ICESAT-2 COMPARISON WITH FIELD HEIGHT IN BRAZILIAN PASTURES

Maria O'Healy Hunter[1], *Jairo Matos da Rocha*[2], *Nathália Monteiro Teles*[3], *Wilton Ladeira da Silva*[4], *João Vitor Silva Costa* [5] *and Laerte Guimarães Ferreira*[6]

[1]LAPIG, UFG - Universidade Federal de Goiás - Campus Samambaia Almeida Palmeiras, s/n - Chácaras Califórnia, GO, 74001-970 and maria.hunter@ufg.br, [2]LAPIG, UFG and jairo@jairomr.com.br; [3]LAPIG, UFG and nathalia.teles@ufg.br, [4]EVZ, UFG and wiltonladeira@ufg.br, [5]LAPIG, UFG and joaovsc17@gmail.com and [6]LAPIG, UFG and laerte@ufg.br

ABSTRACT

ICESat-2 is a satellite lidar system launched in 2018 that has global coverage of terrestrial and ice-covered regions. This satellite provides a potential source of information regarding vegetation height and structure, but has not previously been applied specifically to pasture areas. Here we present results from fieldwork in the Rio Vermelho watershed in Goiás, showing the applicability of ICESat-2 vegetation height and structure metrics to pastures. We analyze metrics calculated from the ATL08 product, which is more readily applied at the large scale, and the ATL03 product, which allows for improved filtering of ICESat-2 measurements. We also test filters based on season, time of day, and laser strength to determine best practices when applying ATL08. We find that median height is a metric appropriate for Brazilian pastures, with an RMSE of 3.3 m when considering 20 m segments and 2.4 m for 100 m segments. Using the filters available in ATL03, the RMSE for 20 m segments decreases to 2.5 m.

Key words – *Cerrado, pasture, height, vegetation structure, Lidar.*

1. INTRODUCTION

Pastures cover 179 million hectares within Brazil, with the majority being classified as natural or semi-natural [1]. Natural and semi-natural grazing areas often include a mixture of herbaceous biomass, together with shrubs and sparse trees. The amount of each vegetation type is important for the productivity of the grazing area, for the above- and below-ground carbon present, and for habitat available for non-grazing species.

Sampling vegetation structure and type across large areas is time consuming and logistically difficult, as well as expensive. Previous work has shown the capacity of satellite lidar to estimate total canopy height and biomass [2–4]. However, non-forest areas are poorly sampled, and metrics of vegetation structure beyond top of canopy height are rarely presented. This is problematic as many regions classify vegetation type and cover through percentage of tree canopy cover and presence of herbacious and shrub layers [5].

The ICESat-2 satelite lidar mission was launched in September 2018 into a near polar orbit with a dual mission objective. The first objective is to measure and monitor ice sheets and glaciers, and the second is to measure the height of forests and other ecosystems worldwide. The ATLAS instrument aboard ICESat-2 has three photon-counting green lasers that are split into 2 beams each with footprints of approximately 11 m diameter [6] spaced 0.7 m along-track. For each laser, there is a power differential between the two beams, with the 'strong' beam including four times more energy than the 'weak' beam.

ICESat-2 has been shown to accurately estimate top of canopy forest height and biomass in forest ecosystems [7, 8], but little research is available for regions with low canopy height or more open ecosystems. However, Purslow and colleagues [9] suggest that stand level structural traits can be derived from ICESat-2 data.

Given the lack of continuous coverage with satellite lidar, efforts to map height or structure over regions must rely on satellite lidar as training data [10] that is incorporated into a modeling framework. ICESat-2 data has broad coverage (24.8 billion 20 m segments globally), but also has many kilometers between sample tracks.

Therefore, the objectives of this study are to (1) test the applicability of ICESat-2 satellite lidar to differentiate structural characteristics in pastures and (2) determine best practices for filtering ICESat-2 based training data for models of short vegetation height.

2. MATERIAL AND METHODS

2.1. Region of Interest

The Rio Vermelho watershed, within the western portion of the state of Goiás is an approximately 11.000 km^2 area typical of the central region of the Brazilian Cerrado, including open woodlands and pastures. In 2023, the land cover for the watershed was dominated by pasture (64%), Savanna (14 %) and Forest (14%) (MapBiomas Collection 9, 2024). It is characterized by high temperatures, with monthly averages ranging between 23°C and 28°C and overall variation between 16°C and 34°C. The total annual rainfall is approximately 1410 mm per year. Monthly rainfall varies greatly, with a wet season between November and February with an average of over 200 mm of rain per month. The months with less than 100 mm of precipitation are from April through September, with less than 10 mm on average in the months of June through August.

2.2. Sampling Design

ICESat-2 data was downloaded from NSIDC coincident with the Rio Vermelho watershed for a one year period spanning September 2022 to September 2023. Eight orbital passes of ICESat-2 occurred during this period. ATL08 segments were filtered based on a minimum number of vegetation returns, occurrence within pastures, and accessibility (within 500 m of any road mapped via OpenStreetMap). The occurrence within pastures was determined based on areas consistently

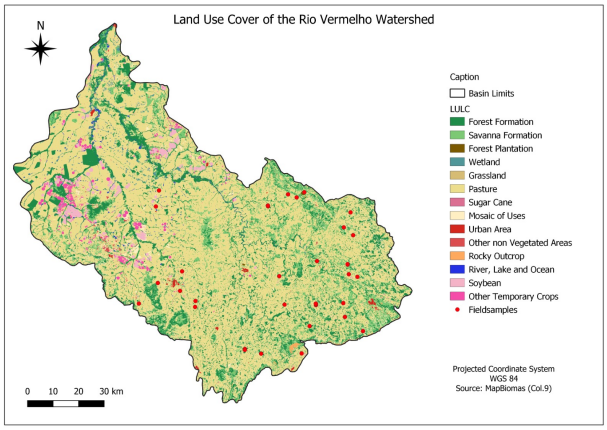

Figure 1: Map of the Rio Vermelho watershed and land use overlaid with field sampling locations.

defined as pastures between 2017 and 2022 by LAPIG (https://atlasdaspastagens.ufg.br/map). Segments within 50 m of roads were excluded to avoid sampling over roadways.

The selection of sample points was based on equal divisions by season of ICESat-2 overpass and vegetation structure. Maximum vegetation height and the presence of canopy layers was taken into account to separate four structural classes. The ATL08 v006 product includes height metrics at 5-percentile intervals based on photons identified as reflected from vegetation measured within 100 m along-track segments. [8, 11] The vegetation heights considered were 1, 5, and 15 m interpreted as grasses (<1 m), shrubs and low trees(<5 m), and two tree classes (5-15 m and >15 m). The four structural classes used in the sample design were: low vegetation (no returns > 5 m), medium vegetation with openings (no returns greater than 15 m and at least one missing layer), dense vegetation (no returns greater than 15 m and all vegetation layers present), and tall vegetation (greater than 15 m in height). Of the 40 locations selected for field sampling, 35 were accessible in the field (Figure 1).

2.3. Field Sampling Procedure

The central point of each sampled ICESat-2 ATL08 segment was located in the field and a 90 m x 90 m area was delimited. Within this area, four N-S transects were sampled (10 and 30 meters to the east and west from the central line). Measurements were made every 5 m along each transect for a total of 18 x 4 measurements of vegetation height. Pasture height was measured at the exact point, whereas shrubs and trees were included within 2 and 5 m of the transect, respectively. In the case of multiple shrubs or trees, multiple heights were recorded. Along each transect, five geolocated photographs were taken to produce a simultaneous record of field conditions and to assist in the geolocation of individual transects.

Coincident with field measurements, a multi-rotor drone, Mavic 3 multispectral, manufactured by DJI, was used to collect aerial images of the drawn plots. The flights were planned to cover areas measuring 90 by 90 meters, with 70% lateral overlap and 80% frontal overlap. The height of the flights was 50 meters, generating images with a 1.44 cm resolution. The photographs obtained were processed in the Pix4d Mapper software version 4.8.4, resulting in the following products: orthorectified mosaic, digital surface model (DSM), and digital terrain model (DTM). Subsequently, the DSM and DTM products were subtracted (DSM - DTM), in QGIS, to obtain the Canopy Height Model (CHM).

2.4. ICESat-2 Filtering and Analysis

The ATL03 product is intended to be a single source product for photon-level data and ancillary information necessary for processing [12]. However, using this data source at larger scales is prohibitive due to file sizes and processing scales that are not commonly available. In contrast, the ATL08 product produces summary statistics for 100 m segments, limited statistics for 20 m segments, and heights of photons together with classification into noise, terrain, canopy and top of canopy [11]. To understand the differences between this product and the summarized information made available within ATL08, we extracted the photon level information from ATL03 files coincident with the ATL08 regions sampled in the field (Figure 2). These were then processed into 20 m segments filtering photons with low reported signal confidence (signal_conf_ph > 1) and high quality (quality_ph = 0). The signal confidence flag indicates whether individual photons are expected to be noise or signal. The specifications used are: unidentified (-1), noise (0), buffer (1), low confidence (2), medium confidence (3), high confidence (4). The quality_ph flag varies from 0 to 3 with zero indicating no signal saturation or internal reflection.

For the secondary objective of identifying high-quality training data with information available for both ATL08 and ATL03, we identified each segment in terms of three characteristics: time of ICESat-2 data acquisition, season of acquisition, and whether the selected segment was scanned using the strong or weak beam of the laser. Time of day was divided into night/day depending on the solar altitude angle and the period from October through April was considered to be wet, and from May through September dry.

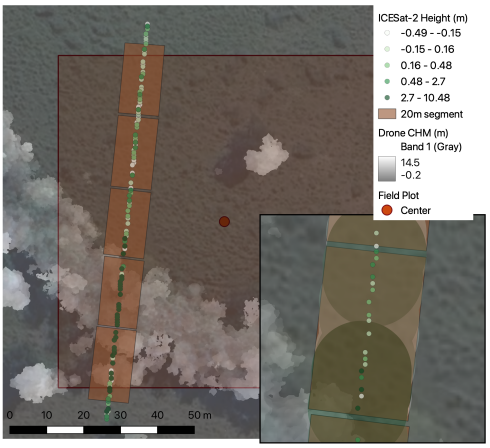

Figure 2: Example plot with field and ICESat-2 data. Inset of center 20m segment and 11m illumination diameter per ICESat-2 height measurement.

2.5. Field Comparison with ICESat-2

Field measurements were aligned with the geolocation of photons from ATL03 to identify which field measurements were closest to the ICESat-2 field of view. Polygons were drawn for both 20 m and 100 m segments and field vegetation measurements greater than 10 cm in height were included to calculate the mean, median, 95th percentile and maximum height of vegetation. Estimates of percent tree cover were calculated using the Canopy Height Model considering the percentage of photons over 5 m versus the total number of non-noise photons returned.

All metrics were compared using the RMSE and the concordance correlation coefficient [13].

3. RESULTS

The additional filtering available within the ATL03 product decreases the variability of measurement and increases the extent to which they are representative of field measurements. For median heights, the ATL03 product decreased the RMSE by an average of 0.7 m with a small difference in improvement seen between the 20 m and 100 m segments. For mean height of 100 m segments, a 0.65 m decrease was observed, similar to median height. The difference grows to 3.8 m for mean height of 20 m segments (Table 1).

	Median	Mean	95th Pct	% Tree
ATL03 20m	2.49	2.11	3.69	11.92
ATL08 20m	3.32	5.88	3.84	18.42
ATL03 100m	1.74	1.49	3.56	5.08
ATL08 100m	2.39	2.14	4.93	10.88

Table 1: RMSE of ATL03 and ATL08 at 20m and 100m resolution with no additional filtering based on laser strength, season or time of day.

The 95th percentile height of field measurements and ICESat-2 photon records were also compared as this is used in forest ecosystems. For 100 m segments, the RMSE is approximately doubled when compared to median height, and more than doubled when compared to mean height. Contrary to other trends, the RMSE decreases when observing a smaller segments for ATL08, and remains stable for ATL03. The percent tree cover shows agreement within 5% only at the 100 m segment and with ATL03 data. All other scales and datasets show RMSE above 10%.

A secondary objective was to determine best practices for selecting data for training models of height over larger areas. Considering the ATL03 dataset, it is important to consider only night data as this reduces the RMSE by almost 0.7 m. This is followed in importance by strong beam data, which also reduced RMSE by more the 60 cm. Taking the season of collection into account causes the least effect on RMSE. For ATL08 data, the strength of the beam is most important, followed by season, and then time of day (Table 2).

Considering the concordance correlation coefficient (ccc), patterns shift. For ATL03 the strongest correlation is found for daytime segments (ccc = 0.65). Segments measured by the strong beam continue to show better correlation between field and estimated values of median height (ccc = 0.61 versus 0.30) and segments recorded by ATLAS during the wet season are better correlated than during the dry (ccc = 0.59 and ccc = 0.27, respectively). For ATL08, correlations are weaker and night shows higher correlation than day, while strong beam and wet season maintain higher correlation coefficients.

Filter	ATL03 Median	ATL08 Median	n
night	1.98	3.19	45
day	2.66	3.39	120
strong	2.29	3.17	115
weak	2.93	3.97	50
dry	2.36	3.07	65
wet	2.57	3.47	100

Table 2: RMSE of ATL03 and ATL08 Median Height estimates for 20 m segments with additional filtering based on time of day (night / day), laser strength (strong / weak), and season (dry / wet).

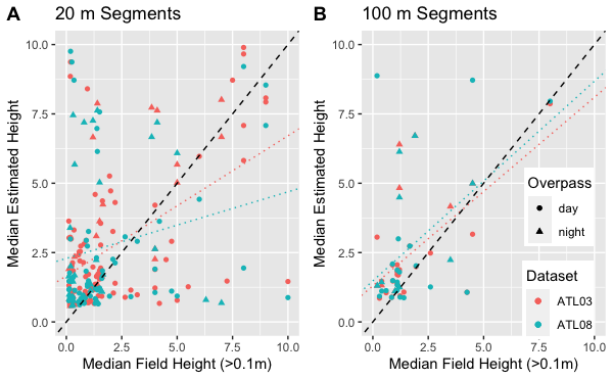

Figure 3: Median Heights estimated within (A) 20 m segments and (B) 100 m segments using ATL08 and ATL03 processing procedures.

4. DISCUSSION

The classification of pastures within the Cerrado can be done using optical imagery such as Landsat or Sentinel in combination with high resolution imagery that allows for visualization of trails, fences and other structures indicative of grazing and human management [1]. However, the extent of tree cover and shrub presence within these areas is difficult to quantify, though recent studies using object oriented modeling are tackling this issue at small scales [14, 15].

Satellite lidar is a source of globally distributed samples of heights. These can be aggregated at either the 20 m segment or 100 m segment from either ATL08 or ATL03 files. Individual ATL03 files are often 5 Gb, whereas an ATL08 file covering the same area is less than 200 Mb. Of the metrics and scales tested, we found that mean height of 100 m segments resulted in the smallest RMSE and the largest correlation coefficient of the height metrics. Estimates of tree canopy cover was the most highly correlated with field data overall.

However, given the polar orbit of the satellite, individual files include relatively small segments within an area of interest, requiring the download and filtering of many individual files. Given the size difference and the number of files involved in even a regional analysis, ATL08 is more

widely used than ATL03. If constrained to data available through ATL08, percent tree canopy cover and median height showed higher correlation than mean height and 95th percentile height for both segment sizes. However, for comparison with Landsat or Sentinel scale satellite imagery, the 20 m segment is preferable as it fits within 1-2 pixels. For this use-case, the median height is the suggested metric.

There is no filter available within the ATL08 files that reduces RMSE to levels seen for ATL03. Our results suggest that all of the tested filters for segments improve the quality of ATL08 measurements. The largest effect is seen between the strong and weak beam with ICESat-2 overflights in the dry season, during night, or with the strong beam. This is consistent with other studies [16, 17]. Milenkovic and colleagues found that both ICESat-2 and GEDI satellite lidar can be used to assess forest regrowth in the Amazon. In this case the ability to aggregate over larger areas and stand-ages allowed for the use of the 100 m segments.

However, our results show the importance of quality flags that are only available through the ATL03 product. Specifically, flags that indicate the confidence that a given photon is signal versus noise (signal_conf_ph) and the likelihood of reflection within the instrument (quality_ph). Overall, the inclusion of these filters improved RMSE by 80 cm for median height, and 3.7 m for mean height. A smaller improvement was present for 95th percentile height, and Tree Canopy Cover was also strongly influenced.

5. CONCLUSIONS

Considering that top of canopy height is often a secondary consideration in pasture and open ecosystems, we conclude that both Median Height and Mean Height are appropriate metrics. Working at large scales, and for training models that use Landsat or Sentinel, we suggest Median Height. This represents the dominant height of plants in the landscape and is less sensitive to scale and available filtering. If larger areas can be aggregated, mean height may be a more suitable metric. In either case, filtering by season, time of day and beam strength improved comparison with field measurements.

The filters available in ATL03 increase the correlation between field data and ICESat-2 metrics at both the 100 m and 20 m scale more than seasonal or temporal filters. The inclusion of a small amount of additional information in subsequent versions could improve the applicability of ICESat-2 data outside forest ecosystems.

6. REFERENCES

[1] L. Parente, L. Sloat, V. Mesquita, and et al. Annual 30-m maps of global grassland class and extent (2000–2022) based on spatiotemporal machine learning. *PREPRINT (Version 3) available at Research Square*, pages https://doi.org/10.21203/rs.3.rs–4514820/v3, 2024.

[2] M.A. Lefsky, W.B. Cohen, and D.J. Parker, G.G.and Harding. Lidar remote sensing for ecosystem studies. *Bioscience*, 52:19–30, 2002.

[3] M.A. Lefsky, D.J. Harding, M. Keller, W.B. Cohen, C.C. Carabajal, F. Del Bom Espirito-Santo, M.O. Hunter, and R. de Oliveira Jr. Estimates of forest canopy height and aboveground biomass using icesat. *Geophysical research letters*, 32(22), 2005.

[4] D.R. Steutker and N.F. Glenn. Lidar measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment*, 102:135–145, 2006.

[5] P. Scarth, J. Armston, R. Lucas, and P. Bunting. A structural classification of australian vegetation using icesat/glas, alos palsar, and landsat sensor data. *Remote Sensing*, 147:11, 2019.

[6] A. Liu, X. Cheng, and Z. Chen. Performance evaluation of gedi and icesat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, 264:112571, 2021.

[7] S. Nandy, R. Srinet, and H. Padalia. Mapping forest height and aboveground biomass by integrating icesat-2, sentinel-1 and sentinel-2 data using random forest algorithm in northwest himalayan foothills of india. *Geophysical Research Letters*, 48:e2021GL093799, 2021.

[8] A. Neuenschwander, E. Guenther, Joanne C. White, L. Duncanson, and P. Montesano. Validation of icesat-2 terrain and canopy heights in boreal forests. *Remote Sensing of Environment*, 251:112110, 2020.

[9] M. Purslow, S. Hancock, A. Neuenschwander, J. Armston, and L. Duncanson. Can icesat-2 estimate stand-level plant structural traits? validation of an icesat-2 simulator. *Science of Remote Sensing*, 7:1000086, 2023.

[10] W. Li, Z. Niu, R. Shang, Y. Qin, and H. Chen. High-resolution mapping of forest canopy height using machine learning by coupling icesat-2 lidar with sentinel-1, sentinel-2 and landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, 92:102163, 2020.

[11] A. Neuenschwander, K. Pitts, B. Jelley, J. Robbins, J. Markel, S. Popescu, R. Nelson, D. Harding, D. Pederson, B. Klotz, and R. Sheridan. Ice, cloud, and land elevation satellite (icesat-2) project algorithm theoretical basis document (atbd) for land - vegetation along-track products (atl08), version 6. icesat-2 project. *DOI: 10.5067/8ANPSL1NN7YS*, 2022.

[12] T.A. Neumann, A. Brenner, D. Hancock, J. Robbins, A. Gibbons, J. Lee, K. Harbeck, and J. Saba. Ice, cloud, and land elevation satellite (icesat-2) project algorithm theoretical basis document (atbd) for global geolocated photons atl03, version 6. icesat-2 project. *DOI: 10.5067/GA5KCLJT7LOT*, 2022.

[13] S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116 (536):2009–2022, 2021.

[14] A. Chen, C. Xu, M. Zhang, J. Guo, X. Xing, D. Yang, B. Xu, and X. Yang. Cross-scale mapping of above ground biomass and shrub dominance by integrating uav and satellite data in temperate grassland. *Remote Sensing of Environment*, 304:114024, 2024.

[15] S.J. Graves, T.T. Caughlin, G.P. Asner, and S.A. Bohlman. A tree-based approach to biomass estimation from remote sensing data in a tropical agricultural landscape. *Remote Sensing of Environment*, 218:32–43, 2018.

[16] M. Milenkovic, J. Reiche, J. Armston, A. Newenschwander, W. De Keersmaecker, M. Herold, and J. Verbesselt. Assessing amazon rainforest regrowth with gedi and icesat-2 data. *Science of Remote Sensing*, 5:1000051, 2022.

[17] V.S. Konduri, D.C. Morton, and N. Andela. Tracking changes in vegetation structure following fire in the cerrado biome using icesat-2. *Journal of Geophysical Research: Biogeosciences*, 128 (5):e2022JG007046, 2023.