Biomass and vegetation index by remote sensing in different caatinga forest areas

Leudiane Rodrigues Luz[1] **Vanderlise Giongo**[2] **Antonio Marcos dos Santos**[1]
Rodrigo José de Carvalho Lopes[1] **Claudemiro de Lima Júnior**[1*]

[1]Programa de Pós-graduação em Ciência e Tecnologia Ambiental, Universidade de Pernambuco (UPE), 56328-900, Petrolina, PE, Brasil.
E-mail: claudemiro.lima@upe.br. *Corresponding author.
[2]Centro de Pesquisa Agropecuária do Trópico Semiárido, Embrapa Semiárido, Petrolina, PE, Brasil.

ABSTRACT: *Continued unsustainable exploitation of natural resources promotes environmental degradation and threatens the preservation of dry forests around the world. This situation exposes the fragility and the necessity to study landscape transformations. In addition, it is necessary to consider the biomass quantity and to establish strategies to monitor natural and anthropic disturbances. Thus, this research analyzed the relationship between vegetation index and the estimated biomass using allometric equations in different Brazilian caatinga forest areas from satellite images. This procedure is performed by estimating the biomass from 9 dry tropical forest fragments using allometric equations. Area delimitations were obtained from the Embrapa collection of dendrometric data collected in the period between 2011 and 2012. Spectral variables were obtained from the orthorectified images of the RapidEye satellite. The aboveground biomass ranged from 6.88 to 123.82 Mg.ha⁻¹. SAVI values were L = 1 and L = 0.5, while NDVI and EVI ranged from 0.1835 to 0.4294, 0.2197 to 0.5019, 0.3622 to 0.7584, and 0.0987 to 0.3169, respectively. Relationships among the estimated biomass and the vegetation indexes were moderate, with correlation coefficients (Rs) varying between 0.64 and 0.58. The best adjusted equation was the SAVI equation, for which the coefficient of determination was R² = 0.50, R²aj = 0.49, RMSE = 17.18 Mg.ha⁻¹ and mean absolute error of prediction (MAE) = 14.07 Mg.ha⁻¹, confirming the importance of the Savi index in estimating the caatinga aboveground biomass.*
Key words: *dry forests, renewable energy, modeling.*

Relação entre biomassa e índices de vegetação por sensoriamento remoto em diferentes áreas de florestas de caatinga

RESUMO: *A exploração contínua e insustentável dos recursos naturais promove a degradação ambiental e ameaça a conservação das florestas secas no mundo. Essa situação expõe a fragilidade e a necessidade de estudar as transformações da paisagem. É necessário considerar o quantitativo de biomassa e o estabelecimento de estratégias para monitorar distúrbios naturais e antrópicos. Assim, este trabalho teve como objetivo analisar a relação entre o índice de vegetação e a biomassa estimada por meio de equações alométricas em diferentes áreas da caatinga brasileira a partir de imagens de satélite. Este procedimento é realizado estimando-se a biomassa de nove fragmentos de floresta tropical seca com o uso em equações alométricas. As delimitações das áreas foram obtidas da coleção da Empresa Brasileira de Pesquisa Agropecuária, a partir dos dados dendrométricos coletados no período entre 2011 e 2012. As variáveis espectrais foram obtidas a partir das imagens ortorretificadas do satélite RapidEye. A biomassa acima do solo variou de 6,88 a 123,82 Mg.ha⁻¹. Os valores de SAVI para os coeficientes L = 1 e L = 0,5, NDVI e EVI variaram de 0,1835 a 0,4294, 0,2197 a 0,5019, 0,3622 a 0,7584 e 0,0987 a 0,3169, respectivamente. As relações entre a biomassa estimada e os índices de vegetação foram moderadas, com coeficiente de correlação (rs) variando entre 0,64 e 0,58. A equação mais bem ajustada foi a Equação do SAVI, com coeficiente de determinação foi R² = 0,50, R²aj = 0.49, RMSE = 17.18 Mg.ha⁻¹ e erro médio absoluto de predição (MAE) = 14.07 Mg.ha⁻¹, confirmando a importância do índice SAVI na estimativa da biomassa aérea da Caatinga.*
Palavras-chave: *florestas secas, energia renovável, modelagem.*

INTRODUCTION

The Brazilian caatinga tropical dry forest is located between Ecuador and the Tropic of Capricorn. It receives abundant light throughout the year with high temperatures ranging from 23 °C to 27 °C and long periods of drought and water scarcity. These factors characterize the floristic diversity of the region predominantly composed of xerophilous, deciduous, usually thorny species, varying according to soil type and rainy season (SAMPAIO et al., 2015).

The unsustainable continuous exploitation of natural resources promotes environmental degradation and threatens caatinga preservation, as 46% from the original area has already been degraded or deforested. Actions such as deforestation, burning, selective logging, mining, agriculture and livestock are some of the main factors which most contribute

to the land use change (SILVA & SAMPAIO, 2008). The partial or total removal of vegetation, results in a reduction of forest biomass and an increase in soil degradation, which constitute factors that lead to the risk of the Biome desertification (COSTA et al., 2009). This situation exposes the fragility of this forest and the necessity to further landscape studies, mainly regarding the amount of aboveground biomass that is removed. These studies can help establish strategies to reduce CO_2 emissions from deforestation and forest degradation, as well as ensure sustainable management of forests and enhance forest carbon stocks, as an essential effort to mitigate climate change (FAO, 2020).

An estimation of the aboveground biomass can be performed by direct and indirect methods. The direct method requires time, high costs and tree cutting to obtain the data. The indirect or non-destructive method is performed through the biomass relationship with dendrometric variables of the trees. This method has reduced time and financial cost, and does not require tree cutting (FERRAZ et al., 2014).

Estimates of Brazilian tropical dry forrest aboveground biomass have shown large spatial and seasonal variation from 2 to 160 Mg.ha^{-1} (SAMPAIO & FREITAS, 2008). This difference is mainly due to the low and irregular rainfall distribution and soil characteristics.

Aboveground biomass estimates can be made using other methods such as using satellite imagery combined to allometric equations which use field-collected data. The variable obtained from applying an allometric equation can be related with reflectance values obtained in satellite images, thereby enabling development of mathematical models which make possible to estimate the aboveground biomass. Such images provide information from areas which are difficult to access, and also require less time and lower operating cost when compared to direct methods (FERRAZ et al., 2014).

Some studies have been carried out with remote sensing to estimate caatinga forest characteristics. TIAN et al. (2017) quantified the temporal trends of the non-photosynthetic woody components (i.e. stems and branches) in global tropical drylands for a period of 12 years using vegetation optical depth (VOD) retrieved from passive microwave observations. ALMEIDA et al. (2014) adjusted models to estimate dendrometric characteristics of the Brazilian dry tropical forest (caatinga) from Landsat 5 TM sensor data, concluding that the metrics derived from the Landsat 5 TM sensor have great potential to explain variation in the mean height of trees and in the wood volume per hectare in tropical dry forest remnant areas in Northeast Brazil.

FORMIGONI et al. (2011) evaluated the temporal behavior of the Enhanced Vegetation Index (EVI) of the Moderate Resolution Imaging Spectroradiometer (MODIS) of different types of vegetation in Northeast Brazil for a period of six years. The results showed that the caatinga vegetation presented the greatest variation in EVI values.

Thus, this research analyzed the relationship between vegetation index and the estimated biomass using allometric equations in different caatinga forest areas from satellite images, helping to improve and redefine non-destructive sustainable forest management strategies.

MATERIALS AND METHODS

Site descriptions

The study carried out in nine preserved native forest fragments, classified as steppe forested savannah. These fragments are located in the municipalities of Sobral-CE, Ribeira do Pombal-BA, Petrolina-PE, Campina Grande-PB, Janauba-MG, Araripina-PE and Nossa Senhora da Glória-SE, Brazil. In addition to these steppe forested savanna fragments, reforested savanna steppe and seasonal deciduous forest in the municipality of Mossoró-RN were also evaluated. All of these cities are located within the limitations of the Brazilian Semi-arid region (Figure 1). Details of location, rainy season, annual average temperature, annual average rainfall, and climate by the Köppen classification (Climate-Data.ORG, 2018) of each site are shown in table 1.

Data analysis
Aboveground biomass inventory

The dataset was obtained from Embrapa based on dendrometric data collected from nine areas from 2011 to 2012. Each area was georeferenced in ten plots of 10 x 20 m, arranged in transects. The areas were selected according to phytophysiognomy and soil type. A total of 1640 living individuals were reportede in these plots, corresponding to 72 tree species, distributed in 19 families. All standing and living plants found in the plots with diameters larger than 3 cm were evaluated, measuring the diameter at breast height (DBH) (1.3 m above the soil surface) and plant height. The allometric equation (Equation 1) developed by SAMPAIO and SILVA (2005) to estimate abovegraund biomass of each tree with DBH ranging from 3 to 30 cm was then applied after obtaining the data from all plants in each plot.

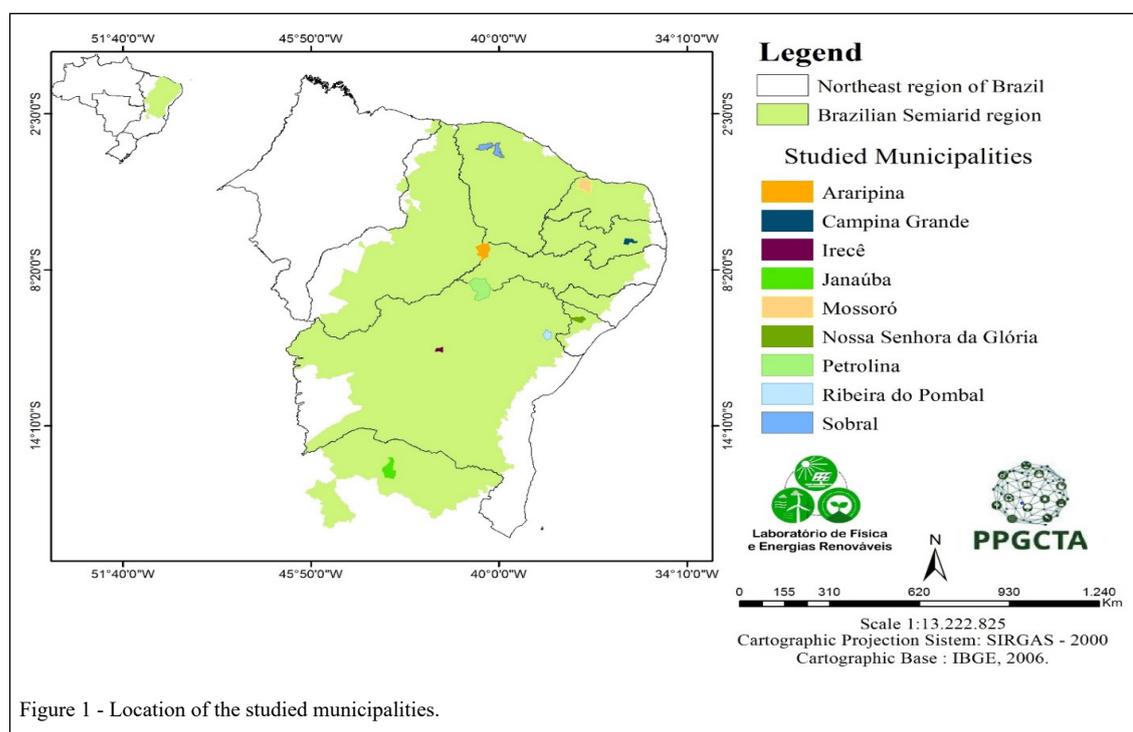$$\widehat{Biomass}(kg) = 0.1730 * DBH^{2.2950} \text{ -}$$ (Eq. 1)

Figure 1 - Location of the studied municipalities.

Plants with DBH larger than 30 cm had their biomass estimated by equation 2:

$$\widehat{Biomass}(kg) = 0.1648 * (SABH * H * d)^{0.9023} \quad \text{(Eq. 2)}$$

In which: SABH = sectional area at breast height; H = total plant height; d = wood density.

Image acquisition and processing

Next, we used the images already in orthorectified format of the RapidEye satellite set made available by the Ministry of Environment (MMA) to the University of Pernambuco (UPE) and Embrapa Semi-arid in order to obtain the vegetation index of

Table 1 - Location, rainy season, annual average temperature, annual average rainfall and climate by the Köppen and Geiger climate classification for each studied area.

Municipality	Location	Rainy season[*]	Climate	Annual average temperature (°C)[*]	Annual average rainfall (mm/year)[*]
Sobral (CE)	3°46'11.4" and 40°19'44.1"	January to May	Aw[**]	27.7	808
Ribeira do Pombal (BA)	10°47'21.4' and 38°32'24.4"	Throughout the year	Aw[**]	24.2	711
Petrolina (PE)	9°11'24.0' and 40°21'00.0"	October to May	BSh[***]	24.8	435
Mossoró (RN)	5°03'36.6" and 37°24'14.7"	December to July	Aw[**]	28	750
Irecê (BA)	11°18'53.7" and 41°49'49.3"	October to April	BSh[***]	23.3	583
Campina Grande (PB)	7°16'42.2" and 35°58'04.9"	December to September	Aw[**]	22.9	765
Janaúba (MG)	15°48'56.7" and 43°17'25.3"	September to April	Aw[**]	23.7	830
Araripina (PE)	7°27'55.7" and 40°24'52.4"	November to May	BSh[***]	23.7	719
Nossa Senhora da Glória (SE)	10°13'02.2" and 37°19'31.5"	Throughout the year	Aw[**]	23.4	763

([*]) data collected from 1982 to 2012. ([**]) Aw: tropical climate with dry winter season; ([***]) BSh: climate of warm low-latitude and high-altitude steppes.

each area. The selection of the RapidEye satellite was motivated by having improved radiometric (12 bit) and temporal (1 day) resolutions, and a spatial resolution of 5 m for accurately corrected images compatible with the 1: 25,000 scale. In addition, the sensors capture images in five spectral ranges: blue (440–510 μm), green (520–590 μm), red (630–685 μm), *red-edge* (690–730 μm) and near-infrared (760–850 μm) (PLANET, 2018).

Scenes used from the RapidEye satellite do not contain systematic radiometric errors such as visible lines or bands, and have little or no cloudiness that would compromise mapping areas of interest. Capturing the scenes coincided with the rainy season of each municipality: Sobral (CE), Acquisition date (Ad) 31/07/2012, Scene (S) 2437308; Ribeira do Pombal (BA), Ad 04/08/2011, S 2434117; Petrolina (PE), Ad 10/09/2011, S 2434908; Mossoró (RN), Ad 27/04/2012, S2436722; Irecê (BA), Ad 20/10/2012, S 2433801; Campina Grande (PB), Ad 17/03/2012, S 2435728; Janaúba (MG), Ad 10/03/2013, S 2331822; Araripina (PE), Ad 18/06/2012, S 2435608; Nossa Senhora da Glória (SE), Ad 28/06/2011, S 2434322. The images were pre-processed for geometric and atmospheric correction to be used in the physical environment mapping. RapidEye constellation images 3A were available with geometric correction so that digital numbers (ND) were converted to reflectance at the top of the atmosphere.

Preprocessing

Next, the digital numbers (ND) of the RapidEye images were converted to physical radiance values and later to reflectance according to the sensor manual hosted on the website https://www.planet.com/products/planet-imagery/ from the company Planets Labs to perform mathematical operations between the spectral bands, aiming to generate vegetation index. The images were processed in ERDAS Imagine 2013 software and the thematic map in ArcGIS 10.2 software. The DN value was then multiplied by a radiometric scale factor provided in the image metadata file to convert digital numbers (DN) to radiance, according to equation 3:

$$RAD(i) = DN(i) * RadiometricScaleFactor(i) \qquad \text{(Eq. 3)}$$

In which: RAD = sensor radiance (W.m^{-2}.sr.μm); DN = original digital value; and RadiometricScaleFactor = radiometric scale factor = 0.01.

Next, the conversion to the reflectance at the top of the atmosphere was performed from the radiance values, considering the distance between the sun and the earth and the geometry of the incident solar radiation, as presented in equation 4:

$$REF(i) = RAD(i) * \frac{\pi * SunDist^2}{EAI(i) * \cos(SolarZenith)} \qquad \text{(Eq. 4)}$$

In which: i = Number of spectral bands; REF = reflectance value; RAD = radiance value; SunDist = Distance between Earth and Sun on the day of astronomical units (UA) image acquisition (this value is not fixed and depends on the day, varying between 0.9832898912 UA and 1.0167103335 AU); EAI = Exo-atmospheric irradiance; SolarZenith = Solar zenith angle (= 90 ° - solar elevation).

Vegetation indexes (VI)

The following vegetation indexes (VI) were estimated aiming to observe the correlation with the estimated aboveground biomass: Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI) and Vegetation Enhancement Index (EVI). The Normalized Difference Vegetation Index (NDVI) is the normalization of the simple ratio and ranges from 1 to -1, and the closer to 1 the greater its photosynthetic activity, i.e. the greater presence of vegetation in the region. Low values approaching zero show that the area has little or no vegetation cover. The negative values represent the water (PONZONI et al., 2012). Normalization is performed by equation 5:

$$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \qquad \text{(Eq. 5)}$$

In which: ρ_{NIR} = near-infrared reflectance value; ρ_R = reflectance value in the red range.

Thus, SAVI (HUETE; 1988) and EVI (HUETE et al; 1997) were proposed due to the limitations observed in NDVI and in order to improve its values. SAVI contains a constant "L", which has the function of minimizing the soil effect on the vegetation signal, especially in less dense areas (SANTOS et al., 2014). The L factor varies according to the vegetation density and the reflectance characteristic of the soil. The value of L = 1 is suggested for use in low density vegetation areas, L = 0.5 for intermediate vegetation and L = 0.25 for large density vegetation areas. After considerations made by HUETE (1988), the L constant was estimated and included in the experimental measurements using the reflectance values in the red and near infrared bands (PONZONI et al., 2012). The index sensitivity in relation to the soil is larger in sparse tree canopies (PONZONI et al., 2012); which consequently, influences the calculation

of the Vegetation Index (SANTOS et al., 2014). The adjustment constants L = 1 and L = 0.5 were analyzed in this research due to the characteristics of the caatinga forest. When L= 0, the SAVI index will be equal to the NDVI index.

The SAVI equation is given as follows:

$$SAVI = \left[\frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R + L)}\right] * (1 + L) \qquad \text{(Eq. 6)}$$

In which: ρ_{NIR} = near-infrared reflectance value; ρ_R = reflectance value in the red range; L = soil adjustment variable.

EVI is less sensitive to substrate and atmosphere contamination compared to NDVI. It was developed to better respond to the vegetation signal, increasing the detection sensitivity in regions with larger biomass densities. This index presents correction factors for soils, atmosphere, and dense aboveground biomass (PONZONI et al., 2012).

The EVI equation is given as follows:

$$EVI = \frac{G(NIR - red)}{L + NIR + C1 * red + C2 * blue} \qquad \text{(Eq. 7)}$$

In which: NIR = reflectance in the near infrared region; red = reflectance in the red region; blue = reflectance in the blue region; C1 and C2 = adjustment coefficients for the effects of aerosols on the atmosphere; and L = the soil adjustment factor; G = the adjusted gain factor. The coefficient values adopted by the EVI algorithm were: L = 1, C1 = 6, C2 = 7.5 and G = 2.5.

The spectral samples were collected after the preprocessing step and estimating the vegetation indexes. The central position of each plot was used to define each sample unit by adding the eight neighboring pixels to it and forming a 3x3 matrix. The average reflectance values of the 9 pixels were calculated, which are used to establish the correlation between the vegetation index and biomass quantity. This process was performed for all indexes of the 90 plots.

Sobral-CE, Petrolina-PE and Irecê-BA municipalities did not present good spectral response and they were disregarded. Therefore, the statistical distribution of biomass was evaluated in 60 plots. However, 58 plots were used in the regression analysis due to outliers.

Statistical analysis

The Shapiro-Wilk test (SHAPIRO & WILK, 1965) was performed to test the normality of data distribution. As the estimated aboveground biomass presented a non-normal distribution, the Spearman's correlation analysis was used. We calculated the coefficient (Rs) and the significance

level (p) between the variables, relating the estimated biomass with each vegetation index. The simple linear regression model was used to analyze which variables best estimated aboveground biomass. The determination coefficient (R^2) in all regression analyzes is the adjusted coefficient of determination (R^2aj), the square root of the mean square error RMSE) and the standard deviation (Sd), which were calculated to identify which model best fits the dependent variable.
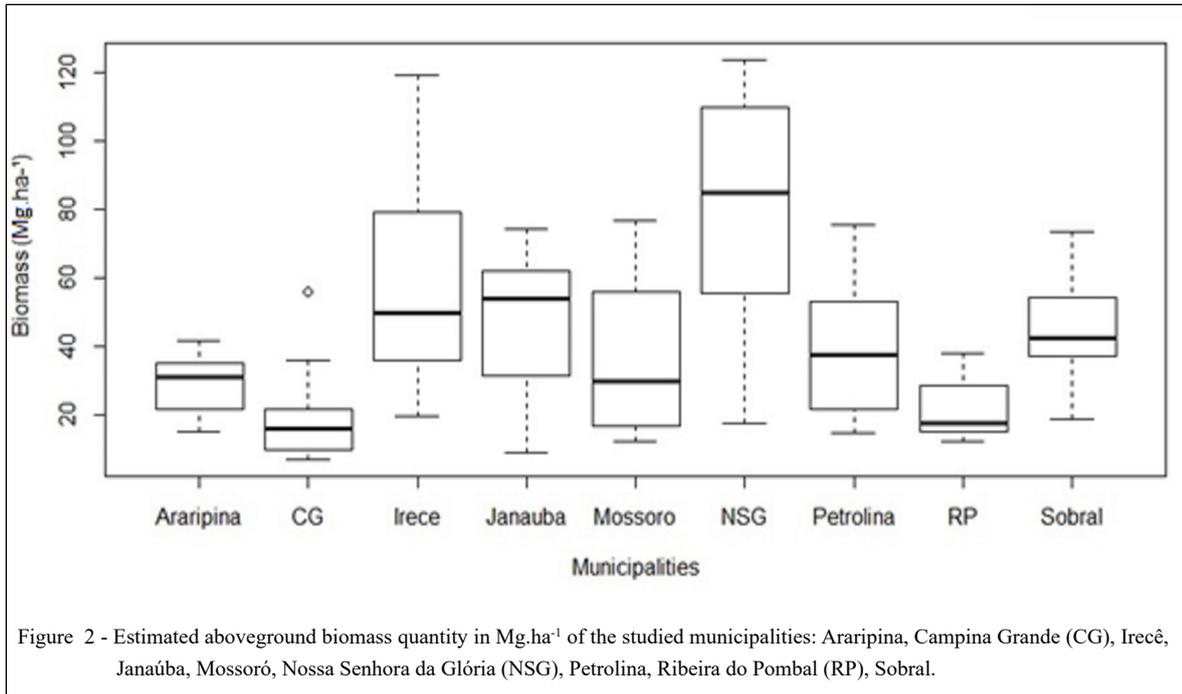
The prediction error calculation, which can be used to predict the average error of the estimates in future researches with the use of these equations, was done by the leave-one-out cross-validation (LOOCV) method. This method consisted of creating a model with (N-1) data at each interaction, considering the total number of data as N (called the training set), and validating the model with the data that were left out (which are called test data), so that only one data set is used as test data and all other data is used as training data for each interaction, thus calculating the mean absolute error (MAE) prediction (ARLOT & CELISSE, 2010; BERGMEIR & BENÍTEZ, 2012). All statistical analyzes were performed using the programming language R (R CORE TEAM, 2017).

RESULTS AND DISCUSSION

The estimated aboveground biomass by the allometric equations for all nine areas ranged from 6.88 to 123.82 Mg.ha^{-1}, whereupon 50% of the analyzed plots are between 16.77 and 55.92 Mg.ha^{-1}. These data corroborated the study by SAMPAIO & FREITAS (2008), who noticed a large variation in biomass stock in different caatinga areas (2 and 160 Mg.ha^{-1}), mainly due to the rainfall distribution at each location. According to the authors, aboveground biomass varies from 30 to 50 Mg.ha^{-1} in most areas, despite the variability. Figure 2 shows the variation of the estimated aboveground biomass in ten plots for each municipality based on the equations of SAMPAIO & SILVA (2005).

The municipalities of Nossa Senhora da Glória-SE and Irecê-BA showed the highest aboveground biomass variability, with minimum and maximum values between 17.41 to 123.82 Mg.ha^{-1} and 19.40 to 119.01 Mg.ha^{-1}, respectively. However, they were lower than those estimated by CASA and SEBAL models, in which the caatinga accumulated above-ground biomass ranging from 170 to 334 Mg.ha^{-1} (BRANDÃO et al., 2007; BRAND, 2017).

The municipalities with the lowest dispersion of estimated biomass data were Ribeira

Figure 2 - Estimated aboveground biomass quantity in Mg.ha⁻¹ of the studied municipalities: Araripina, Campina Grande (CG), Irecê, Janaúba, Mossoró, Nossa Senhora da Glória (NSG), Petrolina, Ribeira do Pombal (RP), Sobral.

do Pombal-BA with 12.38 to 37.81 Mg.ha⁻¹, Araripina-PE with 15.19 to 41.63 Mg.ha⁻¹ and Campina Grande-PB with 6.88 to 36.02 Mg.ha⁻¹. Campina Grande-PB presented an outlier with 55.99 Mg.ha⁻¹, which may have arisen because the sample tended to include high and low values in relation to the other distribution data. These values were higher than those reported by VIRGENS et al. (2017), 29.2 Mg.ha⁻¹, SANTOS et al. (2016), 12 Mg.ha⁻¹, and AMORIM et al. (2005) (20 and 25 Mg.ha⁻¹). The low biomass was justified due to some data which came from managed native vegetation, and due to the caatinga being formed by small and spaced trees.

The estimated biomass in the municipalities of Petrolina-PE, Janaúba-MG, Mossoró-RN and Sobral-CE presented very close variations from 14.79 to 75.28 Mg.ha⁻¹, 8.86 to 74.27 Mg.ha⁻¹, 12.38 to 76.92 Mg.ha⁻¹, and 18.75 to 73.55 Mg.ha⁻¹, respectively. The variabilities were higher, but close to those estimated by LIMA JÚNIOR et al. (2014), who obtained an aboveground biomass estimate between 5.93 and 60.74 Mg.ha⁻¹ for the Municipality of Petrolina-PE.
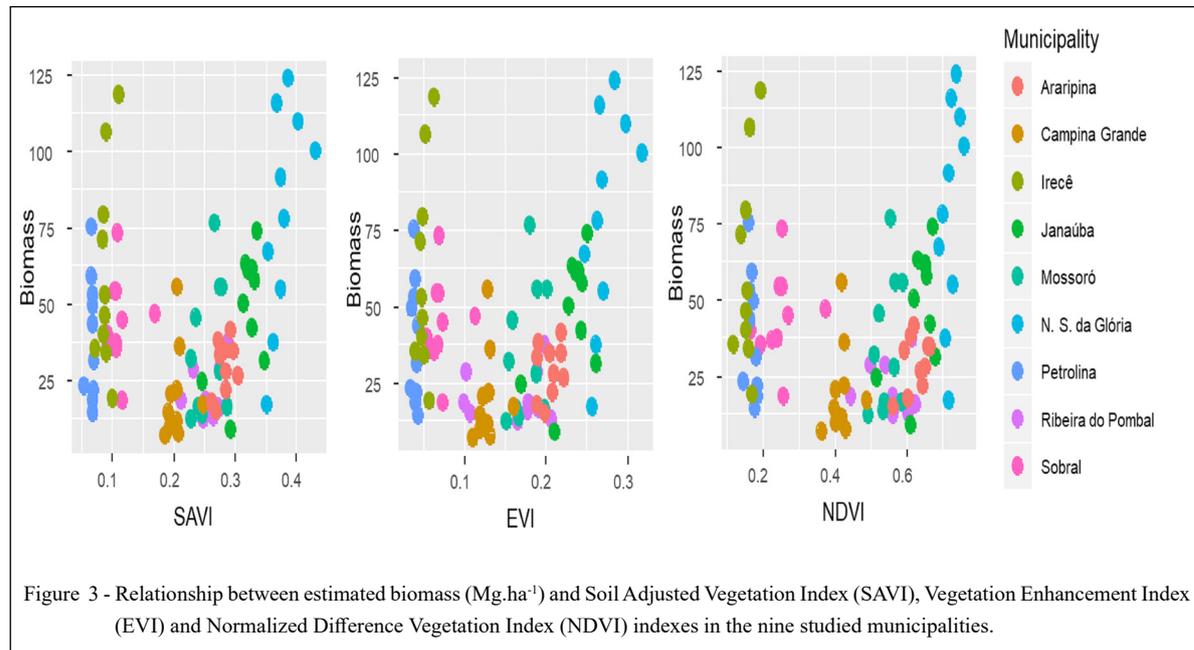
Images during the period of the year when the vegetation is green were not observed among the images available in the MMA collection for the clipping of the municipalities of Petrolina, Irecê and Sobral, so that no good vegetation index values were obtained (Figure 3). Thus, they were removed

from the analysis. Most indexes are based on the vegetation behavior in the red and near infrared bands of the electromagnetic spectrum, in which photosynthetic pigments of leaves absorb radiation in the visible range and reflect in the near infrared (PONZONI et al., 2012).

Next, it was possible to observe that there were two discrepant points in the biomass data-in the boxplot graph for the remaining plots in the municipalities of Araripina, Campina Grande, Janauba, Mossoró, Nossa Senhora da Gloria and Ribeira do Pombal. These outliers were removed from the analysis, leaving 58 plots to be correlated with the vegetation index.

The SAVI values for L=1, SAVI for L=0.5, NDVI and EVI regarding the vegetation varied between 0.1835 to 0.4294, 0.2197 to 0.5019, 0.3622 to 0.7584 and 0.0987 to 0.3169, respectively. All vegetation indexes presented positive and significant correlations at 5% probability level with the aboveground biomass estimated by the alomeric equations, with Spearman's correlation coefficient of 0.64, 0.63, 0.58 and 0.61 for SAVI L = 1, SAVI L = 0.5, NDVI and EVI, respectively. The relations among the estimated biomass and the spectral variables were moderate.

However, the correlation coefficients obtained were lower than those reported by LIMA JÚNIOR et al. (2014) with Pearson's correlation

Figure 3 - Relationship between estimated biomass (Mg.ha^{-1}) and Soil Adjusted Vegetation Index (SAVI), Vegetation Enhancement Index (EVI) and Normalized Difference Vegetation Index (NDVI) indexes in the nine studied municipalities.

coefficient of R = 0.84 for the relationship between NDVI and aboveground biomass of 20 plots in Petrolina-PE. The values were similar to those found by ALMEIDA et al. (2014), in which the authors related the spectral variables to the structural variables (tree basal area, height and volume) and obtained correlations between 0.52 to 0.72 at 5% probability. A correlation coefficient (R) ranging from 0.33 to 0.60 relating the mean height and basal area variables with the spectral variables were generally considered low by Accioly et al. (2002) in similar areas. The vegetation indexes which best correlated with biomass were SAVI with L = 1, followed by SAVI with L = 0.5, EVI and NDVI (p-value less than 0.05. The SAVI performance is attributed to this index, being suitable for studying low density vegetation cover areas, while the L constant has the function of minimizing the soil effect on the vegetation signal, since the sensitivity of the vegetation index in relation to the soil is higher in sparse vegetable canopies (PONZONI et al., 2012). In addition, the L = 1 constant obtained a better result than L = 0.5, because the former is more suitable for places with less dense vegetation cover, such as the caatinga vegetation.

These results are in agreement with those reported by VIGANÓ et al. (2011). The authors analyzed the performance of NDVI and SAVI vegetation indexes in Petrolina-PE in order to identify what best discriminates the vegetation cover. According to the study, SAVI best discriminated the targets, as it presented the largest number of value classes corresponding to the caatinga vegetation area. SILVA & GALVÍNCIO (2012) observed the vegetation behavior during the period from 2001 to 2010 and the different information that the SAVI and NDVI may present on the same surfaces. The authors suggested the use of SAVI for the dry season in caatinga areas instead of NDVI due to soil reflectance which affects the vegetation cover responses.

Despite its adjustment to atmosphere and soil, EVI had a lower correlation than SAVI. This shows that Caatinga vegetation areas have a greater soil influence than atmospheric factors. The improved vegetation index was developed to better respond to the vegetation signal, increasing the sensitivity of its detection in regions with higher biomass densities (PONZONI et al., 2012).

A description of remote sensing prediction models is presented in table 2. All indexes presented good and significant adjustments at 5% probability, but only SAVI with L = 1 met the assumption of homogeneity of variances by the Breusch-Pagan test (BREUSCH & PAGAN, 1979). The linear regression between vegetation index and biomass estimated using allometric equations calculated by the least squares method resulted in equations

Table 2 - Adjusted linear regression models, significant at 5% [*].

| Vegetation indexes | ---------------------------Model------------------------- | | | | | MAE |
	Explanatory Variables	R²	R²$_{aj}$	Sd	RMSE	(loocv)
SAVI L=1	Bio = -49.38 + 307.21 [*] SAVI	0.50[*]	0.49	17.49	17.18	14.07
SAVI L=0.5	Bio = -50.35 + 257.47 [*] SAVI	0.48[*]	0.47	17.86	17.55	14.50
NDVI	Bio = -46.44 + 142.64 [*] NDVI	0.33[*]	0.33	20.20	19.86	16.49
EVI	Bio = -24.555 + 315.31 [*] EVI	0.44[*]	0.43	18.48	18.16	15.29

In which: Bio is the biomass (Mg.ha^{-1}), R² is the coefficient of determination, R²$_{aj}$ is the adjusted coefficient of determination, Sd is the Standard deviation (Mg.ha^{-1}), RMSE is the square root of the mean square error (Mg.ha^{-1}) and LOOCV(MAE) is the mean absolute error (Mg.ha^{-1}) calculated for leave-one-out cross-validation prediction statistical errors.

with determination coefficients (R²), is the adjusted coefficient of determination (R²$_{aj}$), the square root of the mean square error (RMSE) and the standard error deviation (Sd), as well as the mean absolute error (MAE) of prediction in the biomass estimate calculated for statistical forecasting errors of the leave-one-out cross validation of the 57 interactions, which are shown in table 2.

The best adjusted equation was the SAVI equation with L=1, presenting the best coefficient of determination and standard error of the residue. The Shapiro-Wilk test (SHAPIRO & WILK, 1965) for the normality of the residues obtained the p-value = 0.609, assuming the significance level at 5% probability and concluding that the residues have normal distribution.

The adjusted equation from the linear regression for SAVI and biomass presented a R² = 0.50, R²aj = 0.49, RMSE = 17.18 Mg.ha^{-1} and mean absolute error of prediction MAE = 14.07 Mg.ha^{-1}. This result was lower than that reported by LIMA JÚNIOR et al. (2014), who obtained the R² = 0.70 for the estimated biomass ratio and the NDVI for Landsat 5 images in 20 plots in the city of Petrolina-PE. One of the reasons that could explain this difference is the plot selection methodology. In the present study, transects were traced for the choice of plots, while the authors in the cited study split five classes from the NDVI image with different vegetation structures, allowing the selection of 20 plots with vegetation density ranging from low to high aboveground biomass.

BRANDÃO et al. (2007) also observed strong determination coefficients for vegetation in Barbalha-CE, with R² = 0.74 for thin vegetation and R² = 0.90 for dense vegetation, from the relationship between NDVI and biomass. The high determination coefficients may be related to the data collection method, because biomass was calculated by the absorbed photosynthetically active radiation method and the light use efficiency model. However, it was larger than that reportedby SAMPAIO et al. (2015), who adjusted an equation for NDVI and the estimated biomass in 80 plots in an area covering 68 municipalities in the state of Pernambuco, covered by two Resourcesat 1/LISS III satellite images, with a determination coeficient of R² = 0.36. This low adjustment was explained by the authors due to large variations at both ends of the biomass range, below 22 and bigger than 50 Mg.ha^{-1}.

The adjusted multiple linear regression models in which three explanatory variables were used (the red band, the NDVI and the SAVI) were able to explain about 40% of the population variation in height and 60% of variation in wood volume for the municipalities of Porto da Folha and Canindé de São Francisco (SE) (ALMEIDA et al., 2014). In a study conducted at FLONA do Araripe (CE), ACCIOLY et al. (2002) obtained the lowest determination coefficients with R² = 0.36 for the mean height variable related to the simple ratio vegetation index (SR) and the vegetation structure index (SI). Equations with R² = 0.30 were generated for the basal area with the same index. According to LIMA JUNIOR et al. (2014), the diferences in the studies developed in different caatinga areas are due to the edaphoclimatic characteristics of each area, in addition to differences in vegetation type, and different anthropization levels at each site. From the spectral point of view, most of the vegetation indices were developed based on the

reflectance of red and near infrared (EVI, NDVI, among others) due to the bands which comprise the range of red and near infrared being sensitive to the vigor of green vegetation (CURRAN, 1989). Thus, the vegetation indices developed for multispectral data are difficult to relate to the non-photosynthetic fraction of vegetation and the task is even more challenging when there is an attempt to differentiate it from exposed soil, as for caatinga. Under this point of the view, the SAVI index represents an adavance to estimate the aboveground caatinga biomass.

CONCLUSION

The biomass estimates with the allometric equations showed wide biomass variability in the different caatinga vegetation areas, which can be explained according to the variable edaphoclimatic conditions in the semi-arid region.

The vegetation index values related to caatinga vegetation showed a positive and significant correlation with the aboveground biomass values estimated by the allometric equations.

The SAVI model with L = 1 was the best fit in the linear regression analysis between the estimated biomass and the vegetation index calculated by the spectral variables obtained from the orthorectified images of the RapidEye satellite, with a determination coefficient R^2 = 0.50 and and mean absolute error of prediction, MAE = 14.07 Mg.ha^{-1}. Additional studies are needed to increase the accuracy of the estimates, which can be obtained by increasing the number of sample points and using images from other satellites which have better spectral responses to the caatinga vegetation, helping to improve and redefine non-destructive sustainable forest management strategies.

The present study confirmed the importance of the SAVI index in improving the accuracy and reducing the uncertainty about the estimation of aboveground caatinga biomass. The results highlighted the importance of including the soil factor in dry-land forest due to the spectral reflectance of a Caatinga forest being a combination of the reflectance spectra of plants and soil components, governed by the optical properties of these elements and photon exchanges within the caatinga.

As the vegetation changes within the season (rainy season and dry season), the soil contribution increases or decreases, but may still remain significant, depending on plant density. Therefore, we intend to include climate and edaphic variables in future studies, mainly precipitation and soil type.

DECLARATION OF CONFLICT OF INTEREST

We have no conflicts of interest to declare.

AUTHORS' CONTRIBUTION

The authors all equally contributed to the manuscript.

REFERENCES

ACCIOLY, L.J.O. et al. Empirical relationship between vegetation structure and TM/LANDSAT data . **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 6, n. 3, p.492-498, 2002. Available from: <https://doi.org/10.1590/S1415-43662002000300019>. Accessed: Jun. 06, 2021. doi: 10.1590/S1415-43662002000300019.

ALMEIDA, A.Q. et al. Empiric relations between dendrometric characteristics of the Brazilian dry forest and Landsat 5 TM data. **Pesquisa Agropecuária Brasileira**. v.49, n.4, p.306-315, 2014. Available from: <https://doi.org/10.1590/S0100-204X2014000400009>. Accessed: Jun. 06, 2021. doi: 10.1590/S0100-204X2014000400009.

AMORIM, I.L. et al. Flora and structure of the tree and shrub vegetation of the caatinga at Seridó, Rio Grande do Norte State, Brazil. **Acta Botanica Brasilica**. v.19, n.3, p.615-623, 2005. Available from: <https://doi.org/10.1590/S0102-33062005000300023>. Accessed: Jun. 06, 2021. doi: 10.1590/S0102-33062005000300023.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v.4, p. 40-79, 2010. Available from: <https://doi.org/10.1214/09-SS054>. Accessed: Jun. 06, 2021. doi: doi.org/10.1214/09-SS054.

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v.191, p. 192-213, 2012. doi: /10.1016/j.ins.2011.12.028.

BRAND, M.A. Potential use of caatinga forest biomass under sustainable management for energy generation. **Ciência Florestal**. v.27, n.1, p.117-127, 2017. Available from: <https://doi.org/10.5902/1980509826452>. Accessed: Jun. 06, 2021. doi: 10.5902/1980509826452.

BRANDÃO, Z.N. et al. Uso do NDVI para determinação da biomassa na chapada do Araripe. **Simpósio Brasileiro de Sensoriamento Remoto**. Saõ José dos Campos: INPE, [Internet] 2007. p.75-81. Available from: <http://marte.sid.inpe.br/rep/

dpi.inpe.br/sbsr@80/2006/11.15.13.39.22?mirror=ltid.inpe. br/sbsr/2004/09.22.16.44&metadatarepository=dpi.inpe.br/ sbsr@80/2006/11.15.13.39.23>. Accessed: Jun. 06, 2021.

BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**. v.47, n.5, p.1287–1294, 1979. Available from: <https://doi.org/10.2307/1911963>. Accessed: Jun. 06, 2021. doi: 10.2307/1911963.

CLIMATE-DATA.ORG. **Clima**. Available: <https://pt.climate-data.org/>. Accessed: Out. 06, 2018.

COSTA, T.C.C. et al. Analysis of degradation of 'Caatinga' in the desertification nucleus of Seridó - Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13(Suppl.), p.961-974, 2009. Available from: <https://doi.org/10.1590/S1415-43662009000700020>. Accessed: Jun. 06, 2021. doi: 10.1590/S1415-43662009000700020.

FAO (Food and Agricultural Organization of the United Nations). The state of the world's forests 2020. **Forests, biodiversity and people**. Rome. 2020. Available from: <http://www.fao.org/state-of-forests/en/>. Accessed: Jun. 06, 2021. doi.org/10.4060/ca8642en.

CURRAN, P. J. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, v. 30, n. 3, p. 271–278, 1989. . Available from: <https://www.sciencedirect.com/science/article/abs/pii/0034425789900692>. Accessed: Jun. 06, 2021. Doi: 10.1016/0034-4257(89)90069-2.

FERRAZ, A.S. et al. Biomass stock estimate in a forest fragment using orbital data. **Floresta e Ambiente**, v.21, n.3, p.286-296, 2014. Available from: <https://www.scielo.br/j/floram/a/zRYT8gBK4k9CGRvWsSdg8sJ/?format=pdf&lang=pt>. Accessed: Jun. 06, 2021. doi: 10.1590/2179-8087.052213.

FORMIGONI, M.H. et al. Multi-temporal analysis of northeast vegetation by means of modis-evi data. **Ciência Florestal**, Santa Maria, v.21, n.1, p.1-8. 2011. Available from: <https://doi.org/10.5902/198050982740>. Accessed: Jun. 06, 2021. doi: 10.5902/198050982740.

HUETE, A.R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, 25 v.3, p.295-309, 1988. Available from: <https://doi.org/10.1016/0034-4257(88)90106-X>. Accessed: Jun. 06, 2021. doi: 10.1016/0034-4257(88)90106-X.

HUETE, A.R. et al. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**. v.59, n.3, p.440-451, 1997. Available from: <https://doi.org/10.1016/S0034-4257(96)00112-5>. Accessed: Jun. 06, 2021. doi:10.1016/S0034-4257(96)00112-5.

LIMA JÚNIOR, C. et al. Estimation of "caatinga" woody biomass using allometric equations and vegetation índex. **Scientia Forestalis**. V.42, n.102, p.289-298, 2014. Available from: <https://www.ipef.br/publicacoes/scientia/nr102/cap13.pdf>. Accessed: Jan. 30, 2019.

PLANET, **Planet Imagery Product Specifications**, 2018. Available from: <https://www.planet.com/products/planet-imagery/>. Accessed: Set. 01, 2018.

PONZONI, J.F. et al. **Sensoriamento remoto da vegetação**. São Paulo: Oficina de Textos, 2012.

R CORE TEAM (2017). **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical

Computing, 2012. Available from: <https://www.R-project.org/>. Accessed em: Mar. 04, 2017.

SAMPAIO, E.V.S.B; SILVA, G.C. Biomass equations for Brazilian semiarid caatinga plants. **Acta bot. Bras**. v.19, p.935-943, 2005. Available from: <https://doi.org/10.1590/S0102-33062005000400028>. Accessed: Jun. 06, 2021. doi: 10.1590/S0102-33062005000400028.

SAMPAIO, E.V.S.B.; FREITAS, A.D.S. Produção de biomassa na vegetação nativa do semi-árido nordestino. In: Menezes, R.S.C. et al., editors. **Fertilidade do solo e produção de biomassa no semi-árido**. Ed. Universitária da UFPE, 2008. Cap.1, p. 11-25.

SAMPAIO, E.V.S.B. et al. Estimation of phytomass stock through satellite imagery in a semiarid area in Pernambuco state, Brazil. In: Hermann, T. et al., editors. **Land Quality and Landscape Processes**, 2015, p 113-117. ISBN 9789279483103.

SANTOS, F.A. et al. Albedo Seasonal Behavior and vegetation indices in the upper basin Paraíba River. **Revista Brasileira de Geografia Física**. v07, n.5, p. 1015-1027, 2014. Available from: <https://periodicos.ufpe.br/revistas/rbgfe/article/view/233416>. Accessed: Jun. 06, 2021. doi: 10.26848/rbgf.v7.5.

SANTOS, R.C. et al. Estoques de volume, biomassa e carbono na madeira de espécies da Caatinga em Caicó, RN. **Pesquisa Florestal Brasileira**. V.36, n.85, p.1-7, 2016. Available from: <https://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/772>. Accessed: Jun. 06, 2021. doi: 10.4336/2016.pfb.36.85.772.

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, v.52, n.3/4, p.591-611,1965. Available from: <https://doi.org/10.2307/2333709>. Accessed: Jun. 06, 2021. doi: 10.2307/2333709.

SILVA, G.C., SAMPAIO, E.V.S.B. Aboveground biomass of caatinga plants. **Revista ÁRVORE**. V.32 n.3, p.567-575. 2008. Available from: <https://doi.org/10.1590/S0100-67622008000300017>. Accessed: Jun. 06, 2021. doi:10.1590/S0100-67622008000300017.

SILVA, L. C.; GALVÍNCIO, J. D. Comparative Analysis of Changes in NDVI and SAVI in PELD - 22 - Petrolina - PE, in the First Decade of the XXI Century. **Revista Brasileira de Geografia Física**. V.6, p.1446-1456, 2012. Available from: <https://doi.org/10.26848/rbgf.v5i6.232936>. Accessed: Jun. 06, 2021. doi: 10.26848/rbgf.v5i6.232936.

TIAN F. et al. Mapping gains and losses in woody vegetation across global tropical drylands. **Global Change Biology**. v.23, p.1748–1760, 2017. Available from: <https://doi.org/10.1111/gcb.134642>. Accessed: jun. 06, 2021. doi: 10.1111/gcb.13464 2.

VIGANÓ, H.A. et al. Análise do desempenho dos Índices de Vegetação NDVI e SAVI a partir de imagem Aster. **INPE**, Curitiba. 2011. p. 1828 -1834. Available from: <http://marte.sid.inpe.br/col/dpi.inpe.br/banon/2003/12.10.19.30.54/doc/mirror>. Accessed: Jul. 07, 2018.

VIRGENS, A.P. et al. Biomassa de espécies florestais em área de caatinga arbórea. **Pesquisa Florestal Brasileira**, v.37, n.92, p.555-561, 2017. Available from: <https://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/1465>. Accessed: Jun. 06, 2021. doi: 10.26848/rbgf.v5i6.232936. doi: 10.4336/2017.pfb.37.92.1465.