IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 12, Issue 1, January. 2022, || Series -I || PP 01-07

Artificial neural networks in estimating the productivity of a forest harvesting machine

Mayra de Oliveira Belcavello[1], Diogo de Souza Alves[2], Edney Leandro da Vitória[1*], Élcio das Graça Lacerda[3], Tamara Locatelli[4], Ismael Lourenço de Jesus Freitas[4]

[1]Department of Agricultural and Biological Sciences, Federal University of Espirito Santo, São Mateus-ES, Brazil.
[2]Postgraduate Program in Tropical Agriculture, Federal University of Espirito Santo, São Mateus-ES, Brazil.
[3]Federal Institute of Education, Science and Technology of Espirito Santo, Santa Teresa-ES, Brazil.
[4]Secretariat of Education for the State of Espirito Santo, Brazil
**Corresponding Author: vitoria.edney@gmail.com*
Received 11 January 2022; Accepted 27 January 2022

Abstract: The forestry sector, a relevant segment in the Brazilian economy and in the State of Espírito Santo, highlighting the genus Eucalyptus. Forest harvesting is a final and most important operation in the production chain, covering everything from felling to extraction of trees. The harvesting system is evaluated in whole tree (whole trees) or cut to length (short logs, the most used in Brazil), carried out through the use of the harvester, which is one of the most modern forest tractors, whose productivity is influenced by several variables. For variable-based modeling, they have been used as Artificial Neural Networks (ANNs), which precede a meta-heuristic based on the structure and functioning of the brain, which trains one and solves specific specifications. The objective of this work was to estimate the harvester productivity by relating different variables in ANNs. Data used from the capture process in Suzano operations in the northern region of Espírito Santo, Brazil, were used. The Neuroforest software version 4.0.6 was used for training 5 ANNs with backpropagation algorithm and sigmoidal activation function, training calibration groups (80.0%) and validation (20.0%) with random sampling method and criterion of stop was mean error (0.001). Correlation coefficient results were obtained between 0.96 and 0.93 for training and validation, respectively, and RQME approximately 1.35% for training and 2.57% for validation. The conclusion shows the effectiveness of using ANNs to estimate harvester productivity by relating different variables..
Keywords: forestry, harvesting, artificial intelligence, Eucalyptus.

I. INTRODUCTION

Among the trees planted in Brazil, the genus Eucalyptus stands out, with approximately 5.5 million hectares planted and an average productivity of 39 m³/ha/year according to the Brazilian Tree Industry[1]. In Espírito Santo, it is estimated an occupation of around 250 thousand hectares of planted eucalyptus Celulose.

Among the processes in the pulp production chain, forest harvesting is defined as the final operation and the most important in obtaining raw material, encompassing stages ranging from felling to extraction of trees. Mechanized harvesting is a reality in numerous companies in the Brazilian forest sector, as it has lower production costs and lower rates of accidents and occupational diseases, in addition to a low labor force and greater operational efficiency. The better control over operator productivity is also highlighted, as the equipment is equipped with devices that monitor production and effective working time[2].

There are several methods and systems for harvesting and processing wood in the field, according to the forest species, age of the population, purpose of the product and general conditions of the harvesting area[3]. The system is classified . as to the shape of the wood and the length of the logs in the extraction phase: if the tree is extracted in its entirety to the processing location, it is classified as Full-tree (whole trees). have the cut-to-length system (short logs), which are the most used in Brazil and in the harvesting of Eucalyptus, using the Cellulose harvester.

The harvester is considered one of the most modern forest tractors, as it has a head that performs various functions and operates in adverse conditions. Productivity is registered by on-board computers and the information helps in managing the operation. Several parameters can influence the productivity of harvesters, such as the average volume of trees, the slope of the area, sectioning length of the logs and the efficiency of the operator[4].

Among the researches carried out, no articles were found using the concept of artificial neural networks (ANNs) in the relation of harvester productivity with variables such as: date, equipment, shift, operator, production unit, average individual volume, tree height, average increment annual without bark, hours worked, adjusted production, useful area, clone, number of stems/ha, cut volume without bark.

For modeling involving numerous variables, Artificial Neural Networks (ANNs) have been used. The characteristic of ANNs to extract information from data makes them a powerful and flexible tool for predictive and modeling purposes, and offers great potential for applications in different fields[5].

There are several applications of ANNs in forestry, such as forest classification and mapping, forest growth modeling, spatial data analysis, analysis of plant disease dynamics, climate change, and fuel consumption prediction[6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16].

The use of computational tools is important to assist in the task of planning and reducing costs in forest harvesting[17]. ANNs constitute a meta-heuristic based on the structure and functioning of the human brain and central nervous system, being useful in recognizing patterns based on the accumulation of experience or examples, training stored knowledge and solving a specific situation[18].

This work was carried out under the hypothesis that increasing the number of cycles in the modeling of ANNs improves the validation parameters of the productivity estimate. Thus, the main objective of the work was to estimate the harvester productivity relating different number of cycles in ANNs.

II. METHODOLOGY

Data from the forest harvesting process were obtained, compiled and organized in a production monitoring database of a company in the forest sector located in the northern region of Espírito Santo. The data collection period was from September 24th to October 26th, 2020.

The database with 1,197 observations contains information from 10 harvesters, 45 operators in rotation of three shifts lasting 8 hours in a period of 31 days, totaling 6,937.54 hours worked, in 20 production units (UP) that varied the useful area from 10.1 to 121.76 ha with 6 different clones. The individual mean volume (IMV) ranged from 0.15499 to 0.30017 m3, the total height (HT) from 22.21 to 27.73 m and mean annual increment without coat (IMASC) was in the range between 25.99 and 36 .93 m3/ha/year. Data from the forest register (Useful Area, Clone, number of boles/ha and volume of cut without bark (VCSC)) and from the pre-cut inventory (IPC) (UP, VMI, HT and IMASC) were used.

For this work, the short log harvesting system (cut-to-length) and the operation via the Komatsu excavator model PC200 equipment, with a model 370E head from the same manufacturer, were analysed, the variables were obtained through an electronic recording system and classified as quantitative and qualitative (Figure 1). The harvester used for cutting (felling and processing) is made up of a self-propelled base machine with wheels made of metal tracks, hydraulic boom to reach the trees that, together with the head, carry out the operations of felling, delimbing, debarking, and tracing the wood[18].

Figure 1. Forest harvesting machine and data acquisition system

The sequence of operations performed consists of felling, felling the tree; in delimbing, removal of all branches of the tree by means of a horizontal sliding of the tree trunk inside the harvester head; in debarking, removal of the bark simultaneously with the delimbing; in the toragem, sectioning of the shaft in dimensions of seven meters; and in flagging, formation of small piles after sectioning Gonçalves (2017).

The quantitative variables were: the average individual volume (VMI) being the measurement of the volume of each tree (estimated in the pre-cut inventory) with measurement unit in cubic meters; the hours worked indicating the time for carrying out the operations; adjusted production considered as the production given in cubic meters; useful area representing the entire worked area in hectares; the number of stems is the number of shoots conducted in each stump per hectare; cut volume (VCSC) given in cubic meters per hectare, not considering the bark; the total height (HT) of the trees used in meters and the average annual increment without bark (IMASC) characterized in cubic meters per hectare in each year. Qualitative variables: relay shifts; operators; equipment; production units represented by each studied area and distinct clones.

The data survey spreadsheet showed inconsistencies, such as duplicate data, production time values below half an hour and yields that did not match the reality of the field, showing overfitting. Therefore, the spreadsheet was cleaned to adjust the data for a related training. It is recommended to use ANN with the simplest possible configuration and the smallest possible number of neurons in the hidden layer, in order to avoid overfitting. In other words, avoid over-learning the information contained in the data presented to the networks. Otherwise, the trained networks end up copying not only the structural similarity between the variables, but also the noise (relationship error)[18].

The NeuroDAP version 4.0.6 software was used to train the ANNs, which consists of adjusting the network parameters (weights and bias) through[18] learning algorithm for error backpropagation (BackPropagation). The error backpropagation algorithm is one of the best known. This algorithm first adjusts the weights of the output layer according to the difference between the desired output and the one presented by the network. Subsequently, this error is back propagated to the neurons of the anterior layers, and the desired response of a neuron in one layer will be determined as a function of the posterior layers[20].

In this process, the training data were presented to a pre-established architecture containing arrays of neurons in layers for both the hidden eight and the output layer, composed of a sigmoidal activation function, the most used in ANNs. Such a function is defined as strictly increasing, as it exhibits an adequate balance between linear and non-linear behavior. Thus, the number of hidden layers and neurons that networks have will vary according to the problem addressed. The decision on the size of the network that will be used demands experimentation with different configurations, trained more than once to avoid problems associated with local minima, and observing individual performances[20].

The data were separated into training calibration (80.0%) and validation (20.0%) groups, using the random sampling method as well as in the work carried out by Lafetá (2012). The stopping criterion parameter were mean error (0.0001) or number of cycles (3,000 to 15,000), that is, when reaching one of these parameters, the network training was interrupted.

For each ANN, the root mean square error (RMQE%) was calculated using the formula:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{n}}$$

Where n is the total number of observations; c_i the observed value and $\hat{c}_i$ the estimated value.

III. RESULTS

Using the NeuroDAP software, we added data taken from the electronic recording system connected to the harvester and organized it into qualitative and quantitative variables in the software for training the network. It was possible to validate the harvester productivity estimate when presented an interface with different variables in its intermediate layer, presenting us as output the productivity as had been predetermined.

We chose to estimate productivity with the help of software as an artificial intelligence tool as it has significant features and qualities. When choosing different variables for training the network, we understand that not only one factor influences learning, but the set of factors to validate a good estimate through the diagnosis made by the network.

The mean, minimum, maximum and coefficient of variation (CV) values for the quantitative variables are shown in Table 1. The descriptive analysis allowed exploring and analyzing the values of the variables and verifying the existence of outliers or outliers, as well as their influence in the data of measurements of position and dispersion; therefore, when identified, outliers were eliminated from the analysis.

Table 1. Descriptive analysis for quantitative variables of forest harvest.

Variáveis	**Minimun**	**Avarage**	**Maximum**	**CV**
VMI	0.155	0.197	0.300	19.40%
HT(m)	22.210	24.640	27.730	5.43%
IMASC	25.987	32.355	36.929	8.10%
Worked hours	0.635	5.796	13.499	23.18%
Prod. Adjusted	7.875	130.856	329.947	32.33%
Productivity ($m^3\ h^{-1}$)	6.223	22.368	36.744	20.83%
Useful area	10.100	55.752	121.760	61.91%
Num. Fustes/ha	833.330	1041.705	1175.000	7.83%
VCSC ($m^3\ ha^{-1}$)	167.940	193.733	253.530	13.34%

It is observed that this data set (Table 1) has a high amplitude and the CV of observed data is high, indicating that the total of 1,197 observations is capable of representing the variation in the machine's productivity.

Table 2 shows the characteristics and statistical analysis of the 5 ANNs generated in the software, both in training and in validation to estimate harvester productivity. The results of the correlation coefficient were between 0.749 and 0.930 in training and 0.678 and 0.856 for validation. The RQME% presented values from 3.073 to 1.719 for training and 3.512 to 2.469 for validation and the variance in training was from 9.450 to 2.957 while in validation from 12.384 to 6.100. Thus, representing high precision and good correlation for estimation, using all available input variables, both qualitative and quantitative.

Table 2. Characteristics and statistical analysis of artificial neural networks to estimate harvester productivity.

NNA	**Input**		**Output**	**Training**			**Validation**		
	Qualitative	**Quantitative**		**RQME**	**Correation**	**Varriance**	**RQME**	**Correation**	**Varriance**
1	Date. Equipment. Operator. Shift. UP and Clone	VMI. Hours Worked. Prod. Adjusted. VCSC. Useful Area. N° Shafts. HT and IMASC	Productivity ($m^3 h^{-1}$)	3.073	0.749	9.450	3.512	0.678	12.384
2				2.488	0.845	6.198	3.147	0.753	9.938
3				2.194	0.882	4.820	2.921	0.792	8.547
4				1.963	0.907	3.858	2.718	0.823	7.390
5				1.719	0.930	2.957	2.469	0.856	6.100

In principle, our stopping criterion was the mean error at 0.0001 and the number of cycles at 3,000. When reaching one of these values, the network learning is interrupted, changing to the next ANN. However, by placing the network to understand and process this information and present us with the desired output, the network was limited to succinct learning. Seeking better learning results and a good correlation, the stopping criterion was readjusted to an average error of 0.0001 and the number of cycles progressively increasing in 3,000 cycles for each ANN, that is, ANN 1 was trained in 3,000 cycles, the ANN 2 in 6,000 cycles, RNA 3 in 9,000, RNA 4 in 12,000 and RNA 5 in 15,000 cycles.

ANN 5 was higher in correlation in training and validation with values of 0.930 and 0.856 respectively, as well as low values of RMQE%. While the ANN that presented the lowest values in correlation was RNA 1 in 0.749 in training and 0.678 in validation and higher RMQE%, with stopping criterion of the number of cycles in 3,000 provided less learning to the network and higher values of RMQE% in in relation to the 15,000 cycles of ANN 5 (Figure 2).

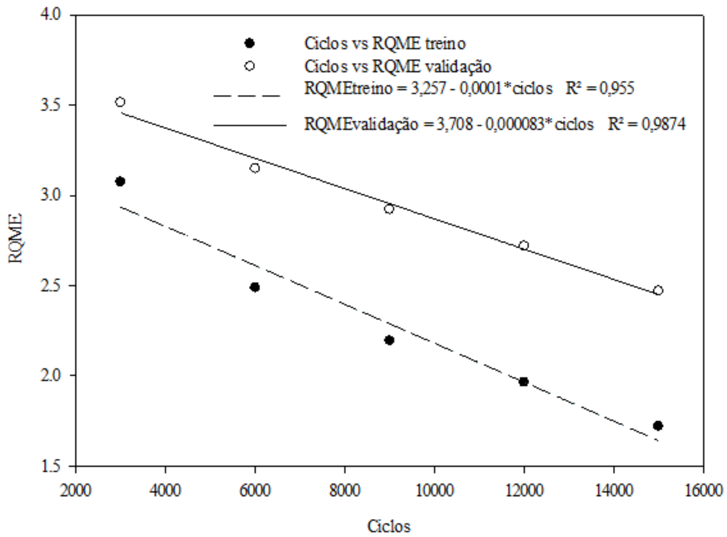
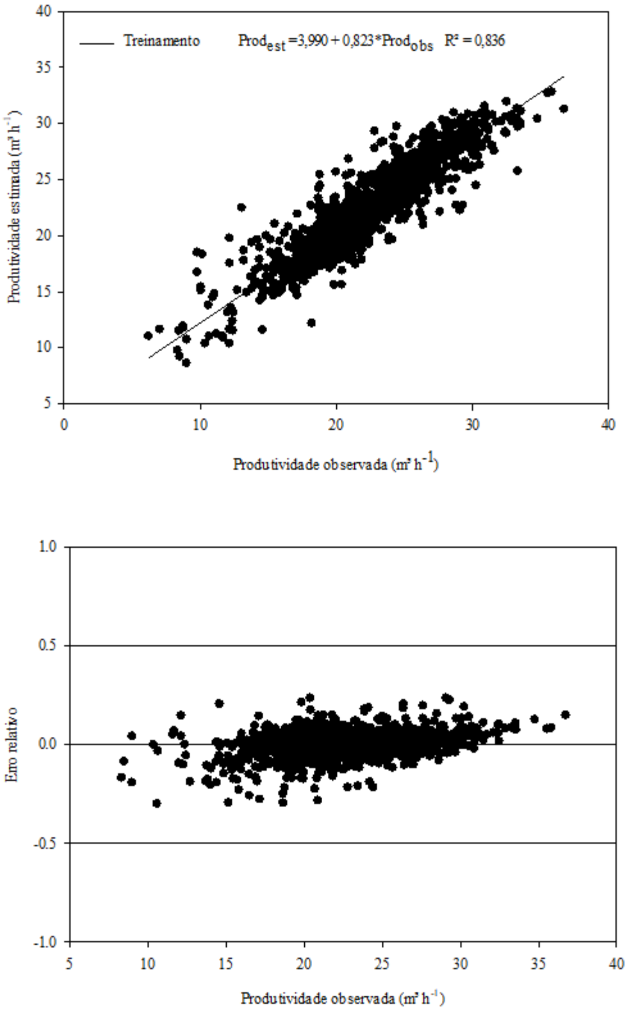


Figure 2. RMQE% for each number of cycles used for training and validation data.

For each variation of a cycle, the error decreases 0.0001 times, so for each cycle we want to work, we multiply the quantity by the error value in the function, that is, by 0.0001. It is observed in Figure 1, as well as in the descriptive analysis, that working with 5 points, that is, 5 ANNs and increasing the number of cycles, we reduced the error value reaching the minimum point.

The relationship of productivity values in m3 h-1, observed and estimated for ANN 5, which presented the highest and lowest RMQE% correlations, are shown in Figure 3 and 4, as well as the productivity error ratio observed for the data selected for training and in Figure 3 for validation.

The coefficient of determination presented a value considered high, it can be inferred that 83.6% of the estimated productivity explained by the ANN can be explained by the observed productivity, the relative errors varied between -30% and 30%.



a

b

Figure 3. Observed versus estimated productivity for data used in training (a) and residuals for the obtained estimate (b).

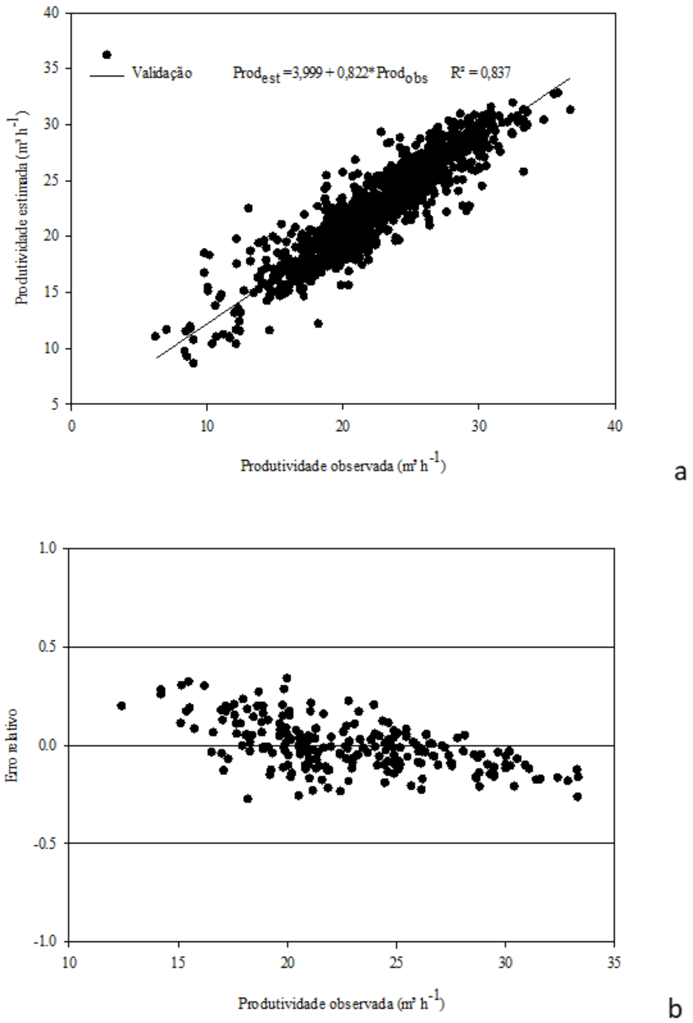


Figure 4. Observed versus estimated productivity for data used in the validation (a) and residuals for the obtained estimate (b).

The application of variables that affect forest harvesting such as date, equipment, operator, shift, VMI, HT, IMASC, hours worked, adjusted production, useful area, number of stems, VCSC, UP and clone using ANNs showed accuracy and efficiency in validating the harvester productivity estimate, as all 5 networks trained in NeuroDAP with backpropagation algorithm and sigmoidal activation function presented low RQME% and good correlation.

IV. DISCUSSION

Several works present a good precision and correlation of training estimates with the use of ANNs. In his doctoral thesis[18], while working on the prediction of specific harvester fuel consumption, proved the usefulness of networks in planning and control.

In his doctoral thesis[17] by performing the mathematical modeling of productivity in mechanized forest cutting using quantitative (individual average volume, volume of wood in the stand and cutting age) and qualitative (spacing, operator experience and management regime) variables) proved the efficiency in the application of ANN to generate harvester productivity estimates. This author obtained networks with high correlation, with values between 0.888 and 0.900 and low RMSE% between 11.55 and 11.91. Corroborating with [18], who found that the accuracy of estimates can be increased by including variables such as: the type of eucalyptus clone, personal status of the operator at the time of operation, and ambient temperature.

In the statistical analysis[5] of the ANNs observed high accuracy in the generalizations of the networks to estimate the productivity of the forwarder in wood extraction, showing the efficiency of the method during training. The values obtained by this author in the validation correlation were higher than 0.7404 using Multilayer Perceptron (MLP) networks in Neuroforest.

V. CONCLUSION

The results obtained in the training of ANNs showed that this technique has considerable potential to estimate harvester productivity in eucalyptus harvesting. The data proved that the ANNs are capable of estimating the harvester's productivity under the conditions used

REFERENCES

[1]. RELATÓRIO ANUAL. **IBÁ,** 2020. Disponível em: <https://iba.org/datafiles/publicacoes/relatorios/relatorio-iba-2020.pdf>. Acesso em: 06 de mar. de 2021.

[2]. Vieira, G. C.; Freitas, L. C.; Cerqueira, P. H. A.; Silva, E. F.; Brito, G. S.; Souza, A. M (2016). Custos Operacionais e de Produção na Atividade mecanizada de Corte Florestal. **Nativa**, Sinop, v.4, n. 5, p. 342-346.

[3]. Simões, D.; Fenner, T.; Esperancini, M. S. T. Avaliação técnica e econômica da colheita de florestas de eucalipto com harvester. **Scientia Forestalis**, Piracicaba, v. 38, n. 88, p. 611-618, 2010. Disponível em: <http://hdl.handle.net/11449/137339>.

[4]. Gonçalves, S. B.; **Modelagem matemática da produtividade do corte florestal mecanizado.** 2017. 63p. Tese (Doutorado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro - ES, 2017.

[5]. Silva, F.; **Aplicação de redes neurais artificiais para estimar a produtividade do forwarder na extração de madeira.** 2015. 50p. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2015.

[6]. Hilbert, D. W; Ostendorf, B (2001). The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. **Ecological modelling**, [S. l.], v.146, n.1, p.311-327.

[7]. Ozçelik, R.; Diamantopoulou, M. J.; Brooks, J. R.; Wiant, H. V (2010). Estimating tree bole volume using artificial neural network models for four species in Turkey. **Journal of environmental management**, [S. l.], v.91, n.3, p.742-753.

[8]. Hickey, C.; Kelly, S.; Carroll, P.; O'connor, J (2015). Prediction of forestry planned end products sing dirichlet regression and neural networks. **Forest Science**, UCD-Ireland, v.61, n.2, p.289-297.

[9]. Binoti, M.L.M.S.; Leite, H.G.; Binoti, D.H.B.; Gleriani, J.M (2015). Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. Cerne, Lavras-MG, v. 21, n. 1, p. 97-105.

[10]. Bhering, L. L.; Cruz, C. D.; Peixoto, L. D. A.; Rosado, A. M.; Laviola, B. G.; Nascimento, M (2015). Application of neural networks to predict volume in eucalyptus. Crop Breeding and Applied Biotechnology, Viçosa-MG, v.15, n.3, p.125-131.

[11]. Silva Ribeiro, R. B.; Gama, J. R. V.; Souza, A. L.; Leite, H. G.; Soares, C. P. B.; Silva, G. F (2016). Métodos para estimar o volume de fustes e galhos na Floresta Nacional do Tapajós. **Revista Árvore**, Viçosa-MG, v.40, n.1, p.81-88.

[12]. Zanuncio, A. J. V.; Carvalho, A. G.; Silva, L. F. D.; Carneiro, A. D. C. O.; Colodette, J. L (2016). Artificial neural networks as a new tool for assessing and monitoring wood moisture content. **Revista Árvore**, Viçosa-MG, v.40, n.3, p.543-549.

[13]. Campos, B. P. F.; Da Silva, G. F.; Binoti, D. H. B.; De Mendonça, A. R.; Leite, H. G (2016). Predição da altura total de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. Pesquisa Florestal Brasileira, [S. l.], v.3, n.88, p.375-385.

[14]. Martins, E.R.; Binoti, M. L. M.S.; Leite, H. G.; Binoti, D. H. B.; Dutra, G. C (2016). Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto. **Revista Brasileira de Ciências Agrárias**, Recife-PE, v.11, n.2, p.117-123.

[15]. Vahedi, A. A (2017). Monitoring soil carbon pool in the Hyrcanian coastal plain forest of Iran: Artificial neural network application in comparison with developing traditional models. **Catena**, Oxford-UK, v.152, p.182-189.

[16]. GONÇALVES, S. B.; **Modelagem matemática da produtividade do corte florestal mecanizado.** 2017. 63p. Tese (Doutorado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro - ES, 2017.

[17]. SANTOS, L.; **Predição do consumo específico de combustível do harvester utilizando modelos lineares mistos e redes neuronais artificiais.** 2018. 58p. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2018.

[18]. Leite, H. G.; Da Silva, M. L. M.; Binoti, D. H. B.; Fardin, L.; Takizawa, F. H (2011). Estimation of inside-bark diameter and heartwood diameter for Tectona grandis Linn. Trees using artificial neural networks. **European Journal of Forest Research**, Viçosa-MG, v.130, n.2, p.263-269.