iForest
Biogeosciences and Forestry

Research Article
doi: 10.3832/ifor3209-012
vol. 13, pp. 48-55

Artificial intelligence associated with satellite data in predicting energy potential in the Brazilian savanna woodland area

João Victor Nobre Carrijo [(1)],
Eder Pereira Miguel [(1)],
Ailton Teixeira Do Vale [(1)],
Eraldo Aparecido Trondoli Matricardi [(1)],
Thiago Campos Monteiro [(2)],
Alba Valéria Rezende [(1)],
Jonas Inkotte [(1)]

The use of artificial intelligence to generate information of the savanna's energy capacity may support sustainable management of those areas. We assessed the efficacy of artificial neural networks (ANNs) combined with satellite data to estimate the energy potential (*Pe*) for *cerradão*, a dense savannah-like vegetation type in Brazil. We conducted a forest inventory for measuring dendrometric variables and sampling woody materials and barks in a *cerradão* area in the state of Tocantins, Brazil. The *Pe* of *cerradão* biomass was estimated based on the observed higher calorific power and drier biomass values. Six vegetation indices were retrieved from a RapidEye image and tested for correlation to choose the optimum vegetation index for biomass modeling. The basal area and the Normalized Difference Vegetation Index were used as predictors in the *Pe* modeling. We estimated an average of 19.234 ± 0.411 GJ ton^{-1} and 19.878 ± 1.090 GJ ton^{-1} for higher heating values of the wood species and barks, respectively, and an average *Pe* of 1022.660 GJ ha^{-1}. The best ANN showed an error of 11.3% by using a structure of two, eight, and one neurons in the input layer, in the hidden layer, and in the output layer, respectively, as well as activation functions of the tangential and sigmoidal types. The validation tests showed no significant difference between the observed and ANN-predicted values. Based on our results, we concluded that *Pe* can be efficiently predicted by combining ANNs and remotely sensed data, which ultimately is a promising tool for forest sustainable management of the *cerrado* ecosystems.

Keywords: Artificial Neural Networks, *Cerrado*, Higher Heating Value, Biomass, Modelling, Forestry

Introduction

The Brazilian savanna (locally known as *Cerrado*) encompasses approximately 24% of Brazilian territory. It is the second largest biome in South America, and is recognized as the savanna with the highest species-diversity on the planet. Its vegetation is characterized by a mosaic of different physiognomies: savannas, fields, and forests, which include the savanna woodlands (also known as *cerradão* – Miguel et al. 2016).

Cerradão has a complex physiognomy and presents species from other formations and physiognomies of the *Cerrado*, being highly associated with regions of interfluvial flat areas (Solórzano et al. 2012). It is a great challenge to acquire dendrometric information in *Cerradão* region due to its complexity and heterogeneity of vegetational structure; indeed, data collection aimed to assess productive aspects of the *Cerrado* is a time and money consuming activity, making this research topic rather scarce in the literature (Miguel et al. 2017).

The *Cerrado* vegetation has an important timber potential to provide raw material for both sawmills and energy purposes (López et al. 2013, Lima et al. 2016). Several studies have emphasized the productive potential of forest areas commonly focusing on the estimation of above-ground biomass and stored wood volume (Miguel et al. 2015, Benítez et al. 2016, Dalponte et al. 2018); however, more research efforts are required to assess the energy potential of those areas for economic and environmental purposes.

Energy density represents the amount of energy to be released after the complete combustion of a given volume of biomass fuel (Protásio et al. 2015). It is a very important parameter in species selection for energy production, since it is directly related to the higher heating value – a mostly important variable for wood selection (Araújo et al. 2018). Additionally, energy density is an indicator of energy potential (*Pe*) per area unit.

Wood and energy production data are essential for improving sustainable management and conservation policies and planning in the *Cerrado* region. This information may be used, for example, in the definition of conservation areas like legal reserves by determining poorly productive areas with relevant environmental and conservative values. However, to ensure the implementation and improvement of such practices, accurate estimates of productive and energetic potential are needed (López et al. 2013).

(1) Department of Forestry, Faculty of Technology, University of Brasília, Brasília, 70910-900 (Brazil); (2) Department of Forestry and Forest Technology, Federal University of Paraná, Av. Prefeito Lothário Meissner, 632 - Jardim Botânico, Curitiba, 80210-170 (Brazil)

@ João Victor Nobre Carrijo (joao.ncarrijo@gmail.com)

Received: Aug 05, 2019 - Accepted: Nov 26, 2019

Citation: Carrijo JVN, Miguel EP, Teixeira Do Vale A, 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 13: 48-55. - doi: 10.3832/ifor3209-012 [online 2020-02-05]

Communicated by: Carlotta Ferrara

Remotely sensed data and geoprocessing techniques have been used to provide information of land use and land cover changes with good accuracy that may decrease time and financial costs of data collection in tropical regions. More recently, the advances in remote sensing and geoprocessing technology allowed to better characterize forest structure, biomass, stored carbon, and other variables (Groot et al. 2015). Several studies using remotely sensed data showed satisfactory results in the indirect estimation of the above-ground biomass in several types of forest (Miguel et al. 2015, Benítez et al. 2016, Dalponte et al. 2018).

Associating remote sensing data with artificial neural networks (ANNs) may result in increasing the predictive power and correcting the estimation errors of these techniques. The ANNs can be defined as parallel distributed processors composed of simple processing units, which present a natural capacity of learning and storing experimental knowledge, thus making their model simulation similar to the human brain (Haykin 2001).

The use of ANNs to predict above-ground biomass has shown a higher accuracy compared to the traditional regression models (Miguel et al. 2015). Studies using remotely sensed data and artificial intelligence tools such as ANNs to predict energy density in tropical regions are fairly uncommon. This study intended to assess the efficiency of ANNs associated with remote sensing data in estimating the timber energy potential for a *Cerradão* in the central region of Brazil.

Materials and methods

Study area

This study was based on a dataset from a forest inventory conducted in 2012 in a study site of 10.15 hectares of *cerradão* (savanna wood land) located within the Lajeado State Park in the state of Tocantins, Brazil (Fig. 1).

According to Köppen classification, the study region is characterized by the Aw climate type (Alvares et al. 2014), with a wet season of an average monthly rainfall of 250 mm between December and February, and a dry season between May and September. The dryer peak period occurs between June and August. The study site is characterized predominantly by deep or very deep alic dystrophic Dark-Red Latosols, with the presence of horizons A, B, and C. The vegetation is composed by 82 tree species of 34 different families.

Data collection

The forest inventory was carried out in eight transects of 20 m width and variable lengths. The transects were spaced 60 m each other. Each transect was subdivided into sample plots of 20 × 20 m (400 m²) dimensions. A total of fifty-four sample plots were measured in a sampled area of 2.16 ha.

All living and dead standing trees showing diameters at breast height (DBH) equal to or greater than 5 cm were identified and their DBH and total height (Ht) were measured within each sample plot. The DBH of each tree was measured using a diameter caliper – these measure and tool were chosen due to their accuracy considering the windingness of trees' stems –, while the Ht was measured using a telescopic ruler of 15 m. The heights of trees taller than 15 m height were eye-estimated by an experienced forester and the DBH were estimated based on a mean value of two perpendicular DBH measures. Trees showing trunk bifurcations below the DBH height were separately measured for the diameter, height, and cross-sectional area.

Wood sampling

A total of 80 individual trees belonging to 34 of the 82 identified tree species in the study area were cut down and used in the lab tests. The reduced number of species in the sampled transects was due to their sparse distribution and to the environmental law that prohibits cutting individuals of protected tree species. The number of individuals of each tree species to be cut was estimated based on the forest inventory. Approximately 3% of trees showing DBH greater than 5 cm and at least one sampled individual of each tree species were cut down and used in the lab tests under a legal permit issued by the Environmental Agency of the State of Tocantins (NATURATINS). The DBH of cut trees varied from 5.3 cm to 49 cm.

Three 5-cm thick, disk-shaped samples of each cut tree were collected from different longitudinal positions of each trunk (base, middle and top), as previously described (Oliveira et al. 2018), to better represent wood properties. Additionally, different branch diameters (thick branch: diameter > 10 cm; medium branch: diameter 3-10 cm; thin branch: diameter < 3 cm) were separated from the bark portion and individually analyzed in the lab.

The 34 tree species that were sampled (cut down) in the field represented approximately 95% of the total individuals sampled in the forest inventory, which was considered a significant sample size reflecting the quantitative characteristic of the tree community (Pakeman & Quested 2007). It is worth to point out that those tree species represent approximately 83% of the Importance Value Index (IVI) in the study area.

Higher heating value

The Higher Heating Value (HHV) of the bark of each tree species sampled was estimated according to the Brazilian regulation NBR 8633. The bark samples were chopped, ground, and then classified using 60 mesh sieves, and the retained fragments were used for composite sampling of each tree species. Those composite sam-

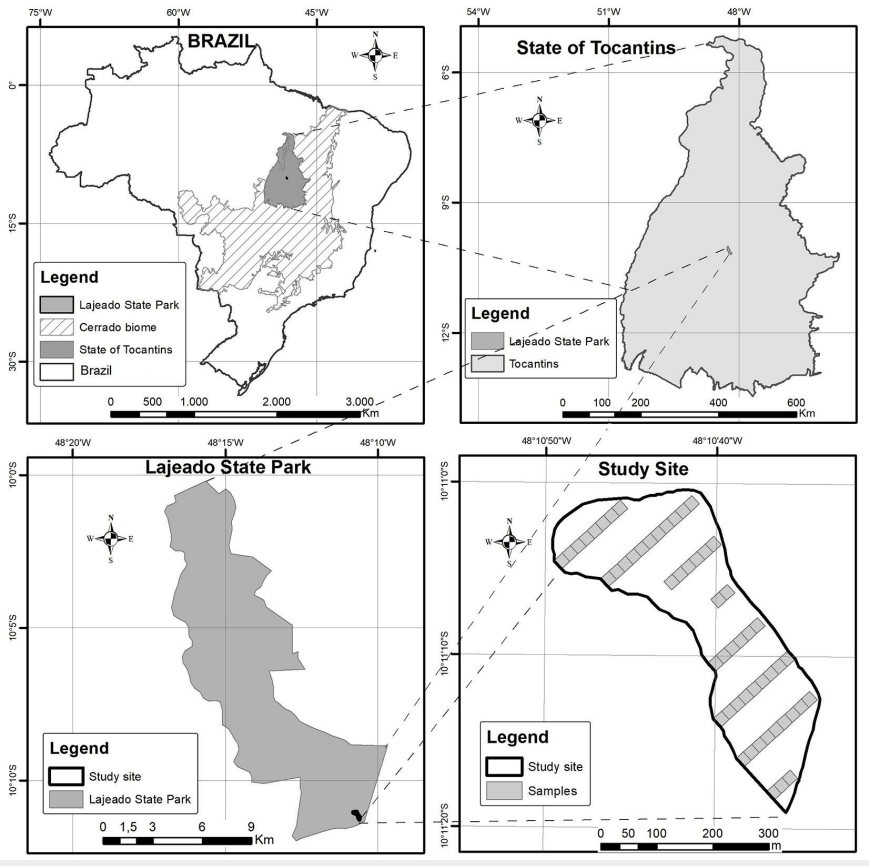

Fig. 1 - Study area location and layout.

ples were dried out in an oven of air circulation at 103 ± 2 °C until reaching constant mass values, and subsequently used as fuel in an adiabatic calorimeter. The wood HHV has been reported by a previous study (Silva 2014) conducted in our study area by applying the same methodology used in this analysis.

Biomass and energy potential

The dry biomass values of wood and bark were estimated for each individual tree using the model of Schumacher & Hall, which was adjusted and validated as described by Miguel et al. (2017 – Tab. 1). The adjusted equation refer to each total tree biomass using the average percentage of wood biomass (71.7%) and bark biomass (21.0%) as estimated by Miguel et al. (2017) for individual trees from the *Cerrado*.

The energy potential (*Pe*) of wood and bark of each sampled tree was separately estimated by the eqn. 1, which is based on the HHV observed in the laboratory for each individual species (eqn. 1):

$$Pe = \frac{HHV \cdot Bm}{238{,}845.9} \quad (1)$$

where *Pe* is the energetic potential of the individual's wood or bark (GJ); HHV is the higher heating value of the wood or bark of the species (kcal kg^{-1}); Bm is the dry biomass at 0% moisture of wood or bark for each tree (kg); and 238,845.9 is a conversion constant from kcal to GJ. For the dead and non-sampled individual trees, the mean value of HHV observed in the lab test was properly considered. The total energy potential of each trees was estimated by adding the wood *Pe* and bark *Pe*.

The *Pe* per hectare was estimated based on the results of each sample plot, as the sum of individual *Pe* of each tree observed in the plot, properly extrapolated to a hectare (eqn. 2):

$$Pe_i = \sum Pe_{ji} \quad (2)$$

where Pe_i is the energy potential per unit area of sample plot *i* (GJ ha^{-1}); Pe_{ji} is the energy potential of tree *j* observed in a sample plot *i* (GJ).

Remote sensing data

A RapidEye image acquired in 2012 covering the study area was geometrically and atmospherically corrected. Subsequently, six vegetation indices were derived from the corrected image using the ERDAS Imagine 2011 software (ERDAS 2011), which allowed to estimate an average vegetation index value for each sample plot used in this analysis. The RapidEye image was chosen based on its high spatial resolution (5 m) and spectral resolution, which features a red edge band in addition to visible and near infrared bands. Located between the red and near-infrared ranges (690-730 nm), the red edge band enhances RapidEye's capabilities for vegetation aspects assessment (Benítez et al. 2016).

Tab. 1 - Models for determination of dry biomass (Miguel et al. 2017). (Bm_w): wood dry biomass; (Bm_b): bark dry biomass; (DBH): diameter at breast height; (Ht): total height; (0.717): constant referring to the average percentage of wood biomass in a *Cerrado* tree; (0.210): constant referring to the average percentage of bark biomass in a *Cerrado* tree.

Model	Equation
Schumacher & Hall - Wood	$Bm_w = 0.0123307 \cdot DBH^{1.79593} \cdot Ht^{1.54701} \cdot 0.717$
Schumacher & Hall - Bark	$Bm_b = 0.0123307 \cdot DBH^{1.79593} \cdot Ht^{1.54701} \cdot 0.210$

We applied the following vegetation indices in this study: (i) NDVI, the normalized difference vegetation index (Rouse et al. 1974 – eqn. 3); (ii) NDRE, the red edge index of normalized difference, which is considered a NDVI variation for red edge band inclusion that increases sensitivity to chlorophyll (Gitelson & Merzlyak 1994 – eqn. 4); (iii) CI_{GREEN}, the chlorophyll index with the GREEN band, which shows high correlation with the chlorophyll content and leaf area index (Gitelson et al. 2003 – eqn. 5); (iv) $CI_{RED\ EDGE}$, the red edge index of chlorophyll, CI_{GREEN} variation for red edge band inclusion (Gitelson et al. 2003 – eqn. 6); (v) Savi, soil-adjusted vegetation index, used to correct the influence of soil exposed to vegetation (Huete 1988 – eqn. 7); and (vi) EVI_2, a modified vegetation index developed as an improvement for NDVI in its sensitivity to photosynthetic activity (Jiang et al. 2008 – eqn. 8):

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (3)$$

$$NDRE = \frac{(NIR - RED\ EDGE)}{(NIR + RED\ EDGE)} \quad (4)$$

$$CI_{GREEN} = \frac{NIR}{GREEN} - 1 \quad (5)$$

$$CI_{RED\ EDGE} = \frac{NIR}{RED\ EDGE} - 1 \quad (6)$$

$$Savi = \frac{(1 + L) \cdot (NIR - RED)}{NIR + RED + L} \quad (7)$$

$$EVI_2 = \frac{2.5 \cdot (NIR - RED)}{(NIR + 2.4 \cdot RED + 1)} \quad (8)$$

where NIR is the reflectance in the near infrared band, RED is the red band reflectance, RED EDGE is the reflectance of the red edge band, GREEN is the reflectance in the green band, L is the soil adjustment constant (0.5), 2.5 is the gain factor, while 2.4 and 1 are constants.

Modeling: independent variable selection

The energy potential prediction was conducted by using two independent variables: (i) the basal area (G – eqn. 9), as it is ease to be measured in the field and shows high correlation with biomass, tree growth and local production (Husch et al. 1982); (ii) the vegetation index retrieved from remotely sensed data. The selection of the optimum vegetation index was conducted by applying a correlation analysis among the six estimated vegetation indices and the *Pe*. Such analysis was preceded by the Shapiro-Wilk normality test, which indicated the correlation method to be subsequently used (parametric or non-parametric).

The basal area G_i of plot *i* (m^2) was calculated as (eqn. 9):

$$G_i = \sum g_{ji} \quad (9)$$

where g_{ij} is the sectional area of tree *j* in plot *i* (m^2), obtained as follows (eqn. 10):

$$g = \frac{\pi \cdot DBH^2}{40.000} \quad (10)$$

where π is the constant equal to 3.14159, DBH (cm) is the diameter at breast height (1.30 m above the ground), and 40.000 is a constant.

The statistical analysis was carried out using the software RStudio ver. 1.0.143 (RStudio 2016).

Modeling: training of neural networks

To select the input and output variables for the network, we applied a supervised training of one thousand neural networks using the Intelligent Problem Solver (IPS) tool available in the software package Statistica® ver. 7.0 (Statsoft Inc. 2007). The top five networks were retained at the end of processing. This tool allows for the optimization of 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 (Vale et al. 2017).

We used the quasi-Newton algorithm developed by Broyden-Fletcher-Goldfarb-Shanno for processing the neural networks in the IPS (BFGS), which has great resolution power for optimization problems and predictions and is the most popular quasi-Newton method (Guerrout et al. 2018).

The trained networks use a Multilayer Perceptron (MLP) architecture with an input layer consisting of two neurons (G and VI), a hidden layer consisting of *n* neurons, and an output layer consisting of one neuron, the energy potential (*Pe*). A multilayer network was used because of its non-linear characteristics, and because it is more suitable for the resolution of real problems and situations when compared to single layer networks with linear characteristics

Tab. 2 - Spearman's correlation matrix among the studied variables. (*Pe*): energy potential per area unit (GJ ha^{-1}); (G): basal area (m^2); (NDVI): normalized difference vegetation index; (NDRE): red edge normalized difference vegetation index; (CI_{green}): clorophyll index; ($CI_{red\ edge}$): red edge clorophyll index; (Savi): vegetation index adjusted to soil; (EVI_2): enhanced vegetation index; (*): p<0.05; (**): p<0.01

Variable	*Pe*	G	NDVI	NDRE	CI_{green}	$CI_{red\ edge}$	SAVI	EVI_2
Pe (GJ ha^{-1})	1	-	-	-	-	-	-	-
G (m^2)	0.8709**	1	-	-	-	-	-	-
NDVI	0.5987**	0.3269*	1	-	-	-	-	-
NDRE	0.5767**	0.2977*	0.9852**	1	-	-	-	-
CI_{green}	0.5922**	0.3160*	0.9893**	0.9813**	1	-	-	-
$CI_{red\ edge}$	0.5788**	0.2996*	0.9854**	0.9998**	0.9812**	1	-	-
SAVI	0.5979**	0.3390*	0.9908**	0.9719**	0.9921**	0.9724**	1	-
EVI_2	0.5979**	0.3390*	0.9908**	0.9719**	0.9921**	0.9724**	1.0000**	1

(Braga et al. 2014).

The neuron is the fundamental element of the ANN structure and is responsible for processing data and information when receiving input values (G and VI) until the result of interest (*Pe*) is achieved. The functionality of the neurons of ANNs is estimated by combining and reproducing information based on the connections between the *n* possible inputs x*n* and the output y. For each input a weight w*n* is defined, which represents the brain synapses and the MLP networks store the learned knowledge (Gorunescu & Belciug 2016). The artificial neuron is mathematically given by (eqn. 11):

$$Y_k = \varphi(V_k) \quad (11)$$

where Y_k is the output of the neuron, φ is the activation function, and V_k is the linear combinator of the output, given by (eqn. 12):

$$V_k = \sum x_m \cdot w_m \quad (12)$$

where V_k is the linear combiner, x_m is the input *m*, and w_m is the *m*-synaptic weight.

Statistical analysis

The ANN training was conducted using data from 40 out of 54 randomly selected sample plots, corresponding to 75% of the total plots. The remaining plots were excluded from the training dataset and used later to validate the results, being these sample plots within the range of 10% to 30% of the total, as suggested by Gujarati & Porter (2011). The estimated quality of the trained neural networks was statistically analyzed by using the following criteria: correlation between the predicted and observed values (R); graphical residue analysis (Draper & Smith 1998); and standard error of the estimate in percentage (RMSE%), calculated as (eqn. 13):

$$RMSE\,\% = \frac{100}{\bar{Y}} \sqrt{\frac{\sum (Y_i \hat{Y}_j)^2}{n}} \quad (13)$$

where $\bar{Y}$ is the mean of the observed values of ρe, Y_i is the observed value of ρe in plot *i*, and $\hat{Y}_i$ is the estimated value of ρe by ANN for plot *i*.

Finally, the ANN that showed the best results was submitted to the validation process using the *t*-test and, subsequently, an aggregate difference in percentage (AD%), a statistical value used as indicator of under- or overestimation (Miguel et al. 2015, Vale et al. 2017). These analyses were conducted using the software Microsoft Excel 2013® (Microsoft Corp., Redmond, CA, USA).

Results

Higher heating value

Based on the tree species sampled and measured in our study area, we estimated that the HHV of the woody material (tree stem and branches) is between 18.282 GJ ton^{-1} and 20.121 GJ ton^{-1}, with an average of 19.234 GJ ton^{-1}. The tree species showing the highest HHV were *Pouteria ramiflora* (20.121 GJ ton^{-1}), *Tachigali vulgaris* (20.105 GJ ton^{-1}), and *Mezilaurus itauba* (19.923 GJ ton^{-1}), while those having the lowest values were *Parkia platycephala* (18.282 GJ ton^{-1}), *Bowdichia virgilioides* (18.434 GJ ton^{-1}), and *Connarus perrottetti* (18.478 GJ ton^{-1} – Silva 2014).

The estimated HHVs for the tree's bark were between 16.805 GJ ton^{-1} and 21.587 GJ ton^{-1}, with an average of 19.878 GJ ton^{-1}. The tree species showing the highest HHV values were *Bowdichia virgilioides* (21.587 GJ ton^{-1}), *Xylopia aromatica* (21.290 GJ ton^{-1}), and *Connarus suberosus* (21.232 GJ ton^{-1}), while those having the lowest values were *Miconia albicans* (16.805 GJ ton^{-1}), *Miconia cuspidata* (16.857 GJ ton^{-1}), and *Qualea parviflora* (17.997 GJ ton^{-1}).

Energy potential

The estimated energy potential per area unit of the *cerradão* (a forest-like vegetation type) were in the range of 391.39 GJ ha^{-1} and 2719.18 GJ ha^{-1}, and the average energy density per hectare estimated for the studied area was 1022.66 GJ ha^{-1} ± 560.89 GJ ha^{-1}.

Modeling: selection of independent variables

The Shapiro-Wilk test showed significant departures from normality for the independent variables and, consequently, we calculated the non-parametric Spearman's correlation (Tab. 2). The results indicated that pairwise correlation coefficients among all vegetation indices (VIs) are moderate and fairly similar, while basal area (G) was highly correlated (ρ > 0.8) with the variable of interest (*Pe*). Despite the high similarity of correlation coefficients among the VIs, the NDVI showed the highest correlation (ρ = 0.5987) and, therefore, it was selected as the optimum VI to be used as predictor in the network training. The NDVI is widely applied and used in studies on forest environment. Notably, the lack of normality in the data did not hinder the modeling, since the ANN does not require assumptions on normality, linearity, and homoscedasticity, in contrast to more conventional regression models (Egrioglu et al. 2015).

Modeling: training of neural networks

The five best performing networks among the 1000 trained ANNs showed satisfactory adjustment and accuracy statistics, correlation coefficient (R) greater than 0.95, and errors of estimates below 15%. However, based on the RMSE% and R values (Tab. 3), the neural network 3 had the best predictive capacity.

The selected network showed an acceptable residue distribution pattern (Fig. 2A), accurate predictions of the variable of interest (Fig. 2B) and, according to the histogram of error classes, it had the largest concentration of residues in the central

Tab. 3 - Characteristics and precision analysis of trained artificial neural networks. (ANN): artificial neural network; (MLP): Multilayer Perceptron; (Hidden Activation): hidden layer activation function; (Output Activation): output layer activation function; (RMSE): Root-Mean-Square Error; (R): correlation coefficient between observed and predicted values.

ANN	Architecture	No. of cycles	Hidden Activation	Output Activation	RMSE (GJ ha^{-1})	RMSE%	R
1	MLP 2-8-1	583	Sigmoidal	Sine	147.91	14.24	0.96
2	MLP 2-7-1	272	Tangent	Sigmoidal	125.96	12.12	0.97
3	MLP 2-8-1	330	Tangent	Sigmoidal	117.05	11.27	0.98
4	MLP 2-9-1	267	Sigmoidal	Sigmoidal	141.60	13.63	0.97
5	MLP 2-8-1	10000	Exponential	Exponential	142.54	13.72	0.97

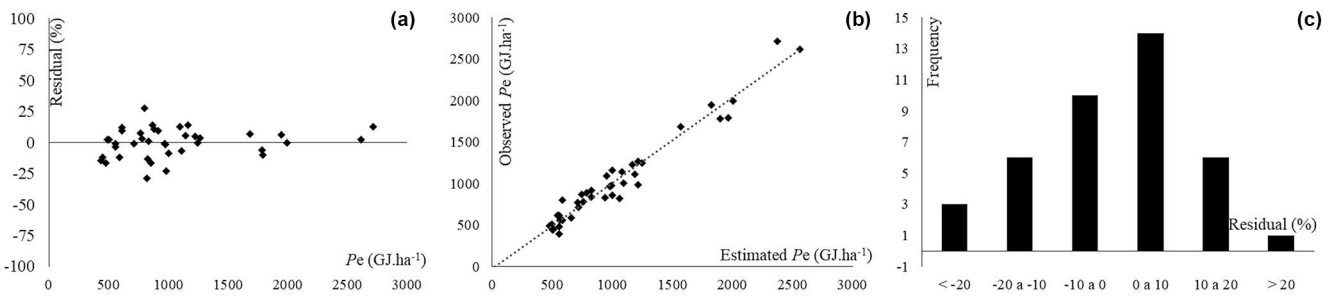

Fig. 2 - Residuals dispersal (A), observed and predicted values (B) e distribution of error classes (C) of energy density estimation.

classes of the graph, which indicates low (<20%) under- or overestimation errors (Fig. 2C). We observed that eight neurons in the hidden layer were activated by a tangent function, whereas the output layer is activated by a sigmoidal function (Fig. 3). Both the above functions have similar behavior, being suitable for biological studies and widely used in ANNs (Haykin 2001, Vendruscolo et al. 2015).

Statistical analysis

Based on the Student's *t*-test (*t* Stat = -0.358; *t* Critical two-tails = 2.160), we estimated a p-value far greater than 0.05 (p = 0.726) for the selected network, which indicates that there were no statistically significant differences between the predicted values by the neural network and the observed values from the forest inventory (validation plots). Additionally, the aggregate difference showed a slight overestimation (AD% = -4.63%), which indicates that the neural network training had a good predicting precision.

Discussion

Higher heating value and energy potential

The forest biomass HHV of the present study (Silva 2014) was between 18.282 GJ ton^{-1} and 20.121 GJ ton^{-1}, as corroborated by Quirino et al. (2004), which estimated HHV between 16.040 GJ ton^{-1} and 22.291 GJ ton^{-1}, with an average of 19.812 GJ ton^{-1} in a study conducted with approximately 200 tree species. In addition, studies conducted in specific formations of other forest types in the *Cerrado* reported HHVs ranging from 18.807 GJ ton^{-1} to 22.713 GJ ton^{-1} (Vale et al. 2002, Machado Neto et al. 2015). Tropical forests in countries such as Ecuador, Suriname, Cameroon, Côte d'Ivoire, Gabon, and Indonesia showed HHVs between 18.045 GJ ton^{-1} and 21.646 GJ ton^{-1} (Doat 1977). Finally, European tree species showed HHVs ranging between 18.290 GJ ton^{-1} and 20.370 GJ ton^{-1} (Strandberg et al. 2015, Yildiz et al. 2015). Therefore, the average HHV of wood for forest species is approximately 19.00 GJ ton^{-1} ± 4.00 GJ ton^{-1}.

It is known that there is a low significant variation in HHV among forest species (Machado Neto et al. 2015), which suggests a small confidence interval, as observed in the above-mentioned studies. Accordingly, we use the mean values of HHV (both wood and bark) for non-sampled tree species and dead trees (Pakeman & Quested 2007).

The variation in the energetic properties of forest biomass are also possibly due to differences in physical, chemical, and anatomical properties among species, within the same species, and even within each tree. According to Zobel & Jett (1995), the amplitude of wood variations reflects genetic and environmental factors and their interaction.

In this study, the estimated HHV for bark ranged across species from 16.805 GJ ton^{-1} to 21.587 GJ ton^{-1}, which was the energy potential expected for this type of material. Tree species of the *Cerrado* and temperate forests in North America show similar variation of bark HHV, ranging from 17.530 GJ ton^{-1} to 24.024 GJ ton^{-1}, and 17.254 GJ ton^{-1} to 24.242 GJ ton^{-1}, respectively (Corder 1976, Vale et al. 2002). All tree species in this study showed HHV within the limits observed in the literature.

We observed that 75% of the studied tree species showed higher HHV for bark than for wood. This may be explained by the differences in the carbon content between wood and bark of individual trees (average: 20.73% vs. 25.19%, respectively – Vale et al. 2002), as well as by differences in the lignin content, which can reach 28% in stem wood and 36% in the bark (Telmo & Lousada 2011).

The *Pe* value per unit area estimated in this study is approximately 25% lower than the value of 1378.54 GJ ha^{-1} reported by Silva & Vale (2018) for the same area. This discrepancy may be related to the different models applied, which may have affected the results and errors in the estimation. However, it is important to mention that the *Pe* value per unit area represent the maximum energy potential of the area, which takes into account the dry biomass at 0% moisture (lab condition), which is not actually observable in the field. Therefore, the energy gain does not pay off the energy necessary to dry out biomass up to the 0% moisture point (Brand et al. 2011). Further, the presence of water in the biomass will reduce its heating values, which decreases 2 MJ kg^{-1} for every 10% moisture increase in the biomass (Górnicki et al. 2016).

Modeling: selection of independent variables

In this study, the estimated correlation coefficients between vegetation index and energy potential (*Pe*) were moderate but significant (Tab. 2). Similar results have been reported by several studies focused on the relationship among VIs and different field variables from the forest inventory, such as trunk circumference, wood volume, and biomass (Miguel et al. 2015, Zhu & Liu 2015, Machado et al. 2017). Nonetheless, the observed correlations between the VIs and forest variables may be affected by remote sensing externalities, such as the effects of the atmosphere, radiation sources, and soil effects, as well as inherent vegetation characteristics such as architecture and conformation of the tree

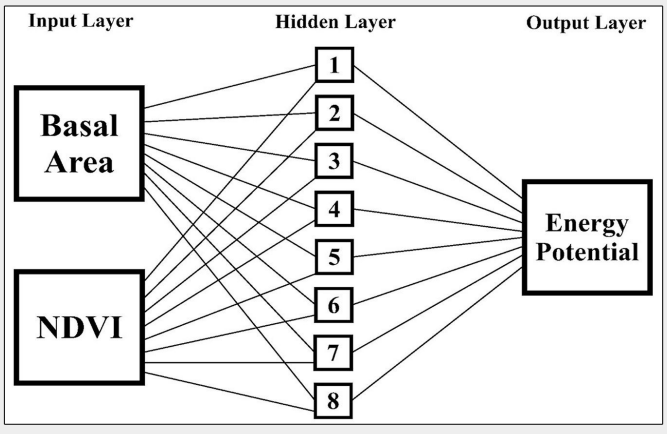

Fig. 3 - Architecture of the selected ANN for prediction of energy density per unit area in *cerradão*: 2-8-1.

top and canopy (Miguel et al. 2015, Machado et al. 2017).

The positive and significant correlation between the VIs and *Pe* (Tab. 2) reflects the correlation of vegetation indices with radial stem growth and leaf area, which are directly related to the accumulation of biomass, carbon sequestration and, consequently, carbon content of the wood (Vicente-Serrano et al. 2016), while core and summer wood of the stem are characterized by a larger presence of lignin (Klitzke et al. 2008).

The presence and content of lignin is one of the determinants of vegetation reflectance for wavelengths in the infrared region (Padolfi et al. 2018), which was used to estimate most of the VIs applied in our analysis. This may have affected our results because carbon and lignin content shows a strong positive correlation with the HHV and, consequently, with the energy density of the wood (Telmo & Lousada 2011).

Modeling: training of neural networks

All trained ANNs in our study showed acceptable adjustment and precision statistics, with high correlation coefficient (R ≥ 0.96) and errors of estimation below 15% (Tab. 3). More specifically, the chosen network showed very satisfactory results (R = 0.98, RMSE% = 11.27%) when considering the complexity of the studied variables and the great variability observed in the sampled area, which may increase the limit of error up to 20% acceptable.

Several studies demonstrated the superiority of neural networks against the classic regression models in the estimation of several forest attributes and proved the great potential of their use associated with remote sensing data (Miguel et al. 2015, Vacchiano et al. 2018). The results obtained in this study also confirmed that the association of artificial intelligence and remote sensing techniques allowed the ANNs to accurately model the great complexity and variety of vegetation and other environmental aspects (Felfili et al. 2007).

Neural networks have also been shown to efficiently predict wood intrinsic characteristics such as moisture content (Ozsahin & Murat 2018), basic wood density (Silva et al. 2018), higher heating value (Estiati et al. 2016), and energy density (Vale et al. 2017), for which the aforementioned authors obtained better results than those reported in this study (RMSE% = 1.45%, R = 0.98, AD% = 0.14). However, those authors used basic wood density as a predictive variable. The predictive power of the ANNs was also tested for the estimation of the quality of products such as medium density fiber (MDF) boards and their different bonding strength processes (Tiryaki et al. 2016).

The graphical analysis of the residues is fundamental to corroborate the results of precision statistics (Draper & Smith 1998), as it can reveal trends in the errors that may not be detected by such statistics. Moreover, errors of interpretation could be due to the occurrence of overlapping points in the graphs. In this study the residual dispersion graph (Fig. 2A) showed adequate distribution of errors, no visible trends, compact and well distributed points along the regression line, and maximum errors limited to ± 30%. The relation between the observed and predicted values (Fig. 2B) reflects the adherence of the selected neural network to the real data. The frequency of errors was mostly concentrated between -10% and 10% (Fig. 2C), with few errors greater than ± 20%.

Statistical analysis

We found no significant difference between the observed values of energy potential and their predictions obtained by ANN, as the use of network allows for consistent adjustment and adaptation to the observed data. Therefore, we conclude that the ANN is a suitable tool for the estimation of energy potential for the phytophysiognomy under study. Similar conclusions were reported by Serpen & Gao (2014), who stated that ANNs have a great ability to learn and extract patterns from one set of data, generalize and apply them to other data sets without losing accuracy.

Due to the variability of species, structure, canopy architecture, canopy format, and the physical and chemical characteristics of different species' wood in the *Cerrado*, further studies are needed to investigate the relationships among forest field variables and VIs. Different configurations and network architectures should be tested to improve the accuracy of estimates for different biomes with different phytophysiognomies, which may strengthen and validate the ANNs application in the prediction of energy density per unit area by using remotely sensed data.

Conclusion

The *cerradão* showed a maximum energy potential of 1022.66 ± 560.89 GJ ha^{-1} at 0% moisture dry biomass

Artificial neural networks of the Multilayer Perceptron type were generated using tangential and sigmoidal activation functions and the BFGS training algorithm. They were implemented using vegetation indices derived from remote sensing data and basal area as input variables, and provided accurate and efficient estimations of the energy potential of forest biomass per area unit of *cerradão*.

The result of this study can help identifying areas with greater energy content, decreasing the time and cost of forest surveys aimed to estimate the energy potential (*Pe*). Based on our results, we conclude that it is possible to accurately estimate energy potential by using NDVI derived from remotely sensed data and the basal area as variable input of an ANNs.

Abbreviations

Pe: energy potential; BFGS: Broyden-Fletcher-Goldfarb-Shanno; Bm; Biomass; DBH: diameter at breast height; AD%: aggregate percentage difference; G: basal area; g: sectional area; Ht: total height; IPS: intelligent problem solver; VI: vegetation index; IVI: importance value index; MLP: multilayer perceptron; NATURATINS: Environmental Agency of the State of Tocantins; HHV: higher heating value; R: correlation coefficient; RMSE: standard error of estimate; ANN: artificial neural network.

Acknowledgements

Financial support was provided by the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Brazilian National Council for Scientific and Technological Development (CNPq).

References

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G (2014). Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22 (6): 711-728. - doi: 10.1127/0941-2948/2013/0507

Araújo ACC, Costa LJ, Braga PPC, Guimarães Neto RM, Rocha MFV, Trugilho PF (2018). Propriedades energéticas da madeira e do carvão vegetal de *Cenostigma macrophyllum*: subsídios ao uso sustentável [Wood and charcoal energy properties of *Cenostigma macrophyllum*: subsidies to sustainable use]. Pesquisa Florestal Brasileira 38: 1-9. [in Portuguese] - doi: 10.4336/2018.pfb.38e201701546

Benítez FL, Anderson LO, Formaggio AR (2016). Evaluation of geostatistical techniques to estimate the spatial distribution of aboveground biomass in the Amazon rainforest using high-resolution remote sensing data. Acta Amazonica 46: 151-160. - doi: 10.1590/1809-43922015012 54

Braga AP, Carvalho APLF, Ludemir TB (2014). Redes neurais artificiais: teoria e aplicações [Artificial neural network: theory and applications] (2nd edn). LTC, Rio de Janeiro, Brazil, pp. 226. [in Portuguese]

Brand MA, Muñiz GIB, Quirino WF, Brito JO (2011). Storage as a tool to improve wood fuel quality. Biomass and Bioenergy 35: 2581-2588. - doi: 10.1016/j.biombioe.2011.02.005

Corder SE (1976). Properties and uses of bark as an energy source. Research Paper 31, Forest Research Laboratory, School of Forestry, Oregon State University, Corvallis, OR, USA, pp. 21. [online] URL: http://ir.library.oregonstate.edu/concern/technical_reports/1z40kv16c?locale=en

Dalponte M, Frizzera L, Orka HO, Gobakken T, Gianelle D (2018). Predicting stem diameters and aboveground biomass of individual trees using remote sensing data. Ecological Indicators 85: 367-376. - doi: 10.1016/j.ecolind.2017.10.066

Doat J (1977). Le pouvoir calorifique des bois tropicaux [The calorific value of tropical woods]. Revue Bois et Forêts Tropicaux 172: 33-48. [in French] - doi: 10.19182/bft1977.172.a19325

Draper NR, Smith H (1998). Applied regression analysis (3rd edn). John Wiley and Sons, New York, USA, pp. 407.

Egrioglu E, Yolcu U, Aladag CH, Bas E (2015). Recurrent multiplicative neuron model artificial neural network for non-linear time series forecasting. Neural Processing Letters 41: 249-258. -

doi: 10.1007/s11063-014-9342-0
ERDAS (2011). ERDAS Imagine 2011. Hexagon Geospatial, Madison, AL, USA.
Estiati I, Freire FB, Freire JT, Aguado R, Olazar M (2016). Fitting performance of artificial neural networks and empirical correlations to estimate higher heating values of biomass. Fuel 180: 377-383. - doi: 10.1016/j.fuel.2016.04.051
Felfili JM, Rezende AV, Silva MCJ, Silva PEN, Walter BMT, Encinas JI, Silva MA (2007). Fitossociologia da vegetação arbórea [Phytosociology of tree vegetation]. In: "Biogeografia do bioma Cerrado: vegetação e solos da Chapada dos Veadeiros" [Cerrado biome biogeography: vegetation and soils of Chapada dos Veadeiros] (Felfili JM, Rezende AV, da Silva Jr. MC eds). Editora Universidade de Brasília, Finatec, Brasília, Brazil, pp. 45-96. [in Portuguese]
Gitelson AA, Merzlyak MN (1994). Quantitative estimation of chlorophyll-*a* using reflectance spectra: experiments with autumn chestnut and maple leaves. Journal of Photochemistry and Photobiology B: Biology 22: 247-252. - doi: 10.1016/1011-1344(93)06963-4
Gitelson AA, Gritz Y, Merzlyak MN (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology 160: 271-282. - doi: 10.1078/0176-1617-00887
Górnicki K, Kaleta A, BryA, Winiczenko R (2016). Thin-layer drying of sawdust mixture. Polish Journal of Chemical Technology 18: 65-70. - doi: 10.1515/pjct-2016-0072
Gorunescu F, Belciug S (2016). Boosting backpropagation algorithm by stimulus-sampling: application in computer-aided medical diagnosis. Journal of Biomedical Informatics 63: 74-81. - doi: 10.1016/j.jbi.2016.08.004
Groot A, Cortini F, Wulder MA (2015). Crown-fibre attribute relationships for enhanced forest inventory: progress and prospects. The Forestry Chronicle 91: 266-279. - doi: 10.5558/tfc2015-048
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. Journal of Experimental and Theoretical Artificial Intelligence 30: 415-427. - doi: 10.1080/0952813X.2017.1409280
Gujarati DN, Porter DC (2011). Econometria básica [Basic econometrics] (5th edn). AMGH, New York, USA, pp. 924. [in Portuguese]
Haykin S (2001). Redes neurais: princípios e prática [Neural networks: principles and practice] (2nd edn). Bookman, Porto Alegre, Brazil, pp. 900. [in Portuguese]
Huete A (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25 (3): 295-309. - doi: 10.1016/0034-4257(88)90106-X
Husch B, Miller CL, Beers TW (1982). Forest mensuration (3rd edn). John Wiley, New York, USA, pp. 402.
Jiang Z, Huete AR, Didan K, Miura T (2008). Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment 112: 3833-3845. - doi: 10.1016/j.rse.2008.06.006
Klitzke RJ, Savioli DL, Muñiz GIB, Batista DC (2008). Caracterização dos lenhos de cerne, alburno e transição de jatobá (*Hymenaea* sp.) visando ao agrupamento para fins de secagem convencional [Characterization of heartwood, sapwood and transition wood of jatoba wood (*Hymenaea* sp.) aiming at its grouping in conventional kiln drying]. Scientia Forestalis 36: 279-284. [in Portuguese]
Lima AS, Santos KPP, Castro AAJF (2016). Aspectos socioambientais da produção de carvão vegetal de origem nativa em uma área de cerrado em Jerumenha, Piauí/Brasil [Social and environmental aspects of the charcoal production of native origin in an area of savannah in Jerumenha, Piauí, Brazil]. Espacios 37: 18. [in Portuguese]
López L, Villalba R, Bravo F (2013). Cumulative diameter growth and biological rotation age for seven tree species in the Cerrado biogeographical province of Bolivia. Forest Ecology and Management 292: 49-55. - doi: 10.1016/j.foreco.2012.12.011
Machado IES, Santos MM, Giongo M, Carvalho EV, Ganassoli Neto E (2017). Modelos para estimativa de variáveis florestais com a utilização de imagens multiespectrais [Models of forest variables estimation using multispectral images]. Pesquisa Florestal Brasileira 37: 171-181. [in Portuguese] - doi: 10.4336/2017.pfb.37.90.1380
Machado Neto AP, Brandão CFLS, Duarte B, Almir J, Marangon LC, Feliciano ALP (2015). Densidade e poder calorífico como base para prevenção de incêndios florestais sob linhas de transmissão [Density and calorific power as a basis for forest fires prevention under transmission lines]. Nativa 3: 10-15. [in Portuguese] - doi: 10.31413/nativa.v3i1.1813
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 [Artificial neural network for modeling wood volume and aboveground biomass of tall Cerrado using satellite data]. Pesquisa Agropecuaria Brasileira 50: 829-839. [in Portuguese] - doi: 10.1590/S0100-204X2015000900012
Miguel EP, Rezende AV, Leal FA, Pereira RS, Melo RR (2016). Floristic-structural characterization and successional group of tree species in the Cerrado biome of Tocantins state, Brazil. Revista Caatinga 29: 393-404. - doi: 10.1590/1983-21252016v29n216rc
Miguel EP, Rezende AV, Pereira RS, De Azevedo GB, Mota FCM, De Souza AN, Joaquim MS (2017). Modeling and prediction of volume and aereal biomass of the tree vegetation in a Cerradão area of central Brazil. Interciencia 42: 21-27. [online] URL: http://www.redalyc.org/pdf/339/33949290004.pdf
Oliveira GMV, Mello JM, Scalon JD, Scolforo JRS, Monteiro TC (2018). Amostragem de discos e uso de equações para estimar a densidade básica da madeira em diversas fitofisionomias [Disc sampling and use of equations for estimate the basic wood density in several phytophysiognomies]. Ciência Florestal 28: 1615-1626. [in Portuguese] - doi: 10.5902/1980509835127
Ozsahin S, Murat M (2018). Prediction of equilibrium moisture content and specific gravity of heat treated wood by artificial neural networks. European Journal of Wood and Wood Products 76: 563-572. - doi: 10.1007/s00107-017-1219-2
Padolfi AS, Ramaldes GP, Dos Santos OL (2018). Análise de índice de vegetação através de imagens obtidas por VANT [Vegetation index analysis through UAV obtained images]. Revista Científica da FAESA 14: 145-165. [in Portuguese] - doi: 10.5008/1809.7367.130
Pakeman RJ, Quested HM (2007). Sampling plant functional traits: what proportion of the species need to be measured? Applied Vegetation Science 10: 91-96. - doi: 10.1111/j.1654-109X.2007.tb00507.x
Protásio TP, Trugilho PF, De Siqueira HF, De Melo ICNA, Andrade CR, Guimarães JBJ (2015). Caracterização energética de pellets in natura e torrificados produzidos com madeira residual de *Pinus* [Energy characterization of fresh and torrified pellets produced from *Pinus* waste wood]. Pesquisa Florestal Brasileira 35: 435. [in Portuguese] - doi: 10.4336/2015.pfb.35.84.843
Quirino WF, Do Vale AT, De Andrade APA, Abreu VLS, Azevedo ACS (2004). Poder calorífico da madeira e de resíduos lignocelulósicos [Calorific value of wood and wood residues]. Biomassa & Energia 1: 173-182. [in Portuguese]
Rouse JW, Haas RH, Schell JA, Deering DW (1974). Monitoring vegetation systems in the great plains with ERTS. In: Proceedings of "3rd ERTS-1 Symposium". Washington (DC, USA) Dec 1973. NASA, Washington, DC, USA, pp. 309-317. [online] URL: http://ntrs.nasa.gov/search.jsp?R= 19740022614
RStudio (2016). RStudio: integrated development for R 1.0.143. RStudio Inc., Boston, MA, USA. [online] URL: http://www.rstudio.com/
Serpen G, Gao Z (2014). Complexity analysis of multilayer perceptron neural network embedded into a wireless sensor network. Procedia Computer Science 36: 192-197. - doi: 10.1016/j.procs.2014.09.078
Silva CJ (2014). Densidade básica e potencial energético de espécies lenhosas do Cerrado do estado de Tocantins [Basic density and energy potential of woody species of the Cerrado of Tocantins state]. Doctoral thesis, Forestry Department, Universidade de Brasília, Brasília, Brazil, pp. 66. [in Portuguese]
Silva CJ, Vale AT (2018). Energy density model for forest species from cerrado. Revista Caatinga 31: 396-404. - doi: 10.1590/1983-21252018v31n216rc
Silva JPM, Cabacinha CD, Assis AL, Monteiro TC, Araújo Júnior CA, Maia RD (2018). Redes neurais artificiais para estimar a densidade básica de madeiras do cerrado [Artificial neural network to estimate the basic density of cerrado wood]. Pesquisa Florestal Brasileira 38: 1-10. [in Portuguese] - doi: 10.4336/2018.pbf.38e201801656
Solórzano A, Pinto JR, Felfili JM, Hay JD (2012). Perfil florístico e estrutural do componente lenhoso em seis áreas de cerradão ao longo do bioma Cerrado [Structural and floristic profile of the woody component of six cerradão areas]. Acta Botanica Brasilica 26: 328-341. [in Portuguese] - doi: 10.1590/S0102-33062012000200009
Statsoft Inc. (2007). Statistica (data analysis software system) release 7. Statsoft Inc., Tulsa, OK, USA. [online] URL: http://www.statsoft.com/

Strandberg M, Olofsson I, Pommer L, Wiklund-Lindström S, Aberg K, Nordin A (2015). Effects of temperature and residence time on continuous torrefaction of spruce wood. Fuel Processing Technology 134: 387-398. - doi: 10.1016/j.fuproc.2015.02.021

Telmo C, Lousada J (2011). The explained variation by lignin and extractive contents on higher heating value of wood. Biomass and Bioenergy 35: 1663-1667. - doi: 10.1016/j.biombioe.2010.12.038

Tiryaki S, Bardak S, Aydin A, Nemli G (2016). Analysis of volumetric swelling and shrinkage of heat treated woods: experimental and artificial neural network modeling approach. Maderas. Ciencia y Tecnología 18: 477-492. - doi: 10.4067/S0718-221X2016005000043

Vacchiano G, Berreti R, Motta R, Mondino EB (2018). Assessing the availability of forest biomass for bioenergy by publicly available satellite imagery. iForest 11: 459-468. - doi: 10.3832/ifor2655-011

Vale AT, Brasil MAM, Leão AL (2002). Quantificação e caracterização energética da madeira e casca de espécies do cerrado [Energetic quantification and caracterization of wood and bark of species of cerrado]. Ciência Florestal 12 (1): 71-80. [in Portuguese] - doi: 10.5902/198050981702

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. African Journal of Agricultural Research 12: 856-862. - doi: 10.5897/AJAR2016.12083

Vendruscolo DGS, Drescher R, Souza HS, Moura JPVM, Mamoré FMD, Siqueira TAS (2015). Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais [*Eucalyptus* height estimation by nonlinear regression and artificial neural networks]. Revista Brasileira de Biometria 33: 556-569. [in Portuguese] - doi: 10.13140/RG.2.1.1742.5684

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 Sensing of Environment 187: 14-29. - doi: 10.1016/j.rse.2016.10.001

Yildiz G, Ronsse F, Venderbosch R, Van Duren R, Kersten SRA, Prins W (2015). Effect of biomass ash in catalytic fast pyrolysis of pine wood. Applied Catalysis B: Environmental 168: 203-211. - doi: 10.1016/j.apcatb.2014.12.044

Zhu X, Liu D (2015). Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS Journal of Photogrammetry and Remote Sensing 102: 222-231. - doi: 10.1016/j.isprsjprs.2014.08.014

Zobel BJ, Jett JB (1995). Genetics of wood production. Springer-Verlag, Berlin, Germany, pp. 337.