GUILHERME SILVERIO AQUINO DE SOUZA

APRENDIZADO DE MÁQUINA EM APLICAÇÕES DE MANEJO FLORESTAL

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

Orientador: José Marinaldo Gleriani

VIÇOSA - MINAS GERAIS

2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
S729a
2019

Souza, Guilherme Silverio Aquino de, 1991-
Aprendizado de máquina em aplicações de manejo florestal / Guilherme Silverio Aquino de Souza. – Viçosa, MG, 2019.
51f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Inclui apêndice.
Orientador: José Marinaldo Gleriani.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Máquinas de vetor de suporte. 2. Redes neurais (Computação). 3. Algoritmos . I. Universidade Federal de Viçosa. Departamento de Engenharia Florestal. Programa de Pós-Graduação em Ciência Florestal. II. Título.

CDD 22 ed. 634.9905

GUILHERME SILVERIO AQUINO DE SOUZA

APRENDIZADO DE MÁQUINA EM APLICAÇÕES DE MANEJO FLORESTAL

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

APROVADA: 16 de setembro de 2019.

Assentimento:

Guilherme Silverio Aquino de Souza

Autor

José Marinaldo Gleriani

Orientador

Aos meus pais e à minha querida irmã...

À minha tia Shirle...

AGRADECIMENTOS

À minha família, por todo apoio psicológico e financeiro;
Aos meus bons amigos de Viçosa-MG, levarei a lembrança de vocês com muito carinho;
Aos colegas de laboratório;
Aos meus professores e orientadores, em especial professores: Marinaldo, Helio e Cibele, pelos grandes ensinamentos técnicos e de vida, e as pelas motivações na pesquisa;
À secretaria da pós-graduação em Ciência Florestal (UFV), Dilson e Alexandre, por serem sempre muito solícitos;
Ao programa de pós-graduação em Ciência Florestal (UFV);
À banca do presente trabalho, pelo aceite do convite para avaliação;
Às minhas treinadoras de *bike*, Bruna e Lígia, que me auxiliam na evolução de um esporte que também é o descanso da rotina da vida acadêmica.
À Carla, uma companheira cujo o carinho e lembranças desses anos de doutorado levarei com muita gratidão para toda vida... "Dançamos juntos, enquanto a música tocou"; e
A todos aqueles que me acompanharam desses anos de doutorado e contribuíram para esta pesquisa.
O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001

“O que eu não havia percebido até aquele momento [1965] era que esse tema de pesquisa haveria de se tornar uma paixão de vida. ”

(Antal Kozak, 2004, em “Minhas últimas palavras sobre equações de afilamento”)

“What I did not realize at the time [1965] was that this research topic would become a lifetime passion”

(Antal Kozak, 2004, in “My last words on taper equations”)

RESUMO

SOUZA, Guilherme Silverio Aquino de, D.Sc., Universidade Federal de Viçosa, setembro de 2019. **Aprendizado de máquina em aplicações de manejo florestal.** Orientador: José Marinaldo Gleriani.

Os algoritmos de aprendizagem de máquina (*machine learning*), constituem algumas das técnicas de inteligência artificial capazes resolver problemas mais complexos e de relações não lineares entre variáveis. Esses algoritmos vêm ganhando espaço em aplicações florestais mostrando-se eficientes em diversas aplicações florestais, retornando ganhos de precisão e redução de custos de processos em empresas. Além das ANN, que acumulam já uma considerável quantidade de estudos em aplicações florestais, alguns outros algoritmos mostraram potencial para a melhoria da precisão e acurácia de trabalhos de modelagem, tais como regressão de vetor de suporte (SVR) e o *random forest* (RF). O objetivo do presente trabalho foi de comparar o desempenho dos algoritmos citados em algumas aplicações florestais, buscando entender o comportamento das predições bem como os melhores modelos para os casos estudados. O primeiro caso, primeiro capítulo, teve o objetivo de avaliar o desempenho de índices de vegetação óticos e radarmétricos, provindos dos sensores ALOS-AVNIR-2 e ALOS-PALSAR, respectivamente, para predição do volume de plantios de eucalipto usando os três algoritmos supracitados. Cinco principais índices contribuíram, em diferentes níveis para as predições de volume: NDVI e R (índices opticos), e Pt, VSI, BMI (índices radarmétricos), provando a complementariedade da informação de ambos sensores. RF foi o algoritmo mais apropriado, com um R^2 de 0.778 e RMSE de 11.561(4.578%). No segundo capítulo, investigou-se o uso dos algoritmos para a predição de diâmetros e alocação ótima de fustes árvores de eucalipto em toras para diferentes usos, comparando-os com equações de afilamento. A equação de Kozak (1988) e as ANN apresentaram as estimativas mais acuradas e desempenho similar. RF gerou estimativas inexatas, gerando curvas de perfil de árvores na forma de "degraus". Em ambos estudos, os três algoritmos testados (ANN, SVR e RF) mostraram desempenho ou igual ou superior as abordagens convencionais. O RF se mostrou um algoritmo muito flexível para os casos de regressão, especialmente para a predição de volume por sensoriamento remoto. Entretanto os modelos gerados são limitados em predizer em uma amplitude e intervalo dado das mensurações das amostras. Para estimar o diâmetro no fuste, a não ser que mensurações sejam tomadas em intervalos menores e grandes amplitude de classes de tamanho de árvores amostras, o algoritmo RF se mostrou inapropriado. Os algoritmos SVR e ANN

preservaram a continuidade das funções, mostrando-se apropriadas para estimativas fora do intervalo de mensuração, especialmente para o caso das funções de afilamento. Entre esses dois algoritmos, a ANN se mostrou muito mais flexível para lidar com a modelagem quantitativa (regressão), especialmente quando são envolvidas variáveis categóricas com muitos fatores (estratos e classes).

Palavras-chave: Máquina de vetor de suporte. Redes Neurais (Computação). Algoritmos.

ABSTRACT

SOUZA, Guilherme Silverio Aquino de, D.Sc., Universidade Federal de Viçosa, September, 2019. **Machine Learning in Forest Management Applications**. Adviser: José Marinaldo Gleriani.

Machine learning algorithms constitute one of the techniques of artificial intelligence that can solve problems with complex data and non-linear relation between variables. This algorithm has been conquering space on forest modelling being efficient for management of planted and natural areas, gaining precision and reducing costs. Artificial neural networks already accumulate a great amount of studies on forestry. Some other algorithms has been shown potential for precision and accuracy of estimate, such as support vector regression (SVR) and random forest (RF). The main objective of this thesis was to compare ANN, SVR and RF in some forest case studies, attempting to understand behavior of predictions and best models. The first case, first chapter, aimed to assess the performance of optical and L-Band SAR vegetation indices from ALOS-AVNIR-2 and ALOSPALSAR, respectively, for eucalyptus stand volume retrieval in eastern Brazil, using three different machine-learning algorithms. Five main indices contributed, in different levels, to volume predictions of eucalyptus stands using the different machine learning algorithms: NDVI and R (optical indices), and Pt, VSI, BMI (SAR indices), proving the complementarity of both sensors information. Random Forest algorithm were the most appropriate machine-learning algorithm for data analysis yielding an R^2 value of 0,778 and RMSE of 11,561 (4,578%), outperforming ANN and SVM. In the second chapter, objective was to evaluate if machine-learning algorithms can bring improvement on diameter estimations and consequent log allocation on initial age of eucalyptus trees in Brazil. We analyzed eight taper models for ages: 40, 55 and 72 months. Variable exponent equation of Kozak (1988) and Artifitial Neural networks outperformed the comparison, showing estimated diameters statistically equal to real values. Both models produced comparable predictions. Random Forest generated misleading diameter estimations affecting optimization algorithm for log allocation. Tree profile derived from RF model presented “step way” behavior. In both studies, the three machine learning algorithms showed comparable or superior accuracy than conventional approaches. RF showed great flexibility for regression cases. However, RF models are restricted to a given range and interval of measurements. For diameter estimation, unless measures were taken in small intervals and with a wide range of size classes, RF is not appropriate. SVR and ANN preserved continuity of the predictive function, with ANN showing more

plasticity, specially when categorical variables are used with a great amount of factors (strata and classes).
Keywords: Support vector machine. Artificial neural networks. Algorithms.

SUMÁRIO

INTRODUÇÃO

A ciência florestal, assim como qualquer outra área que estuda o meio natural, aborda fenômenos complexos e relações muitas vezes não lineares entre variáveis. A abordagem tradicional da modelagem baseada em probabilidade, estatística clássica, é consagrada na literatura e em suas aplicações, por usa eficácia. Entretanto, estudos vêm mostrando precisão inferior com relação as novas técnicas de modelagem e otimização por inteligência artificial.

Os algoritmos de aprendizagem de máquina (*machine learning*) constituem algumas dessas técnicas que são capazes resolver os problemas mais complexos de modelagem, sem mesmo a pressuposição de distribuição dos dados. Esses algoritmos, em especial as redes neurais artificiais (ANN), vêm ganhando espaço em aplicações florestais mostrando-se eficientes para o manejo de florestas equianeas e inequianeas, retornando ganhos de precisão e redução de custos de projetos e empresas.

Além das ANN, que acumulam já uma considerável quantidade de estudos em aplicações florestais, alguns outros algoritmos começam a ser testados e mostram potencial para a melhoria da precisão e acurácia de trabalhos de modelagem, tais como as máquinas de vetores de suporte ou regressão de vetor de suporte (SVM ou SVR) e o algoritmo *random forest* (RF).

O objetivo do presente trabalho é comparar o desempenho dos algoritmos citados (ANN, SVR e RF) em algumas aplicações florestais, buscando entender o comportamento das predições bem como os melhores modelos para os casos estudados.

CAPÍTULO 1

OPTICAL AND SAR VEGETATION INDICES FOR EUCALYPTUS VOLUME MODELING: A MACHINE LEARNING APPROACH

1. INTRODUCTION

In 2016, Brazil had around 6 million hectares of eucalypt forestry plantations (IBÁ, 2017). These forestry plantations became eligible to participate in global initiatives to mitigate global climate change as in the REDD+ program, due to the relevant storage of carbon and by assisting in some degree against deforestation of native forests. The feasibility of these projects may be affected by efficient methods to quantify forest parameters such as biomass and timber volume. At diverse levels of eucalypt forest planning, volume is the main variable for decision-making (MACDICKEN et al., 2015). Remotely sensed data do not estimate directly the amount of volume within forest stands, but rather, quantify other features such as crown size and canopy density, which are correlated to volumetric measures (BACCINI et al., 2004).

Multispectral optical datasets can be used to estimate volume of eucalypt stands (BERRA et al., 2012), but exhibit some limitation to retrieve high biomass levels because of signal saturation (LU, 2005). Furthermore, data acquisition is highly affected by illumination and atmospheric conditions. SAR datasets are solar radiation independent and can be used to estimate forest wood volume at higher levels than multispectral optical datasets, especially from L-Band backscatter (DOBSON et al., 1992). Nevertheless, SAR datasets present high sensitivity to roughness and dielectric constants of targets (ANTROPOV et al., 2017b), which may limit accuracy of volume estimates. Optical and SAR datasets can complement each other, and working with them in synergy is a good strategy to overcome both sensor limitations (SHAO; ZHANG, 2016).

Factors such as soil, terrain and climatic conditions may also influence on volume accumulation rates, development of forest structure, and consequently how we retrieve these variables through a remote sensing approach. To reduce the effect of those factors in the forest

canopy spectral response, single-band information are generally combined into vegetation indices (SANO et al., 2005).

(VAFAEI et al., 2018) show that machine-learning algorithms can deal with nonlinear and complex data, and are more robust in comparison to conventional statistical methods of modelling to retrieve forest parameters via remotely sensed data. Therefore, the aim of this study was to assess the performance of optical and L-Band SAR indices for eucalyptus stand volume retrieval using three different machine-learning algorithms: Artificial Neural Network, Random Forest and Support Vector Machines. Specific objectives were: (a) what are the most suitable vegetation indices to retrieve wood volume of eucalypt stands? We hypothesize that combining optical and SAR indices can lead to a better accuracy; (b) what is the most appropriate machine-learning algorithm for the modeling procedure? Although some author endeavor the use of specific machine-learning algorithm for classification and regression cases, the three tested algorithm have been shown in literature a potential to deal with non-linear cases.

2. MATERIAL AND METHODS

The eucalyptus (*Eucalyptus grandis*) forest plantations of this study are located at the eastern region of Minas Gerais State. The plots were distributed over an area of 837.12 km², consisting of clonal stands with 4-9 years of age and a density of 1666 trees per hectare (3x2 meters of spacing). The average total height of trees was 25.9 meters. The dataset of this study derives from (OLIVEIRA, 2011) thesis database, where a stepwise multiple regression was used to predict different biophysical forest parameters based on ALOS satellite imagery (Fig.1).

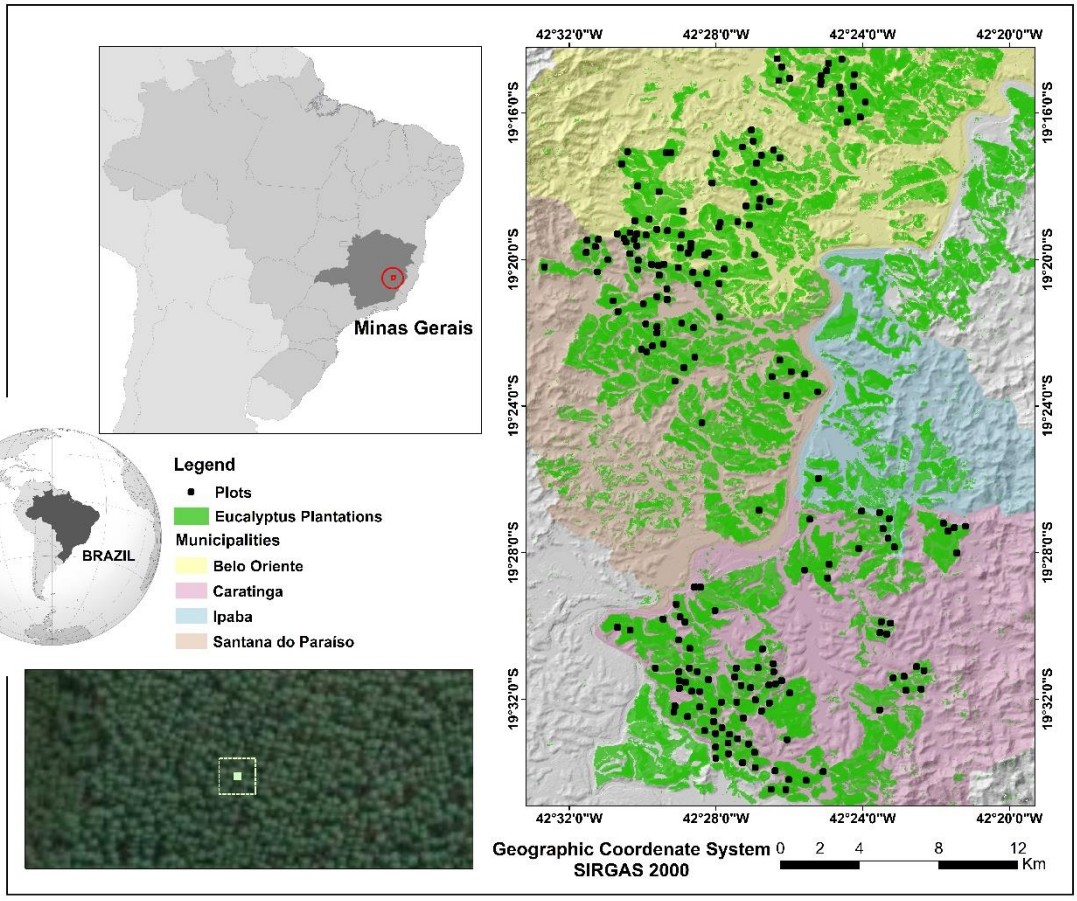

Fig 1. Location of the study area comprising inventory plots of commercial Eucalyptus (*Eucalyptus grandis*) forest plantations from eastern Minas Gerais in the southeastern Brazil. In detail, the plot mask mechanism used to extract pixel values. The map was built using software ArcGIS 10. 3 (http://ww.esri.com/software/arcgis). (SOUZA et al., 2019).

Optical and L-Band SAR backscatter data were acquired by AVNIR-2 and PALSAR sensors on board of ALOS satellite in May of 2009. Multispectral optical dataset, 10m spatial resolution, were geometrically and atmospherically corrected. Data from PALSAR, about 12 m

of spatial resolution, which comprised backscatter of four polarizations (L_{HH}, L_{HV}, L_{VV} and L_{VH}) at 21.5º of incident angle, were converted to backscatter coefficients (σ°) according to (SHIMADA et al., 2006).

Field plots comprised measurements taken from February to September 2009. Stand volume were estimated based on diameter at 1.3 meters and total height of trees within each plot. A number of 206 plots were selected for this study, with an average volume of 254 m3/ha, ages ranging from four to eight years old, mean dbh between 14.60 cm and 17.77 cm, mean total height from 22.47 to 28.23 m and stand density from 892 to 1149 trees per hectare.

The DNs and σ° were extracted using a mask with similar shape and area of field plots. Field plots covered an area of 341 m² (18,46 m x 18,46 m) and encompassed more than one pixel. Data were extracted by weighted average of pixel values in plot mask area.

The Vegetation Indices are presented as follows:

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (1)$$

$$R = \frac{NIR}{Red} \quad (2)$$

$$Rp = \frac{L_{HH}}{L_{VV}} \quad (3)$$

$$Rc = \frac{L_{HH}}{L_{HV}} \quad (4)$$

$$Pt = L_{HH} + L_{HV} + L_{VV} + L_{VH} \quad (5)$$

$$BMI = \frac{L_{HH}+L_{VV}}{2} \quad (6)$$

$$CSI = \frac{L_{VV}}{L_{VV}+L_{HV}} \quad (7)$$

$$VSI = \frac{\frac{L_{HV}+L_{VH}}{2}}{\frac{L_{HV}+L_{VH}}{2}+\frac{L_{VV}+L_{HH}}{2}} \quad (8)$$

A supervised training was employed for model development, where models were trained using known field plot cases to predict unseen data. We tested three machine learning algorithms: Artificial Neural Networks, Random Forest and Support Vector Machines.

ANN models utilizes a number of neurons in parallel to model a specific relationship and its accuracy is dependent on training dataset (HAYKIN; SIMON, 1994). 500 different MLP net architectures were trained with a range of 4 to 12 neurons in hidden layer, using Resilient Backpropagation algorithm, testing the following activation functions: logistic, Gaussian, identity, hyperbolic tangent, exponential.

Random Forest is an ensemble algorithm that works with resampling methods on training dataset (bagging or bootstrap aggregation) and trees are constructed based on a random subset

of samples of training data. Finally, a set of individually trained decision trees along the levels of response variable are combined (HASTIE et al., 2009). In this study, after primary tests, we fixed 90 trees, 60% of training dataset for resampling and 7 and 3 inputs to be drawn by node for all input models and most important input models, respectively.

Support Vector Regression work on training dataset determining an acceptance zone or margin along levels of response, restricting the flatness of this margin to be the maximum and delineating the regression based on this region. This mechanism ensures robustness to deal with unseen data. In non-linear cases, SVR employs kernel functions to map the data into a new feature space expanding the dimension of problem in an attempt to linearize the dataset (SMOLA; SCHÖLKOPF, 2004; HASTIE et al., 2009). For the present work, RBF (Radial Basis Function) was used as kernel and hyper-parameters, capacity (C) and kernel parameter (γ), were optimized via 10-cross-validation.

To ensure the representativeness of all levels on training and test data, the dataset were stratified in six classes before splitting for training. Training and test subsets corresponding to 65% and 35% from the dataset, respectively. To identify the most suitable predictors, a stepwise procedure was employed. The input relative importance was assessed by removal-based approach, i.e., running each training with all but one input. Obtained error were normalized based on RMSE, and the most important input were that which resulted in the highest error value when removed from the database (KATTENBORN et al., 2015).

We assessed the performance of models using determination coefficient (R^2) from regression of predicted and observed values; root mean squared error (RMSE); and graphical analysis.

3. RESULTS

Figure 1 shows that five main indices contributed, in different levels, to volume predictions of eucalyptus stands using the different machine learning algorithms: NDVI, BMI, Pt, R and VSI. The first index mentioned was the most important input for all models.

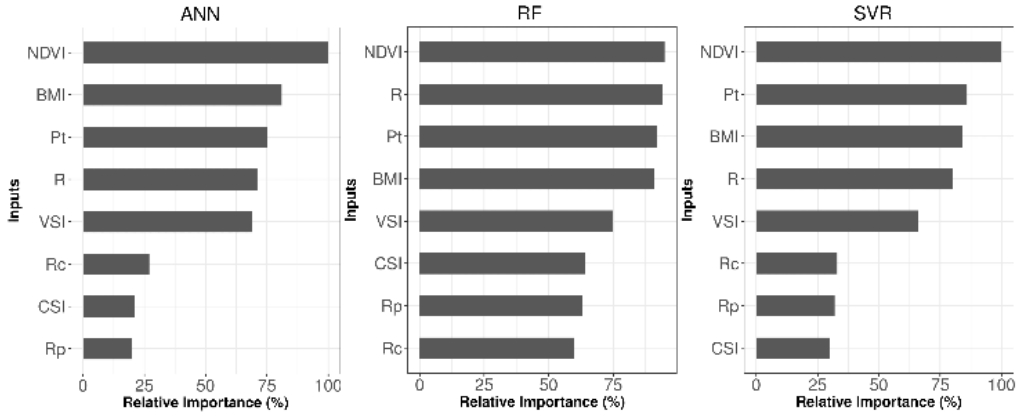

Figure 1. Relative importance of inputs for models of three different machine-learning algorithms (Artificial Neural Network, Random Forest and Support Vector Machine) for volume predictions of eucalyptus stands. Inputs comprised optical and SAR vegetation indices from ALOS/AVNIR-2 and ALOS/PALSAR sensors, respectively.

We trained models with all and the most important independent variables (inputs), in order to verify the effect of collinearity and dimensionality on results. With the most important inputs, models yielded results not significantly different from those with all indices as inputs (Table 1). Random Forest models yielded the most accurate predictions, with a substantial value of coefficient of determination (R^2 = 0,778) and RMSE of 11,561m³/ha (4,578% of the mean volume of tested stands) for testing subset of models with the most important inputs. ANN showed intermediate results, with R^2 values up to 0,658 for testing subsets. SVM yielded the less accurate volume predictions of eucalyptus stands using optical and SAR vegetation indices, with R^2 up to 0,608 and RMSE about 6,130% of generalizing data.

TABLE 1. Validation statistics of three machine-learning algorithms models (Artificial Neural Networks - ANN, Random Forest - RF and Support Vector Machines - SVM) using all vegetation indices and the most import inputs (Ntrain = 134; Ntest = 72)

Validation Statistics	ANN		RF		SVM	
	Train	Test	Train	Test	Train	Test
All Inputs						
R^2	0,739	0,658	0,906	0,777	0,688	0,608
RMSE	14,686	14,385	8,830	11,599	16,036	15,399
RMSE%	5,725	5,700	3,445	4,590	6,240	6,134
*M.I.I.**						
R^2	0,787	0,652	0,898	0,778	0,674	0,604
RMSE	13,250	14,498	9,184	11,561	16,385	15,459
RMSE%	5,158	5,758	3,584	4,578	6,357	6,132

* M.I.: Most Important Inputs (NDVI, R, Pt, BMI and VSI)

Figure 2 shows that Random Forest algorithm model yielded more satisfactory results with predictions of stand volume dispersed more closely to the line of intercept equal to zero and slope equal to one. The scatterplot results corroborate with those from Table 1 that represented mean values. ANN and SVR models showed non-biased predictions (Figure 2a, 2b, 2e and 2f). Although RF model yielded the lowest residuals, the model slightly overestimated stands with the lowest levels of volume (<220m3/ha).

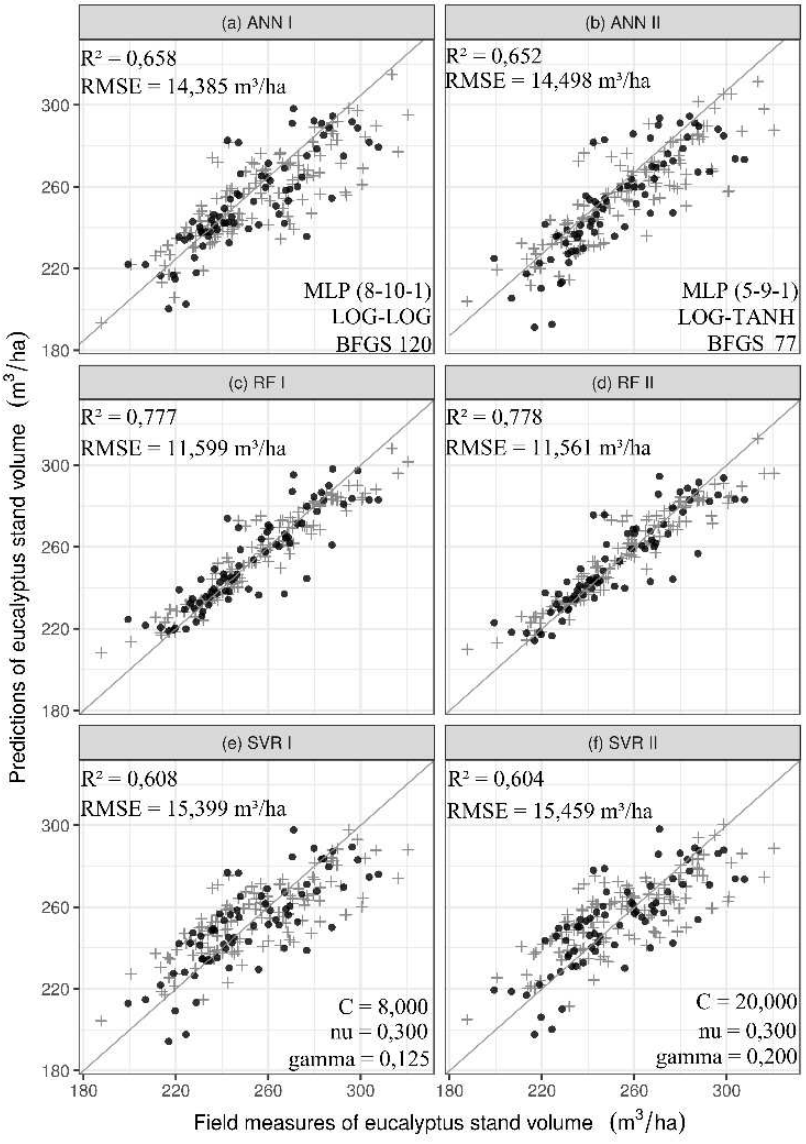

Figure 2. Scatterplot of field measures versus predicted values of volume by three machine learning algorithms (Artificial Neural Network-ANN, Random Forest-RF, and Support Vector Machines-SVM) using all inputs (I; 2a, 2c and 2e) and the most important inputs: NDVI, BMI, Pt, R and VSI (II; 2b, 2d and 2f). Validation statistics for test subset are presented. Black dots represent testing dataset. Gray marks represent training dataset.

4. DISCUSSION

The high correlation of NDVI and LAI (Leaf Are Index) can bring information of forest vertical components of stands (MAIRE, LE et al., 2011). However, from a nadir-viewing remote platform, one may find some difficulty to measure the complexity of stands using only NDVI or R, since NIR band reflectance stabilizes at values of 7 to 8 of LAI (ALLEN; RICHARDSON, 1968). The complexity of forest stand structure is the main factor making the volume estimation challenging (LU, 2005). Unlike multispectral optical data that interacts with canopy at upper portions, L-band SAR data can penetrate the canopy surface and bring information from inside components of stands, especially from vertical structures with greater amount of volume contents (WOLTER; TOWNSEND, 2011).

The high relative importance of SAR indices such as Pt, RMI and VSI in all assessed algorithms proved the complementarity of both sensors information. Overall, we therefore conclude that L-band SAR indices could provide important information that is not present in optical datasets, and their combination is valuable for accurate volume estimation.

(FASSNACHT et al., 2014; SHAO; ZHANG, 2016) state that prediction method had a substantial effect on accuracy and was generally more important than sample size. ANN and SVM showed similar results with a medium correlation and greater values of error, this result could be attributed by similar mechanisms to predict for each level of response variable. Both algorithms restrict regression complexity by minimizing error values. On the other hand, RF, via decision trees, can section the levels deliberately and averages the results of all trees (HASTIE et al., 2009). RF showed a greater potential for modelling the dataset, but one must consider grouping representative samples of levels on training subsets, primarily the extreme levels (the lowest and greatest values) for a more stable resampling. (REIS, DOS et al., 2018) assessed ANN, RF, SVR and Multilinear Regression to estimate volume using optical Landsat TM data. The authors' results corroborated with the present work, with a relative RMSE of 12,88%, 10,41% and 4,77% for ANN, SVR and RF respectively, concluding that Random Forest was the most suitable machine-learning algorithm for this modelling procedure. (SHAO; ZHANG, 2016) combined optical and SAR data to estimate biomass in Inner Mongolia by testing machine-learning algorithms, encountering the best results with Radom Forest Algorithms ($R^2 = 0,82$). (NAIDOO et al., 2014) by testing Multi-frequency SAR datasets found that RF and ANN outperformed SVR, similarly to the present work.

The results of combining optical and SAR vegetation indices and machine-learning algorithms highlighted the important contribution of new sources of space-born data and

artificial intelligence methods to forest management. Concerning the current optical and SAR datasets available, Sentinel-1 (C-band SAR) and Sentinel-2 (multispectral optical bands) show a great potential to perform a more consistent monitoring when used in combination.

5. CONCLUSION

For stand volume retrieval of eucalyptus plantation in eastern Brazil, the most suitable vegetation indices are: NDVI and R as optical indices and BMI, Pt and VSI as SAR indices, used in combination. Random Forest algorithm was the most appropriate machine-learning algorithm for the eucalyptus volume retrieval based on those remotely sensed data.

6. REFERENCES

ALLEN, W. A.; RICHARDSON, A. J. Interaction of Light with a Plant Canopy*. Journal of the Optical Society of America, v. 58, n. 8, p. 1023, 1968. **Optical Society of America**. Disponível em: <https://www.osapublishing.org/abstract.cfm?URI=josa-58-8-1023>. Acesso em: 15/10/2018.

ANTROPOV, O.; RAUSTE, Y.; HÄME, T.; PRAKS, J. Polarimetric ALOS PALSAR Time Series in Mapping Biomass of Boreal Forests. Remote Sensing, v. 9, n. 12, p. 999, 2017. **Multidisciplinary Digital Publishing Institute**. Disponível em: <http://www.mdpi.com/2072-4292/9/10/999>. Acesso em: 28/2/2018.

BACCINI, A.; FRIEDL, M. A.; WOODCOCK, C. E.; WARBINGTON, R. Forest biomass estimation over regional scales using multisource data. **Geophysical Research Letters**, v. 31, n. 10, p. n/a-n/a, 2004. Wiley-Blackwell. Disponível em: <http://doi.wiley.com/10.1029/2004GL019782>. Acesso em: 15/10/2018.

BERRA, E. F.; BRANDELERO, C.; PEREIRA, R. S.; et al. Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite Landsat. **Ciência Florestal**, v. 22, n. 4, p. 853–864, 2012. Disponível em: <http://cascavel.ufsm.br/revistas/ojs-2.2.2/index.php/cienciaflorestal/article/view/7566>. Acesso em: 27/2/2018.

DOBSON, M. C.; ULABY, F. T.; LETOAN, T.; et al. Dependence of radar backscatter on coniferous forest biomass. **IEEE Transactions on Geoscience and Remote Sensing**, v. 30, n. 2, p. 412–415, 1992. Disponível em: <http://ieeexplore.ieee.org/document/134090/>. Acesso em: 5/2/2018.

FASSNACHT, F. E.; HARTIG, F.; LATIFI, H.; et al. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. **Remote Sensing of Environment**, v. 154, p. 102–114, 2014. Elsevier. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425714003022?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&dgcid=raven_sd_recommender_email>. Acesso em: 1/3/2018.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. 2009.

HAYKIN, S. S.; SIMON. **Neural networks : a comprehensive foundation. Macmillan**, 1994.

IBÁ. **Report 2017**. 2017.

KATTENBORN, T.; MAACK, J.; FASSNACHT, F.; et al. Mapping forest biomass from space – Fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models. **International Journal of Applied Earth Observation and Geoinformation**, v. 35, p. 359–367, 2015. Elsevier. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0303243414002311>. Acesso em: 15/10/2018.

LU, D. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. **International Journal of Remote Sensing**, v. 26, n. 12, p. 2509–2525, 2005. Taylor & Francis Group. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/01431160500142145>. Acesso em: 1/3/2018.

MACDICKEN, K. G.; SOLA, P.; HALL, J. E.; et al. Global progress toward sustainable forest management. **Forest Ecology and Management**, v. 352, p. 47–56, 2015. Elsevier. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378112715000560>. Acesso em: 27/2/2018.

MAIRE, G. LE; MARSDEN, C.; NOUVELLON, Y.; et al. MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. **Remote Sensing of Environment**, v. 115, n. 10, p. 2613–2625, 2011. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0034425711002008>. Acesso em: 11/5/2017.

NAIDOO, L.; MATHIEU, R.; MAIN, R.; et al. The assessment of data mining algorithms for modelling Savannah Woody cover using multi-frequency (X-, C- and L-band) synthetic aperture radar (SAR) datasets. 2014 IEEE Geoscience and Remote Sensing Symposium. **Anais...** . p.1049–1052, 2014. IEEE. Disponível em: <http://ieeexplore.ieee.org/document/6946608/>. Acesso em: 15/10/2018.

OLIVEIRA, F. S. DE. **Uso de imagens do satélite ALOS para estimativa do volume de madeira em plantações de Eucaliptos**, 2011. Vicosa Federal University, Vicosa, Minas Gerais, Brazil. Disponível em: <http://alexandria.cpd.ufv.br:8000/teses/ciencia florestal/2011/245688f.pdf>. .

REIS, A. A. DOS; CARVALHO, M. C.; MELLO, J. M. DE; et al. Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data: an assessment of prediction methods. **New Zealand Journal of Forestry Science**, v. 48, n. 1, p. 1, 2018. Nature Publishing Group. Disponível em: <https://nzjforestryscience.springeropen.com/articles/10.1186/s40490-017-0108-0>. Acesso em: 26/2/2018.

SANO, E. E.; FERREIRA, L. G.; HUETE, A. R. Synthetic aperture radar (L band) and optical vegetation indices for discriminating the Brazilian savanna physiognomies: A comparative analysis. **Earth Interactions**, v. 9, n. 15, p. 1–15, 2005. Disponível em: <journals.ametsoc.org/doi/abs/10.1175/EI117.1>. Acesso em: 17/1/2018.

SHAO, Z.; ZHANG, L. Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China. **Sensors** (Basel, Switzerland), v. 16, n. 6, 2016. Multidisciplinary Digital Publishing Institute (MDPI). Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/27338378>. Acesso em: 11/5/2017.

SHIMADA, M.; ITOH, N.; WATANABE, M.; MORIYAMA, T.; TADONO, T. PALSAR initial calibration and validation results. (R. Meynart, S. P. Neeck, & H. Shimoda, Eds.)Proc. SPIE, v. 6361, p. 636103–636112, 2006. **International Society for Optics and Photonics**. Disponível em: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.689363>. Acesso em: 11/5/2017.

SMOLA, A J.; SCHÖLKOPF, B. **A tutorial on support vector regression. Statistics and Computing**, v. 14, p. 199–222, 2004. Disponível em: <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-4043137356&partnerID=40&rel=R8.0.0>.

VAFAEI, S.; SOOSANI, J.; ADELI, K.; et al. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). **Remote Sensing,** v. 10, n. 2, p. 172, 2018. Multidisciplinary Digital Publishing Institute. Disponível em: <http://www.mdpi.com/2072-4292/10/2/172>. Acesso em: 27/2/2018.

WOLTER, P. T.; TOWNSEND, P. A. Estimating forest species composition using a multi-sensor fusion approach. **Remote Sensing of Environment**, v. 115, p. 671–691, 2011.

CAPÍTULO 2

MACHINE LEARNING ALGORITHMS VERSUS STEM TAPER EQUATIONS: MODELING THE EFFECT OF INITIAL AGES ON TAPER AND LOG ALLOCATION OF EUCALYPT TREES IN BRAZIL

1. INTRODUCTION

Eucalypt stands are sources of wood and fiber for pulp and paper in Brazil, covering around 75% of the 7.8 million hectares of forestry plantations established all over the country (IBÁ, 2018). The planted tree industry embodied 1.1 per cent of all the Brazilian GDP in 2017, representing 6.1 per cent of industrial GDP (IBÁ, 2018). These plantations have an impressive productivity in comparison to other countries (~ 36 m³/ha.year), helping meet the high national and international demand for wood, and contributing for natural forests conversation. Most of Brazilian eucalyptus stands were usually established by using high densities of trees (3x3, 2x2m, etc.) with a short rotation periods (~7 years). The optimum exploitation of these fast-growing stands requires for a quick and accurate monitoring of volume. Studies on functional relationships between stem diameter and height are alternate approach to calculate volume of individual trees and different merchantable portions of the stem bole.

The relation of diameter and height of a tree can be expressed mathematically by taper equations. Unlike conventional volume models, that only relates a single diameter measure with and the tree height, stem taper equations can return more accurate volume predictions of a tree or stands capturing the decreasing rate of diameter measures from bottom to the top of trees. These equations has long been of interest in forest management due to the convenience of calculus of stem diameter at any arbitrary height and the calculus of tree height for any portion of stem given specification of log diameters (CAMPOS; LEITE, 2013). Knowing the diameter or volume at any part of stem is strategically reasonable to maximize the income of each tree, given the diverse set of marketplace scenarios in terms of price and demand for wood.

Taper models can be categorized into three major groups: simple mathematical models, segmented models and variable-exponent taper equations. The models of first category relates relative diameter (Y) and relative height (Z) in polynomial (Prodan 1965; Hradeztky, 1972, Kozak, 1969), sigmoid (Garay, 1979; Biging, 1984), and volume compatible (Demaerschalk72, Ormerod88) equations. Aiming more flexibility of models and to diminish bias in the base of stems some authors proposed segmented equations (Max and Burkhart ,1976; Demaerschalk and Kozak, 1977; Parresol et al., 1987). The third category comprise models that attempts to estimate diameter of trees based on different geometric forms along the bole.

A recent review of main taper equations used for managers in Brazil (ANDRADE; SCHMITT, 2017) verified that simple models have being used: Shoepfer, Hradeztky, Demaerschalk, Garay and Biging, when variable exponent models had already been proved to be more accurate (ANDRADE, 2014; SOUZA et al., 2018). Regarding only the simple models, researches underscore the great usage and accurate performances of Garay (SOUZA et al., 2016) and Hradetzky models (RIBEIRO; ANDRADE, 2016). Therefore, the present paper compares Garay and Hadeztky and we also include Biging, a sigmoid funcrion, for its simplicity and integrable porperties. The variable exponent model K88, conceived for big size trees in Canada (MUHAIRWE, 1999), outperformed predictions for eucalypt trees aged with more than 7 years, but a doubt remains on its performance for small size trees (SOUZA et al., 2018). Based on this last study and on Scolforo et al. (2018) we decided to compare variable exponent models of Kozak (1988) and Kozak (2004), given the good performances on both studies.

At the rising age of artificial intelligence (LIU et al., 2018), some machine learning algorithms have been tested to estimate diameters of trees (SCHIKOWSKI et al., 2015, 2018; MARTINS et al., 2016; NUNES; GÖRGENS, 2016). Artificial Neural Networks outperformed taper equations and other machine learning algorithms. Greater part of research focus on residual study, overlooking behavior of relative diameter estimation at the different levels of relative height. Moreover, no prior study has shown the impact of taper models on log allocation for a multiproduct management.

In this context, the present study aims to compare the best performing models on literature for eucalyptus trees to analyze behavior of different methods on diameter estimates and log allocation within small and medium sized Eucalypt trees in Brazil. We used three different categories of taper models: two categories of taper equations (simple and variable exponent models) and three machine learning algorithms. From this last, we for the first time introduce Support Vector Machines algorithm on tree taper application. Models tested were:

simple taper equations of Biging (1984), Garay (1979) and Hradeztky (1976); variable exponent equation of Kozak (1988) and Kozak (2004) (model I); Artifitial Neural Networks, Random Forest and Support Vector Machines algorithms. For log bucking optimization we implement a dynamic programming based on scenario formulated by (CAMPOS; BINOTI, 2014).

2. MATERIAL AND METHODS

2.1. Data

In present study, 158 eucalypt trees were sampled from plots of three ages of even-aged stands with density of 833 stems/ha (5x2.4 m). The sampled trees were distributed in three ages: 45 trees for 40 months; 51 trees for 55 months aged; and 62 for 72 months aged. Table 1 summarizes the statistics related to tree characteristics.

Table 1. Summary statistics for total height and dbh of Eucalypt trees used in this paper.

Stand Age	Number of trees	Total Height (m)				Diameter at 1.3m (cm)			
		mean	min	max	s.d.	mean	min	max	s.d.
Fit Data									
40 months	30	19.183	13.700	22.600	2.997	11.764	6.680	17.030	3.300
55 months	34	25.573	16.460	30.390	4.679	15.374	8.590	22.280	4.158
72 months	41	27.391	16.080	33.980	5.506	17.212	8.280	25.150	4.959
Validation Data									
40 months	15	19.318	13.900	22.950	3.290	11.731	6.680	16.550	3.368
55 months	17	24.933	15.720	30.370	5.359	15.364	8.280	22.280	4.788
72 months	21	27.994	20.000	33.860	4.993	18.060	9.870	24.830	4.729

s.d: standard deviation; min: minimum values; max.: Maximum values.

The tree dataset was divided into size classes based on diameter at breast height (DBH) for representativeness of all set within training data. Then, random selection was applied to each of size class for data splitting (training/testing dataset). For the first age (40 months) 30 trees were selected for model development and 15 for model evaluation. In the same way, for 55 and 72 months aged stands 34 and 41 trees, respectively, were selected for training the models, and 41 and 21, respectively, for testing (Figure 1). The variables total Height (H: m), diameter at breast height (DBH: cm) of each tree were measured. Diameter Outside Bark were measured at heights of 0.5, 1.0, 1.5, 2.0 m and then in intervals of 1m along the remainder portion of stem.

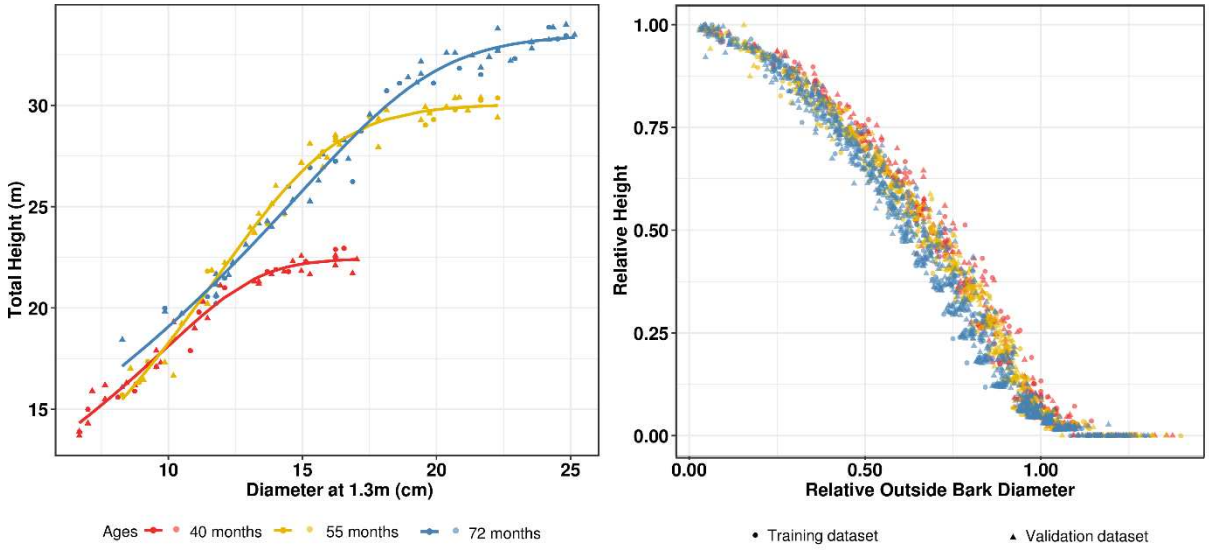

Figure 1. (a) Diameter at breast height and total height relationships and (b) relative tree diameter plotted against relative height of eucalypt trees used in this study from Bahia, Brazil.

2.2. Taper equations

Five taper equations were analyzed in this study, as follow:

$$\boldsymbol{d} = \boldsymbol{dbh}.\left\{\beta_1 + \beta_2 \ln\left[1 - (\boldsymbol{Z})^{\frac{1}{3}}.\left(1 - e^{-\frac{\beta_1}{\beta_2}}\right)\right]\right\} + \varepsilon \quad (1)$$

$$\boldsymbol{d} = \boldsymbol{dbh}.\{\beta_1.[1 + \beta_2.\ln(1 - \beta_3 \boldsymbol{Z}^{\beta_4})]\} + \varepsilon \quad (2)$$

$$\boldsymbol{d} = \boldsymbol{dbh}.\{\beta_0 + \beta_1 \boldsymbol{Z}^{p_1} + \beta_2 \boldsymbol{Z}^{p_2} + \beta_3 \boldsymbol{Z}^{p_3} + \cdots + \beta_n \boldsymbol{Z}^{p_n}\} + \varepsilon \quad (3)$$

$$\boldsymbol{d} = \alpha_0 \boldsymbol{dbh}^{\alpha_1} \alpha_2^{\boldsymbol{dbh}} \left(\frac{1-\sqrt{\boldsymbol{Z}}}{1-\sqrt{ip}}\right)^{\beta_1 Z^2 + \beta_2 \ln(Z+0{,}001) + \beta_3\sqrt{Z} + \beta_4 e^{Z} + \beta_5\left(\frac{\boldsymbol{dbh}}{\boldsymbol{H}}\right)} + \varepsilon \quad (4)$$

$$\boldsymbol{d} = \alpha_0 \boldsymbol{dbh}^{\alpha_1} \left(\frac{1-\sqrt[4]{\boldsymbol{Z}}}{1-\sqrt[4]{ip}}\right)^{\beta_0 + \beta_1\left[\frac{1}{e^{\left(\frac{\boldsymbol{dbh}}{4}\right)}}\right] + \beta_2 \boldsymbol{dbh}^{\left(\frac{1-\sqrt[4]{\boldsymbol{Z}}}{1-\sqrt[4]{ip}}\right)} + \beta_3\left(\frac{1-\sqrt[4]{\boldsymbol{Z}}}{1-\sqrt[4]{ip}}\right)^{\frac{\boldsymbol{dbh}}{\boldsymbol{H}}}} + \varepsilon \quad (5)$$

Where: Z is h/H; d is predicted diameter, p is the power fraction for Hradetzky, *ip* is the inflection point, fixed in 0.25 for Kozak(1988) and 0.10 for Kozak (2004) I; dbh is the diameter at breast height; H is total height, h height at any arbitrary point on the stem; β is the coeficients or parameters of equations, e is neperian base.

2.3. Machine Learning algorithms

Simple taper equations were first conceived to estimate relative diameter by relative height. Thus, with these same models, diameter can be estimated based on relative height and diameter at breast height. In exponent variable models, equations include dbh/H to find a diameter measure. This way, machine learning algorithms models were developed in to

scenarios, diameter estimated by two and three inputs. Further, we analyses estimation behavior on relative diameter of all developed models.

Artificial Neural Network are composed by artificial neurons referred to as simple processing unities (perceptrons), which are linked together and distributed in parallel way to undertake some task (HAYKIN, 2008; LEITE et al., 2011). We trained different architecture of multilayer perceptron ANN (MLP) with the algorithm Resilient Backpropagation, with input layer with independent variables (inputs) in each node, a hidden layer of nodes and a output layer. In the hidden layer we tested 2-10 nodes or neurons and activations function tested were: sigmoid and tangent hyperbolic. Inputs were scaled, and the number of epochs were configurated in 3000 cicles, and the minimum error (RMSE) was 0.0001.

Random Forest is an assembled algorithm that uses bagging algorithm (bootstrap aggregation), generating decision trees from a random number of inputs. The variable number of inputs makes resamples to produce a greater variety of trees (possible results). The results of each level of inputs are averaged producing a mean curve for all dataset (Hastie et al, 2009). For each scenario of fit mechanism (two and three inputs) we required the system to produce 1000 trees.

Support Vector Regresion fits a mean curve along cloud points "ignoring" discrepant data. The algorithm establish an acceptance zone, where data gain weight as far as are from this region. The boundaries of this acceptance zone, refered to as maximum margin, are established by support vectors, which is the more distant points accepted along the curve. The curve is optimized to be the more distant from support vectors (Figure 2).

This acceptance of data, ignoring of discrepant data and calculus of the curve is parameterized by the following equation:

$$Minimize \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\left(\nu\epsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$Sujeito\ \text{à} \quad \begin{cases} y_i - \boldsymbol{w}^T\Phi(x_i) - b \leq \varepsilon + \xi_i \\ \boldsymbol{w}^T\Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0,\ i = 1, \ldots, n, \epsilon \geq 0 \end{cases} \qquad (6)$$

Where C is the regulation term, w the vector of parameters associated with support vectors, b is a constant and ξ the slack variable of error out of ε precision, optimized by ν parameter. The i index labels the n cases. The term Φ(xi) represents the input transformation data by kernel K(xi,xj) at features space, from which (Xi,Xj) = Φ(xi). Φ(xj).

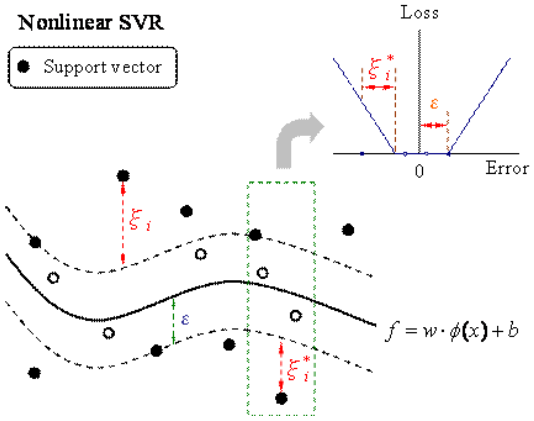

Figure 2. Maximum margin of Support Vector Regression optimized to ignore discrepant data. Fonte: YU et al. (2006). Disponível em: < http://research.ncku.edu.tw/re/articles/e/20080620/3.html >.

For non-linear cases, SVR uses kernel tricks, mapping data into new feature space in order to linearize or simplify data (SMOLA; SCHÖLKOPF, 2004). We used the nu-SVR type and Gaussian Radial Basis (RBF) as kernel function. The RBF function showed a superior performance over linear, polynomial and sigmoid function. A 10-fold cross validation was use for hyperparameters tuning (C and gamma).

2.4. Fit Statistics

For choosing the best taper equation to be compared with machine learning we first used the following fit-statistics:

$$r_{\hat{y}y} = \frac{n^{-1}[\sum_{i=1}^{n}(\hat{y}_i - \hat{y}_m)(y_i - \bar{y})]}{\sqrt{[n^{-1}\sum_{i=1}^{n}(\hat{y}_i - \hat{y}_m)^2][n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2]}} \quad (7)$$

$$Bias = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \quad (8)$$

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m}} \quad (9)$$

$$AAB = \sqrt{\frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (11)$$

Machine Learning Algorithms were compared using the correlation coefficient and RMSE (11).

We also compared error along the relative height of the best ML and taper models. In this analysis, we can verify behavior of models at butt swell of the trees.

2.5.(Comparing models and stem form by L&O test

Then identity statistical procedure proposed by Leite and Oliveira (2002), the L&O test, was used to verify final models performances (predicted x observed values) and similarity between the different models (comparison between methods). This test was also used to compare taper curve of the different ages.

The L&O test comprises three main criteria to determine the equality of two analytical methods: (a) F test proposed by Graybill (1976), (b) t test for significance of mean error (mean bias) and (c) relation of correlation coefficient ($r\hat{y}y$) and the term term (1- $|\bar{e}|$). The modified F test proposed by Graybill (1976), evaluates coefficients of a linear regression between observed and estimated values, under the following null hypothesis H0: [$\beta 0$ $\beta 1$] = [0 1]. It tests whether the intercept is statistically equal to zero and the slope to 1, under the following statistics:

$$Y_j = \beta_0 + \beta_1 Y_1 + \varepsilon \quad (12)$$

where β_0 e β_1 = linear coefficients; ε = random error with normal distribution with mean equals 0.

The F test statistics as follows:

$$F(H_0) = \frac{(\beta - \theta^T)^T (Y_1{}^T Y_1)(\beta - \theta^T)}{2.RMS} \quad (13)$$

where: $\beta = [\beta_0\ \beta_1]$; $\theta^T = [0\ 1]$.RMS = residual mean squared. If F(H0) ≥ Fα (2, n-2 d.f.) the nule hypothesis is rejected. Accepting H0 (i.e., F(H0) < Fα (2, n-2)) implies that real and predicted values are statistically identical.

The second criterion is testing if the bias or mean error ($\bar{e}$) is statistically equal to zero. A t test is applied under the hypothesis, H0: $\bar{e} = 0$, given that $t = (\bar{e} - 0)/S\bar{e}$, where $S\bar{e} = S\bar{e}/(n1/2)$, with /n-1 degrees of freedom. If $t\bar{e} \geq t\alpha(n-1)$, null hypothesis is rejected and predictions are biased. In contrast, if $t\bar{e} < t\alpha(n-1)$, the difference between observed and estimated values follows a random distribution with null mean.

In F test statistics the term “2 RMS” is a denominator. Then, a great amount of small magnitude errors affects inversely the F(H0) value, making the test very sensitive to any error of greater magnitude within dataset. In this case, the authors suggest a comparison between the correlation coefficient ($r\hat{y}y$) and the term (1- $|\bar{e}|$) (Table 1). More details about the test can be found in the authors paper Leite e Oliveira (2002).

In this study, we consider the equal performances if F(H0) test and mean error ($\bar{e} = 0$) is non-significative. When F(H0) test is significant and mean error not, RMS were considered in the analysis, and equality is stablished only if $ry\hat{y}$ were greater than (1 - $|\bar{e}|$) (case 5 in Table 3).

Table 3. Rules to validate and compare predictions of volume from eucalypt stands according to Leite and Oliveira (2002) identity test.

Case	F test (H_0)	$t_{\bar{e}}$	$R_{Y\hat{Y}}$	Decision	
1	n.s.	n.s.	$r_{y\hat{y}} \geq (1 - \lvert\bar{e}\rvert)$	Ideal	$\hat{Y} = Y$
2	n.s.	n.s.	$r_{y\hat{y}} \leq (1 - \lvert\bar{e}\rvert)$	Acceptable	$\hat{Y} = Y$
3	n.s.	*	$r_{y\hat{y}} \geq (1 - \lvert\bar{e}\rvert)$	Not valid	$\hat{Y} \neq Y$
4	n.s.	*	$r_{y\hat{y}} \leq (1 - \lvert\bar{e}\rvert)$	Not valid	$\hat{Y} \neq Y$
5	*	n.s.	$r_{y\hat{y}} \geq (1 - \lvert\bar{e}\rvert)$	Acceptable**	$\hat{Y} \neq Y$**
6	*	n.s.	$r_{y\hat{y}} \leq (1 - \lvert\bar{e}\rvert)$	Not valid	$\hat{Y} \neq Y$
7	*	*	$r_{y\hat{y}} \geq (1 - \lvert\bar{e}\rvert)$	Not valid	$\hat{Y} \neq Y$
8	*	*	$r_{y\hat{y}} \leq (1 - \lvert\bar{e}\rvert)$	Not valid	$\hat{Y} \neq Y$

n.s. = non-significant at probability level of 5%; * = significant at probability level of 5%; ** = only in cases where high correlation greatly reduces the residual variance.

2.6. Volume Predictions and Log Allocation

Some conveniences of taper modelling techniques, as already explained, involves calculating volumes of restricted new datasets, i.e., two measurements (dbh and H) of new trees, and also log bucking. Log bucing is an operation that consists of cutting trees or stems into smaller logs of predefined lengths. Log allocation optimization was undertake on R environment using dynamic programming, where recursive function was used to iterate results for each tree maximizing the income. Log grading rules were based on Campos e Binoti (2014) that characterize a typical scenario of eucalyptus forestry plantations (Table 2).

Table 4. Log grading rules for the present case based on Campos et al. (2013) work.

Log Grade	**D_{min} (cm)**	**D_{max} (cm)**	**Log length (m)**	**Price (US$)***
1. Fuelwood	4.00	40.00	2.20	187.50
2. Pulpwood	8.00	30.00	6.00	243.75
3. Wooden Beam	8.00	25.00	3.50	300.00
4. Sawnwood	15.00	50.00	3.00	562.50

*US$1.00 = R$3.75 (February/2019); D_{min}: minimum sacaling diameter ; D_{max}: maximum scaling diameter

Dynamic Programming used the recursive relation formulated as in (DYKSTRA, 1984) that define the state variable s as the total length of all logs cut from the stem through stage I (that is, including the log cut at stage i). The author defines vi(s, xi) as the value, in dollars, of the log xi associated the state s at stage i, where in the present case xi can be 2.20, 6.00, 3.50, 3.00 m. Than it follows that:

$$f_1(s, x_1) = v_1(s, x_1) \quad (14)$$

$$\text{and } f_i(s, x_i) = v_i(s, x_i) + f^*_{i-1}(s - x_i) \qquad \text{for i >1} \quad (15)$$

Equations can be combined into the following recursive relation:

$$f_i^*(s) = \max_{xi = 2.2,\ 6.0,\ 3.5,\ 3.0} \{v_i(s, x_i) + f_{i-1}^*(s - x_i)\} \quad (16)$$

with $f_0^* \equiv 0$

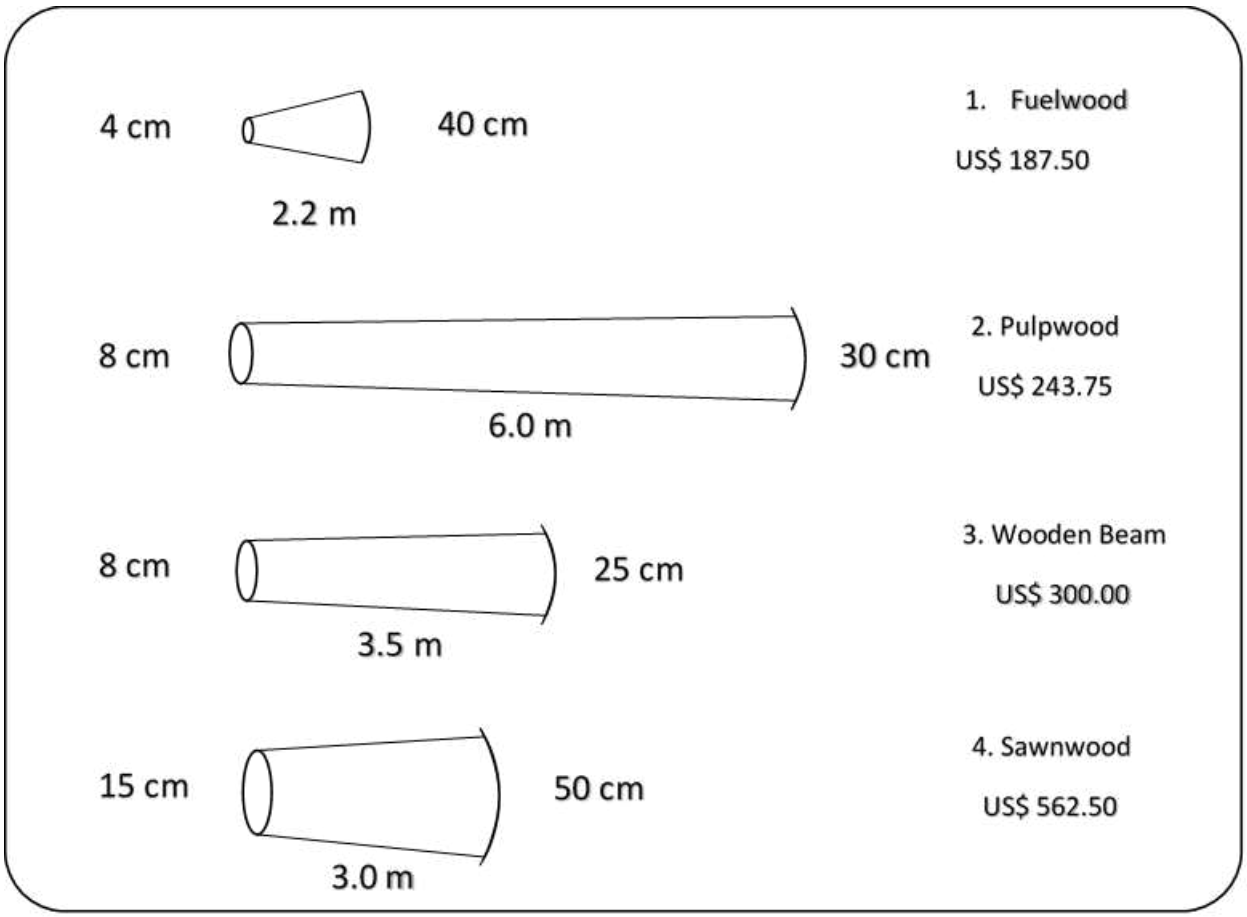

Figure 3. Log gradding rules according to Campos e Binoti (2014) *US$1.00 = R$3.75 (February/2019)

Volume of logs were calculated via estimation of diameter at a given interval of heights. Within this interval, we still estimated diameters by 10 to 10 centimeters and volume of sections was calculates via Smallian formula.

3. RESULTS

3.1. Taper equations

Overall results for taper equations shows that Kozak (1988) was the most accurate model with the highest values of $r_{y\hat{y}}$ and lowest values of error (SEE, Bias and AAB) for diameter predictions (Table 4). Concerning the simple equations, which relates relative diameter in function of relative height, Hradetzky(1972) produced most accurate results for AGE I, and Garay outperformed for AGE II and III. Regarding the variable exponents models, Kozak (2004) I model showed greater errors to estimated diameters of the three ages (Table 4).

Table 4. Fit statistics ($r_{y\hat{y}}$, SEE, Bias and AAB) of diameter inside bark for Eucalyptus Trees from Bahia State, Brazil.

Models	$r_{y\hat{y}}$	SEE	SEE (%)	Bias	AAB
Age 1					
Biging (1984)	0.9958	0.4303	4.4300	**0.0839**	0.3475
Garay (1979)	0.9963	0.4059	4.1788	0.0857	0.3225
Hradetzky (1972)	**0.9964**	**0.4049**	**4.1685**	0.0907	**0.3169**
Kozak (1988)	**0.9977***	**0.2983***	**3.0711***	**0.0100***	**0.2160***
Kozak (2004)	0.9965	0.3668	3.7763	0.0284	0.2562
Age 2					
Biging (1984)	0.9940	0.7162	6.1164	-0.3126	0.5044
Garay (1979)	**0.9946**	**0.6191**	**5.2872**	**0.0708**	**0.4416**
Hradetzky (1972)	0.9945	0.6926	5.9149	-0.3053	0.5083
Kozak (1988)	**0.9955***	**0.5583***	**4.7679***	**0.0464***	**0.3701***
Kozak (2004)	0.9946	0.6170	5.2692	0.0767	0.4457
Age 3					
Biging (1984)	0.9954	0.6634	5.2086	0.0732	0.5080
Garay (1979)	**0.9961**	**0.5910**	**4.6402**	**0.0612**	**0.4522**
Hradetzky (1972)	0.9947	0.7443	5.8438	-0.3814	0.5801
Kozak (1988)	**0.9973***	**0.4634***	**3.6383***	-0.0463	**0.3373***
Kozak (2004)	0.9968	0.4992	3.9313	**0.0027***	0.3638

Variable exponent models showed better performance for diameter estimation in relation to simple equations, with pronounced results for AGE I and AGE III. By density graph of errors (Figure 4) Kozak (1988) model showed greater part of errors at zero percentage outperforming all models at the three different ages. Therefore, we selected Kozak (1988) model for further comparison with machine learning algorithms.

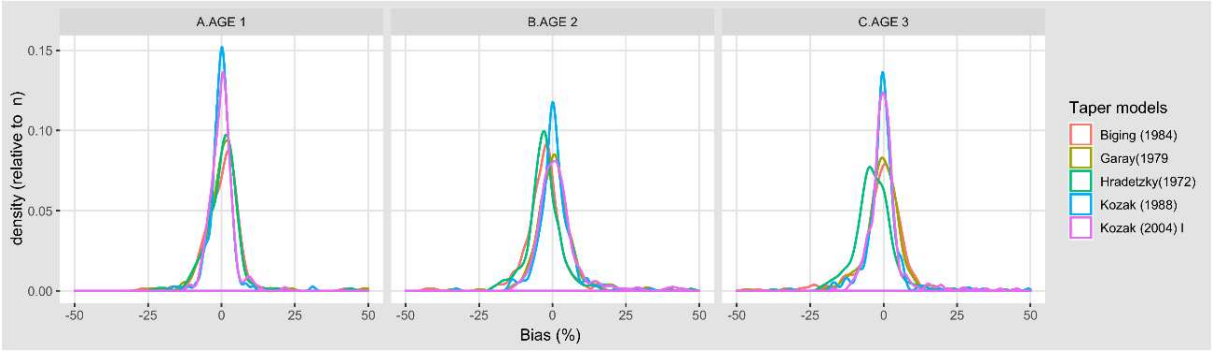

Figure 4. Percentage Bias (Bias%) distribution of eucalyptus tree diameter for testing dataset of five taper models: Biging(1984), Garay(1979), Hradetzky(1972), Kozak(1988) and Kozak (2004) I, in three ages: 40 months (n = 15), 55 months (n = 17) and 72 months (n = 21).

3.2. Machine Learning models

We tested three machine-learning algorithms under two fit mechanism: firstly, models were fit in order to predict relative diameter in function of relative height, simulating simple equations mechanism. Afterwards we fit models aiming to predict directly stem diameter in function of diameter at breast height, total height and height at any point along the stem. This last scenario simulates fit mechanism of variable exponent equations. In the first scenario, RF outperformed for AGE I, while SVR showed greater accuracy for AGE II. Both RF and SVR showed good results for AGEIII. In the second condition, ANN showed better performance for all ages and these results outperformed overall comparison.

Tabela 5. Fit statistics of diameter inside bark estimated by three different machine learning (ML) algorithms: Artificial Neural Networks (ANN), Random Forest(RF) and Support Vector Machines for Eucalyptus Trees from Bahia State, Brazil.

ML models	Age1			Age2			Age3		
	r_{yy}	RMSE	RMSE(%)	r_{yy}	RMSE	RMSE(%)	r_{yy}	RMSE	RMSE(%)
ANN I	0.9966	0.3909	4.0247	0.9947	0.6136	5.2398	0.9963	0.6136	4.8173
RF I	**0.9966**	**0.3755**	**3.8661**	0.9945	0.6104	5.2133	**0.9964**	0.6104	4.7929
SVR I	0.9963	0.4704	4.8425	**0.9947**	**0.6024**	**5.1446**	0.9961	**0.6024**	**4.7297**
ANN II	**0.9975***	**0.2980***	**3.0678***	**0.9956***	**0.5513***	**4.7081***	**0.9977***	**0.3983***	**3.4015***
RF II	0.9975	0.3651	3.7584	0.9945	0.5683	4.8533	0.9974	0.6183	4.8542
SVR II	0.9960	0.3810	3.9228	0.9930	0.5948	5.0796	0.9969	0.6948	5.4549

$r_{y\hat{y}}$: correlation between real and estimated values; RMSE: root mean squared error; I indicates estimations based on 2 inputs: Z and dbh; II indicates estimations based on 3 inputs: h, H and dap; Z is the relative height (h/H), H is total height of a tree, h is the height at any arbitrary point of stem.*Overall best performance.

Figure 5 shows percentage bias distribution of the three ML models under the second condition. ANN presented greater density of error around zero percentage of bias for ages I and III. RF yielded fit statistics close to ANN in all ages, but bias density close to zero was greater in age II. In the second condition, both ANN and RF showed good performances for estimation for test dataset.

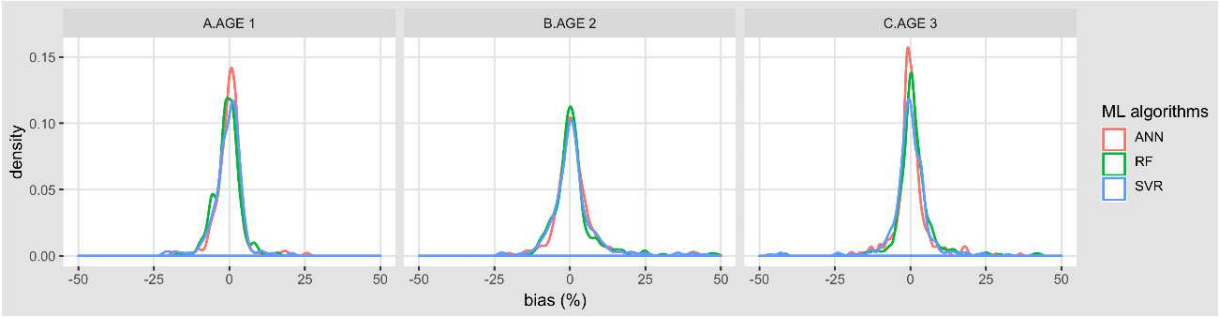

Figure 5. Percentage Bias (Bias%) distribution of eucalyptus tree diameter for testing dataset of three machine learning (ML) algorithms: Artifitial Neural Networks (ANN), Random Forest(RF) and Support Vector Machines, in three ages: 40 months (n = 15), 55 months (n = 17) and 72 months (n = 21).

3.3. Comparison of modeling approaches

We selected Age 3 to analyses behavior of different approaches to model eucalyptus tree tapering. Under the first scenario of fit mechanism, ML models yielded a single curve when relative diameter is plotted against relative height, just as Garay (1979), a simple taper equation model (Figure 6a – 6d). In this scenario, a unique variable is used as input (independent variable) and RF had a pronounced flexibility in comparison to other models. In the second, a cloud of points can observed instead of a single curve in relative diameter versus relative height plot. In this cases, at least three inputs were used: dbh, H/h and dbh/H for Kozak 88 taper equation; and dbh, H and h for ML models. The second fit mechanism produced more than one response by each level of relative height, since more than one variable was used as input. The resulted “cloud of points” produced smaller errors corroborating with performances in the previous analysis of fit statists and graphical analysis.

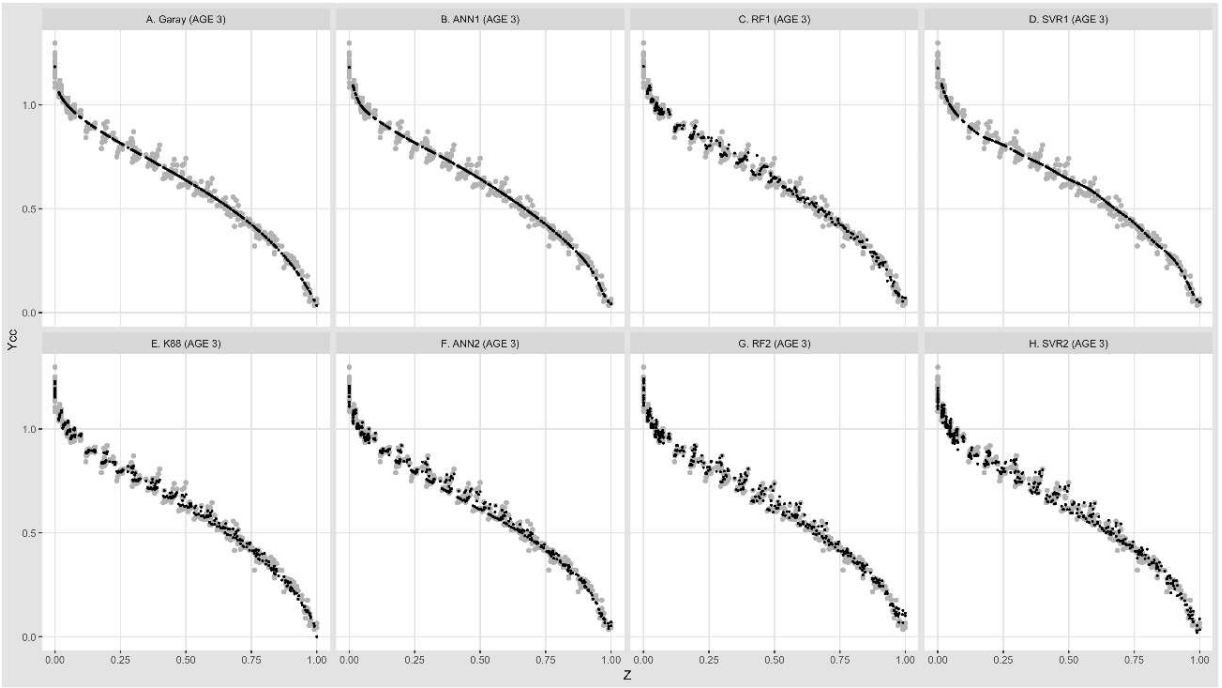

Figure 6. Behavior of different techniques for taper modeling. Figures **(a)** to **(d)** depict taper curves fit using only Z as input by, respectively, Garay (1979), ANN, RF and SVR models. Figures **(e)** to **(h)** depict estimates of relative diameter (Y) with diameter estimated using h, H and dbh as inputs in Kozak (1988), ANN, RF and SVR models, respectively. Gray dots represents real values and black dots estimates of relative diameter.

Figure 7 shows the behavior of methods when they are demanded to predict diameters for tree with a pre-determined dbh and H. We used an average tree of dataset for the analysis (DBH = 15.91 cm , H = 27.41 m). K88, ANN and RF yielded a continuous curve depicting the tree profile. However, RF created a step way curve depicting the behavior of algorithm to deal with the data. Because of step behavior and optimization of log bulking using RF model did not converge in an acceptable time (more than 24 hour), so that we decided to remove RF from analysis. In further analysis, we only consider results of Kozak (1988) model, ANN and SVR models under the second condition of fit, i.e., diameter estimated using three input (dbh, H and h).

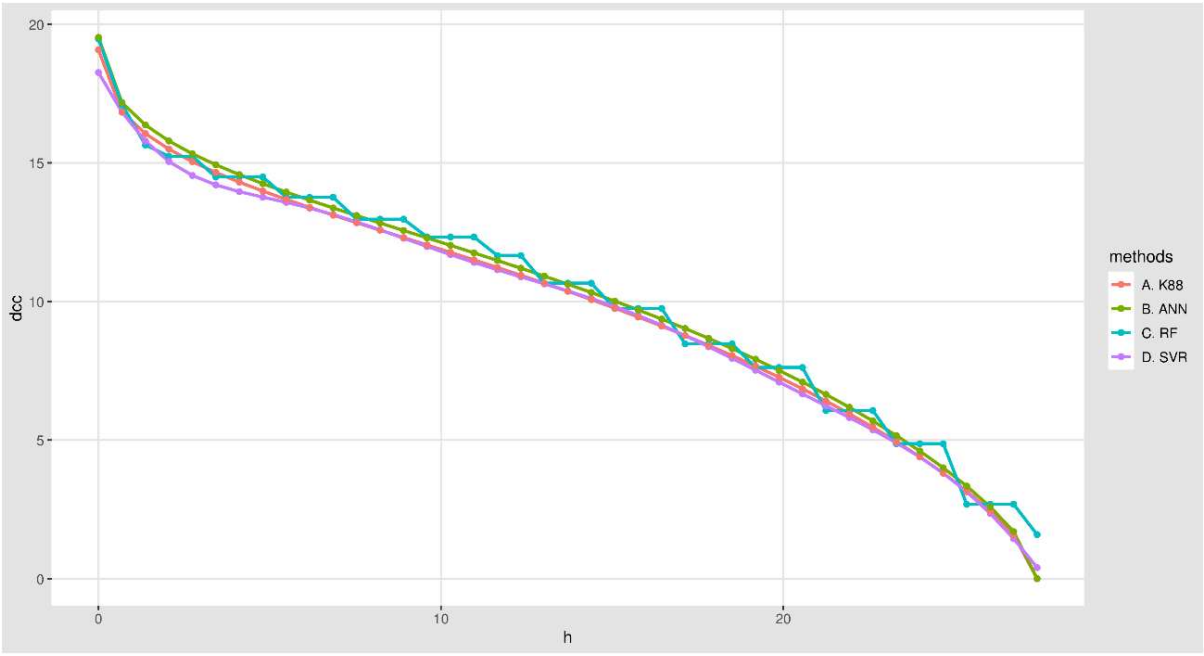

Figure 7. Tree profiles generated using K88, ANN, RF and SVR models for an eucalyptus tree with diameter at breast height 15.92 and total height of 27.41, at 72 months or 6 years aged (average tree of dataset).

We evaluated models performances along the tree bole (Tabel 6 and 7). For Age I, SVR produced biased estimates for base of trees, whereas ANN better predicted this portion. K88 showed less bias for upper portion of trees, expect for the extreme tree top class (Table 6). For Ages II and III, K88 and ANN showed similar behavior and outperformed the majority of tree portions. Concerning the RMSE values, ANN models produced the lowest values of the statistics for majority of height classes, mainly at the bottom portions in the three ages (Table 7).

Table 6. Average Biases of estimating inside bark diameters of eucalyptus trees using the Kozak (1988) (K88), Artificial Neural Networks (ANN) and Support Vector Regression (SVR).

Z	n	40 months			55 months			72 months		
		k88	ann	svr	k88	ann	svr	k88	ann	svr
10%	726	0,016	0,019	-0,066	-0,049	-0,095	-0,211	-0,126	-0,119	-0,190
20%	190	-0,036	-0,002	-0,194	0,130	0,104	-0,097	0,060	0,139	-0,107
30%	190	-0,065	0,045	-0,052	-0,131	0,052	0,086	0,000	0,097	0,101
40%	183	-0,014	0,063	0,096	0,007	0,110	0,175	-0,084	-0,104	-0,063
50%	194	0,051	0,046	0,094	0,132	0,148	0,029	0,045	-0,033	-0,083
60%	180	0,030	-0,099	-0,005	0,302	0,255	0,337	0,059	0,000	0,037
70%	186	0,028	-0,098	0,041	0,170	0,155	0,269	0,080	0,042	0,069
80%	185	0,010	0,022	0,074	0,012	0,052	0,210	0,090	0,128	0,204
90%	198	-0,055	0,145	-0,056	0,078	0,304	0,300	-0,204	0,027	0,139
100%	183	0,180	0,013	-0,072	0,119	0,104	0,097	-0,169	-0,170	-0,115

Table 7. Root mean squared error (RMSE) of estimating inside bark diameters of eucalyptus trees using the Kozak (1988) (K88), Artificial Neural Networks (ANN) and Support Vector Machines (SVM).

Z	n	40 months			55 months			72 months		
		k88	ann	svr	k88	ann	svr	k88	ann	svr
10%	726	0,368	0,341	0,511	0,731	0,684	0,900	0,514	0,431	0,609
20%	190	0,346	0,201	0,367	0,510	0,359	0,532	0,440	0,410	0,449
30%	190	0,379	0,267	0,367	0,460	0,342	0,500	0,504	0,389	0,432
40%	183	0,407	0,307	0,332	0,673	0,492	0,641	0,389	0,309	0,392
50%	194	0,408	0,301	0,254	0,569	0,438	0,402	0,636	0,427	0,443
60%	180	0,370	0,287	0,216	0,691	0,516	0,694	0,406	0,277	0,370
70%	186	0,348	0,262	0,222	0,631	0,526	0,640	0,488	0,402	0,378
80%	185	0,324	0,216	0,239	0,737	0,550	0,690	0,502	0,496	0,460
90%	198	0,306	0,338	0,269	0,676	0,684	0,703	0,627	0,506	0,421
100%	183	1,062	0,252	0,327	0,444	0,322	0,394	0,674	0,429	0,498

According to L&O test SVR models produced diameter estimates statistically different from real values ($p < 0.05$) for all the three ages. In contrast, K88 and ANN models showed results equal to observed values and performances were comparable ($p < 0.05$) (Table9).

Table 9. L&O test performance of three modeling methods – Kozak (1988), Artificial Neural Network (ANN) and Support Vector Regression (SVR).

Ages	Models	$F(H_0)$		$t(\bar{e})$		$r_{\hat{y}y} \geq 1 - \bar{e}$	RMS	n	Results	
	K88	**0,322**	ns	**1,350**	ns	**yes**	**0,086**	**184**	**$\hat{Y} = Y$**	**equal**
	ANN	**0,554**	ns	**0,824**	ns	**yes**	**0,089**	**184**	**$\hat{Y} = Y$**	**equal**
	SVR	0,840	ns	0,243	ns	no	0,145	184	$\hat{Y} \neq Y$	different
I										
	K88 x ANN	**1,862**	ns	**1,007**	ns	**yes**	**0,038**	**184**	**$\hat{Y} = Y$**	**equal**
	K88 x SVR	2,221	ns	2,716	*	yes	0,071	184	$\hat{Y} \neq Y$	different
	ANN x SVR	5,149	*	2,087	*	yes	0,063	184	$\hat{Y} \neq Y$	different
	K88	**3,938**	ns	**2,492**	ns	**yes**	**0,296**	**272**	**$\hat{Y} = Y$**	**equal**
	ANN	**7,345**	ns	**3,014**	ns	**yes**	**0,290**	**272**	**$\hat{Y} = Y$**	**equal**
	SVR	11,406	*	2,854	*	yes	0,448	272	$\hat{Y} \neq Y$	different
II										
	K88 x ANN	**3,307**	*	**0,547**	ns	**yes**	**0,059**	**272**	**$\hat{Y} = Y$**	**acceptable**
	K88 x SVR	8,280	*	1,453	ns	yes	0,208	272	$\hat{Y} \neq Y$	different
	ANN x SVR	6,847	*	1,318	ns	no	0,159	272	$\hat{Y} \neq Y$	different
	K88	**1,815**	ns	**0,526**	ns	**yes**	**0,209**	**352**	**$\hat{Y} = Y$**	**equal**
	ANN	**1,209**	ns	**0,325**	ns	**yes**	**0,172**	**352**	**$\hat{Y} = Y$**	**equal**
	SVR	7,766	*	0,134	ns	no	0,228	352	$\hat{Y} \neq Y$	different
III										
	K88 x ANN	**1,778**	ns	**0,078**	ns	**yes**	**0,061**	**352**	**$\hat{Y} = Y$**	**equal**
	K88 x SVR	8,762	*	0,256	ns	no	0,162	352	$\hat{Y} \neq Y$	different
	ANN x SVR	8,704	*	0,244	ns	no	0,111	352	$\hat{Y} \neq Y$	different

ns and * indicates de significance of L&O test rules at 95% of probability

3.4. Age comparison

We also compared the effect of age on eucalyptus tree tapering using L&O test. The taper curve of the three ages were statically different when predicted using K88 and ANN model ($p < 0.005$). Age 1 (40 months) showed a more cylindrical behavior, whereas age 3 (72 months) presented a more conical form. Age 2 taper curve showed intermediate results between ages 1 and 3.

Table 9. Identity (equality) test of taper curves from three different ages of eucalyptus trees using Kozak(1988) and Artifial Neural Network models at 95% of probability.

Ages	Models	$F(H_0)$		$t(\bar{e})$		$r_{\hat{y}y} \geq 1 - \bar{e}$	RMS	n	Result	
K88	I x II	42,921	*	4,774	*	yes	0,022	40	$\hat{Y} \neq Y$	different
	I x III	85,600	*	9,003	*	yes	0,038	40	$\hat{Y} \neq Y$	different
	II x III	57,216	*	13,345	ns	no	0,024	40	$\hat{Y} \neq Y$	different
ANN	I x II	46,386	*	4,896	*	yes	0,024	40	$\hat{Y} \neq Y$	different
	I x III	71,013	*	8,179	*	yes	0,043	40	$\hat{Y} \neq Y$	different
	II x III	40,257	*	11,365	*	yes	0,023	40	$\hat{Y} \neq Y$	different

3.5. Log allocation via Dynamic Programming

The most accurate models according to previous analysis, K88 and ANN, yielded income of US$16,169.88 and US$ 16,339.02 respectively. From the machine learning models, only ANN and SVR converged results of log bucking. RF did not converge results. SVR produced inferior income than ANN. Regarding the taper models, the simple taper equations returned higher values of income in relation to variable exponent models and ML models. Variable exponent model of K04I also produced higher values of income in comparison to K88. Therefore, we could conclude that choosing a wrong taper model one can overestimate the income of trees in the considered scenario of assortment of wood.

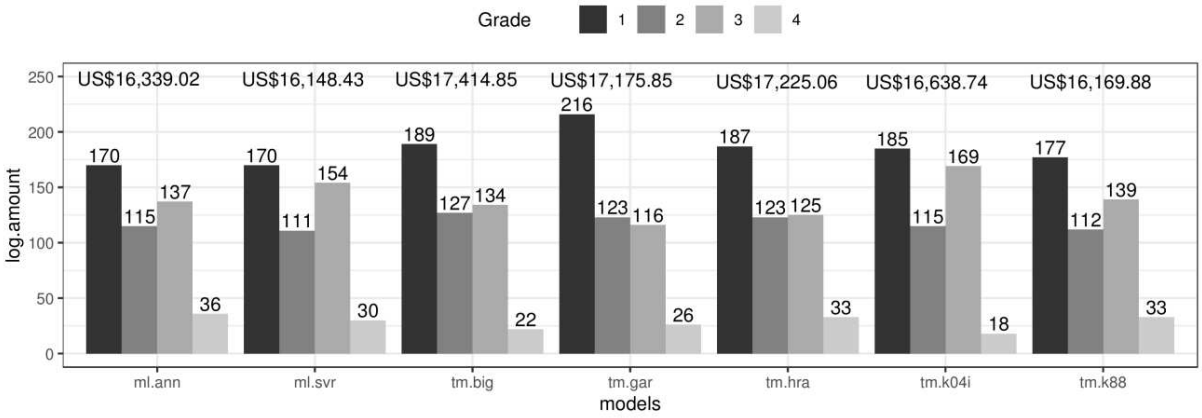

Figura 8. Log allocation of eucalytus trees at 55 months of 6 years aged in Bahia, Brazil. Income and amount of logs optimized via dynimic programming and taper models. Grades indicates log grading rules according to Campo e Binoti (2004) research. Wood destination respetively for: 1 - Fuelwood, 2 - Pulpwood, 3 - Wooden Beam and 4 - Sawnwood.

4. DISCUSSION

We compared diameter prediction and log allocation of eight taper models for eucalyptus trees at initial ages in a plantation from Bahia state in Brazil. We tested three categories of taper models: simple equations, variable exponent equations and machine learning models. Variable exponent equation of Kozak (1988) and ANN models outperformed the comparison and showed comparable behavior for tree tapering for the three initial ages. Model fit with more than two variables as inputs produced best accurate diameter estimates: dbh, Z and dbh/H for Kozak (1988) model, and dbh, h, H for ANN model. Logs were allocated with optimization based on taper models and scenario as in Campos e Binoti (2014). In this scenario, equations with worst fit statistics performances returned greater values of income in relation to the most accurate models.

Studies can use taper equations for capturing variability of all trees within a dataset at stand or holding level investigating any treatments (SOUZA et al., 2018; LEITE; SILVA, M. L M DA; et al., 2011). This convenience of equations has long been used to answer effects of any variable or condition into tapering behavior (NOGUEIRA et al., 2008). In this case, flexibility of equations is a desirable property restricted to different types of taper equations: segmented, compatible, sigmoid, etc. Machine learning algorithms are very flexible methods, but differ from equations in the fit mechanism. Some part of dataset must be separated for training the models, and at least another one for testing. Therefore, for a fair comparison between models, we used the last mechanism of fit, splitting data into training and testing data set.

4.1. Taper Models

Concerning the equations, Kozak (1988) outperformed, in line with previous studies as Andrade (2014) and Souza et al. (2018). Muhairwe (1999) ascribe accuracy of K88 model to the inclusion of dbh/H which is high correlated to live crown ratio, and the inverse of stem height, which seems to improve accuracy on the base deformation. K04I yielded less accurate estimates, notably for the second age. K88 has a fixed inflection point of 0.25, a value used for eucalyptus plantation (MUHAIRWE, 1999). According to Newnham (1992) this is the relative height where neiloidal form of the butt section changes to paraboloidal form of the stem. The inflection point for K04I is of 0.10, a value used for Canadian species dataset, what may explain inferior results of this model. Another possible explanation according to the present study is

that more than two independent variables gives the equation a one more dimension for estimating diameter.

Two well used simple equations known for their good performances were compared in this study: the sigmoid equation of Garay(1979) and fraction polynomial equation of Hradeztky (1972). Hradetzky (1972) as a polynomial model was expected to better predict diameter, but only at the fist age this model outperformed Garay(1979). The last equation showed better results for ages II and III. Garay (1979) equation relates relative diameter and relative height in a sigmoid function, and has produced the most accurate predictions of diameter for eucalyptus trees when compared with simple taper equations (SOUZA et al., 2016). One can say that the only disadvantage of Garay (1979) model is the non-integrability, but volume of any portion of the trunk can be calculated iteratively. Biging (1984) is also a sigmoid function and can be integrated, but with inferior flexibility.

4.2. Machine Learning Algorithms

ANN outperformed machine learning algorithms tested, especially in second fit condition, using three input variables. This performance could be attributed for more explanatory variables and for great flexibility of ANNs. As a high polynomial order curve, they are optimized to approximate a function to the training dataset, restricted to performances on validation dataset, avoiding overfitting.

SVR showed better results in comparison with taper equations, but for diameter predictions using h, H and dbh as inputs (second scenario) was less accurate than ANN and RF. SVR are by nature less flexible than ANN model, since the its development includes restrictions in an attempt to ignore discrepant data, rendering the algorithm a more rigid behavior. Some authors shows that SVR with Radial Basis Function has a very similar behavior of Radial ANN (HAYKIN, 2008). One of SVR models hyperparameter, the band of Gaussion Kenel Function, is optimized to deal with data under no prior assumption. However, if SVR models include a great number of treatments, i.e., classes of categorical data as inputs, optimizing the band of kernel can benefit a number of class and, in the same way, affect classes with different distributions. Therefore, managers must be careful with SVR models at a holding level involving a great number of strata or treatments.

After ANN, RF model showed the best performances for majority of cases. RF algorithm showed the greatest flexibility in comparison to all tested models when only one variable was request to predict relative diameter (first fit condition). This property can be assign

to decision trees within the algorithm that split the input levels into intervals as much as necessary to reach a minimum error in training (HASTIE et al., 2009). However, as measures were taken at least by 0.5 m of interval one only estimate value may be expect to this interval. In the RF model it represents a leaf (final branches). In this study, this property of RF predictions could be observed by the "step way" profile tree curves (Figure 7). The results of RF were disfigured for log allocation algorithm, once volume were calculated in intervals of 0.1 m. Most studies underscore the weakness of RF to forecast out of the range of value from training dataset. The present study confirmed this property with inaccurate RF predictions out of sample intervals.

4.3. Methods Comparison

We verified more than one relative diameter output (Y) for a same level of relative height (Z). Rather than a one response at each level of Z, as in simple equations, K88 relative diameter predictions exhibit a "cloud of points". Arrangement of dbh and dbh/H in equation implicitly add the information if trees are dominant or dominated, since plantations include trees with same dbh and different H or vice versa. Adding a new dimension, new variable, to taper equations increased the accuracy of estimates. This also could observed for machine learning models when more than two variables (h, H and dbh) increased the accuracy of models trained with h/H and dbh.

Beyond the statistical analysis, some operational purposes must be taken into account. K88 equation can be used for research purposed not splitting the dataset and seeking to answers any questions about one or fewer treatments. However, with more than one treatment, equality of parameters must be tested via an identity test (LEITE; OLIVEIRA, 2002). ANN models is also a very flexible method that can capture taper behavior of any treatment, and its major advantage is on operational purposes. ANN can deal with many treatments at a time, and predict for all them with a unique model (LEITE et al., 2011).

4.4. Age effect and Log allocation

The present study shows that from non-integral models (Garay, K88, K04I, ANN, SVR) one can calculate volume of new trees with dbh and H measurements or volume of different portions of trunk iteratively. Essentially, diameter can be estimated for any arbitrary height of interval of heights and volume can be calculate from sections within the interval. We suggest the calculus of sections with height intervals shorter than 10 cm at least. Similarly, height can

be estimated to a given diameter iteratively. That is the mechanism used inside the optimization algorithm to calculate volume an allocate logs.

According to Ferraz Fillho et al. (2018), increment of diameter of eucalyptus trees correspond to an average of 4.9 cm per year, with highest values for unthinned treatments. The authors consider canopy closure and consequently excessive competition between trees at age of 60 months (5 years). Taper curves of present study were significantly different showing effect of initial ages on trunk shape. The first measured age showed more cylindrical form. At the age of 72 months (6 years), eucalyptus trees showed a more conical behavior revealing increment in basal area. Major part of high dense Brazilian eucalyptus stands has rotation periods around 7 years. In this paper, we simulated a scenario of multiproduct management that eucalyptus trees may be subjected. We verified that choosing a wrong taper model, planning can be affect with overestimation of income.

5. CONCLUSION

- Variable exponent equation of Kozak (1988) and Artifitial Neural networks outperformed the comparison, showing estimated diameters equal to real values according to L&O test.
- Random Forest generated misleading diameter estimations affecting optimization algorithm for log allocation. Tree profile derived from RF model presented “step way” behavior.
- Both Kozak (1988) and ANN models showed comparable results according to L&O test.
- For operational purposes, when diameter estimations is demanded with a high amount of treatments ANN model must be preferred. In contrast, simple taper equations should be avoid, unless for didactic purposes.
- Volume calculation of new trees or logs can be calculated iteratively with any model, if computational equipment is available.
- In the considered scenario of densed Brazilian eucalyptus plantations, choosing a wrong taper model, planning can be affect with overestimation of income.
- Taper curves were different for the three ages analyzed. Trees at the first ages were more cylindrical, and trees at 72 months or 6 years of age were more conical.

6. REFERENCES

ANDRADE, V. C. L. DE. Novos modelos de *taper* do tipo expoente-forma para descrever o perfil do fuste de árvores. Pesquisa Florestal Brasileira, v. 34, n. 80, 2014. Disponível em: <http://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/614>. .

ANDRADE, V. C. L.; SCHMITT, T. Modelos de taper empregados em florestas brasileiras nativas e em plantações florestais sem eucalipto e pinus. Advances in Forestry Science, v. 69, n. 1972, p. 89–92, 2017. Disponível em: <https://www.cabi.org/ISC/FullTextPDF/2017/20173202153.pdf>. Acesso em: 21/11/2018.

CAMPOS, B. P. F.; BINOTI, D. H. B. Efeito do modelo de afilamento utilizado sobre a conversão de fustes de árvores em multiprodutos. Scientia Forestalis, v. 42, n. 104, p. 513–520, 2014. Disponível em: <http://www.ipef.br/publicacoes/scientia/nr104/cap05.pdf>. .

CAMPOS, J. C. C.; LEITE, H. G. Mensuração Florestal - Perguntas e respostas. 4th ed. Editora UFV, 2013.

DYKSTRA, D. P. Mathematical programming for natural resource management. Mathematical programming for natural resource management., 1984. McGraw-Hill Book Company. Disponível em: <https://www.cabdirect.org/cabdirect/abstract/19840693575>. Acesso em: 1/7/2019.

FILHO, A. C. F.; MOLA-YUDEGO, B.; GONZÁLEZ-OLABARRIA, J. R.; ROBERTO, J.; SCOLFORO, S. Thinning regimes and initial spacing for Eucalyptus plantations in Brazil. An Acad Bras Cienc, v. 90, n. 1, p. 255–265, 2018. Disponível em: <http://dx.doi.org/10.1590/0001-3765201720150453www.scielo.br/aabc%7Cwww.fb.com/aabcjournal>. Acesso em: 1/7/2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. 2009.

HAYKIN, S. [ebook] Neural Networks and Learning Machines. Prentice Hall/Pearson, 2008.

IBÁ. O SETOR BRASILEIRO DE ÁRVORES PLANTADAS. 2018.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. Tree-Based Methods. . p.303–335, 2013. Disponível em: <http://link.springer.com/10.1007/978-1-4614-7138-7_8>. Acesso em: 1/7/2019.

LEITE, H. G.; SILVA, M. L M DA; BINOTI, D. H. B.; FARDIN, L.; TAKIZAWA, F. H. Estimation of inside-bark diameter and heartwood diameter for Tectona grandis Linn. trees using artificial neural networks. European Journal of Forest Research, v. 130, n. 2, p. 263–269, 2011.

LEITE, H. G.; SILVA, MAYRA LUIZA MARQUES DA; BINOTI, D. H. B.; FARDIN, L.; TAKIZAWA, F. H. Estimation of inside-bark diameter and heartwood diameter for Tectona grandis Linn. trees using artificial neural networks. European Journal of Forest Research, v. 130, n. 2, p. 263–269, 2011. Springer-Verlag. Disponível em: <http://link.springer.com/10.1007/s10342-010-0427-7>. Acesso em: 26/2/2018.

LEITE, H. G.; TAVARES DE OLIVEIRA, F. H. Statistical procedure to test identity between analytical methods. Communications in Soil Science and Plant Analysis, v. 33, n. 7–8, p. 1105–1118, 2002. Disponível em: <http://www.informaworld.com/openurl?genre=article&doi=10.1081/CSS-120003875&magic=crossref%7C%7CD404A21C5BB053405B1A640AFFD44AE3>. Acesso em: 12/5/2017.

MARTINS, E. R.; BINOTI, M. L. M. S.; LEITE, H. G.; BINOTI, D. H. B.; DUTRA, G. C. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences, v. 11, n. 1, p. 33–38, 2016. Disponível em: <http://www.agraria.pro.br/ojs-2.4.6/index.php?journal=agraria&page=article&op=view&path%5B%5D=agraria_v11i1a5354>. Acesso em: 22/11/2018.

MUHAIRWE, C. K. Taper equations for Eucalyptus pilularis and Eucalyptus grandis for the north coast in New South Wales, Australia. Forest Ecology and Management, v. 113, n. 2–3, p. 251–269, 1999.

NOGUEIRA, G. S.; LEITE, H. G.; REIS, G. G.; MOREIRA, A. M. Influência do espaçamento inicial sobre a forma do fuste de árvores de Pinus taeda L. Revista Árvore, v. 32, n. 5, p. 855–860, 2008. Disponível em: <http://www.scielo.br/pdf/rarv/v32n5/10.pdf>. .

NUNES, M. H.; GÖRGENS, E. B. Artificial Intelligence Procedures for Tree Taper Estimation within a Complex Vegetation Mosaic in Brazil. (A. R. Hernandez Montoya, Ed.)PLoS ONE, v. 11, n. 5, p. e0154738, 2016. Public Library of Science. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0154738>. Acesso em: 21/11/2018.

RIBEIRO, J. R.; ANDRADE, V. C. L. DE. Equações de perfil do tronco para Eucalyptus camaldulensis dehnh no centro-sul tocantinense. Floresta e Ambiente, v. 23, n. 4, p. 534–543, 2016. Disponível em: <http://dx.>. Acesso em: 22/11/2018.

SCHIKOWSKI, A. B.; CORTE, A. P. D.; RUZA, M. S.; SANQUETTA, C. R.; MONTAÑO, R. A. N. R. Modeling of stem form and volume through machine learning. Anais da Academia Brasileira de Ciencias, v. 90, n. 4, p. 3389–3401, 2018. Disponível em: <http://dx.doi.org/10.1590/0001-3765201820170569www.scielo.br/aabc%7Cwww.fb.com/aabcjournal>. Acesso em: 30/6/2019.

SCHIKOWSKI, A. B.; CORTE, A. P. D.; SANQUETTA, C. R. Estudo da forma do fuste utilizando redes neurais artificiais e funções de afilamento. Pesquisa Florestal Brasileira Brazilian Journal of Forestry Research, v. 35, n. 82, p. 119–127, 2015. Disponível em: <http://pfb.cnpf.embrapa.br/pfb>. Acesso em: 22/11/2018.

SMOLA, A. J.; SCHÖLKOPF, B. **A tutorial on support vector regression. Statistics and Computing**, v. 14, p. 199–222, 2004. Disponível em: <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-4043137356&partnerID=40&rel=R8.0.0>.

SOUZA, G. S. A.; COSENZA, D. N.; ARAÚJO, A. C. DA S. C.; et al. Evaluation of non-linear taper equations for predicting the diameter of eucalyptus trees. Revista Árvore, v. 42, n. 1, 2018. Sociedade de Investigações Florestais. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-67622018000100201&lng=en&tlng=en>. Acesso em: 21/11/2018.

SOUZA, R. R.; NOGUEIRA, G. S.; JÚNIOR, L. S. M.; et al. forma de fuste de árvores de eucalyptus em plantios com diferentes densidades iniciais. Scientia Forestalis/Forest Sciences, v. 44, n. 109, p. 33–40, 2016. Disponível em: <http://www.ipef.br/publicacoes/scientia/nr109/cap03.pdf>. Acesso em: 22/11/2018.

7. APPENDIX

Appendix A. Estatísticas de Qualidade de Ajuste (mostrar estatística F quando precisar e teste de igualdade de parâmetros)

	α_0	α_1	α_2	β_0	β_1	β_2	β_3	β_4
Age 1								
Biging (1984)	-	-	-	-	1.1781	0.3221	-	-
Garay (1979)	-	-	-	1.207	0.246	0.983	0.261	-
Hradetzky (1972)	1.209	-0.230	-0.563	-1.267	-2.1495	0.8975	2.114	-
Kozak (1988)	0.850	1.077	0.986	-0.070	-0.018	-0.185	0.264	0.142
Kozak (2004)	1.287	0.932	-	0.324	0.204	0.006	-0.3186	-
Age 2								
Biging (1984)	-	-	-	-	1.1386	0.3329	-	-
Garay (1979)	-	-	-	1.212	0.261	0.978	0.255	-
Hradetzky (1972)	1.222	-0.709	-0.180	-0.276	-	-	-	-
Kozak (1988)	0.883	1.012	0.997	-0.028	-0.060	0.3445	0.1036	-0.023
Kozak (2004)	1.277	0.936	-	0.178	0.544	0.006	-0.388	-
Age 3								
Biging (1984)	-	-	-	-	1.1545	0.3532	-	-
Garay (1979)	-	-	-	1.183	0.287	0.966	0.273	-
Hradetzky (1972)	1.181	-0.796	-0.152	0.205	-	-	-	-
Kozak (1988)	0.989	0.934	1.000	-0.407	0.004	-0.560	0.573	-0.008
Kozak (2004)	1.2127	0.9129	-	0.4808	0.0205	0.0001	-0.1705	-

* β_n for Kozak (1988) correspond to β_{n+1}. ** Hradetzky (Y11: 1.000; 10.000; 0.060; 4.000; 0.005; 5.000/ Y12: 1.000; 0.005; 5.000/Y13: 1.000; 0.005; 10.000)

Appendix B. Residual boxplots of volume estimation of eucalypt trees from different dbh classes at ages I (40 months), age II (55 months) and age III (72months). Machine Learning Algortihms (ANN, RF and SVR) models fit using h, H and dbh as inputs and d as output.

CONCLUSÕES GERAIS

Dois estudos de casos forma abordados, o primeiro com a predição de volume de plantios de eucalipto com dados orbitais óticos e radarmétricos, e o segundo com a estimativa de diâmetro do fuste de árvore de eucalipto provindos de plantios. Os três algoritmos testados (ANN, SVR e RF) mostraram desempenho ou igual ou superior a regressão linear múltipla e regressão não-linear (abordagens convencionas).

O RF se mostrou um algoritmo muito flexível para os casos de regressão, especialmente para a predição de volume por sensoriamento remoto. Entretanto os modelos gerados são limitados a predizer em uma amplitude e intervalo dado das mensurações das amostras. Para a estimativa de diâmetro do fuste, a não ser que mensurações sejam tomadas em intervalos pequenos e grandes amplitude de classes de tamanho de árvores amostras, o algoritmo RF se mostrou inapropriado.

O algoritmo SVR configurada com a função kernel RBF, e a ANN configurada com a função de ativação sigmoide, preservaram a continuidade das funções, mostrando-se apropriadas para estimativas fora do intervalo de mensuração, especialmente para o caso das funções de afilamento. Entre esses dois algoritmos, a ANN se mostrou muito mais flexível para lidar com a modelagem quantitativa (regressão), especialmente quando são envolvidas variáveis categóricas com muitos fatores (estratos, classes, etc.)