ESTIMATING CANOPY HEIGHT AND WOOD VOLUME OF EUCALYPTUS PLANTATIONS IN BRAZIL USING GEDI LIDAR 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], Mehrez Zribi[6], Guerric Le Maire[7,8]

[1] CIRAD, CNRS, INRAE, TETIS, Univ Montpellier, AgroParisTech, 34093 Montpellier CEDEX 5, France (e-mail : nicolas.baghdadi@teledetection.fr)
[2] Unesp, Faculdade de Ciências Agronômicas, 18610-034, Botucatu-SP, Brazil (e-mail : jlstape@gmail.com; calcarde@suzano.com.br)
[3] Suzano SA, Estrada Limeira, 391, 13465-970, Limeira, SP, Brazil (e-mail : hscolforo@suzano.com.br)
[4] INRAE, IRD, Institut Agro, LISAH, Univ Montpellier, 34060 Montpellier CEDEX 1, France ;
[5] AgroParisTech, 75005 Paris, France (e-mail: bailly@agroparistech.fr)
[6] CESBIO (CNRS/UPS/IRD/CNES/INRAE), 18 av. Edouard Belin, bpi 2801, 31401 Toulouse CEDEX 9, France (e-mail: mehrez.zribi@ird.fr)
[7] CIRAD, UMR Eco&Sols, F-34398, Montpellier, France (e-mail: guerric.le_maire@cirad.fr)
[8] Eco&Sols, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France

ABSTRACT

Full waveform (FW) LiDAR systems have gained momentum to map forest biophysical variables in the last two decades, owing to their ability to accurately estimate canopy heights and aboveground biomass. Currently, the Global Ecosystem Dynamics Investigation (GEDI) system on board of the International Space Station (ISS) is the most recent FW spaceborne LiDAR instrument for the continuous observation of earth's forests. Here, we assess the accuracy of GEDI FW data for the estimation of stand-scale dominant heights (H_{dom}), and stand volume (V) using linear and non-linear regression models based on several GEDI metrics. The models were calibrated and validated using in-situ data from Eucalyptus plantations in Brazil. Overall, the most accurate estimates of H_{dom} and V were obtained using the stepwise regression, with an RMSE of 1.44 m (R^2 of 0.92) 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.

Index Terms— Lidar, GEDI, Dominant height, Wood volume, Eucalyptus, Brazil

1. INTRODUCTION

The primary source of forest characteristics estimation in tropical forests has become through the use of observations and measurements from different satellite remote sensing platforms in the last couple of decades. Generally, methods based on remotely sensed data, such as from optical imagery, radar, and LiDAR, are less accurate, however, their major advantage is their global coverage and low or free acquisition costs for the end user. And while many studies have been conducted on the use of optical and radar technologies for forest biomass estimation, both technologies suffer from saturation at even medium biomass levels ([1], [2]). In contrast, LiDAR, which is an active optical sensor doesn't suffer from signal saturation problem in high-biomass forests, and it can therefore be used for accurate estimates of forest AGB [3]. The Geoscience Laser Altimeter System (GLAS) on board NASA's Ice Cloud and land Elevation Satellite (ICESat) was the first spaceborne LiDAR system capable of measuring land surfaces. GLAS's main objective was to measure changes in polar ice. However, given that GLAS recorded the changes in laser energy returned from the Earth's surface as a waveform, each recorded waveform over forests contained information on canopy structures, and the waveforms have been successfully used in many studies to estimate many forest variables, including canopy heights and above ground biomass [4].

The most recent spaceborne LiDAR system is GEDI on board the ISS, which launched in December 2018, with on-orbit checkout in 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 [5].

GEDI is comprised of three lasers emitting 1064 nm light, with a firing rate of 242 Hz. One of the lasers' output is split into two beams (half the power of the full laser), called coverage beams, while the other two lasers remain at full power. These four beams are dithered across track to produce eight tracks of data. The 8 produced tracks, henceforth referred to as beams, are separated by ~600 m across-track, with a footprint diameter of ~25 m and a distance between footprint centers of 60 m along-track [5].

GEDI measures vertical structures similarly to GLAS (i.e. waveforms); however, it presents similar improvements over its predecessor. (1) GEDI has a much higher firing rate (242 vs 40 Hz for GLAS), and since it acquires data using four beams, GEDI has a much higher acquisition density in comparison to GLAS. (2) GEDI has a much smaller footprint size (~25 vs ~60 m for GLAS); therefore, GEDI should provide improved measurements over forested areas.

The main objective of this work is to analyze the accuracy on the estimation of both, canopy height and wood volume of Eucalyptus plantations in Brazil from previously established linear regression and random forest regression models using metrics extracted from GEDI waveforms and digital elevation models. To this extent, an in depth analysis of these newly released full waveform LiDAR data will be made. First, by exploring which combination of metrics yields the best forest parameter estimates, and secondly by analyzing the effects of the pre-processing algorithms, with which the metrics have been generated, on the estimation accuracy.

2. DATASET DESCRIPTION

The study area is distributed over five states in Brazil (MatoGrosso do Sul (MS), São Paulo (SP), Espírito Santo (ES), Bahia (BA), and Maranhão (MA)) across a large latitudinal gradient, and covering different climate and soil types (Fig. 1). The study focuses on fast-growing, industrial Eucalyptus plantations by the Brazilian company Suzano Papel e Celulose. The studied plantations are managed in order to produce high yield pulpwood growing at short rotations, with tree densities between 1000 - 1667 trees/ha. Harvest occurs every six to seven years, and very little tree mortality (under 7% from original plantation) are noticed. The annual productivity of the plantations was on average 40 m3/ha/year, with 80% of the stand being between 30-50 m3/ha/year and some stand could reach values as high as 60 m3/ha/year. At harvest time, the stand volume is between 180 and 300 m3/ha, with a dominant height in the 20-35 m range (for 80% of the stands). Tree height is very homogeneous within a stand, with only few dominated Eucalyptus trees that experienced lower growth speed at early growth stages and remained small throughout the whole rotation. The plantation 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. 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.

In order to use only the reliable GEDI data, several filters were applied to the waveforms to remove waveforms contaminated by clouds and other atmospheric artefacts (e.g. [3]). The application of different filters on the GEDI dataset showed that among the 6166 acquired waveforms over ourstudy site between April 2019 and September 2019, the majority, or 5682 (92.2%) provided exploitable waveforms.

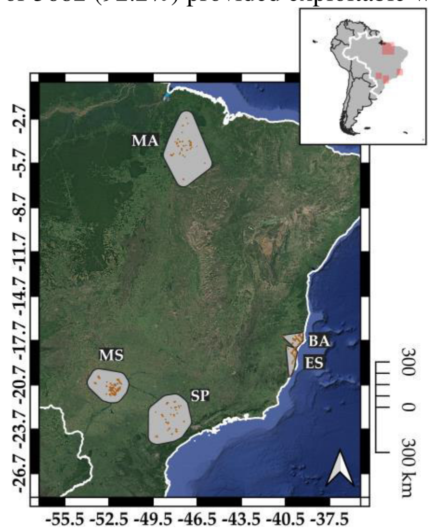


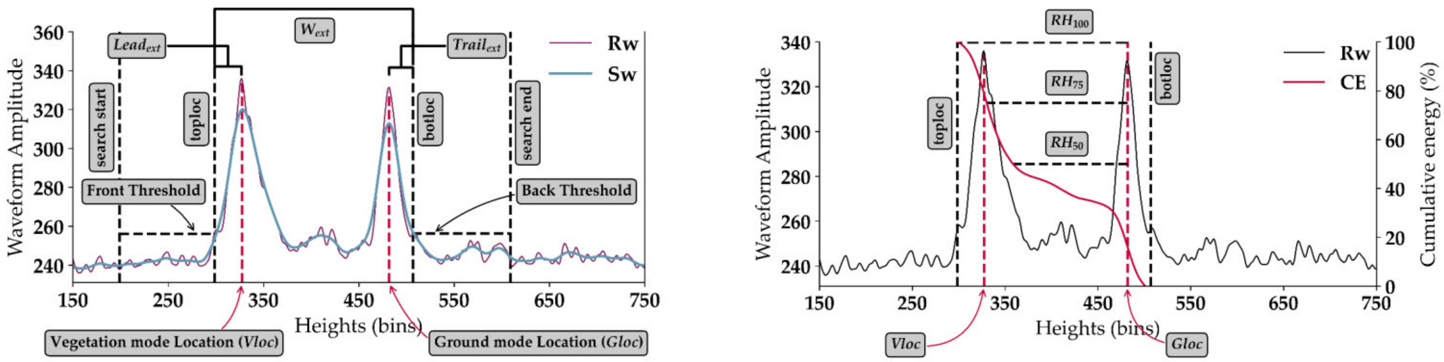
Figure 1. Location of the five study sites

3. FOREST HEIGHT ESTIMATION

Acquired waveforms by GEDI, like all FW systems, consist of a series of multiple connected temporal modes, or peaks, representing the different reflections from an object (e.g. top of canopy cover) or different objects close together (e.g. understory and ground) (Fig. 2). Therefore, the height of the tallest canopy within a waveform is simply the difference between the top location of the signal (toploc, Fig.2) and the ground peak location (Gloc, Fig.2). However, due to noise, the indistinguishable echoes inside the waveform, and terrain and vegetation variabilities, accurate estimation of canopy heights requires the identification of the relevant peaks in the waveform and the subsequent estimation of the metrics that characterizes the vegetation and terrain.

Nonetheless, the success of the different metric-based methodologies to derive forest characteristics lies on the accurate estimation of the waveform metrics. For GEDI, six different algorithms (denoted a1 to a6) are used to preprocess the waveforms, thus, six different values of each metric can be obtained, which can lead to different estimation accuracies on the canopy heights, and the wood volume.

In this study, several linear and linear regression models relying on only waveform metrics or on both waveform metrics and terrain information derived from DEMs were tested. The main waveform metrics used in these models are the waveform extent defined as the height difference between the signal begin and the signal end of a waveform (Wext, in m), the leading edge extent (Leadext, in m) and the trailing edge extent (Trailext, in m). According to Hilbert and Schmullius [6], the leading edge is defined as the elevation difference between toploc and the Vloc, and the trailing edge as the difference between Glov and botloc (Fig. 2). The terrain information used in the regression models are the



(a) (b)

Fig. 2. (a) Example of an acquired GEDI waveform (Rw) over a Eucalyptus stand, its smoothing (Sw) and corresponding waveform metrics. (b) The 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. 1 ns corresponds to 15 cm sampling distance in the waveform.

terrain index (TI, in m) derived from a DEM (we used the SRTM DEM in this study), Terrain roughness (ROUG, in m), and terrain slope (S, in %). TI is defined as the difference between maximum and minimum terrain elevations in a given window centered on each GEDI footprint, while ROUG is defined as standard deviation of the elevation in a 3 × 3 pixel-moving window. Finally, we also used the relative height metric (RH_n) which represents the height between botloc and the location at n% of cumulative energy ($RH_n\,, 10\% \leq n \leq 100\%\,, step\ 10\%$) (Fig. 2b). Several regression models were evaluated based on the previously mentioned GEDI metrics and terrain indices [7]. For our study areas defined by mostly flat terrains, the root mean square error (RMSE) on the estimation of canopy heights ranged between 1.44 and 2.31 m (R^2 between 0.80 and 0.92) with GEDI metrics extracted using algorithm a1. The best fitting results were obtained through a stepwise linear regression model (Fig.3, RMSE = 1.44m, R^2 = 0.92) that relies solely on various relative height metrics, namely RH_{90}, RH_{10}, RH_{80}, and RH_{100}.

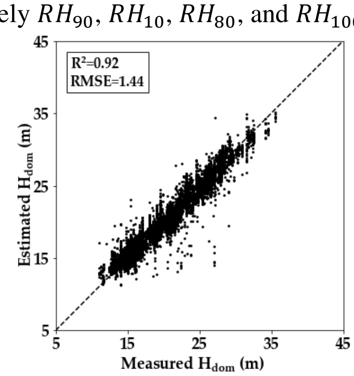


Figure 3. Canopy height estimates (with the best statistical model) in comparison to measured canopy height. RMSE expressed in m.

The estimation of H_{dom} using the models described previously with GEDI metrics extracted from the five remaining algorithms (a2 to a6) has also been tested. The results show that height estimation was worse with the metrics from algorithms a2 through a6 in comparison to the metrics from algorithm a1. Using the best statistical model, the RMSE on the canopy height estimates ranged from 1.46 m (R^2 of 0.92, a2) to 1.67 m (R^2 of 0.88, a5).

4. WOOD VOLUME ESTIMATION

In contrast to canopy heights that are directly measured by FW LiDAR, the estimation of wood volume relies on the allometric relationship to the canopy height, and are generally only accurate to the areas they were developed for [8]. Wood volume estimation using FW LiDAR data can be categorized into two approaches. The first approach first estimates canopy heights from FW data, and then wood volume is estimated by means of an allometric relationship [4]. The second approach relies on regression models, or machine learning algorithms to infer the allometric relationship to canopy heights, and thus allows the estimation of wood volume from FW LiDAR metrics directly [9]. Here, both approaches were used. For the methods relying on direct GEDI metrics, we used the metrics presented in section 3, and as the wood volume (V) increases with canopy height in a non-linear shape, we also calculated RH_n for several power values ($RH_n^p\,, 1 < p \leq 3\,, step\ 0.2$). Wood volume estimation results showed an RMSE between 24.39 (R^2 of 0.90) and 28.19 (R^2 of 0.87), with the methods estimating directly the wood volume from GEDI metrics were being slightly more accurate. Similarly to the estimation of canopy heights, the most accurate wood volume estimates were obtained with the stepwise regressing using solely the relative height metrics (Fig. 4).

The estimation of V was also tested using GEDI metric values extracted using the remaining five algorithms (a2 through a6). The results show that the estimates of V were less accurate with the best statistical model using GEDI metrics from algorithms a2 through a6 compared to a1.

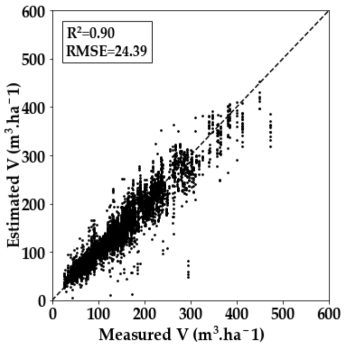


Figure 4. Wood volume estimates (with the best statistical model) in comparison to measured wood volume. RMSE expressed in $m^3.ha^{-1}$.

5. CONCLUSIONS

In this study, 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. Overall, 5517 GEDI shots over 566 Eucalyptus stands were analyzed over our study area.

Model results showed that GEDI acquired waveforms have high accuracy on the estimation of canopy and an RMSE of 1.44 m (R^2 of 0.92) were obtained using a combination of relative height values extracted from each waveform. Moreover, the reliance on only the relative height metrics, alleviates the need to generate previously defined metrics such as lead and trail.

Wood volume estimation accuracy was similarly high, with an RMSE of 24.39 m^3.ha-1 (R^2 of 0.90). However, wood volume is estimated based on allometry rather than directly. Therefore, wood volume model transferability between Eucalyptus plantations exhibiting dissimilar allometric relationship between canopy heights and wood volume, would most likely produce less than ideal wood volume accuracies.

Finally, the choice of the algorithm used to extract the waveform metrics affected sometimes the accuracy, as metrics extracted using algorithm a5 showed ~16% higher RMSE on the estimation of both H_{dom} and V.

6. ACKNOWLEDGEMENTS

The authors would like to thank the GEDI team and the 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 x Suzano project.

7. REFERENCES

[1] P. Ploton, R. Pélissier, N. Barbier, C. Proisy, B. R. Ramesh, and P. Couteron, “Canopy Texture Analysis for Large-Scale Assessments of Tropical Forest Stand Structure and Biomass,” in *Treetops at Risk*, M. Lowman, S. Devy, and T. Ganesh, Eds. New York, NY: Springer New York, 2013, pp. 237–245.

[2] T. Le Toan *et al.*, “The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle,” *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2850–2860, Nov. 2011, doi: 10.1016/j.rse.2011.03.020.

[3] 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 Observations Remote Sensing*, vol. 7, no. 1, pp. 290–299, Jan. 2014, doi: 10.1109/JSTARS.2013.2261978.

[4] 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, p. n/a-n/a, Nov. 2005, doi: 10.1029/2005GL023971.

[5] R. Dubayah *et al.*, “The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography,” *Science of Remote Sensing*, vol. 1, p. 100002, Jun. 2020, doi: 10.1016/j.srs.2020.100002.

[6] C. Hilbert and C. Schmullius, “Influence of Surface Topography on ICESat/GLAS Forest Height Estimation and Waveform Shape,” *Remote Sensing*, vol. 4, no. 8, pp. 2210–2235, Jul. 2012, doi: 10.3390/rs4082210.

[7] I. Fayad *et al.*, “Assessment of GEDI’s Lidar Data for the Estimation of Canopy Heights and Wood Volume,” *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, 2020.

[8] M. Hayashi, N. Saigusa, Y. Yamagata, and T. Hirano, “Regional forest biomass estimation using ICESat/GLAS spaceborne LiDAR over Borneo,” *Carbon Management*, vol. 6, no. 1–2, pp. 19–33, Mar. 2015, doi: 10.1080/17583004.2015.1066638.

[9] 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 Sensing of Environment*, vol. 112, no. 10, pp. 3876–3890, Oct. 2008, doi: 10.1016/j.rse.2008.06.003.