast, RF models trained in the coniferous site relied more heavily on a combination of Sentinel-1, Sentinel-2, and topographic data (VH, band B11, aspect, and slope) (Figure S2). Conversely, the most important variable in training GB models for both forest sites was the B1 Sentinel-2 band.

CART models exhibited the most significant disparity in key predictors between forest types; Sentinel-2 bands (B1, B9) and topographic data (slope, DEM) held the most weight in broadleaf stands (Figure S1), whereas, in the coniferous site, Sentinel-1/2 bands (i.e., B1, B11, VH, and VV) and slope/aspect emerged as the most influential predictors of canopy heights (Figure S2).

5.2. Comparison of Predicted Canopy Height Map with Reference Airborne Laser Scanning-Based Canopy Height Models

When we compared the predicted canopy heights against the site-specific ALS-based CHMs, we found a better fit with coniferous (best model: a RMSE of 29.16% with an R-squared of 0.45; Figure 6) than for broadleaf forest sites (best model: a RMSE of 20.94% with an R-squared of 0.14) (Figure 7). We observed that the RF and GB models outperformed the CART models in the coniferous forest (see Figure 6). The Rh90 and AVG GEDI metrics enabled us to derive more accurate canopy heights in coniferous stands, as indicated by the lower RMSE values, ranging between 29.16% and 37.42% for Rh90 and 30.15% and 38.69% for AVG GEDI, respectively.

Our findings highlighted that the RF and GB models outperformed the CART models when predicting canopy height in the broadleaf site (Figure 7). Leveraging the AVG GEDI metric allowed us to obtain more accurate canopy heights in broadleaf stands, as evidenced by the low RMSE values ranging between 20.94% and 26.54%.

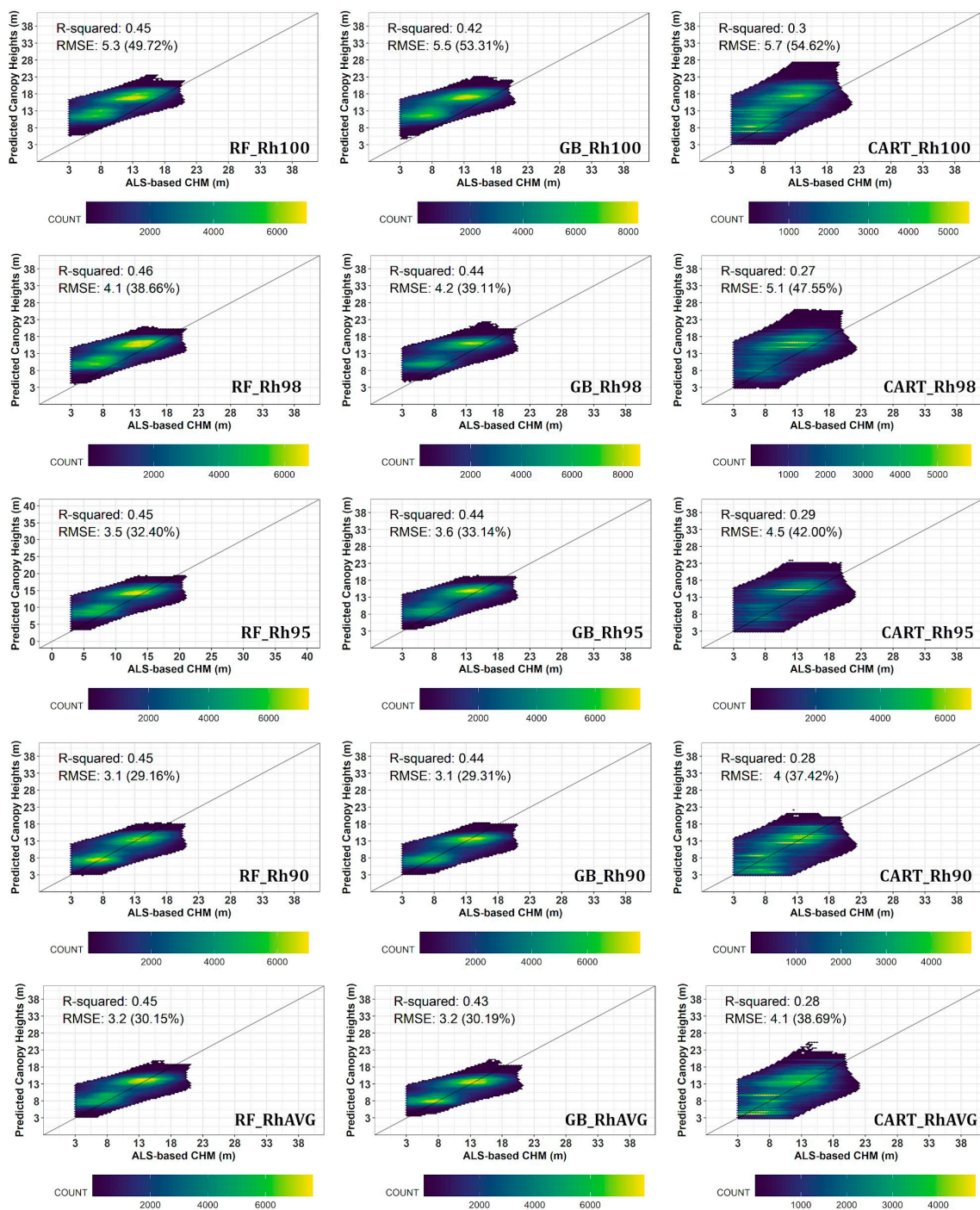


Figure 6. Scatter plots comparing predicted and reference canopy heights in the coniferous site. Five GEDI metrics (Rh90, Rh95, Rh98, Rh100, and the average of Rh75, Rh90, Rh95, and Rh100—AVG) were processed through RF (Random Forests), GB (Gradient Tree Boost), and CART (Classification and Regression Trees) algorithms. Root mean square error (RMSE in meter and percentage) and R-squared values are presented.

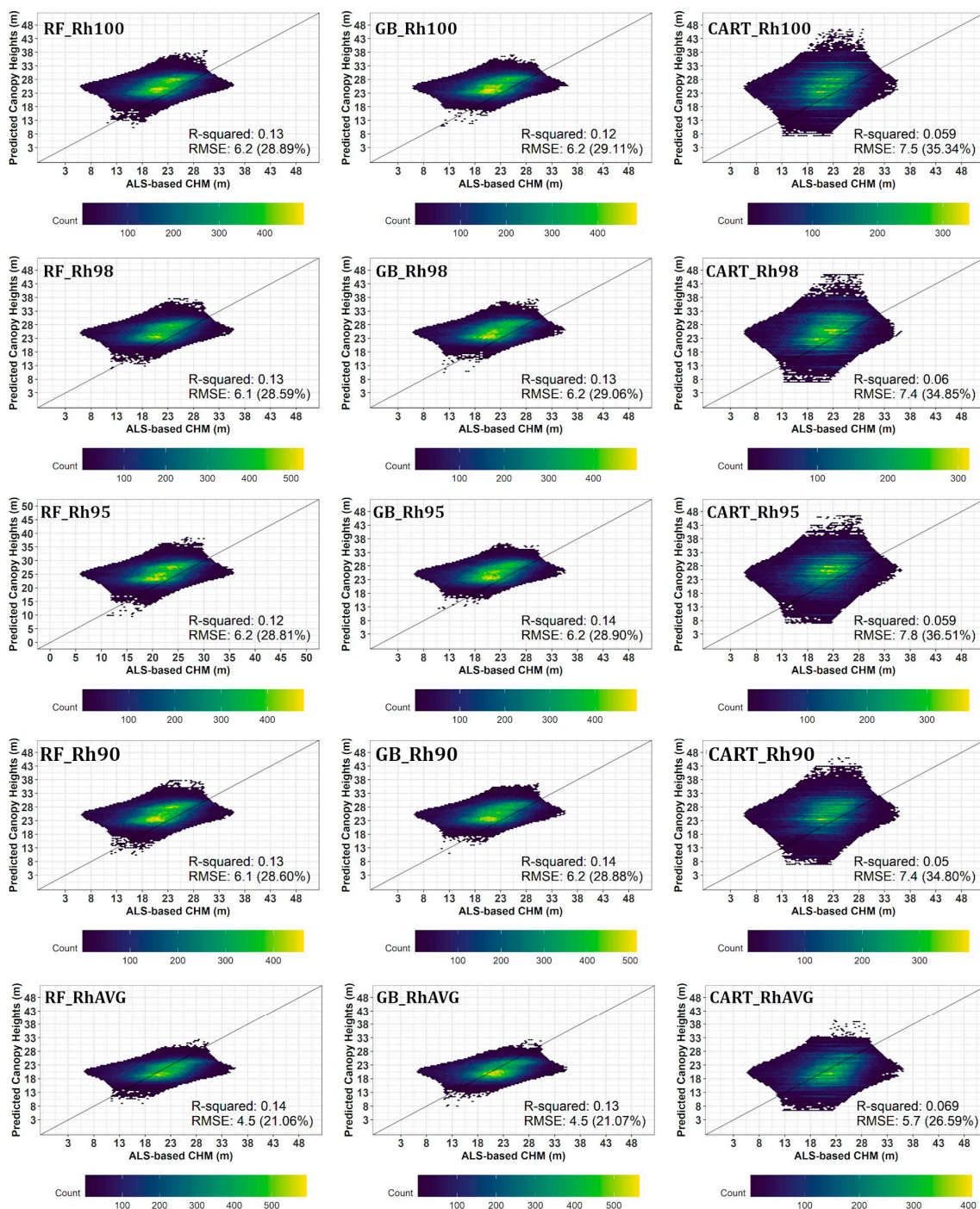


Figure 7. Scatter plots comparing predicted and reference canopy heights in the broadleaf site. Five GEDI metrics (Rh90, Rh95, Rh98, Rh100, and the average of Rh75, Rh90, Rh95, and Rh100—AVG) were processed through RF (Random Forests), GB (Gradient Tree Boost), and CART (Classification and Regression Trees) algorithms. Root mean square error (RMSE in meter and percentage) and R-squared values are presented.

5.3. Comparison of Predicted Canopy Height Maps with Other Existing Global Ecosystem Dynamics Investigation (GEDI)-Derived Canopy Height Maps

Our canopy height predictions were a better fit with the ALS-based CHM in both forest sites than those obtained using the CH-Potapov2019 and CH-Lang2020 maps (see Figures 8 and 9). Contrasting findings emerged for the two available maps across forest sites. For instance, when using canopy heights obtained from the coniferous site, the

CH-Potapov2019 demonstrated a more accurate estimation (RMSE = 49.68%; Figure 8), compared to the CH-Lang2020 (RMSE = 75.67%; Figure 8). In contrast, somewhat similar results were obtained for broadleaf trees using the CH-Lang2020 map (RMSE = 22.54%; Figure 8) and the CH-Potapov2019 map (RMSE = 22.91%; Figure 8). In the coniferous site, both existing maps have a low explanatory power for variability, with R-squared values below 0.19.

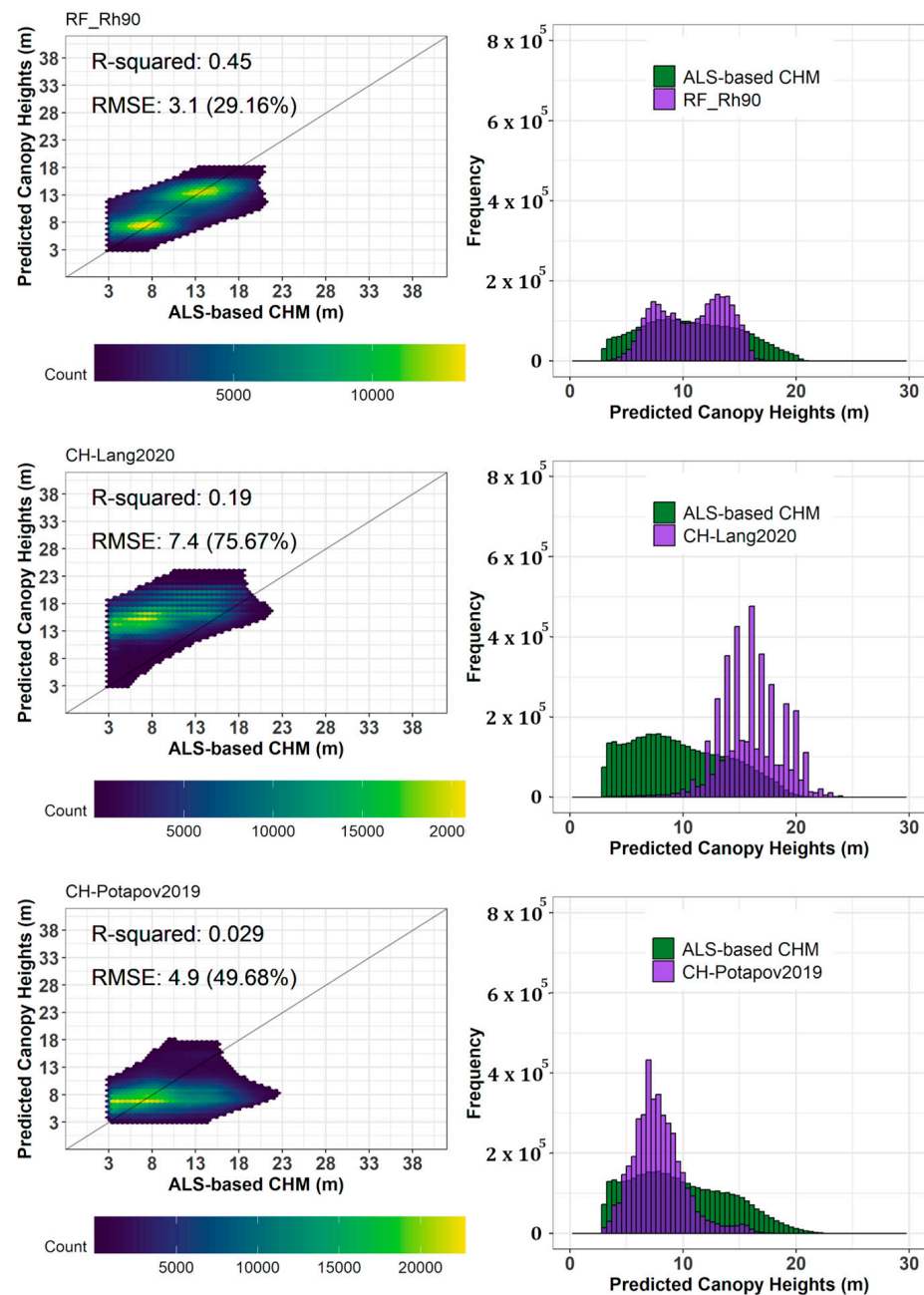


Figure 8. Scatter plot and horizontal pixel frequency distribution comparing predicted canopy heights with reference data in a coniferous forest site. Canopy heights from ALS, RF_Rh90, CH-Potapov2019, and CH-Lang2020 were utilized. Root mean square error (RMSE in m and %) and R-squared values are presented.

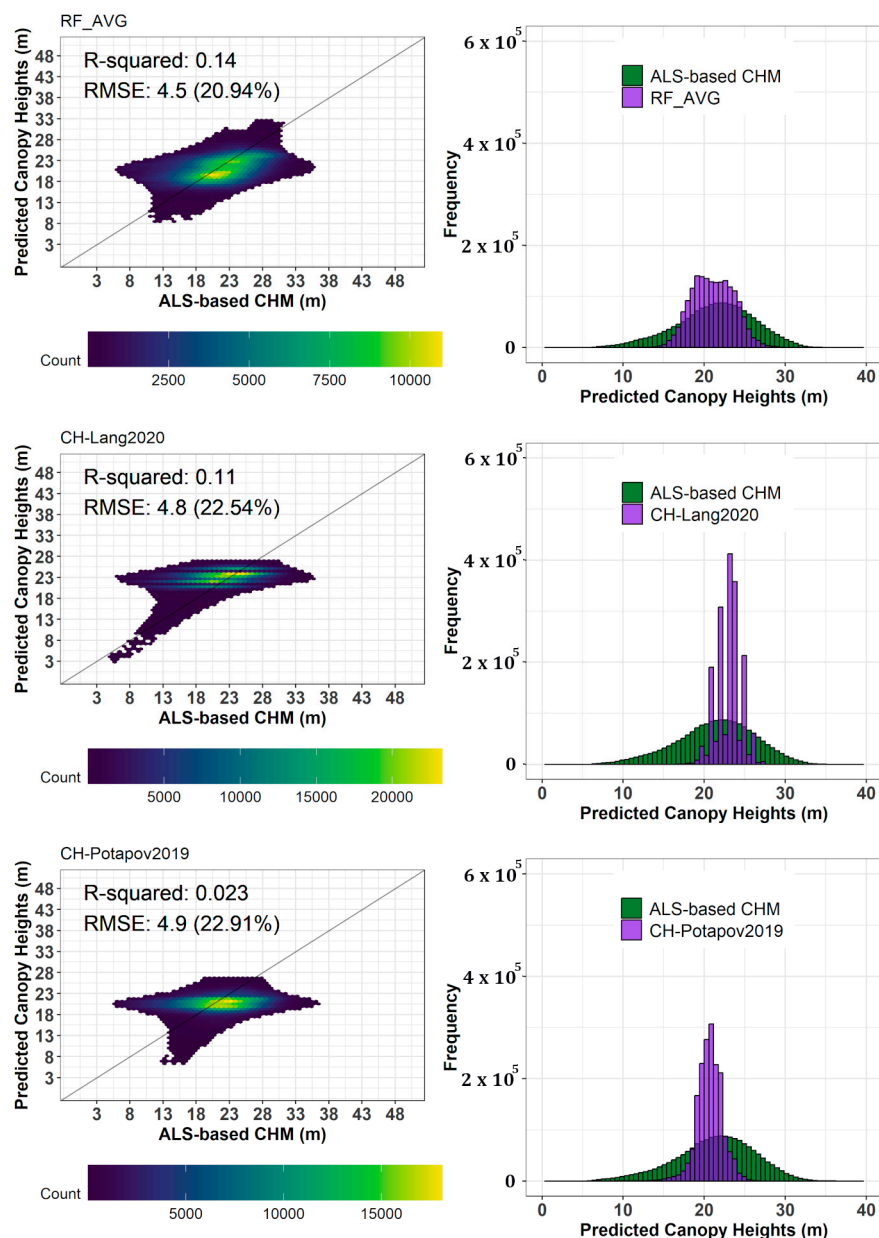


Figure 9. Scatter plot and horizontal pixel frequency distribution comparing predicted canopy heights with reference data in a broadleaf forest site. Canopy heights from ALS, RF_Rh90, CH-Potapov2019, and CH-Lang2020 were utilized. Root mean square error (RMSE in m and %) and R-squared values are presented.

When using canopy heights obtained from the broadleaf site, somewhat similar results were obtained for broadleaf trees using the CH-Lang2020 map (RMSE = 22.54%; Figure 9) and the CH-Potapov2019 map (RMSE = 22.91%; Figure 9). Similarly to the results for the coniferous site, the broadleaf site also showed low explanatory power for the existing maps, with R-squared values falling below 0.11.

6. Discussion

In this study we introduce a novel approach for generating localised high-resolution canopy height maps (10 m) using widely available satellite data and machine learning in GEE. GEDI metrics, considered to represent the top-of-canopy height (Rh90, Rh95, Rh98, Rh100, and AVG), were downscaled and spatialized for two Mediterranean forest sites, using Sentinel-1, Sentinel-2, and topographical data. Three different ML models were

tested (RF, GB, and CART) in structurally and demographically contrasting study sites, and the resulting maps were compared with ALS-based CHMs for validation. We further tested the robustness of the predicted canopy height maps by comparing their performance with that of the global GEDI-derived canopy height maps: CH-Potapov2019 [27] and CH-Lang2020 [28].

6.1. Canopy Heights Map Prediction

Our findings revealed that Rh98, Rh100, and AVG were the most accurate GEDI metrics for predicting real canopy heights derived from ALS-based CHM in both Mediterranean forest sites. However, when comparing the top-performing GEDI metrics (AVG, Rh98, and Rh100) among the ML models, both RF and GB models provided similarly accurate predictions in both Mediterranean forest types (see RMSE % in Figure 3), with only marginal improvements observed in coniferous forest sites.

In line with previous studies, the Rh95 metric was a good predictor of realistic canopy heights [32,72,73]. The AVG metric (the average of Rh75, Rh90, Rh95, and Rh100) emerged as a promising GEDI metric for predicting real canopy heights. Rh95–Rh100 and Rh90 tended to either overestimate or underestimate GEDI canopy height predictions. Accordingly, the Rh75 metric (representing 75% of returned energy between the top of the canopy and the ground surface) was crucial in regulating the overestimations of GEDI metrics higher than Rh90. Despite these considerations, several other factors can influence the effectiveness of Rh95 and AVG. These include the quality of laser pulses, forest canopy structure, topography, and the time difference between GEDI and Sentinel data acquisition. Our study aimed to mitigate the impact of these factors by standardizing the acquisition time of Copernicus data and using power laser shots [32,35].

The quality of raw GEDI data quality is a significant source of uncertainty when estimating canopy height. It can be influenced by factors such as tree density, slope, the type of laser shots used, and canopy heterogeneity [74]. This is reflected in the lower map accuracy of the broadleaf forest site, characterized by its heterogeneous structure, compared to the higher map accuracy of the homogeneous coniferous forest site. Furthermore, even the 10 m pixel resolution of the final downscaled maps is much larger than individual tree crowns, leading to a loss of canopy height data in heterogeneous forest types. Thus, predicting canopy height from GEDI data might be more suitable for forestry purposes, because forestry stands are typically homogeneous and evenly aged. Therefore, it is essential to carefully consider the forest stand structure of the test area when using GEDI data for canopy height prediction, in addition to other secondary factors.

In all cases, the RF and GB models outperformed the CART models (Figure 3). As expected, the RF algorithm performed well in both regression and classification approaches, while GB emerged as one of the most versatile algorithms for regression analysis [60,61]. The superior performance of RF and GB models can be attributed to their capacity to regulate overfitting by employing decision tree techniques [61]. In this regard, the decision tree learning approach increased the efficiency of processing large datasets, as it employs a boosting technique to incorporate random sampling with replacement across weighted data [60]. On the contrary, the reduced predictive potential of the CART models for predicting GEDI metrics can be attributed to their sensitivity to outliers and the influence of numerous observations on the decision tree analysis [75].

Notably, the RF and GB models were more computationally intensive than the CART models, which are known for their lower computational requirements [75]. Moreover, our findings aligned with or exceeded those reported in other studies (Figure 3). For instance, Lang et al. [28] reported an RMSE value equal to 6 m (RMSE = 13%) for global forests and Mediterranean forest class; Potapov et al. [27] reached an accuracy equal to 6.6 m for RMSE and 0.62 for R-squared for global forests. Matasci et al. [36] reported accurate values equal to 2.72 m for RMSE and 0.5 for R-squared for the Canadian boreal zone, and Schwartz et al. [73] indicated a map accuracy equal to 2.98 m for RMSE and 0.73 for R-squared for French coniferous forests.

6.2. Comparison of Predicted Canopy Height Maps with Reference Airborne Laser Scanning (ALS)-Based Canopy Height Model Results

The best-performing RF and GB models achieved an R-squared accuracy of 0.46 in the coniferous site and 0.14 in the broadleaf site, when compared to the ‘realistic’ canopy height derived from the ALS-based CHMs. ALS-based CHMs were selected for this investigation as a proxy for true canopy height because ALS LiDAR is highly correlated with ground truth canopy height [42].

Nevertheless, several hindering factors may have influenced the results obtained for the coniferous and broadleaf study sites. For example, Dubayah et al. [14] note that forests covered by “power beam” GEDI data types are twice as precise as “coverage beams”. This is because power beams can penetrate dense canopies more effectively, thus reducing saturation at high tree densities [14,76,77]. Consequently, if the “power beam” GEDI data are unavailable, this can lead to greater uncertainty in the GEDI Rh metrics for broadleaf stands [47]. Moreover, different species have different canopy architectures: in general, coniferous canopies grow straight and symmetrically in a conical shape, whereas broadleaf canopies frequently display plagiotropic development [78–80]. Although our models used both full power and coverage beam data, we recommend that future studies consider the beam-type data when investigating broadleaf forests. Topographical features are another secondary factor influencing GEDI Rh measurements [81]. However, we integrated topographic ancillary data from 2010 (GMTE-2010) in the predictor’s dataset. Nevertheless, recent land changes may have not been accounted for, potentially affecting the map accuracy in disturbed zones. Further research using recently acquired ALS data in broadleaved forests could help to detect occlusion in this forest type and evaluate the robustness of canopy height maps. Studies indicate that seasonal differences in leaf cover can significantly impact the consistency of canopy height measurements, particularly for trees with thick, crooked, and spreading branches [8,10,49]. Natural processes such as growing, regeneration, fragmentation, and disturbance can provoke variable canopy heights and structure over time, particularly through forest gaps [82,83], which affects the accuracy of forest maps [41,52]. In our studied forests, human-induced disturbance likely affected only coniferous forests, as broadleaf sites were managed for conservation purposes.

6.3. Comparison of Predicted Canopy Height Maps with Other Existing Canopy Height Maps

Our findings demonstrate that locally calibrated canopy height maps are a better fit to the reference ALS-based CHMs, when compared with the global canopy height maps from CH-Potapov2019 and CH-Lang2020. This is expected, since global maps predict canopy height for a wide range of forested biomes. Global maps must deal with a higher degree of variation in satellite data availability, terrain, and canopy structure, in addition to greater uncertainty in raw GEDI data [84]. Consequently, both CH-Potapov2019 and CH-Lang2020 use machine learning techniques. The existing maps use convolutional neural networks (CNNs) and bagged regression trees methods to overcome the issues associated with predicting canopy height at a global scale, which enable contextual learning comparable with the capability of the RF, GB, and CART algorithms. However, it is important to highlight that enhanced model complexity comes with greater computational and programmatic demands. For example, GEE does not offer a freely available package for deep learning with CNNs and, despite the GEE cloud computing infrastructure, model construction is not possible on a global scale. This requires users to have both the skill and hardware to construct a model outside the GEE ecosystem, when for many forest-based research questions this is neither practical nor feasible. This challenge, however, can be overcome through collaboration with qualified coding experts, potentially facilitated by International projects. Notably, we found that estimated canopy heights using the simpler RF and GB models were roughly twice as accurate as CH-Potapov2019 and CH-Lang2020 for the coniferous forest site (see Figures 8 and 9). This highlights the value of using simple but effective machine learning algorithms at smaller spatial scales. The lower

computational burdens of RF, GBs and CART also allow for the inclusion of additional meaningful predictors, i.e., Sentinel-1 and topographical predictors, compared to only Sentinel-2 as in CH-Lang2020 or Landsat in CH-Potapov2019. Our proposed approach differs from CH-Lang2020's in several aspects, including the following: (1) input data predictors (S2 vs. S1, S2, and topographical data); (2) input data target (country-level GEDI shoots vs. local-level GEDI shoots); and (3) forest mask (no-forest mask vs. global forest/non-forest mask (2017–2020) [44]), (4) modelling (deep learning vs. RF/GB or CART), and (5) validation (country-level vs. local-level).

The results reveal that the CH-lang2020 map better fits the ALS-based CHM than CH-Potapov2019, with the best R-squared achieved in the coniferous forest sites. Accordingly, the comparison of canopy height maps from CH-Potapov 2019 and CH-Lang2020 to ALS-based CHMs can be sensitive to noise, lack, and registration errors [82]. In addition to most hindering factors affecting the mapping (Section 4.2), secondary factors, such as the following, may have influenced the comparison pixel-by-pixel approach: (1) ALS point density (60 vs. 300 points/m²); (2) number of GEDI shots (acquired from late 2018 to 2019 vs. 2020); (3) mismatch between GEDI shot dates (2019 vs. 2020) and ALS acquisition date (2016 vs. 2021); (4) forest management plan (anthropic forestry interventions >50 past years vs. 2022); (5) tree species composition (mixed-species vs. unique); (6). forest structural complexity (multi-layered and mono-layered; and (7) canopy conditions. Among the previously listed factors, we believe that the number of GEDI shots considered for constructing those maps was lower than what has been collected up to now, considering that NASA's GEDI mission started collecting data in late 2018. Our forest sites data revealed a time discrepancy between GEDI (acquisitions from 2018 to 2023) and ALS (acquisitions from 2016 to 2021). This gap ranges from 2 to 7 years for broadleaf species and from 2 to 3 years for coniferous species. This difference in acquisition times could significantly impact the agreement between our developed map and CHMs due to potential changes in vegetation structure and species composition [85], particularly for broadleaf trees.

7. Conclusions

This study presents a novel and robust methodology for producing high-resolution canopy height maps at a 10m resolution by integrating GEDI relative height (Rh) metrics with Copernicus and topographical data using machine learning (ML) algorithms. Seven key conclusions emerge from our research.

Firstly, our approach combines GEDI data with Sentinel-1, Sentinel-2, and topographical variables to generate canopy height maps employing random forest (RF), gradient boosting (GB), and classification and regression trees (CART) algorithms. Secondly, RF and GB models consistently outperformed CART models in predicting canopy heights across various GEDI metrics and forest types. Thirdly, the influential predictors for canopy height prediction varied by forest type, with the B1 band from Sentinel-2 and topographical variables being more significant in broadleaf forests, while a combination of Sentinel-1, Sentinel-2, and topographical predictors played a vital role in coniferous forests. Fourthly, the choice of GEDI metric can enhance prediction accuracy, with Rh90 and AVG metrics yielding slightly better results, particularly for coniferous trees. Fifthly, the generated maps exhibited higher accuracy compared to existing ones. Sixthly, our proposed method offers advantages over existing approaches by locally calibrating GEDI footprints, ensuring better fit for predicting canopy heights, and identifying patterns resembling canopy height models (CHMs) derived from airborne laser scanning (ALS). Seventhly, our methodology holds global applicability, and a web-based application leveraging cloud computing platforms could facilitate its widespread adoption.

Additionally, while our approach was validated in two structurally contrasting Mediterranean forest types using a pixel-by-pixel approach, further comprehensive studies focused on similar forest types are warranted to ascertain its potential, particularly in complex broadleaved forests. Nonetheless, our findings hold promise for applications in forest management and environmental monitoring.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16071281/s1>, Figure S1: Variable importance analysis of machine learning models used to predict canopy height in broadleaf sites. Figure S2: Variable importance analysis of machine learning models used to predict canopy height in coniferous sites.

Author Contributions: Conceptualization, C.A., H.O., S.F. and E.B.; methodology, C.A., H.O., S.F. and E.B.; validation, C.A., H.O., S.F. and E.B.; formal analysis, C.A., H.O., S.F. and E.B.; investigation, C.A., H.O., S.F., E.B. and G.C.; data curation, C.A., H.O. and E.B.; writing—original draft preparation, C.A., H.O. and E.B.; writing—review and editing, C.A., H.O., S.F., G.C., M.M. (Marco Marchetti), G.S., B.L., G.C., M.M. (Michela Marignani) and E.B.; supervision, C.A., E.B., S.F. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Acknowledgments: This research was supported by the Financial instrument for the environment (LIFE) program in the framework of the project FRESH LIFE—Demonstrating Remote Sensing Integration in Sustainable Forest Management (LIFE14 ENV/IT/000414). Support has also been provided by the PABLO project (<https://www.gopablo.it/>, accessed on 3 April 2024), within the framework of PSR 2014-2020—MIS. 16—Cooperazione—sottomisura 16.2—Sostegno a progetti pilota e allo sviluppo di nuovi prodotti, pratiche, processi e tecnologie. H.O. was supported by the National Environment Research Council (grant number NE/P012345/1). Special thanks to the Google Earth Engine (GEE) summer school staff at the University of Florence for providing basic and advanced training in GEE coding (<https://geefirenze.wordpress.com/>, accessed on 3 April 2024).

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

References

- Bastin, J.-F.; Finegold, Y.; Garcia, C.; Mollicone, D.; Rezende, M.; Routh, D.; Zohner, C.M.; Crowther, T.W. The Global Tree Restoration Potential. *Science* **2019**, *365*, 76–79. [[CrossRef](#)] [[PubMed](#)]
- R. State of Europe Forests. Summary for Policy Makers State of Europe’s Forest 2020. In Proceedings of the Ministerial Conference on the Protection of Forests in Europe, Bratislava, Slovakia, 14–15 April 2020; Liaison Unit Bratislava: Bratislava, Slovakia, 2020; Volume 4, pp. 64–75.
- Broszofski, K.D.; Froese, R.E.; Falkowski, M.J.; Banskota, A. A Review of Methods for Mapping and Prediction of Inventory Attributes for Operational Forest Management. *For. Sci.* **2014**, *60*, 733–756. [[CrossRef](#)]
- McRoberts, R.; Tomppo, E. Remote Sensing Support for National Forest Inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [[CrossRef](#)]
- McRoberts, R.E.; Naesset, E.; Gobakken, T. Accuracy and Precision for Remote Sensing Applications of Nonlinear Model-Based Inference. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 27–34. [[CrossRef](#)]
- Vizzarri, M.; Chiavetta, U.; Santopuoli, G.; Tonti, D.; Marchetti, M. Mapping Forest Ecosystem Functions for Landscape Planning in a Mountain Natura2000 Site, Central Italy. *J. Environ. Plan. Manag.* **2015**, *58*, 1454–1478. [[CrossRef](#)]
- Proietti, R.; Antonucci, S.; Monteverdi, M.C.; Garfi, V.; Marchetti, M.; Plutino, M.; Di Carlo, M.; Germani, A.; Santopuoli, G.; Castaldi, C.; et al. Monitoring Spring Phenology in Mediterranean Beech Populations through in Situ Observation and Synthetic Aperture Radar Methods. *Remote Sens. Environ.* **2020**, *248*, 111978. [[CrossRef](#)]
- Mulverhill, C.; Coops, N.C.; Hermosilla, T.; White, J.C.; Wulder, M.A. Evaluating ICESat-2 for Monitoring, Modeling, and Update of Large Area Forest Canopy Height Products. *Remote Sens. Environ.* **2022**, *271*, 112919. [[CrossRef](#)]
- 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](#)]
- Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Queinnec, M.; Luther, J.E.; Bolton, D.K.; White, J.C.; Wulder, M.A.; Van Lier, O.R.; Hermosilla, T. Modelling Lidar-Derived Estimates of Forest Attributes over Space and Time: A Review of Approaches and Future Trends. *Remote Sens. Environ.* **2021**, *260*, 112477. [[CrossRef](#)]
- Calders, K.; Adams, J.; Armston, J.; Bartholomeus, H.; Bauwens, S.; Bentley, L.P.; Chave, J.; Danson, F.M.; Demol, M.; Disney, M.; et al. Terrestrial Laser Scanning in Forest Ecology: Expanding the Horizon. *Remote Sens. Environ.* **2020**, *251*, 112102. [[CrossRef](#)]
- Beland, M.; Parker, G.; Sparrow, B.; Harding, D.; Chasmer, L.; Phinn, S.; Antonarakis, A.; Strahler, A. On Promoting the Use of Lidar Systems in Forest Ecosystem Research. *For. Ecol. Manag.* **2019**, *450*, 117484. [[CrossRef](#)]
- Torresan, C.; Berton, A.; Carotenuto, F.; Di Gennaro, S.F.; Gioli, B.; Matese, A.; Miglietta, F.; Vagnoli, C.; Zaldei, A.; Wallace, L. Forestry Applications of UAVs in Europe: A Review. *Int. J. Remote Sens.* **2017**, *38*, 2427–2447. [[CrossRef](#)]

14. Liang, X.; Wang, Y.; Pyörälä, J.; Lehtomäki, M.; Yu, X.; Kaartinen, H.; Kukko, A.; Honkavaara, E.; Issaoui, A.E.I.; Nevalainen, O.; et al. Forest in Situ Observations Using Unmanned Aerial Vehicle as an Alternative of Terrestrial Measurements. *For. Ecosyst.* **2019**, *6*, 20. [[CrossRef](#)]
15. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [[CrossRef](#)]
16. Liang, M.; González-Roglich, M.; Roehrdanz, P.; Tabor, K.; Zvoleff, A.; Leitold, V.; Silva, J.; Fatoyinbo, T.; Hansen, M.; Duncanson, L. Assessing Protected Area's Carbon Stocks and Ecological Structure at Regional-Scale Using GEDI Lidar. *Glob. Environ. Chang.* **2023**, *78*, 102621. [[CrossRef](#)]
17. Francini, S.; D'Amico, G.; Vangi, E.; Borghi, C.; Chirici, G. Integrating GEDI and Landsat: Spaceborne Lidar and Four Decades of Optical Imagery for the Analysis of Forest Disturbances and Biomass Changes in Italy. *Sensors* **2022**, *22*, 2015. [[CrossRef](#)] [[PubMed](#)]
18. Senf, C.; Seidl, R. Mapping the Forest Disturbance Regimes of Europe. *Nat. Sustain.* **2020**, *4*, 63–70. [[CrossRef](#)]
19. Silveira, E.M.O.; Radeloff, V.C.; Martinuzzi, S.; Martinez Pastur, G.J.; Bono, J.; Politi, N.; Lizarraga, L.; Rivera, L.O.; Ciuffoli, L.; Rosas, Y.M.; et al. Nationwide Native Forest Structure Maps for Argentina Based on Forest Inventory Data, SAR Sentinel-1 and Vegetation Metrics from Sentinel-2 Imagery. *Remote Sens. Environ.* **2023**, *285*, 113391. [[CrossRef](#)]
20. Torresani, M.; Rocchini, D.; Alberti, A.; Moudrý, V.; Heym, M.; Thouverai, E.; Kacic, P.; Tomelleri, E. LiDAR GEDI Derived Tree Canopy Height Heterogeneity Reveals Patterns of Biodiversity in Forest Ecosystems. *Ecol. Inf.* **2023**, *76*, 102082. [[CrossRef](#)]
21. Vangi, E.; D'Amico, G.; Francini, S.; Giannetti, F.; Lasserre, B.; Marchetti, M.; McRoberts, R.E.; Chirici, G. The Effect of Forest Mask Quality in the Wall-to-Wall Estimation of Growing Stock Volume. *Remote Sens.* **2021**, *13*, 1038. [[CrossRef](#)]
22. Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S.; et al. Aboveground Biomass Density Models for NASA's Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sens. Environ.* **2022**, *270*, 112845. [[CrossRef](#)]
23. Francini, S.; McRoberts, R.E.; D'Amico, G.; Coops, N.C.; Hermosilla, T.; White, J.C.; Wulder, M.A.; Marchetti, M.; Mugnozza, G.S.; Chirici, G. An Open Science and Open Data Approach for the Statistically Robust Estimation of Forest Disturbance Areas. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102663. [[CrossRef](#)]
24. Tang, H.; Stoker, J.; Luthcke, S.; Armston, J.; Lee, K.; Blair, B.; Hofton, M. Evaluating and Mitigating the Impact of Systematic Geolocation Error on Canopy Height Measurement Performance of GEDI. *Remote Sens. Environ.* **2023**, *291*, 113571. [[CrossRef](#)]
25. Shendryk, Y. Fusing GEDI with Earth Observation Data for Large Area Aboveground Biomass Mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *115*, 103108. [[CrossRef](#)]
26. Di Tommaso, S.; Wang, S.; Lobell, D.B. Combining GEDI and Sentinel-2 for Wall-to-Wall Mapping of Tall and Short Crops. *Environ. Res. Lett.* **2021**, *16*, 125002. [[CrossRef](#)]
27. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [[CrossRef](#)]
28. Lang, N.; Jetz, W.; Schindler, K.; Wegner, J.D. A High-Resolution Canopy Height Model of the Earth. *Nat. Ecol. Evol.* **2023**, *7*, 1778–1789. [[CrossRef](#)] [[PubMed](#)]
29. Li, W.; Niu, Z.; Shang, R.; Qin, Y.; Wang, L.; Chen, H. High-Resolution Mapping of Forest Canopy Height Using Machine Learning by Coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102163. [[CrossRef](#)]
30. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine Learning in Geosciences and Remote Sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]
31. Potapov, P.; Tyukavina, A.; Turubanova, S.; Talero, Y.; Hernandez-Serna, A.; Hansen, M.C.; Saah, D.; Tenneson, K.; Poortinga, A.; Aekakkarakunroj, A.; et al. Annual Continuous Fields of Woody Vegetation Structure in the Lower Mekong Region from 2000-2017 Landsat Time-Series. *Remote Sens. Environ.* **2019**, *232*, 111278. [[CrossRef](#)]
32. Wang, C.; Elmore, A.J.; Numata, I.; Cochrane, M.A.; Lei, S.; Hakkenberg, C.R.; Li, Y.; Zhao, Y.; Tian, Y. A Framework for Improving Wall-to-Wall Canopy Height Mapping by Integrating GEDI LiDAR. *Remote Sens.* **2022**, *14*, 3618. [[CrossRef](#)]
33. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
34. Gomes, V.; Queiroz, G.; Ferreira, K. An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sens.* **2020**, *12*, 1253. [[CrossRef](#)]
35. Mandl, L.; Stritih, A.; Seidl, R.; Ginzler, C.; Senf, C. Spaceborne LIDAR for Characterizing Forest Structure across Scales in the European Alps. *Remote Sens. Ecol. Conserv.* **2023**, *9*, rse2.330. [[CrossRef](#)]
36. Matasci, G.; Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W.; Zald, H.S.J. Large-Area Mapping of Canadian Boreal Forest Cover, Height, Biomass and Other Structural Attributes Using Landsat Composites and Lidar Plots. *Remote Sens. Environ.* **2018**, *209*, 90–106. [[CrossRef](#)]
37. Morin, D.; Planells, M.; Baghdadi, N.; Bouvet, A.; Fayad, I.; Le Toan, T.; Mermoz, S.; Villard, L. Improving Heterogeneous Forest Height Maps by Integrating GEDI-Based Forest Height Information in a Multi-Sensor Mapping Process. *Remote Sens.* **2022**, *14*, 2079. [[CrossRef](#)]

38. Giannetti, F.; Barbati, A.; Mancini, L.D.; Travaglini, D.; Bastrup-Birk, A.; Canullo, R.; Nocentini, S.; Chirici, G. European Forest Types: Toward an Automated Classification. *Ann. For. Sci.* **2018**, *75*, 6. [CrossRef]
39. Barbati, A.; Marchetti, M.; Chirici, G.; Corona, P. European Forest Types and Forest Europe SFM Indicators: Tools for Monitoring Progress on Forest Biodiversity Conservation. *For. Ecol. Manag.* **2014**, *321*, 145–157. [CrossRef]
40. Santopuoli, G.; Di Cristofaro, M.; Kraus, D.; Schuck, A.; Lasserre, B.; Marchetti, M. Biodiversity Conservation and Wood Production in a Natura 2000 Mediterranean Forest. A Trade-off Evaluation Focused on the Occurrence of Microhabitats. *iForest* **2019**, *12*, 76–84. [CrossRef]
41. Marchetti, M.; Vizzarri, M.; Sallustio, L.; Di Cristofaro, M.; Lasserre, B.; Lombardi, F.; Giancola, C.; Perone, A.; Simpatico, A.; Santopuoli, G. Behind Forest Cover Changes: Is Natural Regrowth Supporting Landscape Restoration? Findings from Central Italy. *Plant Biosyst. -Int. J. Deal. Asp. Plant Biosyst.* **2018**, *152*, 524–535. [CrossRef]
42. Roussel, J.-R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.H.; Meador, A.S.; Bourdon, J.-F.; de Boissieu, F.; Achim, A. lidR: An R Package for Analysis of Airborne Laser Scanning (ALS) Data. *Remote Sens. Environ.* **2020**, *251*, 112061. [CrossRef]
43. Roussel, J.-R.; Isenburg, M.; Auty, D.; Marie, P.; de Conto, T. Read and Write “las” and “Laz” Binary File Formats Used for Remote Sensing Data. 2023. Available online: <https://cran.r-project.org/> (accessed on 3 April 2024).
44. Alvites, C.; Santopuoli, G.; Maesano, M.; Chirici, G.; Moresi, F.V.; Tognetti, R.; Marchetti, M.; Lasserre, B. Unsupervised Algorithms to Detect Single Trees in a Mixed-Species and Multilayered Mediterranean Forest Using LiDAR Data. *Can. J. For. Res.* **2021**, *51*, 1766–1780. [CrossRef]
45. Vangi, E.; D’Amico, G.; Francini, S.; Chirici, G. GEDI4R: An R Package for NASA’s GEDI Level 4 A Data Downloading, Processing and Visualization. *Earth Sci. Inform.* **2023**, *16*, 1109–1117. [CrossRef]
46. Kellner, J.R.; Armston, J.; Duncanson, L. Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density. *Earth Space Sci.* **2023**, *10*, e2022EA002516. [CrossRef]
47. Rishmawi, K.; Huang, C.; Zhan, X. Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data. *Remote Sens.* **2021**, *13*, 442. [CrossRef]
48. Puletti, N.; Grotti, M.; Ferrara, C.; Chianucci, F. Lidar-Based Estimates of Aboveground Biomass through Ground, Aerial, and Satellite Observation: A Case Study in a Mediterranean Forest. *J. Appl. Remote Sens.* **2020**, *14*, 044501. [CrossRef]
49. White, J.C.; Arnett, J.T.T.R.; Wulder, M.A.; Tompalski, P.; Coops, N.C. Evaluating the Impact of Leaf-on and Leaf-off Airborne Laser Scanning Data on the Estimation of Forest Inventory Attributes with the Area-Based Approach. *Can. J. For. Res.* **2015**, *45*, 1498–1513. [CrossRef]
50. Dostálová, A.; Lang, M.; Ivanovs, J.; Waser, L.T.; Wagner, W. European Wide Forest Classification Based on Sentinel-1 Data. *Remote Sens.* **2021**, *13*, 337. [CrossRef]
51. Mullissa, A.; Vollrath, A.; Odongo-Braun, C.; Slagter, B.; Balling, J.; Gou, Y.; Gorelick, N.; Reiche, J. Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine. *Remote Sens.* **2021**, *13*, 1954. [CrossRef]
52. Parisi, F.; Vangi, E.; Francini, S.; D’Amico, G.; Chirici, G.; Marchetti, M.; Lombardi, F.; Travaglini, D.; Ravera, S.; De Santis, E.; et al. Sentinel-2 Time Series Analysis for Monitoring Multi-Taxon Biodiversity in Mountain Beech Forests. *Front. For. Glob. Chang.* **2023**, *6*, 1020477. [CrossRef]
53. Cavalli, A.; Francini, S.; Cecili, G.; Coccozza, C.; Congedo, L.; Falanga, V.; Spadoni, G.; Maesano, M.; Munafò, M.; Chirici, G.; et al. Afforestation Monitoring through Automatic Analysis of 36-Years Landsat Best Available Composites. *iForest* **2022**, *15*, 220–228. [CrossRef]
54. Francini, S.; Hermosilla, T.; Coops, N.C.; Wulder, M.A.; White, J.C.; Chirici, G. An Assessment Approach for Pixel-Based Image Composites. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 1–12. [CrossRef]
55. Lefsky, M.A.; Cohen, W.B.; Harding, D.J.; Parker, G.G.; Acker, S.A.; Gower, S.T. Lidar Remote Sensing of Above-Ground Biomass in Three Biomes: Biomass Estimation by LIDAR. *Glob. Ecol. Biogeogr.* **2002**, *11*, 393–399. [CrossRef]
56. Lefsky, M.A.; Harding, D.J.; Keller, M.; Cohen, W.B.; Carabajal, C.C.; Del Bom Espirito-Santo, F.; Hunter, M.O.; de Oliveira, R. Estimates of Forest Canopy Height and Aboveground Biomass Using ICESat: American Geophysical Union, Washington, USA. *Geophys. Res. Lett.* **2005**, *32*, L22502. [CrossRef]
57. Duncanson, L.I.; Niemann, K.O.; Wulder, M.A. Estimating Forest Canopy Height and Terrain Relief from GLAS Waveform Metrics. *Remote Sens. Environ.* **2010**, *114*, 138–154. [CrossRef]
58. Danielson, J.J.; Gesch, D.B. *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073*; Open-File Report; USGS: Reston, VA, USA, 2011; p. 26.
59. Morales, C.; Díaz, A.S.-P.; Dionisio, D.; Guarnieri, L.; Marchi, G.; Maniatis, D.; Mollicone, D. Earth Map: A Novel Tool for Fast Performance of Advanced Land Monitoring and Climate Assessment. *J. Remote Sens.* **2023**, *3*, 3. [CrossRef]
60. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
61. Kotsiantis, S.B. Decision Trees: A Recent Overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [CrossRef]
62. Bozzini, A.; Francini, S.; Chirici, G.; Battisti, A.; Faccoli, M. Spruce Bark Beetle Outbreak Prediction through Automatic Classification of Sentinel-2 Imagery. *Forests* **2023**, *14*, 1116. [CrossRef]
63. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for Land Cover Classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]

64. Chirici, G.; Scotti, R.; Montagni, A.; Barbati, A.; Cartisano, R.; Lopez, G.; Marchetti, M.; McRoberts, R.E.; Olsson, H.; Corona, P. Stochastic Gradient Boosting Classification Trees for Forest Fuel Types Mapping through Airborne Laser Scanning and IRS LISS-III Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *25*, 87–97. [[CrossRef](#)]
65. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
66. Elith, J.; Leathwick, J.R.; Hastie, T. A Working Guide to Boosted Regression Trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
67. Fattorini, L.; Marcheselli, M.; Pisani, C.; Pratelli, L. Design-based Properties of the Nearest Neighbor Spatial Interpolator and Its Bootstrap Mean Squared Error Estimator. *Biometrics* **2022**, *78*, 1454–1463. [[CrossRef](#)] [[PubMed](#)]
68. Francini, S.; Coccozza, C.; Hölttä, T.; Lintunen, A.; Paljakka, T.; Chirici, G.; Traversi, M.L.; Giovannelli, A. A Temporal Segmentation Approach for Dendrometers Signal-to-Noise Discrimination. *Comput. Electron. Agric.* **2023**, *210*, 107925. [[CrossRef](#)]
69. Cook, R.D. Detection of Influential Observation in Linear Regression. *Technometrics* **1977**, *19*, 15–18.
70. Immitzer, M.; Stepper, C.; Böck, S.; Straub, C.; Atzberger, C. Use of WorldView-2 Stereo Imagery and National Forest Inventory Data for Wall-to-Wall Mapping of Growing Stock. *For. Ecol. Manag.* **2016**, *359*, 232–246. [[CrossRef](#)]
71. John, F.; Weisberg, S. *An R Companion to Applied Regression*; Sage Publications: New York, NY, USA, 2019.
72. Kacic, P.; Hirner, A.; Da Ponte, E. Fusing Sentinel-1 and -2 to Model GEDI-Derived Vegetation Structure Characteristics in GEE for the Paraguayan Chaco. *Remote Sens.* **2021**, *13*, 5105. [[CrossRef](#)]
73. Schwartz, M.; Ciaia, P.; Otlé, C.; De Truchis, A.; Vega, C.; Fayad, I.; Brandt, M.; Fensholt, R.; Baghdadi, N.; Morneau, F.; et al. High-Resolution Canopy Height Map in the Landes Forest (France) Based on GEDI, Sentinel-1, and Sentinel-2 Data with a Deep Learning Approach. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *128*, 103711. [[CrossRef](#)]
74. Adam, M.; Urbazaev, M.; Dubois, C.; Schmullius, C. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* **2020**, *12*, 3948. [[CrossRef](#)]
75. Loh, W. Classification and Regression Trees. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
76. Adrah, E.; Wan Mohd Jaafar, W.S.; Omar, H.; Bajaj, S.; Leite, R.V.; Mazlan, S.M.; Silva, C.A.; Chel Gee Ooi, M.; Mohd Said, M.N.; Abdul Maulud, K.N.; et al. Analyzing Canopy Height Patterns and Environmental Landscape Drivers in Tropical Forests Using NASA’s GEDI Spaceborne LiDAR. *Remote Sens.* **2022**, *14*, 3172. [[CrossRef](#)]
77. Lahssini, K.; Baghdadi, N.; Le Maire, G.; Fayad, I. Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests. *Remote Sens.* **2022**, *14*, 6264. [[CrossRef](#)]
78. Rozenbergar, D.; Diaci, J. Architecture of *Fagus Sylvatica* Regeneration Improves over Time in Mixed Old-Growth and Managed Forests. *For. Ecol. Manag.* **2014**, *318*, 334–340. [[CrossRef](#)]
79. Ishii, H.; Asano, S. The Role of Crown Architecture, Leaf Phenology and Photosynthetic Activity in Promoting Complementary Use of Light among Coexisting Species in Temperate Forests. *Ecol. Res.* **2010**, *25*, 715–722. [[CrossRef](#)]
80. Parent, J.R.; Volin, J.C. Assessing the Potential for Leaf-off LiDAR Data to Model Canopy Closure in Temperate Deciduous Forests. *ISPRS J. Photogramm. Remote Sens.* **2014**, *95*, 134–145. [[CrossRef](#)]
81. Spracklen, B.; Spracklen, D.V. Determination of Structural Characteristics of Old-Growth Forest in Ukraine Using Spaceborne LiDAR. *Remote Sens.* **2021**, *13*, 1233. [[CrossRef](#)]
82. Bazzato, E.; Lallai, E.; Caria, M.; Schifani, E.; Cillo, D.; Ancona, C.; Pantini, P.; Maccherini, S.; Bacaro, G.; Marignani, M. Focusing on the Role of Abiotic and Biotic Drivers on Cross-Taxon Congruence. *Ecol. Indic.* **2023**, *151*, 110323. [[CrossRef](#)]
83. Bazzato, E.; Lallai, E.; Serra, E.; Melis, M.T.; Marignani, M. Key Role of Small Woodlots Outside Forest in a Mediterranean Fragmented Landscape. *For. Ecol. Manag.* **2021**, *496*, 119389. [[CrossRef](#)]
84. Mishra, S.; Shrivastava, P.; Dhurvey, P. Change Detection Techniques in Remote Sensing: A Review. *IJWMCIS* **2017**, *4*, 1–8. [[CrossRef](#)]
85. Bazzato, E.; Lallai, E.; Caria, M.; Schifani, E.; Cillo, D.; Ancona, C.; Alamanni, F.; Pantini, P.; Maccherini, S.; Bacaro, G.; et al. Land-Use Intensification Reduces Multi-Taxa Diversity Patterns of Small Woodlots Outside Forests in a Mediterranean Area. *Agric. Ecosyst. Environ.* **2022**, *340*, 108149. [[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.