STUDY OF THE INFLUENCE OF WOOD PROPERTIES ON THE CHARCOAL PRODUCTION: APPLYING THE RANDOM FOREST ALGORITHM

Kaléo Dias Pereira[2]* , Antônio Policarpo Souza Carneiro[3] ,Gerson Rodrigues dos Santos[3] , Angélica de Cassia Oliveira Carneiro[4] , Hélio Garcia Leite[4] and Felipe Pedersoli Borges[2]

[2] Universidade Federal de Viçosa, Programa de Pós-Gaduação em Ciência Florestal, Viçosa, MG- Brasil. E-mail: <kaleoef@gmail.com> and <pedersoli.felipe@gmail.com>.
[3] Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG- Brasil. E-mail: <policarpo@ufv.br> and <prof.gersonrodrigues @gmail.com>.
[4] Universidade Federal de Viçosa, Departamento de Engenharia Florestal, Viçosa, MG- Brasil. E-mail: <cassiacarneiro1@gmail.com> and <hgleite@gmail.com>.
*Corresponding author.

ABSTRACT – The understanding of the relationship between the properties of wood and charcoal makes it possible to improve the production of charcoal. Therefore, the *random forest* algorithm was used in this study to analyze the influence of eucalyptus wood properties on the quality of charcoal as well as the accuracy of the predicted values concerning the results estimated by support vector regression and multiple linear regression. Six properties of wood and six properties of charcoal obtained from the hybrid *Eucalyptus grandis* x *Eucalyptus urophylla* and from twelve clones of *Corymbia torelliana* x *Corymbia critriodora* at the age of seven were measured. In the analysis, the measure of mean decrease in node impurity (residual sum of squares) calculated with the *random forest* and the copula correlation was used to evaluate the relationship between properties of wood and charcoal. The *random forest* was compared to the support vector regression and multiple linear regression through the coefficient of determination, linear correlation between observed and predicted values, mean absolute error and root mean squared error. The accuracy of the *random forest* was greater than that obtained with the support vector regression and multiple linear regression, mainly in terms of the coefficient of determination and the linear correlation between observed and predicted values. The yield and quality of the charcoal produced from clones were mainly influenced by the holocellulose content, heartwood/sapwood ratio, and basic wood density. The apparent relative density of charcoal was the variable in which the *random forest* algorithm reached the best level of explanation of the variability as a function of the properties of wood, while the minor error was observed for the fixed carbon content.

Keywords: Supervised learning, *Corymbia*, *Random forest*.

ESTUDO DA INFLUÊNCIA DAS PROPRIEDADES DA MADEIRA SOBRE A PRODUÇÃO DE CARVÃO VEGETAL: APLICANDO O ALGORITMO RANDOM FOREST

*RESUMO – O entendimento das relações entre as propriedades da madeira e do carvão vegetal possibilita a melhoria da produção de carvão. Sendo assim, neste estudo utilizou-se o algoritmo **random forest** para análise da influência de propriedades da madeira de eucalipto sobre a qualidade do carvão vegetal, bem como avaliou-se a acurácia dos valores preditos em relação às estimativas do **support vector regression** e da regressão linear múltipla. Foram mensuradas seis propriedades da madeira e seis do carvão do híbrido **Eucalyptus grandis** x **Eucalyptus urophylla** e de doze clones de **Corymbia torelliana** x **Corymbia critriodora** com sete anos de idade. Nas análises, foi utilizado a medida de redução média da impureza do nó (soma de quadrados residual) do **random forest** e a correlação de cópulas para avaliar a relação entre propriedades da madeira e do carvão. O **random forest** foi comparado ao **support vector regression** e à regressão linear múltipla por meio do coeficiente de determinação, correlação linear entre valores observados e preditos, erro médio absoluto e raiz quadrada do erro quadrático médio. A acurácia do **random forest** foi maior que a do support vector regression e da*

*regressão linear múltipla, principalmente em termos de coeficiente de determinação e correlação linear entre valores observados e preditos. O rendimento e qualidade do carvão vegetal dos clones foram influenciados, principalmente, pelo teor de holocelulose, relação cerne/alburno e densidade básica da madeira. A densidade relativa aparente do carvão foi a variável em que o **random forest** atingiu o melhor nível de explicação da variabilidade em função das propriedades da madeira, enquanto para o teor de carbono fixo o algoritmo forneceu predições com menor erro.*

Palavras-Chave: aprendizagem supervisionada, ***Corymbia***, ***Random forest***.

1.INTRODUCTION

The heterogeneity of the charcoal quality is one of the main problems faced by the steel industries while controlling the reduction of iron ore process. The charcoal properties are influenced by the carbonization process and by inherent characteristics of the source material, such as species or genotype, chemical characteristics of the wood and age of the trees (Protásio et al. 2012; Soares et al. 2015).

Among the wood properties that influence the quality of charcoal, Oliveira et al. (2010) highlight the basic density, calorific value, chemical constitution and moisture content as the most important characteristics for the selection of genotypes that are suitable for the charcoal production. Soares et al. (2014) emphasize that the knowledge about the relationship between the properties of wood and those of charcoal still needs to be deepened. Furthermore, a better understanding of the chemical reactions that occur in wood during the carbonization process is required.

The relationship between the characteristics of wood and charcoal has been studied by estimating the linear correlation coefficient (Santos et al. 2011; Medeiros Neto et al. 2014; Soares et al. 2014), by simple linear regression (Brito and Barrichelo 1980; Brand et al. 2013; Santos et al. 2016) and by canonical correlation analysis (Protásio et al. 2012; Castro et al. 2013).

There is a more complex method of data analysis that associates statistical principles with computer programming and machine learning. This method allows the implementation of algorithms able to recognize patterns, learn and execute analysis in an automated way (Dantas 2017). Some methods that are already widely used, such as regression analysis and multivariate techniques, can be included in this context (Biamonte et al. 2017). The great contribution of machine learning is related to the systematic increment of programming, which allows the production of advanced and robust algorithms, like the *random forest* (RF).

The RF algorithm is based on decision/regression trees, which, according to Breiman (2001), provides a numerical estimate that is the average of all *k* trees when used for regression. In some cases, the RF proved to be more accurate than other algorithms, such as neural networks and support vector machine (Caruana et al. 2008). The performance of RF has been demonstrated in researches executed in the most diverse areas of knowledge, such as applications in remote sensing (Girolamo Neto et al. 2015), studies of soil physics (Carvalho Júnior et al. 2016), health area (Lento 2017), in the field of electromechanical energy (Lopes 2017) and many others.

In this study, the *random forest* algorithm was used to analyze the influence of some wood properties on the charcoal quality from *Eucalyptus* and *Corymbia clones*, as well as the accuracy of the predicted values concerning the results estimated by support vector regression (SVR) and multiple linear regression (MLR).

2.MATERIALS AND METHODS

The study was performed using data acquired in an experimental field area located in the city of Dionísio, Minas Gerais. This region has a humid subtropical climate, with the average annual temperature between 20 and 23 ºC and the average annual rainfall usually between 1100 and 1400 mm (Motta et al. 1996).

The planting of 13 eucalyptus genotypes, specifically one hybrid of *Eucalyptus grandis* x *Eucalyptus urophylla* and twelve of *Corymbia torelliana* x *Corymbia citriodora*, was carried out in 2008 in a land of Red-Yellow Latosol, with a plant spacing of 3.0 x 2.5 m. The plantings were seven years old when the trees were felled to collect the samples. These samples were obtained for the analysis of wood properties (heartwood/sapwood

ratio, basic wood density, higher calorific value, total lignin content, holocellulose content and extractive content) and charcoal properties (higher calorific value, gravimetric yield in carbonization, friability, apparent relative density, content of volatile materials and fixed carbon content).

2.1 Anatomical and physical properties of wood

The heartwood/sapwood ratio was measured according to Castro et al. (2013). Basic density was measured according to the standard ABNT NBR 11941 (2003) and the higher calorific value was determined according to the ABNT NBR 8633 (1984).

2.2 Analysis of the chemical composition of wood

The total extractive content was determined according to the TAPPI 204 cm-97 standard (TAPPI 1997) and the total lignin content was quantified according to Gomide and Demuner (1986) and Goldshimid (1971). The holocellulose content was calculated by deducting the percentages of total lignin and extractives from 100%.

2.3 Gravimetric yield and charcoal properties

The gravimetric yield of carbonization was determined by the relationship between the charcoal mass and the dry wood mass. Apparent relative density of the charcoal was determined by using the hydrostatic method, according to Vital (1984). Friability was obtained according to the methodology proposed by Oliveira et al. (1982). The contents of volatile materials and ashes were measured according to the NBR 8112 (ABNT 1986). Fixed carbon content was calculated by deducting the contents of volatile materials and ashes from 100% and the higher calorific value was measured according to the NBR 8633 (ABNT 1984).

2.4 Parameterized algorithms

All analysis were performed using the statistical packages of the software R. The RF regression trees were built through the *randomForest* function, which is part of the package with the same name (R Core Team 2018). For the three basic parameters of the function (Breiman 2002) – number of trees, variables randomly selected by node and number of terminal nodes – there were 500 trees, four variables per node and five terminal nodes established, respectively. In

the SVR training, the radial basis function kernel (RBF kernel) was used by means of the train function of the *caret* package (R Core Team 2018). The train function was also applied to adjust the MLR.

The wood properties were classified according to the importance of its influence on the charcoal features using the measure of mean decrease in node impurity (MDNI) of the trees in the RF algorithm, calculated by the residual sum of squares. Furthermore, the level of dependence between the properties of wood and charcoal was measured by the Spearman's rank correlation coefficient (Spearman's ρ) calculated with an elliptical copula, using the *pobs* function (R Core Team 2018) in order to convert data to an uniform distribution. This coefficient is equivalent to the Pearson's correlation coefficient (Pearson's r), with the additional feature of also measuring non-linear correlations (Ding and Li 2013). The training of trees was only repeated in function of the most important variables, aiming the measurement of the influence of these predictor values on the yield and quality of charcoal through the coefficient of determination (R^2).

To evaluate the effectiveness of RF in relation to SVR and MLR, data was randomly divided into 75% for training and 25% for validation. The methodologies were compared based on the performance outcomes on the data of validation, following the criteria used in the papers of Hallak and Pereira Filho (2011), Aitkenhead and Coull (2016), Carvalho Junior et al. (2016) and Malone et al. (2016), in which the coefficient of determination (R^2), the Pearson's correlation coefficient between observed and predicted values (ryŷ), the mean absolute error (MAE) and the root mean squared error (RMSE) were calculated. To calculate the accuracy criteria, the *postResample* function of the caret package was used (R Core Team 2018).

3.RESULTS

3.1 Wood properties vs yield and quality of charcoal

According to the MDNI measurements, it was noticed that the holocellulose content of wood was the variable that mostly influenced the yield of carbonization, volatile materials content and friability of charcoal. Heartwood/sapwood ratio was the most important variable for the fixed carbon content of
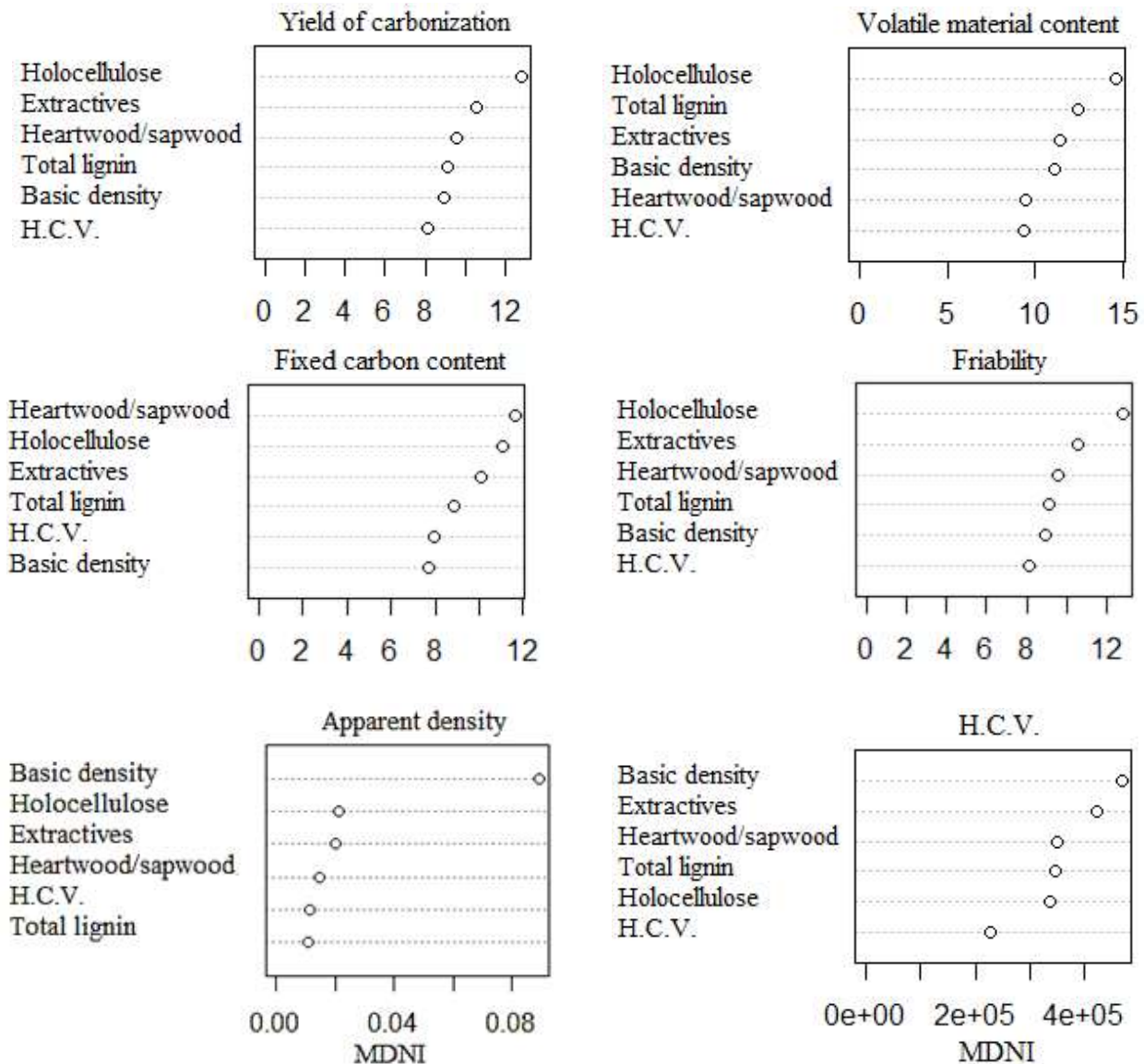
Figure 1 – Influence of wood properties on carbonization yield, volatile material content, fixed carbon content, friability, apparent relative density and higher calorific value (H.C.V.) of charcoal produced from eucalyptus clones. RMIN: mean decrease in node impurity in the *random forest* (RF) algorithm.

*Figura 1 – Influência das propriedades da madeira sobre o rendimento da carbonização, teor de materiais voláteis, teor de carbono fixo, friabilidade, densidade relativa aparente e poder calorífico superior (H.C.V.) do carvão vegetal produzido a partir de clones de Eucalyptus. RMIN: redução média na impureza do nó no algoritmo **random forest**.*

charcoal. Inferring about the apparent relative density and higher calorific value of the charcoal, basic density of wood was the property of greatest influence. The measure of the influence of wood variables on charcoal is shown by the MDNI graph, in whom, the more to the right is the point referring to the variable, the more explanatory the variable is in the algorithm (Fig. 1).

Table 1 shows that, with the exception of friability as a function of holocellulose, and apparent relative density of charcoal as a function of basic wood density, all the other relationships were inversely proportional. By analyzing the coefficients of determination (Table1), it is possible to ensure that most of the variability of the charcoal properties is explained by the most important wood variables in terms of RF modeling.

Table 1 – Spearman's rank correlation coefficient (ρ) and coefficient of determination (R²) of the *random forest* algorithm (RF) implemented only with the most influential properties for the yield and quality of charcoal produced with eucalyptus clones.
Tabela 1 – *Dependência pelo coeficiente de Spearman (ρ) e coeficiente de determinação (R²) do algoritmo **random forest** implementado apenas com as propriedades mais influentes no rendimento e qualidade do carvão vegetal produzido com clones de eucalipto.*

Dependent variable	Predictor[2]	ρ(%)	R²(%)[3]
Yield of carbonization	Holocellulose	-55.19	61.65
Volatile material content	Holocellulose	-37.90	62.84
Fixed carbon content	Heartwood/Sapwood	-12.92	61.74
Friability	Holocellulose	41.59	77.48
Apparent relative density	Basic density	84.84	85.23
Higher calorific value	Basic density	-33.07	80.45

[1]Referring to charcoal; [2]Referring to wood; [3]About validation data.
[1]Referente ao carvão vegetal; [2]Referente à madeira; [3]Considerando dados para validação.

3.2 Performance of the models

The performance of the RF algorithm regarding the measures of the coefficient of determination (R^2), linear correlation between observed and predicted values (ryŷ), mean absolute error (MAE) and root mean squared error (RMSE), was superior to the other algorithms tested for all the charcoal variables (Tables 2 and 3). It is important to point out that the results of performance that are presented in this paper refer to the data intended for validation, which, in this study, includes only 12 observations, therefore featuring a small sample. The values of R^2 and ryŷ regarding to the RF algorithm were relatively high, with most of them presenting values above or close to 90% when including all the independent variables, and always remaining above 60% in the estimates containing only the most influential independent variable (Tables 2 and 3). In both estimation cases, the r values for the RF were significant, which did not occur in the same magnitude for the SVR and MLR (Table 3).

The best results in terms of R^2 and ryŷ were obtained while estimating the apparent relative density of charcoal. On the other hand, considering the measures of MAE and RMSE (Table 4), it is noticed that the accuracy was higher while estimating

fixed carbon. The results of EMA and REQM also suggest that the differences in RF accuracy for SVR and MLR are narrow, mainly in the prediction of fixed carbon and the yield of carbonization. The different interpretations that can be extracted from R^2, ryŷ, MAE and RMSE may imply an unsureness about the more appropriate measure for the comparison of the evaluated methods. In this case, it is important to emphasize that the MAE and RMSE measured the difference between the observed and predicted values directly. Consequently, these error measures become more reliable to compare and evaluate the different prediction methods.

4. DISCUSSION

Unlike the dependence considered high (R^2 = 61.65%) that was observed for the relationship between the carbonization yield and wood holocellulose content, Santos (2008) obtained a low linear correlation (r = -0.15) between these two variables. It is consistent to what was observed in the work of Soares et al. (2014), in which the carbonization yield granted a greater dependence on the higher calorific value of the wood. In practice, the carbonization yield is usually higher when wood has a bigger lignin content, since this macromolecule is

Table 2 – Coefficient of determination (R²) for the *random forest* (RF), *support vector regression* (SVR) and multiple linear regression (MLR) considering six wood predictor variables (1st estimate) and only the most important variable (2nd estimate), using validation data.
Tabela 2 – *Coeficiente de determinação (R²) para o **random forest (RF), support vector regression (SVR)** e regressão linear múltipla (MLR) considerando as seis variáveis preditoras da madeira (1ª estimativa) e apenas a variável mais importante (2ª estimativa), utilizando os dados para validação.*

Dependent variable	RF	SVR	MLR	RF	SVR	MLR
		1st estimation (%)			2nd estimation (%)	
Yield of carbonization	87.10	49.50	55.94	61.65	29.88	49.34
Volatile material content	88.26	33.88	20.00	62.84	27.52	0.2
Fixed carbon content	89.12	0.36	0.09	61.74	22.92	13.18
Friability	89.62	42.76	30.66	77.48	9.51	4.07
Apparent relative density	93.70	64.95	68.56	85.23	70.55	59.93
Higher calorific value	90.30	34.17	4.64	80.45	16.17	24.01

Table 3 – Coefficient of linear correlation between observed and predicted values ($r_{y\hat{y}}$) for the *random forest* (RF), *support vector regression* (SVR) and multiple linear regression (MLR) considering six wood predictor variables (1st estimate) and only the most important variable (2nd estimate), for validation data.

***Tabela 3** – Coeficiente de correlação linear entre valores observados e preditos ($r_{y\hat{y}}$) pelo **random forest** (RF), **support vector regression** (SVR) e regressão linear múltipla (RLM) considerando seis variáveis preditoras da madeira (1ª estimação) e apenas a variável mais importante (2ª estimação), utilizando dados para validação.*

Dependent variable	RF	SVR	MLR	RF	SVR	MLR
	1st estimation (%)			2nd estimation (%)		
Yield of carbonization	93.65*	70.35*	74.80*	78.52*	54.67ns	0.24*
Volatile material content	93.94*	62.54*	43.58ns	79.27*	52.46ns	-4.65ns
Fixed carbon content	94.40*	5.99ns	3.01ns	78.57*	47.87ns	-36.31ns
Friability	94.67*	65.39*	55.37ns	88.02*	30.94ns	20.18ns
Apparent relative density	96.80*	80.59*	82.80*	92.32*	83.99*	77.41*
Higher calorific value	95.03*	58.46*	21.53ns	89.69*	40.83ns	49.00ns

*Significant according to the t test at 5% of significance; ns: not significant.
*Significativo de acordo com o teste t à 5% de significância; ns: não significativo.

more stable to the thermal degradation than the other existent chemical compounds (Trugilho et al. 2001; Pereira et al. 2013).

The relationship between holocellulose and volatile material content in terms of linear correlation was almost null in the study made by Soares et al. (2014), in a way that the main influence reported by the authors on the volatile materials of charcoal was the carbon-hydrogen ratio of wood. According to Oliveira et al. (2010), the volatile content also depends on the carbonization temperature and heating rate.

Brito and Barrichelo (1977) noticed that more lignified eucalyptus wood leads to a charcoal with a higher fixed carbon content, which, according to the authors, is a direct consequence of the composition of lignin, as this polymer contains approximately 65% of elemental carbon. Even though the results indicate an inverse relationship between fixed carbon and heartwood/sapwood ratio, wood that has a greater heartwood/sapwood ratio can increase the fixed carbon content in the charcoal, since heartwood tends to have a lignin content that is greater than – or even equal – to that observed on sapwood (Klitzke et al. 2008; Costa et al. 2017; Fonte et al. 2017). It is also important to emphasize that the production process is the most important factor that interferes with the fixed carbon content of charcoal (Róz et al. 2015).

Table 4 – Mean absolute error (MAE) and root mean squared error (RMSE) for the *random forest* (RF), *support vector regression* (SVR) and multiple linear regression (MLR) using six wood predictor variables (1st estimate) and only the most important variable (2nd estimate), considering validation data.

***Tabela 4** – Erro médio absoluto (EMA) e raiz quadrada do erro quadrático médio (RMSE) do **random forest** (RF), **support vector regression** (SVR) e regressão linear múltipla (RLM) utilizando as seis variáveis preditoras da madeira (1ª estimação) e apenas a mais importante (2ª estimação), considerando dados de validação.*

Dependent variable	RF	SVR	MLR	RF	SVR	MLR
	1st estimation (%)			2nd estimation (%)		
	MAE					
Yield of carbonization	0.88*	1.14	1.26	1.27	1.63	1.45
Volatile material content	1.73	2.21	3.21	2.37	2.90	3.65
Fixed carbon content	0.63	1.18	1.19	0.81	1.09	1.26
Friability	6.84	10.80	12.36	7.43	11.83	13.22
Apparent relative density	3.02	7.50	6.96	4.59	6.53	8.3
Higher calorific value	1.19	2.14	2.35	1.41	2.31	2.32
	RMSE					
Yield of carbonization	1.08	1.71	1.66	1.50	1.99	1.75
Volatile material content	2.29	3.46	3.86	2.74	3.82	4.45
Fixed carbon content	0.78	1.42	1.49	0.97	1.30	1.46
Friability	8.51	15.65	15.27	9.56	17.54	17.65
Apparent relative density	3.68	9.25	8.00	5.43	7.72	9.03
Higher calorific value	1.38	2.71	3.02	1.71	2.97	2.97

*As a percentage of the average.
*Como porcentagem da média.

Previous studies that evaluated the friability of charcoal produced from different species, correlate the highest fines generation mainly to the diameter of the logs, the carbonization time and the wood moisture (Coutinho and Ferraz 1988; Silva 1988; Pinheiro 2013). Coutinho and Ferraz (1988) state that the generation of fines is caused by the formation of internal stresses during the moisture loss process. These stresses are directly influenced by the heartwood area and by the variation of density between the pith and the bark of the wood.

The variation of the apparent density of the charcoal as a function of basic density of the wood is well known, as demonstrated by Brito and Barrichelo (1980) in a simple linear equation with R^2 of 97% for this relationship. The higher is the density of the wood, the greater is the density and mechanical resistance of the charcoal. These characteristics, with the addition of the granulometry, are the most important ones for the operation of the blast furnace in the steel industry (Brito 1993; Pereira 2012).

Although the higher calorific value of the charcoal as a function of the basic density of wood presented a high R^2 value, there is no research that demonstrates any direct relationship between these two variables. According to Couto (2014), a higher calorific value of charcoal is dependent on the elemental chemical composition of the material. An increase in the higher calorific value can be obtained by raising the carbonization temperature, which also results in a rise of the fixed carbon content caused by the elimination of volatile materials (Figueiredo et al. 2018).

Unlike what was observed in this present study, Montaño (2016) obtained better results in the prediction of the volume of *Pinus taeda*, and in the projection of the biomass content and the height of *Acacia mearnsii*, when using the SVR compared to the RF. In the study of Carvalho Junior et al. (2016), better results were obtained with the use of MLR than with RF, in order to estimate the density of the soil in function of the physicochemical properties of the samples. Rodríguez-Lado et al. (2015), on the other hand, obtained better results by applying RF to estimate soil density as a function of organic matter and texture than by applying artificial neural networks and MLR to it.

This divergence of outcomes and lack of unanimity on which technique or algorithm is more efficient considering the most diverse areas of knowledge, highlights the importance of testing different methodologies and parameter settings, so that it turns possible to define a set of procedures that is effectively more appropriate for each research field.

5. CONCLUSIONS

Using the *random forest* algorithm, the yield and quality of the charcoal produced from clones of eucalyptus were mainly influenced by the holocellulose content, heartwood/sapwood ratio and basic wood density.

The apparent relative density of the charcoal was the variable that the *random forest* algorithm reached the best level of explanation of the variability as a function of the properties of the wood, in addition to being the variable with the greatest reduction in the error of the predictions, when compared to the other two tested techniques. The fixed carbon content observed the minor error provided by the algorithm predictions.

The accuracy of the *random forest* algorithm was greater than that obtained with the support vector regression and multiple linear regression. Therefore, the estimation of the yield and quality properties of charcoal as a function of wood properties through machine learning using the *random forest* is possible to be executed. Notwithstanding, considering the error measures, the use of support vector regression and multiple linear regression also leads to accurate predictions for most of the dependent variables.

6. AUTHOR CONTRIBUTIONS

Kaléo D. Pereira: data analyze and text written, Antônio P. S. Carneiro: research supervision and text review, Gerson R. Santos: conception and data analyze, Angélica C. O. Carneiro: technical review, Hélio G. Leite: technical review and Felipe P. Borges: text review and translation.

7. ACKNOWLEDGMENTS

for the Improvement of Higher Education Personnel - CAPES for granting a postgraduate scholarship - Financing Code 001.

8.REFERENCES

ABNT. NBR 11941: Madeira: determinação da densidade básica. Rio de Janeiro: 2003.

ABNT. NBR 8112: Carvão vegetal: analise imediata. Rio de Janeiro: 1986.

ABNT. NBR 8633: Carvão vegetal: determinação do poder calorífico. Rio de Janeiro: 1984.

Aitkenhead MJ, Coull MC. Mapping soil carbon stocks across Scotland using a neural network model. Geoderma. 2016; 262:187–198. doi:10.1016/j.geoderma.2015.08.034

Biamonte JB, Wittek PW, Pancotti NP, Rebentrost PR, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017; 549:195-202. doi: 10.1038/nature23474

Brand MA, Cunha AB, Carvalho AF, Brehmer DR, Küster LC. Análise da qualidade da madeira e do carvão vegetal produzido a partir da espécie *Miconia cinnamomifolia* (De Candolle) Naudin (Jacatirão-açu) na agricultura familiar, em Biguaçu, Santa Catarina. Sci. For. 2013; 41(99): 401-410. ISSN: 2318-1222.

Breiman L. Random forests. Machine Learning. 2001; 45:5–32.

Breiman L. Manual on setting up, using, and understanding random forests V3.1. [Internet]. 2002. [cited 2018 September 26]. Available at: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf.

Brito JO. Reflexões sobre qualidade do carvão vegetal para uso siderúrgico. IPEF. 1993; 181.

Brito JO, Barrichelo LEG. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: 2. densidade da madeira x densidade do carvão. IPEF. 1980; 20: 101-113.

Brito JO, Barrichelo LEG. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: I. densidade e teor de lignina da madeira de eucalipto. IPEF. 1977; 14:9-20.

Caruana R, Karampatziakis N, Yessenalina A. An empirical evaluation of supervised learning in high dimensions. [Internet]. 2008; 25:96-103. [Cited 2018 September 12]. Available https://doi.org/10.1145/1390156.1390169

Carvalho Junior W, Calderano Filho B, Chagas CS, Bhering SB, Pereira NR, Pinheiro HSK. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. Pesquisa Agropecuária Brasileira. 2016; 51(9): 1428-1437. doi: 10.1590/S0100-204X2016000900041

Castro AFNM, Castro RVO, Carneiro, ACO, Lima JE, Santos RC, Pereira BLC, Alves ICN. Análise multivariada para seleção de clones de eucalipto destinados à produção de carvão vegetal. Pesq. agropec. bras. 2013; 48(6): p.627-635. doi: 10.1590/S0100-204X2013000600008

Costa ACS, Leal CS, Santos LC, Carvalho AMML, Oliveira AC, Pereira BLC. Propriedades da madeira de cerne e alburno de *Eucalyptus camaldulensis*. Brazilian Journal of Wood Science. 2017; 8(1): 10-20. doi: 10.15210/cmad.v8i1.8926

Coutinho AR, Ferraz ESB. Determinação da friabilidade do carvão vegetal em função do diâmetro das árvores e temperatura de carbonização. IPEF. 1988; 38: 33-37.

Couto AM. Influência das propriedades anatômicas, químicas e físicas da madeira de *Eucalyptus* e *Corymbia* na qualidade do carvão para uso siderúrgico. [Tese]. Lavras, MG. Universidade Federal de Lavras; 2014.

Dantas PB. Utilização da aprendizagem de máquina e seleção de atributos para o diagnóstico de ceratocone a partir de parâmetros biomecânicos. [Dissertação]. Maceió, AL. Universidade Federal de Alagoas; 2017.

Ding AA, LI Y. Copula correlation: An equitable dependence measure and extension of pearson's correlation. arXiv preprint arXiv:1312.7214. 2013.

Figueiredo MEO, Longue Júnior D, Pereira AKS, Carneiro ACO, Silva CMS. Potencial da madeira de *Pterogyne nitens* Tul. (madeira-nova) para produção

de carvão vegetal. Ciência Florestal. 2018; 28(1): 420-431. doi: 10.5902/1980509831620

Fonte APN, Trianoski R, Iwakiri S, Anjos RAM. Propriedades físicas e químicas da madeira de cerne e alburno de *Cryptomeria japonica*. Revista de Ciências Agroveterinárias. 2017; 16(3): 277-285. doi: 10.5965/223811711632017277

Girolamo Neto CD, Fonseca LMG, Korting TS, Sanches ID, Eberhardt IDR, Bendini HN, Marujo RFB, Tranbaquini K. Classificação automática de áreas cafeeiras utilizando imagens de sensoriamento remoto e técnicas de mineração de dados. [Internet]. 2015. [Cited 2018 August 10]. Available at: http://www.esensing.org/papers/paper_110.pdf

Goldschimid O. Ultraviolet spectra. In: Sarkanen KV, Ludwing CH, editors. Lignins. New York: Wiley Interscience; 1971. p.241-266.

Gomide JL, Demuner BJ. Determinação do teor de lignina em material lenhoso: método Klason modificado. O Papel. 1986; 47:36-38.

Hallak R, Pereira Filho AJ. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: sensibilidade a variações com os esquemas de advecção e assimilação de dados. Revista Brasileira de Meteorologia. 2011; 26(4):591-608. doi: 10.1590/S0102-77862011000400009

Klitzke RJ, Savioli DL, Muñiz GIB, Batista DC. Caracterização dos lenhos de cerne, alburno e transição de jatobá (*Hymenaea* sp.) visando ao agrupamento para fins de secagem convencional. Scientia Forestalis. 2008; 36(80): 279-284.

Lento GC. *Random forest* em dados desbalanceados: uma aplicação na modelagem de churn em seguro saúde. [Dissertação]. Rio de Janeiro, RJ. Fundação Getúlio Vargas; 2017.

Lopes TD, Goedtel A, Palácios RHC, Godoy WF. Aplicação do algoritmo random forest como classificador de padrões de falhas em rolamentos de motores de indução. [Internet]. 2017. [Cited 2018 August 2]. Available at: https://www.ufrgs.br/sbai17/papers/paper_98.pdf

Malone BP, Jha SK, Minasny B, Mcbratney AB. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. Geoderma. 2016; 262: 243–253. doi: 10.1016/j.geoderma.2015.08.037

Medeiros Neto PN, Oliveira Elisabeth, Paes JB. Relações entre as características da madeira e do carvão vegetal de duas espécies da Caatinga Floresta e Ambiente. 2014; 21(4): 484-493. doi: 10.1590/2179-8087.051313

Montaño RANR. Aplicação de técnicas de aprendizado de máquina na mensuração florestal. [Tese]. Curitiba, PR, Universidade Federal do Paraná; 2016.

Motta LP, Machado CC, Soares VP, Ribeiro JC, Ribeiro CAAS. Utilização de sistema de informações e da distância virtual na otimização do transporte florestal rodoviário. Revista Árvore. 1996; 20(3): 381-394. ISSN 0100 – 6762.

Oliveira AC, Carneiro ACO, Vital BR, Almeida W, Pereira BLC, Cardoso MT. Parâmetros de qualidade da madeira e do carvão vegetal de *Eucalyptus pellita* F. Muell. Scientia Forestalis. 2010; 38(87): 431-439.

Oliveira JB, Vivacqua Filho A, Mendes MG, Gomes PA. Produção de carvão vegetal – aspectos técnicos. In: CETEC. Produção e uso do carvão vegetal. Belo Horizonte: CETEC; 1982. p.59-72.

Pereira BLC, Carneiro ACO, Carvalho AMML, Trugilho PF, Melo ICNA, Oliveira AC. Estudo da degradação térmica da madeira de *Eucalyptus através* de termogravimetria e calorimetria. Revista Árvore. 2013; 37(3): 567-576. doi: 10.1590/S0100-67622013000300020

Pereira BLC. Qualidade da madeira de *Eucalyptus* para a produção de carvão vegetal. [Dissertação]. Viçosa, MG. Universidade Federal de Viçosa; 2012.

Pinheiro MA. Influência das dimensões da madeira na secagem e nas propriedades do carvão vegetal. [Dissertação]. Viçosa, MG. Universidade Federal de Viçosa; 2013.

Protásio TP, Trugilho PF, Neves TA, Vieira CMM. Análise de correlação canônica entre características da madeira e do carvão vegetal de *Eucalyptus*. Scientia Forestalis. 2012; 40(95): 317-326.

R Core Team. R: A language and environment for

statistical computing. Vienna: R Foundation for Statistical Computing; 2018.

Rodríguez-Lado L, Rial M, Taboada T, Cortizas AM. A pedotransfer function to map soil bulk density from limited data. Procedia Environmental Sciences. 2015; 27: 45-48. doi: 10.1016/j.proenv.2015.07.112

Róz AL, Ricardo JFC, Nakashima GT, Santos LRO, Yamaji FM. Maximização do teor de carbono fixo em biocarvão aplicado ao sequestro de carbono. R. Bras. Eng. Agríc. Ambiental. 2015; 19(8): 810–814. doi: 10.1590/1807-1929/agriambi.v19n8p810-814

Santos ID. Influência dos teores de lignina, holocelulose e extrativos na densidade básica e contração da madeira e nos rendimentos e densidade do carvão vegetal de cinco espécies lenhosas do Cerrado. [Dissertação]. Brasília, DF. Universidade de Brasília; 2008.

Santos RC, Carneiro ACO, Vital BR, Castro RVO, Vidaurre GB, Trugilho PF, Castro AFNM. Influência das propriedades químicas e da relação siringil/ guaiacil da madeira de eucalipto na produção de carvão vegetal. Ciênc. Florest. 2016; 26(2). doi: 10.5902/1980509822765

Santos RC, Carneiro ACO, Castro AFM, Castro RVO, Bianche JJ, Souza MM, Cardoso MT.

Correlações entre os parâmetros de qualidade da madeira e do carvão vegetal de clones de eucalipto. Sci. For. 2011; 39(90): 221-230.

Silva DA. Qualidade do carvão vegetal produzido com madeiras da região de Manaus em fornos de alvenaria. Acta Amazonica. 1988; 18(1-2): 163-178.

Soares VC, Bianchi ML, Trugilho PF, Höfler J, Pereira AJ. Análise das propriedades da madeira e do carvão vegetal de híbridos de eucalipto em três idades. Cerne. 2015; 21(2): 191-197. doi: 10.1590/01047760201521021294

Soares VC, Bianchi ML, Trugilho PF, Pereira AJ, Höfler J. Correlações entre as propriedades da madeira e do carvão vegetal de híbridos de eucalipto. Revista Árvore. 2014; 38(3): 543-549. doi: 10.1590/ S0100-67622014000300017

TAPPI T 264 cm-97. Preparation of wood for chemical analysis. Atlanta: Tappi Press; 1997.

Trugilho PF, Lima JT, Mori FA, Lino ALL. Avaliação de clones de *Eucalyptus* para produção de carvão vegetal. Cerne. 2001; 7(2): 104-114.

Vital BR. Métodos de determinação da densidade da madeira. Viçosa: Sociedade de Investigações Florestais; 1984.