DANIEL DANTAS

MODELING AND SPATIAL ANALYSIS OF CARBON STOCK AND FOREST ATTRIBUTES USING MIXED-EFFECTS MODELS AND ARTIFICIAL INTELLIGENCE TECHNIQUES

LAVRAS - MG

2022

DANIEL DANTAS

MODELING AND SPATIAL ANALYSIS OF CARBON STOCK AND FOREST ATTRIBUTES USING MIXED-EFFECTS MODELS AND ARTIFICIAL INTELLIGENCE TECHNIQUES

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Doutor.

Prof. Dr. Natalino Calegario

Orientador

LAVRAS - MG

2022

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Dantas, Daniel.
Modeling and spatial analysis of carbon stock andforest attributes using mixed-effects models and artificial intelligencetechniques / Daniel Dantas. - 2022.
110 p. : il.

Orientador(a): Natalino Calegario.

Tese (doutorado) - Universidade Federal de Lavras, 2022.
Bibliografia.

1. Manejo Florestal. 2. Biometria Florestal. 3. Inteligência artificial. I. Calegario, Natalino. II. Título.

DANIEL DANTAS

MODELING AND SPATIAL ANALYSIS OF CARBON STOCK AND FOREST ATTRIBUTES USING MIXED EFFECTS-MODELS AND ARTIFICIAL INTELLIGENCE TECHNIQUES

MODELAGEM E ANÁLISE ESPACIAL DO ESTOQUE DE CARBONO E DE ATRIBUTOS FLORESTAIS POR MEIO DE MODELOS DE EFEITOS MISTOS E TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Doutor.

APROVADA em 13 de julho de 2022.

Dra. Gabriela Paranhos Barbosa – UNIFESSPA

Dra. Marcela de Castro Nunes Santos Terra - UFLA

Dr. Paulo Ricardo Gherardi Hein - UFLA

Dr. Elliezer de Almeida Melo - IFGOIANO

Dr. Samuel José Silva Soares da Rocha - UFLA

Prof. Dr. Natalino Calegario

Orientador

LAVRAS - MG

2022

À minha mãe, Normalice.

Ao meu pai, Alisson.

Aos meus irmãos, Isadora e Davi.

Aos meus avós, Adelice, Oscar (in memorian), Eramita e Darci.

Dedico.

AGRADECIMENTOS

A Deus, por permitir que tudo isso se realizasse, me guiando, protegendo, me dando forças e sendo presença constante na minha vida. “Sem Ti, Senhor, o agora não seria”.

Aos meus pais, Normalice Silva Dantas e Alisson Antunes Dantas, e aos meus irmãos, Isadora Dantas e Davi Lucas Ferreira do Nascimento, por todo amor, compreensão e apoio incondicional. Obrigado pela educação, por acreditarem e confiarem em mim em todos os momentos, isso foi fundamental para que eu conseguisse ir em busca dos meus objetivos. A vocês, meu amor e gratidão eternos!

À toda a minha família, especialmente aos meus avós, que tanto me apoiaram, incentivaram e amaram.

A Gladston Durães Vieira, pelo companheirismo e pelo apoio tão importantes.

Ao meu orientador, professor e amigo Natalino Calegario, pelos muitos ensinamentos transmitidos, pela excelente convivência, paciência e orientação. Muito obrigado pela humanidade e pela compreensão!

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais pelo apoio financeiro.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Engenharia Florestal, pela oportunidade de realizar esta importante etapa da minha vida; aos seus professores, pela minha formação e todo aprendizado adquirido ao longo deste período de mestrado e doutorado. Agradeço em especial à professora Dulcinéia, pela amizade.

À banca examinadora da defesa desta Tese, Dra. Gabriela Paranhos Barbosa, Dra. Marcela de Castro Nunes Santos Terra, Dr. Paulo Ricardo Gherardi Hein, Dr. Elliezer de Almeida Melo e Dr. Samuel José Silva Soares da Rocha, pela participação, pela disponibilidade e pelas contribuições.

À Marcela de Castro Nunes Santos Terra pela amizade, pelas conversas, parceria, apoio e incentivo em diversos e importantes momentos.

Aos funcionários do Departamento de Ciências Florestais, em especial à Chica, por todo carinho e amizade.

Aos colegas do Laboratório de planejamento e manejo intensivo de florestas.

A todos os meus colegas de curso, pelo companheirismo, amizade e pela interação produtiva.

A todos os meus amigos: os velhos, os novos, os distantes e os mais próximos, que, sem citar nomes, sabem a importância que têm na minha vida.

Aos meus grandes amigos da UFLA: Douglas, Lucas, Taiana, Renato, Anny, Isáira, Lorena, Luiz, Vitor, Celina e Laís, por tudo que vivemos e pela cumplicidade ao longo deste período.

A todos que participaram desta minha jornada e contribuíram, direta ou indiretamente, para a conclusão deste trabalho:

O meu muito obrigado!

RESUMO GERAL

Florestas proveem inúmeros serviços ecossistêmicos, como regulação de ciclos biogeoquímicos, controle de poluição, fornecimento de alimentos e o sequestro e estocagem de carbono atmosférico. Esses serviços são cruciais, pois atuam diretamente na mitigação do aquecimento global, sendo de importância estratégica na amenização das mudanças climáticas. Nesse contexto, a quantificação do estoque de carbono presente nos mais variados tipos de florestas, constitui uma ferramenta importante de monitoramento desse serviço ecossistêmico. A estimativa de estoque de carbono por métodos indiretos faz uso de técnicas de modelagem e simulação. Historicamente, a modelagem de atributos florestais se apoiou em abordagens fundamentadas em modelos estatísticos. Essas abordagens dividem hoje espaço com abordagens computacionais de inteligência artificial/aprendizagem de máquina, como redes neurais artificiais, máquinas de vetores de suporte, árvores de decisão, dentre outras, as quais vêm ganhando espaço como ferramentas de análise de dados florestais, modelagem, estimativa de variáveis e prognose da produção. Essas ferramentas têm proporcionado ganhos na qualidade das estimativas e predições. Neste trabalho foram analisadas as distribuições espaciais do estoque de carbono em uma floresta tropical e avaliados os desempenhos de modelos extraídos de técnicas de inteligência artificial para modelar o estoque de carbono em florestas tropicais; além do uso de inteligência artificial e modelos mistos com adoção de estrutura na matriz de variância e covariância para estimativas volumétricas. O estoque total de carbono estimado foi de 267,52 $Mg \cdot ha^{-1}$, sendo 35,23% na biomassa aérea, 63,22% no solo e 1,54% nas raízes. No solo, repetiu-se um padrão espacial do estoque de carbono em todas as profundidades analisadas, com redução da quantidade de carbono à medida que a profundidade aumentava. O estoque de carbono das árvores seguiu o mesmo padrão espacial do solo, indicando uma relação entre essas variáveis. Nas raízes finas, o estoque de carbono diminuiu com o aumento da profundidade, mas o gradiente espacial não seguiu o mesmo padrão do solo e das árvores, o que indicou que o estoque de carbono radicular foi influenciado por outros fatores. As técnicas funcionaram satisfatoriamente na modelagem, com distribuições homogêneas e baixa dispersão dos resíduos. Os critérios de análise de qualidade indicaram o desempenho superior do modelo misto com estrutura Huynh-Feldt da matriz de variância e covariância, que apresentou uma diminuição do erro relativo médio de 13,52% para 2,80%, enquanto as técnicas de aprendizado de máquina tiveram valores de erro de 6,77%. (SVM) e 5,81% (RNA). Este estudo confirma que, embora os modelos de efeitos fixos sejam amplamente utilizados no setor florestal brasileiro, existem métodos mais eficazes para a modelagem de variáveis dendrométricas.

Palavras-chave: Redes neurais artificiais. Máquina de vetor de suporte. Modelos não-lineares. Modelos de efeitos mistos. Geoestatística.

GENERAL ABSTRACT

Forests provide numerous ecosystem services, such as regulation of biogeochemical cycles, pollution control, food supply and the sequestration and storage of atmospheric carbon. These services are crucial, as they act directly in the mitigation of global warming and are of strategic importance in mitigating climate change. In this context, the quantification of the carbon stock present in the most varied types of forests constitutes an important tool for monitoring this ecosystem service. The estimation of carbon stock by indirect methods makes use of modeling and simulation techniques. Historically, the modeling of forest attributes has relied on approaches based on statistical models. These approaches now share space with computational approaches of artificial intelligence/machine learning, such as artificial neural networks, support vector machines, decision trees, among others, which have been gaining ground as tools for forest data analysis, modeling, estimation of variables and production prognosis. These tools have provided gains in the quality of estimates and predictions. In this work, we analyzed the spatial distributions of the carbon stock in a tropical forest and evaluated the performance of models extracted from artificial intelligence techniques to model the carbon stock in tropical forests; in addition to the use of artificial intelligence and mixed models with the adoption of a structure in the variance and covariance matrix for volumetric estimates. The total estimated carbon stock was 267.52 $Mg \cdot ha^{-1}$, of which 35.23% was in aboveground biomass, 63.22% in soil, and 1.54% in roots. In the soil, a spatial pattern of the carbon stock was repeated at all depths analyzed, with a reduction in the amount of carbon as the depth increased. The carbon stock of the trees followed the same spatial pattern as the soil, indicating a relationship between these variables. In the fine roots, the carbon stock decreased with increasing depth, but the spatial gradient did not follow the same pattern as the soil and trees, which indicated that the root carbon stock was most likely influenced by other factors. The techniques performed satisfactorily in modeling, with homogeneous distributions and low dispersion of residuals. The quality analysis criteria indicated the superior performance of the mixed model with a Huynh-Feldt structure of the variance and covariance matrix, which showed a decrease in mean relative error from 13.52% to 2.80%, whereas machine learning techniques had error values of 6.77% (SVM) and 5.81% (ANN). This study confirms that although fixed-effects models are widely used in the Brazilian forest sector, there are more effective methods for modeling dendrometric variables.

Keywords: Artificial neural networks. Support vector machine. Nonlinear models. Mixed models. Geostatistics.

SUMÁRIO

FIRST PART

1 INTRODUCTION

The increase in CO_2 in the atmosphere has accelerated global climate change (SCHEFFER et al., 2006). Scientists are increasingly engaged in investigating possibilities for mitigating the effects of such changes. Forests, especially young forests, are natural CO_2 sinks, as plants carry out carbon biosynthesis in the photosynthesis process (PUGH et al., 2019). Thus, studying how forests work, the factors and mechanisms involved in carbon sequestration and their carbon storage capacity is crucial to understanding this important environmental service provided by forests.

Many studies have focused on generating large-scale results and products, with maps of biomass/carbon distribution in vegetation for large regions, using extensive databases and, often, sensing resources remote control and sophisticated spatialization methods (SILVEIRA et al., 2019a). Other works have focused on investigating the environmental and ecological determinants of biomass in different forests (SORIANO-LUNA et al., 2018) or the threats that human activity poses to plant biomass (SILVEIRA et al., 2019b). However, in the face of so many products and generalizations, the demand for local studies is growing, which investigate specific variations of biomass and carbon stock in forests and offer insights and refinements for the design of general standards and, at the same time, data for the validation of products in large scale (DUNCANSON et al., 2019).

Therefore, local studies that investigate the determinants of biomass in forests are essential. However, the methods used in these investigations must be appropriate, modern and robust, so that the inferences are valid and actually represent scientific gains. In forest modeling, historically, approaches based on statistical models predominated (MELO et al., 2017). Recently, however, it is observed that computational approaches of artificial intelligence/machine learning have gained space as a tool for analyzing forest data (SILVEIRA et al., 2019a). These tools have provided gains in the quality of estimates and predictions (VENDRUSCOLO et al., 2015; MARTINS et al., 2016).

The demand for forest products is increasing, especially due to the growing world population and the COVID 19 pandemic. In this sense, the planning and administration of forest plantations are important and their accurate and efficient evaluation is decisive for adequate forest management actions, meeting the technical and economic objectives of the companies.

The information on the volume of wood in a forest stand is essential for forest management, considering that a large part of the production processes in a company is dependent on production data. This information is obtained through a forest inventory, which consists of determining a sampling method, allocating plots in the area and measuring the variables of interest.

Forest inventories have a high cost due to the time required for surveys, allocation of plots in the area and measurement of a large number of trees. In addition, it is prone to errors. Sampling errors are controllable and are influenced, for example, by the sampling method and sample size. On the other hand, non-sampling errors are difficult to identify, and there may be several sources of this second type of error, including the use of inadequate equations to estimate the height and volume of individual trees.

The forestry sector is constantly searching for techniques that make it possible to achieve excellence in estimating wood production, optimizing processes and maximizing profits. Also noteworthy is the existence of demand for a sampling intensity capable of reducing the number of plots per hectare and at the same time providing a good representation of the population, so that this reduction does not interfere negatively in the quality of the estimates.

In this context, the measurement and data processing procedures are fundamental elements and deserve attention, being necessary studies on methods that provide accurate estimates of the height variable, which are the basis for determining the volume in an area, and then provide subsidies for the decision-making by the forest manager in conducting the forest inventory and processing the data.

This Thesis is divided into two parts. The first seeks to situate the reader about the objectives of this work and to make a theoretical approach to the techniques used, including important terms that were used throughout the second part. The second part was organized into three articles.

1.1 Objective

a) To analyze the spatial distribution and employ machine learning algorithms to model the biomass (carbon stock) of a secondary Semideciduous Seasonal Forest and the total volume of eucalyptus trees.

1.1.1 Specific Objectives

a) Determine the aboveground (tree trunks) and belowground (soil and fine roots, at four depths) carbon stocks in a tropical forest in Brazil and to evaluate the spatial patterns of carbon in the three different compartments and in the total stock;
b) Evaluate the performance of machine learning techniques: support vector machine and to propose a new nonlinear model extracted from the training of an artificial neural network in the modeling the above ground carbon stock in a secondary semideciduous forest;
c) Assess the quality of the volumetric estimation of *Eucalyptus* spp. trees using a mixed-effects model, artificial neural network and support-vector machine.

2 LITERATURE REVIEW

2.1 Geostatistical Model

The need to develop and use statistical tools, capable of characterizing the structure of spatial variability of a given phenomenon, motivated the emergence of the theory of regionalized variables or, simply, geostatistics (NOGUEIRA, 2013). According to Couto et al. (1997), geostatistics is very useful for characterizing and mapping the spatial variation of systems properties.

According to Landim (2003), the roots of geostatistics are in the mining industry, in the mid-1950s, during research developed to calculate mineral reserves in South Africa by Daniel G. Krige and the statistician H. S. Sichel, for gold mining data. It was concluded that the data variances had a structure dependent on the sampling, and that the variances obtained through the classic sampling approach did not make sense if the distances between the samples were not considered. Based on these observations, Matheron (1963) developed the Regionalized Variables Theory, defined as variables whose values are somehow related to the spatial position they occupy, that is, that vary from one place to another with a certain continuity.

This variable operates in a geometric space in which it was defined and where its variation will be studied, presenting location, continuity and anisotropy, which are linked to the natural phenomenon they represent (JOURNEL, 2003). Nogueira (2013) states that regionalized variables have a double characteristic: randomness, since their observed numerical values can vary considerably from one point to another in space, and spatiality, since these values are not entirely independent, showing an apparent spatial continuity, in which they can be captured and modeled mathematically by geostatistical techniques.

According to Mello (2004), the regionalized variables theory assumes that each data $Z(x_i)$ is modeled with a random variable that can be expressed by the sum of three components: a structural component, associated with a constant mean value or a constant trend; a random component, spatially correlated; and a constant random error.

If x represents a position in one, two or three dimensions, then the value of the variable Z in x is given by:

$$Z(x_i) = M(x) + \varepsilon'_x + \varepsilon_i \qquad (1)$$

where $M(x)$ is a deterministic function that describes the structural component Z at x; ε'_x is a stochastic term, which varies locally and spatially depends on $M(x)$; and ε_i is an uncorrelated random noise, with a normal distribution with a mean zero and variance σ^2, acting on a stochastic term ε'_x (MELLO, 2004).

For Vieira (2000), geostatistics aims to identify, in the apparent disorder between the samples, a measure of the spatial correction between the attributes, allowing to evaluate patterns of adequate samples and to carry out estimates of unsampled local values based on some known values in the population, a technique called by Kriging. For Silva et al. (2010), this technique is a geostatistics interpolation method, which uses spatial dependence expressed in the variogram between neighboring samples to estimate values at any position within the field, without trend and with minimal variance.

According to Isaaks and Srivastava (1989), geostatistics is based on the hope that, on average, samples that are close in time and space will be more similar to each other than those that are far away. It is a spatial analysis tool that aims to estimate unsampled values, using probabilistic models associated with the spatial location of sampled points, thus enabling an analysis of the variability between observed values (MELLO, 2004).

Geostatistics has been developing in recent decades and presents itself as a reliable tool for estimating data for the vast majority of natural phenomena, since it is distinguished by capturing the continuity pattern of geopositioned data (SILVA et al., 2010).

The concern with more efficient techniques and estimators, which determine the characteristics of an area, or a portion of the earth's surface, are study targets in several areas of knowledge (WOJCIECHOWSKI et al., 2009).

The agglomeration of statistical methods presents in geostatistics help to describe and analyze data that include simultaneous measurements of many correlated variables, explaining the complexity of a biological system (NEVES, 2013), as they are very useful in the characterization and mapping of the spatial variation of the properties of the systems (COUTO et al., 1997). The techniques of geostatistics and classical statistics enable the storage, treatment, superposition, analysis and visualization, together, of this spatial information, in the form of digital maps, allowing quick, agile and more efficient decision making, minimizing costs, optimizing resources and productive activities.

2.1.1 Semivariogram

According to Correia (2013), the semivariogram is a geostatistical tool that allows verifying and modeling the spatial dependence of a variable, as it allows the quantitative representation of the variation of a regionalized phenomenon in space. According to Salviano (1996), the semivariogram analyzes the degree of spatial dependence between samples within an experimental field, and defines parameters necessary for estimating values for non-sampled locations, using the kriging technique. Zimback (2011) states that it is possible to investigate the magnitude of the correlation between samples and their similarity or not with distance.

For Amaral (2010), the semivariogram mathematically describes the relationship between the variance of pairs of observations (points) and the distance that separates these observations (h), in which the semivariogram depends on the distance *h* between *x* and *x+h* (VIEIRA, 2000).

According to Maciel (2012), the semivariance estimator $\hat{\gamma}(h)$ (2) is equal to half the arithmetic mean of the differences between pairs of experimental values squared, at all points separated by the distance *h*. In this estimator, *N(h)* represents the number of pairs separated by a distance *h*, *Z(x)* is the value of the variable observed at point *x* and *Z(x+h)* is the value of the variable measured at a distance *h* from the point *x* (JOURNEL; HUIJBREGTS, 1978).

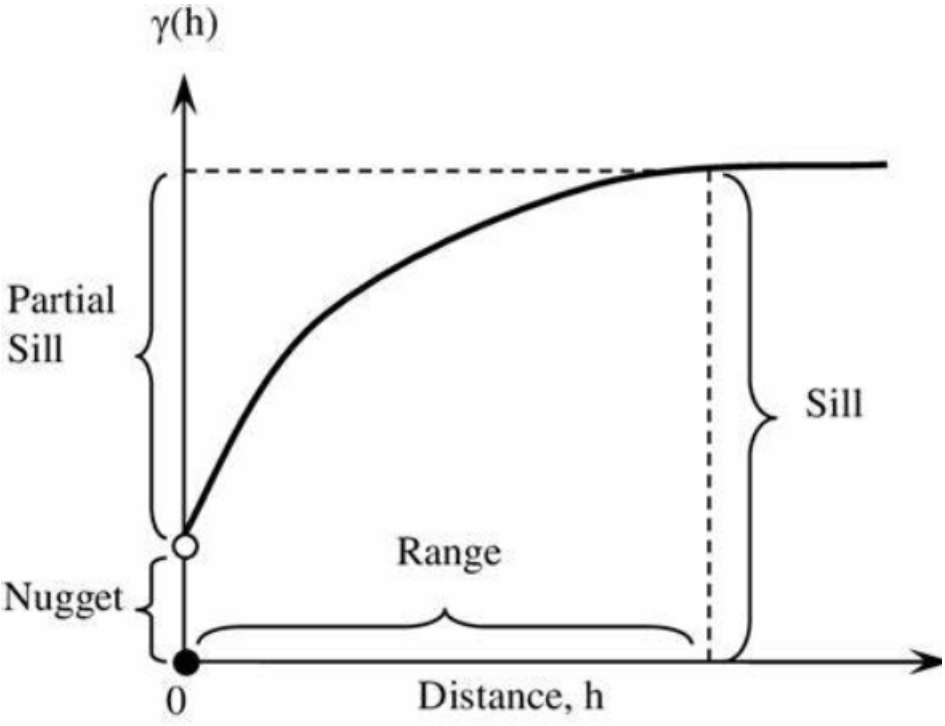
$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(x_i) - z(x_i + h) \right]^2 \qquad (2)$$

where: $\gamma(h)$ is the estimated semivariance; *N(h)* is the number of pairs of measured values separated by a vector $h[Z(x_i) - Z(x_i + h)]$.

According to Vieira (2000), the characteristics of an ideal semivariogram occur when the difference $[Z(x_i) - Z(x_i + h)]$ decreases as the value of *h* decreases. When the distance is zero ($h = 0$), the semivariance value will be zero ($\gamma = 0$). However, this does not occur, since as the distance decreases, the semivariance value tends to a positive value known as the nugget effect (MACIEL, 2012).

According to Cressie (1993), the semivariogram parameters (FIGURE 1) that describes the spatial dependence are: range, threshold, nugget effect and contribution.

Figure 1 - Example of semivariogram.



Source: Camargo (2001).

on what:

Range (a): represents the distance in meters at which the samples are spatially correlated (WOJCIECHOWKI, 2009). It corresponds to the concept of the zone of influence or spatial dependence of a sample, marking the distance from which the samples become independent.

Threshold (c): is the value of the semivariogram corresponding to its range (AMARAL, 2010). It is an estimate of the population variance, and from that point on onwards, it is considered that there is no more spatial dependence between the samples, since the variance of the difference between pairs of samples becomes invariant with distance (AMARAL, 2010).

Nugget effect (Co): represents the discontinuity of the semivariogram for scales smaller than the smallest distance between samples (XAVIER, 2010). Part of this discontinuity may also be due to measurement errors, however, it is impossible to quantify whether the greatest contribution comes from measurement errors or small-scale variability not captured by sampling (WOJCIECHOWKI, 2009).

Contribution (C_1): is the difference between the Level (C) and the Nugget Effect (Co) (CAMARGO, 2001).

Geostatistical estimators can be classified into three types: observed or experimental (obtained from field samples), true (real, but unknown) and theoretical (reference, used to adjust

the model) (VIEIRA, 2000).

The experimental semivariogram is fitted to the curve that provides the maximum possible correlation with the plotted points, while the fitted model is called theoretical (WOJCIECHOWKI, 2009). According to Vieira (2000), it is important that the adjusted semivariogram model represents the trend of the calculated semivariances, in such a way that the estimates obtained from Kriging are more accurate and, therefore, more reliable.

Zimback (2001) states that the sensitivity of semivariograms, in order to detect the spatial variability of the samples, is directly linked to the best fit of the experimental data to the theoretical model of the variogram. Therefore, among the geostatistical tools, the central point is the semivariogram, since it describes quantitative variations and interpolates the unsampled points. It is important to choose the appropriate model to estimate the semivariogram since each model presents different values for range, nugget effect and variance, which are critical parameters for the quality of kriging (RIBEIRO JUNIOR, 1995).

2.1.1.1 Theoretical models of semivariogram

The sensitivity of semivariograms to detect the spatial variability of samples is directly linked to the best fit of the theoretical model to the empirical semivariogram. According to Journel and Huijbregts (2003), the theoretical models of semivariograms are superimposed on the sequence of points obtained in the empirical semivariogram, so that the curve that best fits the points represents the magnitude. Wojciechowki (2009) states that among the basic models, called isotropic by Isaaks and Srivastava (1989), are: spherical models (3); exponential (4); and Gaussian (5).

$$\gamma_{\text{sph}}(h) = C_0 + C_1\left[\frac{3}{2}\left(\frac{h}{\phi}\right) - \left(\frac{1}{2}\left(\frac{|h|}{\phi}\right)^3\right)\right] \quad (3)$$

$$\gamma_{\text{exp}}(h) = C_0 + C_1\left[1 - \exp\left(-3\left(\frac{|h|}{\phi}\right)\right)\right] \quad (4)$$

$$\gamma_{\text{gauss}}(h) = C_0 + C_1\left[1 - \exp\left(-3\left(\frac{|h|}{\phi}\right)^2\right)\right] \quad (5)$$

where: γ = semivariance of the variable of interest; C_0 = nugget effect; C = threshold; ϕ = range; h = distance; exp = exponential.

2.1.2 Spatial estimation method: Kriging

Kriging is the interpolation method of geostatistics, which uses the spatial dependence expressed in the semivariogram between neighboring samples to estimate values at any position within the field, without trend and with minimal variance (SILVA et al., 2010). These two characteristics make kriging an optimal interpolator (ISAAKS; SRIVASTAVA, 1989). According to Correia (2013), this spatial inference method estimates data at unsampled points from sampled points, considering the spatial dependence structure of the phenomenon under study.

According to Soares (2009), there are several interpolators to estimate values at unsampled points, such as simple kriging, ordinary kriging, universal kriging, indicator kriging. Akhavan et al. (2010) state that these methods constitute the anchor of geostatistical procedures.

Maciel (2012) states that in simple kriging the local averages are constant and with a value similar to the population average, that is, the population average is used together with neighboring points to estimate the attribute. Ordinary Kriging, according to Motomiya et al. (2006), is the data interpolation method, which uses the spatial dependence between neighboring samples, expressed in the semivariogram, to estimate values at any position within the analyzed space, to which the semivariogram model was fitted, without bias and with minimum variance. In short, this estimator is nothing more than a weighted average of the observed values. While universal kriging is an efficient method for cases where stationarity does not occur, that is, the process has a tendency. In ordinary kriging, the population mean is disregarded, and only neighboring points are used for the estimate, in addition to considering the stationarity of the data (MACIEL, 2012).

Thus, kriging is a technique used in geostatistics with the objective of estimating variable values for places where they were not measured from adjacent interdependent values, emphasizing that for this tool to be used, it is necessary that there is a spatial dependence, defined by the variogram (SALVIANO, 1996).

2.2 Mixed models

In a modeling process, an attempt is made to explain the variability of a dependent variable through the effects that are attributed to another set of variables, called independent. These effects can be considered as fixed, which are constants to be estimated, or random, which are occurrences of a random variable with a known probability distribution.

According to Searle (1987), linear models in the parameters have at least one random effect, known as experimental error. If a model has all other fixed components, it is called a fixed model; if, however, all other factors are random, the model is called random; when the model presents both random and fixed effects, it is called a mixed model. Random effects can be defined as those that, when taking a random sample from an infinite set of levels, the statistical inferences are valid for the entire population and the fixed effects are those attributed to a finite set of levels of treatments under test in the experiment (SEARLE; CASELLA; MCCULLOCH, 1992).

There are situations in which the data have behavior dependent on each other, that is, one observation depends on another. In general, they are called pooled data, which include longitudinal data, repeated measures, and others. In this sense, the usual models, both linear and non-linear, become inadequate, since the assumption of independence of observations is compromised (BARBOSA, 2009).

The adjustment techniques have allowed more and more flexibility to models with the presence of grouped data, taking into account the possible correlation between the observations, making it possible to model the correlations between and within groups. Such models are called mixed, in which the random part is incorporated through a variance and covariance matrix (LITTEL et al., 1996).

2.2.1 Mixed linear models

Mixed linear models are based on three aspects: adjustments and hypothesis testing on fixed-effect parameters; prediction of random effect parameters; and estimation of variance components. The advantage of this method is strongly associated with the examination of random effects and the modeling of the variance and covariance structure (LAIRD; WARE, 1982; CAMARINHA FILHO, 2003; CUNHA et al., 2013).

The matrix form approach of mixed linear models can be represented, according to Mrode (2015), by:

$$y = X\beta + Zb + \varepsilon \quad (6)$$

where: *y* is a vector $n \times 1$ of observations of the response variable or dependent variable, where n is the number of observations; X is an $n \times p$ matrix of fixed effects incidence, where p is the number of fixed effects parameters; β is a $p \times 1$ vector of fixed effects parameters; Z is an $n \times q$ matrix of the incidence of random effects, where q is the number of parameters of random effects; b is a vector $q \times 1$ of parameters of the random effects; ε is an $n \times 1$ vector of random errors, with:

$$b \sim N\,(0, G)\; e\; \varepsilon \sim N\,(0, R) \quad (7)$$

where G is the variance-covariances random effects matrix and R is the variance-covariances residual matrix, which can assume different structures.

.

In this way, we have

$$y \sim N\,(X\beta, ZGZ' + R) \quad (8)$$

2.2.2 Nonlinear mixed models

Nonlinear mixed effects models have received great attention due to their ability to work with data that have repeated and unbalanced measures. In this class of models, some parameters are allowed to be fixed while others show variations between different groups, through random effects. Furthermore, these models are able to model the existing correlation between observations, within the same group, which is frequently observed in this type of data. The classic regression models assume independence from the observations, therefore, they are not suitable for clustered data, in addition to incorrectly estimating the standard errors in cases where clustering occurs (LITTELL; HENRY; AMMERMAN, 1998).

According to Lindstrom and Bates (1990), the matrix form of nonlinear mixed models can be represented by:

$$y_{ij} = f\left(\phi_{ij}, V_{ij}\right) + \varepsilon_{ij}, i = 1, \ldots, m;\; j = 1, \ldots, n_i. \quad (9)$$

where: f is a general, real and differentiable function of a specific group of vectors of the parameter ϕ_{ij} and a vector of covariants v_{ij} ; and ε_{ij} is the random error normally distributed within the groups; m is the number of groups; n_i is the number of observations in the ith group; n_j is the number of observations in the jth group.

The parameter vector varies from individual to individual. In a second stage, the vector ϕ_{ij} can be expressed by equation 10.

$$\phi_{ij} = A_{ij}\beta_i + B_{ij}\beta_i + B_{ij}\beta_{ij}, \beta_i \approx N(0, \Psi_1)\ e\ \ \beta_{ij} \approx N(0, \Psi_2) \qquad (10)$$

where: $\boldsymbol{\beta}$ is a vector *(p* x *1)* of fixed effects; B_i is a vector *(q_1* x *1)* of independently distributed random effects with a covariance-variance matrix ψ_1; B_{ij} is a vector *(q_2* x *1)* of independently distributed random effects with a covariance-variance matrix ψ_1, and assumed to be independent of the random effects of the first level; A_{ij} and B_{ij} are incidence matrices; ε_{ij} , within groups, are independently distributed and are independent of random effects.

2.2.3 Variance and covariance structures

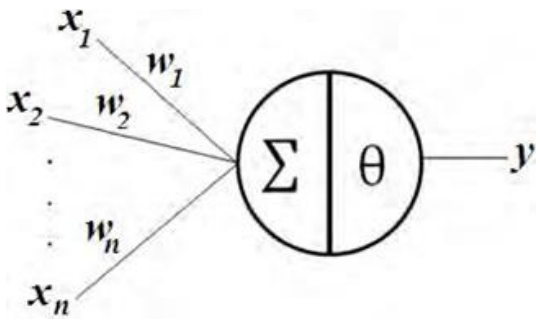
The process of defining the appropriate variance and covariance structure is important in the modeling of mixed models, as it aims to select a parsimonious structure that explains well the variability of the data and the correlation between measures with a small number of parameters. This selection can directly affect parameter estimates, standard errors of fixed and random effects, diagnoses and inferences (TORAL; ALENCAR; FREITAS, 2006) and is dependent on the way the data are structured, empirical information and computational availability. The modeling of these structures for the mixed linear and mixed nonlinear model are similar and more details can be found in the literature, as in Davidian and Giltinan (1995) and Pinheiro and Bates (2000).

The applications of mixed nonlinear models can be diverse. In forest stands, they can be used to estimate and predict growth in height (ADAME et al., 2008; CALEGARIO et al., 2005; CRECENTE-CAMPO et al., 2010; MENDONÇA; CARVALHO; CALEGARIO, 2015), estimation and predict diameter growth (TIMILSINA; STAUDHAMMER, 2013), to evaluate the longitudinal profile of the tree trunk (CARVALHO et al., 2014; PIRES; CALEGARIO, 2007), among others.

2.3 Artificial neural networks

Artificial Neural Networks (ANN) have been used as an alternative to hypsometric models for modeling and projecting forest production. According to Braga et al. (2007), an artificial neural network is a computational processor consisting of a large number of simple processors (artificial neurons) (FIGURE 2), interconnected, based on neurons found in humans.

Figure 2 – General structure of an artificial neuron.

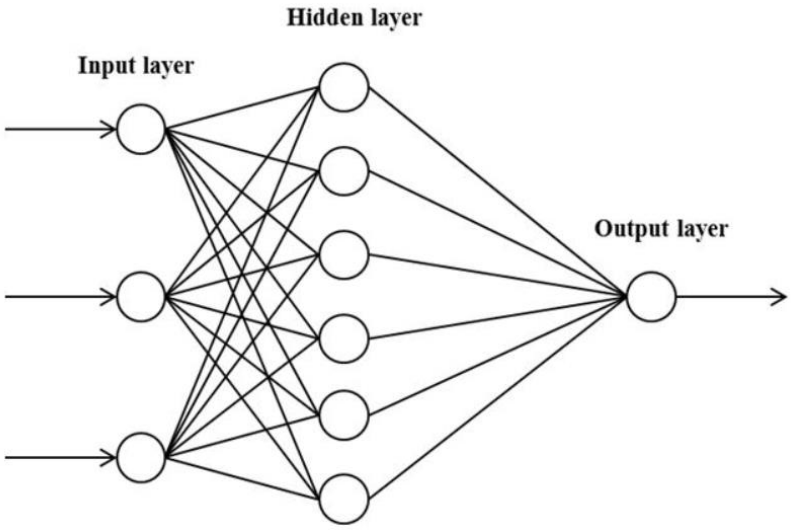


Source: Adapted from Braga et al. (2000).

An artificial neuron is the information processing unit of an artificial neural network. The first artificial neuron model, proposed by McCulloch and Pitts in 1943, is a simplification of what was known about the biological neuron at the time. Consisting of n inputs $x_1, x_2, ..., x_n$ (dendrites) and an output y (axon). The inputs are associated with weights $w_1, w_2, ..., w_n$ that represent the synapses, which can be negative or positive. The weighted sum of the $x_i w_i$ *values* received is calculated and, if the value obtained exceeds the threshold (*threshold*) of the neuron, it "fires", that is, it activates the activation function and generates the output.

Artificial neurons are distributed in layers (FIGURE 3), which justifies the name of the artificial neural network, whose layers are connected by weights, which capture the influence of each input on the variability of the response variable and assign a value to each one, if the weight is high, the input has high influence. Then, the behavior patterns and the relationships between the independent variables and the dependent variable are stored (HAYKIN, 2009). Thus, by adjusting the weights of an artificial neuron, the acquired knowledge becomes available for use, and it is possible to obtain the desired output for specific inputs.

Figure 3 - Structure of an artificial neural network.



Source: Adapted from Braga et al. (2000).

The hidden layers contain the activation functions, whose objective is, in addition to introducing non-linearity to the ANN and finding the function that best adapts to the data distribution, to limit the values of each neuron so that the ANN is not paralyzed by divergent neurons (BRAGA, 2007). The activation functions most used for approximation of functions are threshold, hyperbolic tangent, parabolic and sigmoidal (HAYKIN, 2009).

The characteristics that most attract attention in ANN are their learning and generalization of information. In other words, ANNs, through a learned example, detect the relationships between variables and can generalize the assimilated knowledge to a set of unknown data. Another interesting feature of ANNs is the ease of extracting non-explicit features from a group of information provided to them as examples (DIAMANTOPOULOU, 2005; GORGENS, 2015).

The way a particular problem is internally represented and the parallelism of its architecture can make solving problems with ANN better than conventional solutions (LEK et al., 1996; BRAGA et al., 2000; HAYKIN, 2001; GORGENS, 2006; BRAGA et al., 2000; HAYKIN, 2001; 2006). Other important characteristics of ANNs are: input and output mapping; missing data and noise; adaptability of connection weights to changes in the environment, that is, a trained network can be re-trained or a network can be designed to modify its weights in real time (non-stationary environment); ability to inform the confidence of a decision taken (probability); learning from examples and generalization to unknown data; and neurobiological analogy (HAYKIN, 2001; BULLINARIA, 2009; GORGENS, 2015).

Problems treatable through artificial neural networks fall into learning categories, such as function approximation, pattern classification, data clustering, prediction (time series),

optimization, content retrieval and control (JAIN et al., 1996; HAYKIN, 2001). The prognosis of forest production fits into the function approximation task that consists of designing a neural network that approximates the unknown function *f(*x*)*, which describes the mapping of input-output pairs $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ from a set of n training patterns.

Among the various applications of artificial neural networks, air traffic control can be automated with the location, altitude, direction and speed of each radar used as input to the network (KUMAR et al., 2015); animal behavior and population cycles may be suitable for analysis by neural networks; evaluation and valuation of properties, buildings, automobiles, among others (OZÇELIK et al., 2013); the echo pattern from sonar, radar and magnetic instruments can be used to predict your targets; weather forecast (KUMAR et al., 2015); speech recognition can be obtained by analyzing the audio pattern; traffic flows can be predicted so that signal timing can be optimized (NAGHDI; GHAJAR, 2012); and several other applications.

In recent years, several studies in the forestry sector have tested and applied artificial neural networks as a tool for modeling and forecasting forest growth and production. For Diamatopoulou (2005), the quality of the estimates obtained through the ANN is due to their ability to model several variables and overcome certain problems found in forest data, such as non-linear relationships, non-Gaussian distributions, *outliers* and absences in the data.

Binoti et al. (2013), in a study on the effect of reducing height measurements on the precision obtained by ANN, evaluated the estimates obtained by reducing the number of plots with measured heights and concluded that it is possible to reduce the number of measurements without loss of precision. Also, according to the authors, it is possible to reduce the cost of inventory by applying ANN to estimate tree heights.

REFERENCES

ADAME, P.; DEL-RÍO, M.; CAÑELLAS, I. A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). **Forest Ecology and Management**, Amsterdam, v. 256, n. 1/2, p. 88-98, 2008.

AKHAVAN, R.; ZAHEDI AMIRI, G.; ZOBEIRI, M. Spatial variability of forest growing stock using geostatistics in the Caspian region of Iran**. Caspian Journal of Environmental Sciences**, v. 8, n. 1, p. 43-53, 2010.

AMARAL, L. de P. **Geoestatística na caracterização do solo e da vegetação em Floresta Ombrófila Mista**. 2010. 154 p. Dissertação (Mestrado em Agronomia) – Universidade Estadual do Centro-Oeste, Guarapuava-PR, 2010.

BARBOSA, M. **Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos**. 2009. 119 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Universidade de São Paulo, São Paulo, 2009.

BINOTI, M. L. M. DA S. et al. Aplicação de Redes Neurais Artificiais para Estimativa da Altura de Povoamentos Equiâneos de Eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639–645, 2013.

BRAGA, A. de P.; CARVALHO, A. P. de L. F. de; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações.** Rio de Janeiro, RJ. Editora LTC, 262 p. 2000.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais**: Teoria e Aplicações. Rio de Janeiro: Editora LTC, 2007. 226p.

BULLINARIA, J. A. **Introduction to Neural Computation**. Notas de aula. 2009. Disponível em: <http://www.cs.bham.ac.uk/~jxb/inc.html>. Acesso em: 20 de nov. 2017.

CALEGARIO, N. et al. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos não lineares em multinível de efeito misto. **Ciência Florestal, Santa Maria**, v. 15, n. 3, p. 285-292, 2005.

CAMARINHA FILHO, J. A. Utilização de Modelos Lineares Mistos em Agronomia. **Revista de Matemática e Estatística**. v. 21, p. 34-43, 2003.

CAMARGO, E. C. G. Geoestatística: Fundamentos e aplicações. INPE: São José dos Campos, 2001. Disponível: < http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambien te/5geoest.pdf>. Acesso em 25 de jan. de 2018.

CARVALHO, S. P. C. et al. Modelagem não linear mista para descrever o afilamento de árvores clonais de *Eucalyptus* sp. **Scientia Forestalis**, Piracicaba, v. 42, n. 104, p. 605-614, dez., 2014.

CORREIA, M. D. **Modelagem geoestatística da distribuição de carbono do solo e biomassa de herbáceas em sistema silvipastoril**. 2013. 101 p. Dissertação (Mestrado em Biometria e estatística aplicada) – Universidade Rural de Pernambuco, Recife, 2013.

COUTO, E.G., STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture Ecosystems and Environment**, v. 66, p. 139- 152, 1997.

CRECENTE-CAMPO, F. et al. A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, Amsterdam, v. 259, p. 943–952, 2010.

CRESSIE, N. **Statistics for spatial data**. New York: Wiley, 1993. 900p.

CUNHA, T. A; FINGER, C. A. G.; SCHNEIDER, P. R. Linear mixed model to describe the basal area increment for indivudual cedro (*Cedrela odorata* L.) trees in occidental Amazon, Brazil. **Ciência Florestal**, Santa Maria, v. 23, n. 3, p. 461-470, 2013

DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. New York: Chapman and Hall, 1995. 360 p.

DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. **Computers and electronics in agriculture**, v. 10, p. 235-244, 2005.

DUNCANSON, L. et al. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. **Surveys in Geophysics**, n. 0123456789, p. 1-21, 2019.

GORGENS, E. B. **Estimação do volume de árvores utilizando redes neurais artificiais.** 2006. 84 p. Dissertação (Mestrado em Ciência Florestal). Universidade Federal de Viçosa, Viçosa, MG, 2006.

GORGENS, E. B.; MONTAGHI, A.; RODRIGUEZ, L. C. E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers and Electronics in Agriculture**, v. 116, p. 221-227, 2015.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University Press, 1989.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial Neural Network: A Tutorial. **Computer**, v. 29, n. 3, p. 31-44, 1996.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistcs**. London: Academic, 1978. 600 p.

JOURNEL, A. G.; HUIJBREGTS, C.J. **Mining geoestatistics**. London: Academic, 2003. 600 p.

KUMAR, S.; DESPANDE, S. M., DAS, S. S.; KONWAR, M.; CHAKRAVAERTY, K.; KALAPUREDDY, M. C. R. Temporal and structural evolution of a tropical monsoon cloud

system: A case study using X-band radar observations. **Journal of atmospheric and solar-terrestrial physics**. v. 133, p. 157-168, 2015.

LAIRD, N. M; WARE, J. H. Random effects models for longitudinal data. **Biometrics**, Washington, v. 38, p. 963-974, 1982.

LANDIM, P. M. B**. Análise estatística de dados geológicos**. 2. Ed. Ver e ampl. São Paulo: Editora UNESP, 2003.

LEK, S. et al. Application of neural networks to modelling nonlinear relationships in ecology. **Ecological Modelling**, v. 90, p. 39-52, 1996.

LINDSTROM, M. J.; BATES, D. M. Nonlinear mixed-effects models for repeated measures data. **Biometrics**, Washington, v. 46, p. 673–687, 1990.

LITTEL, R. C. et al**. SAS system for mixed models**. Cary: SAS Institute, 1996. 633p.

LITTELL, R. C.; HENRY, P. R.; AMMERMAN, C.B. Statistical analysis of repeated measures using SAS procedures. **Journal of Animal Science**, v. 76, p. 1216-1231, 1998.

MACIEL, S. M. **Análise espacial do carbono em um fragmento florestal com predominância de *Anadenanthera* sp**. 2012. 135 p. Dissertação (Mestrado em Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras. 2012.

MARTINS, E. R., BINOTI, M. L. M. S., LEITE, H. G., BINOTI, D. H. B. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. **Revista Brasileira de Ciências Agrarias**, v. 11, n. 1, p. 33–38, 2016.

MATHERON, G. Principles of geostatistics. **Economic Geology**. El paso, v. 58, p. 11246-66, 1963.

MCCULLOCH S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathernatical Biophysics**, v. 5, p. 115-133, 1943.

MELLO, J. M. de. **Geoestatística aplicada ao inventário florestal**. 2004. 122 p. Tese (Doutorado em Recursos Florestais) – Universidade de São Paulo, Piracicaba. 2004.

MELO, E. A. et al. Modelagem não Linear da Relação Hipsométrica e do Crescimento das Árvores Dominantes e Codominantes de Eucalyptus sp. **Ciência Florestal**, v. 27, n. 4, p. 1325–1338, 2017.

MENDONÇA, A. R.; CARVALHO, S. P. C.; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus* sp**. Cerne**, Lavras, v. 21, n. 1, p. 107-115, 2015.

MOTOMIYA, A. V. de A.; CORÁ, J. E.; PEREIRA, G. T. Uso da krigagem indicatriz na avaliação de indicadores de fertilidade do solo. **Revista Brasileira de Ciência do Solo**, n. 30, p. 485-496, 2006.

MRODE, R.A. **Linear models for the prediction of animal breeding values**. 3 ed. Wallingford: CAB International, 2015. 343 p.

NAGHDI, R.; GHAJAR, 1. Application of Artificial Neural Network in the Modeling of Skidding Time Prediction. **Advanced Materials Research.** v. 403-408, n. 2012, p. 3538-3543, 2012. doi:10.4028/www.scientific.net/AMR.403-408.3538

NEVES, D. A. **Geoestatística aplicada ao estudo da variabilidade de espécies nativas em Fragmentos do Cerrado Brasileiro no Estado de São Paulo**. 2013. 172 p. Tese (Doutorado em Ciências da Terra e Meio Ambiente) – Universidade de A Coruña, Espanha. 2013.

NOGUEIRA, C. H. **Análise de variância com dependência espacial sob uma abordagem geoestatística**. 2013. 124 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

ÖZÇELIK, R. et al. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. **Forest Ecology and Management**, v. 306, p. 52–60, 2013.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in s and s-plus**. New York: Springer-Verlag, 2000. 528 p.

PIRES, L. M.; CALEGARIO, N. Ajuste de modelos estocásticos lineares e não- lineares para a descrição do perfil longitudinal de árvores. **Revista Árvore**, Viçosa, v. 31, n. 5, p. 845-852, 2007.

PUGH, T. A. M. et al. Role of forest regrowth in global carbon sink dynamics. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 10, p. 4382–4387, 2019.

RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99 p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiros”. 1995.

SALVIANO, A.A.C. **Variabilidade de atributos de solo e de *Crotalaria juncea* em solo degradado do município de Piracicaba-SP**. 1996. 91 p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Universidade de São Paulo.1996.

SCHEFFER, M.; BROVKIN, V.; COX, P. M. Positive feedback between global warming and atmospheric CO_2 concentration inferred from past climate change. **Geophysical Research Letters**, v. 33, n. 10, p. 2–5, 2006.

SEARLE, S.R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536 p.

SEARLE, S.R.; CASELLA, G.; MCCULLOCH, C. **Variance components**. New York: John Wiley, 1992. 501 p.

SILVA, A. F. da.; ZIMBACK, C. R. L.; OLIVEIRA, R. B. de. Cokrigagem na estimativa da evapotranspiração em Campinas (SP). **Tékhne e Lógos**, Botucatu, São Paulo, v. 2, n. 1, 2010.

SILVEIRA, E. M. O. et al. Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. **Forest Ecology and Management**, v. 445, n. May, p. 96–109, 2019a.

SILVEIRA, E. M. O. et al. Estimating Aboveground Biomass Loss from Deforestation in the Savanna and Semi-arid Biomes of Brazil between 2007 and 2017. **Tropical Forests in Transition - The Role of Deforestation and Impacts from Community Composition to Regional Climate Change [Working Title]**, p. 1–17, 2019b.

SOARES, P. **Levantamento fitossociológico de regeneração natural em reflorestamento misto no noroeste de Mato Grosso.** 2009. 50 p. Dissertação (Mestrado em Ciências Florestais e Ambientais) – Universidade Federal de Mato Grosso – Cuiabá. 2009.

SORIANO-LUNA, M. DE LOS Á. et al. Determinants of above-ground biomass and its spatial variability in a temperate forest managed for timber production. **Forests**, v. 9, n. 8, p. 1–20, 2018.

TIMILSINA, N.; STAUDHAMMER, C. L. Individual tree-based diameter growth model of slash pine in Florida using nonlinear mixed modeling. **Forest Science**, Lawrence, v. 59, n. 1, p. 27-37, fev. 2013.

TORAL, F.L.B.; ALENCAR, M.M.; FREITAS, A.R. Arranjos para efeitos fixos e estruturas de (co)variâncias residuais para análises de medidas repetidas do peso de bovinos da raça Canchim. **Revista Brasileira de Zootecnia**, v. 35, n. 5, p. 1951-1958, 2006.

VENDRUSCOLO, D. G. S. et al. Estimativa Da Altura De Eucalipto Por Meio De Regressão Não Linear E Redes Neurais Artificiais. **Revista Brasileira de Biometria**, v. 33, n. 4, p. 556–569, 2015.

VIEIRA, S. R. **Geoestatística em estudos de variabilidade espacial**. 2000.

WOJCIECHOWSKI, J. C. et al. Geoestatística aplicada ao estudo fisíco – químico do solo em áreas de floresta estacional decidual. **Ciência Florestal**, Santa Maria, v. 19, n. 4, p. 383-391, 2009.

XAVIER, A. C.; CECÍLIO, R. A.; LIMA, J. S. S. Módulos em *Matlab* para interpolação espacial pelo método da krigagem ordinária e do inverso da distância. **Revista Brasileira de Cartografia**, n 62/01, 2010.

ZIMBACK, C. R. L. **Análise especial de atributos químicos de solo para fins de mapeamento da fertilidade**. 2001. 114 p. Tese (Livre Docência em Levantamento do solo e fotopedologia), FCA, Botucatu, 2001.

SECOND PART - ARTICLES

ARTICLE 1 - ABOVE AND BELOWGROUND CARBON STOCK IN A TROPICAL FOREST IN BRAZIL

ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA REVISTA ACTA SCIENTIARUM AGRONOMY

Situação: publicado na Revista Acta Scientiarum Agronomy

ABOVE AND BELOWGROUND CARBON STOCK IN A TROPICAL FOREST IN BRAZIL

ESTOQUE DE CARBONO ACIMA E ABAIXO DO SOLO EM UMA FLORESTA TROPICAL NO BRASIL

ABSTRACT

An increase in atmospheric CO_2 levels and global climate changes have led to an increased focus on CO_2 capture mechanisms. The *in situ* quantification and spatial patterns of forest carbon stocks can provide a better picture of the carbon cycle and a deeper understanding of the functions and services of forest ecosystems. This study aimed to determine the aboveground (tree trunks) and belowground (soil and fine roots, at four depths) carbon stocks in a tropical forest in Brazil and to evaluate the spatial patterns of carbon in the three different compartments and in the total stock. We used census data from a semideciduous seasonal forest to estimate the aboveground carbon stock. The carbon stocks of soil and fine roots were sampled in 52 plots at depths of 0-20, 20-40, 40-60, and 60-80 cm, combined with the measured bulk density. The total estimated carbon stock was 267.52 $Mg \cdot ha^{-1}$, of which 35.23% was in aboveground biomass, 63.22% in soil, and 1.54% in roots. In the soil, a spatial pattern of the carbon stock was repeated at all depths analyzed, with a reduction in the amount of carbon as the depth increased. The carbon stock of the trees followed the same spatial pattern as the soil, indicating a relationship between these variables. In the fine roots, the carbon stock decreased with increasing depth, but the spatial gradient did not follow the same pattern as the soil and trees, which indicated that the root carbon stock was most likely influenced by other factors.

KEYWORDS: Carbon sink; biomass; spatial pattern; secondary forest; soil.

INTRODUCTION

The increase in atmospheric CO_2 levels and global climate changes have led to an increased focus on CO_2 capture mechanisms and increased efforts to reduce CO_2 emissions,

such as Reducing Emissions from Deforestation and forest Degradation - REDD+ (Correa, Van der Hoff & Rajão, 2019). The role of soil in CO_2 capture has gained prominence due to its capacity to store approximately four times more carbon than plant biomass and three times more than the atmosphere (Watson, 2001). Therefore, soil organic carbon is the largest carbon sink in the terrestrial biosphere (Jobbágy & Jackson, 2000) and remains one of the main strategies to mitigate increases in atmospheric CO_2 concentrations (Asner & Mascaro, 2014). In addition, the potential to use soil organic carbon as an indicator of soil quality reinforces the importance of having appropriate techniques to accurately measure soil carbon concentrations and to adequately predict soil carbon storage (Franzluebbers, 2002; Bhattacharya et al., 2016).

The soil carbon balance depends on the relationship between the addition of photosynthesized carbon by plants and carbon losses to the atmosphere resulting from the microbial oxidation of organic carbon into CO_2 (King, 2011; Bhattacharya et al., 2016; Crowther et al., 2016). Studies have shown that the storage of organic carbon in soil and its dynamics are determined by factors such as climate, soil type and properties, plant cover, and management practices (Moore et al., 2018; Navarrete-Segueda et al., 2018; Shukla & Chakravarty, 2018).

Forest tree species have a higher capacity for nutrient cycling than annual plants due to the permanent and deep root system that absorbs elements from the subsurface layer, returning them to the surface through litter deposition (Haag, 1985). Tropical ecosystems, such as the Amazon Rainforest and the Atlantic Forest, have high productivity due to the temperature and humidity that favor the decomposition of soil organic matter and to the absence of physical disturbances that allow the formation of large carbon stocks (Shukla & Chakravarty, 2018). Therefore, forests also play a significant role in the storage of global terrestrial carbon in their different components (Van der Sande et al., 2017).

The horizontal distribution of soil organic carbon and its relationship with climate and vegetation have been extensively studied (Jobbágy & Jackson, 2000; Terra, Mello & Mello, 2015). Despite many studies on the geostatistical procedures applied to mapping soil attributes, little spatial information is available on the vertical distribution of soil carbon under forests, especially forests with a notable predominance of one tree genus/species. Knowledge of the variation in depth of soil carbon and the solid study of the spatial patterns of this variable, together with information on carbon present in other forest compartments, can provide a better picture of carbon stocks and a deeper understanding of forest ecosystems. Furthermore, these

data meet the demand for high-quality in situ data (Duncanson et al., 2019) and can certainly provide a background for more effective conservation and management strategies.

Based on the considerations above, there is a clear demand for carbon estimates in the soil and in other forest compartments in different ecosystems. Therefore, this study aimed to (1) determine the soil and root carbon stock at four depths (0-20, 20-40, 40-60, and 60-80 cm) under tropical forest in Brazil, (2) estimate the carbon in tree trunks, and (3) evaluate the spatial patterns of carbon in the three distinct compartments (soil, tree trunks, and fine roots) and in the total stock above and belowground, with the aim of determining the spatial relationships between these variables.

MATERIAL AND METHODS

Study área

The study area is a 1.2-ha secondary forest located in the Lavras municipality, state of Minas Gerais, Brazil, at coordinates 21° 14' S and 45° 00' W and at an average altitude of 900 m. The average annual precipitation and potential evapotranspiration are 1,511 mm and 900 mm, respectively. According to the Köppen classification system, the climate of the region is Cwa (Alvares et al., 2014). There are two well-defined seasons: the dry season (April to September, winter) and the rainy season (October to March, summer). The average annual temperature is 19.4 °C, ranging from 14.4 °C (in July) to 22.5 °C (in January) (Terra et al., 2018; Marques et al., 2019). According to Carvalho et al. (2010), the forest is heterogeneous, with multiannual trees and a predominance of the genus *Anadenanthera*, popularly known as angico. The allelopathic potential of these species has been evaluated in the literature (Oliveira et al., 2005). The forest is bordered by a *Eucalyptus urograndis* plantation to the north and west and by an early secondary, semi-deciduous forest stock to the south. There is a road along the eastern side of the forest.

Soil data

We distributed one hundred and five permanent 10 x 10-m sampling plots over the area. The plots for soil sample collection were systematically chosen (Figure 1). The soil sampling points corresponded to the center of each sampling plot, totaling 52 points, 20 m apart from each other.

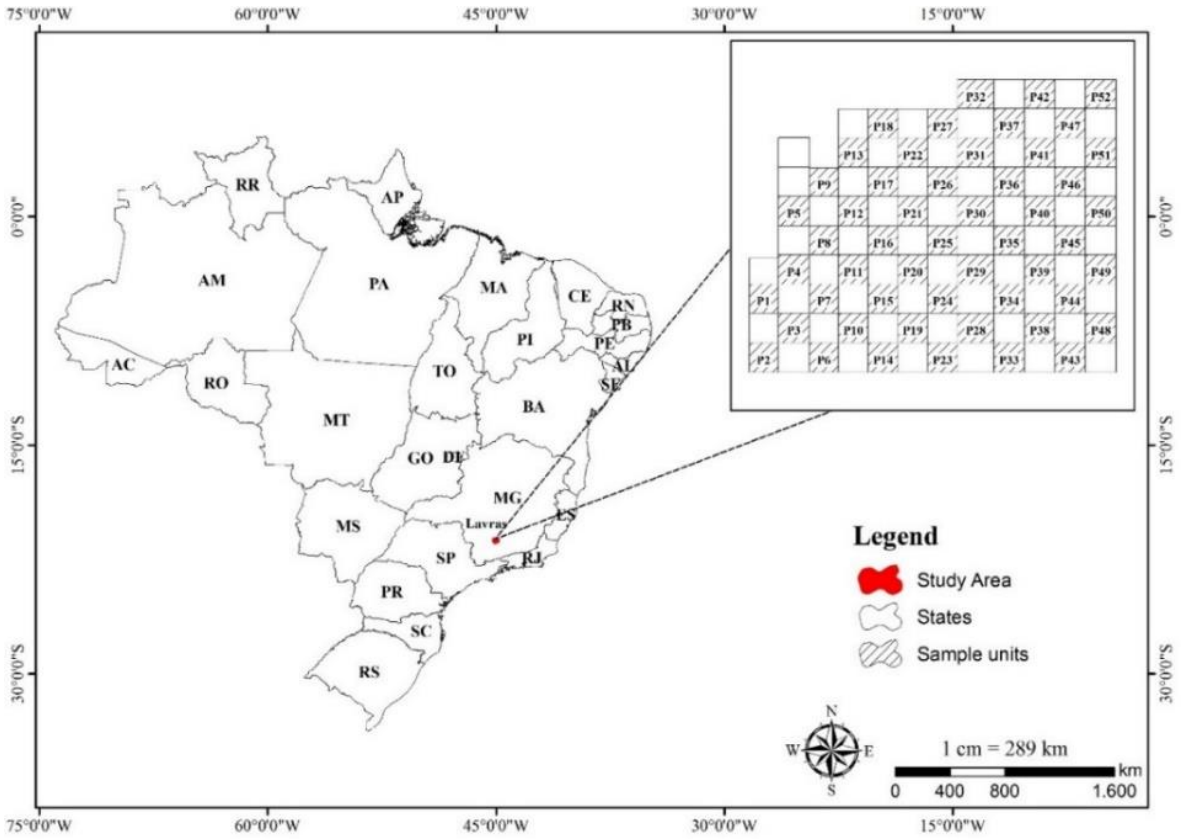


Figure 1. Location and sketch of the study area with the sampled plots.

In all soil sample points, the vegetation cover was removed, and holes 80 cm deep and 20 cm in diameter were made by using a propeller coupled to a chainsaw (Figure 2). We collected soil samples at depths of 0-20, 20-40, 40-60 and 60-80 cm with a post hole digger. Each soil sample was stored in labeled plastic bags, followed by separation of roots and soil using sieves with mesh sizes of 7.95 mm and 2.26 mm. The soil samples were dried and classified using sieves with 200 and 270 mesh sizes. The material was dried in an oven at 60 ± 2 °C for analysis of elemental carbon.

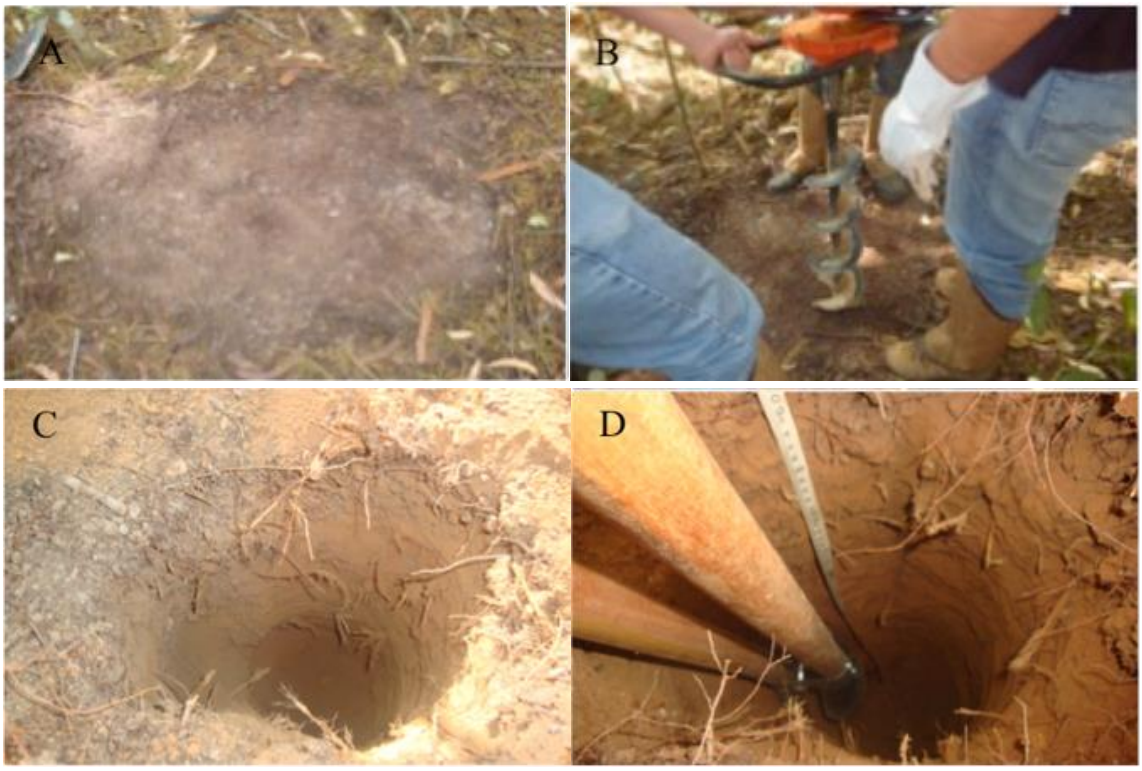


Figure 2. (A) Collection site after cleaning; (B) Instrument used to make the hole; (C) Hole; (D) Depth of the hole and marking of the post hole digger.

Elemental carbon analysis was performed using the CHN elemental analyzer (Vario MICRO Cube), whose basic principle is the combustion of material to determinate the levels of carbon, hydrogen, nitrogen, and sulfur by the difference in oxygen. Two milligrams of each sample were weighed, packed in tin capsules, and then transferred to the apparatus. The soil density was determination by collecting samples at five points representative of the area and at the four depths studied, using a volumetric ring with a 98.174 cm^3 volume. The samples were oven dried at 105 ± 2 °C to determine the dry weight. With the volume and dry weight data, it was possible to determine the mean soil density at each depth.

The expression of Veldkamp (1994) (Equation 1) was used to determine the total carbon stock of the soil at each sampling point and depth:

$$SOCS = OC * BD * e \quad (1)$$

Where: SOCS is the soil organic carbon stock at each point and depth ($Mg \cdot ha^{-1}$); OC is the total organic carbon content at the points and depths sampled (%); BD is the bulk density at each depth ($g \cdot cm^{-3}$); and e is the thickness of the considered layer (cm).

Roots tree data

We separated the roots originating from the soil samples using 7.95 mm and 2.26 mm mesh sieves. After sorting, we washed the samples in running water for complete soil removal. They were then oven dried at 60 ± 2 °C until reaching a constant weight (Figure 3).

Figure 3. (A) Root separation; (B) Root of *Anadenanthera* sp.; (C) Root wash; (D) Samples packed in the oven.

After drying, we weighed the samples to determine the dry weight. Next, they were ground in a Willey mill and classified using 200 and 270 mesh sieves, where the aliquot retained in the 270-mesh sieve was used (Figure 4). Subsequently, the material was oven dried at 65 °C for 24 hours for elemental CHN analysis.

Figure 4. (A) Weighing the samples in a scale; (B) Grinding the samples in Wiley mill; (C) Sampling after grinding; (D) Classification of samples in 200 and 270 mesh sieves.

The elemental CHN analyzer was used to determination the carbon content in each sample. For that purpose, 2 mg of each sample was weighed on an analytical scale, with 0.01 mg precision, placed in the tin capsules, and taken to the elemental analyzer (Figure 5). The carbon values were expressed as percentage. After these steps, it was possible to obtain the amount of carbon by multiplying the carbon content by the rot dry biomass.

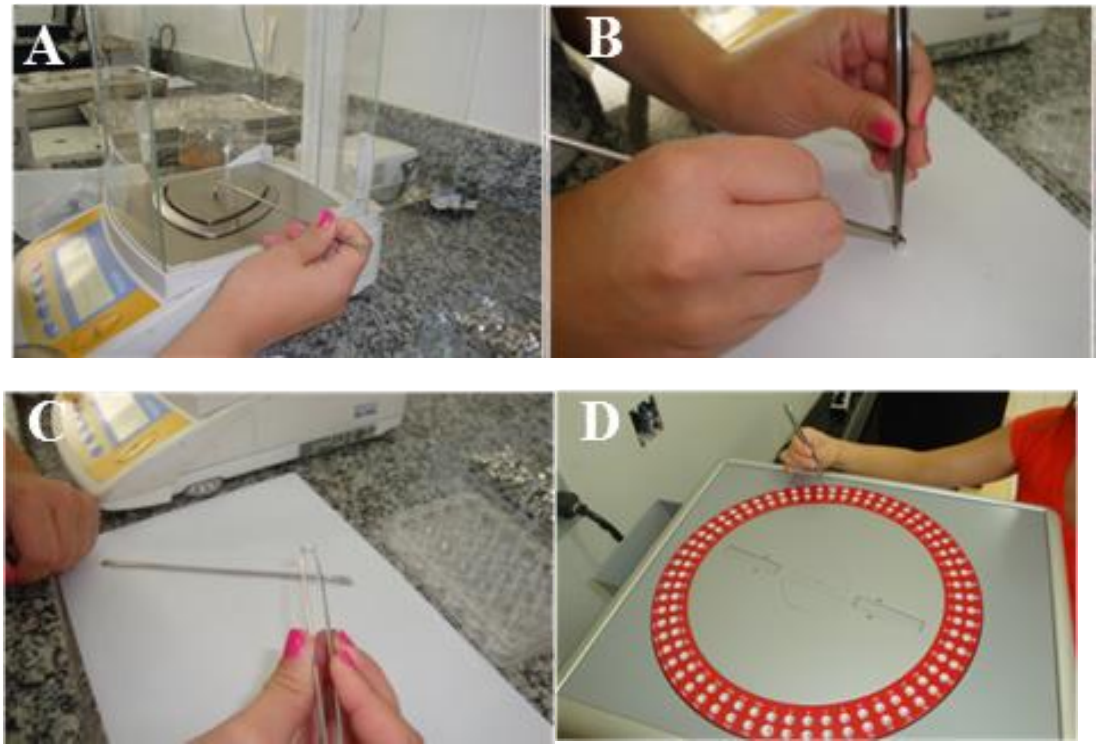


Figure 5. (A) Weighing 2 mg of sample into the tin sample holder; (B) Closing the sample holder; (C) Sample ready for analysis; (D) Placing the samples on the CHN carousel for elemental analysis

Tree data

The tree biomass was obtained from inventory data (diameter at 1.30 m aboveground, known as diameter at breast height or DBH, and total height) combined with wood density estimates. According to Chave et al. (2014), the AboveGround Biomass – AGB was obtained in Mg using the *computeAGB* function of the *Biomass* package (Réjou-Méchain et al., 2017) in software R (R Core Team, 2018), adopting an average wood density of 0.620 $g \cdot cm^{-3}$ obtained by means of the function *getWoodDensity*, from the same package. We transformed the biomass values into carbon by applying the constant 0.471, proposed by Thomas and Martin (2012) as the concentration of carbon in the tissues of tropical angiosperms. The estimate carbon stock for each sampling plot was calculated as the sum of the carbon in each tree trunk in the plot.

Statistical Analysis

The carbon estimation was statistically analyzed for each forest compartment (soil, fine roots, and tree stems) and the different depths (0-20, 20-40, 40-60, and 60-80cm), including the average, variance, standard deviation, and coefficient of variation. The statistical significance of the differences in the average carbon stock between the depths (for soil and roots data) was tested by using Tukey's HSD test. Differences with $p < 0.05$ were considered statistically significant.

We used semivariograms for each variable (soil carbon stock at four depths, tree root carbon at four depths, and tree carbon stock) to test the spatial dependence (SD) of the variable according to the estimator Cressie and Hawkins (1993). The spherical, exponential, and Gaussian models were fitted to the experimental semivariograms by the ordinary least square method. The goodness of fit of each model was evaluated in terms of AIC and degree spatial dependence (Cambardella et al., 1994; Mello et al., 2009). Next, based on the best models for each variable, a carbon stock prediction map was constructed using ordinary kriging (Webster & Oliver, 2007). In the absence of spatial dependence, variables were interpolated using inverse distance weighting (power of 2). According to Cambardella et al. (1994), an SD < 25% indicates weak spatial dependence, SD between 25 and 75% indicates moderate SD and, finally, SD > 75% indicates strong spatial dependence.

Descriptive statistics and geostatistical analyses were performed using the *geoR* package (Ribeiro Júnior & Diggle, 2001) in R (R Core Team, 2018). The spatial variability of the data was considered stationary and isotropic.

RESULTS AND DISCUSSION

Soil carbon stock

Table 1 shows the descriptive statistics for soil carbon stock in the study area.

Table 1. Soil density and descriptive statistics of soil carbon stock at each depth in the study area. *Different letters indicate significant differences between the values according to Tukey's HSD test at 0.05.

STATISTIC	DEPTH			
	1	**2**	**3**	**4**
Soil density ($g \cdot cm^{-3}$)	0.89^{a}	1.07^{b}	1.09^{b}	1.11^{b}
Carbon content (%)	3.09^{a}	2.01^{b}	1.69^{c}	1.53^{c}
Soil carbon stock ($Mg \cdot ha^{-1}$)*	55.0465^{a}	44.6903^{b}	36.7707^{c}	32.6309^{c}
Variance $(Mg \cdot ha^{-1})^2$	264.7700	103.7429	89.9311	67.6550
Standard deviation ($Mg \cdot ha^{-1}$)	16.2718	10.1854	9.3237	8.2253
CV (%)	29.56	22.79	25.36	25.21

A low variation in soil density can be observed at depths between 20 and 80 cm, with values varying ranging from 1.07 to 1.11 $g \cdot cm^{-3}$. In the topsoil, from 0 to 20 cm, there is a lower density, 0.89 $g \cdot cm^{-3}$, which can be explained by the presence of more organic matter in this layer (Parras-Alcántara, Lozano-García & Galán-Espejo, 2015). Generally, there is a negative relationship between soil density and depth as a result of the high organic matter content at the surface because organic matter is less dense than mineral grains (Hossain et al., 2015).

The carbon content, values decrease as the depth increased, from 3.09 to 1.53 %. This result corroborates other studies that found carbon content in the upper 20 cm of soil ranging from 2.9 to 4.1 % (Powers et al., 2011; Ngo et al., 2013). Higher levels of organic carbon in the surface soil in forest environments are due to the presence of the organic mat formed by fallen leaves, tree branches, and bark, and by the higher density of fine roots (Santos et al., 2013, Navarrete-Segueda et al., 2018). Thus, the addition of organic litter material is responsible for the accumulation of carbon in the topsoil layer because it is humified (Mafra et al., 2008), which increases the nutrient cycling in the upper layers of the soil profile (Mora et al., 2018).

Similar to the carbon content, the carbon stock shows a decreasing trend along the soil profile, from 55.05 to 32.63 $Mg \cdot ha^{-1}$, also as a function of the incorporation dynamics of organic matter in the soil profile. The amounts of carbon found in the subsurface layers highlight the importance of subsoils in the terrestrial carbon balances. A similar pattern was observed by Balbinot et al. (2003) when estimating the soil carbon stock under a 5-year *Pinus taeda*

plantation in Rio Grande do Sul, Brazil. The authors found carbon stock values of 83.9; 63.9; 47.6, and 19.6 $Mg \cdot ha^{-1}$ at depths 0-20, 20-40, 40-60, and 60-80 cm, respectively. Morais et al. (2017) found an estimated soil carbon stock of 208 $Mg \cdot ha^{-1}$ in vegetation of the Cerrado (woodland savana) biome, in Brazil. Therefore, it can be inferred that the differences in soil carbon stock observed between these studies are related to the type of forest cover as well as the climatic and soil conditions of each area.

Although both the carbon content and stocking exhibit the same behavior in this area, other studies have found different behaviors for these variables when analyzing them independent. For example, Novaes Filho et al. (2007) observed in a primary forest in Southern Amazonia that, although the carbon content of the soil decreases with depth, the carbon stock did not follow this trend with the same intensity, which was explained by the increase of the bulk density of the soil with depth that compensated for the decrease content. The total soil carbon stock under this forest fragment is 169.11 $Mg \cdot ha^{-1}$.

Roots carbon stock

Table 2 shows the descriptive statistics for the root carbon stock in the study area.

Table 2. Average root carbon stock at each depth in the study area. *Different letters indicate significant differences between the values according to Tukey's HSD test at 0.05.

STATISTIC	DEPTH			
	1	**2**	**3**	**4**
Average ($Mg \cdot ha^{-1}$) *	1.6776^{a}	1.3055^{b}	0.7931^{c}	0.3537^{d}
Variance $(Mg \cdot ha^{-1})^2$	0.0169	0.0154	0.0039	0.0016
Standard deviation ($Mg \cdot ha^{-1}$)	0.1300	0.1242	0.0627	0.0394
CV (%)	7.75	9.51	7.90	11.14

The value found in the top 20 cm of the soil is higher than at the other depths, and there is a significant difference in the root carbon stock between the depths analyzed, which results in a descending gradient of carbon stock in the roots. There is little high variation between the

root carbon stock values throughout the area at each depth, as indicated by the low CV values (≤ 11.14 %).

Paiva and Faria (2007), studying root biomass in a cerrado *sensu stricto* area, observed that the more superficial soil layers concentrated most of the root biomass. According to Laclau et al. (2004), the agglomeration of fine roots in the organic horizon represents a strategy to acquire nutrients in infertile soils with nutrient deficiency. Fine roots (roots up to 2 mm in diameter) mainly fulfil nutritional, metabolic, and symbiotic functions in the upper soil layers and horizons (Hendrick & Pretziger, 1996; Jaloviar et al., 2009), where there is greater porosity (Baker et al., 2001). The root system is a vital part of understanding carbon accumulation, nutrient and water absorption by plants, and groundwater infiltration. The total root carbon stock under this forest fragment is 4.1299 $Mg \cdot ha^{-1}$.

Tree carbon stock

Table 3 shows the descriptive statistics of tree carbon stock in the study area.

Table 3. Average tree carbon stock in the study area.

DESCRIPTIVE STATISTICS	
Average ($Mg \cdot ha^{-1}$)	94.25
Variance ($Mg \cdot ha^{-1}$)	4,495.71
Standard deviation ($Mg \cdot ha^{-1}$)	67.05
CV (%)	71.14

The forest with a predominance of *Anadenanthera sp.* has an average carbon stock of 94.25 $Mg \cdot ha^{-1}$. Aboveground carbon stocks vary widely according to the degree of tree cover, but comparatively, some authors found values similar to those found in this studying other Semideciduous Seasonal Forests. For example, Ribeiro et al. (2009), who quantified the biomass and carbon stock of tree in a mature forest in the Viçosa municipality, Brazil, using the average wood density of the species and found 166.67 $Mg \cdot ha^{-1}$ of biomass and 83.34 $Mg \cdot ha^{-}$

^{1} of carbon. Additionally Figueiredo et al., (2015), who evaluated the dynamics of tree carbon stock in Minas Gerais, Brazil, found an average tree carbon stock of 71.81 $Mg \cdot ha^{-1}$.

Aboveground and belowground carbon stocks

Of the three compartments measured (trees, soil, and roots), most of the carbon stock is belowground. The total estimated carbon stock is 267.5183 $Mg \cdot ha^{-1}$, of which 35.23 % is in aboveground biomass, 63.22 % in soil, 1.54 % in roots. According to Dixon et al. (1994), in tropical forests, approximately 50% of the total carbon is stored in aboveground biomass, and 50% is in the layer extending from the soil surface down to 1 m. However, there is a difference between different regions reported in the literature. For example, Djomo, Knohl and Gravenhorst (2011), when analyzing an African moist tropical forest, found over three times more carbon in the aboveground biomass than in the soil, whereas Gibbon et al. (2010) found twice as much carbon in soil than in aboveground biomass in a Peruvian montane forest.

Ngo et al. (2013) indicated that the contribution of the different compartments to total carbon stock varies markedly between primary and secondary forests. Specifically, in primary forest, the dominant compartment is the aboveground biomass, and the soil contributes less due to a greater number of trees with larger diameters, while the opposite is true in secondary forest. In a secondary tropical forest in Singapore, these authors found a total carbon stock of 274 $Mg \cdot ha^{-1}$, with 38 % in aboveground biomass, 52 % in soil, 6.9 % in coarse roots, 1.5 % in coarse woody debris, and 1.3 % in fine roots.

Spatial Analysis

The results of the geostatistical modeling for the variables with spatial dependence are shown in Table 4. In all cases, the exponential model has performance the best.

Table 2. Parameters of the Semivariogram for the aboveground and belowground carbon stock under a forest fragment with predominance of *Anadenanthera* sp. at different depths.

Carbon stock	Model	Nugget	Partial sill	Range	SD	
Soil (Depth 2)	Exponential	50.84	81.72	142.13	61.65	Moderate
Soil (Depth 3)	Exponential	19.54	64.22	40.48	76.67	Strong
Soil (Depth 4)	Exponential	33.44	53.00	142.13	61.31	Moderate
Roots (Depth 2)	Exponential	0.01	0.01	142.13	51.84	Moderate
Roots (Depth 4)	Exponential	6.10E-04	8.69E-04	26.22	58.75	Moderate
Trees	Exponential	2515.41	3101.73	142.13	55.22	Moderate

Overall, spatial dependence has increased at deeper soil layers. In the deeper soil layers, the soil texture is highly heterogeneous, and the soil is younger and have less carbon (Terra et al., 2015). Therefore, there is a tendency for a more aggregate carbon distribution, increasing the spatial dependence of the process. Furthermore, the surface layers are highly exposed to weathering and fluctuations in the canopy (Parras-Alcántara et al., 2015), which could explain the pure nugget effect found in the layer 1 (0-20 cm) (i.e., spatial dependence is detected) and the moderate/strong spatial dependence in the other layers.

Cambardella et al. (1994) argue that strong spatial dependence can be explained by soil properties, while weak spatial dependence is influenced by external factors. In contrast to our results, Chaves and Farias (2008) observed (in a soil cultivated with sugarcane) that the carbon stock in surface soil layer showed a moderate degree of spatial dependence, while in the subsurface layers, there was a strong degree of dependence. In addition, Novaes Filho et al. (2007) found spatial dependence of soil carbon in the surface and subsurface horizon. These authors concluded that the carbon concentration varies according to the vegetation type, topographical position of the landscape, and intrinsic soil characteristics, such as texture.

Cerri et al. (2004) observed high heterogeneity among soil attributes in the state of Rondônia, Brazil, despite the apparent similarity between the studied sites. The authors used samples at a distance of 25 m from each other and observed a pure nugget effect in most of the soil attributes, indicating high variability within a small space, even at distance shorter than that of the plots.

For the root carbon stock, spatial dependence has not been observed in layers 1 and 3 (0-20 and 40-60 cm, respectively), indicating the occurrence of random root development in these layers. In layers 2 and 4 (20-40 and 60-80 cm, respectively), there is moderate spatial

dependence with a range of 142.13 and 26.22 m, respectively. This finding may be associated with the horizontal and vertical variation of the soil texture. In superficial, layers there is a greater porosity (Baker et al., 2001) and homogeneity soil texture, whereas in deeper layers, there is a greater soil texture heterogeneity (Terra et al., 2015), resulting in a more aggregate distribution of roots.

The tree carbon stock exhibits moderate spatial dependence, with a range of 142.13 m. Spatial dependence of the tree carbon stock was also observed by other authors (Amaral et al., 2010; Terra et al., 2015).

For the variables that exhibited spatial dependence, an interpolation was performed by ordinary kriging, and for those without spatial dependence, an IDW interpolation was performed.

The spatial distribution maps of soil carbon stock at each depth clearly show the presence of a north-south gradient of the soil carbon stock throughout the area at all depths analyzed (Figure 6). The values are higher in the northern part than in the southern part, and a decrease in the amount of soil carbon stock can be observed with increasing depth.

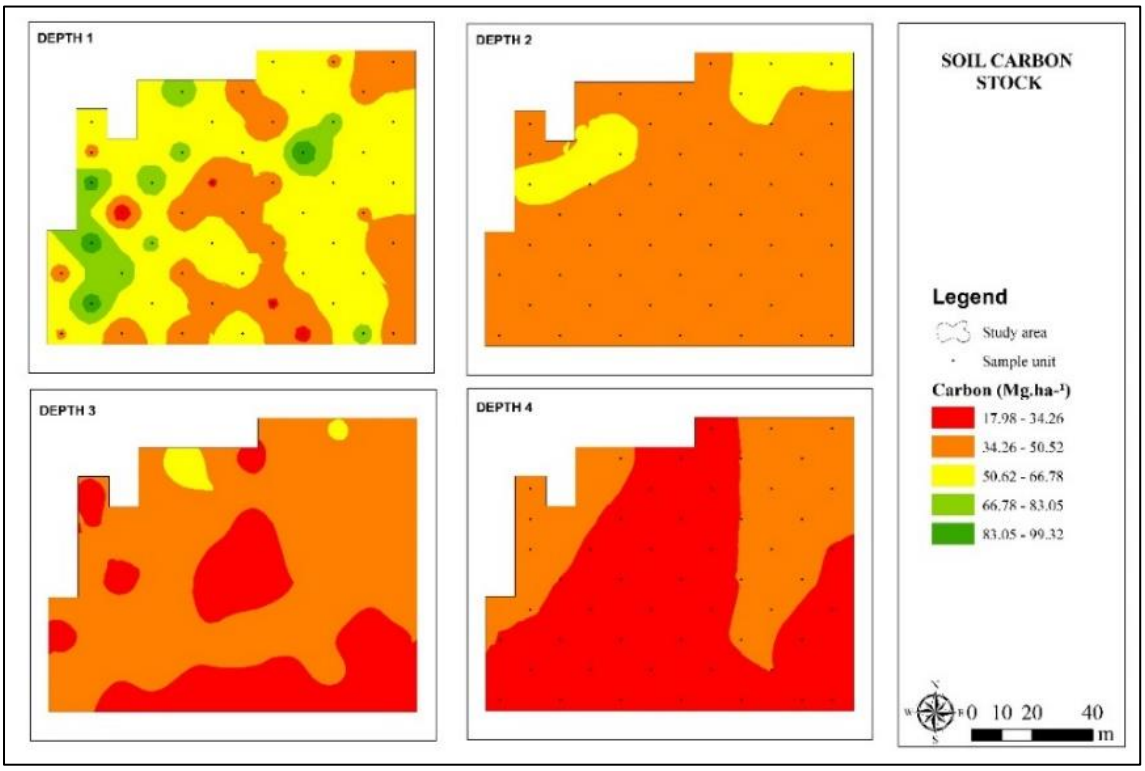


Fig 6. Spatial distribution maps of soil carbon stocks under a forest fragment with a

predominance of *Anadenanthera* sp. at depths 1 (0-20 cm), 2 (20-40 cm), 3 (40-60 cm), and 4 (60-80 cm).

The comparison between the spatial distribution map of the total soil carbon stock and map of tree carbon stock show that these variables have an approximately similar spatial pattern throughout the study area (Figures 7 and 8), indicating a relation between them. Some authors have already shown the existence of a close relationship between the tree carbon stock and soil carbon stock in tropical forests (Mafra et al., 2008; Don, Schumacher & Freibauer, 2010; Fan et al., 2015). Morais et al. (2017) studied the spatialization of carbon stock in the cerrado biome and concluded that its spatial distribution follows a reasonable trend among the studied compartments, where the highest soil carbon stock is observed under the plots with higher tree carbon stocks.

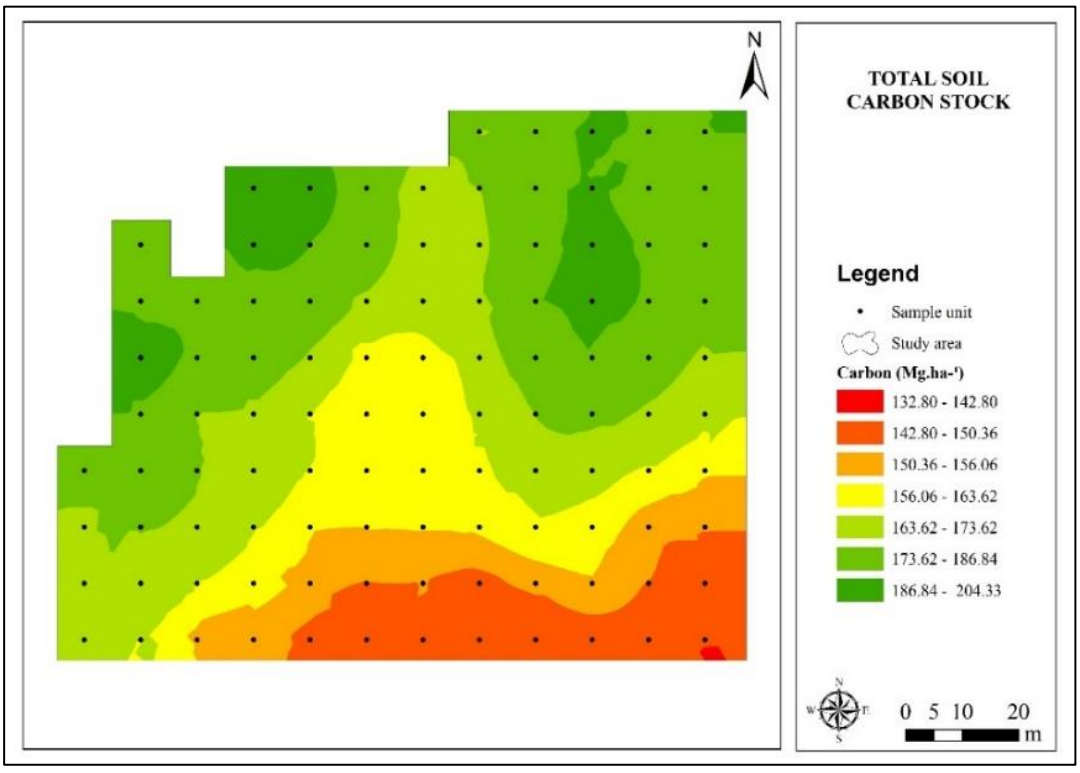


Fig 7. Spatial distribution map of the total soil carbon stock under a forest fragment with a predominance of *Anadenanthera* sp.

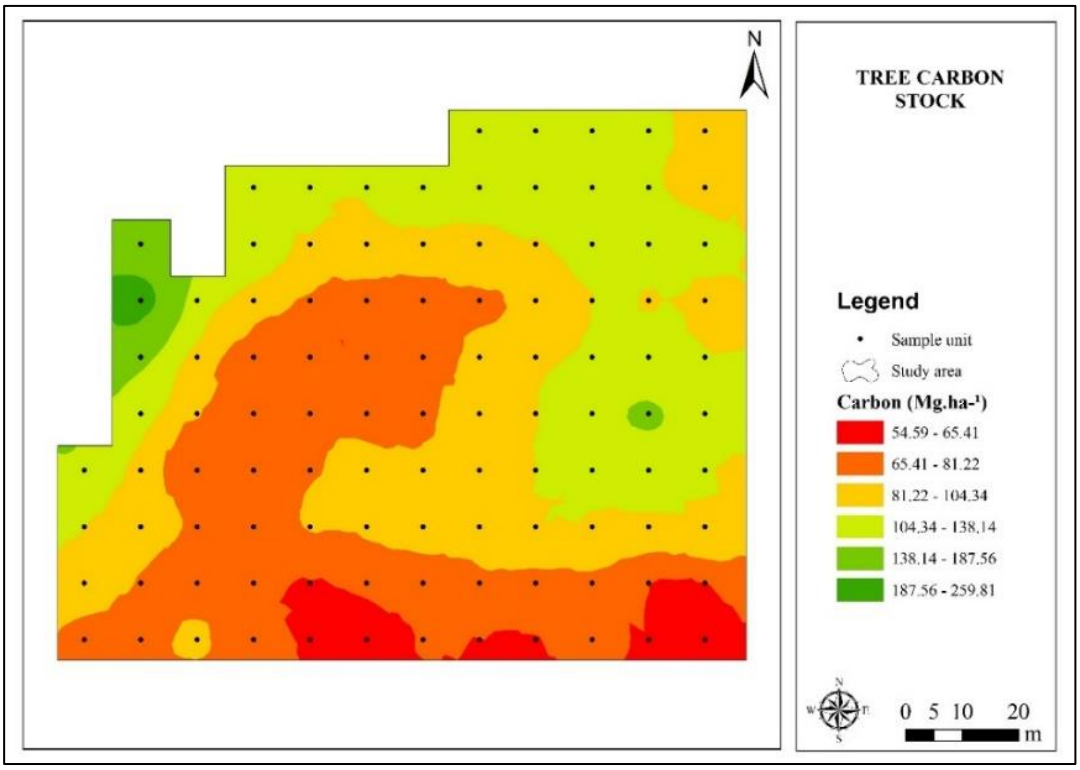


Fig 8. Spatial distribution map of the tree carbon stock in a forest fragment with predominance of *Anadenanthera* sp.

When evaluating the north-south gradient observed in the soil carbon stock, it is important to consider the vegetation around the studied area. On the northern side, there is a mature eucalyptus stand (15 years old) consisting of trees with an average diameter of 23 cm and an average height of 25 m. This eucalyptus stand may explain the greater amount of carbon in the soil on the northern side due to biomass recycling and litter deposition, influencing not only the increase of in the amount of carbon but also nutrients cycling (Wink et al., 2018), which is essential for the growth and increment of vegetation. Dantas et al. (2018), when analyzing the influence of soil on the distribution of tree species in the Cerrado, observed that soil organic matter influences the distribution of species and the number of trees. This finding explains the gradient also observed in the tree carbon stock.

The roots carbon stock maps indicate that there is no horizontal variation at each depth, but there is decreasing trend in root carbon stock depths. The depth increases and the carbon stock decreases. However, when analyzing the total carbon map of the root, there is a spatial structure in the distribution of this variable, which does not follow the gradient observed in the soil carbon stock map nor the tree carbon stock map. Mou et al., (1995) studied the relationship

between aboveground and belowground biomass and found that fine roots had weak and often nonsignificant correlations with local aboveground biomass. The authors suggested that the heterogeneity in soil resources at small scales ay be a key mechanism that controls the spatial distribution of fine roots and that is crucial for understanding belowground interactions in plant communities.

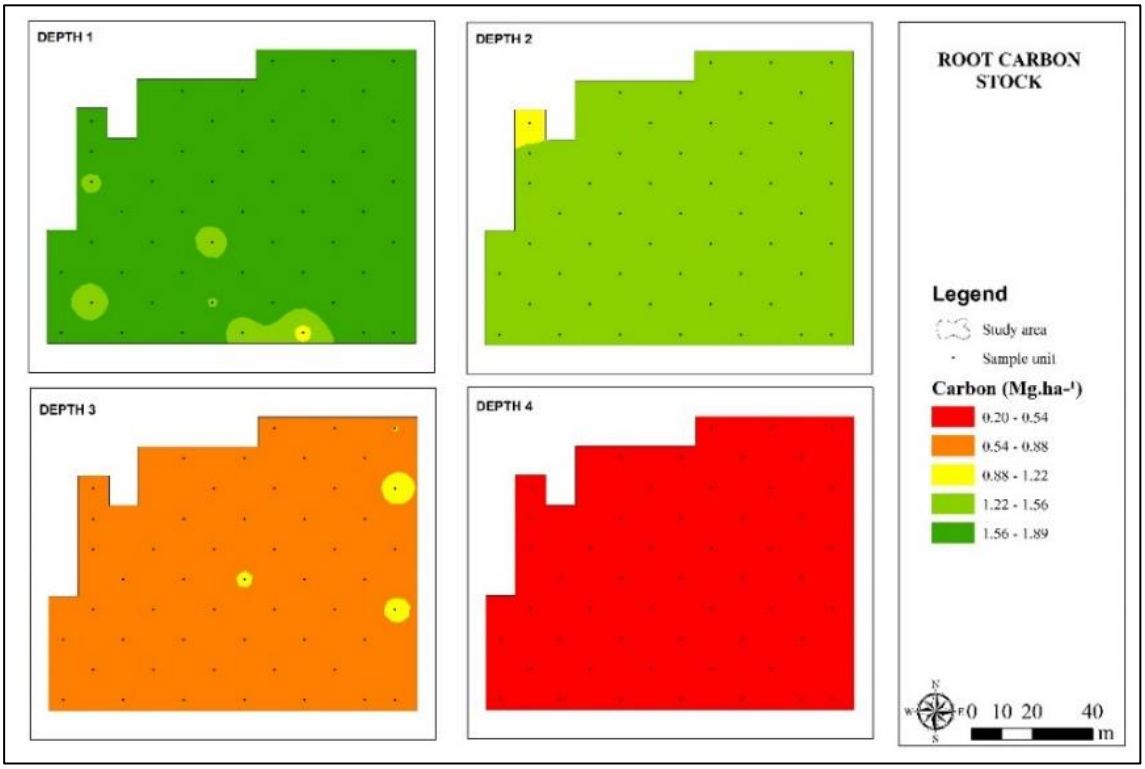


Fig 9. Spatial distribution maps of root carbon stocks under a forest fragment with a predominance of *Anadenanthera* sp. at the depths 1 (0-20 cm), 2 (20-40 cm), 3 (40-60 cm), and 4 (60-80 cm).

There is an east-west horizontal gradient of the total root carbon stock, albeit with a low variation from 4.28 to 3.63 $Mg \cdot ha^{-1}$. The higher root carbon stock observed on the eastside side may be associated with the edge effect in this region, where lower competition may favor greater root development.

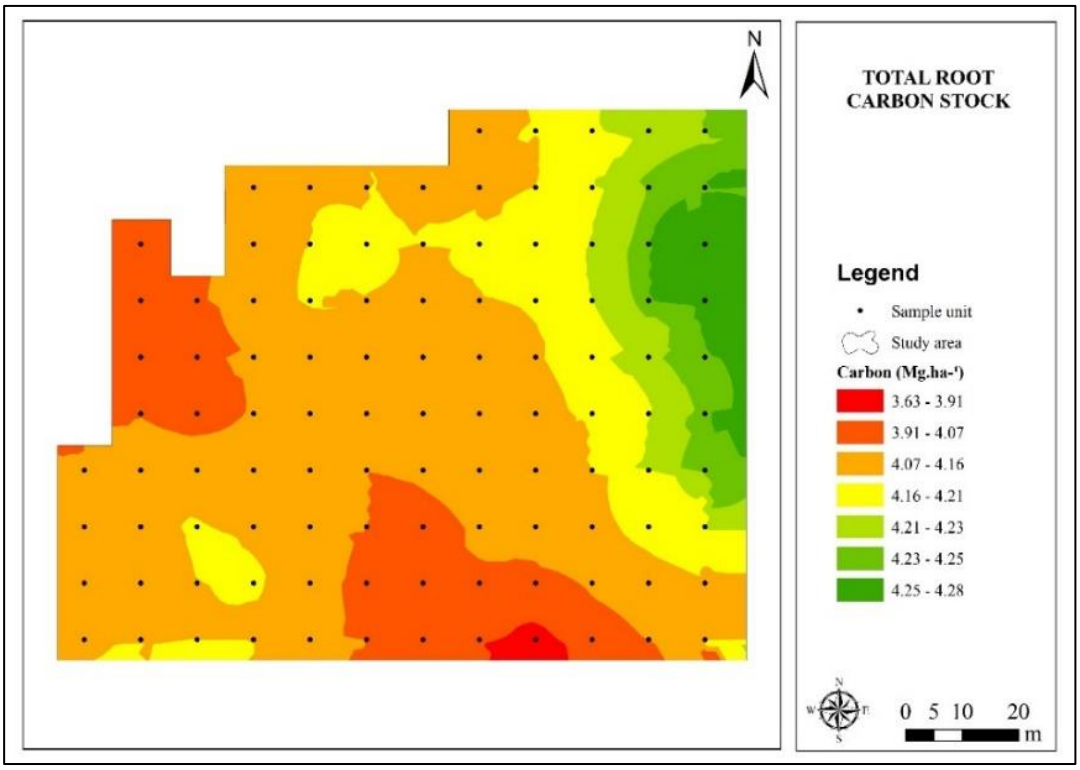


Fig 10. Spatial distribution map of the total root carbon stock under forest fragment with predominance of *Anadenanthera* sp.

The map of total aboveground and belowground carbon stock shows the spatial distribution of the carbon stock in the three compartments analyzed (soil, roots, and trees). The presence of a north-south gradient is evidence, influenced by the higher soil and tree carbon stocks when compared to the root carbon stock. There is high variability in the values found I the entire area, ranging from 201.79 to 423.27 $Mg \cdot ha^{-1}$. The lowest carbon stock values occur along the southern border, where there is a low-carbon, secondary sucessional semi-deciduous forest.

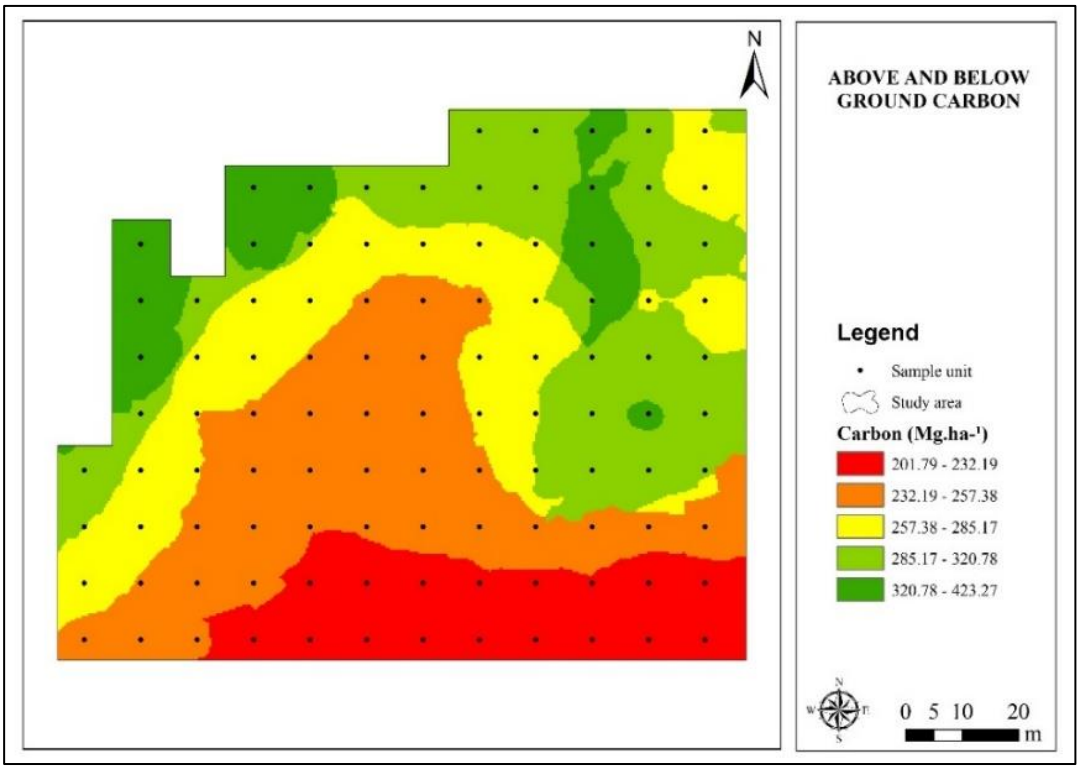


Fig 11. Spatial distribution map of the total aboveground and belowground carbon stock in a forest fragment with predominance of *Anadenanthera* sp.

This study provides high-quality *in situ* data on carbon stock and compartments, which is important for understand carbon stocks and cycling controls, to calibrating global models of the carbon cycle, and supporting regulatory frameworks such as the United Nations REDD+ program.

CONCLUSION

Soil and tree carbon stock have a moderate degree of spatial dependence, while the root carbon stock exhibits a weak degree of dependence.

In the soil, a spatial pattern of the carbon stock is repeated at all depths analyzed, but there is a reduction in the amount of carbon as the depth increases. Tree carbon stock follows the same pattern as the soil, indicating that there is a relationship between these variables. In

the fine roots, the carbon stock decreases with increasing depth, but the spatial gradient does not follow the same pattern as the soil and tree carbon stock.

AKNOLEGMENTS

The authors are grateful for the crucial financial support provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), the Federal University of Lavras, MG, Brazil, and the Science Forest Department.

REFERENCES

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. D. M., Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22, 711–728. doi: 10.1127/0941-2948/2013/0507.

Amaral, L. de P., Ferreira, R. A., Watzlawick, L. F., & Genú, A. M. (2010). Análise da distribuição espacial de biomassa e carbono arbóreo acima do solo em Floresta Ombrófila Mista. *Ambiência*, *6*, 103–114.

Asner, G. P., & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment*, *140*, 614–624. doi: 10.1016/j.rse.2013.09.023

Baker, T. T., Conner, W. H., Lockaby, B. G., Stanturf, J., & Burke, M. K. (2001). Fine Root Productivity and Dynamics on a Forested Floodplain in South Carolina. *Soil Science Society of America Journal*, *66*(2), 671. doi: 10.2136/sssaj2002.6710

Balbinot, R., Shumacher, M. V., Watzlawick, L. F., & Sanquetta, C. R. (2003). Inventário do carbono orgânico em um plantio de Pinus taeda aos 5 anos de idade no Rio Grande do Sul. *Revista Ciências Exatas e Naturais*, *5*(1), 59–68.

Bhattacharya, S. S., Kim, K. H., Das, S., Uchimiya, M., Jeon, B. H., Kwon, E., & Szulejko, J. E. (2016). A review on the role of organic inputs in maintaining the soil carbon pool of the terrestrial ecosystem. *Journal of Environmental Management*, *167*, 214–227. doi: 10.1016/j.jenvman.2015.09.042

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Science Society of America Journal*, *58*(2), 1501–1511. doi: 10.2136/sssaj1994.03615995005800050033x

Carvalho, S. de P. C. e, Mendonça, A. R. de, Lima, M. P. de, & Calegario, N. (2010). Different strategies to estimate the commercial volume of *Anadenanthera colubrina* (Vell.) Brenan. *CERNE*, *16*(3), 399–406. doi: 10.1590/S0104-77602010000300016

Cerri, C. E. P., Bernoux, M., Chaplot, V., Volkoff, B., Victoria, R. L., Melillo, J. M., … Cerri, C. C. (2004). Assessment of soil property spatial variation in an Amazon pasture: Basis for selecting an agronomic experimental area. *Geoderma*, *123*(1–2), 51–68. doi: 10.1016/j.geoderma.2004.01.027

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., … Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190. doi: 10.1111/gcb.12629

Chaves, L. H. G., & Farias, C. H. A. (2010). Variabilidade espacial do estoque de carbono nos Tabuleiros Costeiros da Paraíba: Solo cultivado com cana-de-açúcar. *Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences*, *3*(1), 20–25. doi: 10.5039/agraria.v3i1a235

Correa, J., van der Hoff, R., & Rajão, R. (2019). Amazon Fund 10 Years Later: Lessons from the World's Largest REDD+ Program. *Forests*, *10*(3), 272. doi: 10.3390/f10030272

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: J. Wiley.

Crowther, T. W., Todd-Brown, K. E. O., Rowe, C. W., Wieder, W. R., Carey, J. C., MacHmuller, M. B., … Bradford, M. A. (2016). Quantifying global soil carbon losses in response to warming. *Nature*, *540*(7631), 104–108. doi: 10.1038/nature20150

Dantas, D., Souza, M. J., Vieira, A., Oliveira, M., Pereira, I., Machado, E., … Rocha, W. (2018). Soil influences on tree species distribution in a rupestrian cerrado area. *Floresta e Ambiente*, *25*(4). doi: 10.1590/2179-8087.060517

Dixon, R. K., Brown, S., Houghton, R. A., Solomon, A. M., Trexler, M. C. & Wisniewski. (1994). Carbon Pools and Flux of Global Forest Ecosystems. *Science, 263,* 185-190.

Djomo, A. N., Knohl, A., & Gravenhorst, G. (2011). Estimations of total ecosystem carbon pools distribution and carbon biomass current annual increment of a moist tropical forest. *Forest Ecology and Management*, *261*(8), 1448–1459. doi: 10.1016/j.foreco.2011.01.031

Don, A., Schumacher, J., & Freibauer, A. (2010). Impact of tropical land use change on soil organic carbon stocks - a meta-analysis. *Global Change Biology*, *17*(4), 1658. doi: 10.1111/j.1365- 2486.2010.02336.x

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., … Williams, M. (2019). The Importance of Consistent Global Forest Aboveground Biomass Product Validation. *Surveys in Geophysics*, (May). doi: 10.1007/s10712-019-09538-8

Fan, H., Wu, J., Liu, W., Yuan, Y., Hu, L., & Cai, Q. (2015). Linkages of plant and soil C:N:P stoichiometry and their relationships to forest growth in subtropical plantations. *Plant and Soil*, *392*(1–2), 127–138. doi: 10.1007/s11104-015-2444-2

Figueiredo, L. T. M., Soares, C. P. B., de Sousa, A. L., Leite, H. G., & Da Silva, G. F. (2015). Dinâmica do Estoque de Carbono em Fuste de Árvores de uma Floresta Estacional Semidecidual. *Cerne*, *21*(1), 161–167. doi: 10.1590/01047760201521011529

Franzluebbers, A. J. (2002). Soil organic matter stratification ratio as an indicator of soil quality. *Soil & Tillage Research*, *66*, 95–106. doi: 10.1016/S0167-1987(02)00018-1

Gibbon, A., Silman, M. R., Malhi, Y., Fisher, J. B., Meir, P., Zimmermann, M., … Garcia, K. C. (2010). Ecosystem Carbon Storage Across the Grassland-Forest Transition in the High Andes of Manu National Park, Peru. *Ecosystems*, *13*(7), 1097–1111. doi: 10.1007/s10021-010-9376-8

Haag, H. P. (1985). *Ciclagem de nutrientes em florestas tropicais*. Campinas: Fundação Cargill.

Hendrick, R. L., & Pregitzer, K. S. (1996). Temporal and Depth-Related Patterns of Fine Root Dynamics in Northern Hardwood Forests. *Journal of Ecology*, *84*(2), 167–176.

Hossain, M. F., Chen, W., & Zhang, Y. (2015). Bulk density of mineral and organic soils in the Canada's arctic and sub-arctic. *Information Processing in Agriculture*, 2(3–4), 183–190. doi: 10.1016/j.inpa.2015.09.001

Jaloviar, P., Bakošová, L., Kucbel, S., & Vencurik, J. (2009). Quantity and distribution of fine root biomass in the intermediate stage of beech virgin forest badínsky prales. *Journal of Forest Science*, *55*(11), 502–510.

Jobbagy, E. G., & Jackson, R. B. (2000). The Vertical Distribution of Soil Organic Carbon and its Relation to Climate and Vegetation. *Ecological Applications*, *10*(April), 423–436. doi: 10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

King, G. M. (2011). Enhancing soil carbon storage for carbon remediation: Potential contributions and constraints by microbes. *Trends in Microbiology*, *19*(2), 75–84. doi: 10.1016/j.tim.2010.11.006

Laclau, J., Toutain, F., M'Bou, A. T., Arnaud, M., Joffre, R., & Ranger, J. (2004). The Function of the Superficial Root Mat in the Biogeochemical Cycles of Nutrients in Congolese Eucalyptus Plantations. *Annals of Botany*, *93*(3), 249–261. DOI: 10.1093/aob/mch035

Mafra, Á. L., Guedes, S. de F. F., Klauberg Filho, O., Santos, J. C. P., Almeida, J. A. de, & Dalla Rosa, J. (2008). Carbono Orgânico e Atributos Químicos do Solo em Áreas Florestais. *Revista Árvore*, *32*(2), 217–224. doi: 10.1590/S0100-67622008000200004

Marques, R. R. P. V, Terra, M. C. N. S., Mantovani, V. A., Rodrigues, A. F., Pereira, G. A., Silva, R. A., & Mello, C. R. (2019). Rainfall water quality under different forest stands. *Cerne*, *25*(1), 8–17. doi: 10.1590/01047760201925012581

Mello, J. M. de, Diniz, F. S., Oliveira, A. D. de, Mello, C. R. de, Scolforo, J. R. S., & Acerbi Junior, F. W. (2009). Continuidade espacial para características dendrométricas (número de fustes e volume) em plantios de *Eucalyptus Grandis*. *Revista Árvore*, *33*(1), 185–194. doi: 10.1590/S0100-67622009000100020

Moore, S., Adu-Bredu, S., Duah-Gyamfi, A., Addo-Danso, S. D., Ibrahim, F., Mbou, A. T., … Malhi, Y. (2018). Forest biomass, productivity and carbon cycling along a rainfall gradient in West Africa. *Global Change Biology*, *24*(2), e496–e510. doi: 10.1111/gcb.13907

Mora, F., Jaramillo, V. J., Bhaskar, R., Gavito, M., Siddique, I., Byrnes, J. E. K., & Balvanera, P. (2018). Carbon Accumulation in Neotropical Dry Secondary Forests: The Roles of Forest Age and Tree Dominance and Diversity. *Ecosystems*, *21*(3), 536–550. doi: 10.1007/s10021-017-0168-2

Morais, V. A., Santos, C. A., Mello, J. M., Dadid, H. C., Araújo, E. J. G., & Scolforo, J. R. S. (2017). Distribuição espacial e vertical do carbono em serapilheira e subsolo em vegetação do cerrado brasileiro. *Cerne*, *23*(1), 43–52. doi: 10.1590/01047760201723012247

Mou, R. H. Jones, R. J. M. and B. Z. (1995). Spatial Distribution of Roots in Sweetgum and Loblolly Pine Monocultures and Relations with Above-Ground Biomass and Soil Nutrients *Functional Ecology*, *9*(4), 689–699.

Navarrete-Segueda, A., Martínez-Ramos, M., Ibarra-Manríquez, G., Vázquez-Selem, L., & Siebe, C. (2018). Variation of main terrestrial carbon stocks at the landscape-scale are shaped by soil in a tropical rainforest. *Geoderma*, *313*, 57–68. doi: 10.1016/J.GEODERMA.2017.10.023

Ngo, K. M., Turner, B. L., Muller-Landau, H. C., Davies, S. J., Larjavaara, M., Nik Hassan, N. F. bin, & Lum, S. (2013). Carbon stocks in primary and secondary tropical forests in Singapore. *Forest Ecology and Management*, *296*, 81–89. doi: 10.1016/j.foreco.2013.02.004

Novaes Filho, J. P., Couto, E. G., De Oliveira, V. A., Johnson, M. S., Lehmann, J., & Riha, S. S. (2007). Variabilidade espacial de atributos físicos de Solo usada na identificação de classes pedológicas de microbacias na amazônia meridional. *Revista Brasileira de Ciencia Do Solo*, *31*(1), 91–100. doi: 10.1590/S0100-06832007000100010

Oliveira, M. N. S., Mercadante-Simões, M. O., Ribeiro, L. M., Lopes, P. S. N., Gusmão, E., & Dias, B. A. S. (2005). Efeitos alelopáticos de seis espécies arbóreas da família Fabaceae Allelopathics effects of six trees from the Fabaceae family. *Unimontes Científica*, *7*(2), 121–128.

Paiva, A. O., & Faria, G. E. De. (2007). Estoque de carbono do solo sob cerrado *sensu stricto* no Distrito Federal, Brasil. *Revista Trópica- Ciências e Biológicas*, *1*(1), 59–65. doi: 10.1590/S0100-67622011000300015

Parras-Alcántara, L., Lozano-García, B., & Galán-Espejo, A. (2015). Soil organic carbon along an altitudinal gradient in the Despenaperros Natural Park, southern Spain. *Solid Earth*, *6*(1), 125–134. DOI: 10.5194/se-6-125-2015

Powers, J. S., Corre, M. D., Twine, T. E., & Veldkamp, E. (2011). Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation. *Proceedings of the National Academy of Sciences*, *108*(15), 6318–6322. doi: 10.1073/pnas.1016774108

R Development Core Team. (2018). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., & Hérault, B. (2017). biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods in Ecology and Evolution*, *8*(9), 1163–1167. doi: 10.1111/2041-210X.12753

Ribeiro Júnior, P. J., & Diggle, P. J. (2001). GeoR: a package for geostatistical analysis. *R News*, *1*(2), 15–18.

Ribeiro, S. C., Jaconive, L. A. G.m Soares, C. P. B., Martins, S. V., Souza, A. L. & Nardelli, A. M. B. (2009). Quantificação de biomassa e estimativa de estoque de carbono em uma floresta madura no município de viçosa, Minas Gerais. *Árvore*, *33* (5), 917-926. doi: 10.1590/S0100-67622009000500014

Santos, M. de C. N. S., Mello, J. M., Mello, C. R., & Ávila, L. F. (2013). Spatial Continuity of Soil Attributes in an Atlantic Forest Remnant in the Mantiqueira Range, MG. *Ciência e Agrotecnologia*, *37*(1), 68–77. doi: 10.1590/S1413-70542013000100008

Shukla, G., & Chakravarty, S. (2018). Biomass, Primary Nutrient and Carbon Stock in a Sub-Himalayan Forest of West Bengal, India. *Ournal of Forest and Environmental Science*, *34*(1), 12–23. doi: 10.7747/JFES.2018.34.1.12

Terra, M. C. N. S., Mello, C. R., Mello, J. M., Oliveira, V. A., Nunes, M. H., Silva, V. O., … Alves, G. J. (2018). Stemflow in a neotropical forest remnant: vegetative determinants, spatial distribution and correlation with soil moisture. *Trees*, *32*(1), 323–335. doi: 10.1007/s00468-017-1634-3

Terra, M. C. N. S., Mello, J. M., & Mello, C. R. (2015). Spatial relationship of vegetation Carbon stock and soil organic matter at Serra da Mantiqueira. *Floresta e Ambiente*, *22*(4), 446–455. doi: 10.1590/2179-8087.059713

Thomas, S. C., & Martin, A. R. (2012). Carbon content of tree tissues: A synthesis. *Forests*, *3*(2), 332–352. doi: 10.3390/f3020332

van der Sande, M. T., Poorter, L., Kooistra, L., Balvanera, P., Thonicke, K., Thompson, J., … Peña-Claros, M. (2017). Biodiversity in species, traits, and structure determines carbon stocks and uptake in tropical forests. *Biotropica*, *49*(5), 593–603. doi: 10.1111/btp.12453

Veldkamp, E. (1994). Organic Carbon Turnover in Three Tropical Soils under Pasture after Deforestation. *Soil Science Society of America Journal*, *58*, 175–180. doi: 10.2136/sssaj1994.03615995005800010025x

Watson, R. T. (Ed.). (2001). *Climate change 2001: synthesis report: third assessment report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press.

Webster, R., & Oliver, M. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: J. Wiley.

Wink, C., Lange, A., Araújo, K. Z., Silveira, A. P., Behling, M., Wruck, F. J., & Zandonadi, R. S. (2018). Biomassa e Nutrientes de Eucalipto Cultivado em Sistema Agrossilvipastoril. *Nativa*, *6*, 754. doi: 10.31413/nativa.v6i0.5987

ARTICLE 2 - MACHINE LEARNING FOR CARBON STOCK PREDICTION IN A TROPICAL FOREST IN SOUTHEASTERN BRAZIL

ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA REVISTA BOSQUE

Situação: Publicado na Revista Bosque

Machine learning for carbon stock prediction in a tropical forest in southeastern Brazil

SUMMARY

The increasing awareness of global climate change has drawn attention to the role of forests as mitigators of this process as they act as carbon sinks to the atmosphere. Understanding the process of carbon storage in forests and its drivers, as well as presenting consistent models for their estimation, is a current demand. In this sense, the aim of this study was to evaluate the performance of machine learning techniques: support vector machine (SVM) and to propose a new nonlinear model extracted from the training of an artificial neural network (ANN) in the modeling the above ground carbon stock in a secondary semideciduous forest. The SVM and ANN construction and training process considered independent variables selected by stepwise: minimum DBH (diameter of breast height - 1.3 m), maximum DBH, mean DBH, total height and number of trees, all by plot. The SVM and the model extracted from ANN were applied to the data set intended for validation. Both techniques presented satisfactory performance in modeling carbon stock by plot, with homogeneous distribution and low dispersion of residues and predicted values close to those observed. The analysis criteria indicated superior performance of the model extracted from the artificial neural network, which presented a mean relative error of 6.94 %, while the support vector machine presented 13.52 %, combined with lower bias values and a higher correlation between predictions and observations.

Key words: artificial intelligence, artificial neural networks, support vector machines, forest biomass.

INTRODUCTION

Forests provide numerous ecosystem services, such as regulation of biogeochemical cycles, pollution control and food supply. Among the most acclaimed ecosystem services

provided by forests are the atmospheric carbon (CO_2) sequestration and its storage (Canadell and Raupach 2008, Chazdon *et al.* 2016). This service is of strategic importance in mitigating ongoing climate change because it acts directly in controlling global warming (Bonan 2008).

In this context, the quantification of the carbon stock present in the most varied types of forests constitutes an important tool for monitoring this ecosystem service (Scolforo *et al.* 2015). Many initiatives have been taken to quantify carbon stocks in forests, both by direct means (Dantas *et al.* 2021) and through estimates from related data (indirect methods) (Cordeiro *et al.* 2018).

Carbon stock estimation by indirect methods employs modeling and simulation techniques. Historically, modeling of forest attributes has relied on approaches based on statistical models (e.g. Melo *et al.* 2017). These approaches share space today with computational approaches of artificial intelligence/machine learning, such as artificial neural networks, support vector machines, decision trees, among others, which have been gaining space as tools for forest data analysis, modeling, estimation, and production prognosis. These tools have provided gains in the quality of estimates and predictions (Vendruscolo *et al.* 2015).

Artificial Neural Network (ANN) is a processor consisting of simple processing units (artificial neurons), based on neurons found in the human brain, that calculate certain functions. These units are layered and connected to each other by weights that store experimental knowledge and weight the inputs of each unit. Thus, the acquired knowledge becomes available for use (Braga *et al.* 2007, Dantas *et al*. 2020a).

The most notable features in ANN are their ability to learn and to generalize information. In other words, ANN are able, through a learned example, to generalize assimilated knowledge to an unknown data set. Another interesting feature of ANN is the ability to extract non-explicit features from a set of information provided as examples (Haykin 2001).

Support vector machines (SVM) have also proven to be an interesting alternative for mathematical modeling of complex systems (Heddam and Kisi 2018). They are simple techniques in their conceptual basis and capable of solving extremely complex real problems. SVM is a supervised learning technique that is trained to classify different categories of data from various disciplines (Haykin 2001). These have been used for two-class classification problems and are applicable on both linear and non-linear data classification tasks. SVM creates a hyperplane or multiple hyperplanes in a high-dimensional space, and the best hyperplane in

them is the one that optimally divides data into different classes with the largest separation between the classes (Steinwart and Christmann 2008).

Initially, SVM techniques were successfully applied as a data classification methodology (Tong and Koller 2001). They were later extended to the regression tasks through the following approaches: support vector regression (SVR) and least-square support vector machines (LS-SVM) (Cherkassky and Mulier 1998, Dantas 2020b).

Compared to ANN, SVM has the advantage of leading to an exact solution, that is, a global optimum (Haykin 2001). However, finding a final SVM model may present computational complexity because it requires solving a quadratic programming model and solving a set of nonlinear equations.

The present study aims to evaluate the performance of the support vector machine and artificial neural network techniques and to propose a new nonlinear model to the modeling of above ground biomass (carbon stock), using dendrometric variables as inputs, in a secondary semideciduous seasonal forest. It is proposed as hypothesis that (a) machine learning techniques are suitable in modeling above ground biomass; (b) it is possible to extract accurate above ground biomass equation from the artificial neural network training process.

METHODS

Study area and data collection.

The study area corresponds to a secondary semideciduous seasonal forest, located in Lavras, Minas Gerais, Brazil, under the coordinates 21° 14' S and 45° 00' W, with an average altitude of 900 m (figure 1). The climate is classified as Köppen's Cwb, with dry winters and mild summers (Alvares *et al.* 2013). The mean annual rainfall is around 1,500 mm and the mean annual temperature is 19.4 °C (Marques et al., 2019). The forest is heterogeneous and presents dominance of tree species of the genus *Anadenanthera*, popularly known as “angico”. The data come from 105 sample plots (10x10 m) launched in the area. We measured in each plot all trees

with diameter at breast height (DBH - 1.3 m from the ground) greater than or equal to 5 cm and their respective heights.

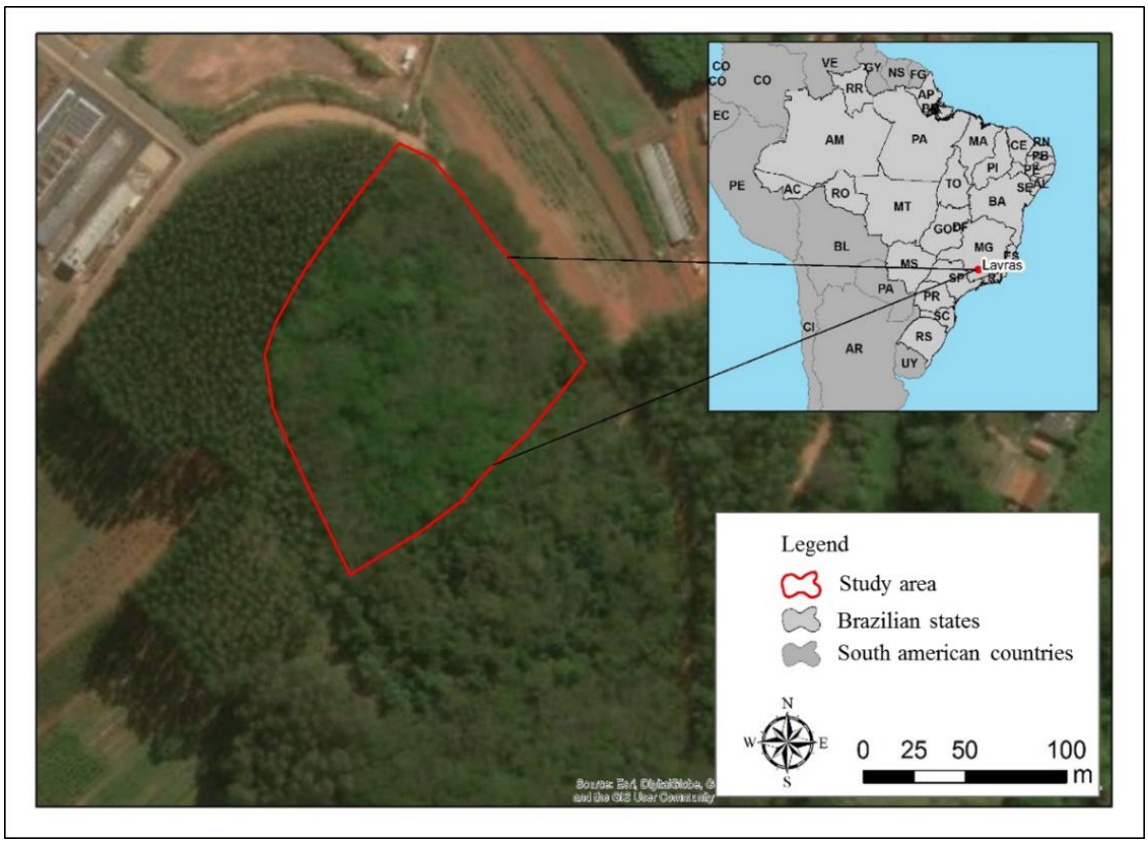


Figure 1. Study area. Secondary semideciduous forest path located in Lavras, Minas Gerais, Brazil. (Adapted from Terra *et al.* 2019)

From the data collected in the field, we obtained the following variables by plot: minimum DBH (DBHmin), mean DBH (DBHmed), maximum DBH (DBHmax), minimum total Height (Hmin), mean total Height (Hmed), maximum total Height (Hmax), Mean Square Diameter (Dq) and Number of Trees (N).

Above Ground Biomass (AGB) was estimated by tree individual according to the equation proposed by Chave *et al*. (2014), using DBH and total tree height and an average basic wood density of 0.620 g cm^{-3}. The estimate was performed using the software R (R Core Team 2018), using the BIOMASS package (Réjou-Méchain *et al*. 2017). The estimate of AGB was converted to carbon stock in Mg ha^{-1}, according to Thomas and Martin (2012), a methodology

consisting in multiplying AGB by 0.471, which according to the authors corresponds to carbon concentration in tropical forest angiosperms tissues.

Independent variables selection.

First, we selected the independent variables by the stepwise method, based on the Akaike Information Criterion (AIC). Thus, the best model combines variables with the lowest AIC. Then the selected variables were used as inputs to model the carbon stock by plot through machine learning techniques.

Machine learning algorithms.

For carbon stock modeling, support vector machines (SVM) and artificial neural networks (ANN) were used. The SVM construction was based on the supervised machine learning process described by Haykin (2001) and Steinwart and Christmann (2008), where there is a set of paired order n samples (**X**, **Y**), where **X** is a matrix of explanatory variables of the sample and Y is the expected value vector of the sample. Based on this information, taking as input a vector of variables, a chosen function predicts the expected value of the sample. A linear function is given by the form f (**X**) = <**W**, **X**> + b, where **W** is a weight vector.

The type IV error function, also known as *eps-regression*, was used, being the RBF (Radial Basis Function) type Kernel function. Kernel functions offer an alternative solution by designing data in a space with large characteristics to increase the computational power of machine learning, making it possible to represent nonlinear phenomena (Cristianini and Shawe-Taylor 2000). This procedure was performed in software R, version 3.4.1, through the *e1071* package (Meyer *et al*. 2019).

The trained ANN were Multilayer Perceptron (MLP), composed of an input layer, an intermediate layer and an output layer. The algorithm used was the resilient backpropagation,

in which the learning rate was automatically defined by the *neuralnet* package, with values ranging from 0.01 to 1.12.

The choice of the number of neurons in the hidden layer was made using k-fold. This methodology randomly subdivides the database into k subgroups (Ali and Pazzani 1996, Cigizoglu and Kisi 2006). The value of k was 10 subgroups, with a proportion of 90 % for training and 10 % for testing (Diamantopolou 2010), applying cross-validation. Different numbers of neurons, ranging from 1 to 20, were tested.

The activation function used was the logistics (or sigmoidal), with a range from 0 to 1, which implies limiting the amplitude of the outputs and inputs. For this reason, data were normalized, which consists of transforming the values of each variable to values between 0 and 1. The linear standardization was obtained through equation [1] (Soares *et al*. 2011) and considers the minimum and maximum value of each variable in the transformation of values, maintaining the original distribution of data (Valença 2010).

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \qquad [1]$$

where: *x'*: normalized value; x: original value; *xmin*: minimum value of the variable; *xmax*: maximum value of the variable; *a*: lower limit of the standardization range; *b*: upper limit of the standardization range.

The stopping criterion for the ANN training process was the maximum number of 100.000 cycles, or the mean square error of less than 1 %, and training was terminated when one of these criteria was met. At the end of the training, the best ANN was selected based on the lowest mean square error.

From the artificial neural network, it was extracted a nonlinear equation for tree biomass prediction. For this, we generated a system of equations with coefficients resulting from the weights generated by the neurons of ANN. This system was used to predict the carbon stock of the plots that comprised the validation database.

Data were divided into two groups: 70 % for ANN training and SVM construction and 30 % for validation of both techniques. Among the data intended for ANN training, 70 % were used in the training phase and 30 % in the test phase.

SVM and ANN performance evaluation.

SVM and ANN performances were evaluated in the training and validation phases. For this, the techniques were used to predict the carbon stock in the data set intended for validation, *i.e.* data that had not been used in training. Prediction quality analysis was performed using Mean Relative Error (MRE %) (equation 2), Bias (equation 3), Root Mean Square Error (RMSE %) (equation 4) (Leite and Andrade 2002, Siipilehto 2000), graphs of residuals distribution, graphs of estimated versus observed carbon stocks and the correlation coefficients between estimated and observed values.

$$MRE\ (\%) = \frac{(\widehat{Yi} - Yi)}{Yi} \times 100 \quad [2]$$

$$Bias\ (\%) = 100 * \frac{1}{n} * \sum_{i=1}^{n} \left[\frac{\left(\widehat{Yi} - Yi \right)}{Yi} \right] \quad [3]$$

$$RMSE\ (\%) = \left[\sqrt{\sum_{i=1}^{n} \frac{\left(\widehat{Yi} - Yi \right)^2}{n - p}} / \bar{Y} \right] * 100 \quad [4]$$

where: *Yi* represents the observed value, $\hat{Y}i$ the estimated value, *n* the number of observations and $\bar{Y}$ the average of the observed values.

RESULTS

The forest with predominance of *Anadenanthera* sp. contained an average of tree carbon stock (AGC) of 94.25 Mg ha^{-1}. Descriptive statistics of the variables used are presented in table 1.

Table 1. Descriptive statistics of dendrometric variables of a secondary seasonal semideciduous forest located in Lavras, Minas Gerais, Brazil.

Variable	Mean	Minimum	Maximum	Standard deviation	CV(%)
C	94.25	15.65	245.09	52.57	71.14
DBHmin	5.84	5.00	9.07	0.90	15.34
DBHmax	29.61	12.50	55.25	8.20	29.96
DBHmed	12.72	8.69	19.83	2.50	20.27
Hmin	6.65	5.24	11.43	1.41	21.12
Hmax	22.00	14.86	29.00	2.39	11.05
Hmed	13.18	9.84	18.83	1.76	13.38
N	14.25	3.00	30.00	4.96	34.90
Dq	14.59	9.16	22.79	3.11	22.74

Where: C = carbon stock ($Mg.ha^{-1}$); DBHmin = minimum breast height diameter (DBH) of the sample plot (cm); DBHmax = maximum DBH of the sample plot (cm); DBHmed = mean DBH of the sample plot (cm); Hmin = minimum total height of the sample plot (m); Hmax = maximum total height of the sample plot (m); Hmed = mean total height of the sample plot (m); N = number of trees in the sample plot; Dq = mean square diameter of the sample plot (cm).

Stepwise method indicated through the Akaike information criterion that the variables minimum DBH, maximum DBH, average DBH, average H and N are those that have a stronger influence on the carbon stock variability, and were therefore selected for modeling. It should be noted that DBHmed and Hmed represent the main tree growth trends of in each plot; DBHmin and DBHmax, the lower and upper limits of diameter growth, respectively; and N represents the density of individuals in each plot. Figure 2 presents the scatter plots between the variables and their respective confidence intervals and the distribution of each variable.

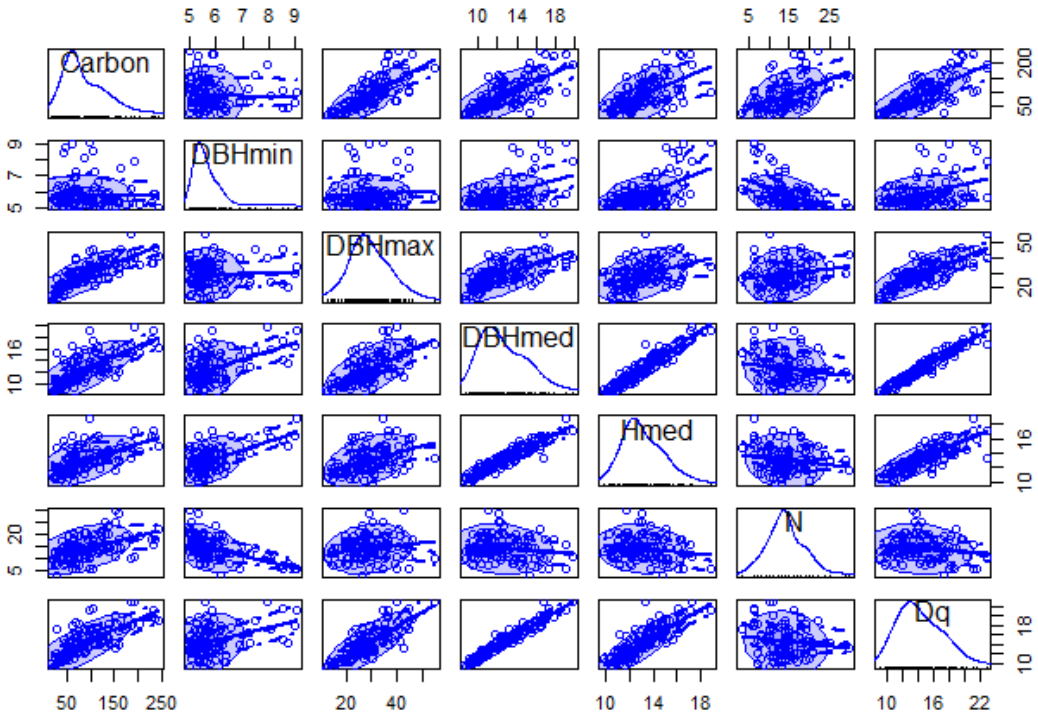


Figure 2. Distribution and dispersion of the independent variables selected by stepwise for carbon stock modelling and their respective confidence intervals.

The configurations obtained with the construction of the support vector machine, which resulted in a machine with 44 support vectors, are presented in table 2.

Table 2. Support Vector Machine Parameters in the estimation of carbon stock in a secondary semideciduous forest.

SVM	Parameters
Type	*eps-regression*
Kernel	Radial basis function
Cost	1
Gamma	0.2
Epsilon	0.1
Numbers of support vectors	40

Regarding the approach by artificial neural networks, figure 3 illustrates the architecture and the weights obtained from the selected ANN that presented the lowest error among the others evaluated, composed by six neurons in the hidden layer.

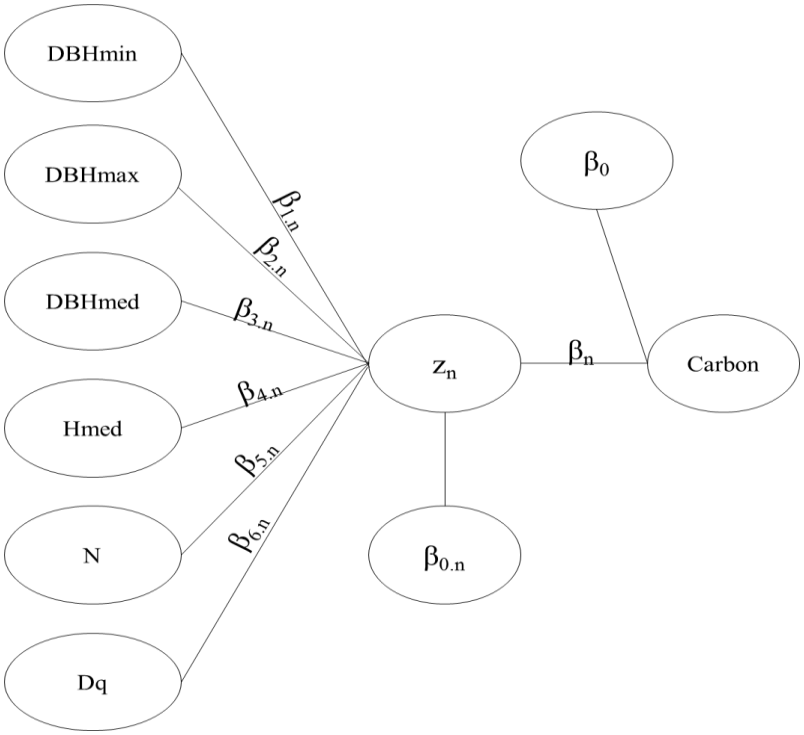


Figure 3. Architecture of the selected artificial neural network, composed of six neurons in the hidden layer. Input variables: DBHmin = minimum plot DBH (cm); DBHmax = maximum DBH of the plot (cm); DBHmed = mean DBH of the plot (cm); Hmed = Average total tree height of the plot (m); N = Number of trees in the plot; Dq = mean square diameter of the sample plot (cm). Output variable: C = carbon stock ($Mg.ha^{-1}$); β_0: bias; β_n: coefficient associated with the neuron n; $\beta_{k.n}$: coefficient of the model between the input variable k and the neuron n; z_n: response of n-th hidden layer neuron.

From the 6-6-1 architecture artificial neural network, an equation system was extracted to predict the carbon stock per plot, with coefficients derived from the weights generated by the neural network neurons. This system was used to predict the carbon stock of the plots that make up the database for validation.

Model [5] expresses the relationship between the hidden layer and the response variable, where β0 is the bias and the other coefficients are the weights related to each neuron. Model [6]

represents the activation function used in each hidden layer neuron, derived from the logistic model. Finally, model [7] is the result of the relationship between the input variables and the respective hidden layer neurons, and a model is generated for each neuron.

$$Carbon' = \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \beta_3 * z_3 + \beta_4 * z_4 + \beta_5 * z_5 + \beta_6 * z_6 \quad [5]$$

$$z_n = \left[\frac{1}{1 + e^{-w_i}}\right] \quad [6]$$

$$w_i = \beta_{0.n} + \beta_{1.n} * DBHmin_i{}' + \beta_{2.n} * DBHmax_i{}' + \beta_{3.n} * DBHmed_i{}' + \beta_{4.n} * Hmed_i{}' + \beta_{5.n} * N_i{}' \quad [7]$$

where: β_0: bias; β_n: model coefficient associated with neuron n; β_n: model coefficient between input variable k and neuron n; z_n: n-th hidden layer neuron response; w_i: sum of products between weights and inputs.

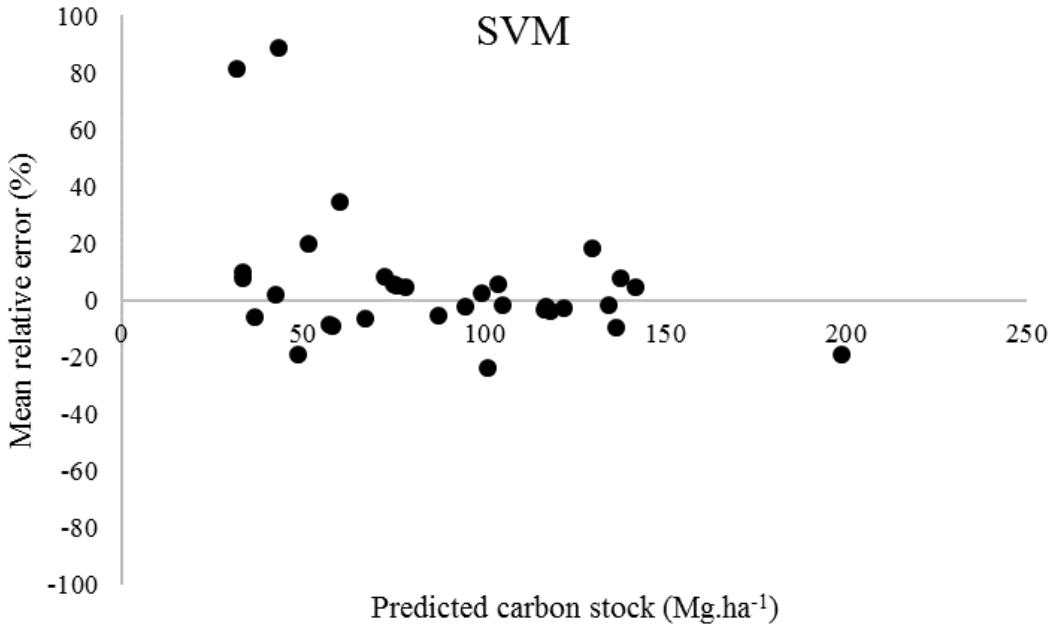
The coefficients of the system of equations extracted from the selected artificial neural network are presented in table 3.

Table 3. Coefficients (β's) of each neuron (N) of the hidden layer and of the ANN output layer.

	β0	β1	β2	β3	β4	β5	β6
ANN	0.8609	-0.7323	0.7619	0.6785	-1.7671	-1.6967	1.0353
N1	-2.3277	2.0301	0.1250	9.7276	-6.8163	6.8954	-
N2	-1.3559	0.6158	2.5005	2.8092	-0.3280	-0.9826	-
N3	-1.7104	2.3549	0.9111	0.6736	-2.8348	6.2268	-
N4	-0.1308	0.1538	0.5098	-1.6662	1.0934	-0.4838	-

N5	0.1174	-1.2125	0.4046	-1.1707	0.2400	-0.3406	-
N6	-0.0674	-2.5130	0.6372	0.7235	-0.2786	1.3868	-

The support vector machine and the model extracted from the artificial neural network were applied to the data set intended for validation. The analyzed techniques presented satisfactory performance in the modeling of carbon stock by plot, due to homogeneous distribution and low dispersion of residues and with predicted values close to those observed, as it can be observed in figures 4 and 5, respectively.



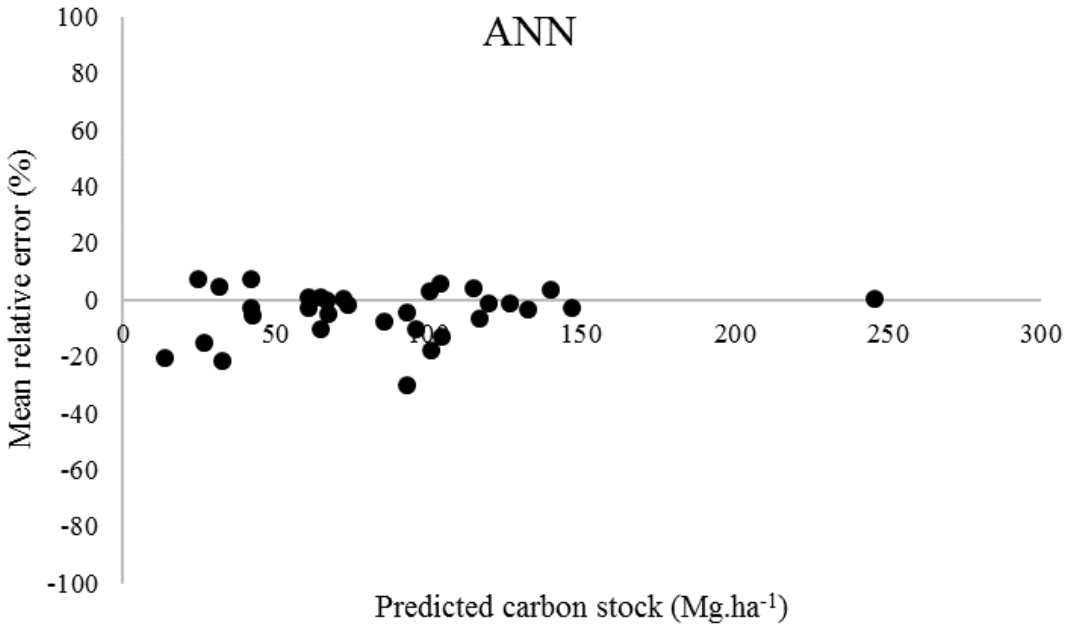


Figure 4. Residuals scatter plots of carbon stock modeling in a secondary semideciduous forest using machine learning algorithms.

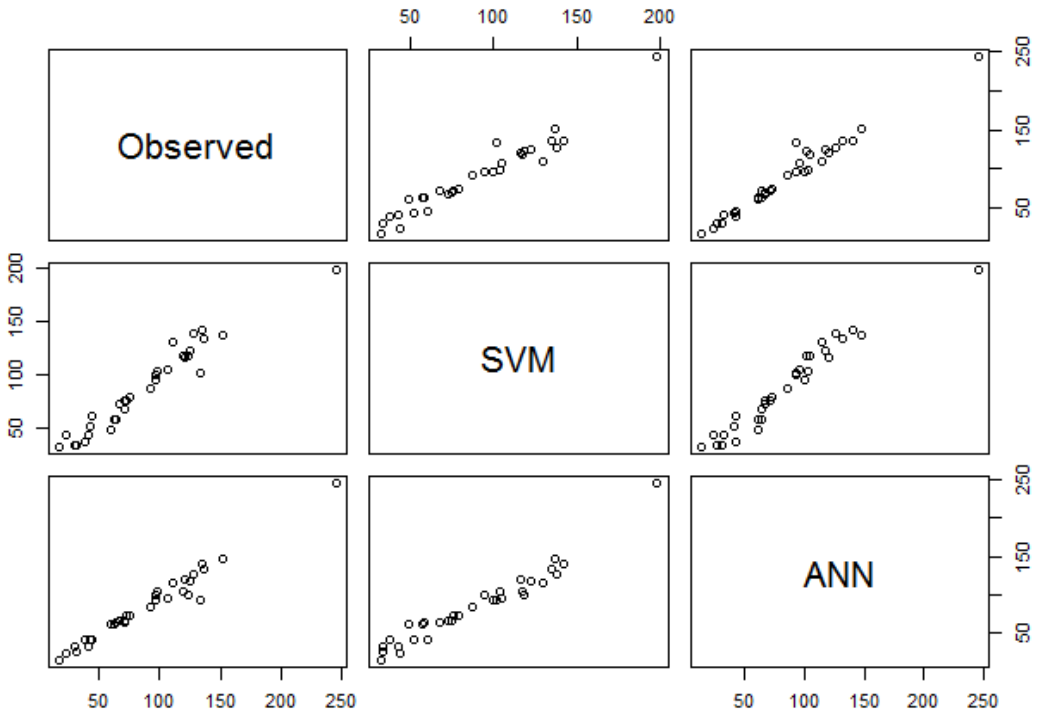


Figure 5. Predicted versus observed graphs of the analyzed techniques (artificial neural network and support vector machine) used to estimate carbon stock in a secondary semideciduous forest.

According to the graphs (figures 4 and 5), there is a slight superiority of ANN, which presented more concentrated residues around zero and predicted values closer to the real ones. SVM presented higher residual dispersion and two relative error values above 80 %, while ANN presented a maximum relative error of 30 %.

The performance evaluation criteria of the analyzed techniques are presented in table 4. The qualities of the estimates made by the SVM and by the model extracted from the ANN were evaluated on both data used in the training and validation dataset.

Table 4. Performance evaluation criteria of the support vector machine (SVM) and artificial neural network (ANN) algorithms used in carbon stock estimation in a secondary semideciduous forest.

	Train		Validation	
	SVM	ANN	SVM	ANN
Mean relative error (%)	6.7689	4.8035	13.5185	6.9375
Bias (%)	-0.2610	-1.2371	5.6479	-4.6217
Root mean square error (%)	7.5858	5.4185	14.5550	10.8191
Coefficient of correlation	0.9900	0.9951	0.9690	0.9828

The estimates of the analyzed techniques were strongly correlated with the observed values, with correlation above 0.99 in the training phase and 0.96 in the validation phase. Error magnitudes, represented by the RMSE, were below 10 % in the training phase and 15 % in the validation. The lower the RMSE, the higher the accuracy of the estimates, and the optimal situation when it is zero (Mehtätalo *et al.* 2006). Bias indicated slight underestimation trends in the training phase whereas in validation there was a tendency of SVM to overestimate the carbon stock values (5.6479 %) and ANN to underestimate them (-4.6217 %). Mean SVM relative error increased from 6.7689 % in the training phase to 13.5185 % in the validation phase whereas in ANN this increase went from 4.8035 % to 6.9375 %.

DISCUSSION

The average of tree carbon stock contained in the forest with predominance of *Anadenanthera* sp. (94,25 Mg ha^{-1}) is above the average carbon stock for this vegetation type in south-central Minas Gerais (55 Mg ha^{-1}, Scolforo *et al*. 2015) and compatible with other local studies in semideciduous seasonal forests in the study region. For instance, Ribeiro *et al.* (2009), quantifying the biomass and tree carbon stock in a mature semideciduous forest in Viçosa, Minas Gerais, Brazil, found 166.67 Mg ha^{-1} and 83.34 Mg ha^{-1} for biomass and carbon stocks, respectively. Also, Figueiredo *et al.* (2015) that evaluated the dynamics of the tree carbon stock in a semideciduous forest in Minas Gerais, Brazil, and found an average carbon stock of 71.81 Mg ha^{-1}.

Although the average carbon stock was relatively high, its variation among plots (CV%) was also significant. This is mainly due to the fact that the carbon stock variable reflects the variations in other dendrometric variables. Moreover, in natural multiage forests, characteristics such as high ecological complexity, spatial variations in species structure and composition, the presence of clearings and other factors can lead to great variability in biomass/carbon stock values (Soriano-Luna *et al*. 2018) among the sample plots.

There is a strong relationship between mean DBH and mean height. Overall, there is a tendency of carbon stock to be positively correlated to the other variables, except for the minimum DBH, in which there was no direct or indirect proportional relationship, which is evidenced by the elliptic shape of the dispersion between these variables. In general, the relationship between carbon and independent variables tended towards linearity. There are, however, some nonlinear behaviors between variables, such as between carbon stock and maximum DBH. In this context, it is worth to emphasize that artificial intelligence/machine learning has the ability to detect implicitly any nonlinear relationship between the response variable and the explanatory variables. In addition to the fact that there are no assumptions regarding the input data, such as independence and normality, and its great capacity for learning and generalization.

The main purpose of using these techniques, as in classical regression, is their application to data that were not used in their training. The ANN presented better performance in the two analyzed phases, training and validation, which indicates its superiority for modeling the carbon stock per plot in the analyzed data set, when compared to SVM.

ANN was able, with the available variables, to explain almost all the variation in carbon stock in the study area. Several studies have demonstrated the superiority of artificial neural networks when compared to other techniques (Özçelik *et al.* 2013, Vendruscolo *et al.* 2015). This superiority can be explained by the ability of neural networks to detect implicit information and nonlinear relationships between the response variable and the explanatory variables provided as examples and to generalize the assimilated knowledge to an unknown data set.

It is worth noting that, although underperforming ANN, SVM was very efficient in estimating carbon stock in the study area. SVM has the advantage over ANN that no evaluation is required after its construction, as it occurs in ANN to select the best network. This is due to the quadratic optimization that occurred during SVM training (Cristianini and Shawe-Taylor 2000) which allows the same result to be obtained for each system configuration whenever applied to the same data set. However, ANN have more elements to be manipulated, besides the initialization of neuron parameters occur at random (Haykin 2001). Thus each trained network will have slight differences in estimates, even if the same architecture is maintained. These differences highlight the practicality of SVM in relation to ANN as SVM excludes the subjectivity of the operator in having to choose the best network to be applied to the database.

Therefore, both approaches were able to explain much of the carbon stock variation in the study area. This is mainly due to the robustness of ANN and SVM. The small part of the carbon stock variation not explained by the variables in question is due to the various factors not considered in the present study that are known to affect the variability of carbon stock in forests, such as species diversity, forest size, degree anthropization, among many others and their interactions (McNicol *et al.* 2018)

The results presented in this study provide insights for future assessments of the use of machine learning techniques to obtain carbon stock estimates. Some examples of potential applications are estimates of carbon stock in other forest compartments, such as soil and tree roots; and carbon stock estimates through the association between machine learning techniques and remote sensing variables; among others.

CONCLUSIONS

The present study brings important contributions in the modeling of carbon stocks in forests through the use of machine learning. The machine learning techniques performed satisfactorily, and a new model extracted from an artificial neural network for carbon stock prediction in Mg ha^{-1} is efficient, with potential application in other secondary semideciduous seasonal forests.

The carbon stock modeling using a model extracted from the artificial neural network training presented better performance than the support vector machine, using the same variables of the analyzed data set.

Determining, modeling, and supplying carbon stock data in forest is a strong scientific and social demand currently, as tree carbon storage is considered a key environmental service in mitigating current climate change by sequestering CO_2 from the forest atmosphere.

AKNOLEGMENTS

The authors are grateful for the financial support provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), the Federal University of Lavras, MG, Brazil, and the Science Forest Department.

REFERENCES

Ali KM, MJ Pazzani. 1996. Error reduction through learning multiple descriptions. *Machine Learning* 24(3):173-202. DOI: 10.1007/BF00058611

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

Bonan GB. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* 320(5882):1444-1449. DOI: 10.1126/science.1155121

Braga AP, APLF Carvalho, TB Ludemir. 2007. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro, Brasil. Editora LTC. 226 p.

Canadell JG, MR Raupach. 2008. Managing forests for climate change mitigation. *Science* 320(5882):1456-1457. DOI: 10.1126/science.1155458

Chave J, M Réjou-Méchain, A Búrquez, E Chidumayo, MS Colgan, WBC Delitti, A Duque, T Eid, PM Fearnside, RC Goodman, M Henry, A Martínez-Yrízar, WA Mugasha, HC Muller-Landau, M Mencuccini, BW Nelson, A Ngomanda, EM Nogueira, E Ortiz-Malavassi, R Pélissier, P Ploton, CM Ryan, JG Saldarriaga, G Vieilledent. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20(10):3177–3190. DOI: 10.1111/gcb.12629

Chave J, M Réjou-Méchain, A Búrquez, E Chidumayo, MS Colgan, WBC Delitti, A Duque, T Eid, PM Fearnside, RC Goodman, M Henry, A Martínez-Yrízar, WA Mugasha, HC Muller-Landau, M Mencuccini, BW Nelson, A Ngomanda, EM Nogueira, E Ortiz-Malavassi, R Pélissier, P Ploton, CM Ryan, JG Saldarriaga, G Vieilledent

Chazdon RL, EN Broadbent, DMA Rozendaal, F Bongers, AMA Zambrano, TM Aide, P Balvanera, JM Becknell, V Boukili, PHS Brancalion, D Craven, JS Almeida-Cortez, GAL Cabral, BD Jong, JS Denslow, DH Dent, SJ Dewalt, JM Dupuy, SM Durán, MM Espírito-Santo, MC Fandino, RG César, JS Hall, JL Hernández-Stefanoni, CC Jakovac, AB Junqueira, D Kennard, SG Letcher, M Lohbeck, M Martínez-Ramos, P

Massoca, JA Meave, R Mesquita, F Mora, R Muñoz, R Muscarella, YRF Nunes, S Ochoa-Gaona, E Orihuela-Belmonte, M Peña-Claros, EA Pérez-García, D Piotto, JS Powers, J Rodríguez-Velazquez, IE Romero-Pérez, J Ruíz, JG Saldarriaga, A Sanchez-Azofeifa, NB Schwartz, MK Steininger, NG Swenson, M Uriarte, MV Breugel, HVD Wal, MDM Veloso, H Vester, ICG Vieira, TV Bentos, GB Williamson, L Poorter. 2016. Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances* 2(5):e1501639. DOI: 10.1126/sciadv.1501639

Cherkassky V, F Mulier. 1998. Learning from data: Concepts, theory, and methods. New York, United States. Wiley. 560 p.

Cigizoglu HK, Ö Kisi. 2006. Methods to improve the neural network performance in suspended sediment estimation. *Journal of Hydrology* 317(3):221-238. DOI: 10.1016/j.jhydrol.2005.05.019

Cristianini N, J Shawe-Taylor. 2000. An introduction to support vector machines: and other kernel-based learning methods. New York, United States. Cambridge University Press. 204 p.

Cordeiro NG, KMG Pereira, MCNS Terra, JM Mello. 2018. Variação temporal do estoque de carbono e volume de madeira em um fragmento de cerrado *sensu stricto*. *Enciclopédia Biosfera* 15(28):931-942. DOI: 10.18677/EnciBio_2018B76

Dantas D, LOR Pinto, MCNS Terra, N Calegario, MLR Oliveira. 2020a. Reduction of sampling intensity in forest inventories to estimate the total height of eucalyptus trees. *Bosque* 41(3):353-364. DOI: 10.4067/S0717-92002020000300353

Dantas D, N Calegario, FW Acerbi Júnior, SPC Carvalho, MAI Júnior, EA Melo. 2020b. Multilevel nonlinear mixed-effects model and machine learning for predicting the volume of *Eucalyptus* spp. trees. *Cerne* 26(1):48-57. DOI: 10.1590/01047760202026012668

Dantas D, MCNS Terra, LOR Pinto, N Calegario, SM Maciel. 2021. Above and belowground carbon stock in a tropical forest in Brazil. *Acta Scientiarum. Agronomy* 43:e48276. DOI: 10.4025/actasciagron.v43i1.48276

Diamantopoulou MJ. 2005. Artificial neural networks as an alternative tool in pine bark volume estimation. *Computers and Electronics in Agriculture* 48(3):235-244. DOI: 10.1016/j.compag.2005.04.002

Figueiredo LTM, CPB Soares, AL Sousa, HG Leite, GF Silva. 2015. Dinâmica do Estoque de Carbono em Fuste de Árvores de uma Floresta Estacional Semidecidual. *Cerne* 21(1):161–167. DOI: 10.1590/01047760201521011529

Haykin, S. 2001. Redes neurais: princípios e prática. 2.ed. Porto Alegre, Brasil. Bookman. 900 p.

Heddam S, O Kisi. 2018. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology* 559:499-509. DOI: 10.1016/j.jhydrol.2018.02.061

Leite HG, VCL Andrade. 2002. Um método para condução de inventários florestais sem o uso de equações volumétricas. *Revista Árvore* 26(3):321-328. DOI: 10.1590/S0100-67622002000300007

Marques RFPV, MCNS Terra, VA Mantovani, AF Rodrigues, GA Pereira, RA Silva, CR Mello. 2019. Rainfall Water Quality Under Different Forest Stands. *Cerne* 25(1):8–17. DOI: 10.1590/01047760201925012581

Mcnicol IM, CM Ryan, KG Dexter, SMJ Ball, M Williams. 2018. Aboveground carbon storage and its links to stand structure, tree diversity and floristic composition in south-eastern Tanzania. *Ecosystems* 21(4):740-754. DOI: 10.1007/s10021-017-0180-6

Mehtätalo L, M Maltamo, A Kangas. 2006. The use of quantile trees in the prediction of the diameter distribution of a stand. *Silva Fennica* 40(3):id333. DOI: 10.14214/sf.333

Melo EA, N Calegario, AR Medonça, EL Possato, JA Alves, MAI Júnior. 2017. Modelagem não Linear da Relação Hipsométrica e do Crescimento das Árvores Dominantes e Codominantes de *Eucalyptus* sp. *Ciência Florestal* 27(4):1325–1338. DOI: 10.5902/1980509829895

Meyer D. 2019. e1071: Misc Functions of the Department of Statistics, Probability Theory Gr oup (Formerly: E1071), TU Wien. R package version 1.7-1. Consultado 24 aug. 2019 . Disponible en https://CRAN.R-project.org/package=e1071

Özçelik R, MJ Diamantopolou, F Crecente-Campo, U Eler. 2013. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management* 306:52–60. DOI: 10.1016/j.foreco.2013.06.009

R Development Core Team. 2018. R: a language and environment for statistical computing. V iennaR Foundation for Statistical Computing, 2018. Consultado 24 aug. 2019. Dispon ible en https://www.r-project.org/

Réjou-Méchain M, A Tanguy, C Piponiot, J Chave, B Hérault. 2017. biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods in Ecology and Evolution* 8(9):1163–1167. DOI: doi.org/10.1111/2041-210X.12753

Ribeiro SC, LAG Jacovine, CPB Soares, SV Martins, AL Souza, AMB Nardelli. 2009. Quantificação de biomassa e estimativa de estoque de carbono em uma floresta madura no município de viçosa, Minas Gerais. *Revista Árvore* 33(5):917-926, 2009. DOI: 10.1590/S0100-67622009000500014.

Scolforo HF, JRS Scolforo, CR Mello, JM Mello, ACF Filho. 2015. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-arid woodland. *PLoS ONE* 10(6):1-20. DOI: 10.1371/journal.pone.0128781

Siipilehto J. 2000. A comparison of two parameter prediction methods for stand structure in Finland. *Silva Fennica* 34:331-349. DOI: http://dx.doi.org/10.14214/sf.617

Soares FAAMN, EL Flores, CD Cabacinha, GA Carrijo, ACP Veiga. 2011. Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. *Computers and Electronics in Agriculture* 78(1):19-27. DOI: 10.1016/j.compag.2011.05.008

Soriano-Luna MLA, et al. 2018. Determinants of above-ground biomass and its spatial variability in a temperate forest managed for timber production. *Forests* 9(8):1–20. DOI: 10.3390/f9080490

Steinwart I, A Christmann. 2008. Support vector machines. Berlin, Germany. Springer. 601 p.

Terra MCNS, D Dantas, LOR Pinto, N Calegario, SM Maciel. 2019. Estimativa do estoque de carbono em floresta semidecidual: uma comparação entre regressão e redes neurais ar tificiais. In Oliveira AC ed. Fontes de Biomassa e Potenciais de Uso. Ponta Grossa, B rasil. Editora Atena. p. 24-35.

Thomas SC, AR Martin. 2012. Carbon content of tree tissues: A synthesis. *Forests* 3(2):332-352. DOI: 10.3390/f3020332

Tong S, D Koller. 2001. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research* 2:45-66. DOI: 10.1162/153244302760185243

Valença M. 2010. Fundamentos das redes neurais: exemplos em Java, ver. ampli. 2 ed. Olinda, Brasil. Livro Rápido. 386 p.

Vendruscolo DGS, R Drescher, HS Souza, JPVM Moura, FMD Mamoré, TAS Siqueira. 2015. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. *Revista Brasileira de Biometria* 33(4):556–569, 2015. DOI: 10.13140/RG.2.1.1742.5684

ARTICLE 3 - PREDICTING STEM VOLUME OF EUCALYPTUS TREE SPECIES WITH MULTI-LEVEL MIXED-EFFECTS MODEL AND MACHINE LEARNING METHODS

ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA REVISTA CERNE

Situação: Publicado na Revista Cerne

PREDICTING STEM VOLUME OF EUCALYPTUS TREE SPECIES WITH MULTI-LEVEL MIXED-EFFECTS MODEL AND MACHINE LEARNING METHODS

Multilevel random-effects in the Schumacher and hall model improved volume predictions.

Huynh-Feldt structure in the variance and covariance matrix provided better model fit.

Machine learning techniques are suitable in modeling tree volume.

It was extracted accurate volume equation from neural network training process.

ABSTRACT: Volumetric equations is one of the main tools for quantifying forest stand production, and is the basis for sustainable management of forest plantations. This study aimed to assess the quality of the volumetric estimation of *Eucalyptus* spp. trees using a mixed-effects model, artificial neural network (ANN) and support-vector machine (SVM). The database was derived from a forest stand located in the municipalities of Bom Jardim de Minas, Lima Duarte and Arantina in Minas Gerais state, Brazil. The volume of 818 trees was accurately estimated using Smalian's Formula. The Schumacher and Hall model was fitted by fixed-effects regression and by including multilevel random effects. The mixed model was fitted by adopting 14 different structures for the variance and covariance matrix. The best structure was selected based on the Akaike Information Criterion, Maximum Likelihood Ratio Test and Vuong's Closeness Test. The SVM and ANN training process considered diameter at breast height and total tree height to be the independent variables. The techniques performed satisfactorily in modeling, with homogeneous distributions and low dispersion of residuals. The quality analysis criteria indicated the superior performance of the mixed model with a Huynh-Feldt structure of the variance and covariance matrix, which showed a decrease in mean relative error from 13.52% to 2.80%, whereas machine learning techniques had error values of 6.77% (SVM) and 5.81% (ANN). This study confirms that although fixed-effects models are widely used in the Brazilian forest sector, there are more effective methods for modeling dendrometric variables.

Keywords: Artificial intelligence, Artificial neural network, Forest Management, Schumacher and Hall Model, Support-vector machine.

INTRODUCTION

One of the most important parameters for assessing the productivity potential of a specific forest stand or region is log volume production. Models describing log volume productivity potential of a forest stand or region are invaluable tools for forest management planning at any level for both forest managers and researches, underpinning decisions related to silvicultural treatments, logging and timber transportation. Therefore, accurate log volume estimates are essential.

The volumetric estimation of *Eucalyptus* spp. clones is usually based on equations in which diameter at breast height (DBH) and total tree height (H) are the independent variables. According to Campos and Leite (2009), the Schumacher and Hall model stands out in tree volume estimation. However, classical regression models assume independence between observations and homogeneity of variance, which, in some cases, may not be true. Particularly, when dealing with correlated data or repeated measurements (Lappi, 1991; Gregoire et al., 1995; Littell et al., 2006; Uzoh and Oliver, 2008; Huang et al., 2009; Uzoh, 2017)

An alternative for analyzing correlated data, in space and/or in time, and for explicitly modeling their covariance structure is to use mixed models. Some of the possible approaches with mixed models include generalizing correlation and variance structures (Fu et al., 2017; Cropper and Putz, 2017; Özçelik et al., 2018; Wang et al., 2019). Mixed models are sophisticated regression techniques, and the study by Lappi (1991) pioneered their use in forestry research.

Mixed models are fitted by the inclusion of random effects variables that reflects the error structure imposed by the sampling scheme thus accounts for the nonzero correlation (see, e.g., Littell et al., 2006; Uzoh and Oliver, 2008; Huang et al., 2009; Uzoh, 2017), and fixed effects variables (Pinheiro and Bates, 2000). Additionally, mixed models can describe incomplete blocks, split plots, spatiotemporal data and random coefficients, as well as polynomial and growth curves (see, e.g., Lappi, 1991; Gregoire et al., 1995; Littell et al., 2006; Uzoh, 2017). Nevertheless, MNLM approach has one drawback. However not always, but sometimes, when generated coefficients are used to make predictions outside the range of the dataset, MNL may outperform MNLM (see, Monleon, 2004).

Usually, in forest sciences, mixed models are applied to nonlinear problems, such as height growth, basal area increment of *Eucalyptus* spp. stands and genetic evaluations (Calegario et al., 2005; Barantal et al., 2019; Wang et al., 2019; Sharma et al., 2019).

In addition, computational approaches of artificial intelligence/machine learning, including artificial neural networks (ANN) and support-vector machines (SVM), have been increasingly used as tools for forest data analysis, modeling, variable estimation and production prognosis (Silveira et al., 2019). Those tools have provided gains in the quality of estimates and predictions (Vendruscolo et al., 2015; Martins et al., 2016).

An ANN is an algorithm based on simple processing units (artificial neurons), mimicking the neurons found in the human brain, which calculate specific functions. Those units are layered and connected to each other through weights, which store experimental knowledge and weigh the inputs of each unit. Therefore, the acquired knowledge becomes available for use (Haykin, 2001; Silva et al., 2018).

The most striking features of ANNs are their learning and generalization capacities. In other words, ANNs are able to, through a learned example, generalize knowledge assimilated for an unknown dataset. Another interesting feature of ANNs is their ability to extract nonexplicit characteristics from a dataset provided as examples (Gorgens et al., 2009).

Support-vector machines (SVMs) have also become an interesting alternative for the mathematical modeling of complex systems. They are simple techniques, in terms of their conceptual framework, capable of solving extremely complex and real problems. In an SVM, input space vectors are mapped nonlinearly to a characteristic space of high dimensionality where a linear decision surface is constructed, constituting an optimal separation hyperplane, for example, the binary separation between data that has positive and negative labels, such that the separation margin is maximum (Shao et al., 2014).

Initially, SVM techniques were successfully applied as data classification methods (Cherkassky and Mulier, 1998; Yu et al., 2019). Subsequently, they were expanded to regression tasks using the following approaches: support vector regression (SVR) and least-squares support vector machines (LS-SVMs) (Karthik et al., 2016; Zeng et al., 2018; Souza et al., 2019; Sivasankar et al., 2019).

Considering the relevance and sophistication of these techniques, this study aimed to assess the performance of the nonlinear volumetric model by Schumacher and Hall fitted by fixed and mixed regression with variance and covariance matrix modeling, by training an artificial neural network and by constructing support-vector machine to estimate the log volume of *Eucalyptus* spp. trees.

MATERIALS AND METHODS

DATA

The study area comprises 21 management units planted with *Eucalyptus* spp. hybrid clones, located in the municipalities of Bom Jardim de Minas, Lima Duarte and Arantina, in Minas Gerais, Brazil, totaling 1,090 ha of inventoried area.

The climate of the region is characterized as subtropical highland climate, Cwb type, according to the Köppen climate classification, with an average annual temperature of 20.1°C, with dry and cold winters, with frost occurring in some areas, and with rainy summers with moderately high temperatures. The average annual total precipitation is 1,456 mm (Alvares et al., 2013).

The data used in this study were collected by accurately estimating the cubic volume of 818 trees of different ages and sizes. The study parameters were total height (*Ht*), measured in meters; diameter at breast height (*DBH*), measured in centimeters; and diameters at the base (at 0.1 m of height) and at heights of 0.5 m, 1 m, 1.5 m and 2 m and, every 2 m from then on. Log volumes were calculated using Smalian's Formula. Descriptive statistics of the data are reported in Table 1.

Table 1. Descriptive statistics of the data on the log volume of *Eucalyptus* spp. trees

Variable	Minimum	Mean	Maximum	Standard deviation	CV%
Volume (m³)	0.0066	0.1785	0.7752	0.1335	74.79
DBH (cm)	4.19	13.60	27.00	4.45	32.72
Height (m)	8.10	21.82	33.08	5.10	23.37
Age (years)	3	7.56	8.5	1.23	16.27
N (trees/ha)	840.34	1064.96	1215.08	76.18	7.15

where DBH is the diameter at 1.30 from the ground.

Scatter plots of log volume as a function of *DBH* and *Ht* are shown in Figure 1, indicating a nonlinear relationship between these variables.

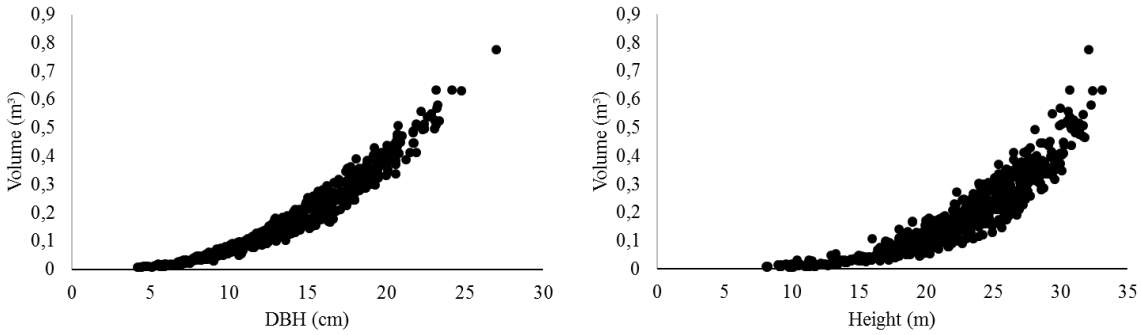


Figure 1. Scatter plots of *Eucalyptus* spp. log volume as a function of *DBH*, *Ht*.

VOLUMETRIC MODEL

The nonlinear volumetric model by Schumacher and Hall (1933) (1) was fitted to the volume data. The processing was performed using the software environment for statistical computing R, version 3.4.1 (2017).

$$V_i = \beta_0 * DBH_i^{\beta_1} * Ht_i^{\beta_2} + \varepsilon_i \quad (1)$$

where *V* is the volume in m³; *DBH* is the diameter, in cm, at 1.30 m of height; *Ht* is the total height in m; β_0, β_1 and β_2 are the parameters of the model; and ε is the random error.

MULTILEVEL MIXED-EFFECTS MODEL FOR VOLUMETRIC ESTIMATION

Subsequently, to evaluate the use of mixed-effects models to estimate the log volume of *Eucalyptus* spp., the Schumacher and Hall model was refitted by incorporating the variability of each tree and management unit, thereby generating a multilevel nonlinear mixed-effects model with fixed and random parameters. The model was fitted using the maximum likelihood method, proposed by Fisher, according to Searle (1987), which consists of obtaining estimators that maximize the probability density function of observations with respect to fixed effects and variance components.

In nonlinear mixed models (2) the response variable y_{ij} represents the random groups *i* and *j*, where *i* are the i-th management units and *j* the j-th trees. Then, $i = 1,...,$ m, and $j = 1,...,$ n_i, where m is the total number of management units and n_i is the number of trees within the i-th management unit; *f* is a general, real and differentiable function of a specific group of parameter vectors ϕ_{ij} and a covariant vector v_{ij}; and ε_{ij} is the random error normally distributed within groups (Pinheiro and Bates, 2000).

$$y_{ij} = f\left(\phi_{ij}, V_{ij}\right) + \varepsilon_{ij}, i = 1, \dots, m;\ j = 1, \dots, n_i. \qquad (2)$$

The parameter vector varies from individual to individual. In a second stage, the vector ϕ_{ij} can be expressed by equation 3.

$$\phi_{ij} = A_{ij}\beta_i + B_{ij}\beta_i + B_{ij}\beta_{ij}, \beta_i \approx N(0, \Psi_1) \text{ and } \beta_{ij} \approx N(0, \Psi_2) \qquad (3)$$

where β is a fixed-effects vector *(p x 1)*; B_i is a vector *(q_1 x 1)* of random effects independently distributed with a covariance-variance matrix ψ_1; B_{ij} is a vector *(q_2 x 1)* of random effects independently distributed with a covariance-variance matrix ψ_1 and presumed independent of first-level random effects; A_{ij} and B_{ij} are incidence matrices; and ε_{ij}, within groups, are independently distributed and independent of random effects.

In modeling mixed models, a key step is the definition of the variance and covariance structure because it aims to obtain a parsimonious structure that explains the data variability

and the correlation between the measurements and a small number of parameters (Clark and Linzer, 2012). This choice may directly affect parameter estimates, standard errors of fixed and random effects, diagnoses and inferences. This selection depends on data structures, empirical information and computational availability.

Variance and covariance structures were entered in the nonlinear mixed model by Schumacher and Hall because trees belonging to the same management unit are likely more correlated with each other than otherwise. This processing was performed using the package *nlme*, software environment R, and its function *correlation* (Pinheiro and Bates, 2000). In total, 14 structures were used, including Variance Components (VC), UNstructured (UN), Compound Symmetry (CS), First-Order Autoregressive (AR(1)), Heterogeneous First-Order Autoregressive (ARH(1)), Heterogeneous Compound Symmetry (HCS), Toeplitz (TOEP), First-Order AutoRegressive Moving Average (ARMA(1,1)), Heterogeneous Toeplitz (TOEPH), First-Order Ante-dependence (ANTE(1)), UNstructured Correlations (UNR), Spatial Power (SP(POW)(c-list)), Banded Main Diagonal (UN(1)) and Huynh-Feldt (H-F). More details on these structures are available from Pinheiro and Bates (2000)

The Akaike Information Criterion (AIC) (Sakamoto et al., 1986) (4), for which the best model is that which has the lowest AIC value; the Maximum Likelihood Ratio Test (MLRT) (Pinheiro and Bates, 2000); and Vuong's Closeness Test (1989) were used to choose the structure of the variance and covariance matrix.

$$AIC = -2 * \ln(ml) + 2p \qquad (4)$$

where *AIC* is the Akaike Information Criterion, ln is the Napierian logarithm, *ml* = maximum likelihood value, and *p* is the number of parameters of the model.

The maximum likelihood ratio test (MLRT) (5) consists of comparing models pairwise, calculating its value as the difference between the values of its likelihood functions (Pinheiro and Bates, 2000).

$$MLRT = 2 * [ln(ml_2) - log(ml_1)] \qquad (5)$$

where *ln* is the Napierian logarithm, ml_2 is the value of the maximum likelihood function of model 2, and ml_1 is the value of the maximum likelihood function of model 1.

To compare the models, through the likelihood ratio test - Vuong's T_{RLNN} (1989) (6), the statistical equation stated below was used:

$$T_{RLNN} = \frac{1}{\sqrt{n}} \frac{\log\left(\widehat{ml}, \hat{\beta}\right)}{\widehat{\omega}^2} \tag{6}$$

where

$$\widehat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(log\frac{f\left(y_i|x_i,\widehat{ml}\right)}{g\left(y_i|x_i,\hat{\beta}\right)}\right)^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(log\frac{f\left(y_i|x_i,\widehat{ml}\right)}{g\left(y_i|x_i,\hat{\beta}\right)}\right)\right]^2 \tag{7}$$

is an estimator of the variance of $\frac{1}{\sqrt{n}}\log\left(\widehat{mv}, \hat{\beta}\right)$ and $\log\left(\widehat{mv}, \hat{\beta}\right)$ is the likelihood ratio test. The statistic has, asymptotically,

$$T_{RLNN} \xrightarrow{d} N(0,1) \tag{8}$$

under

$$H_0: log\frac{f\left(y_i|x_i,\widehat{mv}\right)}{g\left(y_i|x_i,\hat{\beta}\right)} = 0 \tag{9}$$

That is, the models are equivalent to the significance level α, and $Z_{\alpha/2}$ is the critical value of the standard normal distribution, rejecting the null hypothesis if $|T_{RLNN}| > Z_{1-\alpha/2}$.

MACHINE LEARNING

SVMs and ANNs were used as machine learning approaches. SVM construction was based on the supervised machine learning process described in detail by Haykin (2001), with a set of n samples represented as an ordered pair (**X**, **Y**), where **X** is a matrix of explanatory variables of the sample and **Y** is the vector of expected values of the sample. Based on this information, a function that predicts the expected value of the sample, using a vector of

characteristics as input dataset, is chosen. This linear function is represented by f(**X**) = <**W**,**X**> + b, where **W** is a vector of weights.

The type IV error function, also known as *eps-regression*, was used, and the Kernel function was a radial basis function (RBF). Kernel functions provide an alternative solution by projecting data into a space with high-dimensional features to increase the computational power of learning machines, making it possible to represent nonlinear phenomena (Granata et al., 2016). This procedure was performed in the software environment R, version 3.4.1, using the package e1071 (Meyer et al., 2019).

The trained ANNs were Multilayer Perceptron (MLP) networks, consisting of an input layer, an intermediate layer, and an output layer. The algorithm used was *resilient backpropagation*, where the learning rate was set automatically by the package *neuralnet*, with values ranging from 0.01 to 1.12. The number of neurons in the intermediate layer was chosen using the *k-fold*. This methodology randomly subdivides the database into *k* subgroups (Wong et al., 2017). The *k* value was 10 subgroups, with 90% for training and 10% for testing (Diamantopolou, 2010), applying cross validation. Different numbers of neurons, ranging from 1 to 20, were tested.

The activation function used was logistic (or sigmoid), with an interval from 0 to 1, which limits the amplitude of outputs and inputs. Therefore, the data were normalized, which consisted of transforming the values of each variable into values ranging from 0 to 1, using equation (10) (Soares et al., 2011). This equation considers the minimum and maximum value of each variable in the value transformation, maintaining the original data distribution (Valença, 2010).

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \qquad (10)$$

where *x'*: normalized value; *x*: original value; x_{min}: minimum value of the variable; x_{max}: maximum value of the variable; *a*: lower limit of the normalization range; and *b*: upper limit of the normalization range.

The stopping criterion of the ANN training process was a maximum number of 100,000 cycles, or a mean squared error less than 1%, stopping the training when meeting one of the

criteria. At the end of the training, the best ANN was selected, based on the smallest mean squared error.

The data were divided into two groups: model building datasets, using 70% to fit the nonlinear volumetric models, in their fixed and mixed forms, to construct the SVM and to train the ANN; and model validation datasets, using 30% to generalize of the techniques. Among the data intended for ANN training, 70% were used in the training phase and 30% in the testing phase.

ESTIMATE QUALITY ASSESSMENT

The nonlinear volumetric models by Schumacher and Hall were assessed, in their fixed and mixed forms, by adopting a structure of the variance and covariance matrix, artificial neural network and support-vector machine based on Mean Relative Error (MRE) (Equation 11), Bias (Equation 12), residual standard error (Syx) (Equation 13), Root Mean Square Error (RMSE) (Equation 14), correlation between estimated and observed values (r) (Equation 15) and residual scatter plots.

$$MRE\ (\%) = \frac{\sum \frac{\left|\hat{Y} - Y\right|}{Y} \times 100}{n} \tag{11}$$

$$\text{Bias (\%)} = 100 * \frac{1}{m} \sum_{j=1}^{m} \left[\frac{1}{n_i} \sum_{i=1}^{n_i} \left[(Y_{ij} - \hat{Y}_{ij}) / Y_{ij} \right] \right] \tag{12}$$

$$Syx\ (\%) = \frac{\sqrt{\frac{RSS}{n-p-1}}}{\bar{Y}} * 100 \tag{13}$$

$$\text{RMSE (\%)} = \left[\frac{1}{m} \sum_{j=1}^{m} \sqrt{\frac{1}{n_i} \sum_{i=1}^{n_i} \left[(Y_{ij} - \hat{Y}_{ij})^2 \right] / Y_{ij}} \right] * 100 \tag{14}$$

$$r = \sqrt{1 - \frac{\sum\left(\widehat{Y}\imath - \mathrm{Y}_i\right)^2}{\sum\left(\mathrm{Y}_i - \overline{\mathrm{Y}}_i\right)^2}} \quad (15)$$

where Y_{ij} and $\hat{Y}_{ij}$ are the actual and predicted log volume of the i^{th} tree in the j^{th} management unit, n_i is the number of trees in the j^{th} management unit, m is the number of management unit, p is the number of parameters and RSS is the residual sum of squares.

RESULTS AND DISCUSSION

MULTILEVEL MIXED-EFFECTS MODEL

Table 2 reports the fitted parameters of the nonlinear model by Schumacher and Hall, showing that all coefficients were significant at the 0.05 level, according to Student's t-test.

Table 2. Fitted parameters for the nonlinear model by Schumacher and Hall

	Parameters	**Standard error**	**t value**	**Pr (>t)**	
β0	$4.15e^{-5}$	$2.81e^{-6}$	14.77	<0.0000	*
β1	1.7111	0.0246	69.59	<0.0000	*
β2	1.2040	0.0358	33.64	<0.0000	*

where β_i: regression coefficients; *: significant at 0.05.

After the initial fit of the Schumacher and Hall nonlinear model, the mixed models were fitted by including multilevel random effects and considering 14 structures of the variance and covariance matrix.

Table 3 presents the selection criteria for structures of the variance and covariance matrix used in this study. Among them, the structure that best fit the volumetric estimation of *Eucalyptus* spp. was Huynh-Feldt (H-F), which had the lowest AIC value and the highest likelihood logarithm (LogLik) value. In addition, by performing the likelihood ratio test and by assuming such a structure as an alternative hypothesis, Huynh-Feldt was compared with the other structures.

Table 3. Selection criteria for structures of the variance and covariance matrix in fitting the multilevel nonlinear mixed-effects volumetric model

Model	AIC	LogLik	Test	MLRT	p-value
1. Toeplitz	4201.92	2802.12	1 vs 7	555.58	<0.000
2. Banded Main Diagonal	4149.67	2804.17	2 vs 7	502.92	<0.000
3. Compound Symmetry	4155.61	2802.47	3 vs 7	518.33	<0.000
4.UNstructured Correlations	4139.32	2804.45	4 vs 7	773.58	<0.000
5. Unstructured	3951.12	3330.80	5 vs 7	2.13	<0.632
6. Heterogeneous Compound Symmetry	3949.75	3330.45	6 vs 7	2.00	<0.563
7. Huynh-Feldt	3940.31	3350.68	-	-	-

Table 3 indicates that almost all null hypotheses were rejected (*p-value* < 0.05), except for the Heterogeneous Compound Symmetry (HCS) (*p-value* = 0.563) and UNstructured (UN) (*p-value* = 0.632) structures, which exhibited no significant difference from HF. The results indicate that using any of these three structures of the variance and covariance matrix would be adequate for the dataset. According to West et al. (2015), the HF structure is characterized by unequal variances between management units and covariances determined by calculating the arithmetic mean of variances and by subtracting λ; λ is the difference between mean variance and mean covariance. In the HCS structure, different variances and some unequal covariances, fitted by a correlation coefficient ρ between individuals, are applied. In the Unstructured structure, different variances and covariances are also assigned for each occasion.

The HF structure was chosen considering the AIC and the LogLike.

According to West et al. (2015), variance and covariance matrix structures increase the flexibility of correlations. In any data analysis, the correct structure that is most appropriate and parsimonious for these matrices should be chosen, based on observed data and on the relationships between observations of each sample unit, in this case, the management unit, because different variance and covariance numbers should be estimated.

Table 4 outlines the results from the fit (fixed parameters) of the Schumacher and Hall nonlinear mixed model using the maximum likelihood method, adopting the HF structure of the variance and covariance matrix, which shows that all parameters were significant at the 0.001 level, according to maximum likelihood's test.

Table 4. Significance of fixed parameters fitted for the Schumacher and Hall nonlinear mixed-effects model

	Parameters	Standard error	t value	P-value	
β0	$3.4645e^{-5}$	$1.59e^{-6}$	21.81	<0.0000	*
β1	1.8184	0.0200	90.85	<0.0000	*
β2	1.1655	0.0282	41.38	<0.0000	*

where β_i: regression coefficients; *: significant at 0.001.

MACHINE LEARNING

The configurations of the SVM construction are reported in Table 5.

Table 5. SVM parameters

SVM	Parameter
Type	*eps-regression*
Kernel	Radial basis function
Cost	0.002
Gamma	0.2441
Epsilon	0.1535
Number of support vectors	420

The optimization of the parameters of an SVM model is fundamental for the development of the final model with high prediction performance. Modifying the gamma and epsilon parameters in the radial basis function enhances model performance. Epsilon regulates the function, minimizing the residues, while gamma is the parameter responsible for determining the base length of the radial basis function, reducing or increasing the complexity of the search process (Cherkassky and Mulier, 1998).

The combination of the two step grid search approach and cross validation was used for the global optimization of these parameters in this work. For each combination of modeling parameters, a mean square error (MSE) was calculated and the optimal parameters that produced the smallest MSE were selected. The optimal values of epsilon and gamma were 0.2441 and 0.1535, respectively.

Regarding the approach by artificial neural network, the ANN that presented the smallest error, among the others evaluated, was obtained with 23041 iterations. The architecture and weights of the selected ANN with lowest error among all networks evaluated, consisting of seven neurons in the hidden layer, is shown in Figure 2.

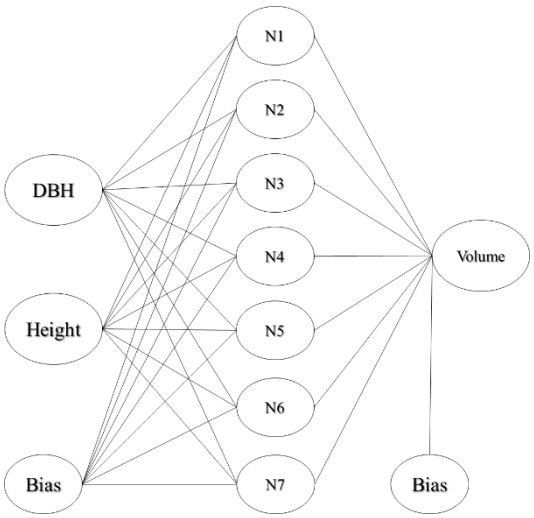


Figure 2. Architecture of the selected ANN, with seven neurons in the hidden layer.

From the artificial neural network with architecture 2-7-1 an equation system was extracted to predict the individual volume of *Eucalyptus* spp. trees, with coefficients resulting from the weights generated by the ANN. This system was used to predict the volume of the trees that made up the database intended for validation.

Model (16) expresses the relationship between the hidden layer and the response variable, where β0 is the bias, and the other coefficients are the weights related to each neuron. Model (17) represents the activation function used in each neuron of the hidden layer, derived from the logistic model. Finally, the model (18) is the result of the relationship between the input variables and the respective hidden layer neurons, being generated a model for each neuron.

$$Volume' = \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \beta_3 * z_3 + \beta_4 * z_4 + \beta_5 * z_5 + \beta_6 * z_6 + \beta_7 * z_7 \quad (16)$$

$$z_n = \left[\frac{1}{1 + e^{-w_i}}\right] \quad (17)$$

$$w_i = \beta_{0.n} + \beta_{1.n} * DBH_i' + \beta_{2.n} * H_i' \quad (18)$$

where β_0: bias; β_n: coefficient of the model associated with neuron *n*; $\beta_{k.n}$: coefficient of the model between input variable *k* and neuron *n*; z_n: response of the n-th neuron of the hidden layer; w_i: sum of the products between the weights and the inputs.

Table 6 presents the coefficients of the system of equations extracted from the artificial neural network.

Table 6. Parameters (β's) of the artificial neural network. N represents the neuron.

	β0	**β1**	**β2**	**β3**	**β4**	**β5**	**β6**	**β7**
RNA	0.8609	-0.7323	0.7619	0.6785	-1.7671	-1.6967	1.0353	-0.7712
N1	-2.3277	2.0301	-	-	-	-	-	-
N2	-1.3559	0.6158	-	-	-	-	-	-
N3	-1.7104	2.3549	-	-	-	-	-	-
N4	-0.1308	0.1538	-	-	-	-	-	-
N5	0.1174	-1.2125	-	-	-	-	-	-
N6	-0.0674	-2.5130	-	-	-	-	-	-
N7	2.5005	2.8092	-	-	-	-	-	-

ESTIMATE QUALITY ASSESSMENT

We applied the techniques analyzed to the dataset for validation and performed satisfactorily in modeling *Eucalyptus* spp. trees volume, with homogeneous distributions and low residual dispersion, as shown in Figure 3.

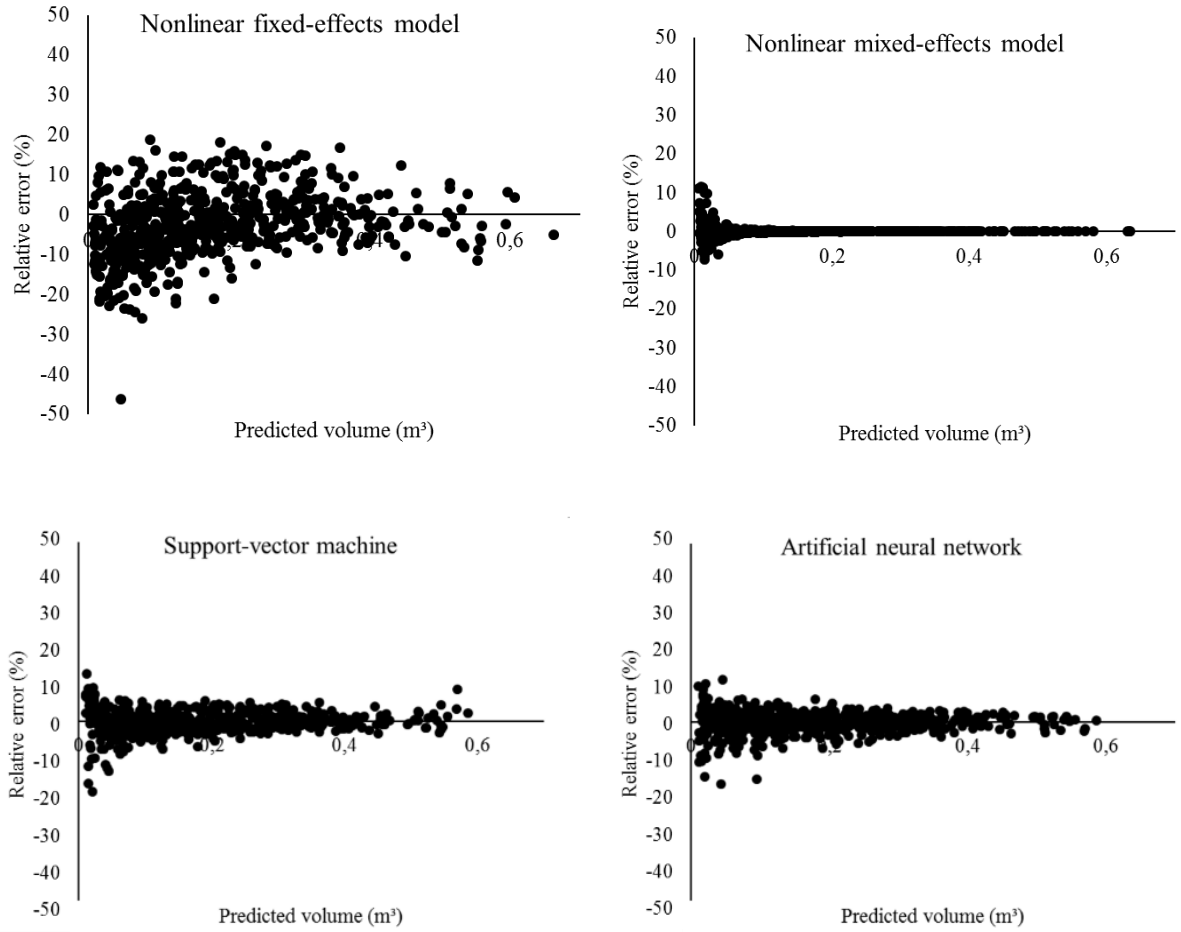


Figure 3. Scatter plots of residuals distribution.

The graphical analysis highlighted the efficiency of the nonlinear mixed-effects model with an HF structure of the variance and covariance matrix and the machine learning techniques, still little spread in the Brazilian forest sector, which indicates the potential of using these techniques, due to the high gain in prediction accuracy.

The inclusion of variability between and within each individual and between management units in the nonlinear mixed-effects regression model provided better performance than the nonlinear fixed-effects regression model. It is verified that the nonlinear mixed-effects regression model was considerably more accurate than the nonlinear fixed-effects regression model, with well-distributed residuals and with a mean around zero. The lowest values of log volume tended to exhibit a higher error, albeit within the range of -10 to 15%, this may possibly be overcome by fitting to a more complete data set, properly calibrating the model. Conversely, the nonlinear fixed-effects model, although it presented all significant parameters at a level of 0.05, by the t-Student test, it exhibited high dispersion and greater heterogeneity of residual

variance, with a tendency to overestimate the lowest values of log volume and to underestimate intermediate values. Residues of the nonlinear fixed-effects model were generally between -30 and 20%, with a value of -46%.

Thus, nonlinear mixed-effects models are important techniques for growth and production modeling. These models explain the different degrees of hierarchy within a dataset and can provide individual predictions specific to each hierarchy (Temesgen et al. 2008; Ou et al., 2016). These models also provide information from various sources of heterogeneity and correlations that are present in the data (Hall and Clutter, 2004), making them an efficient option for those interested in forest volume and biomass estimates.

Comparing machine-learning techniques, according to the graphs, both showed residuals within the range of -20 and 10%. However, the plots show that the ANN was slightly better than the SVM, which presented more concentrated residues around zero, while SVM showed a higher tendency of error in the smallest and largest volume values.

The performance evaluation criteria for the techniques analyzed in this study are reported in Table 7. The qualities of the predictions performed by the techniques in the data intended for validation were evaluated.

Table 7. Performance evaluation criteria of the predictions of nonlinear fixed-effects (MNL) and mixed-effects (MNLM) models, support-vector machine (SVM), and artificial neural network (ANN)

	MNL	**MNLM**	**SVM**	**ANN**
Mean relative error (%)	13.5185	2.8035	6.7689	5.8125
Bias (%)	-6.2610	-1.2371	2.6479	3.6217
Residual standard error (m^3)	0.0142	0.0003	0.0024	0.0012
Residual standard error (%)	7.9830	0.1516	1.2048	0.606
Root-mean-square error (%)	18.5858	4.4185	8.5550	7.8191
Correlation coefficient	0.9700	0.9851	0.9831	0.9838

The predictions of the techniques analyzed in the validation phase were strongly correlated with the observed values, with a value of 0.97 for the nonlinear fixed-effects model and values above 0.9831 for the other techniques. The error values, represented by the root-mean-square error (RMSE), were lower than 9% for the nonlinear mixed-effects model, ANN

and SVM, while for the fixed-effects model the REQM was 18.5858%. The lower the RMSE is, the higher the accuracy of the estimates will be, and the optimal situation is when RMSE is zero. Bias indicated slight overestimation trends for nonlinear models and underestimate for machine learning techniques. The nonlinear mixed-effects model presented the lowest bias value, -1.2371%, indicating that it is a balanced (non-biased) and effective tool. The mean relative error was greater than 13% for the nonlinear fixed-effects model and less than 7% for the other techniques, with the lowest value presented by the nonlinear mixed-effect model, 2.8035%.

One advantage of random-effects models over fixed-effects models is the reduced residual standard error. Calegario et al. (2005) studied the basal area of *Eucalyptus* spp. clones and observed a decreased of approximately 15 times. In this study, a marked decrease in residual standard error was also observed, from 0.0142 in the nonlinear fixed-effects regression model to 0.0003 in the nonlinear mixed-effects regression model, that is, a 53-fold decrease. Carvalho et al. (2011), by applying mixed-effects modeling in basal area and volume prediction, found a decrease in error from 15% to 12% in basal area prediction and from 26% to 4% in volume prediction.

The ANN was able to explain almost all variation in the *Eucalyptus* spp. trees volume with the available variables. Several studies have shown satisfactory performances of artificial neural networks (Özçelik et al., 2013; Vendruscolo et al., 2015; Martins et al., 2016). This superiority results from the ability of neural networks to detect implicit information and nonlinear relationships between the response variable and the explanatory variables provided as examples and to generalize the assimilated knowledge to an unknown dataset.

It should be noted that despite performing worse than the ANN, the SVM was highly efficient at estimating *Eucalyptus* spp. trees volume. The SVM has the advantage over the ANN of requiring no evaluation after its construction, which is performed in the ANN to select the best network, thanks to the quadratic optimization that occurs during the SVM training (Yang et al., 2015). This optimization provides the same result for each system configuration, whenever applied to the same dataset. Conversely, the ANN has more elements for manipulation, in addition to the random initialization of neural parameters (Haykin, 2001). Thus, each trained network will exhibit small differences in estimates, even when maintaining the same architecture. These differences highlight the practicality of SVMs over ANNs by preventing operator subjectivity in having to choose the best network to apply to the database.

Both machine learning and nonlinear mixed-effects model approaches were efficient in modeling the log volume of *Eucalyptus* spp. trees. The small variation in *Eucalyptus* spp. trees volume unexplained by the study variables results from the various factors disregarded in the present study that are known to affect tree volume variability in forests, such as biotic and abiotic factors and their interactions (Tanaka et al., 2017).

The nonlinear Schumacher and Hall multilevel mixed-effects model (MNLM) with an HF structure of the variance and covariance matrix resulted in a better fit in modeling the log volume of Eucalyptus spp. clones than traditional nonlinear regression analysis approach (MNL), artificial neural network (ANN), and support-vector machine (SVM). This is because MNLM allow parameter vectors to vary from management unit to management unit. Also, allow variance components to be simultaneously broken down into a) the fixed-effect prediction common to the population, b) the random-effect prediction that provides the management unit-specific relationship that captures special characteristics of each management unit (Uzoh and Oliver, 2008; Meng and Huang, 2010; Crecente-Campo et al., 2010; Uzoh, 2017).

Studies have indicated the importance and the gain in precision generated by the inclusion of random effects in the modeling of forest structures. Ou et al. (2016) concluded that the addition of topographic variables (elevation, slope, and appearance) as a random-effect for a nonlinear mixed-model using height and DBH as predictors improved the AIC and BIC values. Huff et al. (2018) analyzed the performance of nonlinear mixed-effect models to predict total above ground biomass and the results showed that the inclusion of shrub species as a random-effect provided better performance than nonlinear fixed-effects models.

Many forest management decisions are based on knowledge of wood volume availability in a stand and growth and productivity projections, and the use of nonlinear mixed-effects models can be employed for accurate future predictions involving repeated measurements over time. This study confirms that although fixed-effects models are widely used in the Brazilian forest sector, there are more effective methods for modeling dendrometric variables.

CONCLUSION

The present study considerably improves the modeling of the log volume of *Eucalyptus* spp. trees, using nonlinear multilevel mixed-effects model and machine learning. The techniques performed satisfactorily, and the nonlinear multilevel mixed-effects model by Schumacher and Hall with an HF structure of the variance and covariance matrix more accurately predicted the log volume of *Eucalyptus* spp. trees than the fixed-effects regression model alone, the artificial neural network and the support-vector machine. The ability to explain various sources of heteroscedasticity found in the data through random effects makes the nonlinear mixed-effects model an efficient option for *Eucalyptus* spp. trees volume prediction, and its application is recommended due to the expressive gain in precision.

AKNOLEGMENTS

The authors are grateful for the financial support provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), the Federal University of Lavras, MG, Brazil, and the Science Forest Department.

REFERENCES

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES, G.; LEONARDO, J.; SPAROVEK, G. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift, v. 22, p. 711-728, 2013.

BARANTAL, S.; CASTAGNEYROL, B.; DURKA, W.; IASON, G.; MORATH, S.; KORICHEVA, J. Contrasting effects of tree species and genetic diversity on the leaf-miner communities associated with silver birch. Oecologia, v. 189, n. 3, p. 687-697, 2019.

CALAMA, R.; MONTERO, G. Interregional nonlinear height–diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research, v. 34, p. 150–163, 2004.

CALEGARIO, N.; MAESTRI, R.; LEAL, C. L.; DANIELS, R. F. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos lineares em multiníveis de efeitos mistos. Árvore, v. 29, n. 2, p. 251 - 261, 2005.

CAMPOS, J. C. C.; LEITE, H. G. Mensuração florestal: perguntas e respostas. 3 ed. Viçosa, MG: UFV, 2009. 548p.

CARVALHO, S. P. C.; CALEGARIO, N.; FONSECA, F.; BORGES, L. A. C.; MENDONÇA, A. R.; LIMA, M. P. Modelos não lineares generalizados aplicados na predição da área basal e volume de *Eucalyptus* clonal. Cerne, v. 17, n. 4, p. 541-548, 2011.

CHERKASSKY, V.; MULIER, F. Learning from data: Concepts, theory, and methods. New York: Wiley, 1998. 560p.

CLARK, T. S.; LINZER, D. A. Should I use fixed or random effects? Political Science Research and Methods, v. 3, n. 2, p. 399–408, 2015.

CRECENTE-CAMPO, F.; TOMÉ, M.; SOARES, P.; DIEGUEZ-ARANDA, U. A generalized nonlinear mixed-effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. Forest Ecology Management, v. 259, n. 5, p. 943–952, 2010.

CROPPER JR, W. P.; PUTZ, F. E. Tree diameter increments following silvicultural treatments in a dipterocarp forest in Kalimantan, Indonesia: A mixed-effects modelling approach. Forest ecology and management, v. 396, p. 195-206, 2017.

DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. Computers and Electronics in Agriculture, v. 48, n. 3, p. 235-244, 2005.

FU, L.; ZHANG, H.; SHARMA, R. P.; PANG, L.; WANG, G. A generalized nonlinear mixed-effects height to crown base model for Mongolian oak in northeast China. Forest ecology and management, v. 384, p. 34-43, 2017.

GORGENS, E. B.; LEITE, H. G.; SANTOS, H. N.; GLERIANI, J. M. Estimação do volume de árvores utilizando Redes Neurais Artificiais. Árvore, v. 33, n. 6, p. 1141-1147, 2009.

GRANATA, F.; GARGANO, R.; de MARINIS, G. Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA's storm water management model. Water, v. 8, n. 3, p. 69-82, 2016.

GRÉGOIRE, T. G.; SCHABENBERGER, O.; BARRETT, J. P. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Canadian Journal Forest Research, v. 25, p. 137–156, 1995.

HALL, D. B.; CLUTTER, M. Multivariate multilevel nonlinear mixed effects models for timber yield predictions. Biometrics, v. 60, n. 1, p. 16–24, 2004.

HAYKIN, S. Redes neurais: princípios e prática. 2.ed. Porto Alegre: Bookman, 2001. 900p.

HUANG, S.; MENG, S. X.; YANG, Y. Using nonlinear mixed model technique to determine the optimal tree height prediction model for black spruce. Modern Applied Science, v. 3, p. 3–18, 2009.

HUFF, S.; POUDEL, K. P.; RITCHIE, M.; TEMESGEN, H. Quantifying aboveground biomass for common shrubs in northeastern California using nonlinear mixed effect models. Forest ecology and management, v. 424, p. 154-163, 2018.

KARTHIK, C.; VALARMATHI, K.; RAJALAKSHMI, M. Support vector regression and model reference adaptive control for optimum control of nonlinear drying process. If you had

to choose a kaolin supplier based on one quality, which would you choose? v. 6, p. 111-126, 2016.

LAPPI, J. Calibration of height and volume equations with random parameters. Forest Science, v. 37, n. 3, p. 781-801, 1991.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D.; SCHABENBERGER, O. SAS for Mixed Models. Second Ed. SAS Institute, Cary, NC. 2006.

MARTINS, E. R.; BINOTI, M. L. M. S.; LEITE, H. G.; BINOTI, D. H. B. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. Revista Brasileira de Ciências Agrarias, v. 11, n. 1, p. 33–38, 2016.

MENG, S. X.; HUANG, S. Incorporating correlated error structures into mixed forest growth models: prediction and inference implications. Canadian Journal of Forest Research, v. 40, n. 5, p. 977–990, 2010.

MEYER, D.; DIMITRIADOU, E.; HORNIK, K.; WEINGESSEL, A.; LEISCH, F. e1071: Mis c Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-1, 2019. Available in : <https://CRAN.R-project.org/package=e1 071>. Access in : August, 24, 2019.

MONLEON, V. J. A hierarchical model for tree height prediction. Proceedings of the 2003 M eeting of the American Statistical Association, Section on Statistics & the Environment, Alexa ndria, v. 1, p. 2865-2869, 2004.

OU, G.; WANG, J.; XU, H.; CHEN, K.; ZHENG, H.; ZHANG, B.; SUN, X.; XU, T.; XIAO, Y. Incorporating topographic factors in nonlinear mixed-effects models for aboveground biomass of natural *Simao pine* in Yunnan, China. Journal of Forestry Research, v. 27, n. 1, p. 19–131, 2016.

ÖZÇELIK, R.; DIAMANTOPOULOU, M. J.; CRECENTE-CAMPO, F.; ELER, U. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. Forest Ecology and Management, v. 306, p. 52–60, 2013.

ÖZÇELIK, R., CAO, Q. V.; TRINCADO, G.; GÖÇER, N. Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey. Forest Ecology and Management, v. 419, p. 240-248, 2018.

PINHEIRO, J. C.; BATES, D. M. Mixed-efects models in S and S-PLUS. New York: Springer-Verlag, 2000. v. 1. 528p.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL<https://www.R-project.org/>. 2017.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel, v. 81, 1986.

SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. Journal of Agricultural Research, v. 47, n. 9, p. 719-734, 1933.

SEARLE, S. R. Linear models for unbalanced data. New York: John Wiley, 1987. 536p.

SHAO, Y. H.; CHEN, W. J.; DENG, N. Y. Nonparallel hyperplane support vector machine for binary classification problems. Information Sciences, v. 263, p. 22-35, 2014.

SHARMA, R. P.; VACEK, Z.; VACEK, S.; KUČERA, M. A Nonlinear Mixed-Effects Height-to-Diameter Ratio Model for Several Tree Species Based on Czech National Forest Inventory Data. Forests, v. 10, n. 1, p. 70-101, 2019.

SILVA, J. P. M.; CABACINHA, C. D.; ASSIS, A. L.; MONTEIRO, T. C.; JÚNIOR, C. A. A.; MAIA, R. D. Redes neurais artificiais para estimar a densidade básica de madeiras do cerrado. Pesquisa Florestal Brasileira, v. 38, e201701656, 2018.

SILVEIRA, E. M. O.; SANTO, F. D. E.; WULDER, M. A.; ACERBI JÚNIOR, F. W.; CARVALHO, M. C.; MELLO, C. R.; MELLO, J. M.; SHIMABUKURO, Y. E.; TERRA, M. C. N. S.; CARVALHO, L. M. T.; SCOLFORO, J. R. Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. Forest Ecology and Management, v. 445, p. 96–109, 2019.

SIVASANKAR, T.; LONE, J. M.; SARMA, K. K.; QADIR, A.; RAJU, P. L. N. Estimation of Above Ground Biomass Using Support Vector Machines and ALOS/PALSAR data. Vietnam Journal of Earth Sciences, v. 41, n. 2, p. 95-104, 2019.

SOARES, F. A. A.; FLÔRES, E. L.; CABACINHA, C. D.; CARRIJO, G. A.; VEIGA, A. C. P. Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. Computers and Electronics in Agriculture, v. 78, n. 1, p. 19-27, 2011.

SOUZA, G. S. A.; SOARES, V. P.; LEITE, H. G.; GLERIANI, J. M.; AMARAL, C. H.; FERRAZ, A. S.; SILVEIRA, M. V. F.; SANTOS, F. C.; VELLOSO, G. S.; DOMINGUES, G. F.; SILVA, S. Multi-sensor prediction of *Eucalyptus* stand volume: A support vector approach. ISPRS Journal of Photogrammetry and Remote Sensing, v. 156, p. 135-146, 2019.

TANAKA, N.; LEVIA, D.; IGARASHI, Y.; YOSHIFUJI, N.; TANAKA, K.; TANTASIRIN, C.; NANKO, K.; SUZUKI, M.; KUMAGAI, T. O. What factors are most influential in governing stemflow production from plantation-grown teak trees? Journal of Hydrology, v. 544, p. 10-20, 2017.

TEMESGEN, H.; MONLEON, V. J.; HANN, D. W. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. Canadian Journal of Forest Research, v. 38, n. 3, p. 553–565, 2008.

UZOH, F. C. C.; OLIVER, W. W. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western United States using a multilevel linear mixed effects model, Forest Ecology Management, v. 256, n. 3, p. 438–445, 2008.

UZOH, F. C. C.; MORI, S. R. Applying survival analysis to managed even-aged stands of ponderosa pine for assessment of tree mortality in the western United States. Forest Ecology Management, v. 285, p. 101-122, 2012.

UZOH, F. C. C. Height-Diameter Model for Managed Even-aged Stands of Ponderosa Pine for the Western United States Using Hierarchical Nonlinear Mixed-Effects Model. Australian Journal of Basic and Applied Sciences, v. 11, n. 14, p. 69-87, 2017.

VALENÇA, M. Fundamentos das redes neurais: exemplos em Java. 2 ed. Olinda: Livro Rápido, 2010. 386p.

VENDRUSCOLO, D. G. S.; DRESCHER, R.; SOUZA, H. S.; Moura, J. P. V. M.; MAMORÉ, F. M. D.; SIQUEIRA, T. D. S. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. Revista Brasileira de Biometria, v. 33, n. 4, p. 556–569, 2015.

VUONG, Q. Likelihood ratio tests for model selection and nontested hypotheses. Econometrica, v. 57, p. 307-333, 1989.

WANG, W.; CHEN, X.; ZENG, W.; WANG, J.; MENG, J. Development of a Mixed-Effects Individual-Tree Basal Area Increment Model for Oaks (*Quercus* spp.) Considering Forest Structural Diversity. Forests, v. 10, n. 6, p. 474-493, 2019.

WEST, B. T.; WELCH, K. B.; GATECKI, A. T. Linear mixed models: a practical guide using statistical software. 2 ed. Ann Arbor: Chapman and Hall/CRC, 2015. 405p.

WONG, T. T.; YANG, N. Y. Dependency analysis of accuracy estimates in k-fold cross validation. IEEE Transactions on Knowledge and Data Engineering, v. 29, n. 11, p. 2417-2427, 2017.

WYCKOFF, W. R. A basal area increment model for individual conifers in the northern Rocky Mountains. Forest Science, v. 36, n. 4, p. 1077–1104, 1990.

YANG, D.; LIU, Y.; LI, S.; LI, X.; MA, L. Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm. Mechanism and Machine Theory, v. 90, p. 219-229, 2015.

YU, H.; LIANG, Y.; LIANG, H.; ZHANG, Y. Recognition of wood surface defects with near infrared spectroscopy and machine vision. Journal of Forestry Research, v. 1, p. 1-8, 2019.

Zeng, W.; Zhang, D.; Fang, Y.; Wu, J.; Huang, J. Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data. Journal of Applied Remote Sensing, v. 12, n. 2, p. 022204, 2018.

FINAL CONSIDERATIONS

Determining, modeling, and supplying forest carbon stock data is a strong scientific and social demand currently, as tree carbon storage is considered a key environmental service in mitigating current climate change by sequestering CO_2 from the forest atmosphere. In this Thesis, we determined the soil and root carbon stock in four depths under a tropical forest in Brazil; estimated carbon in tree stems; and assessed carbon spatial patterns in the three different compartments (soil, tree stem, fine roots) and in total above-belowgroung stock, seeking for possible spatial relationships among these variables.

The knowledge of in-depth variation of soil carbon and the solid spatial study of this variable together with information on the carbon present in other forest compartments might provide a better picture of the carbon stocking and a deeper comprehension of forest ecosystems. Additionally, this study provides high-quality in situ data on carbon stock and pools, meeting the demand to understand controls on carbon stocks and cycling, to calibrate global carbon cycle models, and to support regulatory frameworks such as the United Nations REDD program (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries).

The performance of the support vector machine (SVM) techniques and proposed a new nonlinear model extracted from the training of an artificial neural network (ANN) in the modeling the above ground carbon stock in a secondary semideciduous forest were proposed and evaluated. The results brings important contributions in the modeling of carbon stocks in forests through the use of machine learning. The machine learning techniques performed satisfactorily, and a new model extracted from an artificial neural network for carbon stock prediction in $Mg.ha^{-1}$ is efficient, with potential application in other secondary semideciduous seasonal forests. The carbon stock modeling using a model extracted from the artificial neural network training presented better performance than the support vector machine, using the same variables of the analyzed data set.

We assessed the performance of the nonlinear volumetric model by Schumacher and Hall fitted by fixed effects regression and by incorporating multilevel random effects with variance and covariance matrix modeling; a volume equation extracted from an artificial neural network training process; and the construction of a support-vector machine to estimate the log volume of Eucalyptus trees, one of the most important parameters for knowing the forest production potential of a region.

This study considerably improved the volume modeling of Eucalyptus tree logs, with a marked 53-fold decrease in residual standard error in the predicted volumes by the nonlinear multilevel mixed-effects model by Schumacher and Hall with a Huynh-Feldt structure of the variance and covariance matrix. This study confirms that although fixed-effects models are widely used in the Brazilian forest sector, there are more effective methods for modeling dendrometric variables.