An Acad Bras Cienc (2024) 96(3): e20221041 DOI 10.1590/0001-3765202420221041
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

FORESTRY SCIENCE

Artificial Neural Network and Remote Sensing combined to predict the Aboveground Biomass in the Cerrado biome

PAULA L.G. OLIVEIRA, ERALDO A.T. MATRICARDI, EDER P. MIGUEL, BEN HUR MARIMON JÚNIOR & ALBA VALÉRIA REZENDE

Abstract: Cerrado is the second largest biome in Brazil, and it is responsible for providing us several ecosystem services, including the functions of storing Carbon and biodiversity conservation. In this study, we developed a modeling approach to predict the Aboveground biomass (AGB) in Cerrado vegetation using Artificial Neural Networks (ANNs), vegetation indices retrieved from RapidEye satellite imagery, and field data acquired within the Federal District territory, Brazil. Correlation testing was performed to identify potential vegetation index candidates to be used as input in the AGB modeling. Several ANNs were trained to predict the AGB in the study area using vegetation indices and field data. The optimum ANN was selected according to criteria of mean error of the estimate, correlation coefficient, and graphical analysis. The best performing ANN showed a predictive power of 90% and RMSE less than 17%. The validation tests showed no significant difference between the observed and ANN-predicted values. We estimated an average AGB of 16.55± 8.6 $Mg.ha^{-1}$ in shrublands in the study area. Our study results indicate that vegetation indices and ANNs combined could accurately estimate the AGB in the Cerrado vegetation in the study area, showing to be a promising methodological approach to be broadly applied throughout the Cerrado biome.

Key words: ANN-MLP, Brazilian savanna, remote sensing, biomass.

INTRODUCTION

The Cerrado comprises the second largest Brazilian biome and the most diverse tropical savanna in the world (Klink & Machado 2005). The history of land use and land cover changes in the Cerrado in the last decades reveals an intensive anthropogenic pressure on that biome, where the unsustainable use of natural resources has been causing several ecosystem disturbances, making it one of the most threatened biome in Brazil (Oliveira et al. 2019). Approximately 50% of native vegetation of the Cerrado biome had been cleared by 2022, a total of 99,5 million hectares (Mapbiomas 2024). The land demands for expansion of urban centers, mechanized agricultural crops, and pastures are the main causes of deforestation in the Cerrado biome (Roquette 2018).

The Cerrado biome presents a crucial role for maintaining the global environmental quality, responsible for providing several environmental services, including the ability to store expressive amounts of carbon on its biomass (Oliveira et al. 2019). Biomass measurements have been supporting definition of environmental policies by the United Nations Framework Convention on Climate Change (UNFCCC), as countries must regularly monitor emissions related to biomass gains and losses. In

addition, the biomass data are extremely important at the national level to support the formulation of public policies aiming to curb deforestation (Herold et al. 2019).

The increasing interest in monitoring land cover changes has been widely demanding spatial data of vegetation biomass (Yang et al. 2018, Yu & Saatchi 2016). Remote sensing data integrated with forest inventories have been broadly applied to estimate aboveground biomass (Barbosa et al. 2014). Although vegetation biomass estimates based on field measurements are more accurate, they are more time and money consuming and normally unsuitable for large-scale and remote areas (Kumar & Mutanga 2017).

Therefore, there are an increasing demand for developing new methodological approaches to estimate biomass at low cost applied to large scale and long-term monitoring study areas (Barbosa et al. 2014, Roitman et al. 2018). In this case, remote sensing can provide substantial and valuable amounts of data for extensive areas, with high repeatability and relatively low cost (Meneses et al. 2019).

Optical imagery is the most widely used remotely sensed data for natural resources mapping due to its global availability and the diversity of orbiting sensors (Mendes et al. 2019). Similarly, active systems like Synthetic Aperture Radar (SAR), InSAR (Interferometric Synthetic Aperture Radar), and LiDAR (Light Detection and Ranging) have been successfully applied to estimate aboveground biomass and forest structure parameters (Hyde et al. 2006, Duncanson et al. 2020, Zolkos et al. 2013). However, the most widely used remotely sensed data for biomass estimation are reflectance and vegetation indices (Cassol et al. 2014, Ferraz et al. 2014).

The vegetation indices are mathematical combinations of different spectral bands aiming to improve the information inherent to the target´s reflectances by extracting the variability caused by the vegetation characteristics and minimizing the effects of soil and atmosphere (Viña et al. 2011). Moreover, remote sensing data have been used to estimate aboveground biomass through existing correlations between the spectral responses of targets and the field-based variables (Herold et al. 2019, Kumar et al. 2016, Mendes 2019).

The use of remote sensing data to quantify aboveground biomass can be applied to large spatial scale by extrapolating the biomass distribution throughout a surface territory of interest (Cassol et al. 2014, Zhu & Liu 2015). Regression analysis is the most common tool of prediction approaches when relating satellite data to dendrometric variables (Günlü et al. 2019, Yu 2019). However, in recent years, artificial neural networks have been widely used modeling tools since they can provide better solution for complex and nonlinear problems using universal estimation attributes (Günlü et al. 2019, Panda et al. 2010).

The ANN does not necessarily assume that the data have normal distribution and linear relationships. Thus, they can model different data sources, such as satellite imagery and ground measurements, including several independent variables (Günlü et al. 2019). In general, an ANN consists of simple processing units, which have natural ability to store experimental knowledge and make it available for modeling applications. To achieve superior performance, ANNs employ a massive interconnection of computational cells, called “neurons” or processing units (Lorrentz 2015).

The ANN technique includes training, verification, and testing stages using subsets of data which are made up of data values randomly selected from the data set (Ercanli et al. 2018). Based on the learning process, the ANNs can approximate any nonlinear relationship that exists between a set

of input data and its corresponding set of output data (Panda et al. 2010). Some studies have been showing effective results in predicting forest parameters by using ANNs and remotely sensed derived data (Carrijo et al. 2020, Ercanli et al. 2018, Ferraz et al. 2014, Günlü et al. 2019, Silva et al. 2019).

In this analysis, we tested and developed a model to estimate the aboveground biomass in shrublands in the Cerrado biome, Brazil, using remotely sensed data and ANN combine in study area located in the Federal District, Brazil. Our study results showed to be promising approach to provide accurate data of aboveground biomass of Cerrado vegetation in Brazil.

MATERIALS AND METHODS

Study area

The study area is spatially located in the Federal District of Brazil (Figure 1), encompassing a total of 578,919 hectares within the Cerrado biome (Rosa 2016).

The Latosols predominantly occur in the Cerrado biome, covering 46% of its territory, followed by the Quartzarenic Neosols, which cover 15% of its extension. In general, deep, well-drained, acidic, nutrient-poor, and aluminum-saturated soils predominantly occur in that biome (EMBRAPA 2018).

According to Köppen climate classification, the study area is characterized by the Aw climate type, showing a wet season between October and March, and a dry season between April and September (Ferreira et al. 2003). The annual average precipitation is 1,500mm (Silva et al. 2008).

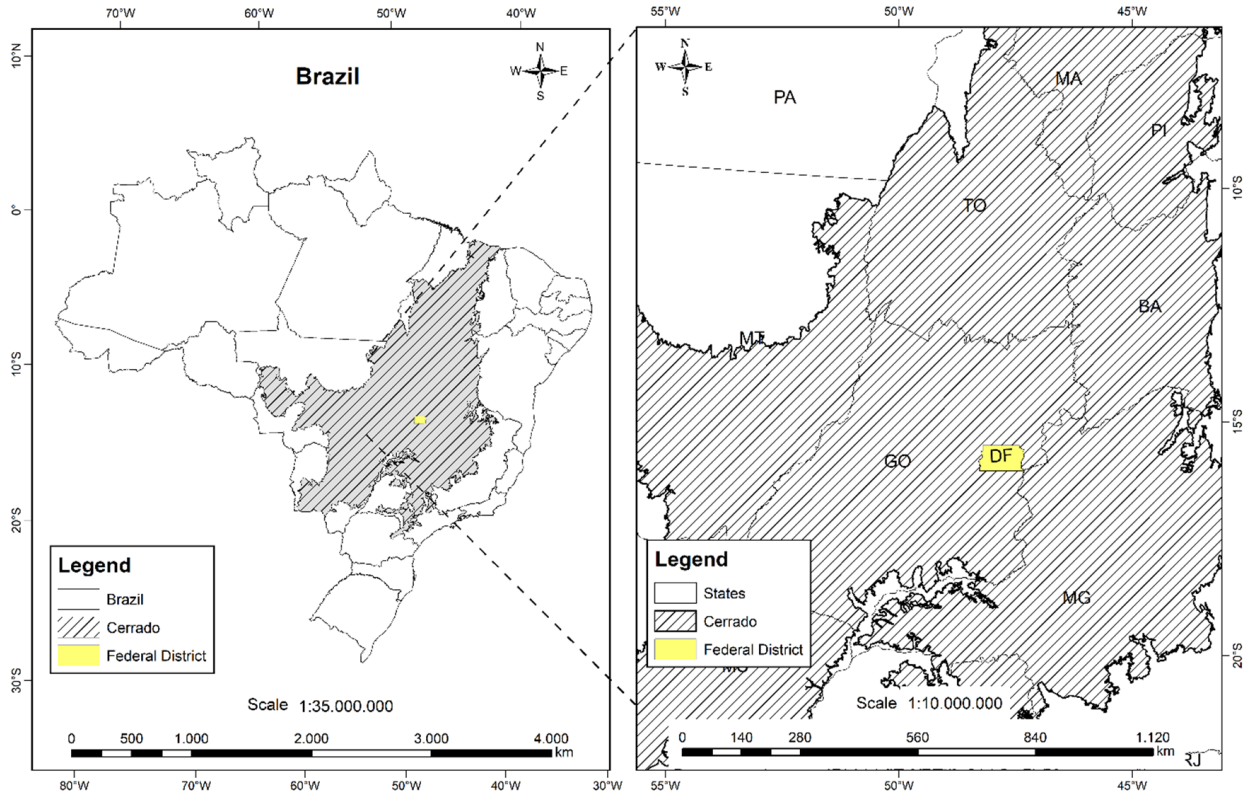

Figure 1. Spatial location of the study area (the Federal District, Brazil)

The Cerrado *sensu stricto*, a savanna-like vegetation type, encompasses about 70% of the entire Cerrado biome. It is characterized by a herbaceous stratum predominantly featuring native grasses, with tree and shrub cover varying between 10% and 60% (Coelho et al. 2020). In 2022, approximately 24% of the total territory of the Federal District was characterized by Savanna formations, totaling 141,964 hectares. Forest formations covered 42,227 hectares, while Grassland formations occupied 41,758 hectares, equivalent to 7.3% and 7.2% of the Federal District territory (Figure 2), respectively (Mapbiomas 2024).

Field observed details of the Cerrado vegetation in our study area are shown in Figure 3.

Data collection

The forest inventory was conducted in 2011 by the National Forest Inventory (NFI) team. The NFI is one of the main forest survey conducted by the federal government to produce information on Brazilian forest resources (Rosa 2016). The NFI data collection was based on systematic sampling plots over a national grid of sampling units. Based on this systematic sampling grids, field variables are collected throughout the country territory and organized by State (Rosa 2016). A total of 42 sampling plots were established within savanna vegetation in the Federal District, selected from 15 conglomerate samplings, and field variables were collected (Figure 4). Each sampling plot used in the National Forest Inventory (NFI) was a subset of conglomerate samplings, consisting of 4 sub-samples of 0.1

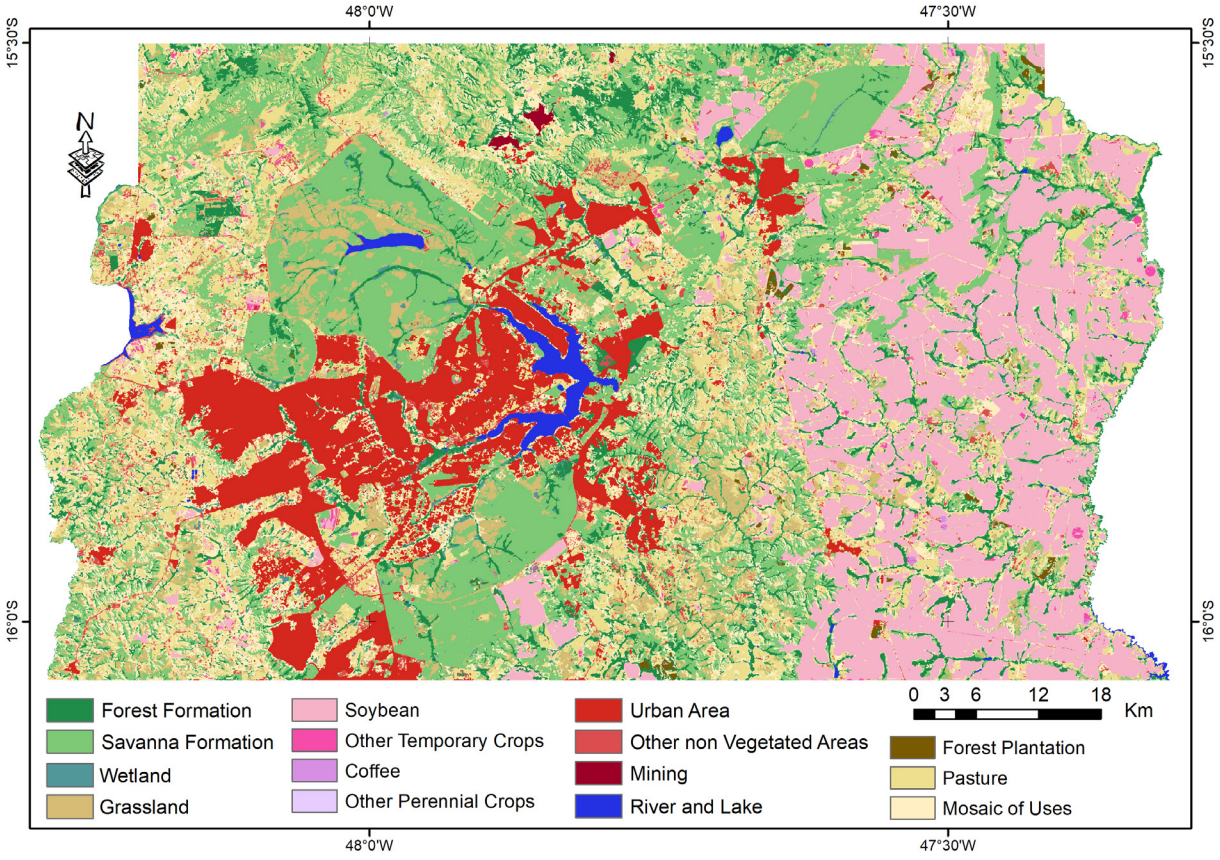

Figure 2. Land use and land cover in the Federal District territory by 2022. Source: Mapbiomas (2024).

hectares (20m x 50m) each. Some of the conglomerate sub-samples were excluded as they were not situated within savanna vegetation.

The sampling subunits were divided into another 10 subplots of 10m x10m dimensions. All standing trees showing diameters at base height (DB – 30cm from the ground) equal to or greater than 10cm were identified, and their DB and total height (Ht) were measured within each sample plot. Additionally, a sample plot of 10x10m was set within each conglomerate subsample, where shrubs and trees showing diameters between 5 and 10cm were measured.

In addition to data provided by the NFI, a dataset of forest inventory conducted in the Água Limpa Experimental Farm (FAL), University of Brasilia, was also used. The FAL dataset was based on field data collected in 2005 in 18 permanent plots of 0.1ha (20x50m) spatially located within an experimental area of the FAL (Figure 4). Each field plot was divided into 10 subsamples of 10x10m where diameter and height of all trees and shrubs were measured for individuals showing DB (base diameter at 30cm) equal to or greater than 5cm. In this case, we assume that there were no significant changes in vegetation biomass, as no fires or severe climate events had been reported in that study site between 2005 and the acquisition date of the remotely sensed data in 2011.

Biomass prediction

Assessment of aboveground biomass requires in situ biomass estimates for calibration and validation of biomass prediction algorithms. In this study, we estimated biomass using dendrometric variables (basal area and tree heights) based on field measurements. We relied our analysis on the model accuracy of the relationship between diameter and height to estimate the aboveground biomass.

Although wood density may influence estimates of aboveground biomass, we did not consider it due to potential factors that could affect it. Wood density can vary significantly among different tree species, among individuals of the same species, and even within a single tree, depending on factors

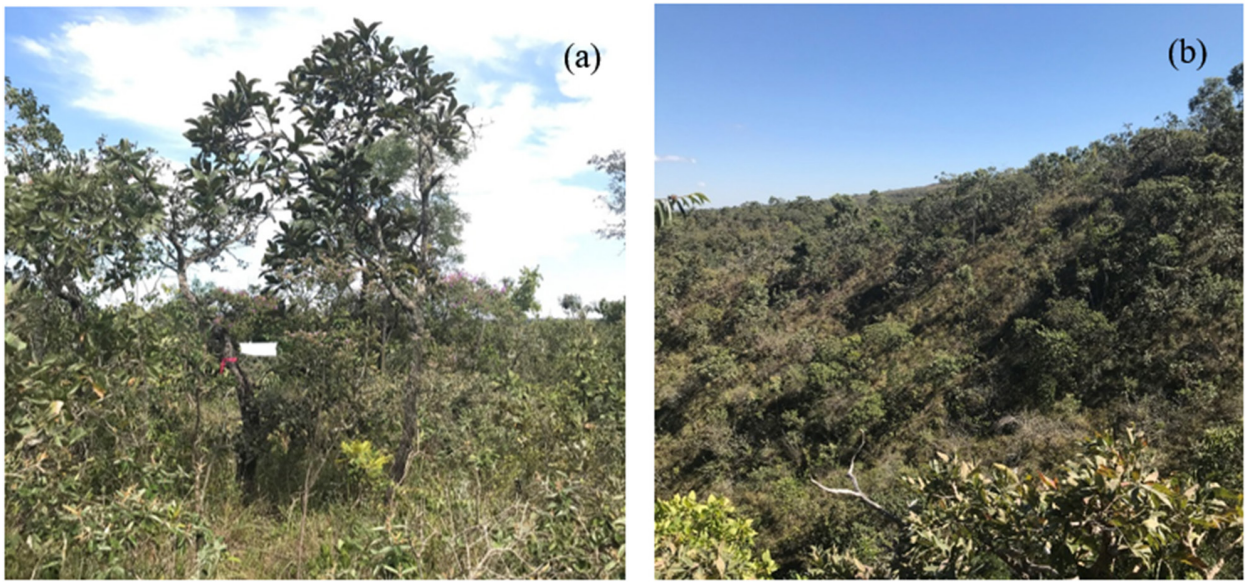

Figure 3. a) close view of Cerrado vegetation (location: 15º43'11"S; 47º56'02"W) b) panoramic views of Cerrado vegetation (location: 15º40'35"S; 47º53'14"W), both within the Brasília National Park (PNB). Photos taken by Paula Lopes Germano Oliveira, April 2021.

such as age, growth, size, and environmental conditions (Chave et al. 2005, Anderson et al. 2009, Henry et al. 2009). This variation poses a significant challenge when using wood density in biomass equations, especially in extensive studies involving multiple tree species or environments (Anderson et al. 2009, Henry et al. 2009). Additionally, in some cases, wood density may be either unavailable or difficult to accurately measure for all samples or species considered in specific study sites. Therefore, since in certain biomass analyses, the primary objective may be to discern general trends in vegetation growth, biomass production, or forest dynamics on a broader scale, wood density may not be as critical for these broader goals.

The basal area (Gi) was obtained according to the equation:

$$G_i = \Sigma g_{ji}\,(x) \quad (1)$$

Where g_{ij} is the sectional area of tree j in plot i (m²), obtained from the equation $g = (\pi * DB^2)/40.000$ Where π is a constant is a constant equal to 3.14159 and DB (cm) is the diameter at the base of the tree (30 cm above the ground).

In this study, the Aboveground Biomass (AGB) was estimated for all individuals showing Db ≥ 5cm using an allometric equation adjusted for cerrado sensu stricto vegetation, a savanna-like vegetation type (Rezende et al. 2006).

$$AGB = 0{,}4913 + 0{,}0291 * DB^2 * Ht \quad (2)$$

R^2 = 98,28% e Syx = 25,79%

Where DB (cm) is the diameter at the base of the tree and Ht is the height in meters.

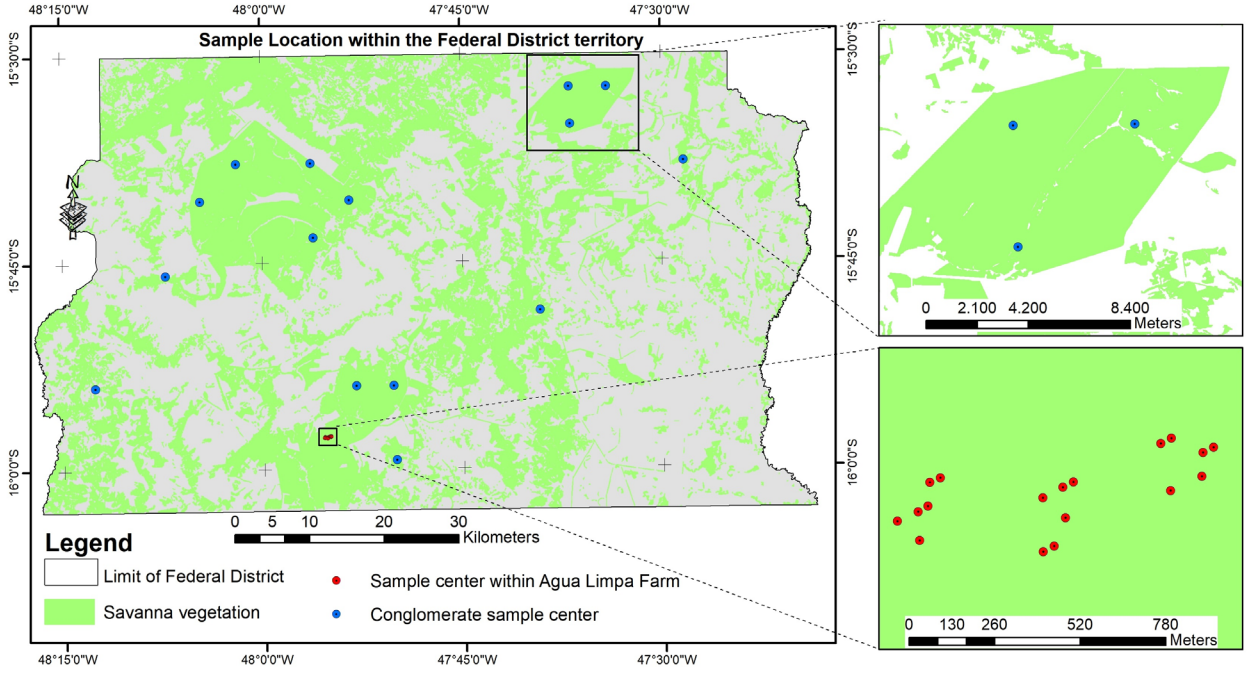

Figure 4. Sample locations within savanna vegetation in the Federal District territory.

The final biomass estimate was calculated using the allometric equation (Eq. 02) for each individual tree and its specific density value. All individual tree biomasses of each sample plot were summed and normalized to estimate the total aboveground biomass ($Mg.ha^{-1}$) of each sample unit of this study.

Remote Sensing data

RapidEye satellite scenes acquired in September and October 2011 were provided by the Brazilian Forest Service, covering the entire study area. We selected this RapidEye imagery because it has been widely and successfully used to estimate above-ground biomass (e.g., Gascón et al. 2019, Luz et al. 2022, Rana et al. 2023, Movchan et al. 2023) and is freely provided by the Brazilian Forest Service in Brazil, with images acquired between 2011 and 2014. Additionally, it includes the red-edge spectral band, a temporal resolution of 5 days, and a spatial resolution of 5m. The multispectral scanner on the RapidEye satellites acquires data in five spectral bands: blue (0.44-0.51 µm), green (0.52-0.59 µm), red (0.63-0.68 µm), red-edge (0.69-0.73 µm), and near-infrared (0.76-0.85 µm), which are widely applied in vegetation assessments (Kross et al. 2015) (Table I).

Vegetation indices

The reflectances of vegetation are related to plant pigments that are responsible for absorbing, reflecting, and transmitting incident electromagnetic radiation (Neto 2018). Vegetation indices are commonly applied to enhance the reflectance of vegetation by combining certain regions of the electromagnetic spectrum (Huete et al. 2002). In this study, vegetation indices were retrieved from the RapidEye images at top of atmosphere reflectance using ERDAS IMAGINE® 2011 software (Table II).

The NDVI is noted as one of the pioneer index applied to vegetation studies (Bannari et al. 1995). NDVI is highly correlated with vegetation greenness and is the most widely used index in research related to vegetation cover dynamics (Ghosh & Behera 2018, Guerini Filho et al. 2019). The NDRE and GNDVI are considered variations of NDVI. The NDRE includes the red edge band to increase sensitivity

Table I. Acquisition dates and RapidEye Scenes used in this analysis.

RapidEye Scenes	**Acquisition date**
2231727_2011-09-01T142332_RE5_3A-NAC_9878530_139445_0_2	01/09/2011
2231827_2011-06-20T143004_RE3_3A-NAC_9787921_138758_0_11	20/06/2011
2231828_2011-09-01T142328_RE5_3A-NAC_9828364_139008_13	01/09/2011
2331702_2011-09-01T142331_RE5_3A-NAC_9878542_139445_0_8	01/09/2011
2331803_2011-09-20T142153_RE5_3A-NAC_9837818_139076_0_0	20/09/2011
2331903_2011-09-20T142149_RE5_3A-NAC_9837824_139076_0_7 2231927_2011-09-01T142325_RE5_3A-NAC_11809590_154387_0_1 2231928_2011-09-01T142325_RE5_3A-NAC_9879244_139445_4 2331701_2011-09-01T142332_RE5_3A-NAC_9878508_139445_4 2331703_2011-06-22T143244_RE5_3A-NAC_9833172_139013_10 2331704_2011-09-20T142156_RE5_3A-NAC_10911877_148037_0_1 2331802_2011-09-01T142328_RE5_3A-NAC_9878512_139445_0_15 2331804_2011-09-20T142152_RE5_3A-NAC_9837860_139076_0_0 2331902_2011-09-01T142324_RE5_3A-NAC_9878562_139445_0_5 2331904_2011-10-09T141949_RE5_3A-NAC_9789181_138761_0_9	20/09/2011 01/09/2011 01/09/2011 01/09/2011 22/06/2011 20/09/2011 01/09/2011 20/09/2011 01/09/2011 09/10/2011

to chlorophyll variations (Gitelson & Merzlyak 1996) and the GNDVI includes the green and the near infrared bands of the electromagnetic spectrum, which is also sensitive to chlorophyl variation in the vegetation and has a higher saturation point than NDVI. CIgreen shows high correlation with chlorophyll content and leaf area index and the CIrededge is sensitive to chlorophyll content variation using the red edge band (Gitelson et al. 2005).

The performance assessment of the vegetation indices to predict biomass in our study area was conducted based on the correlation matrix analysis between the five calculated indices and the biomass measured in the field samples. The correlation matrix analysis was preceded by the Shapiro-Wilk normality test, which can be used to indicate the correlation method to be subsequently applied (parametric or non-parametric). Additionally, the correlation between explanatory variables (multicollinearity) was assessed by the Variance Inflation Factor (VIF) test. In this study, the statistical analysis was carried out using the software RStudio 1.0.143.

Modeling: Artificial Neural Networks

The stepwise methodological approach of biomass prediction using ANN and vegetation indices retrieved from remotely sensed data is presented in Figure 5. The steps are detailed as follows.

A supervised training of 500 neural networks was applied using the Intelligent Problem Solver (IPS) tool available on the software package Statistica ver. 7 (Statsoft 2007). The input and the output variables were defined in the network through the IPS tool (Carrijo et al. 2020). This tool standardizes the data between 0 – 1 and tests several architectures, which allowed us to optimize the ANN architecture by defining the best number of neurons in the hidden layer and the best activation functions of the hidden and output layers (Binoti et al. 2015, Vale et al. 2017), aiming to achieve the lowest error rate of the predictions (Miguel et al. 2018). We used the quasi-Newton algorithm developed by Broyden-Flecther-Goldfarb-Shanno (BFGS) in the IPS to process the neural networks (Broyden 1970, Fletcher 1970, Goldfarb 1970, Shanno 1970), which is the mostly applied quasi-Newton method and shows great resolution power for optimization and prediction problems (Guerrout et al. 2018).

Our trained networks used the Multilayer Perceptron (MLP) architecture in which the input layer consisted of the vegetation indices, the hidden layer consisted of “n” neurons, and the output layer consisted of one neuron representing the aboveground biomass (Carrijo et al. 2020, Gonçalves et al. 2021). The MLP architecture is distinguished by the presence of one or more hidden layers, where the

Table II. Vegetation Indices.

Index	Acronym	Formula	Reference
Normalized Difference Vegetation Index	NDVI		Rouse et al. (1974)
Normalized Difference Red Edge Index	NDRE		Fitzgerald *et al* (2010the canopy chlorophyll content index (CCCI)
Green Normalized Difference Index	GNDVI		Gitelson & Merzlyak (1996)
Red Edge Chlorophyll Index	$CI_{rededge}$		Gitelson et al. (2005)
Green Chlorophyll Index	CI_{green}		Gitelson et al. (2005)

function of the hidden neurons is to combine and improve the input and output network. By adding one or more hidden layers, the network will be able of extracting high-order statistics (Lorrentz 2015).

The neuron is the fundamental component of the ANN structure since it is responsible for processing data and information by receiving the input values until reaching the desirable results in the data output. The functionality of ANN neurons is estimated by combining and reproducing information based on connections between n possible inputs (input variables) x1, x2, ..., xn and the output (output variable) y. For each input a weight w is defined that represents the synapses where MLP networks store knowledge, enabling the learning/training process to reach more precise estimates and optimal values of the parameter (Carrijo et al. 2020, Gonçalves et al. 2021, Gorunescu & Belciug 2016). The artificial neuron is given mathematically, by:

$$Y_k = \varphi (V_k) \quad (3)$$

Where Y_k is the output of the artificial neuron; φ is the activation function; and V_k is the combinator of the output, given by:

$$V_k = \Sigma X_m W_m \quad (4)$$

In this analysis, we first randomly generated the weights for all networks. Then, each individual updated value evolved during the learning process based on the error function (Miguel et al. 2018). The network training run until the error rate was reduced to an acceptable range between the predicted values and the observed values provided to the network, known as the delta rule, or until the maximum number of cycles was reached (Miguel et al. 2018).

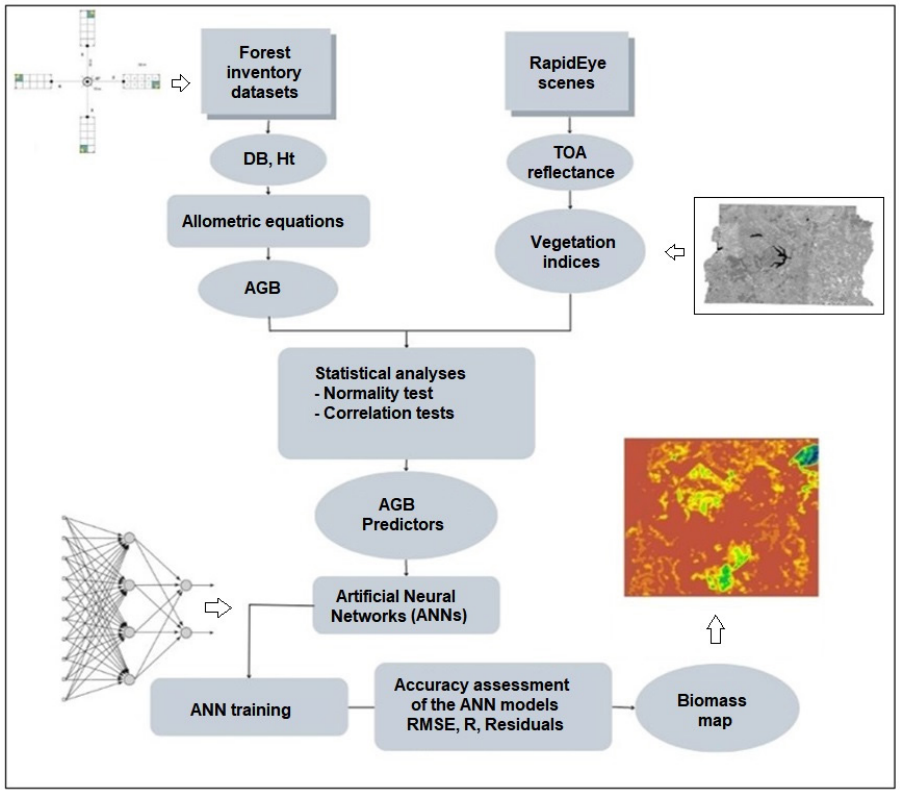

Figure 5. Flow diagram for biomass prediction using ANN (Artificial Neural Network) and vegetation indices retrieved from remotely sensed data in the Federal District, Brazil.

Statistical Analysis

Forty-two of the sixty sample plots (corresponding to 70% of the total samples) were randomly selected for training the neural network. The remaining 18 sample plots were excluded from the training dataset and later used to validate the modeling results (Zucchini 2000).

The model performances were statistically analyzed based on the root mean square error in percent (RMSE%), the correlation between estimated and observed values (R), and the graphical residual analysis (Chen et al. 2019). These model performance indicators have been used to estimate performance of biomass models (Yu et al. 2019).

$$RMSE\,\% = \frac{100}{\bar{Y}} * \sqrt{\frac{\Sigma(\hat{Y}_i)^2}{n}} \quad (5)$$

Where $\bar{Y}$ is the average of the observed AGB; Y_i is the observed AGB value in the sample plot i; and$\hat{Y}_i$ is the estimated AGB value by the RNA for each sample plot i. In general, an R value closer to one and lower RMSE% values indicate a better estimate of model performance (Li et al. 2019).

The ANN showing the best performance results was submitted to a validation process using Student's t-test and, subsequently, an aggregate difference in percentage (DA%) that statically indicates underestimation and overestimation errors (Carrijo et al. 2020, Gonçalves et al. 2021, Miguel et al. 2015, Vale et al. 2017). These analyses were conducted using the Microsoft Excel 2016 ® software.

Autocorrelation assessment

To assess autocorrelation in the dataset, we utilize the autocorrelation function (ACF), a statistical measure that examines the correlation between a series' values and its own past values at various lags. This test enables us to pinpoint autocorrelation patterns within the data, offering insights into its underlying structure. When applied to the residuals of diverse estimation procedures, the test aids in diagnosing the presence of autocorrelation in those residuals. Residual autocorrelation occurs when a model's residuals display systematic patterns after the completion of modeling, suggesting that the model might not fully capture the data structure (Kemp 2002).

RESULTS

Forest Inventory

We estimated the dendrometric variables based on the data collected in the study area (Table III).

Modelling: selection of independent variables

The Shapiro-Wilk test indicated statistically significantly non-normality in all independent variables and, consequently, we applied the non-parametric Spearman's correlation (Table IV).

The results indicated that pairwise correlation coefficients among all vegetation indices (VIs) are moderate and similar. Despite the high similarity of correlation coefficients among the tested Vegetation Indices, the GNDVI and the CIgreen showed the highest correlation (Table IV).

Based on the variance inflation factor (VIF), we observed that there was multicollinearity between the GNDVI and CIgreen indices. Multicollinearity indicates that two or more explanatory variables show a strong linear relationship and, therefore, one variable does not add any effect to another. In

Table III. Dendrometric variables estimated from a forest inventory conducted in the Cerrado vegetation in the study area.

	DB	Ht	G	AGB
Minimum	7,72	2,26	2,63	5,12
Maximum	22,64	8,87	19,95	39,46
Mean	11,57	3,52	10,32	16,55
Variance	9,70	1,79	12,85	73,76
SD	3,11	1,34	3,58	8,59
CV%	26,91	37,94	34,75	51,89

Legend: DB = diameter at breast height (cm), Ht = total height (m), G = basal area ($m^2.ha^{-1}$), AGB = above ground biomass ($Mg.ha^{-1}$).

Table IV. Spearman correlation matrix between the dependent and independent variables.

	AGB	G	NDVI	NDRE	GNDVI	$CI_{red\ edge}$	CI_{green}
AGB	1						
G	0,76**	1					
NDVI	0,71**	0,56**	1				
NDRE	0,67**	0,53**	0,97**	1			
GNDVI	**0,77****	0,58**	0,82**	0,76**	1		
$CI_{red\ edge}$	0,67**	0,52**	0,97**	1,00**	0,76**	1	
CI_{green}	**0,77****	0,57**	0,82**	0,75**	1,00**	0,76**	1

****Significant at 99% probability (α =0.01). Where AGB = Aboveground Biomass ($Mg.ha^{-1}$). G = Basal Area (m^2); NDVI = Normalized Difference Vegetation Index; NDRE = Normalized Difference Vegetation Index with red edge band; GNDVI = Normalized Difference Vegetation Index with green band; CIgreen= Chlorophyll Index with green band; CIred edge = Chlorophyll Index with red edge band.**

our analysis using remote sensing variables, it may occur because two indices are retrieved from the same spectral bands.

Tests based on stepwise regression were applied to choose the best indices for estimating AGB in Cerrado vegetation, considering the lowest possible p-value. In this analysis, the GNDVI and NDRE indices were retained for modeling.

Modelling: training of neural networks

The five networks showing the best initial performance among the 500 ANNs trained in this analysis showed satisfactory fit and precision statistics for modeling AGB in cerrado *sensu stricto* vegetation. The Networks 2, 3, and 4 showed the best results, with correlation coefficient (R) values higher than 0,84 and estimation errors below 18%. However, the Neural Network 2 showed the best predictive ability (Table V).

The Network 2 (ANN 2) was selected out of five networks showing the highest performances in this analysis. The ANN 2 was chosen because it showed an acceptable pattern of residual distribution and high accuracy in predicting the AGB in the study area. The error class histogram showed a higher

concentration of residuals in the central classes of the graph and low levels of underestimation and overestimation (Figure 6).

The best network architecture was composed by three layers, described as following: input layer, consisting of two neurons, determined by the predictor variables (GNDVI and NDRE); a hidden layer, which is responsible for processing the data and it is activated by a tangent function, and an output layer, consisting of the variable of interest (AGB) and activated by a tangent function (Figure 7). The hidden layer consisted of eight neurons determined by the Intelligent Problem Solver (IPS) tool (MIGUEL et al. 2018).

Table V. Characteristics and accuracy statistics of the Artificial Neural Networks selected for predicting the Aboveground Biomass in Cerrado sensu stricto vegetation.

ANN	Architecture	Activation Function		Adjustment		Validation	
		Hidden Activation	Output Activation	RMSE (%)	R	RMSE (%)	R
1	MLP 2-9-1	Logistic	Tangent	18,85	0,839	18,07	0,884
2	**MLP 2-8-1**	**Tangent**	**Tangent**	**16,52**	**0,908**	**13,36**	**0,914**
3	MLP 2-8-1	Logistic	Logistic	18,15	0,856	13,76	0,916
4	MLP 2-6-1	Logistic	Identity	17,49	0,844	15,02	0,902
5	MLP 2-9-1	Tangent	Logistic	19,05	0,82	20,38	0,861

Legend: ANN = Artificial Neural Networks; MLP = Multilayer perceptron; RMSE = Root Mean-Square Error; r = correlation between observed and estimated values. Bold text indicates the best ANN that showed an acceptable residual pattern and higher accuracy.

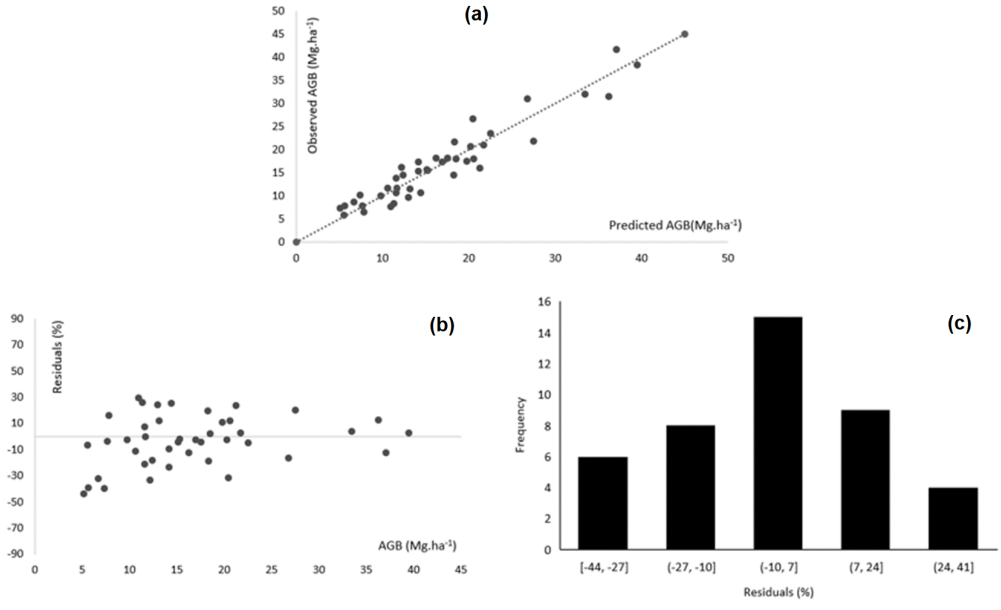

Figure 6. Observed and predicted values (a), distribution of residuals (b), and distribution of error classes (c).

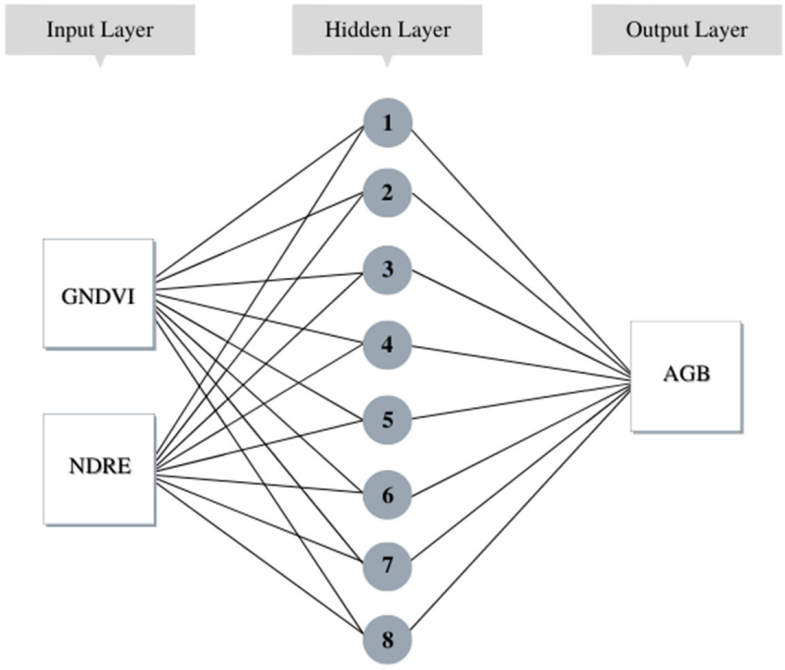

Figure 7. Architecture of the ANN (Artificial Neural Network) selected for aboveground biomass prediction in the DF Cerrado.

Statistical analysis

Based on the results of the Student's t-test, we observed that the selected ANNs showed a p-value greater than α = 0,05 (p = 0,893), which indicates that there were no statically significant differences between the values predicted by the ANN and the values observed in the validation sample plots measured in the forest inventory. The aggregate difference showed a slight overestimation bias of the ANN for biomass prediction (AD% = -0,65), which corroborates the accuracy results observed for training the ANNs.

Based on our selected ANN model, we extrapolated it to predict the AGB for the entire study area covered by native vegetation. We observed that those areas showing highest amount of AGB predominantly occurred within protected areas in the study area, such as National Park and Ecological Stations. Complementarily, small vegetation patches tend to show lower amount of AGB than those larger patches (Figure 8).

Complementarily, we examined the Autocorrelation Function (ACF) plot of the residuals, we checked for significant autocorrelation at any lag residuals. Relevant autocorrelations in the residuals may indicate that the model fails to fully capture the data structure, suggesting the need for a more sophisticated model or the inclusion of additional explanatory variables. Conversely, the absence of significant autocorrelation in the residuals indicates that the model is well-developed and adjusted. In our study, we observed an autocorrelation of -0.13, indicating no significant spatial autocorrelation at the 95% confidence level (Figure 9).

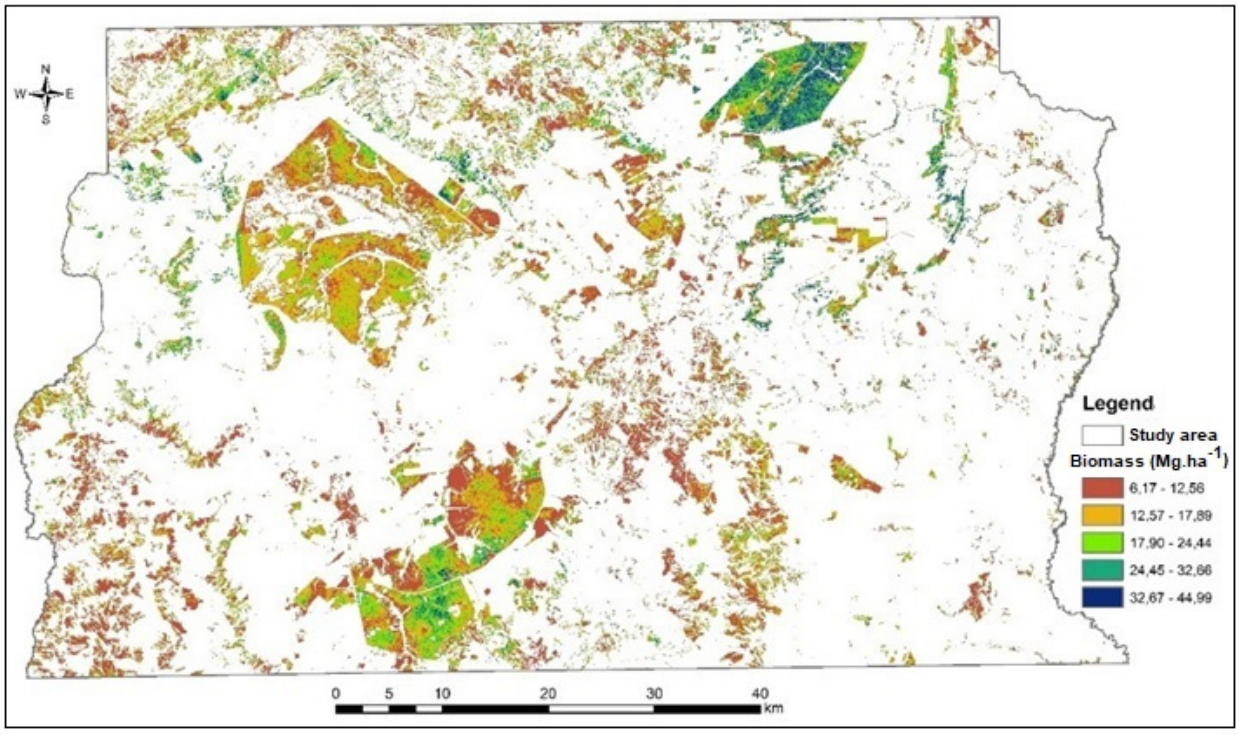

Figure 8. Predicted Above Ground Biomass (AGB) for Cerrado vegetation in the Federal District, Brazil, using a modeling approach based on the Artificial Neural Network and remotely sensed data combined.

DISCUSSION

Forest Inventory

The mean aboveground biomass (AGB) estimated in this study (16.55 ± 8.5 Mg.ha^{-1}) for cerrado vegetation areas closely aligns with the average AGB of 22.9 Mg ha^{-1} reported by Roitman et al. (2018) for the Cerrado biome. Our findings are also consistent with results observed by (BISPO et al. 2020) in the Rio Vermelho watershed, state of Goiás, and by (COSTA et al. 2021) in protected areas within the Cerrado biome, including Serra do Cipó National Park, Chapada dos Veadeiros National Park, Paraopeba National Forest, and Forest of the University of São João Del-Rei. Bispo et al. (2020) and Costa et al. (2021) estimated AGB mean distributions of 18.66 Mg ha^{-1} and 19.72 Mg.ha^{-1}, respectively, in their study areas.

Morandi et al. (2020) estimated an average AGB of 20,4 ± 6,0 Mg ha^{-1} for the central area of the Cerrado biome using data from field plots allocated within protected areas and legal reserves in private properties. Vale & Felfili (2005) estimated an average AGB equal to 12,4 Mg.ha^{-1} in cerrado areas in the Água Limpa farm in the Federal District of Brazil. Oliveira et al. (2019) estimated an average 10,9 Mg.ha^{-1} in a Cerrado area in the municipality of Jaborandi, western Bahia. Based on it, we observed that the estimates of AGB in the Cerrado biome may vary according to the specific characteristics of the study sites, which indicates structural heterogeneity of the Cerrado vegetation.

Spatial variations of ABG occur mainly as a function of physical differences in the environment, such as luminosity, relief, and soil conditions (Fitzgerald et al. 2010, Oliveira et al. 2019), or disturbance regime, such as fires and deforestation (Roitman et al. 2018). Physiographic heterogeneity is reflected

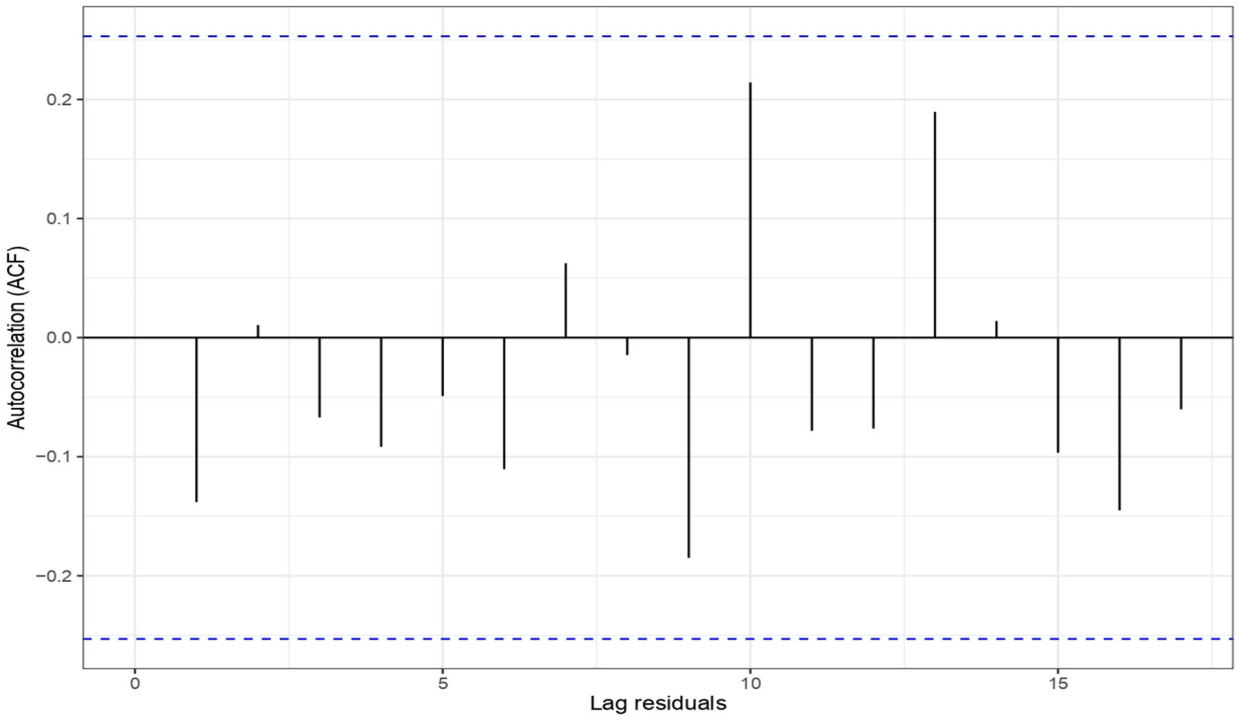

Figure 9. Autocorrelation of AGB (Aboveground Biomass) residuals as a function of NDRE (Normalized Difference Red-Edge index) and GNDVI (Green Normalized Difference Vegetation Index) from the selected ANN (Artificial Neural Network) model.

by the size of individual trees and, consequently, the carbon storage capacity, since tree trunks show the highest aboveground biomass storage capacity (Roquette 2018, Vale & Felfili 2005).

Field measurements in native vegetations are difficult, time and money consuming because of the heterogeneous and great floristic, physiognomic and phenological diversity, which make their biomass quantification a complex task with a little availability of quantitative data (Ferraz et al. 2014). Those data gaps can be improved by conducting regular monitoring of the Cerrado using remotely sensed data as an alternative approach for estimating biomass (Roitman et al. 2018).

Modeling: selection of the independent variables

The GNDVI and CIgreen were identified as the vegetation indices with the strongest correlation with biomass in the study area. The GNDVI is similar to NDVI, however, it is retrieved from remotely sensed data using the green and near infrared spectral bands and is directly related to the proportion of absorbed photosynthetically active radiation, leaf area index (LAI) and biomass (Candiago et al. 2015). On the other hand, the CIgreen, developed by Gitelson et al. (2005), enhances the canopy chlorophyll content and the leaf area index.

However, we observed multicollinearity between GNDVI and CIgreen, suggesting that one index does not substantially contribute additional information beyond the other. This is likely because they are retrieved from the same spectral bands. This observation may explain why the best ANN model to predict biomass in our study area used the GNDVI and NDRE indices as input layers. These indices are retrieved from different spectral bands, green and red edge, respectively, enabling them to capture a wider range of vegetation effects in the prediction model.

In this analysis, we observed a medium correlation among vegetation indices and biomass, which also indicates a moderate model performance to predict AGB, similar to that observed in several researches involving relationships of vegetation indices and variables derived from field measurements, such as biomass, wood volume and trunk circumference (Machado et al. 2017, Miguel et al. 2015, Panda et al. 2010, Zhu & Liu 2015). It is likely that it occurs due to potential atmospheric and soil effects, radiation sources, and inherent aspects of the vegetation architecture on the reflectance of solar radiation detected by the sensors (Carrijo et al. 2020, Gonçalves et al. 2021, Miguel et al. 2015, Ponzoni et al. 2012).

Canopy growth and conformation are irregular and variable in native forests and savannas, resulting in interference of those large tree canopies on the solar reflectance of smaller trees and, consequently, on relationship of field and remotely sensed data-retrieved variables (Machado et al. 2017). Canopy structure, canopy leaf area, and canopy background reflectance are conditioning factors of vegetation reflectance changes in remote sensing images, which will vary according to specific characteristics of each phytophysionomy in each region of interest. In addition to the heterogeneity of individual trees, altitude and relief may affect the relationship of field and remote sensing data. Some studies have shown low correlations between dendrometric variables and vegetation indices in sites of high spatial variations of edaphic attributes (Canavesi et al. 2010, Castillo et al. 2017).

The significant and positive relationships observed between VIs and biomass can be explained by the positive relationship of VIs and radial stem growth and leaf area (Almeida et al. 2015, Babst et al. 2014a, b, Vicente-Serrano et al. 2016). These parameters have a direct relationship with biomass accumulation, carbon sequestration, and, consequently, wood carbon content (Vicente-Serrano et al. 2016). Although vegetation indices cannot directly estimate biomass (Herold et al. 2019), it can be directly related to the amount of photosynthetic material and forest productivity (Mendes et al. 2019).

Modeling: training of neural networks

In our analysis, the training of the artificial neural networks was able to accurately predict the AGB in the Cerrado using the GNDVI index as an addition to the NDRE index effects in the model prediction. Both indices are mathematically derived from NDVI, which has been a robust and widely used vegetation index (Bayma & Sano 2015, Zhu & Liu 2015). The GNDVI is retrieved from the green and near infrared bands, which improves visualization of canopy sanities and the amount of chlorophyll content in the leaves (Gitelson & Merzlyak 1996). The NDRE is retrieved from the red edge band (where solar radiation is strongly affected by chlorophyll pigments) and near infrared band (where solar radiation is mostly affected by the leaf structure) (Kanke et al. 2016).

The accuracy of forest biomass estimates can be assessed by calculating the root mean square error (RMSE) and Pearson's correlation coefficient of estimated and observed values. This method is directly responsible for the quality of the estimates (Congalton 2001). In our analysis, the trained ANNs provided satisfactory fit and accuracy statistics, showing correlation coefficients of $R \geq 0.90$ and estimation errors (RMSE) between 13 and 20%, in which the network 2 showed the best results ($R=0.91$ and $RMSE=13.36$). It indicates that we have reached consistent results given the complexity of the study variables and the vegetational variability observed in the Cerrado biome (Rezende et al. 2006).

The heterogeneity of vegetation structures affects prediction accuracy of AGB when using methodological approaches based on remotely sensed data. It occurs because remote sensing

images are not sensitive to certain vegetation changes. In addition, the saturation of spectral signals captured by optical sensors, the contribution of understory vegetation, soil and shade effects may cause biomass modeling errors (underestimate and overestimate) (Kumar et al. 2016, Lu et al. 2017).

The predictive power of our AGB model is demonstrated by the ratio between observed and predicted values (Figure 6), which indicates a good accuracy of the chosen neural network to estimate the aboveground biomass in the Cerrado biome. Complementarily, the graphical analysis of residuals is crucial to assess the precision statistics since trend errors may not be detected by the previously mentioned statistics. In our case study, the scatter plot of residuals showed an adequate error distribution showing no apparent trends and compact and well-distributed points along the regression line. The frequency of errors indicates that most of the values were concentrated between –10% and 7% for training nets. The analysis of residuals in the histograms can minimize error interpretation based on the occurrence of overlapping points (Campos & Leite 2013).

Our study results demonstrates the efficiency of neural networks in predicting AGB and corroborates with other studies found in the literature, such as (Silva et al. 2019), who observed RMSE ranging between 12% and 17% in estimating volume and AGB in a study site in Montes Claros municipality, state of Minas Gerais. Similarly, Vahedi (2016) estimated RMSE ranging between 8% and 15% in predicting AGB in Iran. Miguel et al. (2015) pointed a standard error of estimate of 8% for biomass estimation in a Cerrado area in Palmas municipality, state of Tocantins.

In this study, we observed similar results to those reported in the literature conducted in other regions of the Cerrado biome, indicating that the artificial intelligence (AI) approach applied in this analysis can be used to estimate biomass in the Brazilian Cerrado. The high accuracy observed in our study can be attributed to the ability of AI approaches to capture the nonlinearity observed in the data, as they can approximate complex functions (Silva et al. 2019), which is an important characteristic for modeling forest parameters, since they often show non-linear behavior (Vieira et al. 2018).

Statistical Analysis

The ANN showing the best results was submitted to Student's t-teste for validation. The t-test results returned a p-value > 0.05 indicating that the null hypothesis of equality between the observed and predicted values must not be rejected, which demonstrates the high estimation accuracy of the model. According to Serpen & Gao (2014), the ANNs have a great capacity to learn and extract patterns from a data set and then generalize and apply them without losing accuracy. The effectiveness and accuracy of the selected ANNs was also confirmed by the aggregate difference (AD%), which showed a small percentage value (-0.65%) of overestimation in the estimate of AGB in the cerrado vegetation of the Federal District.

Our developed ANN modeling approach was effective and accurate to predict AGB using vegetation indices retrieved from remotely sensed data. Our approach is an important contribution and has great potential to be applied for predicting forest biomass in the Cerrado biome, which can support assessment of ecoservices and definition of environmental policies to mitigate global climate changes.

Spatial distribution of the Aboveground Biomass

We observed that small patches of native vegetation closer to urban areas showed lower amount of AGB, ranging between 6 and 13Mg.ha^{-1}. Our results indicates that protected areas in the Federal District were able to provide an important ecosystem service to climate regulation by storing Carbon on the ABG. The Ecological Station of Água Emendadas (ESAE) presented the highest mean value of AGB, showing 26.94 ± 8.41 Mg.ha^{-1}. Jacon et al. (2021) found a similar average of biomass in Cerrado areas (34.5 ± 2.7 Mg.ha^{-1}) using different hyperspectral metrics and machine learning models. The highest average of biomass in the ESAE can be explained by the presence of dense Cerrado vegetation and more than 50% vegetation cover.

In addition to the inherent heterogeneity of vegetation, a possible cause of the lower AGB values in those small patches is related to higher habitat fragmentation. The discontinuity of native vegetation cover into isolated fragments results in population separation, reduced habitat quality, and increased edge area. The increase in the proportion of edge related to the area that makes the vegetation fragments more susceptible to anthropic disturbances, such as fire, hunting, exotic species, logging and invasive species (Coelho et al. 2020, Gamarra et al. 2021)

CONCLUSIONS

Our modeling approach for predicting AGB biomass in Cerrado vegetation based on ANN and vegetation indices retrieved from RapidEye imagery showed high accuracy and effectiveness, with a predictive power of 90% and RMSE less than 17%. It is a promising methodological approach to be broadly applied in the Cerrado biome. The Green Normalized Difference Index using the green and near infrared bands (GNDVI) and the Normalized Difference Red-Edge Index using the red edge and near infrared bands (NDRE) were able to provide the best model fits for predicting the AGB in our study area.

Complementarily, the Multilayer Perceptron artificial neural networks, using tangential activation functions and the BFGS training algorithm, were accurate and efficient in associating vegetation indices in the estimation of AGB in the Cerrado vegetation.

Most of knowledge gaps related to biomass estimation in the Cerrado biome are due to the difficulty and cost of field measurements conducted by forest inventories. Therefore, our study results are useful for civil society and can improve definition of strategic plans of conservation of native vegetation in large extents biomes like the Brazilian Cerrado.

Our analysis was primarily focused on constructing and evaluating artificial neural network (ANN) models to predict aboveground biomass in the shrublands of the Cerrado biome. However, we acknowledge the significant potential of alternative approaches involving machine learning algorithms, such as Random Forest and Support Vector Machine, in combination with hyperspectral remotely sensed data, Synthetic-Aperture Radar (SAR), and/or LiDAR (Light Detection and Ranging) sensors. These alternatives should be further investigated in future research within this field.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, Finance Code 001, and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grants n. 401892/2021-2 e n. 311155/2020-0. Special thanks to the Brazilian Forest Service, Dr. Humberto de Mesquita Júnior and Dr. José Humberto Chaves for providing us valuable field data from the National Forest Inventory conducted in the Federal District of Brazil. We also sincerely thank the anonymous reviewers and members of the editorial team for their comments and contributions. All primary public domain satellite remote sensing data are available from the corresponding author. All derived products are available in standard GIS data layer formats from the corresponding author. Data analysis models and overlay protocols are described in the article or by inquiry of the corresponding author. Validation datasets are available from the corresponding author.

REFERENCES

ALMEIDA AQ, RIBEIRO A, DELGADO RC, RODY YP, OLIVEIRA AS & LEITE FP. 2015. Índice de área foliar de eucalyptus estimado por índices de vegetação utilizando imagens TM - Landsat 5. Floresta e Ambiente 22(3): 368-376. DOI 10.1590/2179-8087.103414.

ANDERSON LO, MALHI Y, LADLE RJ, ARAGÃO LEOC, SHIMABUKURO Y, PHILLIPS OL, BAKER T, COSTA ACL, ESPEJO JS & HIGUCHI N. 2009. Influence of landscape heterogeneity on spatial patterns of wood productivity, wood specific density and above ground biomass in Amazonia. Biogeosci 6: 1883-1902.

BABST F, BOURIAUD O, ALEXANDER R, TROUET V & FRANK D. 2014a. Toward consistent measurements of carbon accumulation: A multi-site assessment of biomass and basal area increment across Europe. Dendrochronol 32(2): 153-161. DOI 10.1016/j.dendro.2014.01.002.

BABST F ET AL. 2014b. Above-ground woody carbon sequestration measured from tree rings is coherent with net ecosystem productivity at five eddy-covariance sites. New Phytol 201(4): 1289-1303. DOI 10.1111/nph.12589.

BANNARI A, MORIN D, BONN F & HUETE AR. 1995. A review of vegetation indices. Remote Sens Rev 12(1-2): 95-120. DOI 10.1080/02757259509532298.

BARBOSA JM, BROADBENT EN & BITENCOURT MD. 2014. Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. Int J For Res: DOI 10.1155/2014/715796.

BAYMA AP & SANO EE. 2015. Séries temporais de índices de vegetação (NDVI e EVI) do sensor Modis para detecção de desmatamentos no bioma Cerrado. Boletim de Ciências Geodésicas 21(4): 797-813. DOI 10.1590/S1982-21702015000400047.

BINOTI MLMS, LEITE HG, BINOTI BDH & GLERIANI JM. 2015. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. Cerne 21(1): 97-105. DOI 10.1590/01047760201521011153.

BISPO PC ET AL. 2020. Estimating the Above Ground Biomass of Brazilian Savanna using multi-sensor approach. EGUGA, p. 20531. DOI 10.5194/egusphere-egu2020-20531.

BROYDEN CG. 1970. The convergence of a class of double-rank minimization algorithms 1. General Considerations. IMA J Appl Math 3(1): 76-90. DOI 10.1093/imamat/6.1.76.

CAMPOS JCC & LEITE HG. 2013. Mensuração Florestal: pergunta e respostas. Editora UFV, 5ª ed., Viçosa, Brasil, 605 p.

CANAVESI V, PONZONI FJ & VALERIANO MM. 2010. Stand volumes estimate Eucalyptus spp. plantations in forests using hyperspectral and topographic data. Revista Arvore 34(3): 539-549. DOI 10.1590/s0100-67622010000300018.

CANDIAGO S, REMONDINO F, DE GIGLIO M, DUBBINI M & GATTELLI M. 2015. Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. Remote Sensing 7(4): 4026-4047. DOI 10.3390/RS70404026.

CARRIJO JVN, MIGUEL EP, DO VALE AT, MATRICARDI EAT, MONTEIRO TC, REZENDE AV & INKOTTE J. 2020. Artificial intelligence associated with satellite data in predicting energy potential in the Brazilian savanna woodland area. iForest - Biogeosci Forestr 13(1): 48. DOI 10.3832/IFOR3209-012.

CASSOL HLG, SALDANHA DL & KUPLICH TM. 2014. Inventário de carbono em fragmento de Floresta Ombrófila Mista por detecção remota. Floresta 44(4): 697. DOI 10.5380/rf.v44i4.33014.

CASTILLO JAA, APAN AA, MARASENI TN & SALMO SG. 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. ISPRS J Photogramm Remote Sens 134: 70-85. DOI 10.1016/J.ISPRSJPRS.2017.10.016.

CHAVE J ET AL. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145: 87-99.

CHEN L, WANG Y, REN C, ZHANG B & WANG Z. 2019. Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data. Remote Sens 11(4): 414. DOI 10.3390/RS11040414.

COELHO AJP, MAGNAGO LFS, MATOS FA R, MOTA NM, DINIZ ÉS & MEIRA-NETO JAA. 2020. Effects of anthropogenic disturbances on biodiversity and biomass stock of Cerrado, the Brazilian savanna. Biodivers Conserv 29(11-12): 3151-3168. DOI 10.1007/S10531-020-02013-6/TABLES/1.

CONGALTON RG. 2001. Accuracy assessment and validation of remotely sensed and other spatial information. Int J Wildl Fire 10(4): 321-328. DOI 10.1071/WF01031.

COSTA MBT ET AL. 2021. Beyond trees: Mapping total aboveground biomass density in the Brazilian savanna using high-density UAV-lidar data. Forest Ecol Manag 491: 119155. https://doi.org/10.1016/J.FORECO.2021.119155.

DUNCANSON L, NEUENSCHWANDER A, HANCOCK S, THOMAS N, FATOYINBO T, SIMARD M, SILVA CA, ARMSTON J, LUTHCKE SB & HOFTON M. 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California.Remote Sens Environ 242: 111779. DOI 10.1016/j.rse.2020.111779.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. 2018. Sistema Brasileiro de Classificação de Solos. Ed. Embrapa Solos, 5ª ed., Brasília. DF, 355 p.

ERCANLI İ, GÜNLÜ A, SENYURT M & KELES S. 2018. Artificial neural network models predicting the leaf area index: A case study in pure even-aged crimean pine forests from Turkey. Forest Ecosy 5(1): 1-12. DOI 10.1186/S40663-018-0149-8.

FERRAZ AS, SOARES VP, SOARES CPB, RIBEIRO CAAS, BINOTI DHB & LEITE HG. 2014. Estimativa do estoque de biomassa em um fragmento florestal usando imagens orbitais. Floresta e Ambiente 21(3): 286-296. DOI 10.1590/2179-8087.052213.

FERREIRA LG, YOSHIOKA H, HUETE A & SANO EE. 2002. Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: An analysis within the Large-Scale Biosphere-Atmosphere Experiment in Amazônia (LBA). Remote Sens Environ 87(4): 534-550. DOI 10.1016/J.RSE.2002.09.003.

FITZGERALD G, RODRIGUEZ D & O'LEARY G. 2010. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). Field Crops Res116(3): 318-324. DOI 10.1016/J.FCR.2010.01.010.

FLETCHER RA. 1970. new approach to variable metric algorithms. The Computer Journal 13(3): 317-322. DOI 10.1093/comjnl/13.3.317.

GASCÓN LH, CECCHERINI G, HARO FJG, AVITABILE V & EVA H. 2019. The potential of high resolution (5 m) RapidEye optical data to estimate above ground biomass at the national level over Tanzania. Forests 10(107): DOI: 10.3390/f10020107.

GAMARRA RM, HIGA LT, GAMARRA MC, CARRIJO MG, MOTA JS, NOTARI F, RODRIGUES AG, DALMAS FB & PARANHOS FILHO AC. 2021. Fragmentation of vegetation in protected area in the cerrado region. Res Soc Develop 10(7): 12. DOI 0000-0002-0049-0009.

GHOSH SM & BEHERA MD. 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. Appl Geogr 96: 29-40. DOI 10.1016/j.apgeog.2018.05.011.

GITELSON AA & MERZLYAK MN. 1996. Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. J Plant Physiol 148(3-4): 494-500. DOI 10.1016/S0176-1617(96)80284-7.

GITELSON AA, VIÑA A, CIGANDA V, RUNDQUIST DC & ARKEBAUER TJ. 2005. Remote estimation of canopy chlorophyll content in crops. Geophys Res Lett 32(8): 1-4. DOI 10.1029/2005GL022688.

GOLDFARB DA. 1970. family of variable-metric methods derived by variational means. Math Comput 24(109): 23.

GONÇALVES FC, MIGUEL EP, MATRICARDI EAT, EMMERT F & SANTANA CC. 2021. Artificial intelligence associated with Sentinel-2 data in predicting commercial volume in Brazilian Amazon Forest. J Appl Remote Sens 15(4): 044511. DOI 10.1117/1.JRS.15.044511.

GORUNESCU F & BELCIUG S. 2016. Boosting backpropagation algorithm by stimulus-sampling: Application in computer-aided medical diagnosis. J Biomed Inf 63: 74-81. DOI 10.1016/J.JBI.2016.08.004.

GUERINI FILHO M, KUPLICH TM & QUADROS FLFD. 2019. Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. Int J Remote Sens 41(8): 2861-2876. DOI 10.1080/01431161.2019.1697004.

GUERROUT EH, AIT-AOUDIA S, MICHELUCCI D & MAHIOU R. 2018. Hidden Markov random field model and Broyden-Fletcher-Goldfarb-Shanno algorithm for brain image segmentation. J Exper Theoret Artif Intel 30(3): 415-427. DOI 10.1080/0952813X.2017.1409280.

GÜNLÜ A, ERCANLI İ, ŞENYURT M & KELEŞ S. 2019. Estimation of some stand parameters from textural features from WorldView-2 satellite image using the artificial neural network and multiple regression methods: a case study from Turkey. Geocart Int 36(8): 918-935. DOI 10.1080/10106049.2019.1629644.

HENRY M, TITTONELL P, MANLAY RJ, BERNOUX M, ALBRECHT A & VANLAUWE B. 2009. Biodiversity, carbon stocks and sequestration potential in aboveground biomass in smallholder farming systems of western Kenya.

Agric Ecosyst Environ 129: 238-252. DOI 10.1016/j.agee.2008.09.006.

HEROLD M ET AL. 2019. The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy. Survey Geophys 40(4): 757-778. DOI 10.1007/s10712-019-09510-6.

HUETE A, DIDAN K, MIURA T, RODRIGUEZ EP, GAO X & FERREIRA LG. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 83(1-2): 195-213. DOI 10.1016/S0034-4257(02)00096-2.

HYDE P, DUBAYAH R, WALKER W, BLAIR JB, HOFTON M & HUNSAKER C. 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sens Environ 102(1-2): 63-73. DOI 10.1016/j.rse.2006.01.021.

JACON AD, GALVAO LS, DALAGNOL R & DOS SANTOS JR. 2021. Aboveground biomass estimates over Brazilian savannas using hyperspectral metrics and machine learning models: experiences with Hyperion/EO-1. GIScience & Remote Sens 58(7): 1112-1129. DOI: 10.1080/15481603.2021.1969630.

KANKE Y, TUBAÑA B, DALEN M & HARRELL D. 2016. Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. Precis Agric 17(5): 507-530. DOI 10.1007/S11119-016-9433-1/TABLES/7.

KEMP F. 2002. Modern Applied Statistics with S. 4th ed. In: Venables B & Ripley B (Eds), New York. Springer, xii + 496 p.

KLINK CA & MACHADO RB. 2005. A conservação do Cerrado brasileiro. Megadiversidade 1(1): 147-155.

KROSS A, MCNAIRN H, LAPEN D, SUNOHARA M & CHAMPAGNE C. 2015. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. Int J Appl Earth Obs Geoinf 34(1): 235-248. DOI 10.1016/J.JAG.2014.08.002.

KUMAR KK, NAGAI M, WITAYANGKURN A, KRITIYUTANANT K & NAKAMURA S. 2016. Above Ground Biomass Assessment from Combined Optical and SAR Remote Sensing Data in Surat Thani Province, Thailand. J Geograph Inf Syst 8: 506-516. DOI 10.4236/jgis.2016.84042.

KUMAR L & MUTANGA O. 2017. Remote sensing of above-ground biomass. Remote Sens 9(9): 1-8. DOI 10.3390/rs9090935.

KUMAR L, SINHA P, TAYLOR S & ALQURASHI AF. 2005. Review of the use of remote sensing for biomass estimation to support renewable energy generation. J Appl Remote Sens 9(1): 097696. DOI 10.1117/1.JRS.9.097696.

LI Y, LI C, LI M & LIU Z. 2019. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. Forests 10(12): 1073. DOI 10.3390/F10121073.

LORRENTZ P. 2015. Artificial neural systems: Principle and practice. Bentham Science Publishers, 249 p.

LU X, ZHENG G, MILLER C & ALVARADO E. 2017. Combining multi-source remotely sensed data and a process-based model for forest abovground biomass updating. Sensors 17(2062): 1-21. DOI 10.3390/s17092062.

LUZ LR, GIONGO V, SANTOS AM, LOPES RJC & LIMA JÚNIOR C. 2022. Biomass and vegetation index by remote sensing in different Caatinga forest areas. Forestr Sci 52(2). DOI 10.1590/0103-8478cr20201104.

MACHADO IES, SANTOS MM, GIONGO M, CARVALHO EV & NETO EG. 2017. Modelos para estimativa de variáveis florestais com a utilização de imagens multiespectrais. Pesquisa Florestal Brasileira 37(90): 171-181. DOI 10.4336/2017.pfb.37.90.1380.

MAPBIOMAS. 2024. Cerrado Project - Collection 8.0 of annual land cover and land use maps. 2024. Available at https://mapbiomas.org/estatisticas. Last accessed on Feb 9th, 2024.

MENDES FS. 2019. Monitoring forest fragmentation and carbon storage in the Cerrado biome of Mato Grosso using optical and SAR satellite images,143 p. George-August University School of Science (GAUSS).

MENDES FS, BARON D, GEROLD G, LIESENBERG V & ERASMI S. 2019. Optical and SAR Remote Sensing Synergism for Mapping Vegetation Types in the Endangered Cerrado/Amazon Ecotone of Nova Mutum - Mato Grosso. Remote Sens 11(10): 1161. DOI: 10.3390/rs11101161.

MENESES PR, ALMEIDA T & BAPTISTA GMM. 2019. Reflectância dos materiais terrestres - Análise e Interpretação. Brasília, DF, Ed. Oficina de textos, Ed. 1, 334 p.

MIGUEL EP, MELO RR, SERENINI JUNIOR L & MENEZZI CH. 2018. Using artificial neural networks in estimating wood resistance. Maderas. Ciencia y Tecnología 20(4): 531-543. DOI 10.4067/S0718-221X2018005004101.

MIGUEL EP, REZENDE AV, LEAL FA, MATRICARDI EAT, VALE AT & PEREIRA RS. 2015. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. Pesquisa Agropecuaria Brasileira 50(9): 829-839. DOI: 10.1590/S0100-204X2015000900012.

MORANDI OS ET AL. 2020. Tree diversity and above-ground biomass in the South America Cerrado biome and their conservation implications. Biodivers Conserv 29(5): 1519-1536. DOI 10.1007/S10531-018-1589-8.

MOVCHAN D, BILOUS A, YELISTRATOVA L, APOSTOLOV A & HODOROVSKY A. 2023. Application of various approaches of multispectral and radar data fusion for modelling of aboveground forest biomass. Folia Forestalia Polonica 65(2): 55-67. DOI: 10.2478/ffp-2023-000.

NETO CDG. 2018. Identificação de fitofisionomias de Cerrado no Parque Nacional de Brasília utilizando Random Forest aplicado a imagens de alta e média resoluções espaciais, 212 p. Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, SP. Disponível em: http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2018/10.03.18.55/doc/publicacao.pdf. Last accessed on: 19 May. 2024.

OLIVEIRA CP, FRANCELINO MR, PAULA MD, LELES PSS & ANDRADE FC. 2019. Comparação de modelos estatísticos para estimativa da biomassa de árvores, e estimativa do estoque de carbono acima do solo em Cerrado. Ciência Florestal 29(1): 255. DOI: 10.5902/1980509827065.

PANDA SS, AMES DP & PANIGRAHI S. 2010. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. Remote Sens 2(3): 673-696. DOI 10.3390/RS2030673.

PONZONI FJ, SHIMABUKURO YE & KUPLICH TM. 2012. Sensoriamento remoto da vegetação. Ed. Oficinas de Textos, São Paulo - SP, ed. 2, 160 p.

REZENDE A, VALE AT, SANQUETTA CR, FIGUEIREDO FILHO A & FELFILI JM. 2006. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. Scientia Forestalis 71: 65-76.

ROITMAN I ET AL. 2018. Optimizing biomass estimates of savanna woodland at different spatial scales in the Brazilian Cerrado: Re-evaluating allometric equations and environmental influences. Plos One 13(8): e0196742. DOI 10.1371/JOURNAL.PONE.0196742.

ROQUETTE JG. 2018. Distribution of biomass in Cerrado and its importance for Carbon storage. Revista Ciência Florestal 28: 121. DOI: 10.5902/1980509833354.

ROUSE JW, HAAS RH, SCHELL JA, DEERING DW & HARLAN JC. 1974. Monitoring the vernal advancement and retrogradation (greenwave effect of natural vegetation). Texas A&M Univ. College Station, TX, United States, No. E75-10354.

RANA P, POPESCU S, TOLVANEN A, GAUTAM B, SRINIVASAN S & TOKOLA T. 2023. Estimation of tropical forest aboveground biomass in Nepal using multiple remotely sensed data and deep learning. Int J Remote Sens 44(17). DOI: 10.1080/01431161.2023.2240508.

SERPEN G & GAO Z. 2014. Complexity Analysis of Multilayer Perceptron Neural Network Embedded into a Wireless Sensor Network. Procedia Comput Sci 36: 192-197. DOI: 10.1016/J.PROCS.2014.09.078.

ROSA CMM. 2016. Inventário Florestal Nacional: principais resultados, Serviço Florestal Brasileiro - SFB, Brasília, DF, Série Relatório Técnico, 51 p.

SHANNO DF. 1970. Conditioning of quasi-Newton methods for function minimization. Math Comput 24(111): 647-656. DOI: 10.2307/2004840.

SILVA FAM, ASSAD SM & EVANGELISTA BA. 2008. Caracterização climática do bioma Cerrado. Cerrado ecologia e flora. Brasília, DF, Embrapa Informação Tecnológica, p. 69-88.

SILVA JP ET AL. 2019. Computational techniques applied to volume and biomass estimation of trees in Brazilian savanna. J Environ Manag 249: 109368. DOI 10.1016/J.JENVMAN.2019.109368.

SILVA JPM ET AL. 2019. Computational techniques applied to volume and biomass estimation of trees in Brazilian savanna. J Environ Manag 249. DOI: 10.1016/j.jenvman.2019.109368.

STATSOFT I. 2007. Statistica: Data analysis software system, version 8.0.

SYDOW JD, SANQUETTA CR, CORTE APD, SANQUETTA MNI & FILHO AF. 2017. Comparação de métodos e processos de amostragem para inventário em Floresta Ombrófila Mista. BIOFIX Sci J 2(1): 60-68. DOI 10.5380/BIOFIX.V2I1.50761.

VAHEDI AA. 2016. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. Biomass Bioenerg 88: 66-76. DOI 10.1016/J.BIOMBIOE.2016.03.020.

VALE AT & FELFILI JM. 2005. Dry biomass distribution in a cerrado sensu stricto site in Brazil central. Revista Árvore 29(5): 661-669. DOI 10.1590/s0100-67622005000500001.

VALE AT, MIGUEL EP, MOREIRA ACO, LIMA CM, ORELLANA BBMA, FORTES MM, MACHADO MPO, GONÇALEZ JC & MARTINS IS. 2017. Artificial neural networks in predicting energy density of Bambusa vulgaris in Brazil. Afric J Agric Res 12(10): 856-862. DOI 10.5897/AJAR2016.12083.

VICENTE-SERRANO SM, CAMARERO JJ, OLANO JM, MARTÍN-HERNÁNDEZ N, PEÑA-GALLARDO M, TOMÁS-BURGUERA M, GAZOL A, AZORIN-MOLINA C, BHUYAN U & EL KENAWY A. 2016. Diverse relationships between forest growth and the Normalized Difference Vegetation Index at a global scale. Remote Sens Environ 187: 14-29. DOI 10.1016/J.RSE.2016.10.001.

VIEIRA GC, MENDONÇA AR, SILVA GF, ZANETTI SS, SILVA MM & SANTOS AR. 2018. Prognoses of diameter and height of trees

of eucalyptus using artificial intelligence. Sci Total Environ 619-620: 1473-1481. DOI 10.1016/J.SCITOTENV.2017.11.138.

VIÑA A, GITELSON AA, NGUY-ROBERTSON AL & PENG Y. 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. Remote Sens Environ 115(12): 3468-3478, DOI 10.1016/J.RSE.2011.08.010.

YANG S, FENG Q, LIANG T, LIU B, ZHANG W & XIE H. 2018. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. Remote Sens Environ 204: 448-455. DOI 10.1016/j.rse.2017.10.011.

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 Sens 11(12): 1437. DOI 10.3390/RS11121437.

YU Y & SAATCHI S. 2016. Sensitivity of L-Band SAR Backscatter to Aboveground Biomass of Global Forests. Remote Sens 8(6): 522. DOI 10.3390/rs8060522.

ZHU X & LIU D. 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS J Photogram Remote Sens 102: 222-231. DOI 10.1016/J.ISPRSJPRS.2014.08.014.

ZOLKOS SG, GOETZ SJ, DUBAYAH RA. 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens Environ 128: 289-298. DOI 10.1016/j.rse.2012.10.017.

ZUCCHINI W. An Introduction to Model Selection. J Math Psychol 44(1): 41-61. DOI 10.1006/JMPS.1999.1276.

How to cite

OLIVEIRA PLG, MATRICARDI EAT, MIGUEL EP, MARIMON JÚNIOR BH & REZENDE AV. 2024. Artificial Neural Network and Remote Sensing combined to predict the Aboveground Biomass in the Cerrado biome. An Acad Bras Cienc 96: e20211041. DOI 10.1590/0001-3765202420221041.

Manuscript received on September 16, 2023;
accepted for publication on April 5, 2024

PAULA L.G. OLIVEIRA[1]
https://orcid.org/0000-0003-0173-3399

ERALDO A.T. MATRICARDI[1]
https://orcid.org/0000-0002-5323-6100

EDER P. MIGUEL[1]
https://orcid.org/0000-0001-6259-4594

BEN HUR MARIMON JÚNIOR[2]
https://orcid.org/0000-0002-6359-6281

ALBA VALÉRIA REZENDE[1]
https://orcid.org/0000-0002-1167-9798

[1]University of Brasilia - UnB, College of Technology, Forestry Department, Campus Darcy Ribeiro, 70910-000 Brasília, DF, Brazil

[2]Mato Grosso State University - UNEMAT, College of Agrarian, Biological and Applied Social Sciences, Campus Nova Xavantina, Parque Municipal Mário Viana, 78690-000 Nova Xavantina, MT, Brazil

Correspondence to: **Eraldo Aparecido Trondoli Matricardi**
E-mail: ematricardi@gmail.com

Author Contributions

P.L.G.O., E.A.T.M, and E.P.M designed the study and contributed to the manuscript; P.L.G.O., E.A.T.M, B.H.M., A.V.R., and E.P.M led the writing and editing of the manuscript. P.L.G.O., E.A.T.M, and A.V.R. conducted the remote sensing and field data analysis; P.L.G.O., E.A.T.M, E.P.M, and A.V.R. conducted the data analysis; P.L.G.O. and E.P.M conducted the ANN modeling and analysis; P.L.G.O and E.A.T.M. produced maps of vegetation indices and biomass; all authors contributed and provided comments on the manuscript.