remote sensing

MDPI

Article

Terrain Slope Effect on Forest Height and Wood Volume Estimation from GEDI Data

Ibrahim Fayad [1,*], Nicolas Baghdadi [1], Clayton Alcarde Alvares [2,3], Jose Luiz Stape [2], Jean Stéphane Bailly [4,5], Henrique Ferraço Scolforo [3], Italo Ramos Cegatta [3,6], Mehrez Zribi [7] and Guerric Le Maire [8,9]

1 CIRAD, CNRS, INRAE, TETIS, AgroParisTech, University Montpellier, CEDEX 5, 34093 Montpellier, France; nicolas.baghdadi@teledetection.fr
2 Unesp, Faculdade de Ciências Agronômicas, Botucatu 18610-034, Brazil; calcarde@suzano.com.br (C.A.A.); jlstape@gmail.com (J.L.S.)
3 Suzano SA, Estrada Limeira 391, Limeira 13465-970, Brazil; hscolforo@suzano.com.br (H.F.S.); italocegatta@gmail.com (I.R.C.)
4 Institut Agroalimentaire, LISAH, University Montpellier, INRAE, IRD, CEDEX 1, 34060 Montpellier, France; bailly@agroparistech.fr
5 AgroParisTech, 75005 Paris, France
6 ESALQ—Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Av. Pádua Dias 11, Piracicaba 13418-900, Brazil
7 CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, 31400 Toulouse, France; mehrez.zribi@ird.fr
8 CIRAD, UMR Eco&Sols, 34398 Montpellier, France; guerric.le_maire@cirad.fr
9 Institut Agroalimentaire, Eco&Sols, University Montpellier, CIRAD, INRAE, IRD, CEDEX 5, 34093 Montpellier, France
* Correspondence: ibrahim.fayad@inrae.fr

Citation: Fayad, I.; Baghdadi, N.; Alcarde Alvares, C.; Stape, J.L.; Bailly, J.S.; Scolforo, H.F.; Cegatta, I.R.; Zribi, M.; Le Maire, G. Terrain Slope Effect on Forest Height and Wood Volume Estimation from GEDI Data. *Remote Sens.* **2021**, *13*, 2136. https://doi.org/10.3390/rs13112136

Academic Editor: Lin Cao

Received: 7 April 2021
Accepted: 25 May 2021
Published: 28 May 2021

Abstract: The Global Ecosystem Dynamics Investigation LiDAR (GEDI) is a new full waveform (FW) based LiDAR system that presents a new opportunity for the observation of forest structures globally. The backscattered GEDI signals, as all FW systems, are distorted by topographic conditions within their footprint, leading to uncertainties on the measured forest variables. In this study, we explore how well several approaches based on waveform metrics and ancillary digital elevation model (DEM) data perform on the estimation of stand dominant heights (H_{dom}) and wood volume (V) across different sites of Eucalyptus plantations with varying terrain slopes. In total, five models were assessed on their ability to estimate H_{dom} and four models for V. Results showed that the models using the GEDI metrics, such as the height at different energy quantiles with terrain data from the shuttle radar topography mission's (SRTM) digital elevation model (DEM) were still dependent on the topographic slope. For H_{dom}, an RMSE increase of 14% was observed for data acquired over slopes higher than 20% in comparison to slopes between 10 and 20%. For V, a 74% increase in RMSE was reported between GEDI data acquired over slopes between 0–10% and those acquired over slopes higher than 10%. Next, a model relying on the height at different energy quantiles of the entire waveform (HT_n) and the height at different energy quartiles of the bare ground waveform (HG_n) was assessed. Two sets of the HG_n metrics were generated, the first one was obtained using a simulated waveform representing the echo from a bare ground, while the second one relied on the actual ground return from the waveform by means of Gaussian fitting. Results showed that both the simulated and fitted models provide the most accurate estimates of H_{dom} and V for all slope ranges. The simulation-based model showed an RMSE that ranged between 1.39 and 1.66 m (between 26.76 and 39.26 $m^3 \cdot ha^{-1}$ for V) while the fitting-based method showed an RMSE that ranged between 1.26 and 1.34 m (between 26.78 and 36.29 $m^3 \cdot ha^{-1}$ for V). Moreover, the dependency of the GEDI metrics on slopes was greatly reduced using the two sets of metrics. As a conclusion, the effect of slopes on the 25-m GEDI footprints is rather low as the estimation on canopy heights from uncorrected waveforms degraded by a maximum of 1 m for slopes between 20 and 45%. Concerning the wood volume estimation, the effect of slopes was more pronounced, and a degradation on the accuracy (increased RMSE) of a maximum of 20 $m^3 \cdot ha^{-1}$ was observed for slopes between 20 and 45%.

Keywords: GEDI; LiDAR; terrain slope; canopy height; wood volume

1. Introduction

In the last couple of decades, due to its accurate Earth observation capabilities, remote sensing has increasingly been used for the estimation, on local and global scales, of forest biophysical characteristics, namely forest heights and above ground biomass (AGB). The estimation of forest characteristics is not restricted to a particular remote sensing technique, as it has been obtained using either passive optical sensing such as from optical imagery, or using active sensors such as synthetic aperture radar (SAR) or light detection and ranging (LiDAR) data. Nonetheless, LiDAR has proven to be better suited for the estimation of AGB and canopy heights than SAR (with available wavelengths to date: L, C and X bands), Global Navigation Satellite System Reflectometry [1], and optical imagery [2,3]. LiDAR data show lower signal saturation with AGB than optical and radar data. In general, the literature reports saturation thresholds with optical and radar imageries (with X, C and L-bands for SAR data) rarely exceeding 150 Mg/ha [4,5] whereas LiDAR data have shown AGB estimation capabilities up to 1200 t/ha [6]. Yet, the AGB estimation levels from LiDAR data are based on canopy vertical structure metrics, and the relationship between height structure and ABG itself may saturate at high AGB, thus sometimes a start of under-estimation of AGB using LiDAR data is observed [7].

LiDAR systems capture vertical structures by measuring the time taken for an emitted laser pulse to return. Over vegetated areas, the emitted pulse will start reflecting as soon as it hits the top of canopy (given a large enough top of canopy surface), and the final return will theoretically be from the ground (if the laser pulse can penetrate through the gaps [8]). The representation of such vertical structures depends on the LiDAR system used. Discrete LiDAR systems usually have small footprints (<1 m) and record multiple returns representing different targets within the travel path. The returned laser echoes are next recorded as a series of 3D coordinates known as point clouds. On the other hand, full waveform (FW) LiDAR systems record all the reflected signals over a given footprint, and not only the first one. They therefore provide a continuous vertical profile representing the heights of the different objects within their footprints, which is usually larger than 10 m. Therefore, FW LiDAR systems provide much richer information about the spatial arrangement of structures within their waveforms [9]. The recorded vertical echoes of objects within the waveform are represented as a series of multiple connected temporal peaks. These peaks might therefore represent reflections from a single object (e.g., top of canopy cover) or different objects with relatively the same heights (e.g., short understory and the ground) [10]. To measure vegetation characteristics, the vegetation and ground portion of the waveform need to be identified and separated. As such, given the wide footprint of FW LiDAR systems, a source of uncertainty on the estimation of forest characteristics such as canopy heights and biomass could occur based on the local terrain morphology. For instance over terrain with a high relief, the ground return might get mixed with the vegetation leading to an overestimation of the relevant vegetation characteristics [11].

Over the last years, many studies were carried out on FW data acquired by the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud, and land Elevation satellite (ICESat) for the estimation of forest characteristics. ICESat/GLAS was the first spaceborne FW LiDAR system, and operated from 2003 until 2009 [12]. ICESat/GLAS pulsed one of its three 1064 nm lasers at a time, at 40 Hz, producing ellipsoidal shaped footprints with a diameter of ~60 m at ~170 m along-track intervals, and several kilometers across tracks. ICESat/GLAS was used during its operational and post-operational period in many studies for the precise estimation of forest characteristics such as canopy height and biomass [11,13–19]. However, most of the previously mentioned studies focused on forests with relatively flat terrain since ICESat/GLAS with its large footprint size was

susceptible to overestimating forest characteristics (e.g., canopy heights, and AGB) over terrains with high relief [20,21]. Nonetheless, a number of studies have been carried out to address this particular issue, and presented several methods to minimize the effects of slope on the waveforms. These studies can be grouped into three categories. The first group of studies attempted to retrieve slope information from the waveforms in the form of waveform metrics, such as the trailing edge extent (representing terrain variability) and the leading edge extent (representing vegetation variability) [11,13,20,22] and terrain indices (range of ground surface elevations within a sampling window) retrieved from a digital elevation model (DEM). This technique provided increased accuracies on the estimation of forest heights over sloping terrain [11], but the squared correlation coefficient (R^2) decreased with increased slope values, and was only viable for slopes lower than 15° (R^2 = 0.63) [23]. The second type of studies, such as the study of Yang et al. [24], minimized the effects of slope on large footprint LiDAR by modifying the geometric optical and radiative transfer (GORT) vegetation LiDAR model [25] to take into account the impacts of surface topography. Their approach showed that slope-corrected ICESat/GLAS footprints had an accuracy on the estimation of canopy heights of 2.2 m (R^2 of 0.77). More recently, Wang et al. [26] developed a method based on new waveform metrics to minimize the effects of slopes on the estimation of forest AGB. The developed method relies on quantile heights (the height above ground at which *n*% of the waveform energy falls below) [27], and quantile heights of the ground return only. The method developed by Wang et al. [26] can be decomposed into three major steps. (1) A LiDAR waveform over bare grounds but with similar slope as the studied waveform is first simulated in order to derive a ground return. This was necessary in the study of Wang et al. [26] as they worked on ICESat/GLAS waveforms with a ~70 m diameter footprints alongside simulated waveforms with 25 m diameter footprints. For ICESat/GLAS, the ground return disappeared with slopes bigger than 15° [28] (i.e., returns from ground and vegetation become mixed up over sloping terrain). Next, the simulated waveform is aligned with the studied waveform at the signal end, and finally the related metrics were derived. The methodology developed by Wang et al. [26] gave significant increase in accuracy in comparison to previous methodologies. Indeed, for the estimation of above ground biomass, a decrease of 20.83 Mg/ha in terms of RMSE (32% increase in R^2) was observed for slope ranges of 0–40°.

ICESat/GLAS was succeeded 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. In contrast to ICESat/GLAS, ATLAS is equipped with a single 532 nm wavelength laser that emits six beams (arranged into three pairs). Beam pairs are separated by ~3 km across-track with a pair spacing of 90 m. The nominal footprint of ATLAS is 17 m with a spacing interval of 0.7 m along-track. Moreover, unlike ICESat/GLAS, ATLAS uses a photon counting system rather than a full waveform system, and has the ability to detect single echoed photons. However, given the wavelength of the equipped laser (532 nm), ATLAS has lower reflectance over vegetation in comparison to ice [29], coupled with the low number of reflected photons, ICESat-2 might be limited over tropical forests for direct canopy height retrievals due to the canopy cover [29].

Recently, the Global Ecosystem Dynamics Investigation (GEDI) on board the International Space Station (ISS) started collecting LiDAR data globally since April 2019. GEDI's mission 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 [30]. GEDI shows many similarities to ICESat/GLAS, however, given GEDI's higher sampling rate (242 vs. 40 Hz for ICESat-1), and its much smaller footprint size (~25 vs. ~60 m for ICESat-1), GEDI will provide a highly improved coverage, and improved measurements over forested areas with high relief in comparison to ICESat/GLAS. Nonetheless, given that GEDI is a FW LiDAR system, it is expected to also be affected by relief. However, since GEDI has a much smaller footprint than ICESat/GLAS while getting equivalent vertical resolution (1 ns), the effect of slopes on the waveform should be less pronounced.

As far as we know, no studies have yet been dedicated to analyze the effects of slopes on GEDI data. Therefore, the objective of this study is two-fold. First, the effects of the terrain slope on the estimation of both canopy heights and wood volume of fast-growing Eucalyptus plantations in Brazil will be analyzed. Next, slope-effect minimization techniques from previous literature will be used in order to determine which methodology yields the best forest characteristics estimates over sloping terrain. The choice of the in situ dataset was decided since the physical structure of Eucalyptus is simple enough, and very homogeneous within the GEDI footprint, thus reducing uncertainties unrelated to the GEDI sensor itself.

The rest of this paper is organized as follows: Section 2 describes the study area and lists the data used. Section 3 presents a thorough description of the methods for the estimation of canopy heights and wood volume over sloping terrain. Sections 4 and 5 present the results and the discussion, respectively. Finally, in Section 6, we summarize and conclude our study.

2. Study Area and Dataset

2.1. Study Area

The study area is located in Brazil, in five regions across a large latitudinal gradient (Figure 1), covering different climate and soil types. Maranhão (MA) is located in a typical equatorial region with different intensities of monsoon rainfall (1200 to 2500 mm/year. Bahia (BA) and Espírito Santo (ES) are located in a tropical coastal region with strong rainfall anisotropy (800 to 1500 mm/year) that directly affects wood productivity from the nearshore towards the hinterland. Mato Grosso do Sul (MS) is located in a tropical region (1200 to 1500 mm/year) but with some subtropical features (rare frost), it is the most environmentally homogeneous among the five study areas, resulting in less variation in wood productivity within the region. São Paulo (SP) is mainly in a subtropical region (1100 to 2000 mm/year with orographic effects), heavy frost-days are frequent in the southern part, has complex relief and a wide range of deep and shallow tropical soils, resulting in a huge variability in wood productivity across the region. Across the five regions, 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. The wood 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 dominant height of around 80% of the stands is in the 20–35 m range with a stand volume between 180 and 300 m^3/ha. The plantations were managed locally by stand units (~50 ha), where the same management is applied for each stand: planting, harvesting and weed control, genetic material, soil preparation and fertilization. The plantations are generally characterized by a sparse understory and herbaceous strata. Eucalyptus 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. The "soil strata" is mainly constituted of litter accumulation of branches and leaves, with some patches of herbaceous species. More than 82% of the Brazilian Eucalyptus plantations are cultivated on flat to gentle slopes due to huge harvesting and logging operation costs on high slopes [31]. In some areas, such as the states of Minas Gerais, Paraná, Santa Catarina, and São Paulo (Paraíba Valley), high slopes are however the rule.

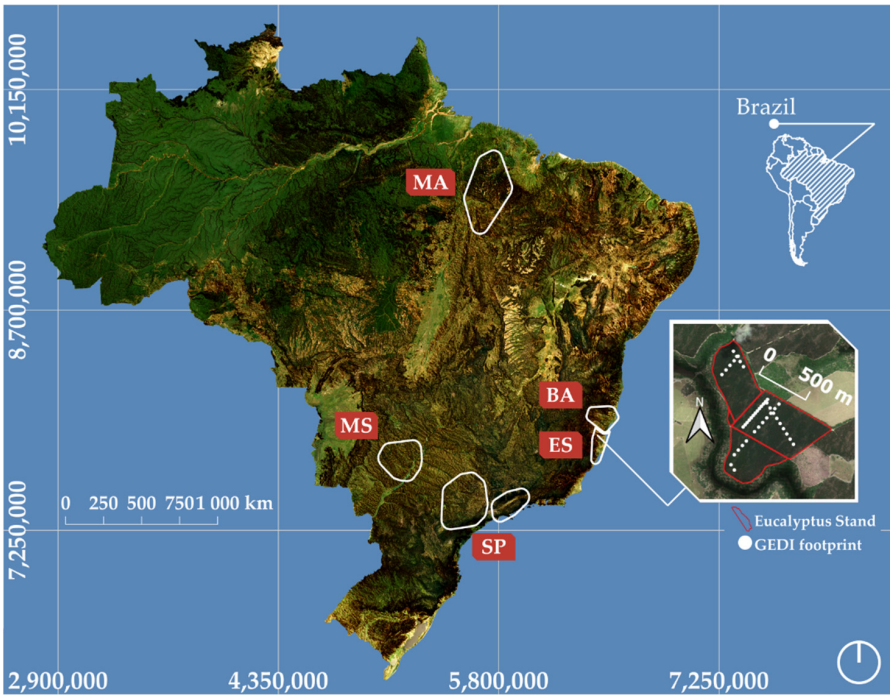

Figure 1. Location of the five study sites; the zoomed-in rectangle shows an example of GEDI tracks over some stands.

2.2. In Situ Data

A total of 168 Eucalyptus stands were selected with field inventories performed between 1 April 2019 and 1 June 2019. The stands were selected due to the presence of acquired GEDI shots within this period. Moreover, these stands are generally located on a terrain with slopes of varying degrees. Specifically, 86.4% of stands are located over slopes lower than 10%, 11.2% of stands are located on sloping terrain with slope ranges between 10 and 20%, and the remainder of the stands (2.4%) are located on a terrain with slopes between 20 and 45%. 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. Permanent inventory plots had an area of approximately 400 m^2 and were systematically distributed throughout the stand with a density of one plot per 10 ha. They included 30 to 100 trees with an average of 58 trees. 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 four largest trees in terms of DBH (dominant trees) were measured. The mean height of these dominant trees defined the dominant height of the plot (H_{dom}). H_{dom}, basal area and age on the inventory date were then used in local volume equations to estimate the plot's total and merchantable volume (the merchantable volume is a tree's volume up to 6 cm stem diameter with bark). Table 1 shows the distribution of field measurements of H_{dom} and wood volume.

Table 1. Distribution of dominant canopy height (**a**) and wood volume densities (**b**) from field inventories over the 168 Eucalyptus stands.

H_{dom} Classes (m)	Stand Distribution (%)	V Classes ($m^3 \cdot ha^{-1}$)	Stand Distribution (%)
[10–15[	10	[0–75[	19
[15–20[	32	[75–150[	33
[20–25[	28	[150–255[	26
[25–30[	24	[255–300[	16
[30–35]	7	[350–450]	6
(**a**)		(**b**)	

2.3. GEDI Dataset

GEDI is equipped with three 1064 nm onboard lasers designed to sample the Earth's surface at a ~60 m interval along the track with a cross track separation of ~600 m. One of the lasers is split into two beams. These four beams are then dithered across track to produce eight parallel tracks of observations. The GEDI lasers fire at a frequency of 242 Hz and, on the ground, measure 3D structures over a 25-m wide footprint. The received 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 [30].

GEDI data are already processed by the Land Processes Distributed Active Archive Center (LP DAAC, https://lpdaac.usgs.gov/tools/data-pool/ (accessed on 27 May 2021)). GEDI data are provided through three data products, L1B [32], L2A [33], and L2B [34]. The L1B data product, contains among other, the geolocated (longitude, latitude, and elevation) raw transmitted and received waveforms as well as information on mean and standard deviation of the noise, and acquisition time. The L2A product 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, Leaf Area Index (LAI), and foliage height diversity (FHD).

The extracted metrics from each waveform are the results of several processing steps [32,33]. First, the raw received waveforms are smoothed to reduce the noise in the signal, and thus permitting the determination of the useful part of the waveform within the corresponding footprint [32]. Waveform smoothing is performed by means of a Gaussian filter with a current width of 6.5 ns. The smoothing permits the determination of *searchstart* and *searchend*, which are, respectively, the first and last positions in the signal where the signal intensity is above the following threshold [32]:

$$threshold = mean + std.v \quad (1)$$

where "*mean*" is the mean noise level, "*std*" is the standard deviation of the noise of the smoothed waveform, and "*v*" is a constant currently set at 4. Inside the window delimited by *searchstart* and *searchend*, the highest (*toploc*) and lowest (*botloc*) detectable returns are determined (Figure 2) [32]. *toploc* and *botloc* respectively represent the highest and lowest locations inside the waveform extent where two adjacent intensities are above a threshold. The threshold equation used to determine *toploc* and *botloc* is the same as Equation (1), with "*v*" an integer fixed at 2, 3, 4 and 6. Two values of "*v*" are used to determine the *toploc* ("Front_threshold") and *botloc* ("Back_threshold"). Finally, the location of the distinctive peaks or modes in the waveform, such as the ground peak, or top of canopy peaks are 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 (Figure 2) [33]. The width of the second Gaussian filter ("Smoothwidth_zcross") is fixed to either 3.5 or 6.5 ns. Currently, the LP DAAC provides six configurations (a1 to a6) for the estimation of the waveform metrics. The difference in these configurations are the values used for the thresholds presented earlier. For studies on Eucalyptus plantations, Fayad et al. [7] determined that algorithm a1 (Smoothwidth_zcross = 6.5, Front_threshold = 3, Back_threshold = 6) provided the best metrics for the estimation of canopy heights and wood volume.

In this study, variables from both L1B and L2A were extracted. From L1B, we extracted the raw received waveforms, their geolocation (longitude, and latitude), as well as their acquisition date and times. From the L2A data product, we extracted the following variables: (1) the position within the waveform of *toploc* and *botloc*, and (2) the relative height metrics at 10% intervals from *botloc* (0%) to *toploc* (100%) ($RH_n, 10\% \leq n \leq 100\%, step\ 10\%$). RH_n represent the height between *botloc* and the location at n% of cumulative energy (Figure 2) [33].

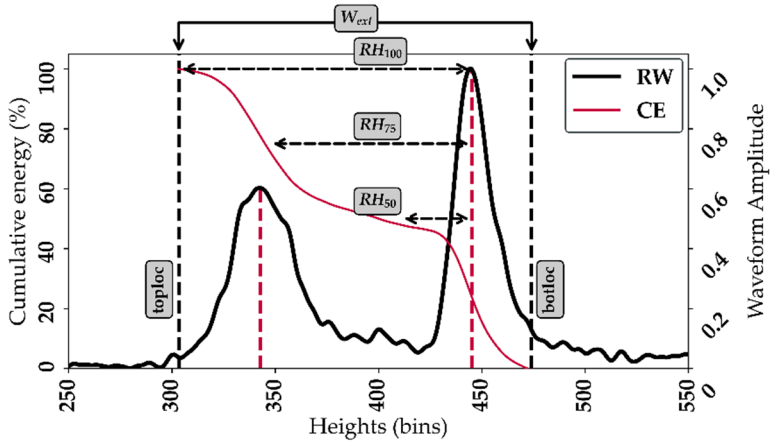

Figure 2. Example of an acquired GEDI waveform (RW) over an Eucalyptus stand (H_{dom} = 20.4 m; V = 129.6 $m^3 \cdot ha^{-1}$). The cumulative energy of the waveform (CE) between *botloc* and *toploc* and the corresponding relative heights (RH_n) at different quantiles "*n*". The left and right red dashed lines represent, respectively, the position of the vegetation (*Vloc*) and ground peaks (*Gloc)*. Note that 1 ns corresponds to 15 cm sampling distance in the waveform. The waveform amplitudes are counts from the analog to digital converter (ADC) on the instrument and normalized to be between 0 and 1.

Initially, 2128 GEDI shots acquired over the 168 Eucalyptus stands were selected. These GEDI shots corresponded to acquisitions within a window of two months (before or after) the stand inventory. The two-month window is necessary to overcome tree growth differences that occur between inventory times and acquisition times. In fact, on these fast-growing plantations, a 2-month difference could result in an up to 50 cm growth in H_{dom} and 10 $m^3 \cdot ha^{-1}$ in V (depending on the genetic material, pedoclimatic conditions and age). However, this reasonable hypothesis allows keeping a large number of stands, including a large variability of age and growing conditions.

Finally, 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:

- Waveforms with reported elevations that are significantly higher than the corresponding elevations in the SRTM DEM [15]. In essence, we removed all waveforms where the absolute difference is higher than 100 m;
- Waveforms with a difference between waveform extent (W_{ext}, height between *toploc* and *botloc*, [13]) and (*Gloc-Vloc*) higher than 400 bins (corresponding to 60 m).

After filtering, 1477 (~69.4%) GEDI shots were retained and used.

2.4. Digital Elevation Model Metrics

The Shuttle Radar Topography Mission's (SRTM) Digital Elevation Model (DEM) with a spatial resolution of 30 m was used is this study. Two variables were derived from the DEM for each GEDI footprint: slope (*S*), and surface Roughness (*ROUG*). The surface roughness map was obtained by computing the standard deviation of the elevation in a 3 × 3 pixel-moving window.

3. Methodology

3.1. Stand Scale Dominant Heights Estimation

In this section, we evaluate five models for the estimation of stand-scale dominant heights (H_{dom}) from GEDI data acquired over terrain with different slopes. Two models are based on GEDI metrics and terrain information from the SRTM DEM and H_{dom} is estimated through linear regression, while the remaining three will be exclusively based on GEDI

metrics where H_{dom} is estimated using random forest regression algorithms. The first tested model is based on the formulation by Lefky et al. [13]:

$$H_{dom} = a \cdot W_{ext} - b \cdot S + c \quad (2)$$

where W_{ext} is the waveform extent (height difference between *botloc* and *toploc*) in meters, S is the terrain slope in degrees, a is the coefficient applied to the waveform height index, b is the coefficient applied to the slope, and c is a constant.

In the study of Fayad et al. [7], it has been shown that the relative height metric (RH_{100}) (Figure 2) is better correlated to in situ H_{dom} than the waveform extent. Therefore, a modified version of Equation (2) will also be tested. Equation (3) will use the RH_{100} instead of the W_{ext}:

$$H_{dom} = a \cdot RH_{100} - b \cdot S + c \quad (3)$$

We will also attempt to estimate H_{dom} 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 over-fitting [35]. Compared to linear models, RF is advantageous for being able to model also nonlinear relationships (threshold effect) between the variable to explain and the explanatory variables. The suggested model will use the relative height metrics $RH_n\{10\% \leq n \leq 100\%, step10\%\}$, as well as the terrain roughness (*ROUG*), and terrain slope (S).

The models integrating terrain information, such as the models provided by Lefsky et al. [13], have been proven to increase the accuracy on H_{dom} estimation over sloping terrain [11]. However, the effect of slopes is not completely eliminated. For instance, in the study of Xing et al. [23], the accuracy of the method proposed by Lefsky decreased with the increasing slope. Indeed, the RMSE increased from of 2.87 m (R^2 of 0.89) for slope ranges 0–5° to 5.97 m (R^2 of 0.08) for slope ranges 0–30°. Therefore, Wang et al. [26] proposed a method relying on new waveform metrics in order to reduce the effects of slopes. The method of Wang et al. [26] comprises three steps in order to generate the new waveform metrics. First, a waveform over bare grounds with the same slope value as the studied waveform is simulated. The simulated waveform is based on a Gaussian function, resembling the Laser pulse used by FW LiDARs, and thus has the following form (assuming a nadir-viewing angle):

$$y = A \cdot exp\left(\frac{-x^2}{2\sigma^2}\right) \quad (4)$$

where σ and A are respectively the standard deviation and amplitude of simulated echoed waveform, and x the waveform sample locations at 1 ns (15 cm) intervals. For simplicity, the amplitude (A) is set to one. The standard deviation of the waveform, which affects its width, is dictated by the characteristics of the LiDAR system used. Over flat bare grounds, the standard deviation of the echoed waveform is defined as follows:

$$\sigma = \sigma_b = \sqrt{\frac{-(c \cdot t)^2}{2ln(PT)}} \quad (5)$$

where c is the speed of light ($3 \cdot 10^8$ m/s), t is the pulse width of the LiDAR system (t = 15.6 ns for GEDI [30]), and PT = 0.5 (half the amplitude points of the pulse width).

Over sloping terrain, the waveform extent will increase with the terrain slope even if the canopy is the same, and the echoed waveform will exhibit a broadening of the ground return (Figures 3 and 4). Therefore, the standard deviation value to simulate a bare ground

return (Equation (5)) should be broadened in accordance to the terrain slope. The standard deviation on a slope terrain σ_s is computed from σ_b by adding a broadening effect:

$$\sigma = \sigma_s = \sigma_b + \beta \cdot \gamma \cdot tan\theta \quad (6)$$

where σ_b is obtained from Equation (5), β = 0.5 for waveforms simulated over forest stands [36], γ is the footprint diameter (γ = 25 m for GEDI), and finally, θ is the terrain slope in degrees (°).

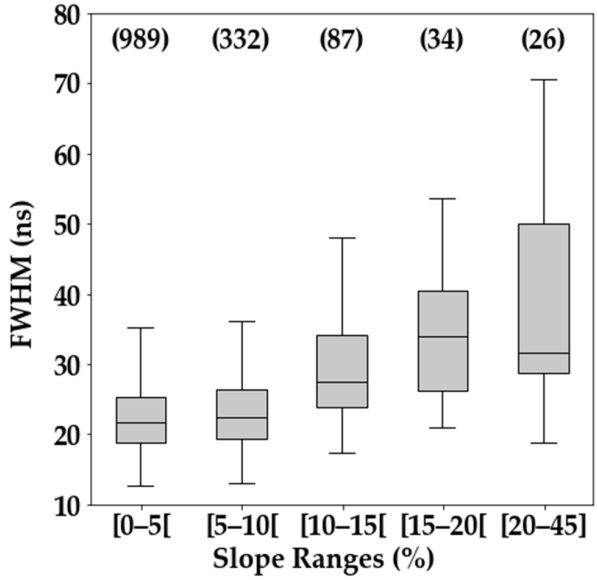

Figure 3. Box-whiskers plots of the full width at half maximum (FWHM) of the fitted ground return across five slope ranges. Numbers in parenthesis represent the number of points in each slope class.

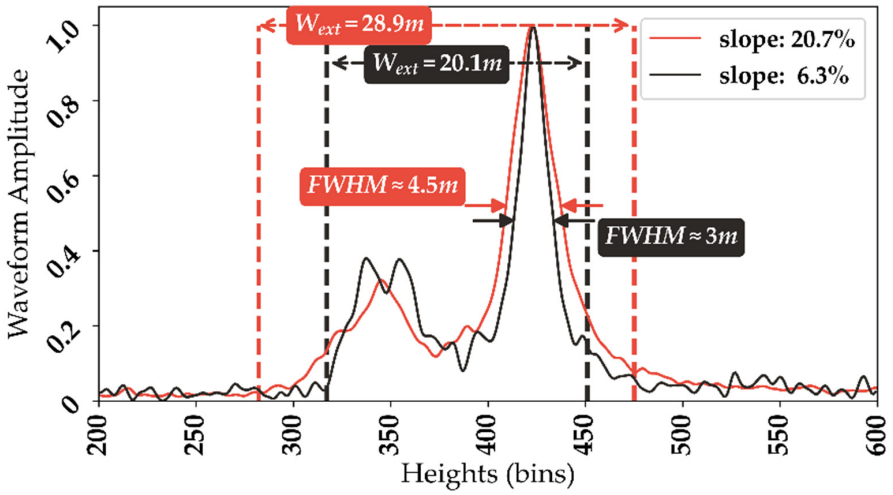

Figure 4. GEDI waveforms acquired over two Eucalyptus stands with the same H_{dom} (~16.75 m) but different terrain slopes. The dashed lines represent the *toploc* and *botloc* positions. The full width at half maximum (FWHM) is indicated for the ground peak. Both waveform amplitudes were normalized to be between 0 and 1 for comparison purposes.

After determining the shape of the waveform over a bare ground with known slope (Figure 5a), the simulated ground return is superposed over the studied GEDI waveform by the position of the signal end (*botloc*) of both waveforms (Figure 5b). For the simulated waveform, *toploc* is determined as the first position where the amplitude of the simulated waveform is above zero, while *botloc* is the last position where the amplitude is above zero.

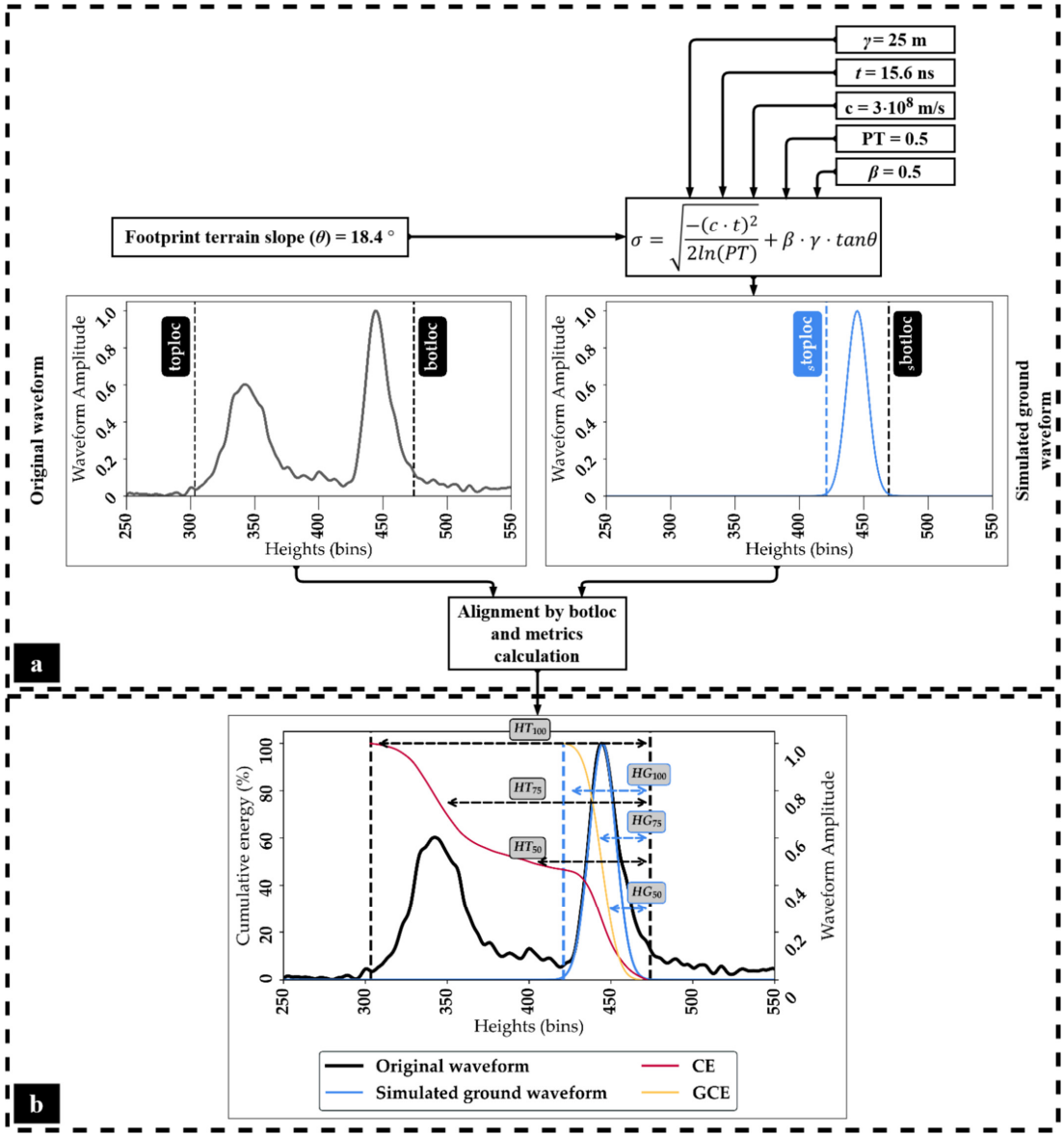

Figure 5. (**a**) Overview of the methodology for the simulation of the ground return as defined by Wang et al. [26] for a GEDI footprint acquired at 18.4° terrain slope. (**b**) The extracted canopy height metrics. CE represents the cumulative energy (from *botloc* to *toploc*) of the original waveform. GCE represents the cumulative energy of the simulated ground waveform. The dashed lines represent the *toploc* and *botloc* positions.

After the superposition of the original waveform and the simulated ground waveform, the ${}_sRHT_n$ metrics ("s" for simulated) can be generated, where ${}_sRHT_n = HT_n - {}_sHG_n$, and HT_n is the height between *botloc* and the position at n% energy of the original waveform, and ${}_sHG_n$ is the height between *botloc* and the position at n% energy of the simulated ground waveform (Figure 5b).

To estimate the AGB, Wang et al. [26] relied on a linear regression model using four ${}_sRHT_n$ (${}_sRHT_{25}$, ${}_sRHT_{50}$, ${}_sRHT_{75}$, and ${}_sRHT_{100}$). However, since random forests are more accurate than linear regression models for the estimation of forest characteristics (e.g., canopy heights, AGB) [7], in this study, ${}_sRHT_n$ and ${}_sHG_n$ will be used in a random forest regression model using nine ${}_sRHT_n$ and nine ${}_sHG_n$ values ($20\% \leq n \leq 100\%$, $step10\%$).

The method proposed by Wang et al. [26] relies on slope information from the SRTM DEM which has a resolution of 30 m while the diameter of GEDI is 25 m. Therefore, in order to analyze the pertinence of the Wang method, we propose a methodology that relies on the same RHT_n metrics, but instead of simulating a ground return, the ${}_sHG_n$ metrics

will be calculated from the ground return of the original waveform (henceforth referred to as ${}_fHG_n$, "f" for fitted). The original ground return will be fitted by means of an automated Gaussian decomposition of the original waveform [37]. Figure 6 shows the difference over the ground return between a simulated ground return, and fitted ground return.

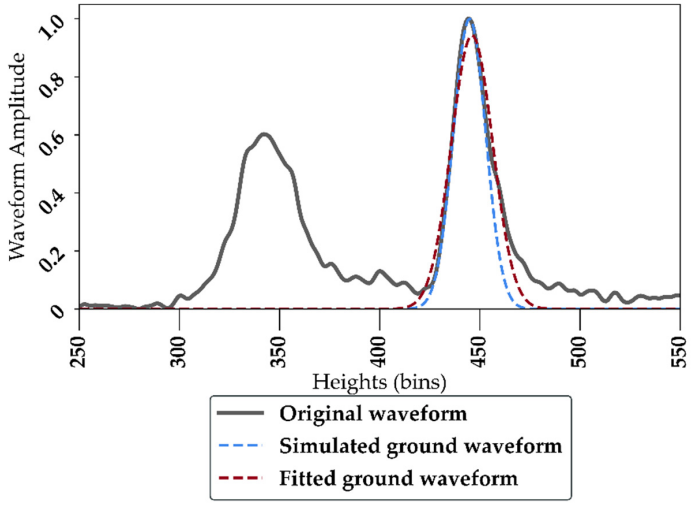

Figure 6. Difference between a simulated ground waveform (dashed blue line) and fitted ground waveform (dashed red line) over an Eucalyptus stand, with a terrain slope of 18.36° (33.2%).

A summary of the models that will be tested for the estimation of H_{dom} are presented in Table 2. For the random forest-based models, they are built using a set of 500 trees (higher tree count slightly increased the model accuracy), with a tree depth equal to the square root of the number of available factors. Model performance is assessed using a 5-fold cross validation, and the Eucalyptus stands used for training or validation were selected randomly regardless of the terrain slope. In addition, we imposed that footprints belonging to the same stand were assigned exclusively to one of the data partitions (train or test) with the aim to avoid possible non-independence of the data due to spatial proximity in the evaluation procedure. Finally, model performances are evaluated by means of the coefficient of determination (R^2), the root mean square error (RMSE), the bias (difference between estimated and in situ variables), and the root mean squared percentage error (*RMSPE*). *RMSPE* is defined as:

$$RMSPE = 100 \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{Y_t - Y_p}{Y_t} \right)^2} \quad (7)$$

where Y_t is the measured variable and Y_p is the predicted variable.

Table 2. List of the models used for the estimation of H_{dom}.

ID	Metrics Used	Model
MH1	W_{ext}, S	$H_{dom} = a \cdot W_{ext} - b \cdot S + c$
MH2	RH_{100}, S	$H_{dom} = a \cdot RH100 - b \cdot S + c$
RFH_{RH}	$RH_n\{10\% \leq n \leq 100\%, step10\%\}$, Slope ($S$), and terrain roughness ($ROUG$)	Random Forests
${}_sRFH_{RHT+HG}$	${}_sRHT_n\{20\% \leq n \leq 100\%, step10\%\}$ ${}_sHG_n\{20\% \leq n \leq 100\%, step10\%\}$	Random Forests
${}_fRFH_{RHT+HG}$	${}_fRHT_n\{20\% \leq n \leq 100\%, step10\%\}$ ${}_fHG_n\{20\% \leq n \leq 100\%, step10\%\}$	Random Forests

3.2. Wood Volume Estimation

Four models will be tested for the estimation of wood volume. The first model is adapted from Saatchi et al. [38] and uses a power law relationship between the volume and Lorey's height. The model will also use the terrain slope (S) in order to compensate for the terrain slope effect:

$$V = aH_L{}^b + c \cdot S + d \quad (8)$$

where H_L is Lorey's height and S is the terrain slope. In this study, the relationship defined in Equation (8) was used by replacing Lorey's height with RH_{100} (representing the dominant height H_{dom}) 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) [15].

The second model that will be tested is based on a random forest regressor using the relative height metrics $RH_n\{10\% \leq n \leq 100\%, step10\%\}$, as well as the terrain roughness ($ROUG$), and terrain slope (S).

Finally, the metrics generated by Wang et al. [26] will be used in a random forest regressor in order to estimate the wood volume. Two sets of metrics will be used: (1) the ${}_sRHT_n$ and ${}_sHG_n$ as defined by Wang et al. [26], and (2) the ${}_fRHT_n$ and ${}_fHG_n$ which rely on a Gaussian decomposition. A summary of the tested models for the estimation of wood volume is presented in Table 3.

Table 3. List of the models used for the estimation of wood volume (V).

ID	Metrics Used	Model
MV1	RH100, S	$V = aRH100^b + c \cdot S + d$
RFV_{RH}	$RH_n\{10\% \leq n \leq 100\%, step10\%\}$, slope ($S$), and terrain roughness ($ROUG$)	Random Forests
${}_sRFV_{RHT+HG}$	${}_sRHT_n\{20\% \leq n \leq 100\%, step10\%\}$ ${}_sHG_n\{20\% \leq n \leq 100\%, step10\%\}$	Random Forests
${}_fRFV_{RHT+HG}$	${}_fRHT_n\{20\% \leq n \leq 100\%, step10\%\}$ ${}_fHG_n\{20\% \leq n \leq 100\%, step10\%\}$	Random Forests

4. Results

4.1. Estimation of Dominant Stand Heights (H_{dom})

The results presented in Figure 7 and Table 4 show the accuracy of the estimation of H_{dom} from the models presented in Table 2 over three slope ranges (0–10%, 10–20%, and between 20 and 45%). For slope ranges 0–10%, the accuracy of the estimation of H_{dom} using MH1 was the lowest with an RMSE of 2.06 m (R^2 = 0.81). For the remaining models, the RMSE of the H_{dom} estimates were similar and ranged between 1.35 m (R^2 = 0.93,RFH_{RH}) and 1.42 m (R^2 = 0.93, ${}_fRFH_{RHT+HG}$). For slope ranges 10–20%, the models did not show any decrease in performance due to slope, except for the Wang model (${}_sRFH_{RHT+HG}$) which had a 30 cm increase in RMSE (Table 4) and a 1% increase in RMSPE. For slopes higher than 20%, all models except for (${}_fRFH_{RHT+HG}$) had a decrease in accuracy with increased slopes. Indeed, for slopes higher than 20%, the RMSE ranged between 1.65 m (R^2 of 0.86, RFH_{RH}) and 3.26 m (R^2 of 0.45, Model 1) (Table 4). Moreover, the bias (estimated H_{dom}-in situ H_{dom}) was ~1.2 m using Models MH1 and MH2, and 0.65 m for (RFH_{RH}). The (${}_sRFH_{RHT+HG}$) and (${}_fRFH_{RHT+HG}$) models were the two models that did not show high sensitivity to terrain slopes (no bias) even for slopes higher than 20% (Figure 7, Table 4). Nonetheless, the model (${}_fRFH_{RHT+HG}$) was slightly more accurate for slope ranges higher than 10%, where the RMSPE on the estimation of H_{dom} remained 6% (RMSE of 1.26 m, R^2 of 0.92), with a slight bias of 0.11 m.

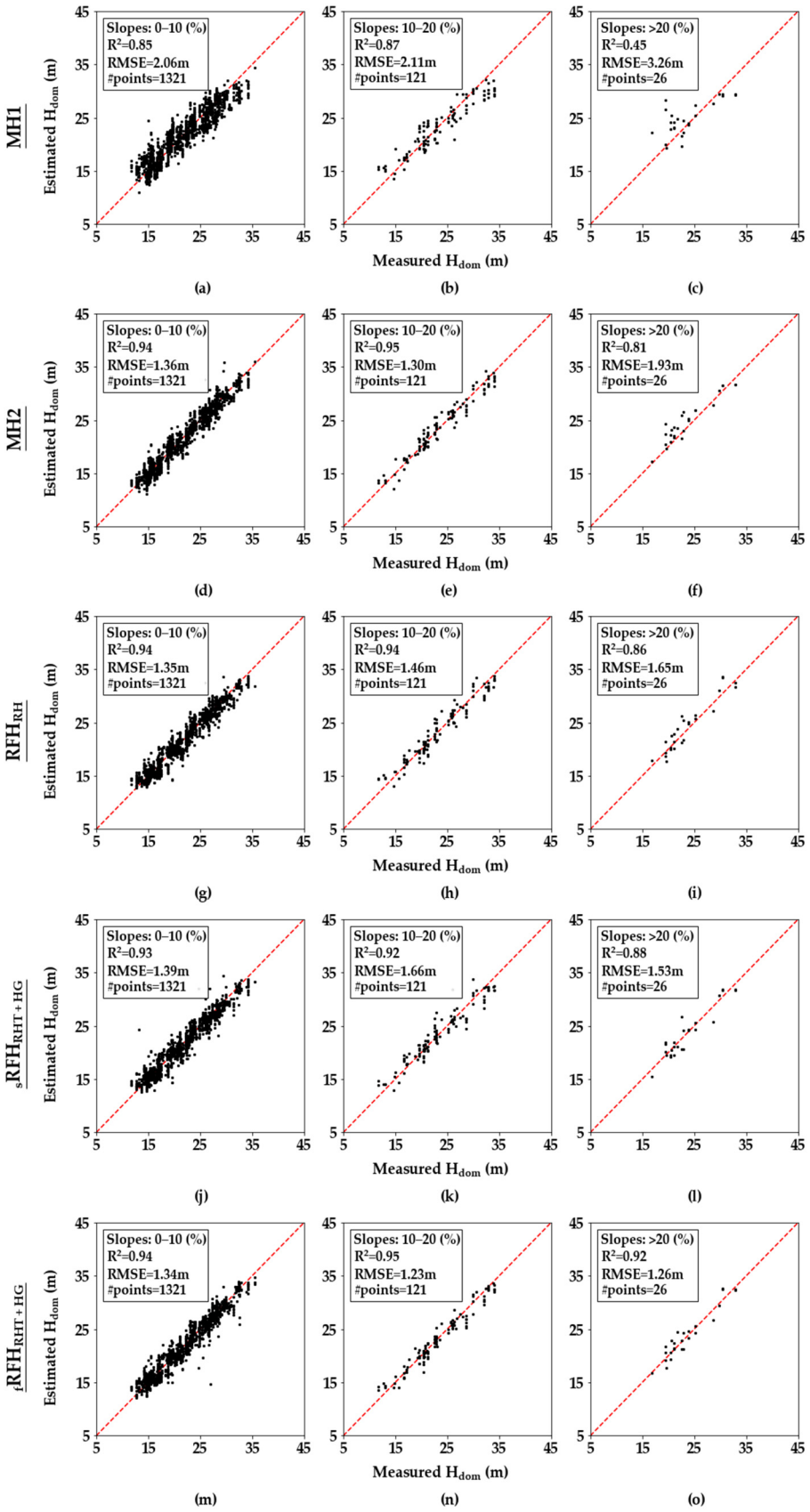

Figure 7. (**a**–**o**) Comparison of the measured vs. estimated dominant height from the models presented in Section 3.1 (Table 2) using GEDI metrics extracted with the processing configuration "a1" of GEDI data.

Table 4. Model performance for the estimation of Eucalyptus stand dominant height (defined in Section 3.1). Bias = estimated H_{dom}-in situ H_{dom}.

	Slope Ranges (%)											
	0–10				10–20				>20			
ID	RMSE (m)	RMSPE (%)	R^2	Bias (m)	RMSE (m)	RMSPE (%)	R^2	Bias (m)	RMSE (m)	RMSPE (%)	R^2	Bias (m)
MH1	2.06	11	0.85	0.07	2.11	11	0.87	−0.16	3.26	16	0.45	1.22
MH2	1.36	7	0.94	−0.08	1.30	7	0.95	0.15	1.93	9	0.81	1.23
RFH_{RH}	1.35	7	0.94	−0.04	1.46	7	0.94	0.05	1.65	7	0.86	0.65
${}_sRFH_{RHT+HG}$	1.39	6	0.93	−0.02	1.66	7	0.92	−0.01	1.53	8	0.88	−0.14
${}_fRFH_{RHT+HG}$	1.34	6	0.94	−0.02	1.34	6	0.95	−0.2	1.26	6	0.92	0.11

An analysis of the slope effects on the accuracy of H_{dom} is presented in Figure 8, which shows the variability of the difference between the estimated and in situ H_{dom}. The results shown in Figure 8 indicate that the slope effect on H_{dom} estimates using the models (MH2) and (RFH_{RH}) were not completely eliminated for slopes higher than 15%. Indeed, for both models, the median difference between estimated and in situ H_{dom} increased from −0.08 and −0.06 for, respectively, the models (MH2) and (RFH_{RH}) for slope ranges]10–15%] to 1.2 and 0.59 m for slope ranges higher than 20% (between 20 and 45%). In contrast, the slope effects on H_{dom} estimates using the Wang-based methodology (${}_sRFH_{RHT+HG}$, or ${}_fRFH_{RHT+HG}$) was greatly reduced, and the median difference between estimated and in situ H_{dom} ranged between −0.096 and 0.0.12 m for (${}_sRFH_{RHT+HG}$) and between −0.25 and 0.17 m for (${}_fRFH_{RHT+HG}$).

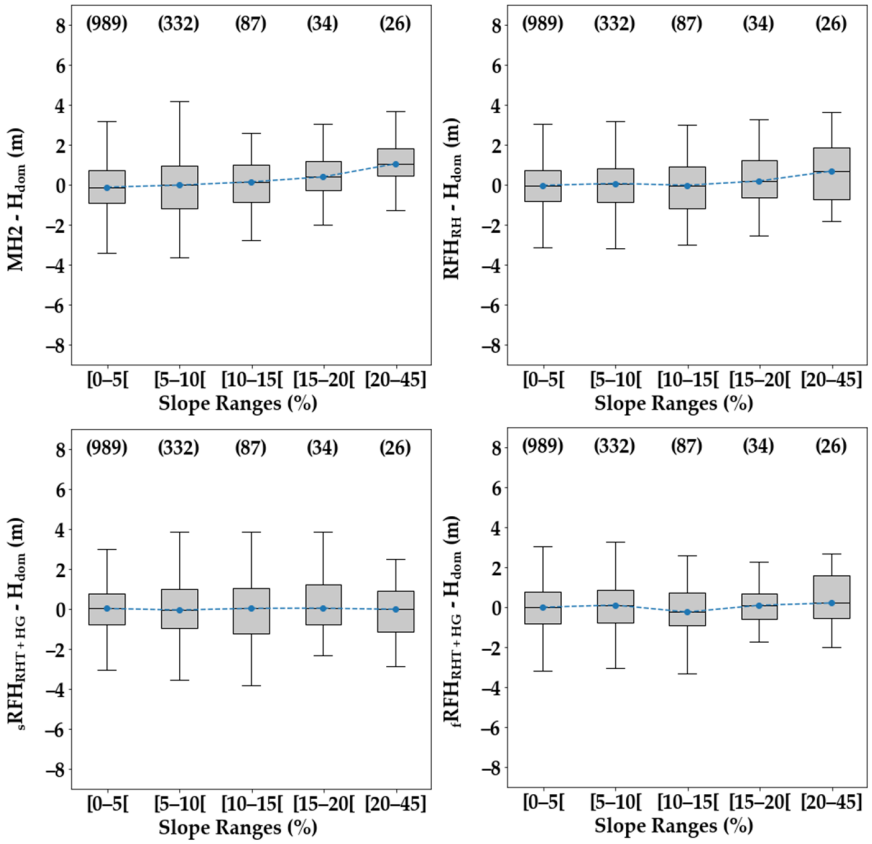

Figure 8. Box-whiskers plots of the difference between estimated and in situ H_{dom} using four estimation models across five slope ranges. Numbers in parenthesis represent the number of points in each slope class.

4.2. Estimation of Wood Volume (V)

Four models were tested for the estimation of the wood volume (V) in Eucalyptus stands. The results presented in Figure 9 and Table 5 show that the models (MV1) and (RFV_{RH}) are sensitive to slopes with increasing RMSE on the estimation of V (decrease of R^2) with increasing slope. For both models, the RMSE increased from about 27.5 $m^3 \cdot ha^{-1}$ (R^2~0.9) to about 48.7 $m^3 \cdot ha^{-1}$ (R^2~0.75) when the terrain slope increases from [0–10%] to slope values higher than 20% (between 20 and 45%). The mean difference between estimated V and in situ V (bias, Table 5) also decreased from, respectively, 0.21 $m^3 \cdot ha^{-1}$ and −1.60 $m^3 \cdot ha^{-1}$ for MV1 and RFV_{RH} for slopes between 0–10% to, respectively, 13.95 $m^3 \cdot ha^{-1}$ and 11.42 $m^3 \cdot ha^{-1}$ for MV1 and RFV_{RH} for slopes higher than 20%. The Wang-based methodology ($_sRFH_{RHT+HG}$) did not show an increased overestimation by increased slopes (bias between −2.66 and −5.91 $m^3 \cdot ha^{-1}$, Table 5). Nonetheless, the ($_sRFH_{RHT+HG}$) model showed lower accuracy for slopes higher than 10%. Indeed, for slopes between 0–10%, the RMSE of the estimation of V using the ($_sRFH_{RHT+HG}$) model was 27.57 $m^3 \cdot ha^{-1}$ (R^2 of 0.91) and decreased to 53.82 $m^3 \cdot ha^{-1}$ (R^2 of 0.79) for slopes between 10–20% and 49.12 $m^3 \cdot ha^{-1}$ (R^2 of 0.74) for slopes higher than 20%. Finally, the modified Wang model ($_fRFH_{RHT+HG}$) which relies on HG_n metrics derived from the fitted ground return of the waveforms, similarly to H_{dom} estimates, was the most accurate model. Indeed, the results presented in Figure 9 and Table 5 show that the estimation of V using the model ($_fRFH_{RHT+HG}$) was the most accurate in comparison to the three other models with an RMSE ranging between 26.78 $m^3 \cdot ha^{-1}$ (RMSPE of 20%, R^2 of 0.92) for slopes between 0–10% to 36.29 $m^3 \cdot ha^{-1}$ (RMSPE of 20%, R^2 of 36.29 $m^3 \cdot ha^{-1}$) for slopes higher than 20%. Moreover, the mean difference between the estimated V and in situ V using the model ($_fRFH_{RHT+HG}$) remained relatively stable with a mean difference of 1.32 $m^3 \cdot ha^{-1}$ for slopes between 0–10% to −2.65 $m^3 \cdot ha^{-1}$ for slopes higher than 20% (Table 5).

The variability of the difference between the estimated and in situ V for the four tested models across five slopes ranges is presented in Figure 10. As seen previously, MV1 and RFV_{RH} both show sensitivity to slopes higher than 10%. This is evident by the increased median difference between the estimated and in situ V. For the model (MV1, Figure 10), the median difference between the estimated and in situ V increased from −0.23 $m^3 \cdot ha^{-1}$ for slopes between 0–5% to 24.32 $m^3 \cdot ha^{-1}$ for slopes higher than 20%. Similarly, for the model (RFV_{RH}), the median difference between the estimated and in situ V increased from −1.53 $m^3 \cdot ha^{-1}$ for slopes between 0–5% to 20.20 $m^3 \cdot ha^{-1}$ for slopes higher than 20%. The models $_sRFV_{RHT+HG}$ and $_fRFV_{RHT+HG}$ were both insensitive to slopes with a median difference between the estimated and in situ V ranging from 0.58 $m^3 \cdot ha^{-1}$ (slopes ∈ [0–5]%) and 5.42 $m^3 \cdot ha^{-1}$ (slopes > 20%) for the model $_sRFV_{RHT+HG}$ and from 1.21 $m^3 \cdot ha^{-1}$ (slopes ∈ [0–5]%) and 2.16 $m^3 \cdot ha^{-1}$ (slopes > 20%) for the model $_fRFV_{RHT+HG}$ (Figure 10). Nonetheless, the model $_sRFV_{RHT+HG}$ showed higher variability on the estimates of V for slopes higher than 10% in comparison to the model $_fRFV_{RHT+HG}$ (Figure 10).

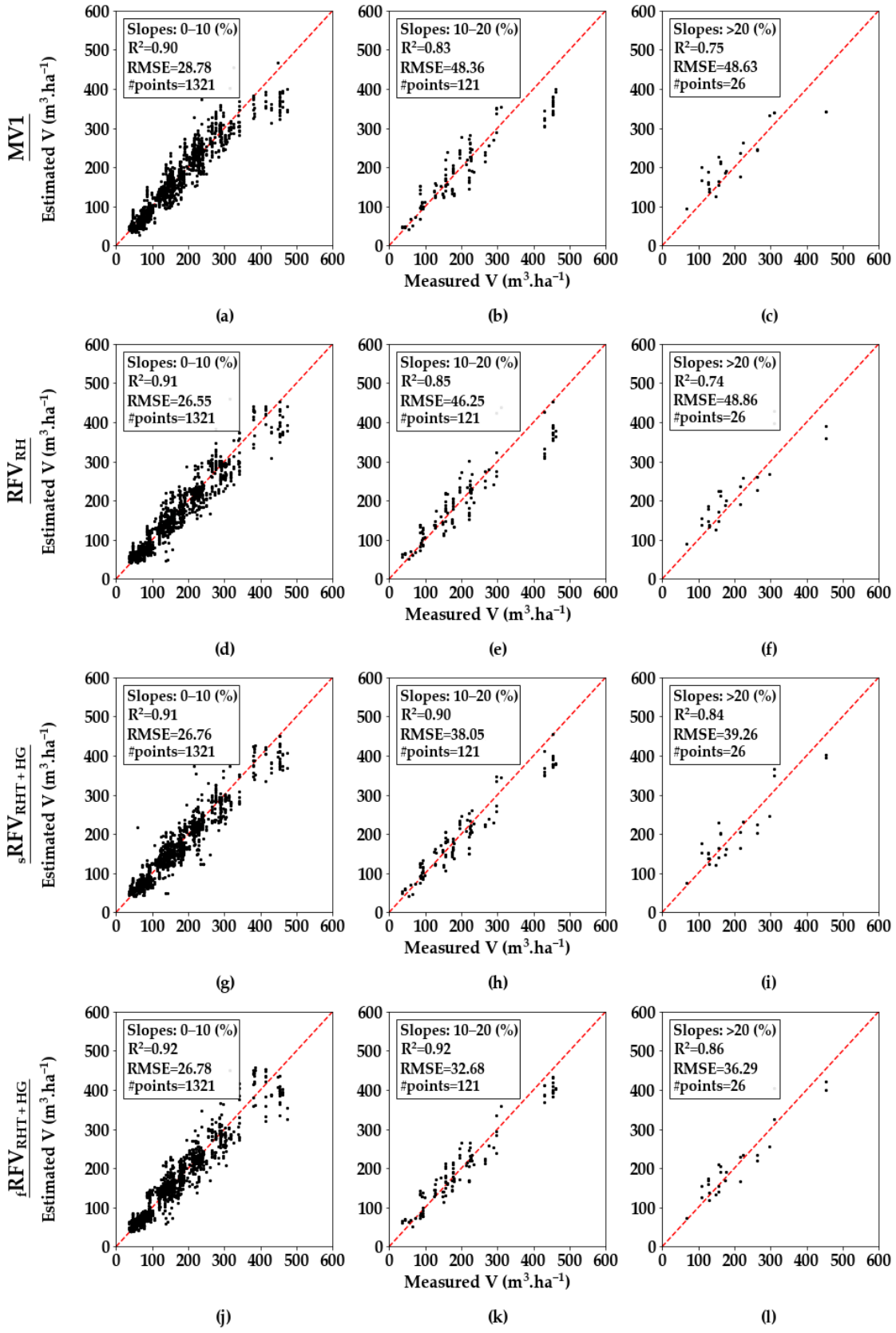

Figure 9. (**a**–**l**) Comparison of measured vs. estimated wood volume from the models presented in Section 3.2 using GEDI metrics extracted using algorithm a1. RMSE is expressed in $m^3 \cdot ha^{-1}$.

Table 5. Model performance for the estimation of Eucalyptus wood volume. The models are described in Section 3.2. Bias = estimated V-in situ V.

	Slope Ranges (%)											
	0–10				10–20				>20			
ID	RMSE ($m^3 \cdot ha^{-1}$)	RMSPE (%)	R^2	Bias ($m^3 \cdot ha^{-1}$)	RMSE ($m^3 \cdot ha^{-1}$)	RMSPE (%)	R^2	Bias ($m^3 \cdot ha^{-1}$)	RMSE ($m^3 \cdot ha^{-1}$)	RMSPE (%)	R^2	Bias ($m^3 \cdot ha^{-1}$)
MV1	28.78	19	0.90	0.21	48.36	22	0.83	9.15	48.63	29	0.75	13.95
RFV_{RH}	26.55	19	0.91	−1.60	46.25	23	0.85	9.45	48.86	24	0.74	11.42
${}_sRFV_{RHT+HG}$	26.76	20	0.91	−2.57	38.05	22	0.90	−5.16	39.26	22	0.86	−1.49
${}_fRFV_{RHT+HG}$	26.78	20	0.92	1.32	32.68	24	0.92	−2.22	36.29	20	0.86	2.65

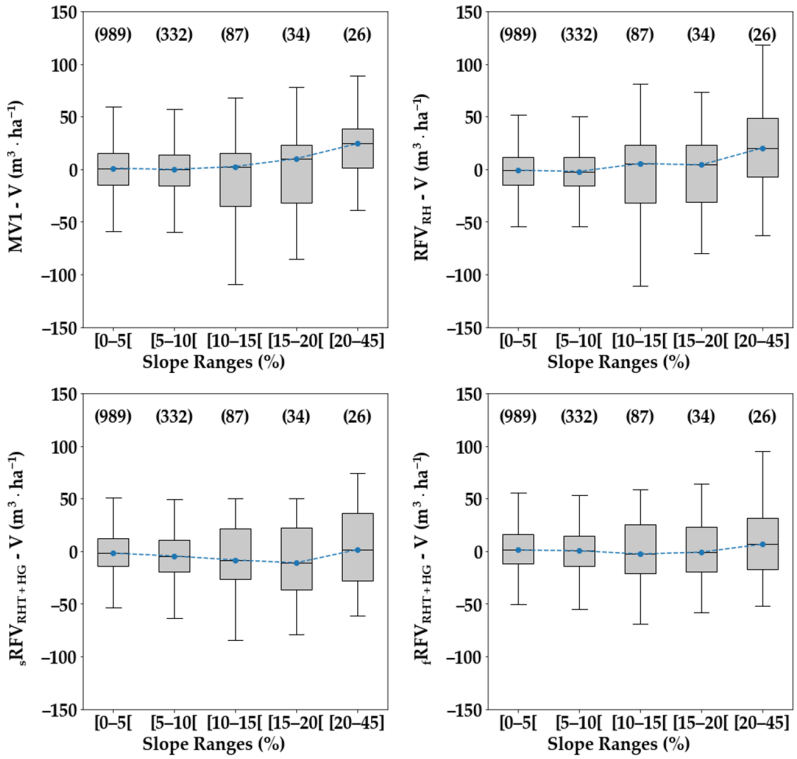

Figure 10. Box-whiskers plots of the difference between the estimated and in situ *V* using four estimation models across five slope ranges. Numbers in parenthesis represent the number of points in each slope class.

5. Discussion

The results presented in this study show that the models relying on the SRTM DEM generated terrain metrics provide a limited slope-effect correction of H_{dom} and wood volume (V) estimates from GEDI data, especially for high slope values (e.g., higher than 20%). Indeed, for MH2 and RFH_{RH} the effect of the slopes was minimal for slope ranges 0–20% and increased for slopes higher than 20%, with an increase on the RMSE of 63 cm for model 2 and 19 cm for the model RFH_{RH}. Nonetheless, the effect of the terrain slope was more pronounced on the estimation of V, and a decrease in accuracy (RMSE) was observed for slopes as low as 10% (e.g., for the RFV_{RH} model, the RMSE decreased from 26.55 $m^3 \cdot ha^{-1}$ for terrain slopes lower than 10% to more than 46 $m^3 \cdot ha^{-1}$ for terrain slopes higher than 10%). These results indicate that the SRTM DEM with its 30-m resolution is not adequate for the 25-m wide GEDI footprints. Therefore, a finer resolution DEM, for example 10 m, is required for the 25-m GEDI footprints. However, the results showed that for V, the R^2 remained high and that the highest volumes were underestimated for high values for all slope classes. The difference in RMSE between the slope classes may therefore come from the fact that there were proportionally more data with high V for the two slope classes 10–20% and 20–45% than for the 0–10% class.

The methodology proposed by Wang using metrics generated from either simulated or fitted ground returns provided the best results, as both approaches showed that the effects of slopes were minimized for all slope classes available in this study. Indeed, among all the generated metrics from GEDI waveforms, the ${}_sHG_n$ and ${}_fHG_n$ metrics were the only metrics that were independent from the slope. Figure 11 shows that the distribution of errors (GEDI-in situ) of RH_{100} (Figure 11a) and RHT_{100} (Figure 11b) was slope dependent, especially for slopes higher than 10% with an intercept of, respectively, ~0.04 and ~0.07. In contrast, the distribution of errors for ${}_sRHT_{100}$ and ${}_fRHT_{100}$ (Figure 11c,d) was constant across all slope gradients (intercept ≈ 0.003). Nonetheless, the estimation of H_{dom} and V using ${}_sRHT_n$ and ${}_sHG_n$ was slightly less accurate than ${}_fRHT_n$ and ${}_fHG_n$. The uncertainties on the estimation of both variables using the ${}_sRHT_n$ and ${}_sHG_n$ metrics could be attributed to two factors. Firstly, the simulated ground return used to calculate the ${}_sRHT_n$ and ${}_sHG_n$ is dependent on the slope of the studied waveform, which is calculated from the SRTM DEM. Given the 30-m resolution of the SRTM DEM, and the 25-m footprint diameter, terrain information within the footprint could not be accurately calculated. This is evident in Figure 12 which compares the full width at half maximum (FWHM) of the simulated and fitted ground returns. The results presented in Figure 12 show that the FWHM of both the simulated and fitted ground returns are highly correlated for slopes lower than 10% but for slopes higher than 10%, the FWHM of the simulated ground returns are lower than the FWHM of the fitted ground returns, leading to the uncertainties on the estimation of H_{dom} and V. Secondly, in the formulation of Wang et al. [26], the ground returns were simulated using a symmetrical Gaussian function [39,40] assuming a plane slope within the 25-m footprint. Nevertheless, this formulation is not always satisfactory as the return echo components recorded by FW LiDAR systems are asymmetrical. Thus, the fitting process has lower accuracy when the echoed asymmetrical components are fitted using a symmetrical Gaussian function. Moreover, GEDI waveforms display a sharp rising part and a slower descending one. As such, a lognormal function, which is characterized by a short rise time and a tailing, could be a better fit for the simulation of the ground echo than a Gaussian.

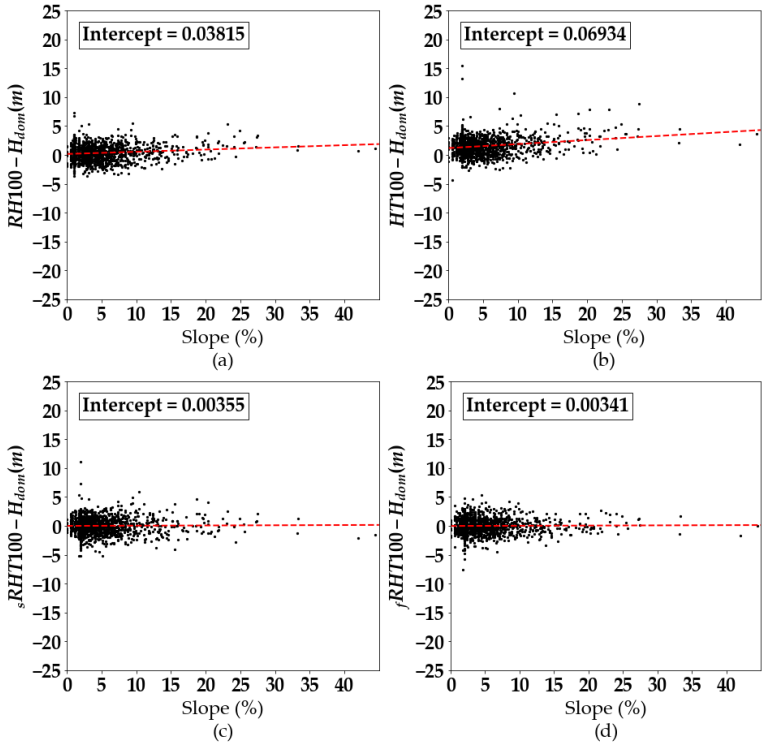

Figure 11. (**a**–**d**) Distribution of several GEDI metric height errors by slope gradient.

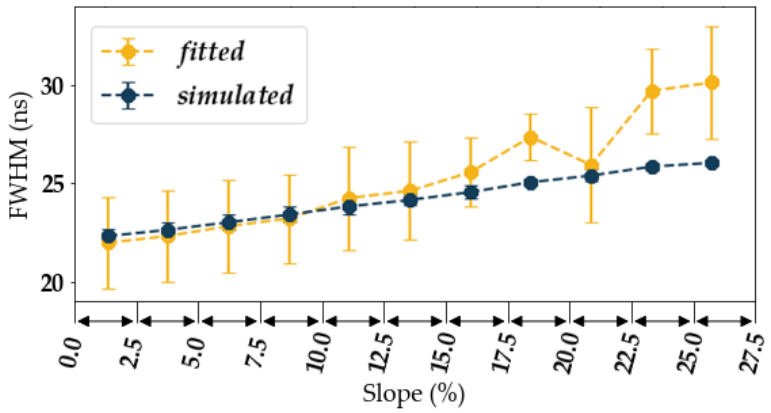

Figure 12. Difference between the mean FWHM of simulated and fitted ground return by slope class. The vertical bars represent one standard deviation.

Finally, while the metrics proposed by Wang et al. [26] seem to greatly minimize the effect of slopes over our study area, there are several uncertainties still unaccounted for. First, within the Eucalyptus stands, tree heights are very homogeneous (i.e., the tree height is evenly distributed within the footprints). However, it was reported by Hyde et al. [41] that the uneven distribution of canopy structure could increase the uncertainty on the estimation of canopy structure with FW LiDAR as the dominant trees at the edge of the LiDAR footprint might not be detectable due to the lower laser energy at the edge of the footprint in comparison to its center. Therefore, in a future work, the metrics proposed by Wang et al. [26] should be assessed with GEDI acquisitions over natural forests. Another source of uncertainty could be present for slopes higher than 20%. In this study, there were very few acquisitions with slopes higher than 20%, and while the slope effects were greatly reduced for slope ranges 0–20%, it is necessary to analyze the pertinence of the Wang et al. [26] methodology for high slopes.

6. Conclusions

In this study, several approaches were tested in order to minimize the uncertainties due to the presence of slopes on the estimation of forest canopy heights and wood volume from GEDI data. The tested approaches can be classified into two groups. The first group are methods incorporating traditional GEDI waveform metrics (e.g., W_{ext}, RH_n) and ancillary SRTM DEM data (e.g., terrain slope *S*, surface roughness *ROUG*), while the second group is based solely on new GEDI waveform metrics. The results showed that the methods relying on ancillary SRTM DEM data provided limited correction capabilities for slopes higher than 20% and this for both canopy heights and wood volume estimates. Indeed, the random forest regression model (RFH_{RH}) using the relative height metrics (RH_n) as well as the *S* and *ROUG* variables extracted from the SRTM DEM presented an increase of 14% in terms of RMSE (8% decrease in R^2) on the H_{dom} estimates for acquisitions over terrain slopes between 10–20% and slopes higher than 20%. Moreover, the same model when used to estimate the wood volume (RFV_{RH}) showed a decrease in accuracy (increase in RMSE) for slopes higher than 10%. Indeed, for the model RFV_{RH} an RMSE increase on the estimation of the wood volume of 74% was reported between GEDI acquisitions with terrain slopes between 0–10% and GEDI acquisitions with terrain slopes higher than 10%. These results indicate that the 30-m resolution SRTM DEM is not suitable for the 25-m wide GEDI footprints.

Next, building on the model of Wang et al. [26], two sets of metrics were generated for GEDI waveforms, the first set of metrics ($_sRHT_n$ and $_sHG_n$) were generated using a simulated ground return that varied based on the slope of the GEDI footprint. The second set of metrics ($_fRHT_n$ and $_fHG_n$) were generated using a fitted Gaussian from the ground return. Estimation approaches of H_{dom} and V using these sets of metrics and the Random Forest technique provided the most accurate estimates of H_{dom} and V for all

terrain slope ranges. The ($_sRHT_n$ and $_sHG_n$)-based model showed an RMSE that ranged between 1.39 and 1.66 m (between 26.76 and 39.26 $m^3 \cdot ha^{-1}$ for V) while the ($_fRHT_n$ and $_fHG_n$)-based method showed an RMSE that ranged between 1.26 and 1.34 m (between 26.78 and 36.29 $m^3 \cdot ha^{-1}$ for V). Moreover, the dependency of the GEDI metrics on slopes (e.g., intercept of 0.069 for HT_{100}) was greatly reduced for the two set of metrics (e.g., intercept of $\approx$ 0.003 for both $_sRHT_{100}$ and $_fRHT_{100}$). Nonetheless, the model based on the ($_sRHT_n$ and $_sHG_n$) performed slightly worse than the ($_fRHT_n$ and $_fHG_n$)-based model for the estimation of both forest variables. The decrease in accuracy of the ($_sRHT_n$ and $_sHG_n$)-based model is due to the use of the 30-m SRTM DEM (the only available global DEM) for the 25-m GEDI footprints. On the other hand, the ($_fRHT_n$ and $_fHG_n$) metrics rely on the presence of a distinct ground return in the received waveform which might not always be detectable over high sloping terrain. Therefore, it is recommended that for the estimation of H_{dom} and V over moderately sloping terrain (i.e., presence of distinct ground peak) the ($_fRHT_n$ and $_fHG_n$) metrics should be used, while the ($_sRHT_n$ and $_sHG_n$) metrics should be used for high sloping terrain (i.e., undetectable ground return).

Finally, the effect of slopes on the 25-m GEDI footprints is rather low as the estimation on canopy heights degraded by a maximum of 1 m for slopes between 20 and 45%. In regard to the wood volume estimation, the effect of slopes was more pronounced, and a degradation on the accuracy (increased RMSE) of a maximum of 20 $m^3 \cdot ha^{-1}$ was observed for slopes between 20 and 45%.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, data Curation, visualization, writing—original draft: I.F. Conceptualization, methodology, validation, formal analysis, data curation, writing—original draft: N.B. Conceptualization, validation, writing—review and editing: C.A.A. Conceptualization, validation, writing—review and editing: J.L.S. Validation, writing—review and editing: J.S.B. Conceptualization, validation, writing—review and editing: H.F.S. Validation, writing—review and editing: M.Z. Software, validation: I.R.C. Conceptualization, methodology, validation, writing—review and editing: G.L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2021 project) and the National Research Institute for Agriculture, Food and the Environment (INRAE). Suzano SA Company supported the forest-field data collection.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: GEDI data were downloaded free of charge from the Land Processes Distributed Active Archive Center.

Acknowledgments: 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, Renan Tarenta Meirelles Brasil and Carla Foster Feria for their technical support and the CIRAD 2020 Suzano project.

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

References

1. Zribi, M.; Guyon, D.; Motte, E.; Dayau, S.; Wigneron, J.P.; Baghdadi, N.; Pierdicca, N. Performance of GNSS-R GLORI Data for Biomass Estimation over the Landes Forest. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *74*, 150–158. [CrossRef]
2. Nelson, R.; Ranson, K.J.; Sun, G.; Kimes, D.S.; Kharuk, V.; Montesano, P. Estimating Siberian Timber Volume Using MODIS and ICESat/GLAS. *Remote Sens. Environ.* **2009**, *113*, 691–701. [CrossRef]
3. Dubayah, R.O.; Sheldon, S.L.; Clark, D.B.; Hofton, M.A.; Blair, J.B.; Hurtt, G.C.; Chazdon, R.L. Estimation of Tropical Forest Height and Biomass Dynamics Using Lidar Remote Sensing at La Selva, Costa Rica: Forest dynamics using lidar. *J. Geophys. Res.* **2010**, *115*, G2. [CrossRef]
4. Ploton, P.; Pélissier, R.; Barbier, N.; Proisy, C.; Ramesh, B.R.; Couteron, P. Canopy Texture Analysis for Large-Scale Assessments of Tropical Forest Stand Structure and Biomass. In *Treetops at Risk*; Lowman, M., Devy, S., Ganesh, T., Eds.; Springer: New York, NY, USA, 2013; pp. 237–245; ISBN 978-1-4614-7160-8.

5. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Vaglio Laurin, G.; Saah, D. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *Int. J. For. Res.* **2012**, *2012*, 436537. [CrossRef]
6. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. Lidar Remote Sensing for Ecosystem Studies. *BioScience* **2002**, *52*, 19. [CrossRef]
7. Fayad, I.; Baghdadi, N.; Alcarde, C.; Stape, J.-L.; Bailly, J.S.; Scolforo, H.F.; Zribi, M.; le Maire, G. 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 Obs. Remote Sens.* **2021**. (under revision).
8. Anderson, K.; Hancock, S.; Disney, M.; Gaston, K.J. Is Waveform Worth It? A Comparison of LiDAR Approaches for Vegetation and Landscape Characterization. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 5–15. [CrossRef]
9. Alexander, C.; Tansey, K.; Kaduk, J.; Holland, D.; Tate, N.J. Backscatter Coefficient as an Attribute for the Classification of Full-Waveform Airborne Laser Scanning Data in Urban Areas. *ISPRS* **2010**, *65*, 423–432. [CrossRef]
10. Sumnall, M.J.; Hill, R.A.; Hinsley, S.A. Comparison of Small-Footprint Discrete Return and Full Waveform Airborne Lidar Data for Estimating Multiple Forest Variables. *Remote Sens. Environ.* **2016**, *173*, 214–223. [CrossRef]
11. Chen, Q. Retrieving Vegetation Height of Forests and Woodlands over Mountainous Areas in the Pacific Coast Region Using Satellite Laser Altimetry. *Remote Sens. Environ.* **2010**, *114*, 1610–1627. [CrossRef]
12. Schutz, B.E.; Zwally, H.J.; Shuman, C.A.; Hancock, D.; DiMarzio, J.P. Overview of the ICESat Mission. *Geophys. Res. Lett.* **2005**, *32*, L21S01. [CrossRef]
13. Lefsky, M.A.; Harding, D.J.; Keller, M.; Cohen, W.B.; Carabajal, C.C.; Del Bom Espirito-Santo, F.; Hunter, M.O.; de Oliveira, R. Estimates of Forest Canopy Height and Aboveground Biomass Using ICESat: ICESat estimates of canopy height. *Geophys. Res. Lett.* **2005**, *32*. [CrossRef]
14. Fayad, I.; Baghdadi, N.; Guitet, S.; Bailly, J.-S.; Hérault, B.; Gond, V.; El Hajj, M.; Tong Minh, D.H. Aboveground Biomass Mapping in French Guiana by Combining Remote Sensing, Forest Inventories and Environmental Data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 502–514. [CrossRef]
15. Baghdadi, N.; le Maire, G.; Fayad, I.; Bailly, J.S.; Nouvellon, Y.; Lemos, C.; Hakamada, R. 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 Obs. Remote Sens.* **2014**, *7*, 290–299. [CrossRef]
16. Boudreau, J.; Nelson, R.; Margolis, H.; Beaudoin, A.; Guindon, L.; Kimes, D. Regional Aboveground Forest Biomass Using Airborne and Spaceborne LiDAR in Québec. *Remote Sens. Environ.* **2008**, *112*, 3876–3890. [CrossRef]
17. El Hajj, M.; Baghdadi, N.; Labrière, N.; Bailly, J.-S.; Villard, L. Mapping of Aboveground Biomass in Gabon. *Comptes Rendus Geosci.* **2019**, *351*, 321–331. [CrossRef]
18. Pourrahmati, M.R.; Baghdadi, N.N.; Darvishsefat, A.A.; Namiranian, M.; Fayad, I.; Bailly, J.-S.; Gond, V. Capability of GLAS/ICESat Data to Estimate Forest Canopy Height and Volume in Mountainous Forests of Iran. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 5246–5261. [CrossRef]
19. Rajab Pourrahmati, M.; Baghdadi, N.; Darvishsefat, A.A.; Namiranian, M.; Gond, V.; Bailly, J.-S.; Zargham, N. Mapping Lorey's Height over Hyrcanian Forests of Iran Using Synergy of ICESat/GLAS and Optical Images. *Eur. J. Remote Sens.* **2018**, *51*, 100–115. [CrossRef]
20. Harding, D.J. ICESat Waveform Measurements of Within-Footprint Topographic Relief and Vegetation Vertical Structure. *Geophys. Res. Lett.* **2005**, *32*, L21S10. [CrossRef]
21. Duncanson, L.I.; Niemann, K.O.; Wulder, M.A. Estimating Forest Canopy Height and Terrain Relief from GLAS Waveform Metrics. *Remote Sens. Environ.* **2010**, *114*, 138–154. [CrossRef]
22. Pang, Y.; Lefsky, M.; Andersen, H.-E.; Miller, M.E.; Sherrill, K. 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.* **2008**, *34*, S471–S484. [CrossRef]
23. Xing, Y.; de Gier, A.; Zhang, J.; Wang, L. An Improved Method for Estimating Forest Canopy Height Using ICESat-GLAS Full Waveform Data over Sloping Terrain: A Case Study in Changbai Mountains, China. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 385–392. [CrossRef]
24. Yang, W.; Ni-Meister, W.; Lee, S. Assessment of the Impacts of Surface Topography, off-Nadir Pointing and Vegetation Structure on Vegetation Lidar Waveforms Using an Extended Geometric Optical and Radiative Transfer Model. *Remote Sens. Environ.* **2011**, *115*, 2810–2822. [CrossRef]
25. Ni-Meister, W.; Jupp, D.L.; Dubayah, R. Modeling Lidar Waveforms in Heterogeneous and Discrete Canopies. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1943–1958. [CrossRef]
26. Wang, Y.; Ni, W.; Sun, G.; Chi, H.; Zhang, Z.; Guo, Z. Slope-Adaptive Waveform Metrics of Large Footprint Lidar for Estimation of Forest Aboveground Biomass. *Remote Sens. Environ.* **2019**, *224*, 386–400. [CrossRef]
27. Næsset, E. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [CrossRef]
28. Lee, S.; Ni-Meister, W.; Yang, W.; Chen, Q. Physically Based Vertical Vegetation Structure Retrieval from ICESat Data: Validation Using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sens. Environ.* **2011**, *115*, 2776–2785. [CrossRef]

29. Neuenschwander, A.; Pitts, K. The ATL08 Land and Vegetation Product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259. [CrossRef]
30. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [CrossRef]
31. Gonçalves, J.L.d.M.; Rocha, J.H.T.; Alvares, C.A. Manejo do solo em sistemas de cultivo de Eucalipto e Pinus. In *Manejo e Conservação do Solo e da Água*; 2019; Volume 1, pp. 1081–1117. Available online: https://www.researchgate.net/publication/339782827_Manejo_do_solo_em_sistemas_de_cultivo_de_Eucalipto_e_Pinus (accessed on 28 May 2021).
32. Dubayah, S.L. GEDI L1B Geolocated Waveform Data Global Footprint Level V001. 2020. Available online: https://doi.org/10.5067/DOC/GEDI/GEDI_WFGEO_ATBD.001 (accessed on 27 May 2021).
33. Dubayah, S.L. GEDI L2A Elevation and Height Metrics Data Global Footprint Level V001. 2020. Available online: https://doi.org/10.5067/DOC/GEDI/GEDI_WF_ATBD.001 (accessed on 27 May 2021).
34. Dubayah, S.L. GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V001. 2020. Available online: https://doi.org/10.5067/DOC/GEDI/GEDI_FCCVPM_ATBD.001 (accessed on 27 May 2021).
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
36. Ranson, K.J.; Sun, G. Modeling Lidar Returns from Forest Canopies. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 2617–2626. [CrossRef]
37. Riener, M.; Kainulainen, J.; Henshaw, J.D.; Orkisz, J.H.; Murray, C.E.; Beuther, H. GAUSSPY+: A Fully Automated Gaussian Decomposition Package for Emission Line Spectra. *A&A* **2019**, *628*, A78. [CrossRef]
38. Saatchi, S.S.; Harris, N.L.; Brown, S.; Lefsky, M.; Mitchard, E.T.A.; Salas, W.; Zutta, B.R.; Buermann, W.; Lewis, S.L.; Hagen, S.; et al. Benchmark Map of Forest Carbon Stocks in Tropical Regions across Three Continents. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9899–9904. [CrossRef] [PubMed]
39. Wagner, W.; Ullrich, A.; Ducic, V.; Melzer, T.; Studnicka, N. Gaussian Decomposition and Calibration of a Novel Small-Footprint Full-Waveform Digitising Airborne Laser Scanner. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 100–112. [CrossRef]
40. Lin, Y.-C.; Mills, J.P.; Smith-Voysey, S. Rigorous Pulse Detection from Full-Waveform Airborne Laser Scanning Data. *Int. J. Remote Sens.* **2010**, *31*, 1303–1324. [CrossRef]
41. Hyde, P.; Dubayah, R.; Peterson, B.; Blair, J.; Hofton, M.; Hunsaker, C.; Knox, R.; Walker, W. Mapping Forest Structure for Wildlife Habitat Analysis Using Waveform Lidar: Validation of Montane Ecosystems. *Remote Sens. Environ.* **2005**, *96*, 427–437. [CrossRef]