Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil

Ibrahim Fayad, Nicolas N. Baghdadi, Clayton Alcarde Alvares, Jose Luiz Stape, Jean Stéphane Bailly, Henrique Ferraço Scolforo, Mehrez Zribi, and Guerric Le Maire

***Abstract*—Over the past two decades spaceborne LiDAR systems have gained momentum in the remote sensing community with their ability to accurately estimate canopy heights and aboveground biomass. This article aims at using the most recent global ecosystem dynamics investigation (GEDI) LiDAR system data to estimate the stand-scale dominant heights (H_{dom}), and stand volume (V) of Eucalyptus plantations in Brazil. These plantations provide a valuable case study due to the homogenous canopy cover and the availability of precise field measurements. Several linear and nonlinear regression models were used for the estimation of H_{dom} and V based on several GEDI metrics. H_{dom} and V estimation results showed that over low-slopped terrain the most accurate estimates of H_{dom} and V were obtained using the stepwise regression, with an root-mean-square error (RMSE) of 1.33 m (R^2 of 0.93) and 24.39 $m^3.ha^{-1}$ (R^2 of 0.90) respectively. The principal metric explaining more than 87% and 84% of the variability (R^2) of H_{dom} and V was the metric representing the height above the ground at which 90% of the waveform energy occurs. Testing the postprocessed GEDI metric values issued from six available different processing algorithms showed that the accuracy on H_{dom} and V estimates is algorithm dependent, with a 16% observed increase in RMSE on both variables using algorithm a5 vs. a1. Finally, the choice to select the ground return from the last detected mode or the stronger of the last two modes could also affect the H_{dom} estimation accuracy with 12 cm RMSE decrease using the latter.**

Manuscript received November 13, 2020; revised January 11, 2021 and May 5, 2021; accepted June 15, 2021. Date of publication June 28, 2021; date of current version July 26, 2021. This work was supported in part by the French Space Study Center (CNES, TOSCA 2020 project), and in part by the National Research Institute for Agriculture, Food, and the Environment (INRAE). *(Corresponding author: Ibrahim Fayad.)*

Ibrahim Fayad and Nicolas N. Baghdadi are with the French National Research Institute for Agriculture, Food and the Environment (INRAE), CIRAD, CNRS, TETIS, AgroParisTech, Université de Montpellier, 34093 Montpellier, France (e-mail: ibrahim.fayad@inrae.fr; nicolas.baghdadi@inrae.fr).

Clayton Alcarde Alvares is with the UNESP, Faculdade de Ciências Agronômicas Botucatu 18610-034, Brazil, and also with the Suzano SA, Limeira 13465-970, Brazil (e-mail: calcarde@suzano.com.br).

Jose Luiz Stape is with the UNESP, Faculdade de Ciências Agronômicas, Botucatu 18610-034, Brazil (e-mail: jlstape@gmail.com).

Jean Stéphane Bailly is with the INRAE, IRD, Institut Agro, LISAH, Université de Montpellier, 34060 Montpellier, France, and also with the AgroParisTech, 75005 Paris, France (e-mail: bailly@agroparistech.fr).

Henrique Ferraço Scolforo is with the Suzano SA, Limeira 13465-970, Brazil (e-mail: hscolforo@suzano.com.br).

Mehrez Zribi is with the Center for the Study of the Biosphere from Space (CNRS/UPS/IRD/CNES/INRAE), 31401 Toulouse, France (e-mail: mehrez.zribi@ird.fr).

Guerric Le Maire is with the CIRAD, UMR Eco&Sols, 34398 Montpellier, France, and also with the Eco&Sols, CIRAD, INRA, IRD, Montpellier SupAgro, Université de Montpellier, 34060 Montpellier, France (e-mail: guerric.le_maire@cirad.fr).

Digital Object Identifier 10.1109/JSTARS.2021.3092836

***Index Terms*—Brazil, dominant heights, eucalyptus, global ecosystem dynamics investigation (GEDI), LiDAR, wood volume.**

I. Introduction

In THE last couple of decades, global concerns on the increased atmospheric concentration of greenhouse gases, such as CO_2 has risen the interest in quantifying the state and change of forest resources due to the key role of forests in the global carbon cycle [1], [2]. Forests sequester a large quantity of carbon in their woody biomass where they store around 70% to 90% of the global terrestrial biomass ranging from $385x10^9$ to $650x10^9$ Mg [3]. Hence, the accurate estimation of forest biomass is needed to better determine its precise role in the global carbon cycle [4], [5]. Forest plantations represent a small fraction (6.9%) of the total forested land ([6]) but are becoming increasingly important around the world, economically, socially and environmentally ([7], [8]).

The primary source of above ground biomass (AGB) estimation in tropical forests at large scales came in the last years from observations and measurements from different satellite remote sensing platforms. Methods based on remotely sensed data are less accurate than field measurements, however, their major advantages are their global and frequent coverage and the low or free acquisition costs for the end user. Currently optical, radar, and LiDAR are the three main sources of remotely sensed data used in AGB estimation techniques. Nonetheless, current data sources are either limited to low AGB levels (<150 Mg/ha) (sensor saturation at certain biomass levels with radar and optical data) or have a limited spatial coverage (e.g., airborne LiDAR data). LiDAR systems either airborne or spaceborne have the capability to capture the horizontal and vertical structure of vegetation comprehensively [9], and can thus estimate biomass with better precision in comparison to the techniques using radar or optical data [10], [11]. To date, there have been only three satellite LiDAR missions. The first mission was the Ice, cloud, and land elevation satellite (ICESat-1) which carried the geoscience laser altimeter system (GLAS) from 2003 until 2009 [12]. Although GLAS's ~60 m diameter footprint was larger than the ideal resolution for forest observations [13], its capability to estimate forest parameters (e.g., canopy heights and biomass) has been exploited in numerous studies during its operational and post-operational periods [5], [14]–[20].

ICESat-1 was followed in 2018 by ICESat-2 that carried the advanced topographic laser altimeter system (ATLAS) with a goal to measure ice-sheet topography, cloud and atmospheric properties and global vegetation. However, the wavelength of the equipped laser (532 nm) has a spectral region of high radiation absorption by the vegetation. This results in a low number of reflected photons measured by ATLAS over vegetation [21], and limits its ability to estimate forest canopy heights [21].

The most recent spaceborne LiDAR system is GEDI on board the ISS, which was launched in December 2018 with on-orbit checkout in April 2019. GEDI's mission objective is to provide information about canopy structure, biomass and topography, and is estimated to acquire 10 billion cloud free shots in its two years mission [22]. GEDI measures vertical structures similar to ICESat-1 (i.e., waveforms). However, given GEDI's higher sampling rate (242 versus 40 Hz for ICESat-1), and the much smaller footprint size ($\sim$25 versus $\sim$60 m for ICESat-1), GEDI provides a highly improved coverage and waveform precision.

GEDI's ability to estimate forest height and wood volume on different types of forest ecosystems, topography and latitudes is of paramount importance. GEDI datasets are organized in different levels of products, from raw acquisition data to more elaborated data obtained by performing signal analysis and metrics extraction from the waveforms. This results in a large number of metrics for each acquired footprint, from which, many different models could be used to retrieve canopy heights and wood volume. While direct metrics could be used as good proxies, it is however acknowledged that combining different metrics yields higher accuracies. For instance, such algorithms make use of linear or non-linear regression models applied on sets of metrics extracted from GEDI waveforms, and eventually combined with digital elevation models (DEMs). The full waveform LiDAR data can potentially give access to more information on canopy structure than the basic "top" and "bottom" return signals, being itself potentially informative for canopy height and volume prediction. Therefore, it is critical to explore which metrics, or combination of metrics, and with which type of models (e.g., linear versus nonlinear) provide the best forest parameter estimates. It is also important to evaluate the effect of the uncertainty of the metrics estimation themselves, which results from differences in preprocessing algorithms, as well as other acquisition characteristics that may influence the final models, such as beam acquisition angles.

Precise evaluation of forest height and volume is not an easy task. One of the main issues is that uncertainties in field measurements can propagate through the models and create larger uncertainties in the estimates [23]. For example, Saarela *et al*. [24] and Holm *et al*. [25] found that not accounting for errors in field measurements could underestimate the uncertainty in final satellite-based AGB maps by a factor of three or more. Feldpausch *et al.* [26] and Kearsley *et al.* [27] found that uncertainties in tree height measurements led to increased bias in the biomass and carbon stock estimates. Other obstacles include: the influence of tree growth during the timespan between the field measurements and satellite acquisitions which cannot be neglected [28]; the comprehensive model validation limited by the sparsity of *in situ* data [29]; and the method used to measure tree heights [30].

Consequently, in such evaluation and comparison of processing algorithms and models for forest height and volume estimation, the field dataset plays a critical role. In this article, we want to focus on the uncertainty coming from the GEDI metrics and models, minimizing the influence of the uncertainty on *in situ* measurements. To reach this objective, we analyzed a large dataset of forest plantations in Brazil, which has many advantages to serve as a test case: large number of sites, different climate and topographical environments, numerous and frequent measurements, precise measurements of tree heights, good allometric relationships for wood volume and homogeneous canopies, etc. (see description in Section II-C). And even though not representative of all forests, the results obtained on forest plantations can give a notion of the *reachable* precision on height and wood volume estimation using GEDI data on a more structurally simple forest than natural forests, while also removing part of the errors due to *in situ* measurements.

The main objectives of this article are therefore summarized in the following questions.

1) What are the more important GEDI metrics linked to canopy height and volume?
2) Are linear and nonlinear models using subsets of metrics more efficient in predicting height and volume?
3) What is the importance of the different pre-processing algorithms on the final uncertainties?
4) Is there an influence of other acquisition characteristics, such as viewing angle on the estimated forest characteristics?
5) Are other stand information, such as data from DEM or age of the stand relevant for the estimation of height and volume on forest plantations?

The manuscript presents first the GEDI dataset, followed by the processing of GEDI data and the main metrics that will be used for the estimation of canopy heights and wood volume. Next, a description of the used methods for the estimation of the forest characteristics is presented in Section II. Finally, the results, discussions, and main conclusions are presented in Sections IV, V, and VI, respectively

II. Study Site and Datasets

A. Study Area

The study area is located in four regions in Brazil, (Bahia & Espírito Santo, Mato Grosso do Sul, São Paulo, and Maranhão) across a large latitudinal gradient (see Fig. 1) and covering different climate and soil types. The studied plantations are managed in order to produce high yield pulpwood growing at short rotations. Clonal seedlings of mainly *E. grandis* (W. Hill) and *E. urophylla* (S.T. Blake) and different types of hybrids are planted in rows at a density of 1000–1667 trees/ha, rationally fertilized with nitrogen, phosphorus, and potassium and micronutrients to alleviate any nutritional limitations. Harvest occurs every six to seven years, and very little tree mortality (under 7% from original plantation) is noticed. The annual productivity of the plantations was on average 40 m^3/ha/year, with 80% of the stands being between 30–50 m^3/ha/year and some stands could reach values as high as 60 m^3/ha/year. At harvest time, the stand volume is between 180 and 300 m^3/ha, with a dominant height

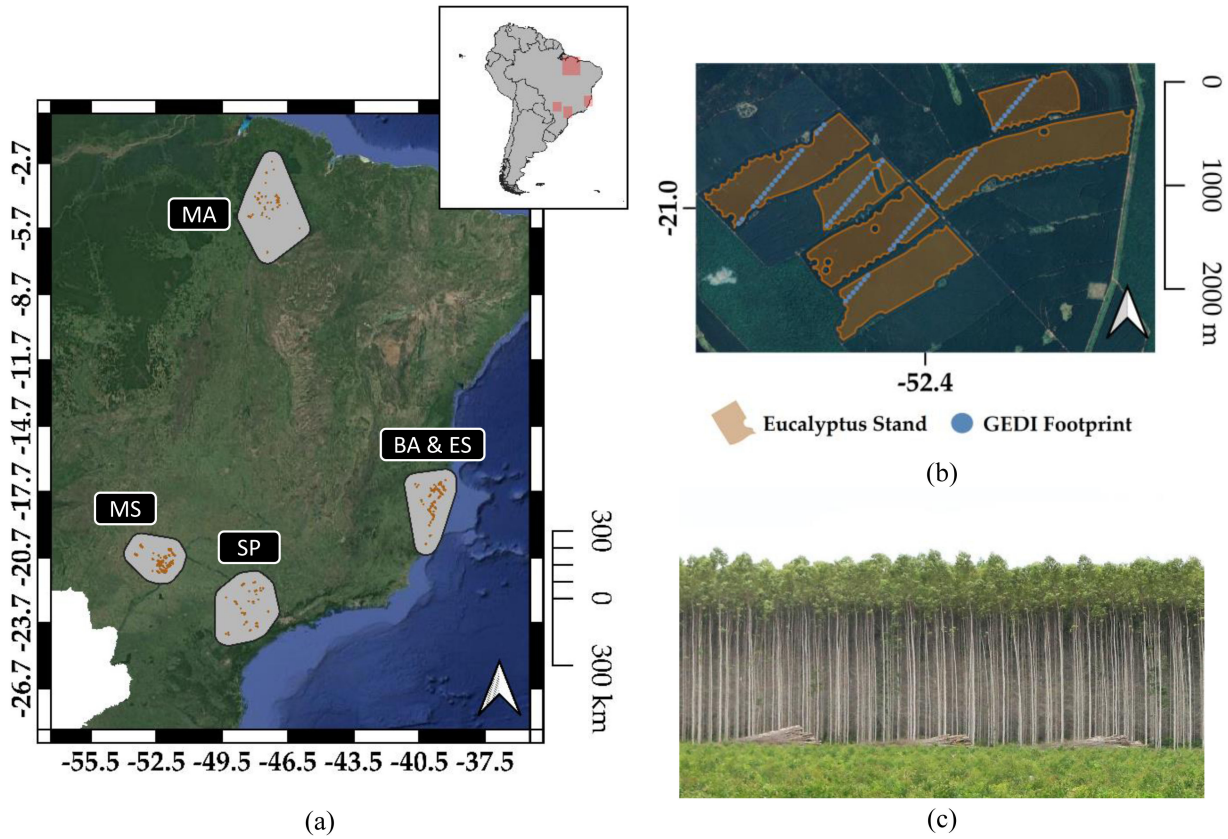


Fig. 1. (a) Location of the four study sites. (b) Example of GEDI tracks over some stands. (c) Eucalyptus stand during harvest (approx. 30 m high) illustrating the clearly separated crown and trunk strata.

of 20 to 35 m range (for 80% of the stands). These plantations were managed locally by stand units, generally around 50 ha, where the same management is applied: planting, harvesting, weed control, genetic material, soil preparation and fertilization. There are generally sparse understory and herbaceous strata in these plantations, as result of chemical weeding the first year, the closing of the canopy, and the high competitive strength of Eucalyptus. Tree height is very homogeneous within a stand, with 95% of the trees having heights at +/- 10.5% around the average tree height in plot inventories. The plantations exhibit a simple structure, with a tree crown strata of 3 to 10 m in width above a "trunk strata" with few Eucalyptus leaves and few understories [see Fig. 1(c)]. The "soil strata" is mainly constituted of litter accumulation of branches and leaves, with some patches of herbaceous species.

B. GEDI Data

1) Processing of GEDI Waveforms: GEDI uses three on-board lasers that produce eight parallel tracks of observations. GEDI lasers illuminate a surface or footprint on the ground with a 25 m diameter, at a frequency of 242 Hz, over which 3D structures are measured. The footprints are separated by ~60 m (center to center) along the track, and the tracks are separated by ~600 m across. Moreover, GEDI has the ability to rotate up to six degrees, allowing the lasers to be pointed as much as 40 km on either side of the ISS's ground track [22]. GEDI measures vertical structures using a 1064-nm laser pulse, and the echoed waveforms are digitized to a maximum of 1246 bins with a vertical resolution of 1 ns (15 cm), corresponding to a maximum of 186.9 m of height ranges, with a vertical accuracy over relatively flat, non-vegetated surfaces of ~3 cm [31].

As described in the algorithm theoretical basis document (ATBD) [32], [33], the received waveforms are first smoothed to reduce the noise in the signal, and thus permitting the determination of the useful part of the waveform within the corresponding footprint. Waveform smoothing is performed by means of a Gaussian filter with various widths. As mentioned in the ATBD, currently a width of 6.5 ns was used for the Gaussian filter (Smooth width). After smoothing, two locations in the waveform denoted as *search start* and *search end* are determined [see Fig. 2(a)]. *search start* and *search end* are, respectively, the first and last positions in the signal where the signal intensity is above the following threshold:

$$\text{threshold} = \text{mean} + \sigma . v \tag{1}$$

where "mean" is the mean noise level, "σ" is the standard deviation of noise of the smoothed waveform, and "v" is a constant currently set at 4. After determining the locations of *search start* and *search end*, the region between them, denoted as the waveform extent, is extended by a predetermined number of sample bins, currently set to 100 bins at both sides. Within the waveform extent, the highest (*toploc*) and lowest (*botloc*) detectable returns are determined [see Fig. 2(a)]. The metrics *toploc* and *botloc* respectively represent the highest and lowest locations within the waveform extent where two adjacent

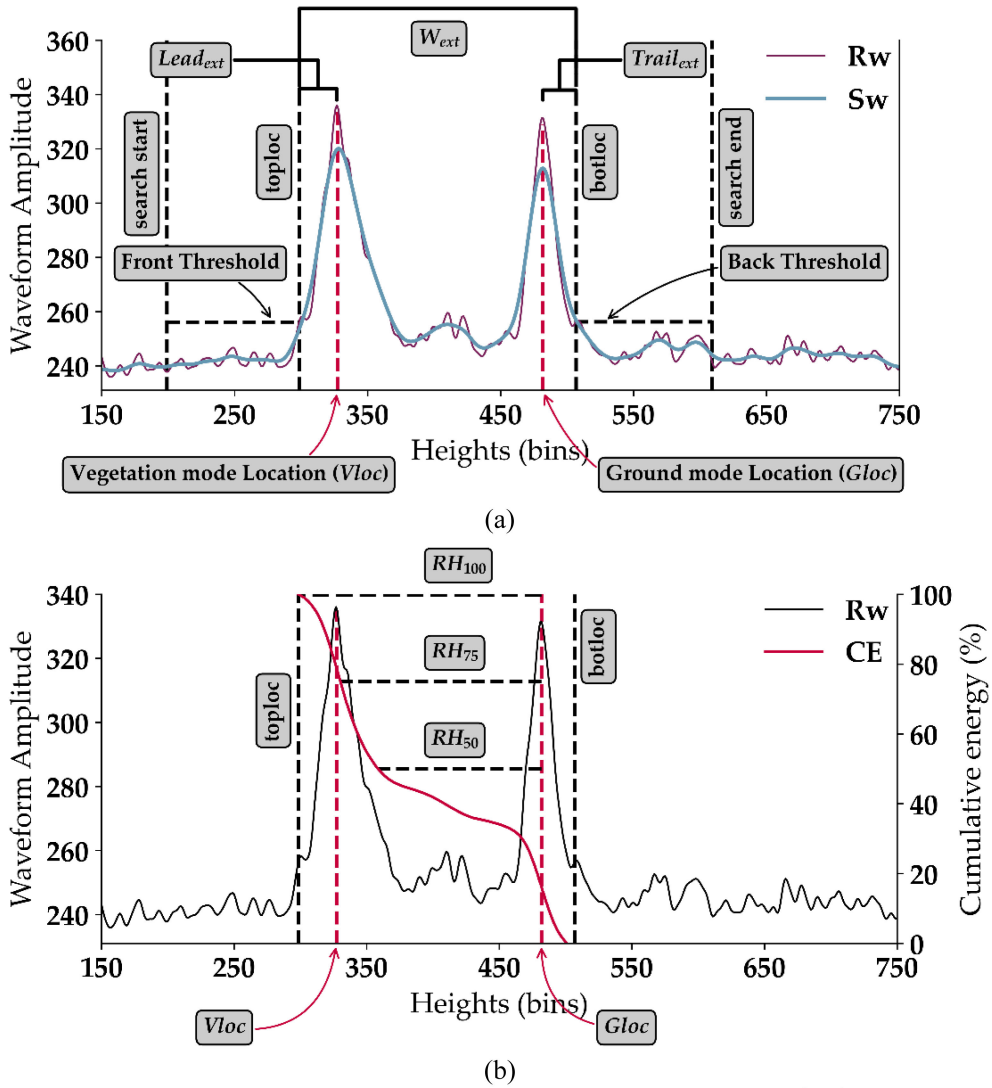


Fig. 2. (a) Example of an acquired GEDI waveform (Rw) over a Eucalyptus stand (H_{dom}= 25.9m; V= 230.7 $\mathrm{m}^3.\mathrm{ha}^{-1}$), its smoothing (Sw) and corresponding waveform metrics. (b) Cumulative energy of the waveform (CE) between *botloc* and *toploc* and the corresponding relative heights (RH_n) at different percentages "n" for the same waveform. One (1) ns corresponds to 15 cm sampling distance in the waveform. The waveform amplitudes are counts from the analog to digital converter on the instrument.

intensities are above a threshold. The threshold equation used to determine *toploc* and *botloc* is the same as (1), with "v" an integer fixed at 2, 3, 4, or 6 (depending on the processing configuration). In the ATBD, the value of "*v*" used to determine *toploc* is named "Front_threshold" and "back_threshold" for *botloc*. Waveform metric values are extracted using thresholds on Smoothwidth_zcross**,** front_threshold, and back_threshold. Currently, there are six configurations (henceforth referred to as algorithms) of different thresholds on these variables, which are used to determine waveform metrics with high precision for a variety of acquisition scenarios (see Table I). Finally, the location of the distinctive peaks or modes in the waveform, such as the ground peak, or top of canopy peaks is determined using a second Gaussian filtering of the waveform section between *toploc* and *botloc*, and then finding all the zero crossings of the first derivative of the filtered waveform [see Fig. 2(a)]. The width of the second Gaussian filter ("Smoothwidth_zcross") is fixed to either 3.5 or 6.5 ns (based on the algorithm used). Finally, the position of the ground return within the waveform is determined using the position of the last detected mode. The six different algorithms, noted a1 to a6 correspond to different values of the above-mentioned parameters (see Table I) and lead to different estimates of the waveform metrics, and could in turn lead to six different canopy height estimates. Over forest stands, the recorded waveforms are multimodal in shape, with each mode representing a reflection from a distinct surface height. Fig. 2(a) shows a typical waveform over a Eucalyptus forest stand on relatively flat terrain. Over flat terrain, the first Gaussian corresponds to a reflection from the top of the canopy while the last Gaussian mostly refers to the lowest point in the footprint, i.e., the ground surface.

GEDI data used in this article have been already processed and published by the land processes distributed active archive center (LP DAAC). Currently, three products (L1B, L2A, and L2B) are available for download. The L1B data product [32] contains detailed information about the transmitted and received

TABLE I
DIFFERENT THRESHOLDS USED IN EACH OF THE SIX ALGORITHMS FOR THE ANALYSIS OF THE RECEIVED WAVEFORMS

Algorithm	Smooth width	Smoothwidth_zcross	Front_ threshold	Back_ threshold
a1	6.5	6.5	3	6
a2	6.5	3.5	3	3
a3	6.5	3.5	3	6
a4	6.5	6.5	6	6
a5	6.5	3.5	3	2
a6	6.5	3.5	3	4

waveforms, the location and elevation of each waveform footprint and other ancillary information, such as mean and standard deviation of the noise and acquisition time. The L2A product [33] contains data of elevation and height metrics of the vertical structures within the waveform. These height metrics are issued from the processing of the received waveforms from the L1B product. Finally, the L2B data product [34] provides footprint-level vegetation metrics, such as canopy cover, vertical profile metrics, plant area index and foliage height diversity.

In this article, the received waveforms, their geolocation (longitude, and latitude), as well as their acquisition times were extracted from the L1B data product. In the L2A data product, the derived metrics are also grouped by algorithm. Therefore, for each beam, the metrics derived from each of the six algorithms, as well as the parameters used for each algorithm are available. Therefore, we extracted from L2A for each beam and for each of the six algorithms (a1 through a6), the following variables.

1) The position within the waveform of *toploc* and *botloc*.
2) The amplitude of the smoothed waveforms lowest detected mode (zcross_amp).
3) The quality flag for each waveform (quality_flag).
4) The number of detected modes (num_detectedmodes).
5) The position and amplitude of each detected mode.
6) The Relative height metrics at 10% intervals from *botloc* (0%) to *toploc* (100%) ($\mathrm{RH}_n,\ 10\% \leq n \leq 100\%,\ \mathrm{step}\ 10\%$). RH_n represents the height between *botloc* and the location at n% of cumulative energy [see Fig. 2(b)]. No metrics were extracted from the L2B product, as they were not relevant to this article.

2) Calculation of Relevant GEDI and Terrain variables: Several linear and non-linear regression models will be tested in order to estimate the stand dominant height H_{dom} (m) and stand merchantable wood volume *V* ($\mathrm{m}^3.\mathrm{ha}^{-1}$) from GEDI data. The models were tested with *a priori* variables that were extracted from GEDI waveforms. These variables represent canopy features, such as canopy top, canopy trunks, ground, or a mix of these elements. In addition to the available GEDI waveform metrics described in the previous section, several additional metrics were also extracted. The first is the waveform extent (W_{ext}) which is the height difference between *botloc* and *toploc*. Next, to remove the effects of canopy height variability and terrain slope, two indices relying on waveform structure were determined. The leading edge extent ($\mathrm{Lead}_{\mathrm{ext}}$) is, as defined by Hilbert and Schmullius [35], the difference between the position of the first mode (*Vloc*) and *toploc*, while the trailing edge extent ($\mathrm{Trail}_{\mathrm{ext}}$) is the difference between *botloc* and the ground return (*Gloc*) [35]. Two methods will be used to determine the ground position: the position of the mode selected as the lowest non-noise mode from the L2A data product, and the position of the highest mode between the last two detected modes. The viewing angle of GEDI, which represents the angle between the looking direction of the instrument and nadir at acquisition time, and for each shot, has also been calculated using the geolocation of the GEDI instrument available from the L1B data product. The viewing angle has been demonstrated in Urban *et al*. [36] to increase elevation errors for ICESat-1 GLAS when the viewing angle deviates from nadir due to precision attitude determination.

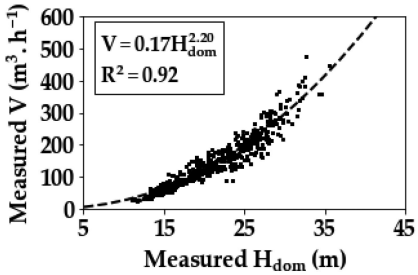


Fig. 3. Allometric relation between *in situ V* and H_{dom}.

Finally, as the wood volume (V) increases with canopy height in a nonlinear shape (see Fig. 3), we calculated RH_n for several power values ($\mathrm{RH}_n^p,\ 1 < p \leq 3, \mathrm{step}\ 0.2$).

All the used variables for the estimation of the stand dominant height H_{dom} (m) and stand merchantable wood volume *V* ($\mathrm{m}^3.\mathrm{ha}^{-1}$) are given in Table II.

The values of the extracted waveform metrics vary by the algorithm used for the processing of the waveforms. Therefore, in this article, all GEDI metrics were determined for the six available algorithms. The variability of the metric values based on the processing algorithm is given in Table III.

3) Filtering of GEDI Waveforms: Not all GEDI acquisitions are viable, as atmospheric conditions (e.g., clouds) can affect them. Therefore, a waveform was not investigated further if it met any of the following criteria.

1) Waveforms with reported elevations that are significantly higher or lower than the corresponding elevations from the SRTM DEM [16]. In essence, we removed all waveforms where the absolute difference is higher than 100 m.

TABLE II
LIST OF ALL THE VARIABLES CALCULATED FROM GEDI WAVEFORMS

Variable	Unit	Abbreviation
Search Start	bin	-
Search End	bin	-
Vegetation mode location	bin	$Vloc$
Ground mode location	bin	$Gloc$
Sginal start (Top of canopy)	bin	*toploc*
Signal end	bin	*botloc*
Waveform extent	m	W_{ext}
Trailing edge extent [44]	m	$Trail_{ext}$
Leading edge extent [44]	m	$Lead_{ext}$
Relative canopy height (*botloc* (0%) to *toploc* (100%)) at n% of cumulative energy	m	$(RH_n, 10\% \leq n \leq 100\%, step\ 10\%)$
3x3 pixels terrain index	m	T_I
3x3 pixels terrain roughness	m	$ROUG$
Terrain slope	%	S
Viewing angle	°	VA

Variables to be used as predictor variables in the canopy height and wood volume estimation models are highlighted in gray.

TABLE III
MEAN AND STANDARD DEVIATION OF SOME GEDI METRIC VALUES FROM EACH OF THE SIX PROCESSING ALGORITHMS USING ALL GEDI SHOTS OVER THE 566 SELECTED EUCALYPTUS PLANTATIONS

Variable	Unit	Algorithm											
		a1		a2		a3		a4		a5		a6	
		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
W_{ext}	(m)	25.00	5.50	28.74	6.66	25.00	5.50	23.91	6.01	32.99	8.70	26.78	5.91
$Lead_{ext}$	(m)	4.89	2.23	4.34	1.08	4.61	2.05	3.80	1.01	4.34	1.08	4.38	1.21
$Trail_{ext}$	(m)	4.24	1.60	6.40	2.84	3.86	1.65	4.24	1.60	1.93	1.89	3.45	2.15
RH_{100}	(m)	20.73	5.32	22.31	5.97	21.11	5.40	19.64	5.72	31.02	9.43	23.30	6.42

TABLE IV
DISTRIBUTION OF GEDI SHOTS ACROSS THE FOUR STUDY REGIONS

Site	GEDI count	*in situ* H_{dom} (m)	*in situ* V ($m^3.ha^{-1}$)
BA&ES	1220	25.7	207.9
MA	808	27.2	208.0
MS	2586	29.4	300.0
SP	1068	32.4	380.6

in situ H_{dom} and *in situ* V represent the 95th percentile of *in situ* values for each site.

2) Waveforms with a difference between waveform extent (W_{ext}) and (Gloc–Vloc) higher than 400 bins (corresponding to 60 m)

A total of 6166 footprints were acquired over our reference stands between April 2019 and September 2019, with the majority of these footprints (92.15%) providing exploitable waveforms. Table IV gives the distribution of GEDI shots across the four regions.

GEDI data accessible through NASA's LP DAAC contain a quality flag (quality_flag) for each acquired waveform. A waveform with a quality flag set to "1" indicates that the waveform meets certain criteria based on energy, sensitivity, amplitude, and real-time surface tracking quality, and thus can be processed further. However, in this article waveforms with either value of the quality_flag were analyzed.

C. Inventory Measurements

A total of 566 Eucalyptus stands were selected, corresponding to stands where GEDI footprints acquired between April 20, 2019 and September 4, 2019 were totally included. An additional 50 m internal buffer strip from the stand borders was used to account for any footprint geolocation errors and to avoid footprints that match the boundary between the stand of interest and the surrounding medium. These 566 Eucalyptus stands were also selected because they had field inventories performed by the company close to GEDI's acquisition date (time difference fewer than two months). Field inventories are performed on several permanent inventory plots within each stand. These inventory plots are systematically distributed throughout the stand with a density of one plot per 10 ha (i.e., a 20 ha stand will have two inventory plots while an 80 ha plot will contain eight inventory plots). These permanent inventory plots had each an area of approximately 400 m^2 including 30 to 100 trees (average of 58

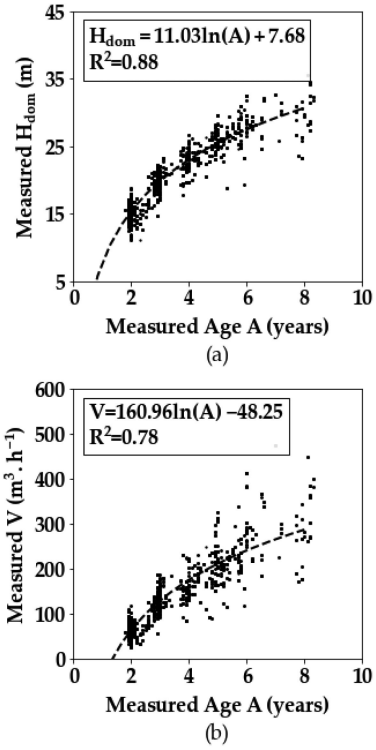


Fig. 4. Relationship between stand age and (a) H_{dom} (m) and (b) Volume (m^3.ha^{-1}).

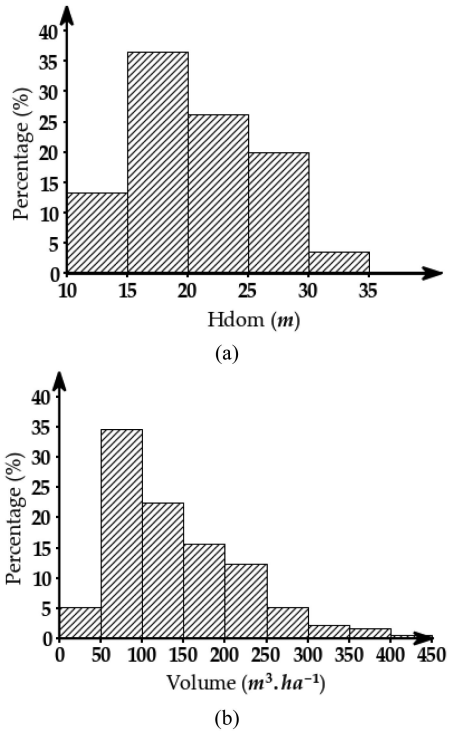


Fig. 5. Distribution of measurements of (a) dominant canopy heights and (b) wood volume from field inventories of the 566 Eucalyptus stands. The *Y*-axis represents the percentage of samples within each height range (a) and wood volume (b).

trees) in function of the inventory plot size and planting density. During a field inventory, the diameter at breast height (DBH, 1.3 m above the ground) of each tree in the inventory plot, the height of a central subsample of 10 trees and the height of the 10% largest trees in terms of DBH (dominant trees) were measured. The mean height of the 10% of the largest trees defined the dominant height of the plot (H_{dom}), while the mean height of all trees in the plot (measured + estimated) defined the average height of the plot (H_{mean}). H_{dom}, basal area and age on the inventory date were then used in local volume equations to estimate the plot total and merchantable volume (merchantable volume is the tree volume up to the diameter outside bark of 6 cm). Stem biomass was then estimated from the stem volume using age-dependent estimates of wood biomass density.

As the dates of the inventory measurements were different from GEDI acquisition dates, only data with a difference fewer than two months in the date between GEDI acquisitions and inventory were used. In fact, on these fast growing plantations, a two-month difference could result in an up to 50 cm growth in H_{dom} [see Fig. 4(a)] and 10 m3.ha^{-1} in *V* [see Fig. 4(b)]. However, this reasonable compromise allows keeping a large number of stands including a large variability of age and growing conditions. Fig. 5 shows the distribution of field measured H_{dom} and volume.

D. Digital Elevation Model

The DEM with a spatial resolution of 30 m, derived from the Shuttle Radar Topography Mission (SRTM), was used in this article. Three variables were derived from the DEM: slope (S), Terrain Index (T_I), and surface Roughness (Roug). The T_I map was obtained by calculating the difference between the highest and lowest altitude in a 3 × 3 pixel-moving window. The surface roughness map was obtained by computing the standard deviation of the elevation in a 3 × 3 pixel-moving window.

III. Methodology

A. Forest Height Estimation

The simplest method to estimate H_{dom} from a GEDI waveform over forest stands with a gently sloping terrain is the height difference between signal start (*toploc*) and the ground position (*Gloc*) [37]

$$H_{\mathrm{dom}} = \mathrm{toploc} - \mathrm{Gloc}. \tag{2}$$

For the previous ICESat-1 GLAS waveforms, the ground return was assumed to be the stronger of the last two detected modes [15]. Therefore, in this article, the stronger between the last two detected modes, as well as the position of the mode in the field "selected_mode (SM)" from the L2A data product will be considered as the ground return separately.

Estimating canopy heights using (2) has several caveats. For example, over sloping terrain, the ground peak becomes wider, and the returns from ground and vegetation can be mixed in the case of large footprints, making the identification of the ground peak return difficult and the estimation of forest height inaccurate ([5], [37]). To remove or minimize the terrain slope effect on the waveforms, as well as the vegetation variability,

statistical approaches have been developed and used in several studies to predict canopy heights from GLAS data (e.g., [5], [15], [35], [38], [39]). These approaches proposed regression models based either on only waveform metrics or on both waveform metrics and terrain information derived from DEMs.

The first statistical model was developed by Lefsky *et al.* [5] to estimate the maximum canopy height (H_{dom}) from GLAS waveforms

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - bT_I. \tag{3}$$

The coefficients *a* and *b* are fitted using least squares regression (H_{dom} given by inventory measurements, W_{ext} is derived from the GEDI waveform, and T_I is calculated from the SRTM DEM, see Section II-D). For our dataset, T_I values calculated from the SRTM DEM ranged from 1 to 46 m. The incorporation by Payn *et al.* [6] of the waveform leading edge extent in (4) showed a slight improvement on canopy height estimation

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - bT_I + c\mathrm{Lead}_{\mathrm{ext}}. \tag{4}$$

Over sloping terrain, Lefsky *et al.* [38] observed that the waveform extent is insufficient for estimating canopy heights. Hence, a new model based on the waveform extent, leading edge extent, and trailing edge extent was proposed. However, Pang *et al*. [39] observed inaccurate estimates of canopy heights with the improved model by Lefsky *et al*. [38], especially for small waveform extents, and thus proposed a simpler model to estimate canopy heights using the following equation:

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - \{b\left(\mathrm{Lead}_{\mathrm{ext}} + \mathrm{Trail}_{\mathrm{ext}}\right)\}^c. \tag{5}$$

The nonlinear model by Pang *et al.* [39] was further simplified by Chen [15]

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - b\left(\mathrm{Lead}_{\mathrm{ext}} + \mathrm{Trail}_{\mathrm{ext}}\right) \tag{6}$$

Baghdadi *et al.* [16] tested additional models for the estimation of canopy heights using ICESat-1 GLAS waveforms, of which, two will be tested in this article. The first model uses the $\mathrm{Trail}_{\mathrm{ext}}$ and T_I

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - bT_I + c\mathrm{Trail}_{\mathrm{ext}}. \tag{7}$$

The second model uses exclusively GEDI metrics

$$H_{\mathrm{dom}} = aW_{\mathrm{ext}} - b\mathrm{Lead}_{\mathrm{ext}} - c\mathrm{Trail}_{\mathrm{ext}} + d. \tag{8}$$

In addition to the previously described models, a stepwise multilinear regression aimed at estimating canopy dominant heights (SRH) and automatically choosing a set of predictive explanatory variables among all possible variables presented in Table II. The choice of adding or removing a variable from the SRH model is based on the increase or decrease of the mean squared error (MSE).

We also estimated canopy dominant heights through nonlinear nonparametric regressions by means of a random forest regressor (RFH). Random forests are an ensemble of machine learning algorithms used for classification or regressing by fitting a number of decision trees on various sub-samples of the dataset, and use averaging to improve the predictive accuracy and control overfitting [40]. Compared to linear models, RF is advantageous for being able to model also nonlinear relationships (threshold effect) between the explanatory variables. For this article, the number of trees in the RF were set to 100 trees (higher tree count slightly increased model accuracy), with a tree depth equal to the square root of the number of available factors.

Finally, since random forests are nonlinear and nonparametric, we only used the original relative heights without modification (i.e., $\mathrm{RH}^1_n,\ 10\% \leq n \leq 100\%, \mathrm{step}\ 10\%$.)

B. Wood Volume Estimation

The estimation of aboveground biomass has been proven to be successful using ICESat-1 GLAS waveforms as demonstrated by several studies ([15]–[17]). In this article, four models were tested to estimate wood volume from GEDI waveforms based on Hdom estimates. The first model was adapted from Lefsky *et al.* [5] for the estimation of wood volume (instead of AGB in its original formulation), using the squared dominant canopy heights (H_{dom})

$$V = a + bH^2_{\mathrm{dom}}. \tag{9}$$

The second tested model was adapted from Saatchi *et al.* [41], and uses a power law relationship between the volume and Lorey's height

$$V = aH_L{}^b \tag{10}$$

where H_L is Lorey's height which weighs the contribution of trees (all trees >10 cm in diameter) to the stand height by their basal area. In this article, the relationship defined in (10) was used by replacing Lorey's height with the dominant height as both height values were similar (H_L was lower than H_{dom} by a maximum of 0.9 m at the end of the rotation of the Eucalyptus plantation) [16]. For both models (9 and 10), the coefficients a and b were first fitted using *in situ* measurements of dominant height and wood volume (see Fig. 6), and then, the calibrated equations were used to estimate wood volume using the dominant height predicted from GEDI footprints (best model from Section III-A).

Similarly to Section III-A, a stepwise linear regression model (SRV) and a random forest regressor (RFV) were used to estimate the wood volume.

C. Model Assessment

To assess how the tested models generalize to an independent data set, a five-fold cross validation was used. Large k-fold values mean less bias towards overestimating the true expected error (as training folds will be closer to the total dataset). Moreover, since there are several GEDI footprints inside each stand, and the stands are very homogeneous, the five-fold splitting was also done along the stands in order to reduce fitting bias. In essence, GEDI footprints inside the same stand were used for either training or validation. Finally, models' performance were assessed using the coefficient of determination (R^2), the bias (measured—estimated), the root-mean-square error (RMSE), the root mean squared percentage error (RMSPE), and the Akaike information criterion (AIC). R^2, RMSE, and RMSPE

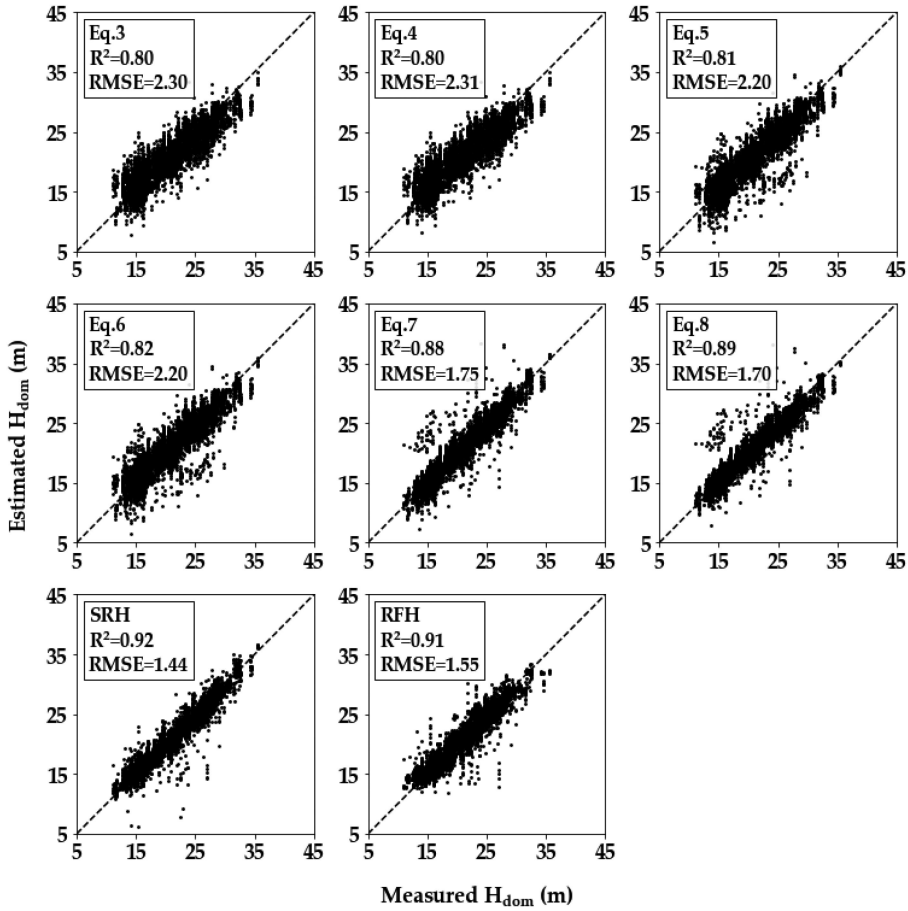


Fig. 6. Comparison of measured vs. estimated H_{dom} from the models presented in Section III-A using GEDI metrics extracted with algorithm a1 (see Table I). RMSE is expressed in meters (m).

are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \widehat{y_i})^2} \tag{12}$$

$$\mathrm{RMSPE} = 100 \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{y_i - \widehat{y_i}}{y_i} \right)^2} \tag{13}$$

where y_i is the observed value, $\widehat{y_i}$ the estimated value, $\bar{y}$ is the mean of all the observed values, and *n* is the sample size.

The AIC proposed by Akaike [42] is a measurement of the relative goodness of fit of a statistical model to the true values. By calculating AIC values for each model, the most performant model based on the lowest AIC values can be identified.

IV. Results

A. Canopy Height Estimation

We start our model performance analysis using GEDI metrics extracted from algorithm a1 (see Table I), and the ground location as determined from the SM field from the L2A dataset (last detected nonnoise mode). The estimation of the canopy dominant heights (H_{dom}) using the linear regression models [(3) through 8] with five-fold cross validation shows an accuracy (RMSE) between 1.70 and 2.31 m with a coefficient of determination (R^2) between 0.80 and 0.89 (see Fig. 6). Moreover, the contribution of the trailing edge extent appeared to be higher than that of the leading edge extent [see (7) versus (4), Table V]. However, the best model between (3) through (8) was (8) (RMSE = 1.70 m and R^2 = 0.89) which uses both Leading and Trailing edge extents, with an independent coefficient fitted for each variable. The introduction of terrain information in the linear regression models did not show any significant improvements on the accuracy of the estimations.

The stepwise linear regression model (see Fig. 6, SRH) showed slightly better accuracy for the estimation of canopy heights (RMSE = 1.44 m, R^2 = 0.93) in comparison to Eq.8. However, unlike Eq.8 which relied on W_{ext} , $\mathrm{Lead}_{\mathrm{ext}}$, and $\mathrm{Trail}_{\mathrm{ext}}$, the most contributing variables for the estimation of the canopy heights using the SRH model were RH_{90}, followed by RH_{10}, RH_{80}, and RH_{100}. Meanwhile, the other metrics (e.g., $\mathrm{Lead}_{\mathrm{ext}}$, $\mathrm{Trail}_{\mathrm{ext}}$, TI, etc.,) were not necessary.

Furthermore the estimation of canopy heights using only RH_{90} (by linear fitting) showed an RMSE of 1.63 m with an R^2 of 0.90, and this accuracy could be improved to an RMSE of 1.5 m (R^2 of 0.91) by only adding $\mathrm{RH}_{10}^{1.8}$. The estimation of

TABLE V
MODELS' PERFORMANCE AND THE FITTED LINEAR EQUATIONS FOR ESTIMATING EUCALYPTUS STAND DOMINANT HEIGHTS

Model	RMSE (m)	RMSPE (%)	Bias (cm)	R^2	AIC
$EQ.3$: $H_{dom} = 0.83W_{ext} - 0.02T_l$	2.30	12.4	1.7	0.80	9217.1
$EQ.4$: $H_{dom} = 0.82W_{ext} + 0.03T_l - 0.03Lead_{ext}$	2.31	12.5	1.9	0.80	9254.6
$EQ.5$: $H_{dom} = 0.91W_{ext} - \{0.26(Lead_{ext} + Trail_{ext})\}^{0.90}$	2.20	11.6	10.1	0.81	8710.3
$EQ.6$: $H_{dom} = 0.90W_{ext} - 0.21(Lead_{ext} + Trail_{ext})$	2.20	11.6	7.3	0.82	8683.9
$EQ.7$: $H_{dom} = 0.97W_{ext} + 0.041T_l + 0.90Trail_{ext}$	1.75	9.5	8.0	0.88	6174.2
$EQ.8$: $H_{dom} = 0.91W_{ext} + 0.06Lead_{ext} - 0.98Trail_{ext} + 1.83$	1.70	9.2	1.7	0.89	5858.0
SRH: $H_{dom} = 0.64RH_{90}^{1.2} - 0.18RH_{10}^{1.8} - 0.75RH_{80} + 0.9RH_{100} + 6.42$	1.44	7.1	0.4	0.92	4245.1
RFH	1.55	7.8	3.1	0.91	4887.6

The variables are described in Section II-B-2, with the models described in Section III-A

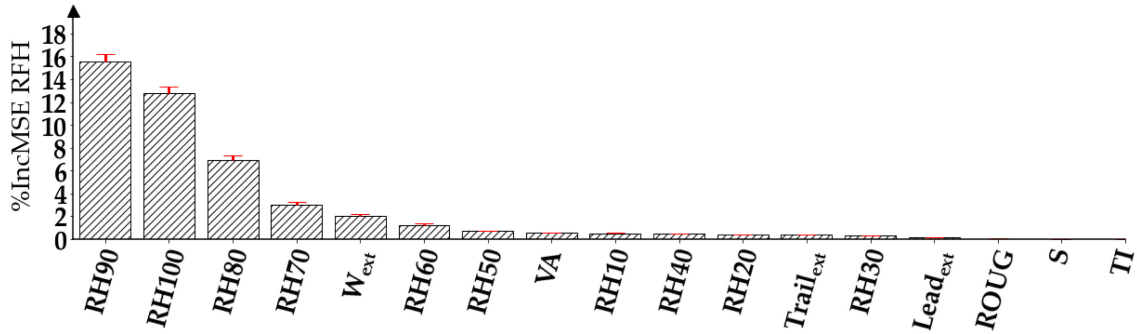


Fig. 7. Classification of the variable importance by decreasing order of importance in the RFH model for stand dominant height estimation. The importance is measured via the average percentage increase of MSE (%IncMSE) over 50 repetitions. The red bars indicates the standard deviation of %IncMSE.

TABLE VI
ACCURACY (RMSE IN M) OF THE MODELS PRESENTED IN SECTION III-A FOR THE ESTIMATION OF H_{dom} USING GEDI METRIC VALUES EXTRACTED USING THE SIX DIFFERENT ALGORITHMS (A1 THROUGH A6)

Model	Algorithm											
	a1		a2		a3		a4		a5		a6	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Eq.3	2.30	0.80	3.82	0.44	2.53	0.75	3.35	0.57	4.96	0.06	3.09	0.64
Eq.4	2.31	0.80	3.77	0.46	2.54	0.75	3.32	0.58	5.01	0.04	3.00	0.66
Eq.5	2.20	0.81	3.32	0.58	2.47	0.77	3.11	0.63	5.06	0.02	3.07	0.64
Eq.6	2.20	0.82	3.52	0.52	2.46	0.77	3.24	0.60	4.69	0.16	3.08	0.64
Eq.7	1.75	0.88	3.50	0.53	2.40	0.78	3.13	0.63	4.49	0.23	2.98	0.66
Eq.8	1.70	0.89	3.08	0.64	2.50	0.76	2.67	0.73	4.06	0.37	2.79	0.70
SRH	1.44	0.92	1.46	0.92	1.48	0.92	1.60	0.90	1.67	0.88	1.49	0.92
RFH	1.55	0.91	1.58	0.90	1.72	0.89	1.75	0.88	1.80	0.88	1.61	0.90

H_{dom} using the random forest regressor (RFH, Fig. 6) with the GEDI metrics in Table II (p in RH_n^p was set to 1 for RFH) as the dependent variables showed an accuracy on the canopy height estimates similar to that of the SRH model.

The variable importance test of the metrics (see Fig. 7) used in RFH showed that the most contributing factors for the estimation of GEDI canopy heights is a combination of RH_{90}, RH_{100}, and to lesser extent RH_{80}. These results show that in a low relief area the use of other metrics in addition to the RH_{90} only slightly improved the precision of the estimation of canopy heights.

The estimation of H_{dom} using the models described previously with GEDI metrics extracted from the five remaining algorithms (a2 to a6, see Table I) has been also tested. The results presented in Table VI show that for the linear regression models [(3) through (8)], canopy height estimation was worst with the metrics from algorithms a2 through a6 in comparison to the metrics from algorithm a1 with an RMSE on the canopy height estimates ranging from 2.40 m (R^2 of 0.78, a3) to 5.06 m (R^2 of 0.02, a5). This is to be expected given the low terrain relief in our study area (mean slope of $4.7 \pm 3\%$).

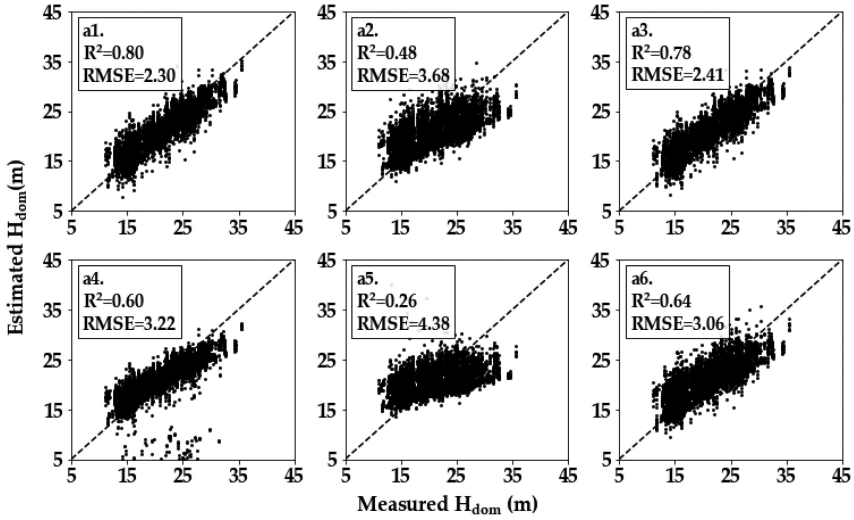


Fig. 8. Comparison between Measured H_{dom} and estimated H_{dom} using only W_{ext} values ($H_{\mathrm{dom}} = \alpha.W_{\mathrm{ext}} + \beta$) from the six algorithms (a1 through a6). RMSE is expressed in meters (m).

The low accuracy obtained with algorithm a5 is due to the low thresholds used for the front and back thresholds ($3.\sigma$ and $2.\sigma$, Table I), which result in larger waveform extents. This is evident when trying to estimate canopy heights based solely on the waveform extent ($H_{\mathrm{dom/insitu}} = \alpha.W_{\mathrm{ext}} + \beta$), with the results in Fig. 8 showing that the metrics extracted using algorithm a5, especially the waveform extent (W_{ext}), were the least correlated to H_{dom} , with an RMSE of 4.38 m (R^2 of 0.26).

In contrast to the linear regression models, canopy height estimation using SRH or RFH with metrics from algorithms a2, a3, a5, and a6 showed accuracies similar to those obtained with algorithm 1 (see Table VI). In contrast, algorithm a5 was slightly less accurate with an RMSE of respectively 1.6 m (R^2 of 0.90) and 1.80 m (R^2 of 0.88) when using SRH and RFH.

Finally, the effects of the method to select the ground return has been studied. The results presented thus far have been based on detecting the ground mode from the SM provided in the L2A data product. SM detects the ground return as being the lowest, nonnoisy mode, which usually refers to the last detected mode. Previous studies that used GLAS waveforms suggested that the mode with the higher amplitude between the last two modes (HL2M) is a better indicator of the ground return [15], [16]. In this article, H_{dom} estimation was also tested using the same models described in Section III-A, with the metrics calculated relatively to the mode with the higher amplitude between the last two modes, as the ground return. The results in Table VII show that the models relying mostly on the relative canopy heights (RH_n), such as SRH, or the trailing edge extent fitted separately [(7) and (8)] had an increase in accuracy on the canopy height estimates between 12 and 20 cm.

TABLE VII
DIFFERENCE IN ACCURACY (RMSE IN M) ON H_{dom} BASED ON THE CHOICE OF THE SELECTED GROUND MODE FOR THE DIFFERENT MODELS DESCRIBED IN SECTION III-A, AND METRICS EXTRACTED USING ALGORITHM A1

Model	Ground Mode Selection			
	SM		HL2M	
	RMSE	R^2	RMSE	R^2
Eq.3	2.30	0.80	2.30	0.80
Eq.4	2.31	0.80	2.31	0.80
Eq.5	2.20	0.81	2.15	0.82
Eq.6	2.20	0.82	2.14	0.82
Eq.7	1.75	0.88	**1.57**	**0.91**
Eq.8	1.70	0.89	**1.50**	**0.91**
SRH	1.44	0.92	**1.33**	**0.93**
RFH	1.55	0.91	1.50	0.91

SM = Ground mode from SM (last detected mode) provided in the L2A data product. HL2M = ground mode corresponding to the higher amplitude between the last two modes.

B. Wood Volume Estimation

Four models were used to estimate the stand volume V. Two power functions as presented in Section III-B [(9) and (10)], a stepwise multilinear regressing model (SRV) and a random forest based model (RFV). The coefficients of the models as presented in (9) and (10) were fitted using *in situ V* as the estimated variable and *in situ* H_{dom} as the predictor on the 566 studied stands. Next, to estimate *V*, the fitted models in (9) and (10) used the estimated H_{dom} values from SRH with GEDI metrics extracted using algorithm a1, while SRV and RFV used the GEDI waveform metrics from Table II extracted with algorithm a1 (p in RH_n^p was set to 1 for RFV). The estimation results of stand volume V (see Fig. 9, Table VIII) show that the four tested models produced similar accuracies on the estimation of *V*, with an RMSE between 24.39 and 27.45 $\mathrm{m}^3.\mathrm{ha}^{-1}$ and coefficient of determination (R^2) between 0.87 and 0.90. Moreover, the results in Fig. 9 also show that the estimations of V were close to the 1:1 for all values of V between 0 and 250 $\mathrm{m}^3.\mathrm{ha}^{-1}$, while they under estimated the volume for V values higher than 250 $\mathrm{m}^3.\mathrm{ha}^{-1}$. For

TABLE VIII
MODELS' PERFORMANCE AND THE FITTED LINEAR EQUATIONS FOR ESTIMATING EUCALYPTUS STAND WOOD VOLUME (V)

Model	RMSE ($m^3.ha^{-1}$)	RMSPE (%)	Bias ($m^3.ha^{-1}$)	R^2	AIC
$Eq.9: V = -12.44 + 0.34{H_{dom}}^{2.2}$	28.19	21.8	0.64	0.87	36844.2
$\text{Eq.10}: V = 0.17{H_{dom}}^{2.2}$	28.04	21.9	0.96	0.87	36786.8
$SRV: V = 0.3RH_{90}^{2.2} - 0.34RH_{10}^{2.8} + 0.02RH_{30}^{2.2} - 0.15RH_{80}^{2.2} + 0.03\text{RH}_{20}^{2.6}$	24.39	20.4	-0.03	0.90	35467.8
RFV	27.48	21.3	0.65	0.88	36789.0

The variables are described in section II.B.2, with the models described in section III.B

TABLE IX
ACCURACY (RMSE IN $M^3.HA^{-1}$) OF THE MODELS PRESENTED IN SECTION III-B FOR THE ESTIMATION OF V USING GEDI METRIC VALUES EXTRACTED USING THE SIX DIFFERENT ALGORITHMS (A1 THROUGH A6)

Model	Algorithm											
	a1		a2		a3		a4		a5		a6	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Eq. 9	28.19	0.87	28.72	0.87	28.59	0.87	29.79	0.86	29.83	0.86	28.12	0.87
Eq.10	28.04	0.87	28.53	0.87	28.42	0.87	29.69	0.86	29.82	0.86	27.92	0.88
SRV	24.39	0.90	27.25	0.88	26.32	0.89	27.66	0.88	28.96	0.87	26.85	0.88
RFV	27.48	0.88	29.01	0.87	31.59	0.84	30.87	0.85	33.36	0.82	29.65	0.86

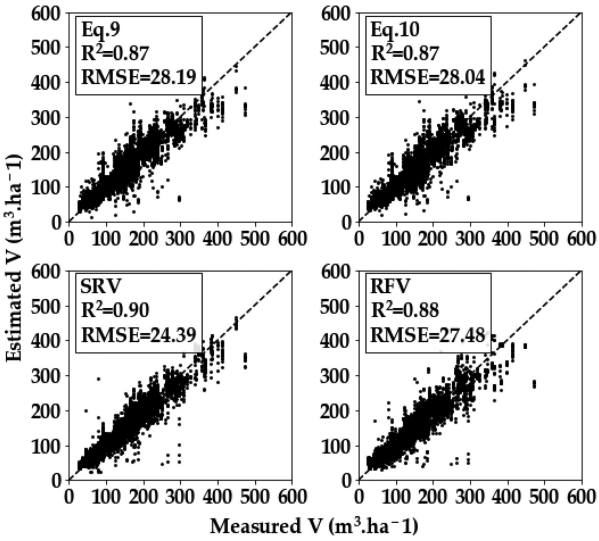


Fig. 9. Comparison of measured vs. estimated wood volume from the models presented in section III.B using GEDI metrics extracted with algorithm a1. RMSE is expressed in $m^3.ha^{-1}$.

the four models, the relative RMSE increased from ∼18% for V less than 250 $m^3.ha^{-1}$ to ∼40% for V higher than 250 $m^3.ha^{-1}$. The bias (mean difference of *in situ* V and estimated V) for V higher than 250 $m^3.ha^{-1}$ was also more apparent, and decreased from 2.2 $m^3.ha^{-1}$ (average bias from all models) for V less than 250 $m^3.ha^{-1}$ to 26.5 $m^3.ha^{-1}$ for V higher than 250 $m^3.ha^{-1}$.

The variable importance test of the GEDI metrics (see Fig. 10) showed that the three most contributing factors on the estimation of V using the random forest regressor (RFV) were the same as those for the estimation of canopy heights, with the highest contributor being RH_{90}, followed by RH_{80} and RH_{100}.

The estimation of V was also tested using GEDI metric values extracted using the remaining five algorithms (a2 through a6). The results presented in Table IX show that the estimates of V were mostly similar with (9) and (10) across all algorithms. On the other hand, SRV and RFV show less accurate estimates of V using GEDI metrics from algorithms a2 through a6 compared to a1.

Finally, the choice of the method for detecting the ground return was also studied. The results, unlike those obtained when estimating H_{dom}, did not show any significant improvements when considering the last detected mode, or the stronger between the last two modes.

V. DISCUSSION

The different tested models in this article showed that GEDI waveform metrics could be used to obtain good accuracies of canopy heights and wood volumes, with a RMSPE of 7.1% on canopy height estimation and 20.4% on wood volume estimation. Moreover, GEDI waveforms appear to be of high quality given the very little variability on the estimation of H_{dom} and V from the individual footprints within a given stand. Indeed, the accuracy (RMSE) on the estimation of H_{dom} using the mean estimates from SRH of the individual footprints was 1.33 m (R^2 of 0.93) versus 1.32 m (R^2 of 0.93) when averaging H_{dom} estimates from GEDI for each stand. Similarly, the accuracy on the estimates of V using the mean estimates from SRV was 24.39 $m^3.ha^{-1}$ (R^2 of 0.90) and 23.93 $m^3.ha^{-1}$ (R^2 of 0.91) for the average of V from GEDI over each stand.

The most important GEDI variable for the estimation of H_{dom} and V is RH_{90}, which explained respectively more than 87% and 84% of the variability of H_{dom} and V. Some of the remaining variability are explained by different GEDI metrics based on the

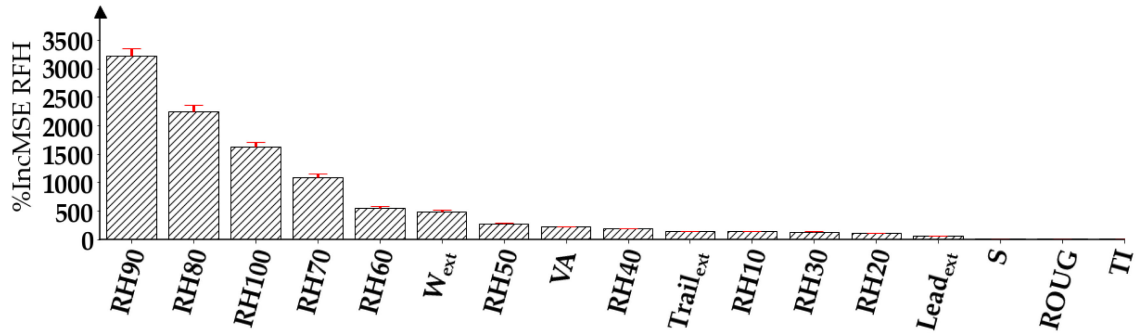


Fig. 10. Classification of the variable importance by decreasing order of importance in the RFV model for stand wood volume estimation. The importance is measured via the average percentage increase of MSE (%IncMSE) over 50 repetitions. The red bars indicates the standard deviation of %IncMSE.

model used. In the case of the stepwise linear regression models, RH_{10}, RH_{80}, and RH_{100} are used for the estimation of H_{dom} and RH_{10}, RH_{30}, RH_{80}, and RH_{20} for the estimation of V.

The success of the models in estimating H_{dom} and V with good accuracy from GEDI metrics relies heavily on the accuracy of extracting such metrics from the raw waveform data, and on the precision of the field measurements. GEDI datasets provide metrics issued from six algorithms, and the same metric can differ in value from one algorithm to another. This was evident from our results where the accuracy (RMSE) in estimating both H_{dom} and V was slightly less accurate with GEDI metrics generated from algorithms a2 through a6 in comparison to a1. Therefore, the choice of algorithm of which the metrics are calculated from, is important. Nonetheless, GEDI L2A datasets already provide a field for each footprint called "selected_algorithm" that recognizes the algorithm selected as identifying the lowest mode (last detected mode) with less noise. This field could be used as an indicator of the algorithm that provides the most accurate metrics. Indeed, for our dataset, the best accuracy on both H_{dom} and V was observed with metrics extracted using algorithm a1, which was also the algorithm suggested by the "selected_algorithm" for more than 99% of the studied footprints.

Another variable that also affects the accuracy of the extracted metrics, such as RH_n, $\mathrm{Trail}_{\mathrm{ext}}$, or $\mathrm{Lead}_{\mathrm{ext}}$ is the location of the ground return. In this article, the ground return was identified either by the SM field in the L2a dataset, or by identifying the mode with the higher amplitude between the last two detected modes. Our results, which were obtained over study sites with mostly homogenous canopy cover and flat terrain, indicated that the second method improved the estimation accuracy (RMSE) on H_{dom} by up to 20 cm. On the other hand, for the estimation of V, the choice of ground return did not have any effect on the accuracy. However, for highly dense vegetated areas as in the case of tropical forests, where the ground return is not easily identifiable, choosing the strongest mode between the last two as the ground return should provide better accuracies on H_{dom} and V [15], [16].

Some of the uncertainties on the estimation of H_{dom} and V can be attributed to some biophysical properties of the canopy that could not be quantified using GEDI data alone, or due to instrumental factors. To understand the effect of these biophysical properties, additional parameters could therefore be required.

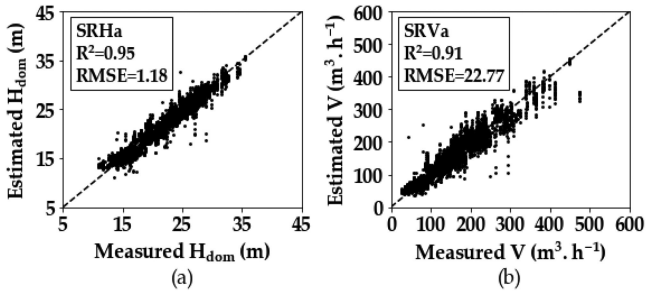


Fig. 11. The effects of adding stand age (*in situ* information) as a predictor variable on the estimation of (a) stand dominant heights H_{dom} (SRHa, m), and (b) stand volume V (SRVa, $\mathrm{m}^3.\mathrm{ha}^{-1}$) in a stepwise regression model.

Moreover, since GEDI data were acquired up to two months before or after *in situ* measurements, stand growth that happened between this time gap added to the uncertainty of our estimations (up to 50 cm difference). Canopy age plays an important role in understanding the variability of both H_{dom} and V. Fig. 4(a) and (b) show that both H_{dom} and V are a log function of stand age. In fact, adding the log transform of stand age to the stepwise linear regression models enhances the accuracy (RMSE) on H_{dom} by $\sim$13% (RMSE of 1.33 m without age and 1.18 m with age) and the volume estimation by $\sim$7% (24.39 versus 22.77 $\mathrm{m}^3.\mathrm{ha}^{-1}$ with age) (see Fig. 11). Moreover, the addition of the age of canopies to the SRH and SRV models also helps reduce the difference between estimates and measurements of some outliers points (see Fig. 11). The interest of using the stand age has also been demonstrated by Le Maire *et al.* [43] in their study over Eucalyptus plantations with MODIS optical data. Other properties of the canopy, such as crown dimension, leaf area index, distribution of leaf angles, tree gaps due to mortality, could also add variability to GEDI waveforms. Therefore, by adequately filtering the stand, one could obtain more accurate estimates of H_{dom} and V. On the opposite, the results presented in this article are specific to Eucalyptus plantations, since most sources of canopy variations are included.

The instrumental factors affecting the estimation accuracy include the viewing angle of GEDI at acquisition time. In fact, GEDI acquires data along eight beams with varying viewing angles (VA), which could affect the estimation accuracy. In this article, the analysis of estimation accuracy on both H_{dom} and V according to the acquiring beam (see Table X) shows that

TABLE X
COMPARISON OF THE ESTIMATION ACCURACY ON H_{dom} AND V BASED ON THE ACQUIRING BEAM

Beam	H_{dom} (m)		V ($m^3.ha^{-1}$)	
	RMSE	R^2	RMSE	R^2
1	1.34	0.92	22.92	0.91
2	1.41	0.93	25.99	0.91
3	1.22	0.95	21.92	0.91
4	1.28	0.94	22.08	0.90
5	1.68	0.90	30.65	0.86
6	1.39	0.93	22.73	0.91
7	1.42	0.91	25.30	0.91
8	1.25	0.93	22.50	0.91

Estimation results are produced using stepwise linear regression models with GEDI metric values extracted using algorithm a1.

the most accurate data were from beams 8, 3 and 4, while the least accurate H_{dom} and *V* estimates were obtained using data from beam 5. Indeed, the difference between the most accurate estimates (beam 3) and the least accurate (beam 5) is 46 cm for H_{dom} and of 9 $m^3.ha^{-1}$ for V. A preliminary analysis of VA values shows that beam 5 acquisitions had on average higher VA values.

A part of the uncertainties on the estimation of H_{dom} and *V* can be attributed to the heterogeneity of the Eucalyptus stands. In fact, the present study compared GEDI acquisitions with stand-scale averaged H_{dom} and V obtained from 1 to up to 10 permanent inventory plots. These *in situ* measurements are therefore the result of only a few observations within the stand. The hypothesis that the stands are homogeneous enough for using stand-scale averages is sometimes challenged. Indeed, some stands have high variability between their inventory plots, which reaches more than 6 m in some cases. A more precise analysis of intra-stand variability of GEDI waveforms variations could help determine which Eucalyptus stand could be compared to stand-scale values of H_{dom} and V.

Another part of the uncertainties stems from the generalized models used in this article to estimate H_{dom} and *V*. Indeed, we used a single model across the entire dataset with disregard to the variability of growing conditions over each study site, leading to different canopy structures and stand-scale allometric relationships. A preliminary analysis of locally trained random forest models shows that for the estimation of H_{dom}, a locally trained model could slightly improve the estimation results (a maximum RMSE decrease of 16 cm was observed, Table S1). However, for the estimation of *V*, a locally trained model could reduce volume estimation errors by as much as 12.8 $m^3.ha^{-1}$ (see Table S2). The difference in accuracies on the estimation of *V* are most probably due to the allometric relations between H_{dom} and V, which could vary greatly between one region and another even for the same tree types, but further analysis is required to confirm these results.

Finally, after the filtering scheme applied to our dataset (see Section II-B-3), the remaining footprints with either value of the quality_flag (either 0 or 1) showed the same accuracy on the estimation of both H_{dom} and V for all the tested models.

VI. CONCLUSION

In this article, we analyzed GEDI data in order to determine its accuracy in estimating stand-scale dominant heights (H_{dom}) and stand volume (*V*) of intensively managed Eucalyptus plantations in Brazil. H_{dom} and *V* values have been estimated using the most accurate models used for estimating forest height and aboveground biomass from ICESat-1 waveforms. The GEDI waveform metrics used in the H_{dom} and *V* estimation models were extracted using algorithms provided by the land processes distributed active archive center (LP DAAC), in addition to already established metrics for ICESat-1 waveforms. Overall, 5517 GEDI shots over 566 Eucalyptus stands were analyzed over our study area.

For our study site defined by flat and gently sloping terrains (average slope $< 5°$), six regression models, a stepwise linear regression model (SRH), and a random forest regressor (RFH) using GEDI waveform metrics were assessed on their accuracy to estimate H_{dom}. Results showed that the most accurate model was SRH with an RMSE on the H_{dom} estimates of 1.33 m (R^2 of 0.93) using metrics extracted with the configuration of algorithm a1 and considering the higher mode between the last two modes as the ground return. For this model, the most relevant metrics for the estimation of H_{dom} was the 90th percentile relative height (RH_{90}), followed by RH_{10}, RH_{80}, and RH_{100}.

Stand wood volume (*V*) was modeled following power law with the canopy height, a stepwise regression model (SRV), and a random forest regressor (RFV). The four tested models showed similar accuracies, with SRV being the most accurate one with an RMSE of 24.39 $m^3.ha^{-1}$ (relative error $\sim$20% of the wood volume average). Similar to SRH, the most relevant metrics for the estimation of *V* using SRV was RH_{90}, followed by RH_{10}, RH_{30}, RH_{80}, and RH_{20}.

The choice of the algorithm used to extract the waveform metrics affected sometimes the accuracy, as metrics extracted using algorithm a5 showed $\sim$16% higher RMSE on the estimation of both H_{dom} and V. Nonetheless, the field "selected_algorithm" from the L2A dataset provides a robust indicator of the algorithm that provides the most accurate metrics. In addition to the accuracy of the metrics on the accuracy of H_{dom}, selecting the higher of the last two modes as the ground return could potentially increase the accuracies on H_{dom} from 12 to 20 cm.

Not all the variability on H_{dom} and *V* could be explained by GEDI alone. For example, including the stand age in the stepwise regression models help decrease the RMSE on the estimation of H_{dom} and *V* by, respectively, $\sim$13% and $\sim$7%. Nonetheless, and despite our efforts, underestimation of *V* using GEDI data was still observed for V greater than 250 $m^3.ha^{-1}$.

Finally, given the high accuracy of GEDI data on the estimation of tree heights and volume, GEDI can provide an excellent source of information to calibrate and validate upcoming and future radar missions, such as the upcoming *P*-band BIOMASS mission. GEDI can also supplement radar data by means of data fusion models in order to obtain high resolution and very accurate wall-to-wall maps of forest properties.

Author Contributions

Ibrahim Fayad–Conceptualization, Methodology, software, validation, formal analysis, data curation, visualization, writing–original draft. Nicolas Baghdadi—conceptualization, methodology, validation, formal analysis, data curation, writing—original draft. Clayton Alcarde Alvares—Conceptualization, validation, writing—review and editing. Jose Luiz Stape–Conceptualization, validation, writing - review and editing. Jean Stéphane Bailly—Validation, writing - review and editing. Henrique Ferraço Scolforo—Conceptualization, validation, writing—review and editing. Mehrez Zribi—Validation, writing - review and editing. Guerric Le Maire—Conceptualization, validation, writing—review and editing.

Acknowledgment

The authors would like to thank the GEDI team and the NASA LPDAAC (Land Processes Distributed Active Archive Center) for providing GEDI data. The authors acknowledge Suzano´s researchers Italo Ramos Cegatta, Renan Tarenta Meirelles Brasil and Carla Foster Feria for their technnical support and the CIRAD Suzano project. Suzano SA Company supported the forest-field data collection.

References

[1] M. Main-Knorn *et al.*, "Monitoring coniferous forest biomass change using a Landsat trajectory-based approach," *Remote Sens. Environ.*, vol. 139, pp. 277–290, Dec. 2013.

[2] A. Peregon and Y. Yamagata, "The use of ALOS/PALSAR backscatter to estimate above-ground forest biomass: A case study in Western Siberia," *Remote Sens. Environ.*, vol. 137, pp. 139–146, Oct. 2013.

[3] R. Houghton, F. Hall, and S. J. Goetz, "Importance of biomass in the global carbon cycle," *J. Geophys. Res., Biogeosci.*, vol. 114, no. G2, 2009.

[4] T. E. Fatoyinbo and M. Simard, "Height and biomass of mangroves in Africa from ICESat/GLAS and SRTM," *Int. J. Remote Sens.*, vol. 34, no. 2, pp. 668–681, Jan. 2013.

[5] M. A. Lefsky *et al.*, "Estimates of forest canopy height and aboveground biomass using ICESat: ICESAT Estimates of Canopy Height," *Geophys. Res. Lett.*, vol. 32, no. 22, Nov. 2005.

[6] T. Payn *et al.*, "Changes in planted forests and future global implications," *Forest Ecol. Manage.*, vol. 352, pp. 57–67, 2015.

[7] P. Elias and D. Boucher, *Planting for the Future. How Demand for Wood Products Could Be Friendly to Tropical Forests*. Cambridge, MA, USA: Union Concerned Scientists, 2014.

[8] R. Pirard, L. Dal Secco, and R. Warman, "Do timber plantations contribute to forest conservation?," *Environ. Sci. Policy*, vol. 57, pp. 122–130, Mar. 2016.

[9] M. A. Lefsky, W. B. Cohen, G. G. Parker, and D. J. Harding, "Lidar remote sensing for ecosystem studies," *Bio. Sci.*, vol. 52, no. 1, pp. 19, 2002.

[10] R. Nelson, K. J. Ranson, G. Sun, D. S. Kimes, V. Kharuk, and P. Montesano, "Estimating Siberian timber volume using MODIS and ICESat/GLAS," *Remote Sens. Environ.*, vol. 113, no. 3, pp. 691–701, Mar. 2009.

[11] R. O. Dubayah *et al.*, "Estimation of tropical forest height and biomass dynamics using Lidar remote sensing at La Selva, Costa Rica: Forest dynamics using LiDAR," *J. Geophys. Res.*, vol. 115, Jun. 2010.

[12] B. E. Schutz, H. J. Zwally, C. A. Shuman, D. Hancock, and J. P. DiMarzio, "Overview of the ICESat mission," *Geophys. Res. Lett.*, vol. 32, no. 21, 2005, Art. no. L21S01..

[13] Y. Pang, M. Lefsky, G. Sun, and J. Ranson, "Impact of footprint diameter and off-nadir pointing on the precision of canopy height estimates from spaceborne LiDAR," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2798–2809, Nov. 2011.

[14] I. Fayad *et al.*, "Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 502–514, Oct. 2016.

[15] Q. Chen, "Retrieving vegetation height of forests and woodlands over mountainous areas in the pacific coast region using satellite laser altimetry," *Remote Sens. Environ.*, vol. 114, no. 7, pp. 1610–1627, Jul. 2010.

[16] N. Baghdadi *et al.*, "Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in eucalyptus plantations in Brazil," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 290–299, Jan. 2014, doi: 10.1109/JSTARS.2013.2261978.

[17] J. Boudreau, R. Nelson, H. Margolis, A. Beaudoin, L. Guindon, and D. Kimes, "Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3876–3890, Oct. 2008.

[18] M. El Hajj, N. Baghdadi, N. Labrière, J.-S. Bailly, and L. Villard, "Mapping of aboveground biomass in Gabon," *Comptes Rendus Geosci.*, vol. 351, no. 4, pp. 321–331, Apr. 2019.

[19] M. R. Pourrahmati *et al.*, "Capability of GLAS/ICESat data to estimate forest canopy height and volume in mountainous forests of Iran," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5246–5261, Nov. 2015, doi: 10.1109/JSTARS.2015.2478478.

[20] M. R. Pourrahmati *et al.*, "Mapping Lorey's height over Hyrcanian forests of Iran using synergy of ICESat/GLAS and optical images," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 100–115, Jan. 2018.

[21] A. Neuenschwander and K. Pitts, "The ATL08 land and vegetation product for the ICESat-2 mission," *Remote Sens. Environ.*, vol. 221, pp. 247–259, Feb. 2019.

[22] R. Dubayah *et al.*, "The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography," *Sci. Remote Sens.*, vol. 1, Jun. 2020, Art. no. 100002.

[23] A. Persson, J. Holmgren, and U. Soderman, "Detecting and measuring individual trees using an airborne laser scanner," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 9, pp. 925–932, 2002.

[24] S. Saarela *et al.*, "Hierarchical model-based inference for forest inventory utilizing three sources of information," *Ann. Forest Sci.*, vol. 73, no. 4, pp. 895–910, Dec. 2016.

[25] S. Holm, R. Nelson, and G. Ståhl, "Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne LiDAR, and space LiDAR," *Remote Sens. Environ.*, vol. 197, pp. 85–97, Aug. 2017.

[26] T. R. Feldpausch *et al.*, "Tree height integrated into pantropical forest biomass estimates," *Biogeosciences*, vol. 9, no. 8, pp. 3381–3403, Aug. 2012.

[27] E. Kearsley *et al.*, "Conventional tree height–diameter relationships significantly overestimate aboveground carbon stocks in the central Congo Basin," *Nat. Commun.*, vol. 4, no. 1, Oct. 2013.

[28] Y. Su, Q. Ma, and Q. Guo, "Fine-resolution forest tree height estimation across the Sierra Nevada through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery," *Int. J. Digit. Earth*, vol. 10, no. 3, pp. 307–323, Mar. 2017.

[29] H. Tang *et al.*, "Deriving and validating leaf area index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA," *Remote Sens. Environ.*, vol. 143, pp. 131–141, Mar. 2014.

[30] Y. Wang *et al.*, "Is field-measured tree height as reliable as believed – A comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a Boreal forest," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 132–145, Jan. 2019.

[31] R. Dubayah *et al.*, "The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography," *Sci. Remote Sens.*, vol. 1, Jun. 2020, Art. no. 100002.

[32] S. L. R. Dubayah, "GEDI L1B geolocated waveform data global footprint level V001," NASA EOSDIS Land Processes DAAC, 2020. Accessed: Jul. 2021. [Online]. Avilable: https://doi.org/10.5067/GEDI/GEDI01_B.001

[33] S. L. R. Dubayah, "GEDI L2A elevation and height metrics data global footprint level V001," NASA EOSDIS Land Processes DAAC, 2020. Accessed: Jul. 2021. [Online]. Avilable: https://doi.org/10.5067/GEDI/GEDI02_A.001

[34] S. L. R. Dubayah, "GEDI L2B canopy cover and vertical profile metrics data global footprint level V001," NASA EOSDIS Land Processes DAAC, 2020. Accessed: Jul. 2021. [Online]. Avilable: https://doi: 10.5067/GEDI/GEDI02_B.001

[35] C. Hilbert and C. Schmullius, "Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape," *Remote Sens.*, vol. 4, no. 8, pp. 2210–2235, Jul. 2012.

[36] T. J. Urban, B. E. Schutz, and A. L. Neuenschwander, "A survey of ICESat coastal altimetry applications: Continental coast, open ocean island, and inland river," *Terr. Atmospheric Ocean. Sci.*, vol. 19, pp. 1–19, 2008.

[37] D. J. Harding, "ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure," *Geophys. Res. Lett.*, vol. 32, no. 21, 2005, Art. no. L21S10.
[38] M. A. Lefsky, M. Keller, Y. Pang, P. B. De Camargo, and M. O. Hunter, "Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms," *J. Appl. Remote Sens.*, vol. 1, no. 1, 2007, Art. no. 013537
[39] Y. Pang, M. Lefsky, H.-E. Andersen, M. E. Miller, and K. Sherrill, "Validation of the ICEsat vegetation product using crown-area-weighted mean height derived using crown delineation with discrete return lidar data," *Can. J. Remote Sens.*, vol. 34, pp. S471–S484, 2008.
[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[41] S. S. Saatchi *et al.*, "Benchmark map of forest carbon stocks in tropical regions across three continents," *Proc. Nat. Acad. Sci.*, vol. 108, no. 24, pp. 9899–9904, Jun. 2011.
[42] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa, Eds. New York, NY, USA: Springer, 1998, pp. 199–213.
[43] G. le Maire *et al.*, "MODIS NDVI time-series allow the monitoring of eucalyptus plantation biomass," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2613–2625, Oct. 2011.

Ibrahim Fayad received the Engineering degree in computer and telecommunications in 2011 and the Ph.D. degree in automatic and microelectronic systems both from the University of Montpellier, Montpellier, France in 2015.

He is currently a Research Engineer with the National Research Institute for Agriculture, Food and the Environment, Montpellier, France. His research interests include machine learning for the retrieval of environmental parameters using remote sensing data.

Nicolas N. Baghdadi received the Ph.D. degree from the University of Toulon, Toulon, France in 1994.

From 1995 to 1997, he was a Postdoctoral Researcher with INRS Ete—Water Earth Environment Research Centre, Quebec University. From 1998 to 2008, he was with the French Geological Survey, Orléans, France. Since 2008, he has been a Senior Scientist with the National Research Institute for Agriculture, Food and the Environment, Montpellier, France. He is currently a Scientific Director with the THEIA Land Data Centre.

His research activities are in the areas of microwave remote sensing, image processing, and satellite and airborne remote sensing data analysis. His main field of interest is the analysis of remote sensing data (SAR, Lidar, optical) and the retrieval of environmental parameters (e.g., soil moisture content, surface roughness, biomass, etc.).

Clayton Alcarde Alvares received the Ph.D. degree in forestry science from the University of São Paulo, São Paulo, Brazil, in 2012.

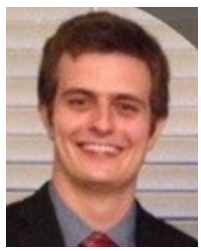
His research focused on mapping and edaphoclimatic modeling of productivity of Eucalyptus plantations in Brazil. Since 2011, he has been a Forest Scientist with Suzano SA, São Paulo, Brazil, where he has been developing in-depth analyzes in climatology, remote sensing, production environments zoning, and applied ecophysiology, and is strongly committed to delivering tools for operational uses in the short, medium and long term. Since 2019, he has been a Professor and an Advisor of the Postgraduate Program in Forest Science, State University "Júlio Mesquita Filho"—UNESP, Faculty of Agronomic Sciences—Campus de Botucatu.

Jose Luiz Stape received the Ph.D. degree in forest ecology from Colorado State University, Fort Collins, CO, USA, in 2002.

He is a Permanent Graduate Professor of Forest Ecophysiology with Sao Paulo State University (UNESP, Brazil), Sao Paulo, Brazil. He was with the University of Sao Paulo, and with the North Carolina State University and across many countries and companies, looking to improve silvicultural recommendations for the sustainability of forest plantations including: clonal deployment; site-preparation; nutrition and spacing. To better evaluate the factors limiting forest productivity and controlling C allocation, he coordinated the establishment, with other scientists, of four large Eucalyptus and Pine cooperative research programs in Brazil via IPEF (BEPP, Eucflux, TECHS and PPPIB) and a research network at Suzano company (G2M2P2). Nowadays the use of remote sensing to improve monitoring, management and modeling planted forests has been a main focus of his research.

Jean Stéphane Bailly received the Engineering degree in agronomy, the M.Sc. degree in biostatistics, and the Ph.D. degree in hydrology from the University of Montpellier, Montpellier, France, in 1990, 2003, and 2007.

He is currently a Senior Lecturer of physical geography and geostatistics in AgroParis-Tech, Paris, France. He is currently a Research Fellow with the UM LISAHLab, Montpellier, France. His research interests include spatial observations and modeling for hydrological issues.



Henrique Ferraço Scolforo received the Ph.D. degree in forest biometrics from the North Carolina State University, Raleigh, NC, USA, in 2018.

His research focused on growth and yield modeling sensitive to climate and clonal variation applied to eucalypt stands in Brazil. Since 2018, he has been leading the biometrics, inventory, growth, and yield studies with Suzano SA, São Paulo, Brazil.

Mehrez Zribi received the B.E. degree in signal processing from the Ecole Nationale Supérieure d'Ingénieurs en Constructions Aéronautiques, Toulouse, France, in 1995, and the Ph.D. degree in signal processing and remote sensing from the Université Paul Sabatier, Toulouse, France, in 1998.

He is currently a Research Director with the Centre National de Recherche Scientifique, Paris, France. In 1995, he was with the Centre d'Etude des Environnements Terrestre et Planétaires Laboratory/Institut Pierre Simon Laplace, Vélizy, France. Since October 2008, he has been with the Centre d'Etudes Spatiales de la Biosphère, Toulouse, France. He has authored or coauthored more than 140 articles in refereed journals. He is currently the Director of Centre d'Etudes Spatiales de la Biosphère, Toulouse, France. His research interests include microwave remote sensing applied to hydrology, microwave modeling for land surface parameters estimations and finally airborne microwave instrumentation.

Guerric le Maire received the M.Sc. degree in agronomy, specialized in ecology, from the National Institute of Agronomy (INA P-G, Paris, France), Paris, France, in 2002 and the Ph.D. degree in plant ecophysiology from Paris XI University, Orsay, France, in 2005.

During 2006–2007, he was a Postdoctoral Researcher with the Le Laboratoire des Sciences du Climat et de l'Environnement Laboratory (Saclay-France). He was a Researcher with the CIRAD, Montpellier, France, in 2008. His research interests include remote-sensing image processing/analysis and process-based ecophysiological forest models development applied to tropical forest plantations.