

Comparison between Quantitative and Qualitative Theme-Feature Forest Biomass Estimation Models built over SAR Data

Carlos Alberto Pires de Castro-Filho^{1,2}, Edilson Bias²

¹Brazilian Army, Brazil

²University of Brasília, Brasília

Received: 11 Jun 2021;

Received in revised form: 12 Jul 2021;

Accepted: 21 Jul 2021;

Available online: 30 Jul 2021

©2021 The Author(s). Published by AI
Publication. This is an open access article
under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>).

Keywords— Amazon Forest, Biomass,
Machine Learning, Remote Sensing, SAR.

Abstract— International organizations are still in need for methodologies that accurately measures forests above ground biomass (AGB). Among the remote sensing technologies, those of Synthetic Aperture Radar (SAR) stands out in the modeling of forest biomass due to their ability to characterize the geometry of the imaged region. The semantic representation, through thematic maps, is one of the main means for the geospatial situational understanding. However, there is a gap of knowledge for models that are built by the analysis of quantitative and qualitative theme-feature in a complementary way. This article aims to develop and compare forest biomass estimation models, through an innovative methodology, over quantitative and qualitative theme-features. To this end, extracted SAR data and specific machine learning (ML) and feature selection techniques are applied for each case. The models developed are based into forest inventories with 128 plots located in two different Brazilian Amazon Forest areas and were built over 231 extracted independent variables. The methodology applied used techniques to categorize numeric data and, afterwards, comparatively evaluate numeric quantitative and categorized qualitative results. The constructions of the models were based on ML algorithms such as Multilayer Perceptron, Support Vector Machine and Random Forest. The results showed that the different study areas had very different vegetation characteristics, significantly impacting the feature selection and ML algorithms. The different biomes of the Amazon Forest and their respective characteristics demanded specific models and techniques, not fitting into a single pattern. importance.

I. INTRODUCTION

In 2016 more than 190 countries participated in the 21st United Nations Conference of the Parties on Climate Change (COP-21), held in Paris. This conference aimed to continue the Kyoto Protocol, expired in 2012, and, consequently, to define goals regarding the emission of polluting gases into the atmosphere. Despite the intense

work, a legally binding treaty, capable of compelling the international community to cut greenhouse gas emissions, has not been signed. Among the reasons for this failure, one of the highlights was the lack of methodologies that accurately measures these cuts and establishes mechanisms for this reduction [1,2].

According to the United Nations Framework Convention on Climate Change – UNFCCC [3] the Article 3.4 of the Kyoto Protocol requires countries to report annually on changes in carbon stocks associated with forest biomass. The Intergovernmental Panel for Climate Change [4] and [5] states that reports with this information must follow a methodology based on the principles of transparency, consistency, comparability, completeness and accuracy.

However, [2,6-7] states that studies quantifying the carbon cycle between the atmosphere and forests are still needed. [2] points out that 53 to 58% of the carbon cycle comes from forests, therefore, accurate data on forest biomass are essential for many purposes, including subsidizing projects for environmental monitoring and Reducing Emissions from Deforestation and Forest Degradation (REDD +). [1,8] also states that forest biomass should be considered as a source of renewable energy and can be a source of income for national economies when used as carbon credit.

Among the remote sensing technologies, those of Synthetic Aperture Radar (SAR) stands out in the modeling of forest biomass due to their ability to characterize the geometry of the imaged region [1,2,6,8-12]. It also allows the monitoring and the verification of the type, direction, intensity and extent of the degradation in different areas, caused by human influence or by natural forest fires [6,13-16]. Due to the good results obtained by researchers, new projects that aims to use SAR data to estimate biomass are under execution or planning [6]. The Japan Aerospace Exploration Agency (JAXA) project, ALOS PALSAR 2, has been underway since 2014 and is a source of significant data for recent researches [14,17-20].

In Brazil, among the projects that aims to generate SAR images and that can be used in biomass estimation, the Amazon Radiography Project developed by the Geographic Service of the Army (DSG) stands out. By 2022, a total area of 1,800,000 km² of the Amazon region will be covered with airborne sensors in the X and P bands [21]. In addition to the 1:50,000 scale mapping, the project also has the potential to generate data to support infrastructure projects and sustainable exploitation of natural resources in the region [22-24].

Due to the large amount of data that can be originated from available SAR sensors, it is necessary to apply techniques that aims to organize and analyze quantitative and qualitative features in an intelligent and automated way [20,25-27]. Machine Learning – ML techniques are able to model knowledge and make associations between

different types of quantitative or qualitative information [28-29]. According to [30], the main advantages of ML are accuracy, since the optimal algorithm is selected from the characteristics of the data and the problem to be solved; automation in learning, which adjusts the models according to the success or failure of the results; processing speed; customization, being suitable in any type of problem; and scalability, as they are processes that adapt to data growth.

One of the possible applications in ML is the development of models involving thematic issues and those resulting in qualitative theme-attributes [28-29]. In these cases, the theme-attribute is commonly used for the construction of thematic maps that includes different areas of human geography, from the spatial representation of health and social geography [31-33], to characteristics related to forest biomass stocks [2,12-13, 16-18]. The semantic representation, through thematic maps, grows in importance, being one of the main means for the geospatial situational understanding and, consequently, the implementation of public administrations [34-35].

Recent published researches referring to biomass estimation presents ML originated models which output results are quantitative theme-attribute, that is, numerical [1,16,18-19]. However, studies that builds and analyzes simultaneously quantitative and qualitative theme-attributes models were not observed. Therefore, it is necessary researches that seeks to cover this gap of knowledge and that aims at building thematic maps models using, in a complementary way, quantitative and qualitative theme-attributes.

This article aims to develop and compare forest biomass estimation models built over quantitative and qualitative theme-feature based on extracted SAR data. To this end, machine learning and feature selection techniques are specifically selected and applied for each case.

II. METHOD

2.1 Study Area and data

The study areas are located in different geographical regions of the Brazilian Amazon rain forest: São Gabriel da Cachoeira (SGC), a municipality located on the banks of the Rio Negro, in the northwest of the state of Amazonas; and the Unini River Extractive Reserve (Unini River ExRes) located in the Unini River basin, in the municipality of Barcelos. The areas, in white, are highlighted in Figure 1, together with the location of some of the inventoried plots, in green.

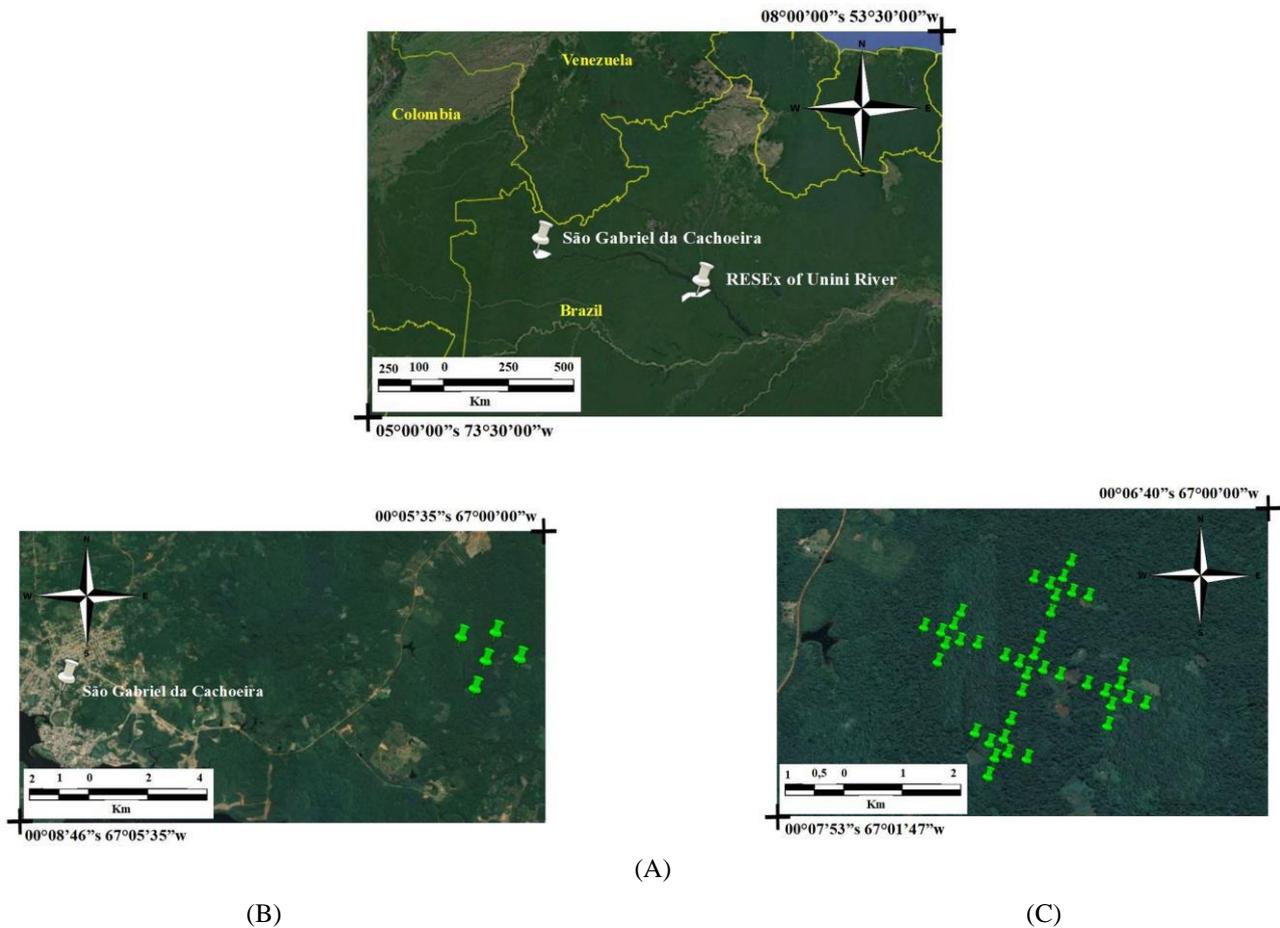


Fig.1:(a) Study areas, highlighted in white; (b) São Gabriel da Cachoeira region; (c) Location of a subset of plots inventoried and arranged in the shape of Maltese Cross.

The areas were selected for two reasons: the distinct phytoecological and land use and occupation situations and the availability of data. The SGC area has hybrid characteristics, composed of anthropized regions together with dense vegetation. In contrast, the Unini River ExRes area is composed only of primary virgin forest vegetation.

According to [31], the vegetation found in the study areas is of forest formation. More specifically, [32] indicates that the vegetation found in the São Gabriel da Cachoeira area is composed by phytoecological forest contact / edaphic formations regions (*campinaranas*). These regions are characterized in three ways:

- (1) dense, submontane forests with dissected relief. [32] states that the average AGB volume in the area is 107.4 m³/ha;
- (2) dense, submontane and undulating forests; and
- (3) dense forests, lowlands and relief with the presence of plateaus.

The Unini River ExRes, in its turn, is an extractive conservation unit with about 833 hectares in length and characterized in [32] as:

- (1) dense tropical forest, referring to the sub-region of the low plateaus of the Amazon; and
- (2) areas of ecological tension with dense alluvial presence.

The remote sensing data was obtained from the ALOS PALSAR 2 sensor and the Amazon Radiography Project. The working areas are comprised between 0° and 1° south latitudes and 67° and 68° west longitudes, for the region of São Gabriel da Cachoeira; and between 1° and 2° south latitudes and 62° and 63 ° west longitudes, for the Unini River ExRes.

The data from ALOS PALSAR 2 were provided by IBAMA and are Level 1.1 – Single Look Complex (SLC) processing images in the quadri-polarized strip-map imaging mode.

The Amazon Radiography Project data were provided by the [21] with the following characteristics:

- (1) amplitude orthoimages in X band HH polarization and P band quadri-polarized, all with 16 bits radiometric resolution and 5 meters spatial resolution;
- (2) digital surface models (DSM) and digital terrain models (DTM) generated, respectively, from the interferometric processing of X and P data, with 32 bits radiometric resolution and 5 meters spatial resolution.

The AGB data were provided by the National Institute of Amazon Researches – INPA, and follow the methods developed by [33] and described by [34]. In addition to the exact same geographical position as the images, the proximity to the region's imaging date was also important as it aims to avoid major changes in the analyzed vegetation.

The given biomass data provided was composed of 128 inventoried plots, 58 plots of São Gabriel da Cachoeira and 70 of Unini River ExRes, presenting the AGB values (ton/ha) and the UTM coordinates of the start and end points of each plot. As pointed out by [35-36], different allometric equations were used to calculate the inventoried plots due to the characteristics of the region. Figure 2 illustrates the format, the start (P1) and end (P2) points and the arbitrary coordinates of each arboreal individual within the plot.

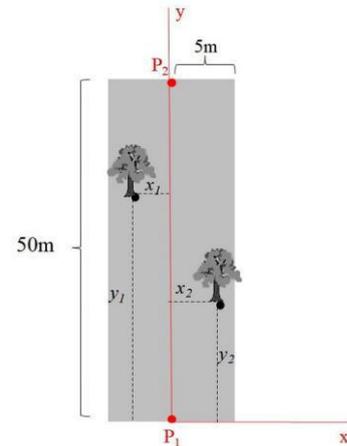


Fig.2:Plot of forest inventory.

Fig.2:Plot of forest inventory.

2.2 Methodological approach

The research was structured according to the flowchart shown in Figure 3. Each step is described in the following subitems.

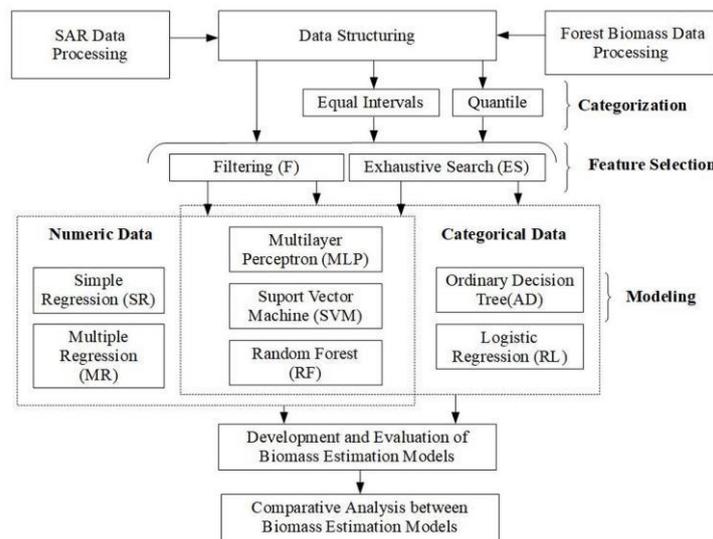


Fig.3: Methodological Flowchart.

2.2.1 Forest Biomass Data Processing

Using analytical geometry techniques, the UTM coordinates of each 4 corners of the inventoried plots were calculated and the respective vector files for each region of interest (ROI) were generated.

2.2.2 SAR Data Processing

In this stage, the ALOS PARSAR 2 images, obtained in SLC format, were processed and the features on the available X, L and P bands were extracted. All processing steps were performed using the Polarimetric SAR Data Processing and Educational Tool (PolSARpro), version 6.0

(Biomass Edition), from the European Space Agency (ESA).

The ALOS PALSAR 2 images were processed according to the flowchart shown in Figure 4. The following parameters were used:

- multilook processing with 2 looks for the rows and 1 look for the columns, as suggested by [19];
- Lee Refined speckle filter with 2 looks and 7x7 size window;
- calculation of the covariance [C] and coherence [T] matrices images, both 3x3;
- geocoding of the coherence matrix image [T], performing the correction of the Range-doppler terrain and the respective georeferencing using the digital elevation model automatically extracted from the Shuttle Radar Topography Mission (SRTM), with 90m spatial resolution;
- polarimetric calibration and conversion to sigma-nought (σ^0) using Equation 1, where the DN is the Digital Number, in amplitude, and CF is the calibration factor in dB for the channels [37]. The value applied for the CF was -83; and
- application of target decomposition techniques.

$$\sigma_0 = 10 * \log_{10} \langle DN^2 \rangle + CF \quad (1)$$

At the end of the SAR data processing, the interferometric, incoherent and coherent features were extracted according to Table 1.

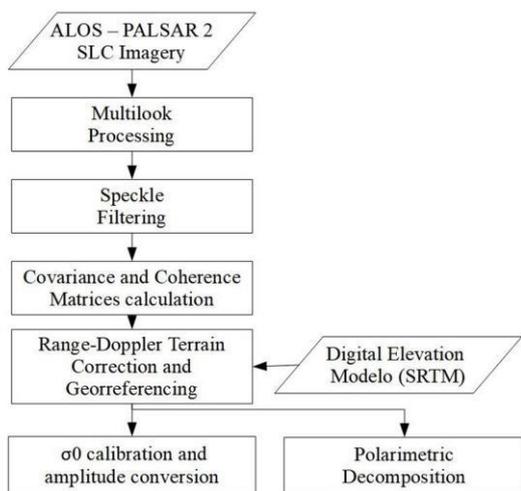


Fig.4: ALOS PALSAR 2 image processing. Adapted from [19]

Table.1: Extracted Features from SAR Data

Symbol	Description
SAR Interferometric Features	
H_{int}	Interferometric height – It is the difference in altitude between the Digital Surface Model (MDS), obtained with the X band, and the Digital Terrain Model (MDT), obtained with the P band. It represents the height of the vegetation.
Decliv	Declivity – It is the slope of the land surface in relation to the horizontal, obtained through the MDT.
Incoherent SAR Features	
Xhh	Amplitude image of the X band in the HH polarization – The backscatter of the forest canopy.
Lhh, Lhv, Lvv	Amplitude image of the L band in the polarizations HH, HV or VV – Represents the main geometric characteristics of arboreal individuals.
Phh, Phv, Pvv	Amplitude image of the P band in the polarizations HH, HV or VV – Associated with the main geometric characteristics of the terrain.
Lhh-Lhv, Lhh-Lvv, Lvv-Lhv	Subtraction between amplitude images in the L band polarizations.
Phh-Phv, Phh-Pvv, Pvv-Phv	Subtraction between amplitude images in the P band polarizations.
PC1L, PC2L, PC3L	Principal Components of the amplitude images in the L bands polarizations.
PC1P, PC2P, PC3P	Principal Components of the amplitude images in the P bands polarizations.
Henderson and Lewis Polarimetric Decomposition Features [38]	
PR_L, PR_P	Ratio between parallel polarizations (Parallel Ratio – PR) in the L or P bands ($PR_{Band} = Band_{vv} / Band_{hh}$) – Associated with the orientation and shape of the backscatter elements in the forest.

CR_L, CR_P	Ratio between crossed polarizations (<i>Crossed Ratio – CR</i>) in the L or P bands ($CR_{Band} = Band_{hv} / Band_{hh}$) – Referring to the volumetric backscatter of the target.
TotPow_L, TotPow_P	Total power of the L or P bands ($TotPow_{Band} = Band_{hh} + Band_{vv} + 2 * Band_{hv}$) – They represent the sum of all backscatter mechanisms occurring in the forest.
Pope Polarimetric Decomposition Features [39]	
BMI_L, BMI_P	Biomass index in bands L or P ($BMI_{Band} = (Band_{hh} + Band_{vv}) / 2$) – Indicator of the amount of woody structure in the forest.
CSI_L, CSI_P	Canopy structure index in the L or P bands ($CSI_{Band} = Band_{vv} / (Band_{vv} + Band_{hh})$) – Compares the vertical structure with the horizontal vegetation.
VSI_L, VSI_P	Volumetric scattering index in the L or P bands ($VSI_{Band} = Band_{hv} / (Band_{hv} + BMI_{Band})$) – Related to the density of the canopy, being directly proportional to the amount of elements that cause multiple type scattering.
Kim and Zyl Polarimetric Decomposition Features [40]	
RVI_L, RVI_P	Radar vegetation index ($RVI_{Band} = 8 * Band_{hv} / (Band_{hh} + Band_{vv} + 2 * Band_{hv})$) – Associated with the proportion of vegetation in the soil.
Haralick Textural Features [41]	
The co-occurrence texture features analyzes the relationship between pixel pairs values within a window and constructs a Grey Level Co-occurrence Matrix (GLCM). In the texture equations, $P(i, j)$ is the co-occurrence probability of each pixel value in column i and row j ; N_g is the number of distinct grey levels in the quantized image; μ is the average value of P ; σ is the x or y deviation pattern of the image.	
J_Me_Band	Mean ($Me = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i * P(i,j)$) value within the GLCM.

J_Va_Band	Variance ($Va = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P(i,j)$) value within the GLCM.
J_Ho_Band	Homogeneity ($Ho = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \frac{1}{1+(i-j)^2}$) is the spatial correlation measurement in the GLCM.
J_Con_Band	Contrast ($Con = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j)(i-j)^2$) is the intensity difference between the reference pixels and its neighbors in the GLCM.
J_Di_Band	Dissimilarity ($Di = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) i-j $) is the amplitude difference between the reference pixels and its neighbors in the GLCM.
J_En_Band	Entropy ($En = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \log(P(i,j))$) value represents the randomness between the elements of the GLCM
J_Se_Band	Second Moment ($Se = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j)^2$) is the second angular moment between the elements of the GLCM.
J_Cor_Band	Correlation ($Cor = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i,j) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$) is the statistical difference between the reference pixels and its neighbors in the GLCM.
Coherent SAR Features	

Cloude and Pottier Polarimetric Decomposition Features [42]	
Alpha	α angle – Dominant type of scattering.
H	Entropy – Proportion in the importance of the dominant type of scattering.
A	Anisotropy – Proportion in the importance of the secondary and tertiary types of scattering.
Freeman and Durden Polarimetric Decomposition Features [43]	
FD_Vol	Volumetric – Contribution of the type of volumetric scattering, simulating the forest canopy.
FD_Dbl	Double Bounce – Result of a set of dihedral corner reflectors.
FD_Odd	Superficial – Contribution of the type of surface scattering.
Touzi Polarimetric Decomposition Features [44]	
TAlfa_S1, TAlfa_S2, TAlfa_S1, TAlfa_Sm	Magnitude (α) - Provides the type of symmetry related to the type of scattering of the target.
TPhi_S1, TPhi_S2, TPhi_S1, TPhi_Sm	Phase (ϕ) - Represents a more complete characterization of the target's scattering type.
TTau_S1, TTau_S2, TTau_S1, TTau_Sm	Helical angle (τ) - Allows the measurement of the target's degree of symmetry, distinguishing symmetric and asymmetric scattering.
TPsi_S1, TPsi_S2, TPsi_S1, TPsi_Sm	Orientation angle (ψ) - Associated with the target's angle of inclination.
Van Zyl Polarimetric Decomposition Features [45]	
VanZ_Vol	Volumetric Scattering – Volumetric scattering proportion.
VanZ_Dbl	Double Bounce Scattering – Double

	bounce scattering proportion.
VanZ_Odd	Odd Scattering – Surface (odd) scattering proportion.
Yamaguchi Polarimetric Decomposition Features [46]	
Yam_Vol	Volumetric Scattering – Volumetric scattering proportion.
Yam_Dbl	Double Bounce Scattering – Double bounce scattering proportion.
Yam_Odd	Odd Scattering – Surface (odd) scattering proportion.

2.2.3 Data Structuring

The data extracted from SAR and the AGB data were organized in a single structured spreadsheet, having the features represented in columns and the instances, referring to each inventoried forest biomass plot, as rows. The AGB feature was defined as the theme-feature (or “result” or “output” feature) of the structured spreadsheet.

For each of the extracted features, the arithmetic mean of the pixels’ value corresponding to the areas of the inventoried AGB plots was calculated.

The numerical data was used in two different ways. First, using the original values of the explanatory feature set $x = (x_1, x_2, \dots, x_p)^T$, so that the multiple regression model would be as shown in Equation 2. Second, with the logarithmic of the original value, as Equation 3. In all cases p is the number of variables, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ is the parameter set, y is the dependent AGB variable and ε is the random error.

$$y = \beta_0 + \beta_1x_1 + \dots + \beta_px_2 + \varepsilon \quad (2)$$

$$\ln(y) = \ln(\beta_0) + \beta_1\ln(x_1) + \dots + \beta_p\ln(x_p) + \varepsilon \quad (3)$$

2.2.4 Categorization

The numerical data of the AGB quantitative feature were categorized and associated with one of the 5 (five) categories of biomass: "Low", "Medium-Low", "Medium", "Medium-High" and "High". The categorization methods, used to transform quantitative to qualitative features, were of the equal intervals and of the quantile.

According to [47], the method of equal intervals is performed by dividing the theme-feature values in the domain range by the number of categories of interest. In Equation 4, K is the number of categories defined by the user, x_{\min} and x_{\max} , respectively, the minimum and

maximum values observed in the theme-feature and δ the value of the widths for each category interval.

$$\delta = (x_{max} - x_{min}) / K \quad (4)$$

In the quantile method, categorization is performed by dividing the total number of instances N by the number of categories of interest K . Therefore, at the end of this method each category will have the same number of objects.

At the end of the categorization stage, the theme-feature was classified in one of three possibilities: numeric (NumThFe), categorical by the “equal intervals” method (EqIntThFe) and categorical by the “quantile” method (QuThFe). Then, all other steps were performed for each of these cases.

2.2.5 Feature Selection

Tests were performed using the filtering type feature selection, in comparison to the exhaustive search including all features extracted from SAR data. The objective was to verify the impacts of the feature selection process on the quality of the final AGB models developed.

The feature selection technique performed was the Correlation-based Feature Subset (CFS) Selection, as described [48]. In this case, the search method used was the greedy Best First, which performs the “hill climb” heuristic in the “forward” direction.

According to [49], the CFS feature selection method is adequate to identify features that are related to the AGB by using the Pearson correlation coefficient method.

2.2.6 Modeling

In the specific cases in which the constructions of the models were based on numerical quantitative data, that is, when the theme-feature has not been categorized, the methods of simple statistical regression – SR and multiple statistical regression – MR were used. On the other hand, for the specific cases of the qualitative categorized data, the methods of logistic statistical regression – LR and ordinary decision tree – ODT were applied.

In addition to these methods, the Multilayer Perceptron – MLP, Support Vector Machine – SVM and Random Forest – RF methods were used for all cases.

The feature selection and the model development steps were carried out entirely in the WEKA (Waikato Environment for Knowledge Analyzes) system, version 3.8.4, and followed algorithms described by [50].

2.2.7 Development and Evaluation of a Biomass Estimation Model

After the development of the models, the evaluation stage is carried out. In the case of the models based on numerical data, such as those of statistical regression, there are several parameters that can be observed and that reflects the assessment. The parameter used in this case was the correlation coefficient (r), described by [51].

In the case of the models based on categorized qualitative data, the assessment was made by building a confusion matrix and calculating the respective Kappa coefficient of agreement [52]. Due to the reduced number of instances, the process of cross-validation divided into 10 folds was used, as suggested by [53].

2.2.8 Comparative Analysis between Biomass Estimation Models

Initially, the selected models were those that obtained the best correlation coefficient, in the case of the numerical quantitative data, and best Kappa coefficient, for the models based on categorized qualitative data.

In order to compare those different type of models, the numerical values resulting from the AGB will follow the process described in the flowchart presented in Figure 5. In this process, numerical quantitative values will be categorized using the equal intervals method, followed by the assessment obtained through the construction of the confusion matrices and calculations of the respective Kappa coefficients.

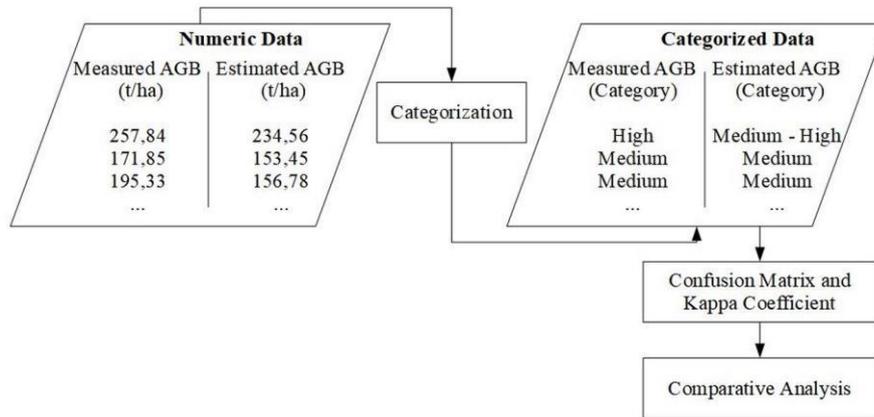


Fig. 5: Categorization process for comparative analysis.

III. RESULTS AND DISCUSSION

3.1 Forest Biomass Data Processing

From the AGB data granted by INPA, 3 sample sets were defined according to the region inventoried: São Gabriel da Cachoeira, Unini River ExRes and the joint regions. The statistics for each set, referring to the number of pixels and AGB in each plot, are shown in Table 2.

Table.2: Statistics for the number of pixels and AGB in the inventoried plots

Set	Joint Regions	
Statistics	Number of Pixels (un)	AGB (t/ha)
Mean	50,59	227,93
Minimum	35	92,21
Maximum	72	351,73
Standard Deviation	7,28	45,21
Number of Plots	128 plots	
Set	São Gabriel da Cachoeira	
Statistics	Number of Pixels (un)	AGB (t/ha)
Mean	50,17	224,95
Minimum	35	92,21
Maximum	69	351,73

Standard Deviation	8,19	52,24
Number of Plots	58 plots	
Set	Unini River ExRes	
Statistics	Number of Pixels (un)	AGB (t/ha)
Mean	50,93	230,40
Minimum	39	153,32
Maximum	72	311,57
Standard Deviation	6,48	38,65
Number of Plots	70 plots	

3.2 SAR Data Processing

Together with the features detailed in Table 1, the textural features were extracted for all available polarimetric bands, that is, Xhh, Phh, Phv, Pvv, Lhh, Lhv and Lv, for 3x3, 5x5 and 7x7 window sizes.

At the end of the SAR data processing, 231 features, or independent variables, were extracted, in addition to the theme-feature.

3.3 Categorization

The categorization by the equal intervals technique obtained a δ of 52 (t / ha). Therefore, the AGB categories were defined as: Low (below 100 t/ha); Medium-Low

(between 100 and 200 t/ha); Medium (between 200 and 250 t/ha); Medium-High (between 250 and 300 t/ha); and High (above 300 t/ha). The number of categorized instances was 2 (two) for the Low class, 38 (thirty-eight) for Medium-Low, 42 (forty-two) for Medium, 40 (forty) for Medium-High and 6 (six) for High.

The categorization by the quantile method obtained 25 (twenty-five) or 26 (twenty-six) instances for each category.

3.4 Feature Selection

The process was carried out separately for numerical quantitative and categorized qualitative data. The results of the 5 (five) selected features, in decreasing order of relevance, are shown in Table 3. In the same table Pearson's correlation values between the selected feature and the respective theme-feature, quantitative or qualitative, was calculated.

In general, the selected features showed low correlation with the biomass theme-feature. The highlight was the H_{int} feature, which achieved a good correlation with the quantitative data, in addition to being selected for both cases.

Table.3: Result of the feature selection process

Quantitative Data		Qualitative Data	
Feature	Correlation	Feature	Correlation
H_{int}	0.449975	PC3	0.1765
Lhh	-0.188703	H_{int}	0.1592
CSI_L	-0.046255	TAlphaS3L	0.1059
FreeOddL	0.125393	7x7_Xhh_S e	0.2772
TPhiS1L	0.10413	7x7_Phh_M e	0.2851

3.5 Development of Biomass Estimation Models

The ML techniques applied in the biomass estimation modeling had the following specific configurations:

(1) SVM – the model applied to numerical quantitative data was the SMOreg, specific for statistical regression, as described by [54]. The complexity parameter c was 1.0 and the Radial Basis Function (RBF) kernel used 0.01 gamma;

(2) MLP – the models not submitted to the feature selection process were built with one (composed of 50 nodes) or two (composed of 50 and 10 nodes) hidden layers. The models submitted to the feature selection process were built with one (composed of 5 nodes) or two (composed of 5 and 5 nodes) hidden layers;

(3) RF – the parameter of 100 trees was used in the construction of the model;

(4) ODT – the minimum quantity of 2 instances per node was applied.

The correlation and kappa coefficients resulting from the tests are shown in Tables 4, 5, 6 and 7 and have the following characteristics:

(1) Tables 4 and 5 refers to models based on numerical quantitative and Tables 6 and 7 to models based on categorized qualitative theme-features;

(2) Tables 4 and 6 refer to the original values and Tables 5 and 7 refer to log values of the features ;

(3) the values before the bars (/) are those obtained by models that have not been submitted to the feature selection process, while the values after the bars are those referring to models with selected features;

(4) the results in MLP models with an asterisk (*) are those obtained with 2 (two) hidden layers and that obtained results superior to those of a single hidden layer;

(5) the results in bold are the best obtained, having been highlighted 2 (two) results for each type of region and for each type of data (quantitative or qualitative).

Table.4: Correlation coefficients of AGB estimation models for numerical quantitative theme-feature and original feature values.

ML Technique	Joint Regions	São Gabriel da Cachoeira	Unini River ExRes
SR	0.42 /0.42	0.39 /0.39	0.35 / 0.43
MR	0.21 /0.40	0.02 /0.41	0.04 /0.38
SVM	0.12 /0.21	0.13 /0.13	0.35 /0.12
MLP	0.07 /0.32*	0.12 / 0.70	0.13 /0.23
RF	0.16 /0.39	0.21 /0.33	0.14 /0.29

Table.5: Correlation coefficients of AGB estimation models for numerical quantitative theme-feature and logarithmic feature values.

ML Technique	Joint Regions	São Gabriel da Cachoeira	Unini River ExRes
SR	0.49 / 0.54	0.49 / 0.58	0.30 /0.30
MR	0.09 /0.41	0.04 /0.25	0.01 /0.31
SVM	0.20 /0.22	0.16 /0.10	0.29 /0.06
MLP	0.33 */ 0.49	0.26 */0.52*	0.06 / 0.36*
RF	0.14 /0.39	0.14 /0.47	0.19 /0.25

Table.6: Kappa index of AGB estimation models for categorized qualitative theme-features and original feature values.

ML Technique	Joint Regions		São Gabriel da Cachoeira		Unini River ExRes	
	Equal Intervals	Quantile	Equal Intervals	Quantile	Equal Intervals	Quantile
LR	0.10 /0.22	0.22 /0.15	0.25 /0.10	0.20 /0.10	0.18 /0.35	0.30 /0.33
MLP	0.22 / 0.38	0.32 /0.15	0.18 /0.02	0.13 /0.07	0.31 /0.29	0.14 /0.19
SVM	0.09 /0.01	0.04 /0.01	0.01 /0.01	0.01 /0.01	0.25 /0.01	0.10 /0.01
ODT	0.09 /0.19	0.11 /0.11	0.09 /0.01	0.04 /0.01	0.22 / 0.48	0.27 /0.21
RF	0.13 /0.28	0.19 /0.25	0.30 /0.16	0.24 /0.01	0.19 /0.38	0.26 /0.28

Table.7: Kappa index of AGB estimation models for categorized qualitative theme-features and logarithmic feature values.

ML Technique	Joint Regions		São Gabriel da Cachoeira		Unini River EsRes	
	Equal Intervals	Quantile	Equal Intervals	Quantile	Equal Intervals	Quantile
LR	0.23 /0.23	0.21 /0.18	0.21 /0.24	0.26 /0.12	0.20 /0.35	0.28 /0.31
MLP	0.36 /0.24	0.18 /0.17	0.30 /0.12	0.22 /0.16	0.36 / 0.47	0.28 /0.32
SVM	0.05 /0.01	0.05 /0.01	0.01 /0.01	0.02 /0.01	0.01 /0.01	0.06 /0.01
ODT	0.11 /0.22	0.18 /0.12	0.07 /0.08	0.08 /0.03	0.21 /0.39	0.18 /0.32
RF	0.24 /0.22	0.22 /0.20	0.26 /0.11	0.26 /0.06	0.24 /0.39	0.31 /0.30

3.6 Comparative Analysis between Biomass Estimation Models

As observed in Tables 4, 5, 6 and 7, in general, there was an emphasis on MLP and SR techniques, corresponding to

58% and 25% of the highlighted results, respectively. MR, RF and ODT techniques achieved results close to the best, however, with a single highlight. The SVM technique

showed results significantly lower than the other techniques.

In the case of the numerical quantitative theme-feature, presented in Tables 4 and 5, only the MLP and SR techniques showed outstanding results. The MR technique was not able to increase the r from the input of new features.

The models developed for the categorized qualitative theme-feature, Tables 6 and 7, showed an increase in results for non-parametric techniques, including MLP, RF and ODT.

The models submitted to the feature selection process showed improvement in 73% of the numerical quantitative theme-feature cases. In these cases, only 10% worsened the results, all of which refers to the SVM technique.

On the other hand, for the case of categorized qualitative theme-feature submitted to the feature selection process, the percentages of improvement, worsening and maintenance of the results were, respectively, 35%, 10% and 55%. In this case, there was no correlation to the ML technique.

Regarding the categorization method, all the best results were obtained using the method of equal intervals. Despite this, considering all cases, there was not a conclusive difference in the results between the categorization methods.

The different areas analyzed also presented different results. For the case of the numerical quantitative theme-feature, the São Gabriel da Cachoeira region obtained the best results, unlike the region of the Unini River ExRes with the worst results. The opposite result was obtained for the case of the categorized qualitative theme-feature. In both cases, the results for the joint regions, as they aggregate data from both study areas, were average.

In order to carry out the comparative analysis, the process shown in Figure 5 was applied. The comparative analysis was performed on data from the same regions (Joint Regions, SGC or Unini River ExRes), separately for quantitative or qualitative data. The results obtained are shown in Tables 8, 9, 10, 11, 12 and 13. In all cases, 3 (three) types of Z hypothesis tests were performed, with a significance level (α) of 0.05:

In order to carry out the comparative analysis, the process shown in Figure 5 was applied. The comparative analysis was performed on data from the same regions (Joint Regions, SGC or Unini River ExRes), separately for quantitative or qualitative data. The results obtained are shown in Tables 8, 9, 10, 11, 12 and 13. In all cases, 3 (three) types of Z hypothesis tests were performed, with a significance level (α) of 0.05:

- test to analyze the hypothesis of Kappa * (value referring to the first selected model) being equal to zero;
- test to analyze the hypothesis of Kappa ** (value for the second selected model) to be equal to zero;
- and test to analyze the hypothesis whether the difference between Kappa * and Kappa ** is significantly greater (or lower) than zero, that is, if both are significantly different.

Table.8: Comparative analysis between confusion matrices: numerical quantitative theme-feature of the joint region.

Categorized	SR over logarithmic values (r=0.54)*					MLP over logarithmic values (r=0.49)**				
	Reference					Reference				
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High
Low	0	0	0	0	0	0	1	0	0	0
Medium-Low	2	6	2	4	0	2	6	1	0	0
Medium	0	9	21	9	3	0	8	18	13	3
Medium-High	0	1	1	1	1	0	1	5	1	0

High	0	0	0	0	0	0	0	0	0	0	1
Kappa*: 0.17; Kappa Variance*: 0.0057 Global Accuracy*: 47%						Kappa**: 0.13; Kappa Variance**: 0.0073 Global Accuracy**: 43%					
Analysis: Hypothesis Z-Test: Kappa* = 0 Kappa is significantly higher than zero (z=2.25; p-value=0.0123; α=0.05) Hypothesis Z-Test: Kappa** =0 Kappa** is significantly higher than zero (z=2.25; p-value=0.0123; α=0.05) Hypothesis Z-Test: Kappa*- Kappa**=0 Kappa*- Kappa** is significantly higher than zero (z=2.25; p-value=0.0123; α=0.05)											

Table.9: Comparative analysis between confusion matrices: numerical quantitative theme-feature, from SGC.

Categorized	MLP over original values (r=0.70)*					RS over logarithmic values (r=0.58)**					
	Reference					Reference					
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High	High
	Low	0	1	0	0	0	0	0	0	0	0
	Medium-Low	2	5	1	0	0	2	5	2	4	0
	Medium	0	10	18	5	0	0	10	17	8	3
	Medium-High	0	0	3	9	1	0	1	3	2	1
	High	0	0	0	0	3	0	0	0	0	0
Kappa*: 0.42; Kappa Variance*: 0.0082 Global Accuracy*: 60%						Kappa**: 0.11; Kappa Variance**: 0.0064 Global Accuracy**: 41%					
Analysis: Hypothesis Z-Test: Kappa* = 0 Kappa is significantly higher than zero (z=4.68; p-value=0.0000; α=0.05) Hypothesis Z-Test: Kappa** =0 Kappa** is not significantly higher than zero (z=1.41; p-value=0.0798; α=0.05) Hypothesis Z-Test: Kappa*- Kappa**=0 Kappa*- Kappa** is significantly higher than zero (z=2.57; p-value=0.0050; α=0.05)											

Table.10: Comparative analysis between confusion matrices: numerical quantitative theme-feature, from Unini River ExRes.

Categorized	RS over original values (r=0,43)*					MLP over logarithmic values (r=0,36)**					
	Reference					Reference					
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High	High
	Low	0	0	0	0	0	0	0	0	0	0
	Medium-Low	0	0	1	0	0	0	0	0	0	0
	Medium	0	16	17	18	0	0	15	18	7	0
	Medium-High	0	0	0	6	2	0	1	0	13	1
High	0	0	0	0	0	0	0	0	4	1	
Kappa*: 0.10; Kappa Variance*: 0.0029 Global Accuracy*: 38%					Kappa**: 0.33; Kappa Variance**: 0.0046 Global Accuracy**: 53%						
<p>Analysis:</p> <p>Hypothesis Z-Test: Kappa* = 0 Kappa is significantly higher than zero (z=1.89; p-value=0.0295; α=0.05)</p> <p>Hypothesis Z-Test: Kappa** = 0 Kappa** is significantly higher than zero (z=4.85; p-value=0.0000; α=0.05)</p> <p>Hypothesis Z-Test: Kappa* - Kappa** = 0 Kappa* - Kappa** is significantly lower than zero (z=-2.62; p-value=0.0045; α=0.05)</p>											

Table.11: Comparative analysis between confusion matrices: categorized qualitative theme-feature, from the joint region

Categorized	MLP over original values*					MLP over logarithmic values**					
	Reference					Reference					
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High	High
	Low	0	0	0	0	0	2	0	0	0	0
	Medium-Low	2	25	12	4	1	0	18	10	10	0
	Medium	0	7	23	13	2	0	13	24	6	1
	Medium-High	0	6	7	23	1	0	7	8	24	2

High	0	0	0	0	2	0	0	0	0	3
Kappa*: 0.38; Kappa Variance*: 0.0039 Global Accuracy*: 57%						Kappa**: 0.36; Kappa Variance**: 0.0042 Global Accuracy**: 55%				
Analysis: Hypothesis Z-Test: Kappa* = 0 Kappa is significantly higher than zero (z=6.00; p-value=0.0000; α=0.05) Hypothesis Z-Test: Kappa** = 0 Kappa** is significantly higher than zero (z=5.60; p-value=0.0000; α=0.05) Hypothesis Z-Test: Kappa* - Kappa** = 0 Kappa* - Kappa** is not significantly different than zero (z=0.19; p-value=0.4255; α=0.05)										

Table.12: Comparative analysis between confusion matrices: categorized qualitative theme-feature, from SGC.

Categorized	RF over original values*					MLP over logarithmic values**					
	Reference					Reference					
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High	High
	Low	2	0	0	0	0	1	0	1	0	0
	Medium-Low	0	10	7	4	0	0	8	8	2	0
	Medium	0	5	14	6	3	1	5	10	4	0
	Medium-High	0	1	1	4	1	0	3	3	7	1
	High	0	0	0	0	0	0	0	0	1	3
Kappa*: 0.30; Kappa Variance*: 0.0088 Global Accuracy*: 52%						Kappa**: 0.30; Kappa Variance**: 0.0091 Global Accuracy**: 50%					
Analysis: Hypothesis Z-Test: Kappa* = 0 Kappa is significantly higher than zero (z=3.16; p-value=0.0008; α=0.05) Hypothesis Z-Test: Kappa** = 0 Kappa** is significantly higher than zero (z=3.20; p-value=0.0007; α=0.05) Hypothesis Z-Test: Kappa* - Kappa** = 0 Kappa* - Kappa** is not significantly different than zero (z=-0.06; p-value=0.4762; α=0.05)											

Table.13: Comparative analysis between confusion matrices: categorized qualitative theme-feature, from Unini River ExRes.

Categorized	ODT over original values*					MLP over logarithmic values**					
	Reference					Reference					
	Category	Low	Medium-Low	Medium	Medium-High	High	Low	Medium-Low	Medium	Medium-High	High
	Low	0	0	0	0	0	0	0	0	0	0
	Medium-Low	0	12	5	4	0	0	14	8	4	0
	Medium	0	3	14	3	0	0	4	9	2	0
	Medium-High	0	7	1	17	0	0	4	3	20	0
High	0	0	0	2	2	0	0	0	0	2	
Kappa*: 0.48; Kappa Variance*: 0.0069 Global Accuracy*: 64%					Kappa**: 0.47; Kappa Variance**: 0.0071 Global Accuracy**: 64%						
<p>Analysis:</p> <p>Hypothesis Z-Test: $Kappa^* = 0$ Kappa is significantly higher than zero ($z=5.76$; $p\text{-value}=0.0000$; $\alpha=0.05$)</p> <p>Hypothesis Z-Test: $Kappa^{**} = 0$ Kappa** is significantly higher than zero ($z=5.61$; $p\text{-value}=0.0000$; $\alpha=0.05$)</p> <p>Hypothesis Z-Test: $Kappa^* - Kappa^{**} = 0$ Kappa* - Kappa** is not significantly different than zero ($z=0.08$; $p\text{-value}=0.4697$; $\alpha=0.05$)</p>											

From the analysis of the results presented in the tables, it is observed that the kappa values obtained by the post-modeling categorization process (Tables 8, 9 and 10), in general, were lower than those obtained in the pre-modeling categorization process (Tables 11, 12 and 13). In both cases, the ML techniques built specific models for quantitative or qualitative data, suffering loss of accuracy in the transformation process between these types of data.

Due to the loss of accuracy in the post-modeling categorization process, the best results obtained are shown in Table 13, with insignificant difference in the kappa values for the ODT (Kappa = 0.48) and MLP (Kappa = 0.47).

The values obtained by the Kappa coefficient, in addition to serving as parameters for comparison between the categorizations, can also be evaluated, being classified

in different linguistic intervals, according to their level of agreement, as shown in Figure 6. In this case, according to [55], the best results obtained in this research are classified as *moderate*.

The *moderate* results obtained may have occurred for several reasons, including: the quantity of biomass samples; the sampling distribution of biomass values; and the low correlation between the biomass theme-feature and extracted the extracted features. Regarding the latter, Table 3 shows the low correlation, including on the selected features.

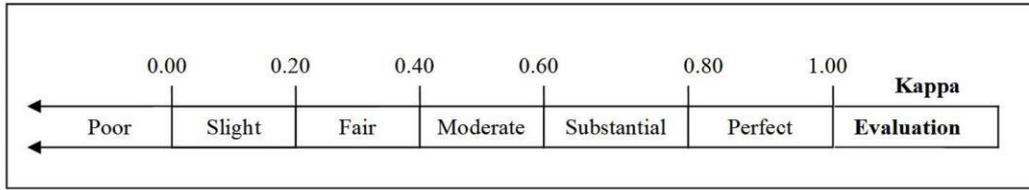


Fig.6: Linguistic evaluation of Kappa coefficient values. Adapted from [55].

IV. CONCLUSION

The present work aimed to develop and compare forest biomass estimation models, from different regions of the Amazon forest, built over numerical quantitative or categorical qualitative theme-feature. For this, ML techniques were applied on polarimetric and interferometric X, L and P bands SAR data extracted features, generating models that were analysed and compared.

In an innovative way, the work presents a methodology that involves:

- the process of feature selection and AGB estimation models development over quantitative and qualitative theme-features. It is noteworthy that, for each case, the feature selection and ML techniques were specific and configured in order to obtain the best results;
- comparative analyses between quantitative and qualitative results. In this case, the post-modeling categorization process and the respective confusion matrices construction was performed, followed by the comparison using hypothesis tests.

The results showed that the different study areas had very different characteristics, significantly impacting the feature selection and ML algorithms. The SGC area, due to the greater variation in AGB inventoried values (between 92.21 and 351.73 t/ha), obtained better results with the numeric quantitative theme-features. On the other hand, Unini’s River ExRes area, that had AGB values with less variation (between 153.32 and 311.57 t/ha), was better suited to categorized qualitative data modelling.

The different biomes of the Amazon Forest and their respective characteristics demanded specific models and techniques, not fitting into a single pattern. This conclusion is in agreement with the research of [2] who affirms that the heterogeneity of tropical forests is one of the main factors for the increasing uncertainty regarding the biomass stocks measurement in the region.

The process of feature selection was unanimous in selecting the interferometric height (H_{ini}) as the most

relevant feature for all areas of study, both in the case of qualitative and quantitative theme-features, in agreement with the results obtained by [23-24,56-57]. Likewise, there was an emphasis on features obtained by target decomposition techniques on the L band, from the ALOS PALSAR 2 sensor. The textural features, on the other hand, did not show significant correlation with the AGB values, different from the results obtained by [58].

As a conclusion of the presented methodology, there was no significant improvement in the AGB estimation process, since the results obtained from Kappa varied between *fair* and *moderate*. Likewise, the post-modeling categorization process did not achieve the expected results, keeping the Kappa value stable and not being able to generalize the AGB values into categories. The result obtained may have occurred due to the low correlation between the biomass theme-feature and the extracted SAR features.

In order to develop more suitable AGB models for different regions of the Amazon Forest, further studies will be carried out aiming to adjust the training parameters of ML techniques. In this case, the possibility of applying search methods and deep learning, commonly used in the Artificial Intelligence area to define such parameters, will be verified.

Analysing the possible reasons that led to the limited results, two factors were identified that may contribute to new research in the area in focus.

The first factor refers to the inventoried forest management plots used as samples. In agreement with the quoted by [59-65], a large number of plots, including areas with greater variations of AGB values, allows a more reliable sample representation and more in-depth statistical analysis.

The second factor is related to the processing of SAR data and the possibility of extracting new polarimetric and interferometric features. Accessing data in SLC format of polarimetric X and P bands would enable the extraction and analysis of the respective target decomposition features. Likewise, through the construction of a digital elevation model in the L band, it would be possible to obtain new interferometric heights involving the

differences between the X-L and L-P bands and the corresponding analyzes.

REFERENCES

- [1] Sinha, S., Jeganathan, C., Sharma, L.K., Nathawat, M.S (2015). A review of radar remote sensing for biomass estimation. *International Journal of Environment Sciences Technologies*. <https://doi.org/10.1007/s13762-015-0750-0>
- [2] Erb, K.H., Kastner, T., Plutzar, C., Bais, A.L.S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., Niedertscheider, M., Pongratz, J., Thurner, M., Luyssaert, S (2018). Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature*, 553, 73–76. <https://doi.org/10.1038/nature25138>
- [3] UNFCCC – United Nation Framework Convention on Climate Change (2008). *Kyoto Protocol Reference Manual on Accounting of Emissions and Assigned Amounts*.
- [4] IPCC – Intergovernmental Panel on Climate Change (2003). *Good practice guidance for land use, land-use changes and forestry*.
- [5] Köhl, M., Lasco, R., Cifuentes, M., Jonsson, Ö., Korhonen, K.T., Mundhenk, P., de Jesus Navar, J., Stinson, G (2015). Changes in forest production, biomass and carbon: Results from the 2015 UN FAO. *Global Forest Resource Assessment*. *Forest Ecology and Management*, 352, 21–34. <https://doi.org/10.1016/j.foreco.2015.05.036>
- [6] Ho Tong Minh, D., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., Phillips, O.L., Feldpausch, T.R., 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 of Environment*, 175, 138–147. <https://doi.org/10.1016/j.rse.2015.12.037>
- [7] Houghton, R.A., Nassikas, A.A (2017). Global and regional fluxes of carbon from land use and land cover change 1850–2015. *Global Biogeochem. Cycles*, 31, 456–472. <https://doi.org/10.1002/2016GB005546>
- [8] Kumar, L., Sinha, P., Taylor, S., Alqurashi, A.F (2015). Review of the use of remote sensing for biomass estimation to support renewable energy generation. *Journal of Applied Remote Sensing*, 9, 1–29. <https://doi.org/10.1117/1.jrs.9.097696>
- [9] Beaudoin, A., Le Toan, T., Goze, S., Nezry, E., Lopes, A., Mougin, E., Hsu, C.C., Han, H.C., Kong, J.A., Shin, R.T (1994). Retrieval of forest biomass from SAR data. *International Journal of Remote Sensing*, 15, 2777–2796. <https://doi.org/10.1080/01431169408954284>
- [10] Furtado, L.F. de A., Silva, T.S.F., Novo, E.M.L. de M. Dual-season and full-polarimetric C band SAR assessment for vegetation mapping in the Amazon várzea wetlands (2016). *Remote Sensing of Environment*, 174, 212–222. <https://doi.org/10.1016/j.rse.2015.12.013>
- [11] Ningthoujam, R.K., Balzter, H., Tansey, K., Feldpausch, T.R., Mitchard, E.T.A., Wani, A.A., Joshi, P.K (2017). Relationships of S-band radar backscatter and forest aboveground biomass in different forest types. *Remote Sensing*, 9, 1–17. <https://doi.org/10.3390/rs9111116>
- [12] Debastiani, A.B., Moura, M.M., Rex, F.E., Sanquetta, C.R., Corte, A.P.D., Pinto, N. Regressões Robusta e Linear para Estimativa de Biomassa Via Imagem Sentinel em uma Floresta Tropical (2019). *BIOFIX Science Journal*, 4, 81–87. <https://doi.org/10.5380/biofix.v4i2.62922>
- [13] Saatchi, S., Harris, N.L., Lefsky, M., Brown, S., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M. & Morel, A (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences, USA*, 2011, 108: 9899–9904.
- [14] Huang, W., Sun, G., Ni, W., Zhang, Z., Dubayah, R (2015). Sensitivity of multi-source SAR backscatter to changes in forest aboveground biomass, in: *International Geoscience And Remote Sensing Symposium (IGARSS)*, Melbourne. <https://doi.org/10.3390/rs70809587>
- [15] Treuhaff, R., Lei, Y., Gonçalves, F., Keller, M., dos Santos, J.R., Neumann, M., Almeida, A. Tropical-forest structure and biomass dynamics from TanDEM-X radar interferometry. *Forests*, 2017, 8, 277–294. <https://doi.org/10.3390/f8080277>
- [16] Le Noë, J., Matej, S., Magerl, A., Bhan, M., Erb, K.H., Gingrich, S (2020). Modeling and empirical validation of long-term carbon sequestration in forests (France, 1850–2015). *Global Change Biology*, 26, 2421–2434. <https://doi.org/10.1111/gcb.15004>
- [17] Avtar, R., Suzuki, R., Sawada, H (2014). Natural Forest Biomass Estimation Based on Plantation Information Using PALSAR Data. *PLOS ONE*, 9 (1). <https://doi.org/10.1371/journal.pone.0086121>
- [18] Berninger, A., Lohberger, S., Stängel, 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, 831–853. <https://doi.org/10.3390/rs10060831>
- [19] Pereira, L.O., Furtado, L.F.A., Novo, E.M.L.M., Sant’Anna, S.J.S., Liesenberg, V., Silva, T.S.F (2018). Multifrequency and Full-Polarimetric SAR assessment for estimating above ground biomass and leaf area index in the Amazon Várzea Wetlands. *Remote Sensing*, 10, 1–23. <https://doi.org/10.3390/rs10091355>
- [20] Camargo, F.F., Sano, E.E., Almeida, C.M., Mura, J.C., Almeida, T (2019). A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sensing*, 11, 1600–1616. <https://doi.org/10.3390/rs11131600>
- [21] DSG – Diretoria de Serviço Geográfico (2008). *Contratação de Serviços de Aerolevantamento na Região Amazônica e Processamento de Dados com Radars de Abertura Sintética Aerotransportados Interferométricos*. Mapping Project.
- [22] Santos, J.R., Freitas, C.C., Araujo, L.S., Dutra, L. V., Mura, J.C., Gama, F.F., Soler, L.S., Sant’Anna, S.J.S (2003). Airborne P-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest. *Remote Sensing of Environment*, 87, 482–493. <https://doi.org/10.1016/j.rse.2002.12.001>

- [23] Neeff, T., Dutra, L.V., Dos Santos, J.R., Da Costa Freitas, C., Araujo, L.S (2005). Tropical forest measurement by interferometric height modeling and P-band radar backscatter. *Forest Science*, 51, 585–594. <https://doi.org/10.1093/forestscience/51.6.585>
- [24] Gama, F.F., Mura, J.C., De Albuquerque, P.C.G., Dos Santos, J.R (2010). Avaliação do potencial da interferometria sar para o mapeamento altimétrico de áreas reflorestadas por eucalyptus sp. *Boletim de Ciências Geodésicas*. <https://doi.org/10.1590/s1982-21702010000400003>
- [25] Del Frate, F., Solimini, D (2004). On neural network algorithms for retrieving forest biomass from SAR data. *IEEE Transation Geoscience on Remote Sensing*, 42, 24–34. <https://doi.org/10.1109/TGRS.2003.817220>
- [26] Enghart, S., Keuck, V., Siegert, F (2012). Modeling aboveground biomass in tropical forests using multi-frequency SAR data – a comparison of methods. *IEEE Journal Selected Topics on Applied Earth Observation Remote Sensing*, 5, 298–306. <https://doi.org/10.1109/JSTARS.2011.2176720>
- [27] Wylie, B.K., Pastick, N.J., Picotte, J.J., Deering, C.A (2019). Geospatial data mining for digital raster mapping. *GIScience Remote Sensing*. <https://doi.org/10.1080/15481603.2018.1517445>
- [28] Quinlan, J.R., (1993). *C4.5: Programs for Machine Learning*, Machine Learning Kluwer Academic Publishers, Boston, Manufactured in The Netherlands. Morgan Kaufmann, California.
- [29] Ng, A., (2018). *Machine Learning Yearning: Technical Strategy for AI Engineers in the Era of Deep Learning [Draft Version]*, deeplearning.ai.
- [30] Brink, H.B., Richard, J.W., Fetherolf, M. (2015). *Real-World Machine Learning*, MEAP Editi. ed, Book. Manning Publication, New York.
- [31] Cavalcante, J. R. & Abreu, A. J. L (2020). COVID-19 in the city of Rio de Janeiro: spatial analysis of first confirmed cases and deaths. *Epidemiologia Serviço de Saúde*, Brasília, 29(3):e2020204, 2020. doi: 10.5123/S1679-49742020000300007.
- [32] Pardo, I. F., Napoletano, B. M., Verges, F. R., Billa, L (2020). Spatial analysis and GIS in the study of COVID-19: a review. *Science of The Total Environment*, 739. <https://doi.org/10.1016/j.scitotenv.2020.140033>.
- [33] Fatima, M., O’Keefe, K. J., Wei, W., Arshad, S., Gruebner, O (2021). Geospatial analysis of COVID-19: a scoping review. *International Journal of Environment Res Public Health*, 18 (5):2336. DOI: <https://doi.org/10.3390/ijerph18052336>.
- [34] Mooney, Peter & Juhász, Levente (2020). Mapping COVID-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*, 10, 265-270. [10.1177/2043820620934926](https://doi.org/10.1177/2043820620934926).
- [35] Li, R (2021). Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps. *Human Behavior & Emerging Technology*, 3, 97– 106. <https://doi.org/10.1002/hbe2.248>.
- [36] Mapbiomas, 2019. Mapeamento Anual da Cobertura e Uso do Solo no Brasil (MapBiomas). Mapeamento Anu. da Cober. e Uso do Solo no Bras. URL <http://mapbiomas.org/map#coverage> (accessed 6.14.19).
- [37] RadamBrasil, 1977. Geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Mapping Project.
- [38] Higuchi, N., Santos, J. dos, Ribeiro, R.J., Minette, L., Biot, Y (1998). Biomassa da parte aérea da vegetação da Floresta Tropical úmida de terra-firme da Amazônia Brasileira. *Acta Amazon*, 28, 153–166. <https://doi.org/10.1590/1809-43921998282166>
- [39] Silva, R.P. (2007). *Alometria, estoque e dinâmica da biomassa de florestas primárias e secundárias na região de Manaus (AM)*. National Institute for Space Research (INPE). PhD Thesis.
- [40] Araújo, T.M., Higuchi, N., Junio, J.A. de C (1999). Comparison of formulae for biomass content determination in a tropical rain forest site in the state of Para, Brazil. *Forest Ecology and Management*, 117, 43–52. [https://doi.org/10.1016/S0378-1127\(98\)00470-8](https://doi.org/10.1016/S0378-1127(98)00470-8)
- [41] Lima, A.J.N., Suwa, R., De Mello Ribeiro, G.H.P., Kajimoto, T., Dos Santos, J., Da Silva, R.P., De Souza, C.A.S., De Barros, P.C., Noguchi, H., Ishizuka, M., Higuchi, N (2012). Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. *Forest Ecology and Management*, 277, 163–172. <https://doi.org/10.1016/j.foreco.2012.04.028>
- [42] Woodhouse, I.H., 2017. *Introduction to Microwave Remote Sensing*, Introduction to Microwave Remote Sensing. Taylor & Francis Group CRC Press, Florida. <https://doi.org/10.1201/9781315272573>
- [43] Henderson, F.M., Lewis, A.J., 1998. *Manual of remote sensing: principles and applications of imaging radars*, 3rd ed. ed. John Wiley and Sons, New York.
- [44] Pope, K.O., Rey-Benayas, J.M., Paris, J.F (1994). Radar remote sensing of forest and wetland ecosystems in the Central American tropics. *Remote Sensing of Environment*, 48, 205–219. [https://doi.org/10.1016/0034-4257\(94\)90142-2](https://doi.org/10.1016/0034-4257(94)90142-2)
- [45] Kim, Y., Van Zyl, J.J (2009). A time-series approach to estimate soil moisture using polarimetric radar data. *IEEE Transaction Geoscience Remote Sensing*, 47, 2519–2527. <https://doi.org/10.1109/TGRS.2009.2014944>
- [46] Haralick, R., Shanmugam, K., Dinstein, I (1973). Textural Features for Image Classification. *IEEE Transaction System Man Cybernetics*, 3, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [47] Cloude, S.R., Pottier, E (1996). A review of target decomposition theorems in radar polarimetry. *IEEE Transactions Geoscience and Remote Sensing*, 34. <https://doi.org/10.1109/36.485127>
- [48] Freeman, A., Durden, S.L (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transaction Geoscience Remote Sensing*, 36, 963–973. <https://doi.org/10.1109/36.673687>
- [49] Touzi, R (2007). Target scattering decomposition in terms of roll-invariant target parameters. *IEEE Transaction*

- Geoscience Remote Sensing, 45, 73–84. <https://doi.org/10.1109/TGRS.2006.886176>
- [50] Van Zyl, J.J., (1992). Application of Cloude's target decomposition theorem to polarimetric imaging radar data, in: Proceedings Society of Photo-Optical Instrumentation Engineers. pp. 184–212. <https://doi.org/10.1117/12.140615>
- [51] Yamaguchi, Y., Yajima, Y., Yamada, H (2006). A four-component decomposition of POLSAR images based on the coherency matrix. IEEE Geoscience Remote Sensing Letters, 3, 292–296. <https://doi.org/10.1109/LGRS.2006.869986>
- [52] Dent, B., Torguson, J., Hodler, T. (2008). Cartography: thematic map design, 6th ed., Cartographic Perspectives. McGraw-Hill Science, New York.
- [53] Hall, M. a., Smith, L. a. (1998). Practical feature subset selection for machine learning, Computer Science. Hamilton, New Zealand.
- [54] Yu, X.; Ge, H.; Lu, D.; Zhang, M.; Lai, Z.; Yao, R (2019). Comparative Study on Variable Selection Approaches in Establishment of Remote Sensing Model for Forest Biomass Estimation. Remote Sensing, 11, 1437. <https://doi.org/10.3390/rs11121437>
- [55] Witten, I.H., Frank, E., Hall, M.A., Pal, C.J. (2016). Data Mining: Practical Machine Learning Tools and Techniques, 2nd Editio. ed, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Massachusetts. <https://doi.org/10.1016/c2009-0-19715-5>
- [56] Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. (2004). Applied Linear Statistical Models, 5th Edition. ed, Journal of Education. McGraw-Hill, Boston.
- [57] Congalton, R.G., Green, K. (2013). Assessing the Accuracy of Remotely Sensed Data Principles and Practices (Second Edition), CRC Press Taylor & Francis Group, Boca Raton, London, New York.
- [58] Sileshi, G. W (2014). A critical review of forest biomass estimation models, common mistakes and corrective measures. Forest Ecology and Management, 329, 237-254. <https://doi.org/10.1016/j.foreco.2014.06.026>.
- [59] Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K (1999). Improvements to the SMO algorithm for SVM regression. IEEE Transaction Neural Networks, 11, 1188–1193. <https://doi.org/10.1109/72.870050>
- [60] Landis, J.R., Koch, G.G (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977, 33, 159–174. <https://doi.org/10.2307/2529310>
- [61] Castro-Filho, C.A.P. de, Freitas, C.D.C., Sant'Anna, S.J.S., Lima, A.J.N., Higuchi, N. (2013). Relating Amazon forest biomass to PolInSAR extracted features, in: International Geoscience and Remote Sensing Symposium (IGARSS). Melbourne. <https://doi.org/10.1109/IGARSS.2013.6721320>
- [62] Schlund, M., Erasmi, S., Scipal, K (2020). Comparison of Aboveground Biomass Estimation from InSAR and LiDAR Canopy Height Models in Tropical Forests. IEEE Geoscience Remote Sensing Letters, 17, 367–371. <https://doi.org/10.1109/LGRS.2019.2925901>
- [63] Sarker, M. L. R., Nichol, J., Iz, H. B., Ahmad, B. B., Rahman (2012), A. A. Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. ISPRS Journal of Photogrammetry and Remote Sensing, 69, 146-166. <https://doi.org/10.1016/j.isprsjprs.2012.03.002>.
- [64] Clark, D.B., Kellner, J.R (2012). Tropical forest biomass estimation and the fallacy of misplaced concreteness. Journal Vegetation Science, 23, 1191-1196. <https://doi.org/10.1111/j.1654-1103.2012.01471.x>
- [65] Santoro, M., Cartus, O (2018). Research pathways of forest above-ground biomass estimation based on SAR backscatter and interferometric SAR observations. Remote Sensing, 10, 1–23. <https://doi.org/10.3390/rs10040608>