Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Special Issue: Artificial Intelligence

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210153/2022

RANDOM FOREST MODEL TO PREDICT THE HEIGHT OF EUCALYPTUS

Elizeu de S. Lima[1], Zigomar M. de Souza[1], Stanley R. de M. Oliveira[1,2], Rafael Montanari[3], Camila V. V. Farhate[4*]

[4*]Corresponding author. Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista/ Jaboticabal - SP, Brasil.
E-mail: camilavianav@hotmail.com | ORCID ID: https://orcid.org/0000-0002-5027-9295

KEYWORDS

Physicochemical variables of soil, machine learning, soil phosphorus content, soil moisture, exchangeable aluminum.

ABSTRACT

Eucalyptus (*Eucalyptus urograndis*) production has significantly advanced over the past few years in Brazil, especially with regard to acreage and productivity. Machine learning has made significant advances in most varied fields of agrarian sciences. In this context, this study aimed to use physicochemical variables of the soil as well as climatic and dendrometric variables of eucalyptus to predict its height using the random forest algorithm. The study was conducted in the municipality of Três Lagoas, in Mato Grosso do Sul, Brazil. The original database consisted of 49 soil physicochemical variables collected at 0–0.20 m and 0.20–0.40 m, two dendrometric and four climatic variables, and one response variable related to the height of eucalyptus. A correlation matrix was applied to select variables. Furthermore, modeling was performed using the random forest algorithm, which performed well ($r = 0.98$, $R^2 = 0.96$) in predicting the height of eucalyptus. Overall, the most important variables to predict the eucalyptus plant height included diameter at breast height (DBH), phosphorus content (P1), gravimetric moisture (GM1) at a soil depth between 0.00 m and 0.20 m, and exchangeable aluminum content (Al2) between 0.20 m to 0.40 m of soil.

INTRODUCTION

High demand for wood for different purposes (sawmill, lamination, charcoal, and cellulose) has led to a substantial increase in the area of planted forests. Consequently, it has contributed to the national economy by generating employment (direct and indirect) and income in primary and secondary sectors (Pichelli & Soares, 2019). In 2019, the total planted forest area in Brazil was 10 million hectares (IBGE, 2019). Of this, eucalyptus cultivation contributed 77% (6.97 million hectares), with an average productivity of 35 m^3 ha^{-1} per year (IBÁ, 2020).

Predictive models for eucalyptus growth have great potential to expand cultivated areas by aiding the decision-making process for regions whose species adaptation and growth conditions are not well established (Santos et al., 2017). Thus, predictive modeling can be useful for assessing growth and production in eucalyptus-cultivated areas to support the management of forests (Castro et al., 2016). According to Scolforo et al. (2013), choosing the ideal management practice for each forest area cultivation is crucial for successful actions.

The total height of trees in forest inventories is important as it is strongly correlated with wood volume (Campos et al., 2016). However, its estimation is a long-term process that demands financial resources and is subject to errors (Souza et al., 2016; Vendruscolo et al., 2015). In this context, hypsometric models that express the relationship between tree diameter and height are commonly used to predict tree height (Martins et al., 2016). However, many factors influence thypsometric relations, including age, edaphoclimatic conditions, cultivar, management system, competition status, and productive

[1] Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas/ Campinas - SP, Brasil.
[2] Embrapa Agricultura Digital/ Campinas - SP, Brasil.
[3] Faculdade de Engenharia, Universidade Estadual Paulista/ Ilha Solteira - SP, Brasil.

Area Editor: Gizele Ingrid Gadotti
Received in: 8-30-2021
Accepted in: 12-20-2021

capacity (Campos & Leite, 2013; Finger, 1992). In addition, it is difficult to find the right relationship between diameter and height. This is because tree trunks have portions that are not usable or are uneven, which lead to overestimation in diameter and underestimation in height; the opposite can also hold true (Ferraz Filho et al., 2018; Hess et al., 2014; Martins et al., 2016).

Machine learning algorithms are a promising approach in the most varied fields of agrarian sciences (Farhate et al., 2018; Marçal et al., 2021; da Silva et al., 2021; Tavares et al., 2018); of these, the random forest (RF) algorithm is considered to be one of the most accurate methods (Biau, 2012; Wang et al., 2016). It is an unsupervised learning method that assesses the performance of a set of independent regression tree-type algorithms using different bootstrap samples of the training data to predict the value of a given variable and express the final results through the mean values of individual trees (Breiman, 2001). RF is advantageous because of its high processing speed, easy implementation, high precision, and ability to handle a large number of input variables without overlap (Biau, 2012).

Considering the difficulty in predicting the stem height in eucalyptus plantations using traditional methods and the high predictive potential of the RF model, this approach can be applied to predict eucalyptus growth using correlated variables. The objective of this study was to predict the growth of eucalyptus via the RF machine learning model, using physicochemical soil components with climatic and dendrometric variables.

MATERIAL AND METHODS

Experiment Location

The experiment was conducted in Três Lagoas (20°27′ S, 52°29′ W), which is a municipality in the state of Mato Grosso do Sul, Brazil. According to the Köppen-Geiger climate classification system (Köppen & Geinger, 1928), it belongs to class Aw and is characterized as rainy in the summer and dry in the winter. Furthermore, this region has a mean annual rainfall precipitation of 1,300 mm and mean temperature of 23.7 °C. According to the Brazilian system of soil classification (Santos et al., 2018) and the Soil Taxonomy System (Soil Survey Staff, 2014), the soil of the experimental area is Neossolo Quartzarênico and Etisols Quartzipsamments, respectively.

Description and history of the experimental area

Fifty years ago, the experimental area was cultivated with degraded pasture. Since 2013, it has been cultivated with *Eucalyptus urograndis*. The present study was conducted over the crop years of 2014–2015.

Analyzed variables

The following dendrometric variables were assessed: individual height of the eucalyptus trees (HGT), diameter at breast height (DBH), and wood volume (VOL). Data on tree height were collected using a 5 m graduated ruler; DBH data were collected at a height of 1.30 m from the soil using a digital caliper rule. Individual VOL was obtained using a standing tree. Cutting down of trees was ruled out because the experimental area was located in a privately owned commercial area. Therefore, Huber's formulation was used to establish the VOL because it assumes that the mean area of a sectioned tree is at its midpoint; however, such an assumption is not always true, which indicates that its accuracy is intermediary (Campos & Leite, 2013). In this scenario, a form factor of 0.4 was used to correct individual wood volume, assuming that wood was not a perfect cylinder (Oliveira et al., 2009). Huber's formula, adapted and described by Péllico Netto (2004), was obtained from the product of half the sectioned area and the section length, and determined using [eq. (1)].

$$VOL = [DBH^2 * (3.14 / 4) * HGT] * 0.4 \quad (1)$$

Where:

VOL is the wood volume (m^3);

DBH is the diameter at breast height (m), and

HGT is the tree height (m).

Additionally, this study assessed the following physicochemical variables: soil penetration resistance (PR), gravimetric moisture (GM), volumetric moisture (VM), bulk density (BD), particle density (PD), total porosity (TP), sand, silt, clay, phosphorus (P), organic matter (OM), potential of hydrogen (pH), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), potential acidity (H^++Al^{3+}), aluminum (Al^{3+}), sum of bases (SB), cation exchange capacity (CEC), base saturation (V), calcium and cation-exchange capacity ratio (Ca/CEC), magnesium and cation-exchange capacity ratio (Mg/CEC), and aluminum saturation (m). All attributes were collected at depths of 0.00–0.20 m and 0.20–0.40 m, and assessed using the methodology proposed by Teixeira et al. (2017), Stolf et al. (2014), and Raij et al. (2001). For BD and TP, samples with preserved structures were collected in stainless steel cylinders having an average volume of 83.70 cm^3 (diameter = 47 cm; height = 50 cm). GM, VM, PD, P, OM, pH, K^+, Ca^{2+}, Mg^{2+}, H^++Al^{3+}, and Al^{3+} were performed on deformed samples.

Temperature and rainfall in the experimental area were monitored using an automatic meteorological station located in the municipality of Três Lagoas-MS, which was ~50 km from the experimental area. The obtained data enabled the evaluation of climatic conditions during the experimental period.

Data mining

For each crop year, 150 plants were sampled and soil samples were collected around each tree. This process was carried out for two crop years, totaling 300 observations. According to Table 1, the original database consisted of 49 physicochemical variables of the soil collected at two depths (0–0.20 m and 0.20–0.40 m), two dendrometric and four climatic variables, as well as one response variable related to the height of *Eucalyptus urograndis* (Table 1).

TABLE 1. Description of the predictive variables (soil physicochemical, dendrometrical, and climatic) used in the database to predict the height of *Eucalyptus urograndis* through the Random Forest (RF) model.

Variable	Description	Unit	Type of variable	Depth of sampling
PR	Penetration resistance	Mpa	Predictive	1 and 2
GM	Gravimetric moisture	kg kg^{-1}	Predictive	1 and 2
VM	Volumetric moisture	m^3 m^{-3}	Predictive	2
BD	Bulk density	Mg dm^{-3}	Predictive	2
PD	Particle density	m^3 m^{-3}	Predictive	1 and 2
TP	Total porosity	m^3 m^{-3}	Predictive	2
Sand	Sand	g kg^{-1}	Predictive	1 and 2
Sil	Silt	g kg^{-1}	Predictive	1 and 2
Clay	Clay	g kg^{-1}	Predictive	1 and 2
P	Exchangeable phosphorus	mg dm^{-3}	Predictive	1 and 2
OM	Organic matter	mg dm^{-3}	Predictive	1 and 2
pH	pH in calcium chloride	-	Predictive	1 and 2
K^+	Exchangeable potassium	$mmol_c$ dm^{-3}	Predictive	1 and 2
Ca^{2+}	Exchangeable calcium	$mmol_c$ dm^{-3}	Predictive	1 and 2
Mg^{2+}	Exchangeable magnesium	$mmol_c$ dm^{-3}	Predictive	1 and 2
H^++Al^{3+}	Potential acidity	$mmol_c$ dm^{-3}	Predictive	1 and 2
Al^{3+}	Exchangeable aluminum	$mmol_c$ dm^{-3}	Predictive	1 and 2
SB	Sum of bases	$mmol_c$ dm^{-3}	Predictive	1 and 2
CEC	Cation-exchange capacity	$mmol_c$ dm^{-3}	Predictive	1 and 2
V	Bases saturation	%	Predictive	1 and 2
Ca/CEC	Ca/CEC ratio	$mmol_c$ dm^{-3}	Predictive	1 and 2
Mg/CEC	Mg/CEC ratio	$mmol_c$ dm^{-3}	Predictive	1 and 2
M	Aluminum saturation	%	Predictive	1 and 2
Prec	Precipitation	mm day^{-1}	Predictive	-
T	Temperature (max, mean, min) minimum and mean	°C	Predictive	-
DBH	Diameter at breast height	cm	Predictive	-
VOL	Wood volume	m^3	Predictive	-
HGT	Height of the *Eucalyptus urograndis*	m	Response	-

Numbers 1 and 2 with each variable refer to the sampling layers: 1 = 0.00–0.20 m and 2 = 0.20–0.40 m of soil depth.

The covariance between two variables is related to their variance with each other. Therefore, a correlation matrix was established to verify the simple linear correlations for the two-to-two combinations of all the variables contained in the database. Positive correlations were expressed through blue staining: the more intense the blue staining, the more positive the correlation degree; in contrast, the more intense the red staining, the more negative the degree of correlation. Null correlations were expressed in the absence of color. The aim was to only select variables that could contribute to the model. Variables with null variance or high correlation with each other were eliminated. In the case of two highly correlated variables, one was randomly maintained and the other was eliminated because it added no additional information to the model.

The RF algorithm (Breiman, 2001) was applied to elaborate the predictive modeling for the height of the eucalyptus plants. It consisted of a set of combined decision trees to solve possible classification and regression issues. Each decision tree was built using random initial data sampling, and each data division used a random subset of

attributes to select the most informative ones (Breiman, 2001; Hastie et al., 2001).

In data mining applications, input predictor variables differ in relevance. Often, few variables have a substantial influence on the response, and most are irrelevant and are discarded. In this context, it is useful to learn the relative importance or contribution of each input variable to predict the response. Each tree was trained on a bootstrap sample, and the optimal variables at each split were identified from a random subset of all variables. The selection criteria for classification and regression problems were the Gini index and variance reduction, respectively (Hastie et al., 2009).

The RF algorithm was implemented in the R program (R Core Team, 2017), while model validation was conducted via the hold-out method, in which 70% of the data were used for training and 30% for testing. The results were graphically expressed through a regression, and the final result was the mean of all results of the regression tree (Breiman, 2001). The model performance was assessed by calculating the correlation between the observed and estimated values through the coefficient of determination (R^2), given by the ratio between the sum of squared regression residuals (SSR) and the total sum of squares (TSS), using the following equation:

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_{i=l}^{n}(\hat{Y}i - \bar{Y})^2}{\sum_{i=l}^{n}(Yi - \bar{Y})^2}$$

Where:

R^2 = coefficient of determination;

Yi = observed value of the dependent variable;

$\hat{Y}i$ = estimated value of the dependent variable, and

$\bar{Y}$ = mean of the dependent variable.

RESULTS AND DISCUSSION

Figure 1B illustrates the selection of the variables using a correlation matrix. Of the 49 predictive variables in the original database, 29 (59%) were eliminated.

Before the selection of variables **After the selection of variables**

(A) **(B)**

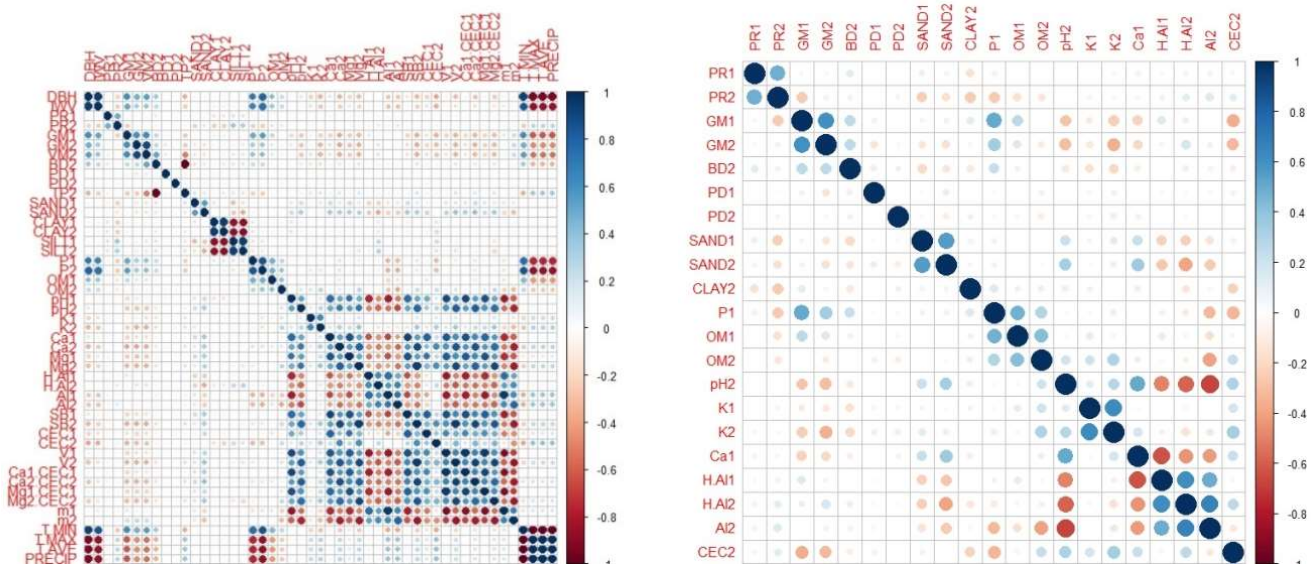

FIGURE 1. (A) Correlation matrix and (B) selected variables through a correlation matrix. Those with null or high correlation variance with each other are eliminated. PR, GM, BD, PD, SAND, CLAY, P, OM, pH, K, Ca, H.Al, Al, CEC represent soil penetration resistance, gravimetric moisture, bulk density, particle density, sand content, clay content, phosphorus content, organic matter, soil pH, potassium content, calcium content, potential acidity, aluminum, cation-exchange capacity, respectively. Number 1 or 2 together with each attribute refer to sampling layers at a soil depth of 0.00–0.20 m and 0.20–0.40 m.

The following dendrometric variables had strong or null correlation and were eliminated from the final model: DBH and VOL; climatic variables: T/min, T/max, T/mean, and Prec; physicochemical variables of the soil at a depth of 0.0–0.20 m: clay, silt, pH, Mg, Al, SB, CEC, V, Ca/CEC, Mg/CEC, and m; and physicochemical variables of the soil at a depth of 0.20–0.40 m: VM, TP, silt, P, Ca, Mg, SB, V, Ca/CEC, Mg/CEC, and m.

In contrast, variables PR1, PR2, GM1, GM2, BD2, PD1, PD2, SAND1, SAND2, CLAY2, P1, OM1, OM2, pH2, K1, K2, Ca1, H.Al1, H.Al2, Al2, and CEC2 were used in the predictive model using the RF algorithm. It must be emphasized that all selected variables are related to the physicochemical attributes of the soil.

Considering that the correlation matrix is a symmetric matrix, that is Corr [X, Y] = Corr [Y, X], Figure 2 shows the values of the correlations between each pair of variables in the upper matrix. The lower matrix refers to the distribution of each pair of variables, while the main diagonal represents the correlation of a variable with itself. The analysis of data dispersion, frequency distribution, and Pearson's correlation coefficient confirmed that null and strongly correlated variables were fully excluded and only those with a correlation coefficient below 70% were retained.

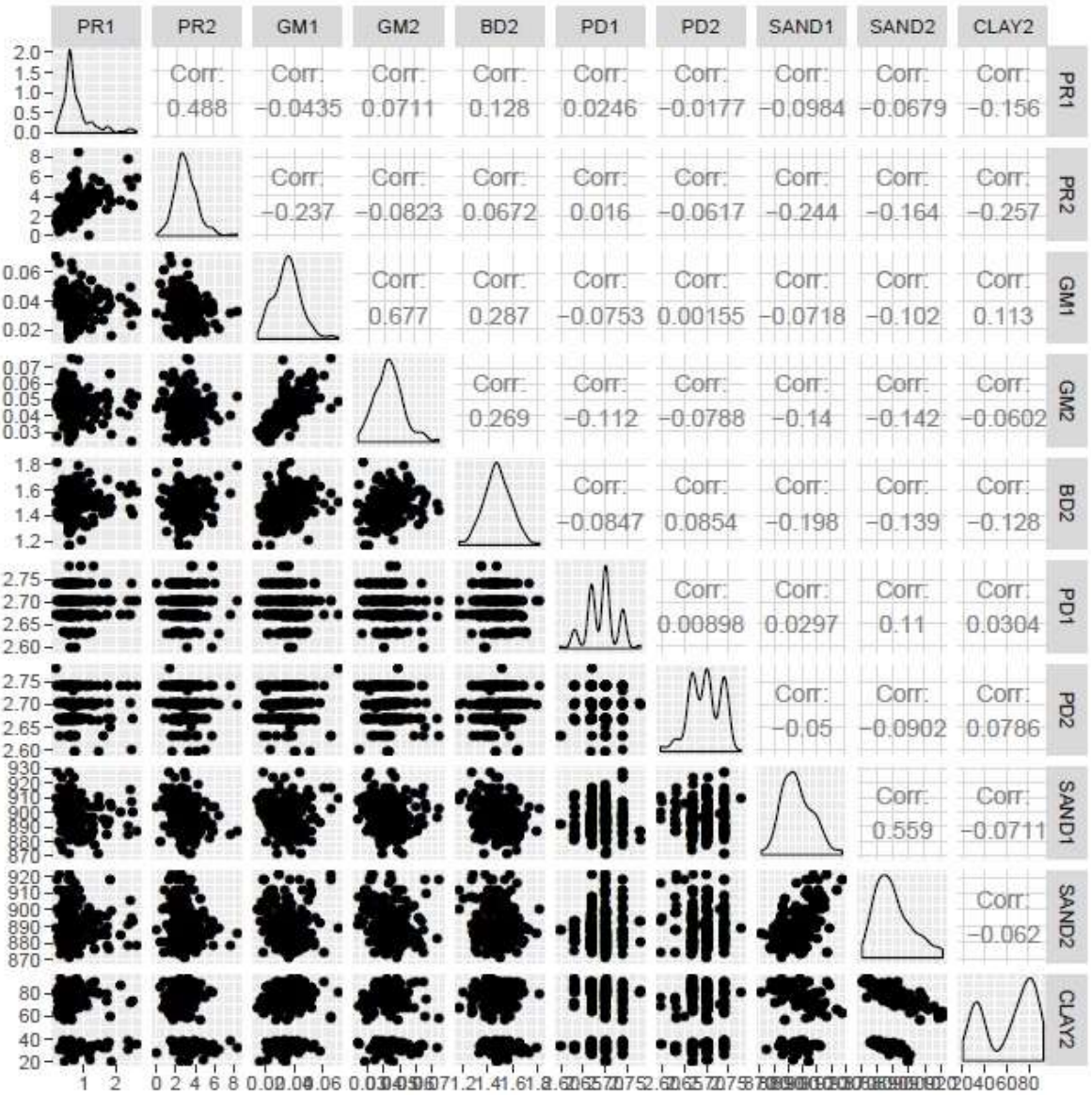

FIGURE 2. Variables selected to be used in the Random Forest (RF) classification model. PR, GM, BD, PD, SAND, and CLAY represent soil penetration resistance, gravimetric moisture, bulk density, particle density, sand content, and clay content, respectively. Numbers 1 or 2 along with each attribute refer to sampling layers at a soil depth of 0.00–0.20 m and 20.0–0.40 m.

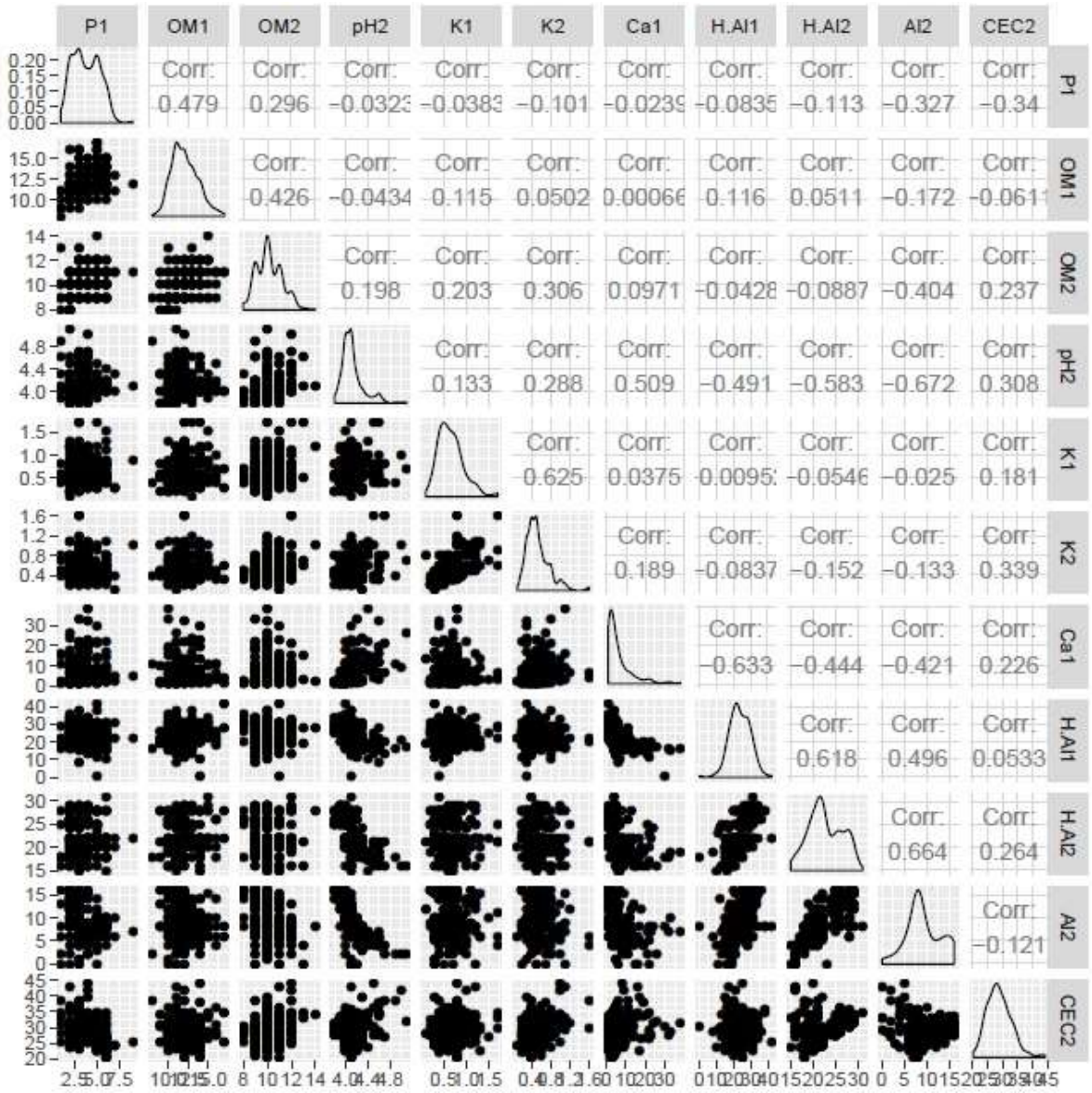

FIGURE 2 (continuation). Variables selected for use in the Random Forest (RF) classification model. P, OM, pH, K, Ca, H.Al, Al, CEC represent phosphorus content, organic matter, soil pH, potassium content, calcium content, potential acidity, aluminum, and cation-exchange capacity, respectively. Number 1 or 2 along with each attribute refer to sampling layers at a soil depth of 0.00–0.20 m and 20.0–0.40 m.

Given the high level of correlation between DBH and certain soil and climatic variables, DBH was discarded in the random variable selection process through the correlation matrix (Figure 1 A and B). However, considering its high influence on eucalyptus height prediction, DBH was reincorporated into the database. This resulted in a hybrid approach, wherein the DBH variable was added to the set of variables previously selected through formal methods.

DBH is the most important variable for predicting eucalyptus height, reaching the maximum value of importance in the predictive process (100% - normalized by the attribute with the highest contribution). This is followed by P1, Al2, and GM1, having degrees of importance ranging between 15 and 19% (Figure 3).

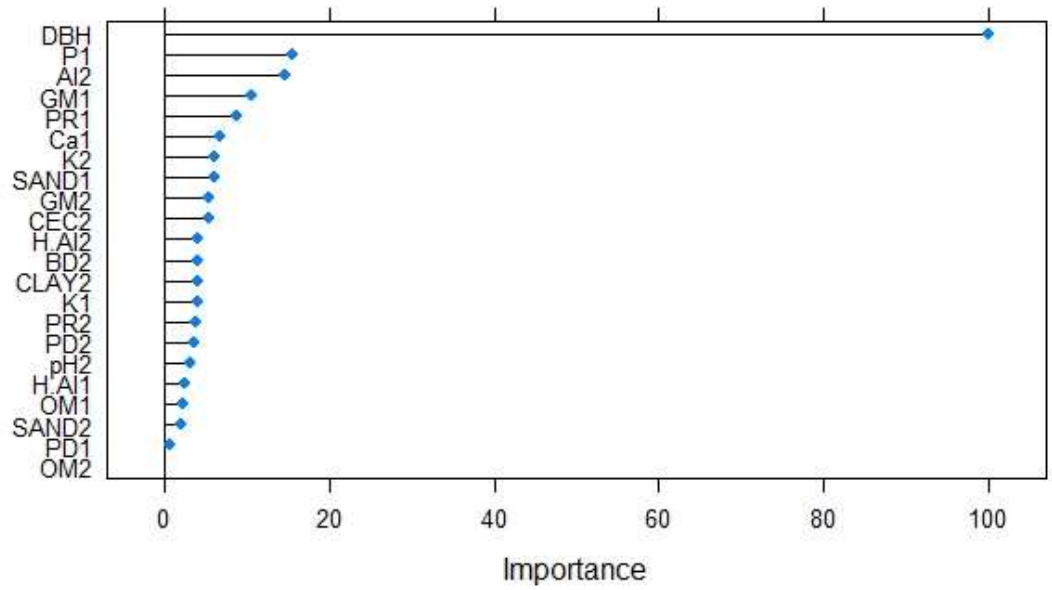

FIGURE 3. Importance of the physicochemical variables of soil used to classify the height of eucalyptus through the Random Forest model.

Finally, the results obtained by validating the RF model showed a high predictive capacity. The correlation between predicted and observed values were 0.98, with R^2 equal to 0.96 (Figure 4). The regression analysis was used to observe the formation of two data clusters, that is above 6 and below 6, which were related to the first and second crop years.

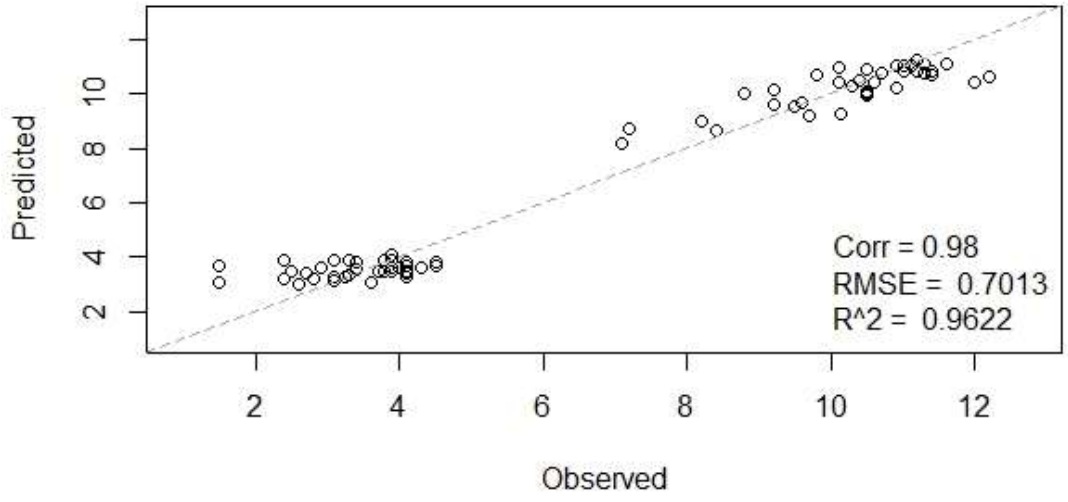

FIGURE 4. Validation of the Random Forest model for predicting the height of Eucalyptus.

The correlation between predicted and observed values was 0.98 and R^2 was 0.96. This revealed significant potential of the RF model in predicting the height of eucalyptus using physicochemical variables of the soil and DBH. The results obtained in this study were superior to those obtained by da Silva et al. (2021), who used different machine learning algorithms (based on spectral indices) to predict the eucalyptus total height, and obtained a correlation coefficient of 0.79.

Several national and international studies have established the efficiency of the RF algorithm in other predictive analyses, with emphasis on its use against other data analysis techniques (Chen et al., 2018; Parmar et al., 2020; Singh et al., 2017). For instance, in a study to predict the basal area and volume in eucalyptus stands using Landsat TM data in Brazil, dos Reis et al. (2018) observed that RF was the best method for multiple linear regression, support vector machine, and artificial neural network. Likewise, to classify the growth of five species of eucalyptus and *Corymbria citriodora*, de Oliveira et al. (2021) reported that the RF algorithm using 24 features was the most accurate (0.76), as compared to other algorithms (0.66).

This study was able to correctly classify data for different height classes, which were induced by the formation of two clusters between the sampling of the first and second crop years. de Oliveira et al. (2021) focused on the classification of eucalyptus species based on their growth (total height and diameter at breast height) and revealed that the development of eucalyptus trees over time induced changes in clusters.

Raudys & Jain (1991) indicated that a small sample size reduced statistical power for pattern recognition. In this

study, the model performed well due to the higher number of observations. The purpose of machine learning algorithms is to learn from the data (Mahesh, 2020). Therefore, the quality of training data has a significant impact on the efficiency, accuracy, and complexity of machine learning tasks (Gupta et al., 2021).

In practice, data are split randomly between 70–30 and 80–20 for training and test datasets, respectively (Dangeti, 2017). This division is necessary to obtain greater reliability of the generated model (Camilo & Silva, 2009). In our study, 70% of the data were used for training and 30% for testing. This was a consistent way to validate the performance of the machine learning model because a portion of the data was separated before developing a model and used only for validation (Vabalas et al., 2019). In addition, the method used to divide the training and test sets was critical. Therefore, representativeness of the original dataset in the samples should be maintained to make the model more efficient and reliable.

When selecting model parameters, datasets with a finite number of training samples require closer attention, including the number of variables used in decision making (Raudys & Jain, 1991). According to Speiser et al. (2019), the prediction efficiency of the model can be improved through variable selection techniques by identifying a subset of predictor variables to be included in a final, simpler model. Although the RF algorithm helps rank variables based on their predictive importance, it is difficult to distinguish relevant from irrelevant variables based on only this ranking (Degenhardt et al., 2019). Our results indicated that the correlation matrix, which is a variable selection method in our study, was highly efficient in selecting a minimum dataset, which was capable of representing the variability of the height of eucalyptus. This was proven by the high values of the correlation coefficient and determination between the predicted and observed data during model validation. These results were consistent with the findings of Everingham et al. (2016), who investigated the accuracy of RF to explain annual variation in sugarcane productivity, and observed that the variable selection process reduced the number of predictor variables in each model and improved the forecast performance.

Soil is an important component for wood production because it is responsible for water and nutrient supply to the plants (Bini et al., 2013). Lima et al. (2010) emphasized that the growth of eucalyptus can strongly influence certain physicochemical characteristics of the soil, namely DBH, P1, Al2, and GM1. Azevedo et al. (2015) used genetic selection in *Eucalyptus camaldulensis* progenies in the savanna area of Mato Grosso State, Brazil, and reported a high correlation (r = 0.72) between DBH and plant height variables. However, Taylor et al. (2016) pointed out the absence of a linear relationship between height and DBH. For most species, variations in height increased with increasing diameter, which led to precision problems in linear regression equations that were designed to estimate the growth of trees.

During early eucalyptus development, phosphorus (P) is directly related to wood productivity; in addition, its highest absorption rates appear during the second year of the tree, that is, the treetop closing (Barros et al., 2000; Melo et al., 2015). Graciano et al. (2006) and Fontes et al. (2013) pointed out that P is the most essential nutrient at an early development stage and for eucalyptus wood productivity. Lack of P in the soil leads to a nutritional imbalance in plants and irreversible falls during the final wood production.

High concentration of aluminum (Al) in the soil reduces the development of roots and diminishes nutrient absorption (Miguel et al., 2010). Although eucalyptus is more tolerant to exchangeable Al than annual crops (Silva et al., 2012), Brazilian forest activity is usually implemented in sandy and low-fertility soils, often with high levels of toxic elements, with emphasis on aluminum (Basso et al., 2007; Guimarães et al., 2015).

As eucalyptus is a fast-growing species, it has high energy expenditure, which leads to higher water consumption (Vital, 2007). Thus, any possible variation in the water supply of the culture reflects directly on plant growth and productivity. Jung et al. (2017) indicated that decreased soil water content reduces the plant water potential, which directly affects its growth in terms of height and diameter, due to reduced cell expansion and cell wall formation. In addition, lower availability of carbohydrates influences the production of plant hormones.

Melo Neto et al. (2017) carried out a study on eucalyptus cultivation and verified a high variability in the mean soil moisture, especially in the superficial layer; homogeneity was observed between 30 and 100 cm of soil depth. They concluded that a steeper moisture reduction in these layers was due to: i) quick response to rain events, ii) demand for soil evaporation being met, and iii) greater exploitation by the root system of the plants.

In addition, the surface layer had higher accumulation of organic matter, which preserved the soil structure and contributed to a higher water flow, both in terms of depth and width. Consequently, this increased the variability of soil moisture (Melo Neto et al., 2017).

Overall, this study provides promising results for forest management purposes as it offers producers and technicians with guidelines to carefully plan the viability of new production areas by allowing the estimation of the height of the culture based exclusively on physicochemical attributes of the soil and identifying areas with high or low production potential. Moreover, the model allows the establishment of predictions in crops previously implemented for the purpose of forest inventories, considering the high cost of direct measurements of eucalyptus height as well as its difficult resolution in the field.

CONCLUSIONS

The random forest (RF) model generated in our study performed well (r = 0.98 and R^2 = 0.96) in predicting the height of eucalyptus using physicochemical variables of the soil and diameter at breast height (DBH). Therefore, this method can be used to support the decision-making process in the management of eucalyptus plantations.

The most important variables to predict the eucalyptus plant height consisted of DBH, phosphorus content (P1), gravimetric moisture (GM1) at a soil depth pf 0.00–0.20 m, and exchangeable aluminum content (Al2) at a soil depth of 0.20–0.40 m.

REFERENCES

Azevedo LPdA, Costa RBd, Martinez DT, Tsukamoto Filho AdA, Brondani GE, Baretta MC, Ajala WV (2015) Genetic selection in Eucalyptus camaldulensis progenies in savanna area of Mato Grosso State, Brazil. Ciência Rural 45(11):2001-2006. DOI: https://doi.org/10.1590/0103-8478cr20131557

Barros NF, Neves JCL, Novais RF (2000) Recomendação de fertilizantes minerais em plantios de eucalipto. In: Gonçalves J.L.M., Benedetti V (ed) Nutrição e fertilização florestal. Piracicaba, IPEF, p269-286

Basso LHM, Lima GPP, Gonçalves AN, Vilhena SMC, Padilha CCF (2007) Efeito do alumínio no conteúdo de poliaminas livres e atividade da fosfatasse ácida durante o crescimento de brotações de Eucalyptus grandis x E. urophylla cultivadas in vitro. Scientia Forestalis 75:9-18

Biau G (2012) Analysis of a random forests model. Journal of Machine Learning Research 13:1063-1095

Bini D, Santos CAd, Bouillet JP, Gonçalves JLdM, Cardoso EJBN (2013) Eucalyptus grandis and Acacia Mangium in monoculture and intercropped plantations: Evolution of soil and litter microbial and chemical attributes during early stages of plant development. Applied Soil Ecology, 63:57-66. DOI: https://doi.org/10.1016/j.apsoil.2012.09.012

Breiman L (2001) Random forests. Machine Learning 45(1):5-32

Camilo CO, Silva JC (2009) Mineração de dados: conceitos, tarefas, métodos e ferramentas. Relatório técnico. Goiania, Instituto de Informática, Universidade Federal de Goiás.

Campos BPF, Silva GFd, Binoti DHB, Mendonça ARd, Leite HG (2016) Predição da altura total de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. Pesquisa Florestal Brasileira 36(88):375-385. DOI: https://doi.org/10.4336/2016.pfb.36.88.1166

Campos JCC, Leite HG (2013) Mensuração florestal: Perguntas e respostas. Viçosa, UFV. 605p

Castro RVO, Araújo RAA, Leite HG, Castro AFNM, Silva A, Pereira RS, Assis Leal FA (2016) Modeling of growth and yield of eucalyptus stands in level of diameter distribution using site index. Revista Árvore 40(1):107-116. DOI: https://doi.org/10.1590/0100-67622016000100012

Chen W, Zhang S, Li R, Shahabi H (2018) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Science of the Total Environment 644(1):1006-1018. DOI: https://doi.org/10.1016/j.scitotenv.2018.06.389

da Silva AKV, Borges MVV, Batista TS, da Silva Junior CA, Furuya DEG, Prado Osco L, Teodoro LPR, Baio FHR, Ramos APM, Gonçalves WN, Marcato Junior J, Teodoro PE, Pistori H (2021) Predicting eucalyptus diameter at breast height and total height with UAV-based spectral indices and machine learning. Forests 12(5):1-13. DOI: https://doi.org/10.3390/f12050582

Dangeti P (2017) Statistics for machine learning. Birmingham: Packt Publishing, 444p

de Oliveira BR, da Silva AAP, Teodoro LPR, de Azevedo GB, Azevedo GTdOS, Baio FHR, Sobrinho RL, da Silva Junior CA, Teodoro PE (2021) Eucalyptus growth recognition using machine learning methods and spectral variables. Forest Ecology and Management 497(1):1-8. DOI: https://doi.org/10.1016/j.foreco.2021.119496

Degenhardt F, Seifert S, Szymczak S (2019) Evaluation of variable selection methods for random forests and omics data sets. Briefings in Bioinformatics 20(2):492-503. DOI: https://doi.org/10.1093/bib/bbx124

dos Reis AA, Carvalho MC, de Mello JM, Gomide LR, Ferraz Filho AC, Acerbi Junior FW (2018) Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data: An assessment of prediction methods. New Zealand Journal of Forestry Science 48(1):1-17. DOI: https://doi.org/10.1186/s40490-017-0108-0

Everingham Y, Sexton J, Skocaj D, Inman-Bamber G (2016) Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for Sustainable Development 36(2):1-9. DOI https://doi.org/10.1007/s13593-016-0364-z

Farhate CVV, Souza ZM, Oliveira SRM, Tavares RLM, Carvalho JLN (2018) Use of data mining techniques to classify soil CO_2 emission induced by crop management in sugarcane field. PLOS ONE 13(3):e0193537. DOI: https://doi.org/10.1371/journal.pone.0193537

Ferraz Filho AC, Mola-Yudego B, Ribeiro A, Scolforo JRS, Loos RA, Scolforo HF (2018) Height-diameter models for *Eucalyptus sp.* plantations in Brazil. CERNE 24(1):9-17. DOI: https://doi.org/10.1590/01047760201824012466

Finger CAG (1992) Fundamentos da biometria florestal. Santa Maria, UFSM/CEPEF/FATEC. 269p

Fontes AG, Gama-Rodrigues AC, Gama-Rodrigues EF (2013) Eficiência nutricional de espécies arbóreas em função da fertilização fosfatada. Pesquisa Florestal Brasileira 33(73):9-17. DOI: https://doi.org/10.4336/2013.pfb.33.73.392

Graciano C, Goya JF, Frangi JL, Guiamet JJ (2006) Fertilization with phosphorus increases soil nitrogen absorption in young plants of Eucalyptus grandis. Forest Ecology and Management 236(2-3):202-210. DOI: https://doi.org/10.1016/j.foreco.2006.09.005

Guimarães CdC, Floriano EP, Vieira FCB (2015) Chemical constraints to initial growth of Eucalyptus saligna in sandy soils of Pampa Gaúcho: A case study. Ciência Rural 45(7):1183-1190. DOI: http://dx.doi.org/10.1590/0103-8478cr20120533

Gupta N, Mujumdar S, Patel H, Masuda S, Panwar N, Bandyopadhyay S, Mehta S, Guttula S, Afzal S, Mittal RS, Munigal V (2021) Data quality for machine learning tasks. Conference on Knowledge Discovery & Data Mining. 1. 4040-4041. DOI: https://doi.org/10.1145/3447548.3470817

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. New York, Springer Series in Statistics Springer.

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: Data mining, inference, and prediction. Science & Business Media.

Hess AF, Braz EM, Thaines F, Mattos PP (2014) Adjustment of the hypsometric relationship for species of Amazon Forest. Ambiência 10(1):21-29

IBÁ (2020) Indústria Brasileira de árvores. Relatório Anual. Available: https://iba.org/datafiles/publicacoes/relatorios/relatorio-iba-2020.pdf. Accessed Nov 22, 2021

IBGE – Instituto Brasileiro de Geografia e Estatística (2019) Produção da extração vegetal e da silvicultura. Rio de Janeiro, IBGE, p1-8

Jung LH, Lopes AS, Oliveira GQ, Oliveira JCL, Fanaya Júnior ED, Brito KRM (2017) Irrigação no desenvolvimento inicial de Eucalyptus urophylla x Eucalyptus grandis e Eucalyptus grandis x Eucalyptus camaldulensis. Ciência Florestal 27(2):655-667. DOI: https://doi.org/10.5902/1980509827750

Soil Survey Staff (2014) Keys to soil taxonomy. Washington, Natural Resources Conservation Service.

Köppen W, Geinge R (1928) Climate der Erde. Gotha, Verlag justus perthes. Wall-map.

Lima CGdR, Carvalho MdPe, Narimatsu KCP, Silva MGd, Queiroz HAd (2010) Atributos físico-químicos de um latossolo do cerrado brasileiro e sua relação com características dendrométricas do eucalipto. Revista brasileira de ciência do Solo 34(1):163-173. DOI: https://doi.org/10.1590/S0100-06832010000100017

Mahesh B (2020) Machine learning algorithms – A review. International Journal of Scientific and Engineering Research 9:381-386

Marçal MFM, Souza ZMd, Tavares RLM, Farhate CVV, Oliveira SRM, Galindo FS (2021) Predictive models to estimate carbon stocks in agroforestry systems. Forests, 12(9):1-15. DOI: https://doi.org/10.3390/f12091240

Martins ER, Binoti MLMS, Leite HG, Binoti DHB, Dutra GC (2016) Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto. Revista brasileira de ciências agrárias - Brazilian Journal of Agricultural Sciences 11(2):117-123. DOI: https://doi.org/10.5039/agraria.v11i2a5373

Melo E, Gonçalves J, Rocha J, Hakamada R, Bazani J, Wenzel A, Arthur J, Borges J, Malheiros R, Lemos C, Ferreira E, Ferraz A (2015) Responses of clonal eucalypt plantations to N, P and K fertilizer application in different edaphoclimatic conditions. Forests 7(12):2-15. DOI: https://doi.org/10.3390/f7010002

Melo Neto JDO, De Mello CR, Silva AMd, De Mello JM, Viola MR, Yanagi SDNM (2017) <b>Temporal stability of soil moisture under effect of three spacings in a eucalyptus stand. Acta Scientiarum. Agronomy 39(3):393-399. DOI: https://doi.org/10.4025/actasciagron.v39i3.32656

Miguel PSB, Gomes FT, Rocha WSD, Carvalho CA, Oliveira AV (2010) Efeitos tóxicos do alumínio no crescimento das plantas: Mecanismos de tolerância, sintomas, efeitos fisiológicos, bioquímicos e controles genéticos. CES Revista 24:1-20

Oliveira TKd, Macedo RLGM, Venturin N, Higashikawa EM (2009) Desempenho silvicultural e produtivo de eucalipto sob diferentes arranjos espaciais em sistema agrossilvipastoril. Pesquisa Florestal Brasileira 60(60):1-10. DOI: https://doi.org/10.4336/2009.pfb.60.01

Parmar A, Katariya R, Patel V (2020) A review on random forest: an ensemble classifier. In: Hemanth J, Fernando X, Lafata P, Baig Z (ed) International Conference on Intelligent Data Communication Technologies and Internet of Things. ICICI 2018. Lecture Notes on Data Engineering and Communications Technologies. Springer. DOI: https://doi.org/10.1007/978-3-030-03146-6_86

Péllico Netto S (2004) Equivalência Volumétrica: Uma nova metodologia para estimativa do volume de árvores. Revista Acadêmica: Ciência Animal 2(1):17-30. DOI: http://dx.doi.org/10.7213/cienciaanimal.v2i1.15003

Pichelli K, Soares S (2019) Perguntas e respostas: Eucalipto. Colombo, Embrapa

R Core Team (2017) R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing. Available: https://www.R-project.org/

Raij B, Andrade JC, Cantarella H, Quaggio JA (2001) Análise química para avaliação da fertilidade de solos tropicais. Campinas, Instituto Agronômico

Raudys SJ, Jain AK (1991) Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(3):252-264

Santos ACA, Silva S, Leite HG, Cruz JPd (2017) Influência da variabilidade edafoclimática no crescimento de clones de eucalipto no Nordeste baiano. Pesquisa Florestal Brasileira 37(91):259-268. DOI: http://dx.doi.org/10.4336/2017.pfb.37.91.1207

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreras JF, Coelho MR, Almeida JA, Araujo Filho JC, Oliveira JB, Cunha TJF (2018) Sistema brasileiro de classificação de solos. Educação em Revista e Ampliada. Brasília, Embrapa.

Scolforo JRS, Maestri R, Ferraz Filho AC, Mello JM, Oliveira AD, Assis AL (2013) Model for site classification of Eucalyptus grandis incorporating climatic variables Dominant H. International Journal of Forestry Research:1-7

Silva MOP, Corrêa GFC, Coelho L, Rabelo PG (2012) Avaliação de dois tratamentos de adubação em plantio de eucalipto clonal em solo arenoso. Biosciênce Journal 28:212-222

Singh B, Sihag P, Singh K (2017) Modelling of impact of water quality on infiltration rate of soil by random forest regression. Modeling Earth Systems and Environment 3(3):999-1004. DOI http://dx.doi.org/10.1007/s40808-017-0347-3

Souza HS, Tsukamoto Filho AA, Vendruscolo DGS, Chaves AGS, Motta AS (2016) Modelos hipsométricos para eucalipto em sistema de integração lavoura-pecuária-floresta. Nativa 4(1):11-14. DOI: http://dx.doi.org/10.14583/2318-7670.v04n01a03

Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Systems with Applications 134:93-101. DOI: https://doi.org/10.1016/j.eswa.2019.05.028

Stolf R, Murakami JH, Brugnaro C, Silva LG, Silva LCFd, Margarido LAC (2014) Penetrômetro de impacto Stolf – Programa computacional de dados em EXCEL-VBA. Revista brasileira de ciência do Solo 38(3):774-782. DOI: https://doi.org/10.1590/S0100-06832014000300009

Tavares RLM, Oliveira SRdM, Barros FMMd, Farhate CVV, Souza ZMd, Scala Junior NL (2018) Prediction of soil CO_2 flux in sugarcane management systems using the random forest approach. Scientia Agricola 75(4):281-287. DOI: https://doi.org/10.1590/1678-992X-2017-0095

Taylor JE, Ellis MV, Rayner L, Ross KA (2016) Variability in allometric relationships for temperate woodland Eucalyptus trees. Forest Ecology and Management 360(15):122-132. DOI: https://doi.org/10.1016/j.foreco.2015.10.031

Teixeira PC, Donagemma GK, Fontana A, Teixeira WG (2017) Manual de métodos de análise de solos. Brasília, Embrapa. 573p

Vabalas A, Gowen E, Poliakoff E, Casson AJ (2019) Machine learning algorithm validation with a limited sample size. PLOS ONE 14(11):e0224365. https://doi.org/10.1371/journal.pone.0224365

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. Revista brasileira de biometria 33(4):556-569. DOI: http://dx.doi.org/10.13140/RG.2.1.1742.5684

Vital MHF (2007) Impacto ambiental de florestas de eucalipto. Revista BNDES 14(28):235-276

Wang L, Zhou X, Zhu X, Dong Z, Guo W (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal 4(3):212-219. DOI: https://doi.org/10.1016/j.cj.2016.01.008