Machine learning techniques to estimate mechanised forest cutting productivity

Saulo Boldrini Gonçalves¹, Nilton Cesar Fiedler², Jeferson Pereira Martins Silva²*, Gilson Fernandes da Silva², Mayra Luiza Marques da Silva³, Luciano José Minette⁴, Daniel Pena Pereira⁴, Domingos Manuel Mendes Lopes⁵, Evandro Ferreira da Silva⁶, Antônio Henrique Cordeiro Ramalho², Jeangalis Silva Santos⁵, Marcelo Otone Aguiar⁵, Jose de Oliveira Melo Neto¹ and Renisson Neponuceno de Araújo Filho¹

The productivity of wood harvesting operations is one of the main viability indicators of the forestry enterprise, which is directly influenced by land, population, and operational planning characteristics. The variables that affect the productivity of harvesting machines are particularly difficult to measure and have complex relationships, making it challenging to predict the productivity of operations. This study generated a model using machine learning (ML) techniques to estimate harvesting productivity in Eucalyptus plantations in southeastern Brazil. The input variables for modelling harvesting productivity were the average individual tree volumes, wood volume in the stand, cutting age, spacing, operator experience, and the management regime. The database was randomly divided into training (70%) and validation (30%) datasets. Boosted, artificial neural network (ANN), and adaptive network-based fuzzy inference system (ANFIS) techniques were used to fit the model and were evaluated through statistics and graphical analysis of the residues. The configurations selected for training and validation to estimate harvester productivity resulted in correlation coefficient values greater than 0.9, and root-mean-square error (RMSE) percentages less than 12.41, indicating a strong correlation and high accuracy between the estimates and the observed values. The boosted technique yielded the best results, with correlation coefficients of 0.98 and 0.97, and RMSE percentages of 6.15 and 6.65 in training and validation, respectively. The worst performance for estimating harvesting productivity was obtained using the ANFIS technique. ML techniques were efficient in modelling the productivity of mechanised forest cutting with a harvesting model.

Keywords: Brazil, forest planning, forestry operations, fuzzy inference, harvester operation, machine productivity, mechanised harvesting

Online supplementary material: Supplementary data for this article are available at https://doi.org/10.2989/20702620.2020.1994342

Introduction

Harvesting of forest resources is defined as a set of activities and pre-established techniques conducted in the forest stand to prepare and dispose of the wood at the transport site (Rodrigues et al. 2018). Owing to the significant capital investment for the acquisition of equipment, high operational complexity, a large number of variables capable of reducing the productivity of the system, and the representativeness of the final cost of wood, harvesting stands out as one of the most important stages in the productivity of the forest system (Hiesl et al. 2015; Lacerda et al. 2017; Leite et al. 2019).

Many companies opt for a high degree of mechanisation in the search for uniform performance, reduction in labor dependence, decreased accident rate, and increased productivity (Schettino et al. 2015). High individual performance of the machines guarantees a direct influence on the rate of return on invested capital (Leite et al. 2014). The extensive number of industries that use wood in production processes led forestry companies to develop different harvesting systems to serve the diverse market. Among the harvesting systems recognised by the Food and Agriculture Organization of the United Nations (FAO), one of the most-used systems in Brazil is cutting the wood to a specific length. This is characterised by conducting all activities related to the cutting stage (felling, delimbing, uncappping, tracing and stacking) within the harvesting system by cutting the trunks into logs up to 7 m long (Malinovski et al. 2014).

The specific characteristics of the harvester heads allow simultaneous performance of all activities inherent to forest cutting; therefore, machines equipped with this feature are the most-used in the short-log system (Erber et al. 2016; Malinen et al. 2016; Norhio et al. 2018). However, because of the high cost of acquisition, the variables that influence the productivity of these machines must be constantly monitored and controlled. Consequently, the application of methodologies capable of predicting the productivity of machines in different land and population conditions through an analysis of the
variables to their behavior in the model becomes essential in forest planning. Some studies have already been conducted to estimate harvester productivity by linear regression (Leite et al. 2013). However, mechanised harvesting is a complex process and involves the interaction of many variables, which include the size of stands, extraction distance, slope of terrain, experience of the operator, type of machine, characteristics of the stand, model and intensity of thinning, and assortments (Malinovski et al. 2006). The relationship of these variables with harvester productivity is hardly a linear function. Therefore, the most appropriate possibility is the use of non-linear models. However, the use of non-linear models requires *a priori* knowledge of the relationship between these variables and the productivity of mechanised harvest. An alternative would be the use of machine learning (ML) methods, such as artificial neural network (ANN), random forest (RF), support-vector machine (SVM), adaptive network-based fuzzy inference system (ANFIS) and boosted techniques, which enable the calculation of the productivity of mechanised harvesting without prior knowledge of the relationship of the variables.

These techniques have been efficient for function approximation in some forestry applications (Vieira et al. 2018; Silva et al. 2019). However, few studies have tested the efficiency and effectiveness for harvester productivity prediction. Thus, the objective of this study was to estimate the productivity of mechanised harvesting through ML techniques.

**Materials and methods**

**Characterisation of the study area**

Data collection was carried out by a forestry company located in the states of Bahia and Espírito Santo, Brazil. These areas have clonal plantations of *Eucalyptus urophylla* S.T.Blake and *E. grandis* W.Hill hybrids, which are distributed in four regions (Figure 1).

The study areas have a flat relief with smooth undulations (maximum slope of up to 5%) and elevations between 10 and 50 m. According to the Köppen classification, the climate is type Aw (tropical wet with a dry winter), with an average annual rainfall of 1,350–1,375 mm, a rainy period from October to December, a dry period from July to September, and irregular rainfall between January and June. In these areas the following soil types predominate: an abrupt yellow Argisol A, moderate Planossol A or prominent A, and Quartzarenic Neossol (Silva 2011).

![Figure 1: Location of the municipalities where the research was carried out](image-url)
Database
The data were collected at the field level (4 941 fields in total), in three work shifts involving 312 operators and 82 harvester forestry machines (base machine + processor head), over a period of three years (1 Jan 2014 to 31 Dec 2016). Thus, the total number of observations was 47 000 (trees harvested). The dimensions of the base machine model for the harvester used in this research are shown in Figure 2. Table 1 presents the technical specifications of the base machine model evaluated in this study. Figure 3 shows the dimensions of the harvester head model evaluated in this research in different positions as used during the activities.

Harvesting system
The harvesting system adopted by the forestry company is cut-to-length and is fully mechanised. The activities related to the cutting stage (felling, deliming, peeling, twisting and flagging) were carried out by the harvester machines. The pre-defined size of the logs was 6 m. The cutting area (groups of lines working together during forest cutting) was made up of four rows of trees, with logs deposited in flags between the first and second lines of the field.

Description of database variables
The quantitative variables considered in this study for modelling the productivity of mechanised harvesting with the harvester model were the average individual tree volume (VMI, in m³), wood volume in plot (VM, in m³), and cutting age (I, in months). The slope of the terrain was not considered in this study because the areas were all flat with

Table 1: Technical specifications of the harvester evaluated in this study (source: Komatsu, http://www.komatsuforest.com.br)

| Item                   | Characteristic                                                                 |
|------------------------|-------------------------------------------------------------------------------|
| Cabin rotation         | 360°                                                                          |
| Weight (kg)            | 21 000                                                                        |
| Engine                 | E-1 six-cylinder, turbo-powered with aftercooler, four-stroke, water-cooled,  |
|                        | direct injection, and 2 000 rpm                                                  |
| Engine power (HP)      | 155                                                                           |
| Maximum speed (km/h)   | 5.5                                                                           |
| Crane reach (m)        | 9.38                                                                          |
| Wheel                  | Mats with central frame in “X”                                                 |

Figure 2: Design of the harvester model evaluated by this research (source: Komatsu, http://www.komatsuforest.com.br)

Figure 3: Harvester head-sizing assessed in the survey (source: Komatsu, http://www.komatsuforest.com.br)
a maximum slope of 5%. Table 2 presents the descriptive analysis of the quantitative variables used in modelling the cutting productivity with the harvester model of the evaluated plots.

A standard deviation of ±1 883 m³ was observed for VM. This variation in the wood volume in the field is due to the diversity of field sizes in the company’s operational areas among the 13 municipalities that make up this research.

The categorical variables considered in this study for the modelling of mechanised harvesting productivity were spacing (ES), operator experience (EO), and management regime (RM). All variables were extracted from the pre-cut inventory data and from the operator’s record which is related to the production record of the operators involved in the cutting of the company’s plots. The categorical variables used for modelling are shown in Table 3. The number of classes of the categorical variables spacing ES, EO and RM were 3, 2 and 3, respectively.

Techniques to estimate harvester productivity and statistical indicators to assess accuracy

In this research, harvester productivity was estimated by the techniques of boosted algorithm, ANN and ANFIS. More details on these techniques are provided in the online supplementary material.

The database was randomly divided into training (70%) and validation (30%) datasets. The evaluation of the techniques analysed to estimate the productivity of the harvester were calculated as follows:

\[
\text{Relative bias: } B(\%) = \frac{100}{\bar{Y}} \sum_{i=1}^{n} \frac{\bar{Y}_i - \bar{Y}}{\bar{Y}} \\
\text{Relative root-mean-square error: } \text{RMSE}(\%) = \frac{100}{\bar{Y}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2} \\
\text{Correlation coefficient: } r_{YY} = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(\bar{Y}_i - \bar{Y}_m)}{\sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 \sum_{i=1}^{n} (\bar{Y}_i - \bar{Y}_m)^2}}
\]

where: \(Y_i\) = observed dependent variable; \(\bar{Y}_i\) = estimated dependent variable; \(\bar{Y}\) = mean of the dependent variable observed; \(\bar{Y}_m\) = mean of the estimated dependent variable; and \(n\) = number of observations.

To complement the statistics, graphs were created relating the values observed and the values estimated by the techniques and graphs of percentage residues. The error of each observation was calculated as a percentage:

\[
E(\%) = \frac{Y_i - \bar{Y}_i \times 100}{\bar{Y}_i}
\]

Results

An important step in prediction models that use ML techniques is the calibration of the parameters. This step can be performed by experimental methods or by exhaustive search. In this study, an exhaustive search was proposed for the best average combination of parameters in a pre-specified interval, whose performance indicators were the RMSE percentage and the correlation coefficient. The supplementary material includes the results of the calibration with the boosted, ANN and ANFIS techniques (Supplementary Figures S3, S4 and S5, respectively).

The variation of the parameters resulted in a total of 1 402, 760 and 700 combinations for the boosted, ANN and ANFIS techniques, respectively. The best configuration for each technique was selected based on the training and validation errors. The obtained parameters and statistical indicators are shown in Table 4.

The configurations selected for the training and validation to estimate harvester productivity had correlation coefficient values of > 0.9 and RMSE percentages of < 12.41, indicating a strong correlation and high accuracy between the estimates and the observed values. The boosted technique presented the best \(r_{YY}\) (0.98 and 0.97) and RMSE percentage (6.15 and 6.65) results for the training and validation datasets, respectively. The worst performance for estimating harvester productivity was obtained using the ANFIS technique.

When analysing the graphs of the wood volume harvested estimated by the boosted, ANN and ANFIS techniques and the actual volume of wood observed, the boosted technique had more points closer to the one-to-one line compared with

| Variable                  | Min.   | Max.   | Average | Standard deviation |
|---------------------------|--------|--------|---------|--------------------|
| VMI (m³)                  | 0.0132 | 0.3798 | 0.1206  | ±0.0448            |
| VM (m³)                   | 7.8    | 12 384 | 2101    | ±1 883             |
| I (months)                | 36     | 120    | 67      | ±11                |

VMI = average individual tree volume
VM = wood volume in plot; I = cutting age

| Variable                  | Class                                          |
|---------------------------|------------------------------------------------|
| Spacing (ES)              | 3 x 2 m; 3 x 2.5 m; 3 x 3 m                   |
| Operator experience (EO)  | Inexperienced or experienced                   |
| Management regime (RM)    | Implantation; driving; remodelling             |

Inexperienced = operators have worked <1 year in the cutting operation with harvester; experienced = operators have worked ≥1 year in the cutting operation with the harvester
the ANN and ANFIS techniques. Therefore, it is reasonable to say that the boosted technique can extract the relationship better between the input variables and the productivity of the harvester, corroborating the values of $r$ shown in Table 4.

There was no tendency to over- or underestimate the harvester’s productivity for the training and validation datasets in which the boosted technique was used, thereby proving its superiority in relation to ANN and ANFIS. When using ANN and ANFIS techniques, the points became more dispersed and tended to overestimate the initial volume classes. It is also worth mentioning that the outliers were maintained to verify which technique had the best adaptability. Notably, the boosted technique proved to be more robust in the estimation process, both in the training dataset and the validation dataset. This is an additional advantage of using this technique to estimate the harvester productivity.

The relative frequency histograms for the training and validation datasets are presented for the boosted technique, and on average about 94% of the errors are contained within an interval of ±10%. ANFIS performed worse compared with the other techniques, with a tendency to overestimate the smallest volumes. In addition, approximately 87.3% of the errors were concentrated in the interval of ±25%.

Discussion

Round timber harvesting and transport operations are the most expensive activities in the planted forest production chain, leading to the decision-making process being a complex task because it involves ecological, productive and economic factors. In addition, the impact of poor decisions can be detrimental to the ecosystem. Thus, having adequate and solid support for the decision-making process is essential.

Although the various factors that influence harvester productivity have been listed in this study, it is almost impossible to isolate or measure all biotic and abiotic components that influence the productivity of forest machines at field scale. In this sense, ML techniques were used to model the productivity of the forest harvest, since the usual modelling, as in the case of regression, is hampered by the inclusion of categorical variables, as well as by the large number of independent variables and by the existing complex relationships.

The boosted, ANN and ANFIS techniques are effective modelling approaches to estimate harvesting productivity. These techniques accurately estimated the amount in cubic meters of wood cut per hour by the harvester model considering different variables. Based on the RMSE and $r$ values (Table 4) with graphical analysis (Figures 4 and 5), it is reasonable to say that the boosted technique had the best performance, followed by ANN and ANFIS.

The productivity of forest harvesting operations is one of the main variables in the viability of wood production. Typically, it is inversely proportional to the cost per cubic meter produced and is directly influenced by land and population variables as well as by planning made by the technicians (Malinovski et al. 2014). Several authors claim that the individual volume of trees is the factor that explains most of the variations in the productivity of forest harvesting machines (Martins et al. 2009; Malinovski et al. 2014). The productivity of the machines is closely related to the size of the tree, and as the volume of the tree increases the productivity increases provided that the processing time remains the same (Akay et al. 2004).

Another important factor that influences the productivity of the harvesting machines is the spatial distance between the trees, especially during the cutting phase, as the harvesters must move between one tree and another. The greater the distance between the trees, the longer the travel time and, consequently, the lower the productivity (Malinovski et al. 2014).

The productivity of the machines depends directly on the skill of the operator, although this variable does not remain constant over time (Purrfürst and Erler 2011). The average performance of operators is extremely influential on productivity, affecting all machines with different intensities, depending on the level of machine technology (Malinovski et al. 2014). Leonello et al. (2012) evaluated the performance of harvesting operators according to their level of experience. Regarding the yield of the volume of wood harvested, they found that there was a significant increase in the first 18 months, then it remained constant until 26 months, followed by a decrease after 44 months due to daily accommodation.

The operations that make up forest harvesting involve high costs of wood production influenced by land characteristics and operational characteristics. The need to estimate productivity is crucial in the planning stage, but difficult to obtain owing to the complex relationship of the variables involved. ML techniques were able to model complex relationships with a large number of variables, including categorical variables. Importantly, these techniques would allow estimation of productivity of the harvester in places without inventory data or without planting, using values for the predictive variables within the maximum and minimum limits of the training set. This provides the possibility to minimise errors in decision-making regarding the length of the harvester’s harvest time, thus serving to develop strategic and tactical management plans. Thus, the use of

Table 4: Statistical indicators for the training and validation sets, and parameters of the settings chosen for the boosted, artificial neural network (ANN), and adaptive network-based fuzzy inference system (ANFIS) techniques

| Technique | Configuration | Training | Validation |
|-----------|---------------|----------|------------|
|           |               | RMSE (%) | B (%) | $r$ | RMSE (%) | B (%) | $r$ |
| Boosted   | Learners 11 Rate 0.3 Cycles 175 | 6.15 | 0.00 | 0.98 | 6.65 | 0.15 | 0.97 |
| ANN       | Function Tansig Algorithm LM Neurons 37 | 9.71 | 0.00 | 0.94 | 9.87 | 0.10 | 0.94 |
| ANFIS     | Cluster 18 Exponent 1415 | 12.41 | -0.02 | 0.90 | 12.33 | 0.04 | 0.90 |

RMSE = relative root-mean-square error $B$ = relative bias $r$ = correlation coefficient $LM$ = Levenberg–Marquardt
computational techniques is important to improve planning efficiency and reduce costs in forest harvesting. With the forecasting model developed for mechanised forest cutting productivity through ML, a forest manager can perform efficient harvest planning. Thus, it provides greater security for wood supply in factories over the short, medium, and long term.

**Conclusions**

ML techniques were efficient in modelling the productivity of mechanised forest cutting with the harvester model as a function of the variables studied. Modelling using the boosted technique was the most accurate in comparison to the ANN and ANFIS techniques. In addition, this methodology can be used in other areas, and low-cost variables can be added to improve the modelling performance.

**Acknowledgments** — We thank the Federal University of Espírito Santo for supporting this research, and the Coordination for the Improvement of Higher Education Personnel (CAPES) of the Brazilian Ministry of Education for financial support given to our students.

**References**

Akay AE, Erdas O, Sessions J. 2004. Determining productivity of mechanized harvesting machines. *Journal of Applied Sciences* 4: 100–105.

Akbarzadeh S, Arof AK, Ramesh S, Khamirzaei MH, Nor RM. 2014. Prediction of conductivity by adaptive neuro-fuzzy model. *PLoS One* 9: e92241.

Bezdek JC. 1981. *Pattern recognition with fuzzy objective function algorithms*. Boston, Massachusetts: Springer.

Braga AP. 2007. *Redes neurais artificiais: teoria e aplicações*. Rio de Janeiro, Brazil: Editora LTC.

Erber G, Holzleitner F, Kastner M, Stamper K. 2016. Effect of multi-tree handling and tree-size on harvester performance in small-diameter hardwood thinnings. *Silva Fennica* 50: article 1428.

---

**Figure 4:** Graphical analysis of the volume of wood harvested by the harvester, estimated by three techniques (boosted, artificial neural network [ANN], and adaptive network-based fuzzy inference system [ANFIS]) versus the observed wood volume (V), graphs of the residuals in percentage versus estimated values, and a histogram of the relative frequency of errors in percentage for the training set.
Figure 5: Graphical analysis of the volume of wood harvested by the harvester estimated by the three techniques (boosted, artificial neural network [ANN], and adaptive network-based fuzzy inference system [ANFIS]) versus the observed wood volume (V), graphs of the residuals in percentage versus estimated values, and a histogram of the relative frequency of errors in percentage for the validation set.
de um harvester trabalhando em diferentes condições de espaçamento e arranjo de plantio em povoamento de eucalipto. *Scientia Forestalis* 37: 253–263.

Mathur N, Glesk I, Buis A. 2016. Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower-limb prostheses. *Medical Engineering and Physics* 38: 1083–1089.

Norihiro J, Ackerman P, Spong BD, Längin D. 2018. Productivity model for cut-to-length harvester operation in South African eucalyptus pulpwood plantations. *Croatian Journal of Forest Engineering* 39: 1–13.

Nunes MH, Görgens EB. 2016. Artificial intelligence procedures for tree taper estimation within a complex vegetation mosaic in Brazil. *PLoS One* 11: e0154738.

Purfürst FT, Erler J. 2011. The human influence on productivity in harvester operations. *International Journal of Forest Engineering* 22: 15–22.

Reis LP, de Souza AL, dos Reis PCM, Mazzei L, Soares CPB, Torres CMME et al. 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the Amazon rainforest. *Ecological Engineering* 112: 140–147.

Rodrigues CK, da Silva Lopes E, de Oliveira D, Sampietro JA. 2018. Influência do volume das árvores no desempenho do processador florestal harvester em povoamento de eucalipto. *Biofix Scientific Journal* 3: 237–242.

Schettino S, Minette LJ, Souza AP. 2015. Correlação entre volumetria de florestas de eucalipto e produtividade e custos de máquinas de colheita de madeira. *Revista Árvore* 39: 935–942.

Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova N. 2017. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomedical Signal Processing and Control* 968: 1–14.

Silva EN, Machado CC, Minette LJ, Souza AP, Fernandes HC, Silva ML, Jacovine LA. 2011. Avaliação técnica e econômica do corte mecanizado de *Pinus* sp. com harvester. *Revista Árvore* 34: 745–753.

Silva JPM, Silva MLM, Silva EF, Silva GF, Mendonça AR, Cabacinha CD et al. 2019. Computational techniques applied to volume and biomass estimation of trees in Brazilian savanna. *Journal of Environmental Management* 249: article 109388.

Simões MG, Shaw IS. 2007. *Controle e modelagem fuzzy*. São Paulo, Brazil: Edgard Blucher.

Su S, Zhao S. 2018. An optimal clustering mechanism based on Fuzzy-C means for wireless sensor networks. *Sustainable Computing: Informatics and Systems* 18: 127–134.

Vieira GC, Mendonça AR, Silva GF, Zanetti SS, Silva MM, Santos AR. 2018. Prognoses of diameter and height of trees of *Eucalyptus* using artificial intelligence. *Science of the Total Environment* 619–620: 1473–1481.

Zamo M, Mestre O, Arbogast P, Pannekoucke O. 2014. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: deterministic forecast of hourly production. *Solar Energy* 106: 792–803.