Exploring the potential of spaceborne LiDAR data for forest biomass estimation in a complex-structured coniferous forest

Nikos Georgopoulos[a]*, Konstantinos Antoniadis[a], Dimitris Stavrakoudis[a], Ioannis Z. Gitas[a]
[a]Laboratory of Forest Management and Remote Sensing, Department of Forestry and Natural Environment, Aristotle University of Thessaloniki, P.O. Box 248, 54124 Thessaloniki, Greece

ABSTRACT

The development of adaptive forest ecosystem management strategies presumes the existence of aboveground biomass estimates at different spatial scales. In recent years, a series of remote sensing technologies have been developed, providing timely, low-cost, and reliable estimates on biomass and other forest attributes, especially over large and/or remote areas. Among the various technologies, Light Detection and Ranging (LiDAR) employed on aerial, satellite, and even ground platforms typically offer accurate forest biomass estimates. This is due to the ability of the emitted pulses to penetrate the forest canopy, thereby providing 3D information about all forest vegetation layers. While ground-based and aerial LiDAR provide a more detailed characterization of the vertical structure of vegetation compared to satellite-based LiDAR, the latter offers obvious advantages in terms of cost-effectiveness. This study aimed to investigate the potential of LiDAR GEDI (Global Environmental Dynamics Investigation) satellite data to reliably estimate the biomass of stem, bark, branches, and needles at the GEDI footprint level, which are the constituent components of forest aboveground biomass. The study area consists of a dense fir forest characterized by complex structure and intense topography. A regression analysis with the random forest (RF) algorithm was performed to develop prediction models for all parts of the aboveground forest biomass using height and crown profile metrics. Accordingly, five random forest models were created and evaluated for their predictive performance using an independent validation sample. In addition, the influence of forest biomass density and topography on the estimates was examined. The results reveal that the biomass of crown parts can be estimated with satisfactory accuracy and an increase of almost 15% in accuracy was observed in areas with gentle slopes and topography. Overall, GEDI data can be used to estimate forest biomass and its key parts in large extents with mild topography, regardless of the aboveground biomass density.

Keywords: Forest biomass, LiDAR, GEDI, regression analysis

1. INTRODUCTION

Forest biomass assessment is a key tool for carbon stock accounting and mitigating the impacts of climate change on ecosystems [1]. It has been estimated that forests globally absorb twice as much carbon as is released by deforestation and other disturbances [2]. Therefore, longitudinal monitoring and inventory of global biomass and carbon stocks is a prerequisite for understanding the global carbon cycle.

Forest biomass can be distinguished into aboveground (trunk, branches, leaves, bark) and belowground (root system) biomass, with aboveground (AGB) biomass being the primary focus of the research community, mainly due to the difficulty of collecting and sampling belowground biomass [3]. Biomass estimation methods can be divided into direct and indirect methods, with destructive sampling (direct) having the highest accuracy [4]. However, these procedures are quite time-consuming, and demanding, and have very high implementation cost [5]. The most widespread method of biomass estimation is the use of allometric equations, which are based on destructive sampling of a statistically significant number of trees and recording their basic structural parameters, such as breast height diameter (DBH) and tree height [6], [7].

Remote sensing technologies can provide accurate forest biomass estimates at large scales, based primarily on active and passive sensors [8]. The use of optical data derived from passive sensors is the most common approach to forest biomass estimation, although high saturation is observed in regions with high biomass density [9], [10]. Synthetic aperture radar (SAR) data have been exploited in several applications involving forest biomass [11], [12], [13], [14], [15] also present significant limitations, including saturation, high acquisition costs, and complex processing methods [16]. The use of LiDAR data has been proven to be the most accurate approach for forest biomass estimation, providing 3D visualizations

Tenth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2024), edited by
A. Christofe, S. Michaelides, D. Hadjimitsis, C. Danezis, K. Themistocleous, N. Kyriakides, G. Schreier,
Proc. of SPIE Vol. 13212, 132120U · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3037268

of forest structure and biophysical characteristics [17]. LiDAR data, depending on the acquisition platform, can be divided into aerial, ground-based and satellite-based.

Although ground-based and aerial LiDAR provide a better characterization of the vertical structure of vegetation than satellite-based ones, the latter has obvious advantages in terms of cost-effectiveness. Several recent studies have found that satellite LiDAR data can offer reliable estimates of forest parameters such as height [18], [19], forest fuels [20], and biomass [21], [22], [23], even on a global scale. Historically, the GLAS (Geoscience Laser Altimeter System) mission was the first attempt to record space-based laser observations of the Earth's surface. Healey et al. (2012) used the measurements from this sensor to estimate AGB in the California region, comparing the results of the derived model to those of the forest inventory. While the estimation errors were quite high, their cost was overwhelmingly lower than that of the forest inventory measurements, highlighting the importance of space-based data in capturing AGB at large scales [24]. In 2018, the launch of the GEDI (Global Environmental Dynamics Investigation) sensor marked the beginning of a new era in biomass observation at the global level, as it promises the possibility of estimating AGB and crown height at large scales [25].

The GEDI sensor is a full-waveform LiDAR system installed on the International Space Station, which provides billions of observations of the vertical vegetation structure in temperate and tropical ecosystems [26]. The generated data were used to extract information such as canopy height, leaf area index (LAI), soil topography and AGB. In addition, the sensor's observations are used to create predictive ecosystem models by fusing other types of remote sensing data [26]. In recent years, a plethora of studies have examined the feasibility of estimating forest biomass using GEDI data [21], [27], [28], [29], [30], [31], [32]. More specifically, Chen et al. (2022) examined the feasibility of combining GEDI and ICEsat-2 data with ALOS-2, Sentinel-1, Sentinel-2, and ALOS-1 data to estimate forest biomass in a heterogeneous mountainous forest. The feasibility of estimating AGB and crown height in a typical Mediterranean forest with GEDI data was examined by Dorado-Roda et al. (2021), highlighting the influence of forest structure on biomass estimates. Ni et al. (2021) focused on the importance of slope-adapted metrics for forest biomass estimation in areas with high relief, as the sensor's signal is significantly affected. Also, Duncanson et al. (2021) utilized the GEDI relative height metrics for biomass estimates in a conifer forest, which were compared with those of the ICEsat-2 sensor. Finally, Kellner et al. (2023) developed the product GEDI04_A, which is the AGB density at a global scale. This product was based on the GEDI02_A data and a series of linear models to estimate forest biomass density at the global scale. However, all these estimates have low accuracy in estimating crown height and biomass, as they are significantly affected by land cover and pulse sensitivity [34].

Most of the aforementioned studies are based on combining data from different sensors to evaluate the accuracy of the estimates provided by GEDI data [32, p. 202]. To the best of our knowledge, no other study has tried to estimate the constituent parts of AGB, such as biomass of stem, branches (live and dead), leaves, and bark, which are relevant to adapting the forest management practices and for deriving economic metrics. To address this gap, this paper aims to investigate the potential of GEDI data to reliably estimate the AGB of forest components (live and dead branches, stem, bark, and needles) in a complex-structured coniferous forest with intense topography. Reference data were obtained from the results of a previous study, which combined aerial LiDAR data with field measurements to estimate the components' biomass at the sampling plot level. The GEDI data are exploited to estimate the biomass of the components using a Random Forest (RF) regressor. Finally, the influence of biomass density and topography on the accuracy of forest biomass estimates is examined.

2. MATERIALS AND METHODS

2.1 Study area

This research was undertaken in the University Forest of Pertouli (Figure *1*), which is located in the eastern side of the mountain range Pindus, in the Prefecture of Trikala, Thessaly (latitude 39° 32' N – 39° 35' N, longitude 21° 33' E – 21° 38' E). The forest is owned by the Greek State that has granted the land to the Aristotle University of Thessaloniki, which has been managing the forest since 1934, aiming to support research and educational purposes. It covers almost 3,300 ha and the altitude ranges from 900 to 2050 m, with an average of 1350 m. Most of the total area (73%) is covered by forested or partially forested areas, 20% by grassland, and the remaining percentage by barren and other areas. The climate is characterized as transitional, combining elements of the Mediterranean and Mid-European climate. The topography of the area plays a decisive role in shaping the climate with an uneven distribution of precipitation throughout the year.

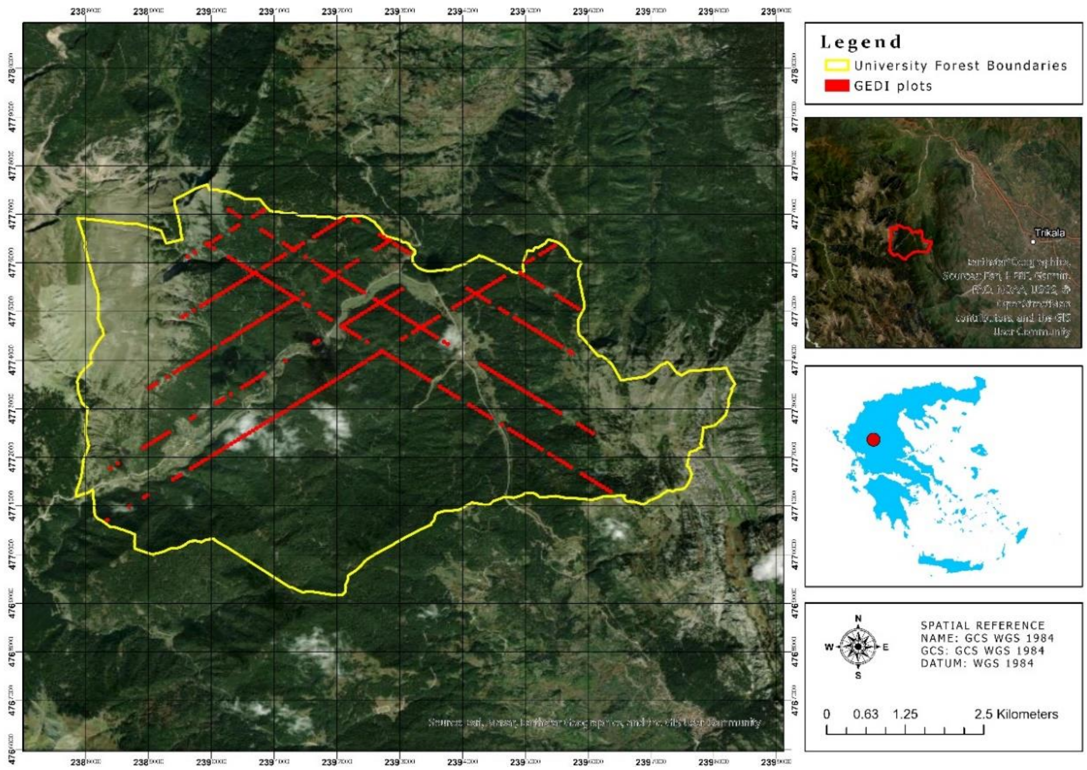

Figure 1. Map of the study area, including the GEDI plots that were valid for biomass estimation.

2.2 Airborne LiDAR data

Airborne LiDAR data were collected in October 2018 at the University Forest of Pertouli, using a RIEGL VQ-1560i-DW multispectral LiDAR sensor. The data comprised information derived from two spectral channels, Green and Near Infrared (NIR) (532 nm and 1064 nm respectively). Prior to data delivery, the vendor (GEOSYSTEMS HELLAS S.A.) performed pre-processing of the LiDAR data, including the conversion of all full waveform information (19 pulses/m2) into discrete returns. The point cloud density for both channels was nearly identical (44.65 points/m2 for Near Infrared and 44.85 points/m2 for Green), with a nominal inter-pulse spacing of 0.15 m and 0.18 m respectively, with up to seven returns per pulse and a scan angle range of −32° to +32°.

2.3 Spaceborne LiDAR data

The data of the GEDI satellite system were used for the basic AGB components estimation in the Pertouli University Forest. The data were acquired through the rGEDI library [35], using the outer boundaries of the study area to search all available data from 2019-2020. A total of 19 strips of the GEDI02_A and GEDI02_B products were acquired. The GEDI02_A products are georeferenced height metrics derived from the analysis of the full waveform data. The process for generating the GEDI02_A product datasets is adapted from the Land, Vegetation, and Ice Sensor (LVIS) algorithm. The specific product is provided in HDF5 format and has a spatial resolution of 25 m. These data provide 156 layers for each bundle, including ground elevation, canopy height, energy, and several other interpreted metrics from the waveform data. The GEDI02_B data provide metrics such as land cover, plant area index (PAI), plant area volume density (PAVD), and foliage height diversity (FHD). Finally, the erroneous measurements of the sensor (e.g. height greater than 1 km, Quality Flag) were removed and the dataset was cropped to the area of interest for the estimation of fir biomass components in the Pertouli University Forest.

2.4 Methods

Figure 2 presents the methodology developed for estimating AGB forest biomass components using the GEDI satellite data. The following subsections detail the steps for processing and analyzing LIDAR data (GEDI and aerial) to estimate the basic forest biomass components in a forest with complex structure and topography.

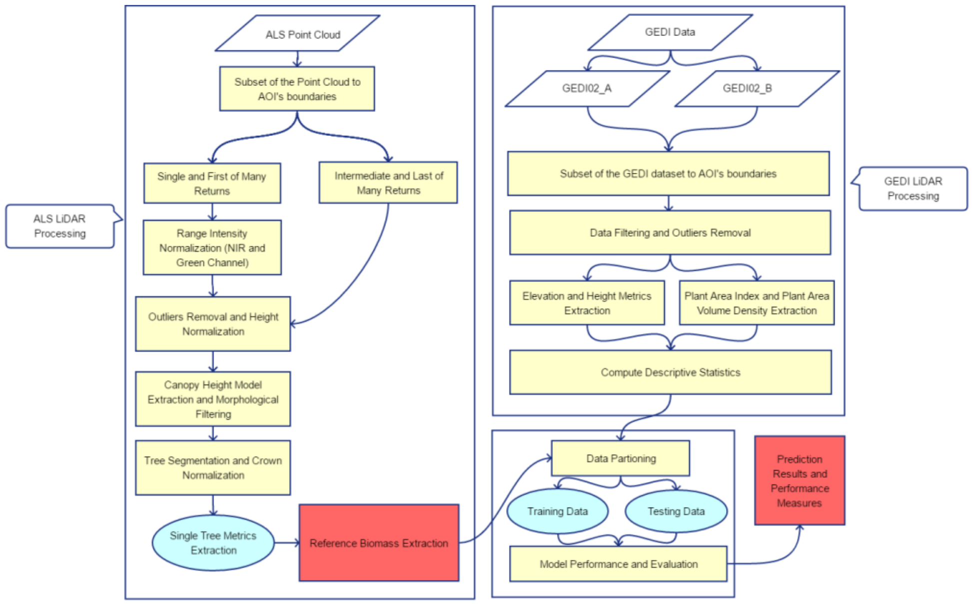

Figure 2. Flowchart of the workflow for the AGB components estimation using spaceborne laser data (GEDI) in a multi-layered coniferous forest with complex topography.

2.4.1 Spaceborne LiDAR data

The GEDI instrument has three lasers, two of which are full power and the other is split into two coverage beams, producing a total of four power beams and four coverage paths. The tracks have an average diameter of approximately 25 m and are spaced approximately 60 m apart along the track.

In this paper, GEDI02_A and GEDI02_B data, obtained for the period 2019-2020, were exploited. The obtained data were checked for validity and pulses that according to the Quality_Flag and Degrade_Flag fields were considered to be of lower quality were removed [36]. In addition, pulses that exhibited extreme height values were removed (height greater than 40 meters), as these values would significantly affect the estimates. Finally, only data taken during the evening hours was considered; however, the amount of data was drastically reduced, compromising the reliability of the results. Consequently, out of a total of 1,900 pulses covering the Pertouli University Forest, only 342 pulses were used, as only these pulses met the above criteria.

The GEDI02_A data provide relative height (RH) measurements, from 0 to 100% of the height in 1% increments. Although several studies examining forest biomass estimation use only selected height metrics [37], [38], in this study, the full set of height metrics was employed to estimate aboveground biomass. The GEDI02_B data include metrics related to crown profile such as plant surface area index (PAI) and vertical plant surface area volume density (PAVD) profile (Table 1).

Table 1. Summary of the GEDI-derived metrics used for the biomass components estimation. The metrics were extracted from two different products, GEDI02_A and GEDI02_B.

Product	Predictors	Description
GEDI02_A	rh0,rh1,…..rh100	Relative height

	height_bin	Height of the first beam pgap_theta_z relative to the ground
GEDI02_B	pavd_z0_5m,……. pavd_z140_145m	Vertical Plant Area Volume Density profile with a vertical step size of dZ
	pai_z0_5m,……. pai_z145m	Vertical PAI profile from canopy height (z) to ground (z=0) with a vertical step size of dZ, where cover(z > z_max) = 0

2.4.2 Airborne LiDAR data

In this work, the aerial LiDAR data and biomass estimation models were produced by Georgopoulos et al. (2021) and Georgopoulos et al. (2023). The analysis was performed using the R software, and more specifically the lidR package [41], [42].

Initially, the footprints of the GEDI dataset were used to clip the point cloud to produce the height and intensity metrics of the airborne LiDAR data. The point cloud, split into the green and NIR channels, was cropped to the regions of interest and used for further processing. Since the generated biomass estimation models utilize height and intensity metrics, it is considered necessary to normalize the intensity of the point cloud. Therefore, the airborne LiDAR point clouds were separated into single and first returns to normalize the intensity, as intermediate returns were excluded from the process [43]. After intensity normalization, point cloud reconstruction was performed to derive the height and intensity metrics for estimating the biomass of the AGB components at the tree level.

The normalized point cloud was further processed to remove erroneous measurements and classified into ground and non-ground points using the Cloth Simulation algorithm [44]. Next, the point cloud height was normalized and the crown height model was generated for the tree segmentation. In this study, the Watershed algorithm was used for the crown delineation process as it provides accurate results of tree crown partitioning. Then, the essential metrics were calculated for each segmented tree in order to generate the reference AGB (Table 2).

Table 2. Summary of the airborne LiDAR-derived metrics extracted for the tree components biomass estimations, in order to create the reference dataset.

LiDAR metrics	Description
p10, p20…. p95, p99	Height percentiles
b10, b20… b95, b99	Height bincentiles
max	Max height
avg	Average height
std	Height standard deviation
int_p10, int_p20…. int_p99	Intensity percentiles
int_std	Intensity standard deviation
int_avg	Average intensity
int_skew	Intensity skewness

The estimated biomass for each of the forest biomass components (branches, dead branches, stem, bark and needles) was first estimated at the tree level and then aggregated at the GEDI footprint level, which represents the GEDI footprint radius (Table 3). Finally, for each footprint, the mean slope was calculated in order to investigate the effect of topography on forest biomass estimation using GEDI data.

Table 3. Reference biomass metrics for each AGB component using the Airborne LiDAR data.

Biomass (n=342)	Mean (kg)	Minimum (kg)	Maximum (kg)
Stem	16839.84	5610.76	40899.52
Dead Branches	219.69	43.11	538.86
Branches	1976.80	245.17	5673.46
Needles	1187.56	147.07	3557.94
Bark	528.34	170.564	1340.73

2.4.3 Random forest regression algorithm and accuracy assessment

The random forest (RF) algorithm was used to estimate forest biomass components using GEDI data. The biomass parts estimated were the stem, bark, dead branches, branches, and needles. The RF algorithm is a widely used machine-learning technique based on decision trees to solve classification and regression problems [45], [46]. This algorithm was chosen as several studies have highlighted its ability to provide reliable forest biomass estimates in different types of ecosystems [31]. To obtain unbiased estimates of the models' accuracy, a repeated runs cross-validation approach was followed for each model trained (one for each tree component). More specifically, ten independent repetitions were performed, each time randomly splitting the data into training (70%) and test (30%) data. In each of these ten repetitions, an internal 3-fold cross-validation is performed on the respective training set to determine the model's hyperparameters [47]. The model hyperparameters that were optimized were the number of trees to be grown (50 to 500 with a step of 50) and the number of independent variables per tree. The performance of each model is ultimately evaluated based on mean absolute error (MAE), bias, coefficient of determination (R^2), and relative root mean square error (RMSE) on the testing sets, averaged over the ten repetitions.

3. RESULTS

3.1 Biomass estimation using spaceborne LiDAR data

The first experiment considers all available GEDI footprints, without differentiating them over any topographic or other feature. The performance of the models for estimating AGB fractions using GEDI satellite data is presented in Table 4. This table reports the mean, minimum, and maximum values of the evaluation metrics over the ten independent runs for each biomass estimation model. All estimation models exhibit a weak correlation between reference and estimated biomass, with R^2 values ranging between 0.36 and 0.4, except the one for the bark biomass estimation, which exhibits even lower performance.

More specifically, the RF model for branch biomass estimation provided the best results compared to all other models, achieving mean values of R^2 at 0.40, MAE at 566.20 kg, rbias at -0.16, and RMSE% at 37.42%. The foliage biomass estimation model provided similar results to that of the branches, with slightly decreased R^2 (0.39), slightly increased RMSE% (37.63%), and rbias (-0.19). Regarding the dead branches model, the metrics revealed a relatively small RMSE% (31.19%) and almost similar R^2 (0.38) to the aforementioned models. However, the trunk and bark models showed the lowest predictive ability. More specifically, the trunk model shows a slightly reduced R^2 (0.36) and rbias (-0.10) compared to the other models, however, a highly reduced RMSE% (27.76%) is observed. More specifically, the trunk model showed a slightly reduced R^2 (0.36) and rbias (-0.10) compared to the other models, but an extremely reduced RMSE% (27.76%) is observed. Similar to the trunk model, the bark model showed the lowest RMSE% and rbias (22.23% and -0.05), in contrast to the worst performance in terms of R^2 (0.12).

Table 4. Testing performance of the RF algorithm of the AGB components. The AVG, MAX, MIN, and STD_STATS represent, respectively, the average, maximum, minimum, and standard deviation over the ten independent runs.

Biomass Component	Statistics	MAE	Bias	Bias%	RMSE	RMSE%	R^2

Stem	AVG_STATS	1659.17	-60.44	-0.10	2104.21	27.76	0.36
	MAX_STATS	1834.15	240.79	-0.06	2329.62	30.67	0.46
	MIN_STATS	1531.11	-337.78	-0.14	1820.39	23.64	0.21
	STD_STATS	114.30	159.15	0.03	180.69	2.41	0.08
Foliage	AVG_STATS	339.66	-18.87	-0.19	442.95	37.63	0.39
	MAX_STATS	370.41	10.18	-0.13	491.59	40.98	0.49
	MIN_STATS	312.54	-53.18	-0.27	399.75	33.87	0.30
	STD_STATS	19.26	23.74	0.05	28.98	2.43	0.06
Dead Branches	AVG_STATS	52.73	-1.82	-0.12	67.81	31.19	0.38
	MAX_STATS	57.52	6.50	-0.08	73.76	33.62	0.46
	MIN_STATS	48.78	-10.63	-0.18	60.91	27.45	0.28
	STD_STATS	2.98	5.46	0.04	4.40	2.17	0.06
Branches	AVG_STATS	566.20	35.88	-0.16	744.30	37.42	0.40
	MAX_STATS	645.06	162.14	-0.07	820.20	41.70	0.47
	MIN_STATS	526.25	-104.68	-0.27	693.83	34.93	0.32
	STD_STATS	39.66	85.61	0.07	42.07	2.00	0.05
Bark	AVG_STATS	89.61	-1.23	-0.05	116.52	22.23	0.12
	MAX_STATS	99.84	13.24	-0.01	138.27	26.60	0.30
	MIN_STATS	78.58	-24.74	-0.09	100.51	19.01	0.02
	STD_STATS	7.41	11.50	0.02	12.30	2.29	0.09

The results above indicate that there are rather weak correlations between the estimated and reference biomass values when considering all available GEDI footprints. Hence, we tried to investigate the influence of two parameters that could theoretically affect the results: biomass density, since the penetration is the LiDAR signal is theoretically expected to decrease in denser forested areas, and slope, which severely affects the signal.

3.2 Effects of biomass density and topography on biomass estimation

To investigate the influence of biomass density, the available GEDI footprints were classified into three different biomass densities (Low, Medium, and High), where Low refers to plots with biomass less than 8626 kg/footprint (0-25%), Medium from 8,626 to 11,950 kg/footprint (25-75%), and High greater than 11,950 kg/footprint (75-100%). Accordingly, a different model was trained for each biomass category and tree component, performing ten independent runs in each case, with a 70%–30% split for training and testing data, respectively, as done in the previous experiment. Table *5* reports the performance metrics obtained, similarly to the representation of Table *4*.

Table 5. Testing performance of the RF algorithm of the AGB components for each biomass density class (i.e., Low, Medium, and High). The reported metrics represent the average value from the 10 iterations for every biomass component and biomass density.

Biomass Component	Biomass Density	MAE	Bias	Bias%	RMSE	RMSE%	R^2
Stem	Low	713.68	-46.13	-0.07	909.72	20.96	0.12
	Medium	880.01	116.75	0.00	1050.91	13.68	0.07

	High	1158.37	-144.05	-0.03	1524.77	13.83	0.06
Foliage	Low	120.36	-4.19	-0.12	151.38	27.99	0.19
	Medium	177.05	9.66	-0.03	222.42	19.17	0.18
	High	300.13	20.78	-0.03	405.26	21.07	0.16
Dead Branches	Low	20.43	-1.33	-0.08	25.37	22.14	0.17
	Medium	27.85	2.16	-0.01	33.97	15.59	0.10
	High	41.64	-0.43	-0.02	54.31	16.34	0.09
Branches	Low	203.81	-10.06	-0.12	259.49	28.49	0.20
	Medium	287.21	14.26	-0.03	361.88	18.62	0.18
	High	472.32	22.06	-0.03	628.23	19.89	0.16
Bark	Low	33.56	-4.30	-0.96	42.49	25.01	0.11
	Medium	43.98	6.89	-0.01	54.22	19.84	0.11
	High	70.85	-4.22	-0.16	88.40	32.55	0.08

Low biomass areas (<8,651 kg/area) show the highest metrics in all biomass categories, with branches having the highest R^2 (0.20) and RMSE% (28.49%). The needle and dead branches biomass estimation models performed marginally lower on the low biomass density surfaces, with R^2 at 0.20 and 0.17 and RMSE% at 27.99% and 22.14% respectively. In the medium and high biomass density plots, lower accuracies were observed in all forest biomass parts, with the branch and needle models showing the highest performance among the others. Finally, the bark and stem models showed the lowest predictive ability across all biomass densities. However, it is worth noting that RMSE% decreased significantly in all models at medium and high biomass densities, except for the bark model.

To investigate the influence of topography, the available GEDI plots were split into two categories based on the footprint's average slope: low slope / gentle topography (inclinations 0–15%) and medium or high slope (inclinations above 15%). Unlike forest biomass density, slopes significantly affected the accuracy of all model estimates (Table 6). All but the bark estimation model exhibited much higher accuracy in areas with low slopes and gentle topography (0-15% slope). Similarly to the previous experiments, the crown parts (foliage and branches) provided the best estimations. More specifically, the branch model showed the highest R^2 (0.47) and satisfactory RMSE% (32.77%) in low-slope regions. The models of live and dead branches showed marginally lower R^2 (0.46 and 0.44 respectively), with the latter showing a rather low—or at least much lower than the previous experiments—RMSE% (28.00%). Stem biomass can be estimated with satisfactory accuracy (R^2 = 0.42) and a rather low RMSE% (24.62%), in contrast to the previous cases, where this biomass category could not be estimated with high confidence. Finally, the bark biomass model showed the lowest predictive performance, with R^2 = 0.15 in low and R^2 = 0.08 in medium gradients, similarly to the previous cases.

Table 6. Testing performance of the RF algorithm of the AGB components for each slope category (i.e., Low and Medium). The reported metrics represent the average value from the 10 iterations for every biomass component and biomass density.

Biomass Component	Slope Category	MAE	Bias	Bias%	RMSE	RMSE%	R^2
Stem	Low (0-15%)	1512.05	147.46	-0.04	1905.50	24.62	0.42
	Medium (>16%)	2230.95	138.11	-0.21	2677.08	35.17	0.13
Foliage	Low (0-15%)	319.44	22.21	-0.11	410.93	34.15	0.46
	Medium (>16%)	449.33	-44.50	-0.41	539.67	46.53	0.20
	Low (0-15%)	48.98	3.65	-0.06	62.18	28.00	0.44

Dead branches	Medium (>16%)	70.68	-3.85	-0.26	84.59	39.35	0.16
Branches	Low (0-15%)	516.32	41.34	-0.10	657.27	32.77	0.47
	Medium (>16%)	733.42	-60.06	-0.40	890.10	45.93	0.20
Bark	Low (0-15%)	14.91	0.82	-0.06	21.43	31.32	0.15
	Medium (>16%)	70.84	-4.20	-0.15	88.40	32.54	0.08

4. DISCUSSION

This study focuses on the potential of GEDI LiDAR satellite data combined with the random forest (RF) algorithm to estimate forest biomass components in a coniferous forest. Airborne LiDAR data were used to generate reference data through the utilization of the stem, bark, branches (live and dead), and needle biomass estimation techniques. The generated data were used to investigate the ability of the GEDI L2A and L2B satellite data to provide reliable biomass estimates through the utilization of height metrics, plant surface profiles, and vertical plant surface volume density profiles. In addition, the RF algorithm was selected for biomass estimation, as it is a fairly widespread algorithm that provides accurate forest biomass estimates. In addition to the basic estimates, the ability of the GEDI data to provide biomass estimates in areas with high biomass density and complex topography was examined.

The first and key part of the analysis involves removing GEDI measurements that are identified as erroneous and of low quality, in order to retain only the high-quality data. Therefore, the 'Quality_Flag' and 'Degrade_Flag' fields were used to remove measurements with low quality and positional errors, and elevation measurements with extremely high values exceeding 45 m in height were removed (Hoffrén et al, 2023). Furthermore, it was observed that a multitude of “rh” field values (rh_1 to rh_10) had negative height values, which occurs mainly in areas with low land cover (Di Tommaso et al., 2021).

In the case of crown biomass (dead and live branches, needles), the RF model using GEDI metrics provided higher estimation accuracies compared to those of the trunk. More specifically, branch biomass was estimated with the highest accuracy (R^2=0.40, MAE=566.20 kg, rbias=-0.16 and RMSE%=37.42%), with the estimation accuracies of needles and dead branches lagging slightly behind. The difference in accuracy between the trunk and canopy parts may be due to direct contact of the signal with the surface of branches and needles [32]. The diameter at breast height and therefore the biomass of the trunk parts is not proportional to the crown area, especially in Abies borisii-regis. Stem biomass can be estimated with reasonable accuracy compared to that of the bark, however, given the importance of this part of biomass in forest management, the accuracy of the results is considered insufficient.

To improve the above results and find the main factors influencing the estimates, the GEDI measurements were categorized by biomass density and slope. The above categorization was chosen based on the literature, as biomass density and slope significantly affect the GEDI signal [27], [31]. Therefore, three different classes of biomass density (Low < 8,626 kg/footprint, Medium from 8,626 to 11,950 kg/footprint, and High > 11,950 kg/footprint) and two classes of slopes (low < 15% and medium to steep slopes >16%) were created. Accordingly, the ability of GEDI satellite data to provide reliable biomass estimates in areas with different biomass densities was examined. Results showed that areas with low biomass density provide more accurate estimates in all forest biomass components, with the highest performance observed at the branch biomass (R^2 = 0.20 and RMSE% = 28.49%). It is worth noting that results in all density classes are significantly lower than the overall results. The large difference in the accuracy of the estimates may be due to the large deviation of the estimated GEDI heights from reality, leading to large variations [48]. Regarding the topography, the results showed the significant influence of slope on the accuracy of the estimates. Areas with mild topography and slopes of up to 15% showed significantly higher results than areas with medium to steep slopes (>16%). More specifically, all major components of forest biomass were estimated with satisfactory accuracy, with branch biomass providing the highest estimation accuracy R^2 (0.47) and RMSE% (32.77 %). Stem biomass on gently sloping surfaces was estimated with satisfactory R^2 (0.42) accuracy, presenting the smallest estimation error with RMSE% (32.77 %). These results are in agreement with previous studies highlighting the significant effect of slope on estimates of woody volume, height, and forest biomass, especially in areas with slopes greater than 10% [31], [48], [49].

Although the GEDI sensor is a quite widespread tool for estimating forest biomass, no study to the best of our knowledge has examined the accuracy of estimating the key parts (stem, bark, dead and live branches, needles) of forest AGB. Hence, any comparisons with previous findings can only done in rather vague terms, e.g., via the approximate ranges of the reported R^2 or RMSE% values for total AGB or timber volume, keeping in mind that the uncertainties of total AGB estimations are not necessarily directly comparable with AGB components estimations. Fayad et al. (2021) used GEDI data to estimate height and woody volume in eucalyptus plantations, showing much higher volume estimation accuracies than the present study, with R^2 exceeding 0.75 across all slope categories. However, the reported RMSE% values were higher, ranging from 26.78 to 48.63%. These results agree with those of the present study, in terms of the influence of topography on the accuracy of the estimates. It is also worth noting that the significant difference observed between the two studies in R^2 is due to the different ecosystem types, as one considers a eucalyptus plantation and the present one considers a multilayered naturally regenerated fir forest. Sun et al. (2022) evaluated the ability of GEDI data to provide reliable measurements of crown height, vertical structure, and AGB in tropical and temperate forests, showing very high estimation accuracy (R^2=0.82). However, the reported statistics refer to training rather than testing accuracy (i.e., they are the results of linear regression correlation analysis rather than the results of a predictive model), in which the results are expected to be much lower. The effect of slope and topography on forest biomass estimation in mountainous forests was examined by Ni et al. (2021) who showed similar results to the present work, with R^2=0.48, which was significantly improved by using geometrically corrected data. Finally, Xu et al. (2023) estimated oak biomass using GEDI data, showing higher results than the present study. Variables extracted from MODIS, such as modis_treecover, are also used, which significantly increases the estimation accuracy.

This research highlights the ability of GEDI data to provide forest biomass estimates with moderate accuracy in a multilayered fir forest with intense topography. A key limitation is the lack of a sufficient number of GEDI measurements in the study area due to the unsuitability of a large percentage of observations. Also, this study did not use any form of GEDI data clearing based on aerial LiDAR data, comparing digital crown model values and estimated GEDI heights. This action may significantly increase the accuracy of the estimates but would possibly provide a false picture of the GEDI's data true ability to estimate forest biomass. In summary, this study demonstrates that GEDI data are capable of providing satisfactory forest biomass estimates in a coniferous forest canopy with relatively low errors. This research highlights the importance of using GEDI data for aboveground forest biomass estimates in coniferous forests, providing estimates for large areas and less accessible areas. In addition, these estimates provide an economical solution for estimating forest biomass at large scales where field measurements and other remote sensing data are quite costly. Finally, this study is the first contribution to the framework for using GEDI data to estimate the key parts of the aboveground forest biomass using aerial LiDAR estimates as reference data, examining the influence of forest biomass density and topography on the estimates.

5. CONCLUSIONS

This study aimed to estimate the biomass of the main parts of the AGB using LiDAR satellite data in a coniferous forest. Specifically, GEDI sensor data were utilized to estimate the biomass of stem, bark, branches (live and dead), and needles using the RF machine learning regression algorithm. The results showed that LiDAR satellite data can provide satisfactory biomass estimates in a conifer forest in low-slope, low-relief areas with gentle topography. More specifically:

- The crown biomass (live and dead branches, needles) can be estimated with satisfactory accuracy and relatively low errors using the RF algorithm.
- Bark and trunk biomasses lag significantly in estimation accuracy compared to crown but exhibit lower estimation errors.
- Forest biomass density did not affect the accuracy of the estimates, as in all biomass density classes the accuracy of the estimates was too low.
- Slope and topography significantly affect biomass estimates, as the GEDI signal has several errors in height metrics at slopes greater than 16%.
- The models estimating all forest biomass parts have relatively low accuracy but could be used to study the distribution of forest biomass at the national level.

Overall, it is evident that LiDAR satellites could successfully contribute to the quantification of forest biomass and carbon stocks in a variety of forest ecosystems. This study contributes to the improvement of forest biomass estimates and a better understanding of the behaviour of satellite LiDAR data in natural ecosystems. Given the limitations of the study described in the previous section, it would be interesting to explore the potential of LiDAR satellite data in biomass estimation at the national level, analysing different vegetation types and large numbers of GEDI measurements. Finally, future research could also focus on the combined use of space-based and airborne LiDAR systems to provide a cost-effective approach for large-scale forest biomass estimation.

REFERENCES

[1] M. L. R. Sarker and J. Nichol, "Forest biomass estimation from the fusion of C-band SAR and optical data using wavelet transform," in *Remote Sensing for Agriculture, Ecosystems, and Hydrology XV*, C. M. U. Neale and A. Maltese, Eds., SPIE, 2013, p. 88870S. doi: 10.1117/12.2029043.

[2] N. L. Harris *et al.*, "Global maps of twenty-first century forest carbon fluxes," *Nature Climate Change*, vol. 11, no. 3, pp. 234–240, Mar. 2021, doi: 10.1038/s41558-020-00976-6.

[3] P. Zeng, W. Zhang, Y. Li, J. Shi, and Z. Wang, "Forest Total and Component Above-Ground Biomass (AGB) Estimation through C- and L-band Polarimetric SAR Data," *Forests*, vol. 13, no. 3, p. 442, Mar. 2022, doi: 10.3390/f13030442.

[4] M. A. Njana, "Indirect methods of tree biomass estimation and their uncertainties," *Southern Forests: a Journal of Forest Science*, vol. 79, no. 1, pp. 41–49, Jan. 2017, doi: 10.2989/20702620.2016.1233753.

[5] S. Luo *et al.*, "Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation," *Ecological Indicators*, vol. 73, pp. 378–387, Feb. 2017, doi: 10.1016/j.ecolind.2016.10.001.

[6] P. Migolet, K. Goïta, A. Ngomanda, and A. P. M. Biyogo, "Estimation of Aboveground Oil Palm Biomass in a Mature Plantation in the Congo Basin," p. 23, 2020.

[7] D. Zianis and M. Mencuccini, "On simplifying allometric analyses of forest biomass," *Forest Ecology and Management*, vol. 187, no. 2–3, pp. 311–332, Jan. 2004, doi: 10.1016/j.foreco.2003.07.007.

[8] D. Lu, Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran, "A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems," *International Journal of Digital Earth*, vol. 9, no. 1, pp. 63–105, Jan. 2016, doi: 10.1080/17538947.2014.990526.

[9] P. Naik, M. Dalponte, and L. Bruzzone, "Prediction of Forest Aboveground Biomass Using Multitemporal Multispectral Remote Sensing Data," *Remote Sensing*, vol. 13, no. 7, p. 1282, Mar. 2021, doi: 10.3390/rs13071282.

[10] X. Wang and H. Jiao, "Spatial Scaling of Forest Aboveground Biomass Using Multi-Source Remote Sensing Data," *IEEE Access*, vol. 8, pp. 178870–178885, 2020, doi: 10.1109/ACCESS.2020.3027361.

[11] A. B. Debastiani, C. R. Sanquetta, A. P. D. Corte, F. E. Rex, and N. S. Pinto, "Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest," p. 14, 2019.

[12] G. Forkuor, J.-B. B. Zoungrana, K. Dimobe, B. Ouattara, K. P. Vadrevu, and J. E. Tondoh, "Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets-A case study," *Remote Sensing of Environment*, vol. 236, p. 111496, 2020.

[13] S. Kaasalainen, "Combining Lidar and Synthetic Aperture Radar Data to Estimate Forest Biomass: Status and Prospects," p. 19, 2015.

[14] M. Santoro and O. Cartus, "Research Pathways of Forest Above-Ground Biomass Estimation Based on SAR Backscatter and Interferometric SAR Observations," *Remote Sensing*, vol. 10, no. 4, p. 608, Apr. 2018, doi: 10.3390/rs10040608.

[15] Yunjin Kim and J. van Zyl, "Comparison of forest parameter estimation techniques using SAR data," in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, Jul. 2001, pp. 1395–1397 vol.3. doi: 10.1109/IGARSS.2001.976856.

[16] S. Sinha, C. Jeganathan, L. K. Sharma, and M. S. Nathawat, "A review of radar remote sensing for biomass estimation," *Int. J. Environ. Sci. Technol.*, vol. 12, no. 5, pp. 1779–1792, May 2015, doi: 10.1007/s13762-015-0750-0.

[17] X. Jiang, G. Li, D. Lu, E. Chen, and X. Wei, "Stratification-Based Forest Aboveground Biomass Estimation in a Subtropical Region Using Airborne Lidar Data," *Remote Sensing*, vol. 12, no. 7, p. 1101, Mar. 2020, doi: 10.3390/rs12071101.

[18] S. P. Healey, Z. Yang, N. Gorelick, and S. Ilyushchenko, “Highly Local Model Calibration with a New GEDI LiDAR Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation,” *Remote Sensing*, vol. 12, no. 17, p. 2840, Sep. 2020, doi: 10.3390/rs12172840.
[19] J. Lee, J. Im, K. Kim, and L. Quackenbush, “Machine Learning Approaches for Estimating Forest Stand Height Using Plot-Based Observations and Airborne LiDAR Data,” *Forests*, vol. 9, no. 5, p. 268, May 2018, doi: 10.3390/f9050268.
[20] R. Hoffrén *et al.*, “Assessing GEDI-NASA system for forest fuels classification using machine learning techniques,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103175, Feb. 2023, doi: 10.1016/j.jag.2022.103175.
[21] L. Chen *et al.*, “Improved Object-Based Estimation of Forest Aboveground Biomass by Integrating LiDAR Data from GEDI and ICESat-2 with Multi-Sensor Images in a Heterogeneous Mountainous Region,” *Remote Sensing*, vol. 14, no. 12, p. 2743, Jun. 2022, doi: 10.3390/rs14122743.
[22] J. R. Kellner, J. Armston, and L. Duncanson, “Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density,” *Earth and Space Science*, vol. 10, no. 4, p. e2022EA002516, Apr. 2023, doi: 10.1029/2022EA002516.
[23] W. Qi, S. Saarela, J. Armston, G. Ståhl, and R. Dubayah, “Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data,” *Remote Sensing of Environment*, vol. 232, p. 111283, Oct. 2019, doi: 10.1016/j.rse.2019.111283.
[24] S. P. Healey, P. L. Patterson, S. Saatchi, M. A. Lefsky, A. J. Lister, and E. A. Freeman, “A sample design for globally consistent biomass estimation using lidar data from the Geoscience Laser Altimeter System (GLAS),” *Carbon Balance Manage*, vol. 7, no. 1, p. 10, Dec. 2012, doi: 10.1186/1750-0680-7-10.
[25] A. Liu, X. Cheng, and Z. Chen, “Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals,” *Remote Sensing of Environment*, vol. 264, p. 112571, Oct. 2021, doi: 10.1016/j.rse.2021.112571.
[26] R. Dubayah *et al.*, “The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography,” *Science of Remote Sensing*, vol. 1, p. 100002, Jun. 2020, doi: 10.1016/j.srs.2020.100002.
[27] I. Dorado-Roda *et al.*, “Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests,” *Remote Sensing*, vol. 13, no. 12, p. 2279, Jun. 2021, doi: 10.3390/rs13122279.
[28] L. Duncanson *et al.*, “Aboveground biomass density models for NASA’s Global Ecosystem Dynamics Investigation (GEDI) lidar mission,” *Remote Sensing of Environment*, vol. 270, p. 112845, Mar. 2022, doi: 10.1016/j.rse.2021.112845.
[29] W. Ni, Z. Zhang, and G. Sun, “Assessment of Slope-Adaptive Metrics of GEDI Waveforms for Estimations of Forest Aboveground Biomass over Mountainous Areas,” *J Remote Sens*, vol. 2021, p. 2021/9805364, Jan. 2021, doi: 10.34133/2021/9805364.
[30] Y. Shendryk, “Fusing GEDI, Sentinel-1, Sentinel-2, and elevation data for seasonal forest biomass mapping across Australia,” display, other, Mar. 2022. doi: 10.5194/egusphere-egu22-2077.
[31] M. Sun *et al.*, “Evaluation of NASA’s GEDI Lidar Observations for Estimating Biomass in Temperate and Tropical Forests,” *Forests*, vol. 13, no. 10, p. 1686, Oct. 2022, doi: 10.3390/f13101686.
[32] L. Xu *et al.*, “Estimation of Quercus Biomass in Shangri-La Based on GEDI Spaceborne Lidar Data,” *Forests*, vol. 14, no. 5, p. 876, Apr. 2023, doi: 10.3390/f14050876.
[33] L. Duncanson *et al.*, “Forest Aboveground Biomass Estimation with GEDI and ICESat-2 in Boreal Forests,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium: IEEE, Jul. 2021, pp. 670–672. doi: 10.1109/IGARSS47720.2021.9553209.
[34] M. Adam, M. Urbazaev, C. Dubois, and C. Schmullius, “Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters,” *Remote Sensing*, vol. 12, no. 23, p. 3948, Dec. 2020, doi: 10.3390/rs12233948.
[35] C. Silva *et al.*, “rGEDI: NASA’s Global Ecosystem Dynamics Investigation (GEDI) Data Visualization and Processing.” 2020.
[36] R. Dubayah, H. Tang, J. Armstrong, S. Luthcke, M. Hofton, and J. B. Blair, “GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V002 [Data set].” NASA EOSDIS Land Processes DAAC, 2021.

[37] D. Morin *et al.*, "Improving Heterogeneous Forest Height Maps by Integrating GEDI-Based Forest Height Information in a Multi-Sensor Mapping Process," *Remote Sensing*, vol. 14, no. 9, p. 2079, Apr. 2022, doi: 10.3390/rs14092079.
[38] C. Sothe, A. Gonsamo, R. B. Lourenço, W. A. Kurz, and J. Snider, "Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel," *Remote Sensing*, vol. 14, no. 20, p. 5158, Oct. 2022, doi: 10.3390/rs14205158.
[39] N. Georgopoulos, I. Z. Gitas, A. Stefanidou, L. Korhonen, and D. Stavrakoudis, "Estimation of Individual Tree Stem Biomass in an Uneven-Aged Structured Coniferous Forest Using Multispectral LiDAR Data," *Remote Sensing*, vol. 13, no. 23, p. 4827, Nov. 2021, doi: 10.3390/rs13234827.
[40] N. Georgopoulos, I. Z. Gitas, L. Korhonen, K. Antoniadis, and A. Stefanidou, "Estimating Crown Biomass in a Multilayered Fir Forest Using Airborne LiDAR Data," *Remote Sensing*, vol. 15, no. 11, p. 2919, Jun. 2023, doi: 10.3390/rs15112919.
[41] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, 2017. [Online]. Available: https://www.R-project.org/
[42] J.-R. Roussel *et al.*, "lidR: An R package for analysis of Airborne Laser Scanning (ALS) data," *Remote Sensing of Environment*, vol. 251, p. 112061, Dec. 2020, doi: 10.1016/j.rse.2020.112061.
[43] D. Gatziolis, "Dynamic Range-based Intensity Normalization for Airborne, Discrete Return Lidar Data of Forest Canopies," *photogramm eng remote sensing*, vol. 77, no. 3, pp. 251–259, Mar. 2011, doi: 10.14358/PERS.77.3.251.
[44] W. Zhang *et al.*, "An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation," *Remote Sensing*, vol. 8, no. 6, p. 501, Jun. 2016, doi: 10.3390/rs8060501.
[45] M. Marabel and F. Alvarez-Taboada, "Spectroscopic Determination of Aboveground Biomass in Grasslands Using Spectral Transformations, Support Vector Machine and Partial Least Squares Regression," *Sensors*, vol. 13, no. 8, pp. 10027–10051, Aug. 2013, doi: 10.3390/s130810027.
[46] S. Tibshirani and H. Friedman, *The Elements of Statistical Learning*. 2008.
[47] J. Wainer and G. Cawley, "Nested cross-validation when selecting classifiers is overzealous for most practical applications," *Expert Systems with Applications*, vol. 182, p. 115222, Nov. 2021, doi: 10.1016/j.eswa.2021.115222.
[48] K. Lahssini, N. Baghdadi, G. le Maire, and I. Fayad, "Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests," *Remote Sensing*, vol. 14, no. 24, p. 6264, Dec. 2022, doi: 10.3390/rs14246264.
[49] I. Fayad *et al.*, "Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 14, pp. 7095–7110, 2021, doi: 10.1109/JSTARS.2021.3092836.