

Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest

Aline Bernarda Debastiani^{1§}, Carlos Roberto Sanquetta¹, Ana Paula Dalla Corte¹, Franciel Eduardo Rex¹, Naiara Sardinha Pinto¹

Debastiani A.B., Sanquetta C.R., Corte A.P.D., Rex F.E., Pinto A.N., 2019. Evaluating SAR-optical sensor fusion for aboveground biomass estimation in a Brazilian tropical forest. Ann. For. Res. 62(1): _-_.

Abstract. The aim of the present study is to evaluate the potential of C-band SAR data from the Sentinel-1/2 instruments and machine learning algorithms for the estimation of forest above ground forest biomass (AGB) in a high-biomass tropical ecosystem. This study was carried out in Jamari National Forest, located in the Brazilian Amazon. The response variable was AGB (Mg/ha) estimated from airborne laser surveys. The following treatments were considered as model predictors: 1) Sentinel-1 Sigma 0 at VV and VH polarizations; 2) (1) plus Sentinel-1 textural metrics; 3) (2) plus Sentinel-2 bands and derived vegetation indices (LAI, RVI, SAVI, NDVI). Our modeling design estimated the relative importance of SAR vs. optical variables in explaining AGB. The modeling was performed with twelve machine-learning algorithms including, neural network and regression tree. The addition of texture and optical data provided a noticeable improvement (3%) over models with SAR backscatter only. The best model performance was achieved with the Random Tree algorithm. Our results demonstrate the potential of freely-available SAR data and machine learning for mapping AGB in tropical ecosystems.

Keywords: Amazon Forest, artificial intelligence, Sentinel 1, Sentinel 2, AGB, carbon

Authors. ¹University of State of Parana, Avenida Rio Grande do Norte, 1525 87701-020 - Centro - Paranavaí - PR, Brazil

[§] Corresponding author: Aline Bernarda Debastiani (aline.debastiani@gmail.com)

Manuscript received May 4, 2019; revised July 24, 2019; accepted July 25, 2019; online first July 30, 2019.

Introduction

Forests are widely distributed vegetation ecosystems on the planet, covering approximately 4000 million hectares (FAO 2015). Information on the status and monitoring of forests is

highly relevant to map ecosystem environmental services, and to quantify their role as terrestrial carbon sinks (Waring & Running 2007). However, forests' role in the global carbon budget remains poorly characterized owing to practical difficulties of measuring the stocks

of forest biomass on a large scale (Minh et al. 2016), especially when dealing with tropical forests located in complex environments.

Commonly, forest above-ground biomass (AGB) can be estimated through three methods: (i) inventory measurements: involve destructive measurements of sample plots which produce reliable information about the AGB at local or regional scales (Chave et al. 2004), (ii) model-based simulations: they usually provide estimates on a local to global scale, based on pre-adjusted allometric models with inventory data (Lu 2006), (iii) estimation from data from satellites: these are usually combinations of remote measurements and based on the measurements obtained by the field inventory (Stickler et al. 2009). Compared to the forest inventory model and approaches, remote sensing techniques significantly improve the efficiency of forest AGB mapping in areas of difficult access (Lu 2006), with better spatial and temporal coverage.

Field measurements are essential for the verification of the indicators obtained by remote sensing; however, there are several potential sources of inventory error, such as variation in plot size and the allometric equations used (Chave et al. 2004). In addition, according to Houghton et al. (2009), current estimates of AGB and carbon stock in tropical forests come from extrapolations from a limited number of field sites, leaving many regions and forest types underrepresented. As with field measurements, some technologies, such as LiDAR, are promising and capable of estimating forest variables for a large scale and with high accuracy (Andersen et al. 2013, Ruiz et al. 2014, Deo et al. 2016, Silva et al. 2017). However, LiDAR technology has the limitation of having a high cost, making it impossible to acquire systematic data for monitoring, thus creating the need for another technology that can produce wall-to-wall maps for large-scale monitoring efforts. Thus, the alternative of Synthetic Aperture Radar (SAR) data emerges for forest investigations and is more

evident after the launch of new satellites that have SAR sensors on board (for example TerraSAR-X, ALOS-2 and Sentinel-1), offering cloud-free images with fine resolution (< 20m) and frequent revisits (Santini et al. 2017).

SARs are active microwave sensors that can operate at any time of day and penetrate clouds, offering significant advantages for monitoring sites with dense cloud cover such as the Amazon. Furthermore, dual - or quad-pol SAR products allow researchers to map vegetation structure by estimating the relative contribution of ground, canopy, and trunks to the received signal (Ningthoujam et al. 2018). The backscatter approach based on SAR data is widely known for the mapping of AGB (Berninger et al. 2018). It consists of the measurement of the energy received by the sensor after the transmission, known as backscatter, in which this measurement is subsequently related to AGB measurements in the field. The backscatter values are sensitive to the amount of biomass at a given location, it means that at a certain point the saturation will occur, in which the sensitivity of the backscatter to AGB is stagnated. This point is related to the wavelength of the sensor (Joshi et al. 2007), the C-band has a wavelength capable of penetrating through the leaves, but scattered with the existence of small branches, the L-band, has a longer wavelength and is spread by trunks and branches thick. The P-band is the most suitable for AGB quantification, because at this wavelength the sensor energy is able to penetrate the forest canopy, but P-band data is not yet available for free (Ghasemi et al. 2011). L-band sensors have been widely employed to estimate forest AGB (Carreiras et al. 2013, Cartus et al. 2012, Saatchi et al. 2011a, Santos et al. 2002). Originating mainly from the satellite ALOS PALSAR satellite that was in operation from 2006 to 2011, leaving a free monitoring and availability gap ever since. Considering the reduction of the series of free data for this range and the indicated limitation for the high AGB forest, characteristic of the

Amazon Forest, there is a need to explore the combination of other freely available data to ensure the continuity of forest cover estimates, such as the case of the C-band SAR data obtained by the Sentinel 1 satellite, texture measurements and optical data, together with texture information as well as optical images. The texture of SAR images can provide information related to the structural and geometric properties of forest canopies (DeGrandi et al. 2009), which may correlate with AGB and be especially useful in discriminating between forest types (Podest & Saatchi 2002).

In this context, the objective of the present study is to evaluate the potential for estimating forest AGB with the combination of SAR, information of texture and optical data and employing machine learning algorithms. A secondary objective is to compare model accuracy across different machine learning approaches and quantifying the relative importance of radar and optical observations as AGB predictors.

Materials and methods

Study area

The study area is located in the Jamari National Forest (Flona) (Figure 1), State of Rondônia, in the north region of Brazil (9° 9' S and 62° 58' W). Jamari Flona has an area of about 220.000 hectares of Dense Rain Forest, designated as a Sustainable Use Conservation Unit of the Amazon Forest.

The flora at our study site is representative of the southwestern Amazon, a region subject to high rates of deforestation. According to the Technical Manual of Brazilian Vegetation (IBGE 2012), the physiognomic-ecological classification is predominantly of Dense Rain Forest, with some portions of Open Ombrophylous Forest, predominantly of palm trees or liana.

Sentinel data acquisition and processing

Synthetic Aperture Radar images from the Sentinel 1 satellite and multispectral Sentinel 2 images of the European Space Agency were downloaded from the Copernicus Sentinels Agency's Scientific Data Center (ESA 2016) and used for this study. A summary of the specifications can be seen in Table 1.

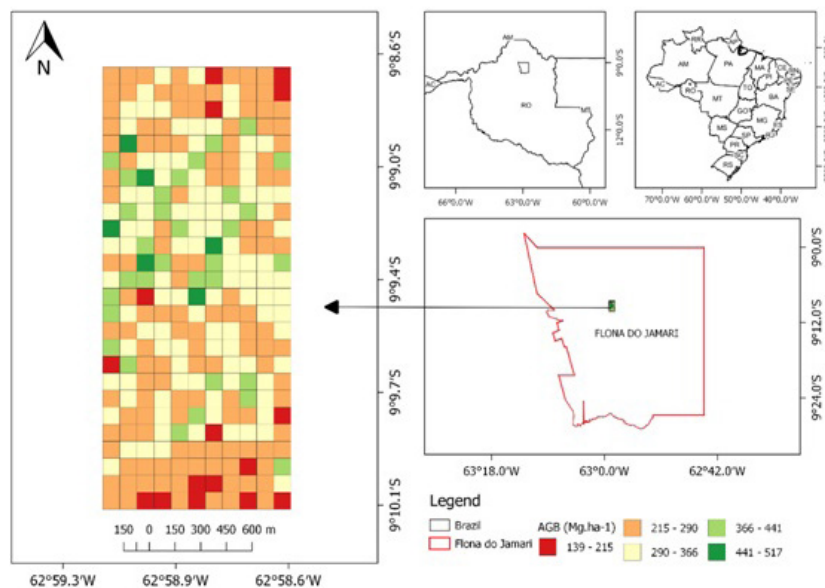


Figure 1 AGB (Mg.ha⁻¹) obtained by means of LiDAR airborne data and location of the study area on Brazil

For the images of the Sentinel 1A satellite, we employed SNAP 5.0 software (Sentinel's Application Platform) (SNAP 2018) to generate Analyses Ready Datasets. The steps include (1) the radiometric calibration for the conversion of the pixel value in backscatter (Sigma 0), (2) apply a Redefined Lee filter for Speckle reduction, and (3) apply the Range-Doppler Terrain correction to geocode the SAR image to UTM WGS84 coordinate system.

SNAP software was also used to derive several texture metrics from the geocoded backscatter. Specifically, we employed a 5 x 5 window (Mahdianpari & Motagh 2017, Oon & Azhar 2019), to generate a Gray Level Co-Occurrence Matrix (GLCM) for the VH and VV bands with the following calculations: contrast, dissimilarity, homogeneity, second angular momentum (ASM), energy, maximum likelihood (MAX), entropy, mean, variance and correlation. The Principal Components (PCA) images were also generated. Detailed descriptions of each texture measure can be obtained in Haralick (1979).

We searched for local precipitation data near our study site, given that humidity has an influence on backscatter values (Kirimi et al. 2016). The meteorological stations closest to our study site were Porto Velho (code A925 113 Kilometers away) and Ariquemes (code A940, 80 km away). Porto Velho station recorded 4 mm precipitation on the day before the Sentinel 1 acquisition and no accumulated

volume in the last 5 days, whereas Ariquemes station (code A940) did not register precipitation in the 10 days prior to imaging. The Sentinel 2A satellite images used here had Level 1C processing applied to them, which includes they have been processed for radiometric and geometric corrections (Sentinel-2_Team 2015). Each granule contained 13 bands with resolution varying between 10-60 m. Atmospheric correction was performed in semi-automatic classification plugin in Qgis software (version 2.18.16).

Vegetation indices were generated from the multispectral image to serve as complementary information for the modeling. They included Soil Adjusted Vegetation Index (SAVI) taking 0.25 for the soil adjustment variable (Huete 1988), Ratio Vegetation Index (RVI), Leaf area index (LAI) and Normalized Difference Vegetation Index (Allen et al. 2002).

We chose to perform our analyses at 100 m scale based on previous studies of LiDAR metrics as a function of spatial resolution. To achieve this, some statistics (minimum, maximum, mean, standard deviation and variance) were calculated, covering the sets allowed pixels within each grid.

Proving with this grid-size decision, Morel et al. (2011) resampling the images to the resolution of 100 m and Urbazaev et al. (2018) used the resampling of the nearest neighbors and the average of blocks to aggregate the pixels in a spatial resolution of 100 m. Morel et

Table 1 Summary of the Sentinel 1 and 2 satellites specifications

Sentinel 1		Sentinel 2	
Acquisition date	2016-02-15	Ingestion Date	2015-08-25
Product type	GRD	Satellite number	A
Instrument mode	IW	Instrument	MSI
Orbit number	9955	Orbit number	908
Relative orbit	83	Relative orbit	53
Polarisation	VV, VH	NSSDC identifier	2015-000A
Pass direction	Descending	Pass direction	Descending
Product Level	L1	Product Level	1C

al. (2011) emphasize that grid size is an important factor for correlating field data with SAR backscattering and that a 100 m grid is a reliable size, being less prone to errors related to the expansion of variables measured in smaller areas and the impact of geolocation errors. Garcia et al. (2018) emphasizes that pixel size also influences model accuracy, for estimating forest canopy height with 100 m spatial resolution.

LiDAR data and processing

The data used in this study were acquired from the Brazilian Sustainable Landscapes Project, supported by the Brazilian Agricultural Research Corporation (EMBRAPA), the United States Forest Service (USFS), the United States Agency for International Development (USAID) US Department of State. The LiDAR data used in this work refer to a flight with an instrument de named Optech Orion, conducted on September 21, 2015. Table 2 presents the survey parameters.

The LiDAR data were processed using FUSION software version 3.60 (McGaughey 2016), which was developed by the US Forest Service. We followed standard protocols for processing cloud LiDAR data, shown on Figure 2. First, the point cloud was cropped to the area of interest and points belonging to ground returns were identified in order to generate a digital terrain model (DTM).

Next, the LiDAR points clouds were subject to a normalization of values, in this step, the real height of each object

were obtained, such as the height of the trees, generating of digital surface model (DSM). Subsequently, several statistics were extracted at 100 m resolution from the LiDAR points clouds, totaling 286 squares (Figure 1). Finally, with the extraction of the statistics the values of AGB were obtained by means of Equation 1 (D’Oliveira et al. 2012):

$$AGB = (3.119 + 0.564 \cdot P25 + 0.062 \cdot Var)^2 + 1.74$$

Table 2 Airborne Specifications with LiDAR

Specifications	2011
Date of acquisition	2011/11/17
Datum	Sirgas 2000
UTM coordinate system	20S
Total area	500ha
Average pulse density/m ²	15.43 ppm ²
Average Flight Altitude	850 m
Field of vision	11.1 °
Scanning frequency	59.8 Hz
Overlay Percentage	65%

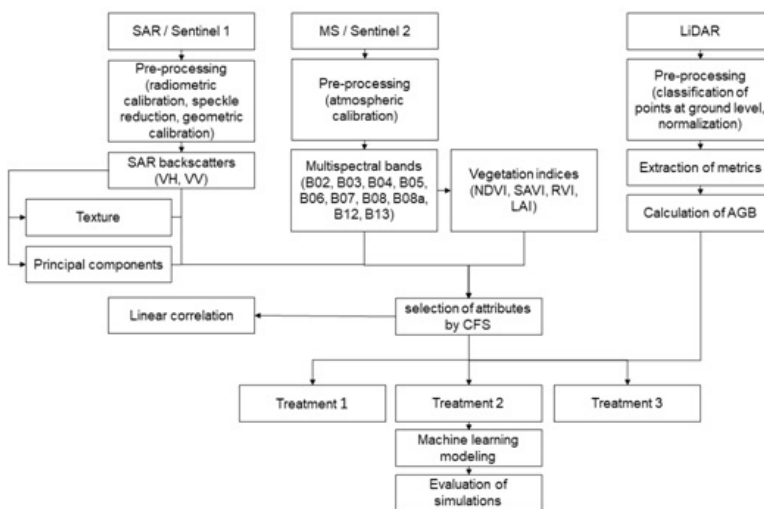


Figure 2 Workflow performed in this study

where: AGB - represents the dry biomass above the soil ($\text{Mg}\cdot\text{ha}^{-1}$, diameter at breast height trees ≥ 10 cm), P25 - equals the returns of the first quartile or 25% percentile, Var - elevation variation.

It was decided to use as a dependent variable the AGB estimation obtained from a survey with LiDAR technology mainly due to the unavailability of field data for the year of interest. The use of biomass data obtained from LiDAR measurements as the dependent variable has previously been used by Englhart et al. (2011) with satisfactory results.

Biomass modeling from Sentinel images

We compared three treatments that aimed at predicting AGB from different combinations of SAR and optical observations: (1) Sentinel-1 Sigma 0 at VV and VH polarizations (2) (1) plus Sentinel-1 textural metrics, (3) (2) plus Sentinel-2 bands (B02, B03, B04, B05, B06, B07, B08, B8a) and derived vegetation indices (LAI, RVI, SAVI, NDVI).

In order to build a parsimonious model, the CFS algorithm (CfsSubsetEval) implemented

in WEKA software (Waikato Environment for Knowledge Analysis) was used to find the best pairs of variables as well as their minimum, maximum, mean, standard deviation, and variance for the three treatments. CFS is a well-known data mining approaches that evaluates the value of a subset of attributes by considering the individual predictive capacity of each characteristic along with the degree of redundancy between them. The main idea of the CFS is that good variables include values highly correlated with the class, but not with each other. Thus, the data subsets are first selected using a heuristic search algorithm coupled with a Pearson correlation function (Hall 1999).

Having the input variables selected, we employed Huber's Robust Regression (Li 1985), using R software and the MASS package (R CORE TEAM 2017). All other modeling was performed in WEKA software version 3.8.2. The trainings / tests with the machine learning models used (Table 3), were constructed with 10 part (k) cross validation. In the cross validation of the WEKA software, the original sample is randomly divided into k subsamples.

Table 3 Machine learning algorithms used for this study available in WEKA software (Adapted from Castillo et al. 2017)

Algorithm	Classifier type	Key description
Multilayer Perceptron	Functions	Backpropagation to classify instances
SMOreg	Functions	Support vector machine for regression
Robust Regression	Functions	Weigh the observations differently based on how well behaved these observations are
Decision Stump	Trees	Building and using a decision stump
Random Forest	Trees	Construction a forest of random trees
Random Tree	Trees	Tree construction based on K-randomly chosen attributes
REP Tree	Trees	Fast decision tree learner
M5P	Trees	Implements the M5' model tree algorithm
IBk	Lazy	K-nearest neighbors classifier
Kstar	Lazy	Instance-based classifier
LWL	Lazy	Locally weighted learning

A single subsample k was maintained to validate the model and the remaining $k-1$ was used for model training. Then, this process was repeated k times, with each of the subsamples for validation. The evaluation of the performance of the algorithms was through the root mean square error (RMSE) in absolute and relative form.

The total error was computed by assuming independent error sources Saatchi et al. (2011b), as shown in Equation 1.

$$\varepsilon_{AGB} = (\varepsilon_{measurement}^2 + \varepsilon_{allometry}^2 + \varepsilon_{sampling}^2 + \varepsilon_{prediction}^2)^{1/2}$$

where: $\varepsilon_{measurement}$ - inventory measurement error, $\varepsilon_{allometry}$ - error associated with allometric equations to obtain AGB , $\varepsilon_{sampling}$ - spatial error

in upscaling lidar-derived estimates, $\varepsilon_{prediction}$ - prediction error.

Results

Database and attribute selection

AGB ranged between 139 to 516 Mg·ha⁻¹, with a mean value of 296 Mg·ha⁻¹ and a coefficient of variation of 20%. The distribution of AGB values is platykurtic (kurtosis 1.075) with right asymmetry (0.672).

Among the variations of measure of predictor variables (backscattering, GLCM textures, Sentinel 2 satellite bands and vegetation indices), the CFS algorithm selected five variables for treatment 1, nine variables for treatment 2

Table 4 Pearson (r) linear correlation between AGB and predictor variables, which were selected by the CFS algorithm

Treatment 1	r	Treatment 2	r	Treatment 3	r
SD (Backscattering band HV)	-0.094	Var (Backscattering band HV)	-0.107	Var (MAX band VH)	-0.098
Var (Backscattering band HV)	-0.107	Var (ASM band VH)	-0.096	Var (Backscattering band VV)	0.154
Min (Backscattering band VV)	-0.106	Var (Backscattering band VV)	0.154	Var (Correlation band VV)	-0.116
SD (Backscattering band VV)	0.122	Var (MAX band VH)	-0.098	Var (SAVI)	0.442
Var (Backscattering band VV)	0.154	Max (PCA band VH)	0.111	Var (band 13)	0.283
		Var (Homogeneity band VV)	-0.117	Var (band 2)	0.272
		Ave (Correlation band VV)	0.117	Var (band 3)	0.357
		SD (Correlation band VV)	-0.117	Var (band 4)	0.301
		Var (MAX band VV)	-0.116	Var (band 5)	0.331
				Min (band 12)	-0.414

Note. Abbreviations: Var - variance representation of the variable covering the grid of 100 m or 100 pixels of 10 m from Sentinel 1 and 2, Ave - represents the average of the variable within the grid of 100 m, SD - standard deviation of the variable, PCA - main components, MAX - maximum probability, Min- minimum, ASM - second angular momentum, CFS - CfsSubsetEval algorithm, AGB - above-ground biomass.

and 10 for treatment 3 (Table 4).

The linear correlation between AGB and the variables selected by the CFS algorithm was weak, and some variables showed a negative correlation, indicating an inverse behavior to that of the forest AGB.

Biomass modeling

The mean AGB for the 286 plots was 297 $\text{Mg}\cdot\text{ha}^{-1}$, and the RMSE for the machine-learning algorithms in treatment 1 ranged from 58.030 $\text{Mg}\cdot\text{ha}^{-1}$ (Robust Regression) to 83.651 $\text{Mg}\cdot\text{ha}^{-1}$ (Random Tree) and for treatment 2 the results were similar, where according to the RMSE the SMOreg algorithm presented the best performance (59.639 $\text{Mg}\cdot\text{ha}^{-1}$) and the worst was by the algorithm Random Tree (85.369 $\text{Mg}\cdot\text{ha}^{-1}$) (Table 5).

However, most of the algorithms fail to capture the range of AGB, producing intermediate values, that are to a large extent insensitive to the values of predictor variables. This behavior of the algorithms ends up affecting the statistics that do not show great average errors and can lead to hasty conclusions on the per-

formance of the algorithms.

For treatment 1 in which only the backscatter variables are used, on average the RMSE is 62.358 $\text{Mg}\cdot\text{ha}^{-1}$ or 21.011%, and for treatment 2, on average the RMSE is lower, 54.569 or 18.386%. However, when analyzing the boxplot (Figure 3), it is possible to see an improvement in the quality of the simulations in most models of machine learning when adding the texture information.

In Table 6 the balance of the algorithms can be set up against the addition of multispectral bands and indexes generated from the Sentinel 2 satellite (treatment 3).

It can be observed that in general the addition of multispectral information increased the performance of machine learning algorithms, with RMSE ranging from 47 to 70 $\text{Mg}\cdot\text{ha}^{-1}$, on average an improvement of 3%.

The algorithm that best estimated AGB for treatment 3, according to performance statistics, was the Robust Regression, which based on the idea of weighing the influential observations and assigning smaller weights to observations that could act as outliers. The algorithm was able to represent only 73% of

Table 5 Performance of the algorithms against the predictive variables of the treatment 1, use of backscattering values of the HV and VV bands of Sentinel 1 satellite and for treatment 2, addition of texture variables

Algorithm	Treatment 1		Treatment 2	
	RMSE ($\text{Mg}\cdot\text{ha}^{-1}$)	RMSE (%)	RMSE ($\text{Mg}\cdot\text{ha}^{-1}$)	RMSE (%)
Multilayer Perceptron	59.928	20.191	64.257	21.650
SMOreg	59.745	20.130	59.639	20.094
Robust Regression	58.030	19.552	57.759	19.461
Decision Stump	60.917	20.525	60.864	20.507
Random Forest	63.889	21.526	62.403	21.026
Random Tree	83.651	28.185	85.369	28.763
MSP	59.887	20.178	61.696	20.787
REP Tree	59.887	20.178	59.994	20.214
IBk	60.001	20.216	60.157	20.269
KStar	60.066	20.238	69.821	23.525
LWL	59.945	20.197	65.717	22.142

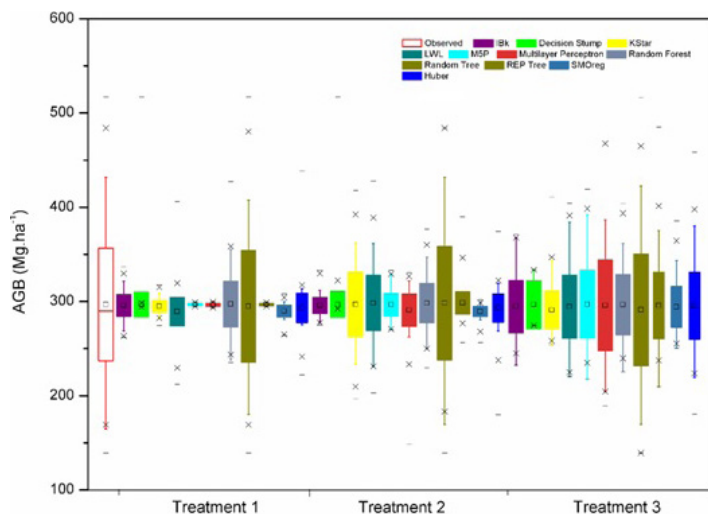


Figure 3 Boxplot of the actual and estimated values in each of the treatments

Table 6 Performance of the algorithms using the predictive variables of the treatment 3. Use of the backscatter, texture and with the addition of bands and by-products of the Sentinel 2 satellite

Algorithm	RMSE (Mg·ha ⁻¹)	RMSE (%)
Multilayer Perceptron	60.721	20.459
SMOreg	51.766	17.441
Robust Regression	47.660	16.058
Decision Stump	54.471	18.353
Random Forest	50.304	16.949
Random Tree	70.996	23.921
MSP	51.660	17.406
REP Tree	56.279	18.962
IBk	52.546	17.704
KStar	52.196	17.586
LWL	51.665	17.408

the range of values, with the Random Tree algorithm being the most suitable for the AGB simulation, being able to represent 100% of the forest variability with RMSE of 70.996 Mg·ha⁻¹ or 23.921%.

According to the boxplot analysis of the modeled AGB values (Figure 3) for the treatment 1 it can be concluded that the algorithm

that best represented the variability of AGB was Random Tree, which was able to represent 91.33% of the variability of values of AGB. For treatments 2 and 3, the same algorithm was applied to the others with 99.81% and 100% of the amplitude values, respectively. In Figure 4 it is possible to observe the relationship between the AGB observed and estimated by the algorithm Random Tree.

For the calculation of the total error, no measurement errors were considered due to the lack of inventory data, We considered the following errors: allometry error of 40.2 Mg·ha⁻¹ or 17.47% (D’Oliveira et al. 2012) which results in an error of 40.2 Mg·ha⁻¹ or 17.47%, when considering the total AGB of 230 Mg·ha⁻¹; sampling error of 22.8% following (Saatchi et al. 2011b) and prediction error of 23.921% for the Random Tree algorithm. The total errors following the methodology of the present study is approximately 41.1%.

Discussion

Analysis of the database

Backscatter values (C band, HV and VV polarizations) were poorly correlated with AGB, as it is believed that saturation of this band

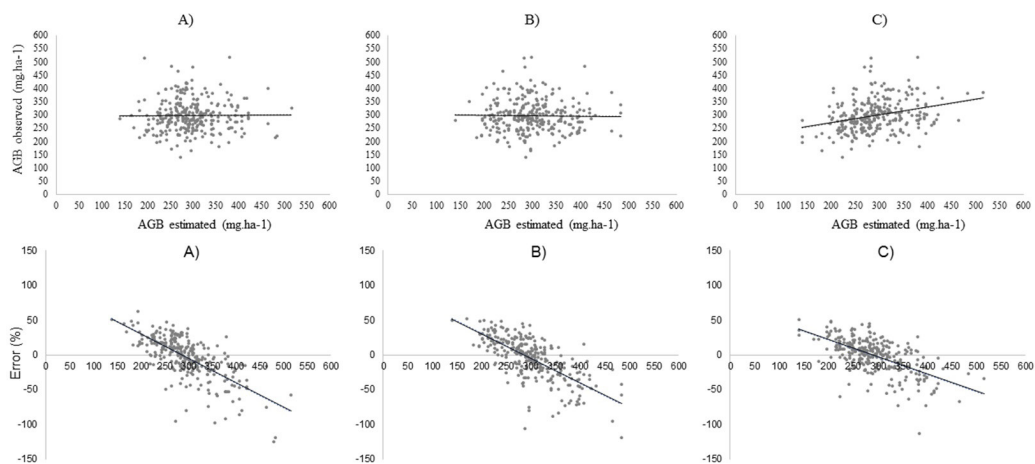


Figure 4 Graphic dispersion and residuals of the values observed and estimated by the Random Tree algorithm, being: A) treatment 1, B) treatment 2 and C) treatment 3, on the axis of the estimated values.

occurred because it was a site of high AGB. The L band presents a more appropriate wavelength compared to the C band for the AGB estimates and this presents the same behavior for these estimates in tropical forests (Lucas et al., 2006), so it is natural that the C band presents form saturation analogous. Similarly, for an area near the city of Manaus, also in the Brazilian rainforest, Cutler et al. (2012) found $r = 0.16$ between the SAR image of the JERS-1 satellite (L-band).

This low correlation behavior is not verified in other low AGB soil cover. As presented by Castillo et al. (2017), which evaluated areas of non-forest uses (agriculture, aquaculture and urban areas) and mangroves, where correlated values of AGB and backscatter were more strongly correlated, ranging from 0.28-0.84 for data from the same satellite.

The measure of variance of the soil-adjusted vegetation index (SAVI) was the variable that presented the highest correlation with AGB ($r = 0.442$), followed by the Sentinel 2 band 12, where the values are inversely related. More strongly correlated results ($r = 0.69-0.72$) were found by Castillo et al. (2017) when linking the multispectral bands of the Sentinel 2 satel-

lite with the AGB of mangroves. It is believed that this fact is related to the complexity of the forest structure in areas of tropical forest, compared to mangrove areas.

Biomass modeling

The use of the GLCM texture associated with the HV and VV bands of the Sentinel 1 satellite provided an improvement in AGB estimates by most of the algorithms. Corroborating with these results, Cutler et al. (2012) adjusted a series of RNAs for the estimation of AGB of the tropical forest using data from Brazil, Malaysia and Thailand. It was used the multispectral data from the Landsat TM satellite, SAR images of the JERS-1 satellite and texture images (wavelet and GLCM analysis). It is possible to conclude that the inclusion of the texture information of the SAR images to the multispectral data, improves the relation with the AGB. Kuplich et al. (2005) also emphasized that the addition of texture to the backscatter images increase the potential of using SAR images to estimate forest biomass.

Thapa et al. (2015) found Lower RMSE values compared to the present study, they used

SAR images derived from the ALOS PALSAR satellite and texture images for estimating the above-ground carbon stock in Indonesia through multiple linear regression. Only with the SAR images did the authors adjust the regression with RMSE of $45.64 \text{ MgC}\cdot\text{ha}^{-1}$, with the addition of the texture variables, the models improved, presenting RMSE between 28.01 and $37.70 \text{ Mg}\cdot\text{C}\cdot\text{ha}^{-1}$. For the treatment 3 in which all variables of greater relevance according to the selection of attributes by the CFS method participated in the modeling, the improvement in results was perceptible, were on average 3% statistically, but it is much more visually representative, where the Random Tree algorithm was able to estimate the AGB in its entire range of values.

Applying similar methodology to this study, Castillo et al. (2017) found lower RMSE values by taking advantage of images of this same Sentinel satellite, linear regression and machine learning to estimate and map the AGB of mangroves and non-forest areas in the Philippines. Using only the Sentinel 1 polarization, it found RMSE ranging from 27 to $55 \text{ Mg}\cdot\text{ha}^{-1}$, and when using only the multispectral bands of the Sentinel 2 satellite, it found RMSE values varying from 27 to $48 \text{ Mg}\cdot\text{ha}^{-1}$, a small improvement when comparing only with the use of SAR data, corroborating with the results obtained in the present study.

These results found by Castillo et al. (2017) can be justify by the conclusion of Urbazaev et al. (2018) who used a machine learning approach, the hybrid tree Cubist, for the estimation of AGB in Mexico. For this, it used the polarized images of ALOS PALSAR, DEM from the SRTM project, Landsat satellite image mosaic and metrics obtained by LiDAR as predictor variables. Was observe an underestimation of the higher AGB values, which justified the insufficient sensitivity of the satellite data to estimate the high AGB.

It is noteworthy the low RMSE for the three treatments and the good performance of the machine learning algorithms, result also found

by Garcia et al. (2018) that emphasize that these techniques are attractive because it does not make explicit assumptions about the distribution of the data and allow to model complex relations between the independent variables on the dependent one. Moreover, these algorithms have the advantage of being versatile and flexible, besides not having to take into account some regression assumptions, such as data normality and homoscedasticity of the variance (Sanquetta et al. 2015).

The quantification of total error considering the methodology of propagation of errors of Saatchi et al. (2011b) was 41.1%. It was believed that a non-quantifiable part of the errors in estimating AGB are related to soil and vegetation moisture conditions that were not fully account for by the lack of weather station in the study area, and which affect backscatter data. Another issue that can influence is the consideration of only part of the vegetation of arboreal size (above a certain minimum diameter of inclusion), leaving, therefore, part of the tree stratum not being considered. They are high values considering the total error, which has practical implications, but they are likely to improve, replacing the method of obtaining the AGB and sampling, which will consequently affect the performance of the prediction algorithms.

It is known that saturation of the backscatter occurs in bands with lower wavelengths in places with high AGB values, being indicated the use of the P band, since this one presents greater penetration and sensitivity with AGB (Saatchi et al, 2017). But unfortunately no sensor from the current satellites provide free images with such a feature. However, the launch of the BIOMASS satellite covering this spectrum range is planned for 2020 and will provide quantitative information on the structure of the forest from space (Le Toan et al. 2011)

Although this study was based on SAR, there are other options based on remote sensing that present great potential for the estimation of AGB and are expected to come into operation,

such as ICESat-2, NISAR, MOLI, and GEDI. One of ICESat-2's products will provide field and canopy height estimates (ATL08), which will facilitate assessments and monitoring of biomass and forest carbon at global scales (Neuenschwander & Pitts 2019). The NISAR satellite will be of double frequency (Band S and L) and free availability. The GEDI and MOLI satellites promise imaging with a laser altimeter, which will greatly contribute to the development of new research.

Conclusions

C-band SAR data present potential for AGB estimation in tropical forest. The use of this technology demonstrates that it is possible to quantify biophysical parameters of the forest with precision (RMSE: 23,9%) and low cost, since these data are freely available, allowing this methodology to be reproduced at regional scales.

The Random Tree algorithm stands out over the others in all treatments of input variables, being able to estimate the AGB in all the variability represented by the tropical forest used as the study area.

This study demonstrates the potential of using free data, allied to machine learning techniques for the estimation of AGB. The methodology allows the large-scale, regular and low-cost monitoring of tropical forest ecosystems.

Acknowledgments

The Sustainable Landscapes Brazil Project of the Brazilian Agricultural Research Company (EMBRAPA), for permission to use the LiDAR data for the preparation of this study. Data acquired with financial support from the US Agency for International Development (USAID) and the US Department of State (USDS) and with technical support from the US Forest Service.

References

- Allen RG., Tasumi M., Trezza R. 2002. SEBAL (Surface Energy balance Algorithms for land). Advanced Training and users' manual. Idaho: Implementation.
- Andersen HE., McGaughey RJ., Reutebuch SE. 2005. Estimating forest canopy fuel parameters using LiDAR data. *Remote Sensing of Environment* 94: 441-449. DOI: 10.1016/j.rse.2004.10.013
- Berninger A., Lohberger S., Stangel M., Siegert F. 2018. SAR-Based estimation of above-ground biomass and its changes in Tropical Forests of Kalimantan using L- and C-band. *Remote Sensing* 10: 1-22. DOI: 10.3390/rs10060831
- Carreiras J., Melo JB., Vasconcelos MJ. 2013. Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using l-band synthetic aperture radar data. *Remote Sensing* 5: 1524-1548. DOI: 10.3390/rs5041524
- Cartus O., Santoro M., Kelldorfer J. 2012. Mapping forest aboveground biomass in the northeastern united states with AlosPalsar dual polarization l-band. *Remote Sensing of Environment* 466-478. DOI: 10.1016/j.rse.2012.05.029
- Castillo JAA., Apan AA., Maraseni TN., Salmo SGS. 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 134: 70-85. DOI: 10.1016/j.isprsjprs.2017.10.016
- Chave J., Condit R., Aguilar S., Hernandez A., Lao S., Perez R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359: 409-420. DOI: 10.1098/rstb.2003.1425
- Chave J., Réjou-Méchain M., Búrquez A., Chidumayo E., Colgan MS., Delitti WBC., Duque A., Eid T., Fearnside PM., Goodman RC., Henry M., Martínez-Yrizar A., Mugasha WA., Muller-Landau HC., Mencuccini M., Nelson BW., Ngomanda A., Nogueira EM., Ortiz-Malavassi E., Pélissier R., Ploton P., Ryan CM., Saldarriaga JG., Vieilledent G. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biological* 20: 3177-3190. DOI: 10.1111/gcb.12629
- Cutler MEJ., Boyd DS., Foody GM., Vetrivel A. 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 66-77. DOI: 10.1016/j.isprsjprs.2012.03.011
- DeGrandi G., Lucas RM., Kropacek J. 2009. Analysis by wavelet frames of spatial statistics in SAR data for characterising structural properties of forests. *IEEE Transactions on Geoscience and Remote Sensing* 47: 494-507. DOI: 10.1109/TGRS.2008.2006183
- Deo R., Russell M., Domke G. 2017. Using Landsat

- Time-series and LiDAR to Inform Aboveground Forest Biomass Baselines in Northern Minnesota, USA. *Canadian Journal of Remote Sensing* 43: 28–47. DOI: 10.1080/07038992.2017.1259556
- D'Oliveira MVN., Reutebuch SE., McGaughey RJ., Andersen HE. 2012. Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sensing of Environment* 124: 479-491. DOI: 10.1016/j.rse.2012.05.014
- Englhart V., Keuck V., Siegert F. 2011. Aboveground biomass retrieval in tropical forest – The potential of combined X and L band Sar data use. *Remote Sensing of Environment* 115: 1260-1271. DOI: 10.1016/j.rse.2011.01.008
- ESA. 2018. Sentinels Scientific Data Hub, European Space Agency. Available online: <https://scihub.copernicus.eu/>. Accessed 01.02.2018
- European Space Agency (ESA) 2015. Sentinel-2 User Handbook, 64 p.
- FAO. 2015. Global Forest Resources Assessment. FAO Forestry Paper n. 1. Rome.
- Garcia M., Saatchi S., Ustin S., Balzter H. 2018. Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation* 66: 159-173. DOI: 10.1016/j.jag.2017.11.017
- Ghasemi N., Reza Sahebi M., Mohammadzadeh A. 2011. A review on biomass estimation methods using synthetic aperture radar data. *International Journal of Geomatics and Geosciences* 1: 776–788.
- Hall MA. 1999. Correlation-based Feature Selection for Machine Learning. Department of Computer Science University of Waikato, Hamilton, New Zealand.
- Haralick RM. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67: 786-804.
- Houghton RA., Hall F., Goetz SJ. 2009. Importance of biomass in the global carbon cycle. *Journal of Geophysical Research* 114: G00E03. DOI: 10.1029/2009JG000935
- Huete AR. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing Environment* 25: 295-309. DOI: 10.1016/0034-4257(88)90106-X
- IBGE. 2012. Manual Técnico da Vegetação Brasileira. 2ª ed. revisada e ampliada. [Technical manual of the Brazilian vegetation], Rio de Janeiro: IBGE- Diretoria de Geociências, 271 p.
- Joshi N., Mitchard ETA., Brolly M., Schumacher J., Fernández-Landa A., Johannsen VK., Marchamalo M., Fensholt R. 2017. Understanding 'saturation' of radar signals over forests. *Scientific Reports* 7: 3505.
- Kirimi F., Kuria DN., Thonfeld F., Amler E., Mubea K., Misana S., Menz G. 2016. Influence of vegetation cover on the oh soil moisture retrieval model: a case study of the Malinda Wetland, Tanzania. *Advances in Remote Sensing* 5: 28-42. DOI: 10.4236/ars.2016.51003
- Kuplich TM., Curran PJ., Atkinson PM. 2005. Relating SAR image texture to the biomass of regenerating tropical forests. *International Journal of Remote Sensing* 6: 4829-4854. DOI: 10.1080/01431160500239107
- Li G. 1985. Robust regression. In: Hoaglin D.C., Mosteller F., Tukey J.W., Exploring data tables, trends, and shapes. Wiley, 527 p.
- Le Toan T., Quegan S., Davidson MWJ., Balzter H., Pailou P., Plummer S., Papathanassiou K., Rocca F., Saatchi S., Shugart H., Ulander L. 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing Environment* 115: 2850-2860. DOI: 10.1016/j.rse.2011.03.020
- Lu DS. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27: 1297–1328. DOI: 10.1080/01431160500486732
- Lucas RM., Cronin N., Moghaddam M., Lee A., Armston J., Bunting P., Witte C. 2006. Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. *Remote Sensing Environment* 100: 388-406. DOI: 10.1016/j.rse.2005.09.020
- Mahdianpari M., Motagh M. 2017. Random Forest Wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSar-X imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 13-31. DOI: 10.1016/j.isprsjprs.2017.05.010
- Mcgaughey RJ. 2016. FUSION / LDV: software para análise e visualização de dados LIDAR. Seattle, WA, USFS, pp. 11.
- Minh DHT., Le Toan T., Rocca F., Tebaldini S., Villard L., Réjou-Méchain M., Phillips OL., Feldpausch TR., Dubois-Fernandez P., Scipal K., Chave J. 2016. SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. *Remote Sensing Environment* 175: 138-147. DOI: 10.1016/j.rse.2015.12.037
- Morel AC., Saatchi SS., Malhi Y., Berry NJ., Banin L., Burslem D., Nilus R., Ong RC. 2011. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management* 262: 1786-1798. DOI: 10.1016/j.foreco.2011.07.008
- Neuenschwander A., Pitts K. 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment* 221: 247-259. DOI: 10.1016/j.rse.2018.11.005
- Ningthoujam RK., Joshi PK., Roy PS. 2018. Retrieval of forest biomass for tropical deciduous mixed forest using ALOS PALSAR mosaic imagery and field plot data. *International Journal of Applied Earth Observation and Geoinformation* 69: 206-2016. DOI: 10.1016/j.jag.2018.03.007
- Oon A., Azhar B. 2019. Assessment of ALOS-2 PALSAR-2 L-band and Sentinel-1 C-band SAR backscatter for discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands. *Remote*

- Sensing Applications: Society and Environment 13: 183-190. DOI: 10.1016/j.rsase.2018.11.002
- Podest E., Saatchi S. 2002. Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing* 23: 1487-1506.
- R Core Team. 2017. R: Uma Linguagem e Ambiente para Computação Estatística. Web: <https://www.-project.org>. Accessed: 01.02.2018
- Ruiz LA., Hermosilla T., Mauro F., Godino M. 2014. Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates. *Forests* 5: 936-951. DOI: 10.3390/f5050936
- Saatchi S., Houghton RA., Dos Santos Alvala RC., Soares JV., Yu Y. 2017. Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology* 13: 816-837. DOI: 10.1111/j.1365-2486.2007.01323.x
- Saatchi S., Marlier M., Chazdon R., Clark D., Russell A. 2011a. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sensing Environment* 115: 2836-2849. DOI: 10.1016/j.rse.2010.07.015
- Saatchi SS., Harris NL., Brown S., Lefsky M., Mitchard ETA., Salas W., Zutta BR., Buermann W., Lewis SL., Hagen S., Petrova S., White L., Silman M., Morel A. 2011b. Supporting Information. In: Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1-9. DOI: 10.1073/pnas.1019576108
- Santi E., Paloscia S., Pettinato S., Fontanelli G., Mura M., Zolli C., Maselli F., Chiesi M., Bottai L., Chirici G. 2017. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. *Remote Sensing Environment* 200: 63-73. DOI: 10.1016/j.rse.2017.07.038
- Santos J., Lacruz M., Araujo L., Keil M. 2002. Savanna and tropical rainforest biomass estimation and spatialization using jers-1 data. *International Journal of Remote Sensing* 23: 1217-1229. DOI: 10.1080/01431160110092867
- Sanquetta CR., Wojciechowski J., Corte APD., Behling A., Péllico S No., Rodrigues AL., Sanquetta MNI. 2015. Comparison of data mining and allometric model in estimation of tree biomass. *BMC Bioinformatics* 16: 247. DOI: 10.1186/s12859-015-0662-5
- Silva CA., Hudak AT., Klauber C., Vierling AA., Gonzalez-Benecke C., Carvalho SOC. 2017. Rodriguez LCE, Cardil A. Combined effect of pulse density and grid cell size on predicting and mapping aboveground carbon in fast-growing Eucalyptus forest plantation using airborne LiDAR data. *Carbon Balance and Management* 12: 1-16. DOI: 10.1186/s13021-017-0081-1
- Singh M., Evans D., Friess DA., Tan BS., Nin CS. 2015. Mapping Above-Ground Biomass in a Tropical Forest in Cambodia Using Canopy Textures Derived From Google Earth. *Remote Sensing* 7: 5057-5076. DOI: 10.3390/rs70505057
- SNAP. 2018. Sentinels Application Platform software ver. 5.0. European Space Agency.
- Stickler CM., Nepstad DC., Coe MT., McGrath DG., Rodrigues HO., Walker WS., Soares-Filho BS., Davidson EA. 2009. The potential ecological costs and cobenefits of REDD: a critical review and case study from the Amazon region. *Global Change Biological* 15: 2803-2824. DOI: 10.1111/j.1365-2486.2009.02109.x
- Thapa RB., Watanabe M., Motohka T., Shimada M. 2015. Potential of high-resolution ALOS-PALSAR mosaic texture for aboveground forest carbon tracking in tropical region. *Remote Sensing Environment* 160: 122-133. DOI: 10.1016/j.rse.2015.01.007
- Urbazaev M., Thiel C., Cremer F., Dubayah R., Migliavacca M., Reichstein M., Schimmlius C. 2018. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance Management* 13: 1-20. DOI: 10.1186/s13021-018-0093-5
- Waring HR., Running SW. 2007. *Forest Ecosystems: Analysis at Multiples Scales*. 3rd edition, Academic Press, San Diego, USA, pp. 440