UFSM

Ciência Florestal

ISSN 1980-5098
OPEN ACCESS

Ci. Fl., Santa Maria, v. 36, e95533, p. 1-19, 2026 • https://doi.org/10.5902/1980509895533
Submitted: 2nd/03/2026 • Approved: 18th/06/2026 • Published: 25th/09/2026

Articles

Prediction of kraft pulp industrial production with generalized additive models and artificial neural networks

Predição de produção industrial de celulose kraft com modelos aditivos generalizados e redes neurais artificiais

Fabiano da Rocha Stein[I], Gustavo Eduardo Marcatti[II], Marcelo Moreira da Costa[I], Cassiano Rodrigues de Oliveira[III], Hélio Garcia Leite[I]

[I]Universidade Federal de Viçosa, Viçosa, MG, Brazil
[II]Universidade Federal de São João del Rei, Sete Lagoas, MG, Brazil
[III]Universidade Federal de Viçosa, Rio Paranaíba, MG, Brazil

ABSTRACT

This work describes the modelling of observed data from kraft pulping variables at industrial scale for several years by generalized additive models and artificial neural networks. Generalized additive models were adjusted with R software, using database "mgcv". Statistical significance tests were applied to each model. In a complementary way, data were analyzed by 100 artificial neural networks in order to relate digester production and alkaline load with several process and pulp variables. At this stage, Statistica software was applied to describe data behaviour. Generalized additive models proved to be a good option to represent the variables from kraft pulp industry processes, showing a correlation of at least 88% between estimated and actual data. This fitting was possible due to the ability of generalized additive models to easily adequate data with high variability and mild control, on which experimental design treatments are not efficient. Neural artificial networks were able to represent at least 60% of the alkaline charge and the digester production.

Keywords: Statistical computing; Wood cooking; Multivariate statistics; Predictive models; Pulping chemistry

RESUMO

No presente trabalho, dados observados em escala industrial, durante alguns anos, foram submetidos à modelagem empregando modelos aditivos generalizados e redes neurais artificiais, como ferramentas para avaliar a influência de algumas variáveis, da madeira e do processo, sobre a produção do digestor e carga alcalina. Os modelos aditivos generalizados foram ajustados utilizando o software R, através da biblioteca "mgcv". Foram aplicados testes de significância para cada modelo ajustado. De forma complementar, os dados foram ajustados em 100 redes neurais artificiais para relacionar a produção do digestor e a carga alcalina, com diversas variáveis de processo e de polpa de celulose. Nesta etapa do trabalho, foi empregado o software Statistica e os dados resultantes mostraram que o modelo aditivo generalizado constitui uma boa opção para representar os fenômenos da indústria de celulose, com correlação de 88% entre os dados reais e os estimados. Este modelo é adequado para variáveis que apresentam alta variabilidade e em que não há um rigoroso controle, diferentemente do que ocorre em dados provenientes de delineamentos experimentais. A aplicação de redes neurais artificiais para estimar a produção do digestor e para carga alcalina também mostrou-se de moderada a forte, com correlações entre os dados reais e estimados que ficaram acima de 60%.

Palavras-chave: Computação estatística; Cozimento de madeira; Estatística multivariada; Modelos preditivos; Química da polpação

1 INTRODUCTION

Currently, the Kraft process is the most widely used to obtain pulp, consisting of the reaction of wood in the form of chips with white liquor (NaOH and Na_2S), which results in lignin dissolution and the release of cellulose fibers. This process has relatively low selectivity with regard to lignin removal, with release of extractives and dissolution of part of the carbohydrates also occurring. At the end of cooking, the carbohydrate removal rate remains considerable, causing production loss while maintaining a residual lignin content (Foelkel, 1977).

In all chemical pulp production processes, the variables time and temperature are considered the most important, as they directly affect lignin removal and final product quality. These variables are inversely related, that is, the higher the temperature, the shorter the cooking time. In addition, the reaction vessels are operationally heterogeneous and complex, since many chemical reactions occur within them, besides having process variables that are difficult to measure.

Pulp manufacturing is influenced by different silvicultural aspects, such as: physical (e.g., basic density), anatomical (fiber dimensions), and chemical (cellulose, hemicellulose, and lignin content) wood properties, which are generally interdependent (Carvalho; Silva; Colodette, 2014). For this reason, it is difficult to change one property without also affecting the others (Shimoyama; Barrichello, 1991). Another factor that drives the optimization of pulp yield is wood cost, which is influenced by genetic improvement and forest management variables.

Since the Kraft process is multivariate, the development of mathematical models for predicting parameters has been the subject of several studies. The use of prediction results from models leads to better performance in reducing pollutant emissions, increasing productivity, and improving the cost-effectiveness of production (Blanco et al. 2013). The use of Artificial Neural Networks (ANN), especially MLP (Multilayer Perceptron) type networks, has been studied through different models (Okan et al., 2015; Correia et al. 2018). However, no studies have been found that apply these models to predict digester production from wood and process properties simultaneously.

The objective of this work is to employ generalized additive models (GAM) and artificial neural networks (ANN) as tools to evaluate the influence of some wood and process variables on digester production using an industrial-scale database.

2 MATERIAL AND METHODS

2.1 Material

The database used in this study comes from the Process Information Management System (PIMS) of Veracel Celulose S.A. (Eunápolis-Bahia-Brazil), based on daily averages. Some data come from laboratory analyses and others from online meters installed at various points in the process. The online meters are calibrated against laboratory analyses using standard procedures as a reference.

The analysis frequency of the parameters generated and made available for this study:

- Apparent chip density – 6x/day;

- Basic wood density (chips) – 3x/day;
- Kappa number of the digester discharge pulp – online measurement (1 determination every 12 minutes);
- Digester production – calculated based on the speed of the screw feeding the chips; online measurement;
- Alkaline load – defined to maintain the target kappa number between 18 and 18.5.

To minimize interference in the analyses from data originating from unstable process conditions, data generated for a digester production below 3,000 tons were excluded from the evaluation. Production below this value occurs only when there are process disturbances and does not reflect reality. In addition, the online meters lose their reference and show values that are completely out of control.

Data on the wood supplying the mill were generated from the Forest Management System (SGF) of Veracel Celulose S.A. The wood properties collected were basic density, age, and rainfall. Density is determined before harvesting through material sampling using the destructive method with laboratory determination (Scan-CM 43:95, 1995). Age is determined during forest harvesting, based on the company's Forest Registry, which contains the planting records for the areas (planting date, genetic material, etc.). To keep the industrial and forest databases on the same time basis, 2 days were added to the wood supply data from the SGF, corresponding to the retention period of the chip pile.

This study used an industrial-scale database covering the period from January 2006 to March 2010, totaling 1,537 observations.

2.1 Methods

The models were fitted using R software (R Development Core Team, 2010), through the "mgcv" library (Wood, 2006), specific for fitting generalized additive models.

The associations between the variables studied were represented graphically, where the solid line represents the values fitted by the model and the dotted lines represent the standard error of the fitted values, interpreted as the "confidence interval". The vertical bars above the x-axis represent the observed values and their dispersion along the fitted curve.

Significance tests were applied to each fitted model (Tables). Significant results (p-value*100 <= adopted significance level) indicate that a significant association exists between the variables studied; the lower the p-value, the greater the intensity of the association. Each table shows the degrees of freedom corresponding to the smooth term, where the lower the smoothing degrees of freedom, the closer to linear the association between the variables.

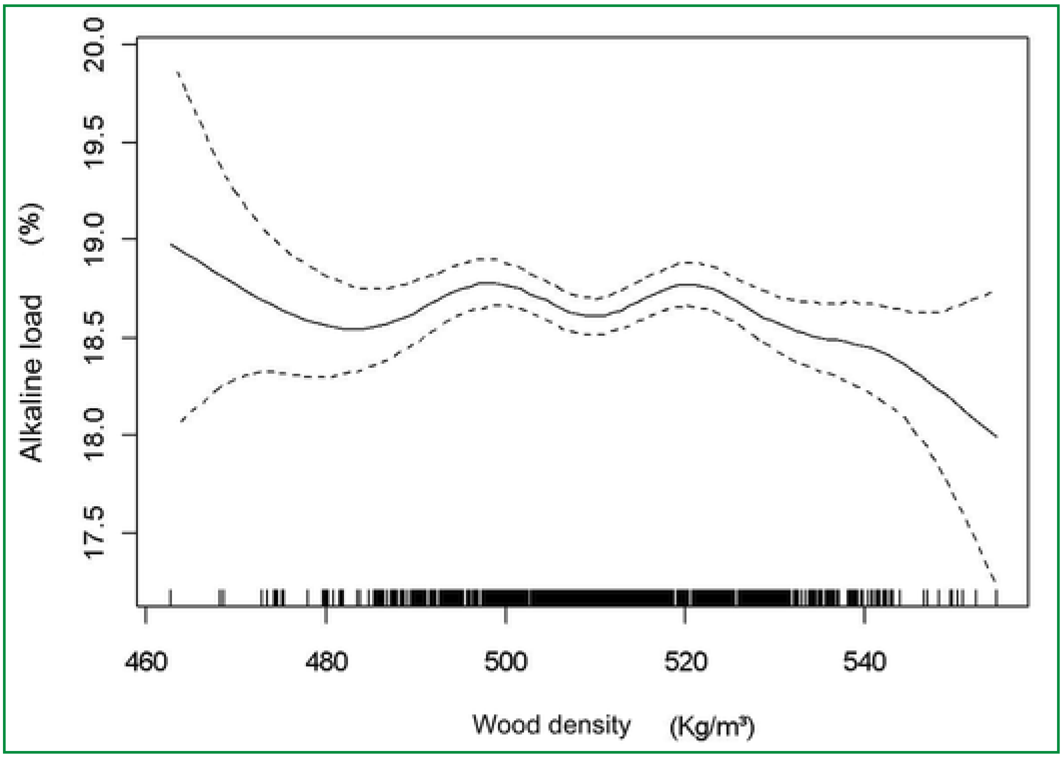
Surface plots were also produced, in which the predictor variables are represented on the graph's Cartesian axes and the contour lines represent the response variable. The lighter the area on the graph, the greater the magnitude of the response variable.

In a complementary way, the data were analyzed using artificial neural networks. The digester production variable was grouped into classes with an amplitude of 50 tons, allowing the correlations to be evaluated with the variables wood density, rainfall, chip dry content, chip density, and Kappa number. Likewise, the alkaline load was grouped into classes with an amplitude of 0.3. A total of 100 artificial neural networks (ANN) were fitted to relate digester production and alkaline load with the variables: wood density, age, rainfall, chip dry content, apparent chip density, basic chip density, pulp viscosity, and kappa. At this stage of the work, the Statistica software was used (Statsoft INC., 2007).

3 RESULTS AND DISCUSSIONS

Kraft pulping performance and the quality of the resulting pulps are influenced by variables associated with the raw material and the pulping process (Fengel; Wegener, 1989; D'Almeida, 1988). In the present work, the relationship between alkaline load and wood density was non-significant ($P>0.05$), that is, no well-defined associations were found (Figures 1 and 2 and Table 1). This finding indicates that an increase or decrease in wood basic density does not allow any inference about alkaline load using generalized additive models. These results agree with those observed by Mokfienski et al. (2003), cited by Gomide et al. (2005) and Carvalho et al. (2015). On the other hand, Lanna et al., (2001), cited by Gomide *et al.*, (2005), did not report this correlation, indicating the complexity of relationships between variables.

Figure 1 – Smoothed curve of alkaline load as a function of wood density



Source: Authors (2026)

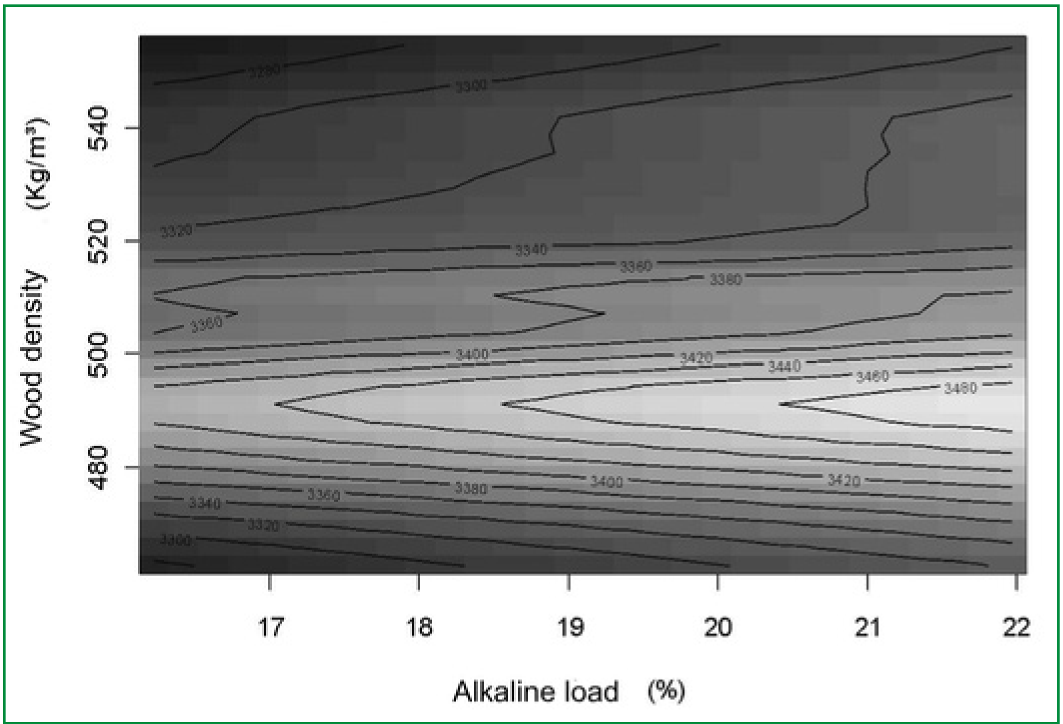
Table 1 – Significance test for the smoothed curve of alkaline load as a function of wood density

Significance of the smooth term	
d.f. s(Wood Density)#	6.7
F	1.93
p-value	0.053ns

Source: Authors (2026)

In where: #Degrees of freedom corresponding to the smooth term; nsNot significant ($P > 0.05$).

Figure 2 – Digester production as a function of wood density and alkaline load



Source: Authors (2026)

When Artificial Neural Networks are applied, a correlation between alkaline load and wood density is observed. For this, the alkaline load variable was grouped into classes with several amplitudes, and an amplitude of 0.3 allowed the following correlations to be verified:

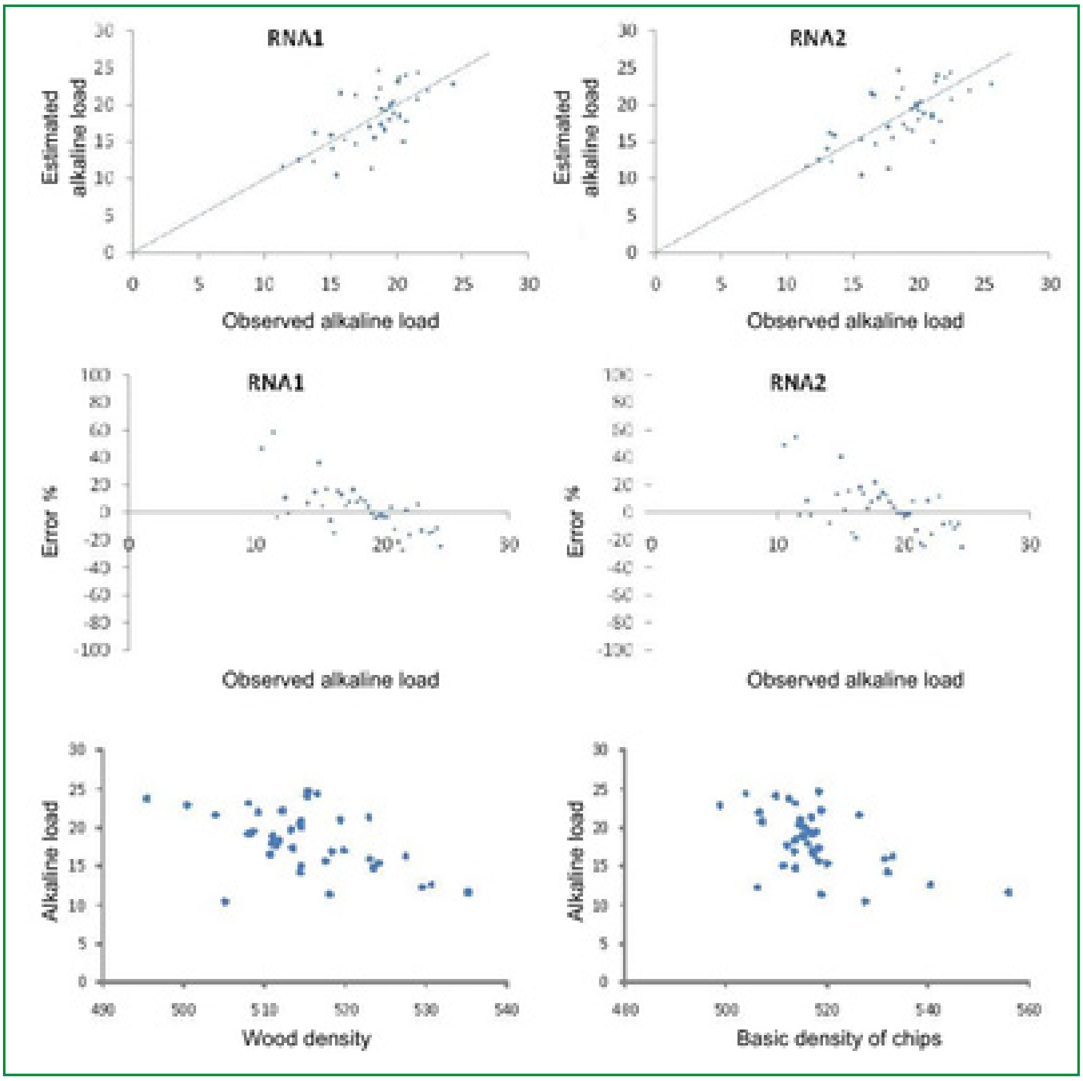
- **Wood density: -0.52**
- Age: -0.20
- Precipitation: -0.36
- Chip dry content: 0.17
- Apparent chip density: 0.25
- **Basic chip density: -0.54**
- Pulp viscosity: 0.05
- Kappa: 0.23

Two ANNs (100 each) were fitted to estimate alkaline load (in a class with an amplitude of 0.3) as a function of the means of the variables: wood density (ANN1) and basic chip density (ANN2). For each correlation, the best networks were: Multilayer

Perceptron, 2 input neurons, 2, 7 and 1 neurons in the hidden layers, 1 output neuron, or MLP 2-2-1, MLP 2:2-7-1:1 (ANN1) with r = 0.67, and MLP 2:2-5-1:1 (ANN2) with r = 0.68. A correlation of 0.60 (60%) between variables is considered **moderate to strong.** For Artificial Neural Networks (ANNs), this value **is acceptable**, allowing the identification of complex patterns and nonlinear relationships that conventional statistical methods (such as simple linear regression) would not easily detect (Costa, 2011). Tests with the other variables were not satisfactory.

Figure 3 shows the validation of the best ANN obtained, in which good response capacity of the trained models is observed, with a trend of estimated values similar to the actual alkaline load values measured in the process.

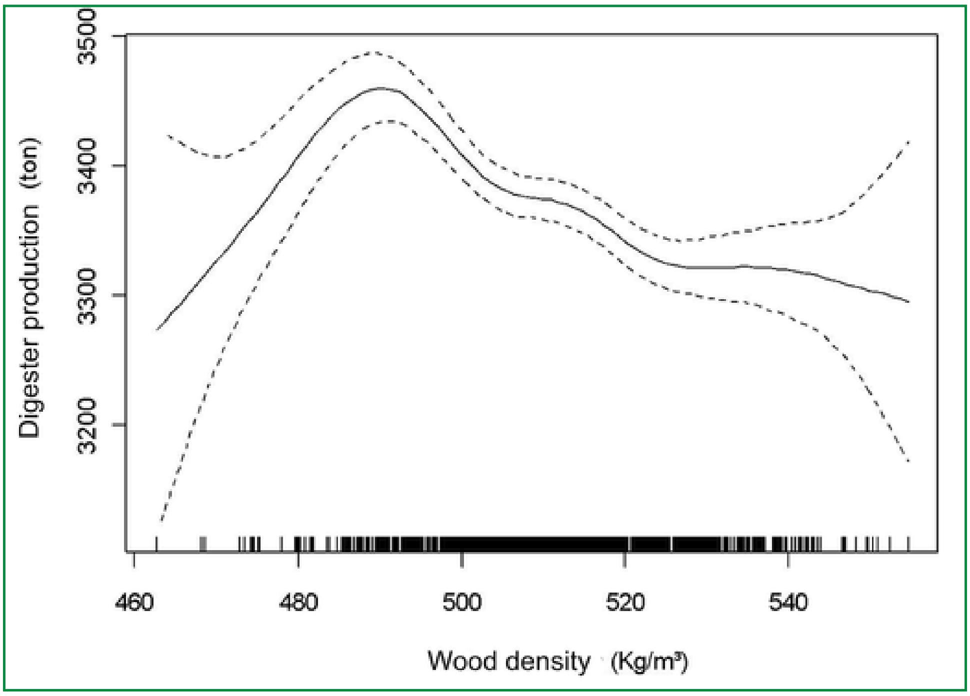
Figure 3 - Alkaline load as a function of wood density and basic chip density



Source: Authors (2026)

The correlation between digester production and wood density is variable (Figure 4), with densities between 490 and 530 kg/m³ being responsible for the highest cellulose pulp production. Establishing a narrow basic density range for digester supply proves ideal for producing pulp with uniform properties when intended for the manufacture of a single class of paper. In a study conducted by Santos and Sansígolo (2007), it was observed that low basic densities of eucalyptus clones are suitable for the production of papers for printing and writing, while high basic densities could be applied to sanitary papers.

Figure 4 – Smoothed curve of digester production as a function of wood density



Source: Authors (2026)

Table 2 – Significance test for the smoothed curve of digester production as a function of wood density

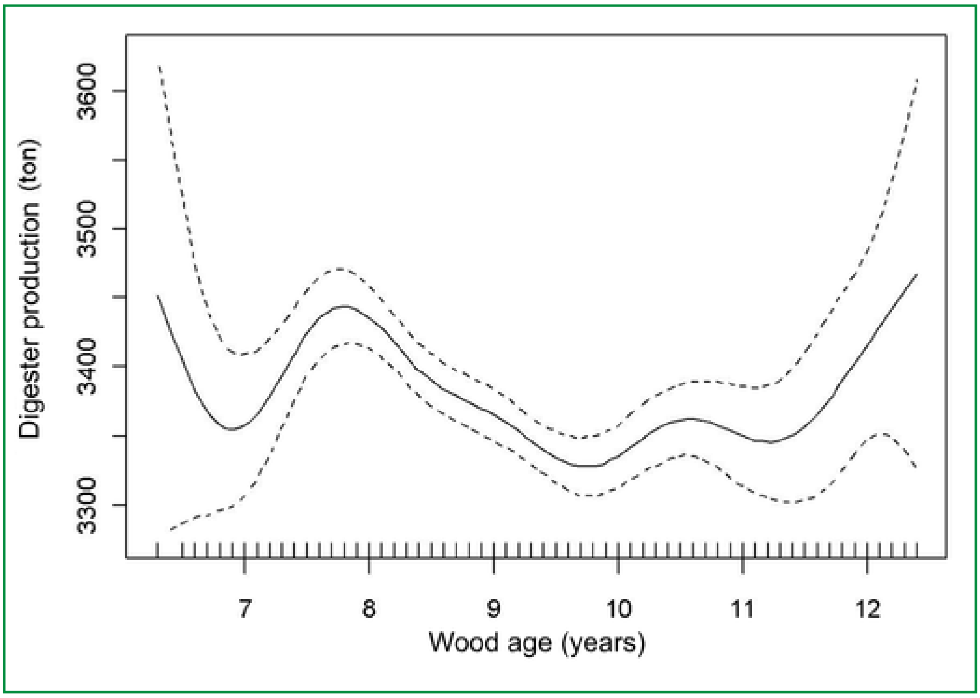
Significance of the smooth term	
d.f. s(Wood Density)#	6.29
F	10.39
p-value	2.75e-13***

Source: Authors (2026)

In where: #Degrees of freedom corresponding to the smooth term; *** Significant (P < 0.001).

The effect of wood age on chemical composition and process variables is increasingly studied, due to the possibility of influencing the final characteristics of the bleached pulp. It can be stated that age influences chemical composition, which stabilizes in the wood once the tree passes the temporal stages of juvenile wood variation (Trugilho et al. 1996; Morais, 2008). Figure 5 shows the correlation between digester production and wood age, identifying higher production between 7.5 and 8.0 years.

Figure 5 – Smoothed curve of digester production as a function of wood age (Source: the authors)



Source: Authors (2026)

Table 3 – Significance test for the smoothed curve of digester production as a function of wood age

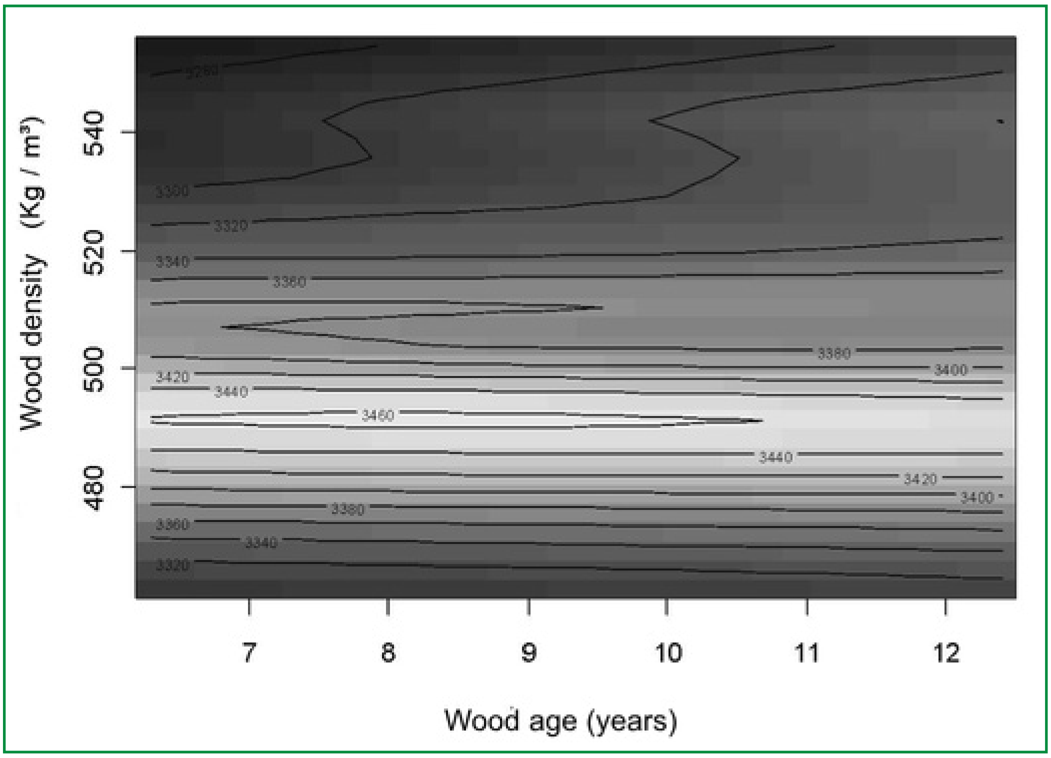
Significance of the smooth term	
d.f. s(Age)#	7.96
F	5.96
p-value	6.35e-08***

Source: Authors (2026)

In where: #Degrees of freedom corresponding to the smooth term; *** Significant ($P < 0.001$).

In the combined evaluation of the influence of age and wood density on digester production (Figure 6), density is observed to be more limiting than age. The ranges of highest production are found within a narrow density range (480-500 kg/m³), while for age this range is wider, varying from 6 to 10 years.

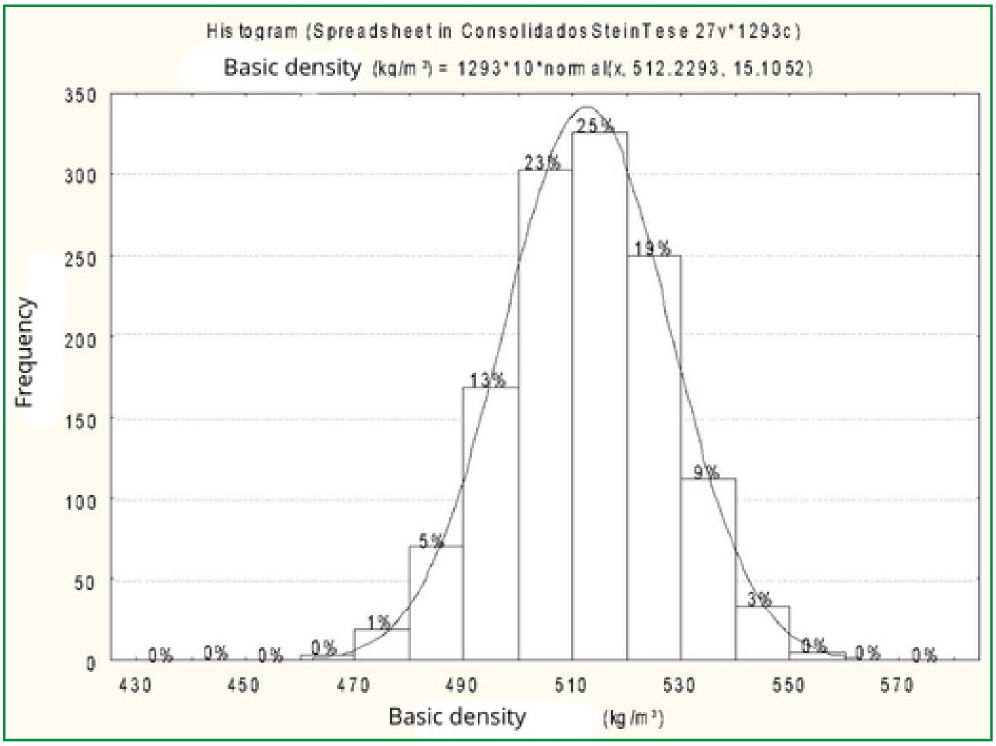
Figure 6 – Digester production as a function of wood density and age



Source: Authors (2026)

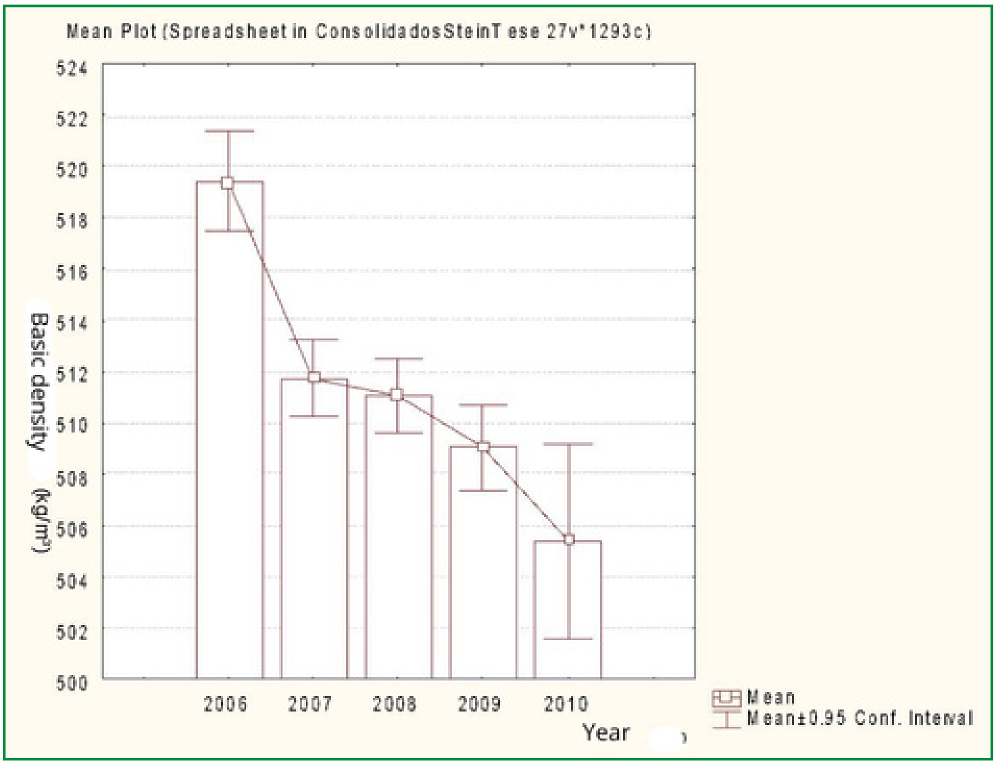
During the period evaluated, the mill's wood supply consisted mostly of wood with basic density ranging from 490 to 530 kg/m³sc (Figures 7 and 8). This was due to the fact that there was a surplus wood stock, raising the average age to the range of 8 to 10 years (Figures 9 and 10). Vital *et al.* (1984) reported that density tends to increase with increasing tree age, as a consequence of increased cell wall thickness and reduced cell width. This result may lead to higher reagent consumption in the stages following cooking, and correlations with kappa number and viscosity should be studied, as these are important for producing quality pulp.

Figure 7 – Distribution of basic density



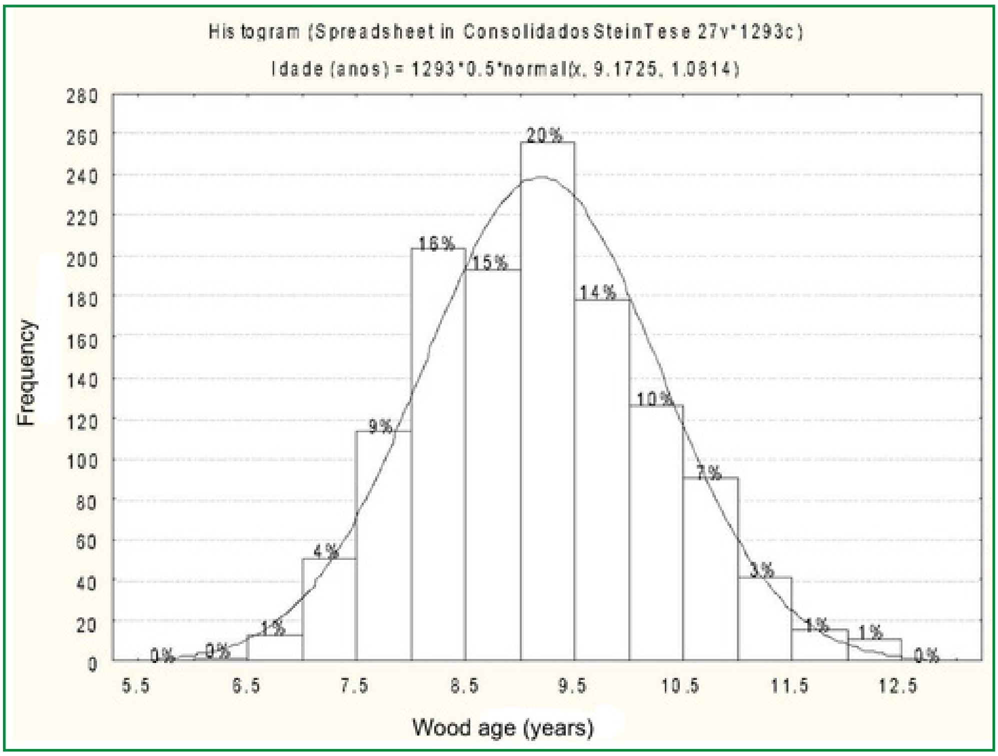
Source: Authors (2026)

Figure 8 – Histogram of basic density by year



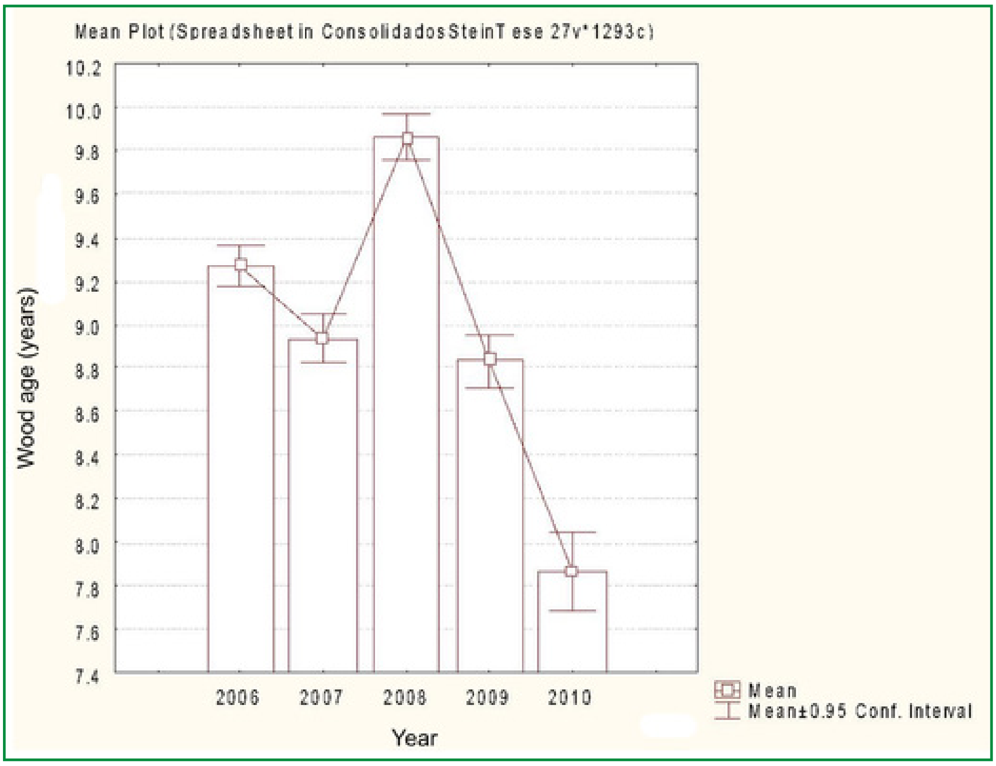
Source: Authors (2026)

Figure 9 – Distribution of wood age



Source: Authors (2026)

Figure 10 – Histogram of wood age by year



Source: Authors (2026)

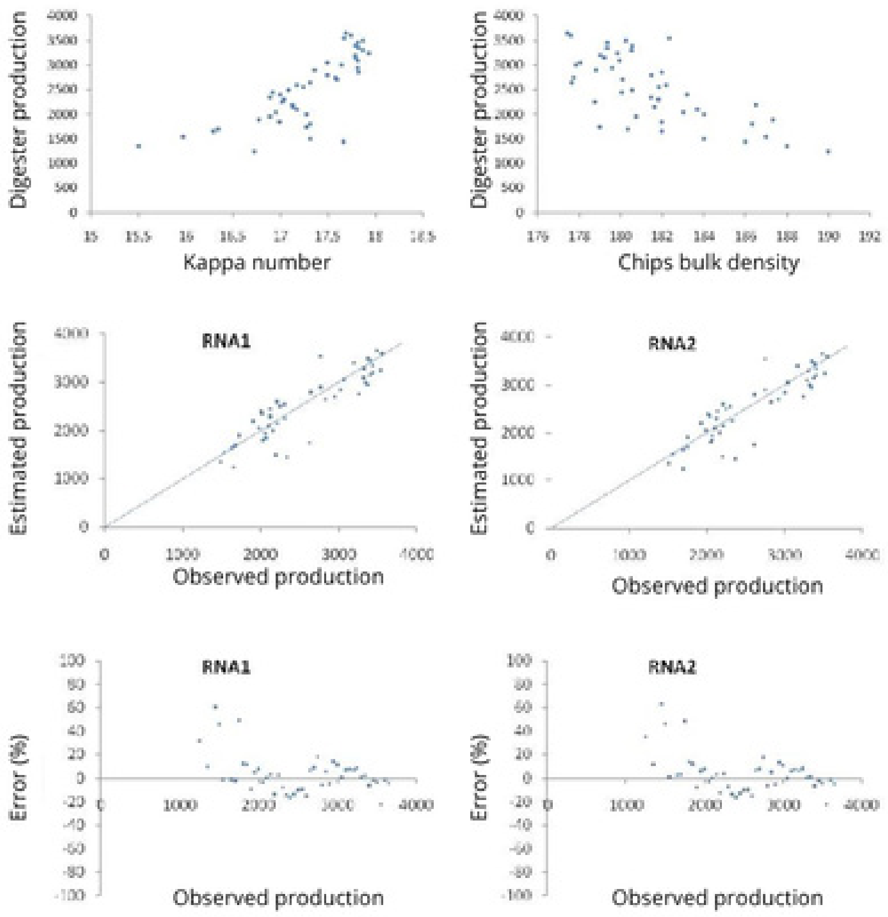
To evaluate the application of artificial neural networks in estimating digester production, this variable was grouped into classes with several amplitudes. With an amplitude of 50 tons (Excel function: MROUND(PRODUCTION;50)), the following correlations were obtained:

- Wood density: -0.07
- Age: -0.26
- Precipitation: 0.31
- Chip dry content: -0.08
- **Apparent chip density: -0.70**
- Basic chip density: -0.05
- Pulp viscosity: 0.20
- **Kappa: 0.70**

A total of 100 ANNs were fitted to estimate digester production (grouped into classes with an amplitude of 50 tons) as a function of the variables (mean values): wood density, age, rainfall, chip dry content, apparent chip density, basic chip density, pulp viscosity, and kappa. The option to optimize variable selection was applied, and the best networks, ANN1 (MLP 2-2-1, with r = 0.885) and ANN2 (MLP 2-2-1, with r = 0.883), selected the variables kappa and apparent chip density, respectively. The ANNs were trained using the supervised “backpropagation” algorithm.

Validation of the networks revealed good response capacity of the trained models, as the values estimated by the ANNs follow the trend of the actual digester production values measured in the process.

Figure 11 – Digester production as a function of kappa number and apparent chip density



Source: Authors (2026)

It is worth noting that apparent chip density is related to chip size and wood basic density. Considering that chip size is uniform in the process, wood basic density has a direct influence on this variable. Therefore, for higher apparent densities, due to an increase in wood basic density, it becomes necessary to adjust the mill process so that there is no loss in the cooking process. To do this, the alkaline load is increased to maintain the target kappa number. Without this adjustment, the kappa number will increase and delignification will decrease.

Rubini and Yamamoto (2006) reported the applicability of ANNs for predicting kappa number in oxygen delignification processes, using seven months of process data, gathering 811 variables to fit the models, with relative errors of at most 8.4% for kappa number after oxygen delignification. In the present work, the error showed absolute values around 20%, due to this property's composition being more variable when compared to pulp after oxygen delignification.

4 CONCLUSIONS

The present work shows that generalized additive models and artificial neural networks can be used efficiently for modelling digester production as a function of variables such as apparent chip density and kappa number in pulp mills. Furthermore, artificial neural networks are efficient for estimating alkaline load as a function of variables such as wood density and basic chip density.

Through the use of prediction techniques, it was found that an increase in alkaline load provides an increase in the delignification level and a reduction in kappa number. However, for alkaline loads above 20%, there is no reduction in kappa number, which shows that the practice of using higher alkaline loads, aiming exclusively at lignin removal, is not justified.

Considering the same final kappa number, the prediction models tested in this work demonstrated that an increase in alkaline load will result in reduced viscosity and cellulose yield. Such predictions can help prevent bottlenecks in subsequent chemical recovery processes, both in causticizing (lack of liquor for cooking) and in the boiler (excess solids for burning), which would result in reduced digester production.

ACKNOWLEDGMENTS

To the Universidade Federal de Viçosa and the Pulp and Paper Laboratory for the opportunity to carry out this work. To the company Veracel Celulose S.A., for providing the data necessary to carry out this work. To the Minas Gerais Chemistry Network (RQ-MG) and the Minas Gerais Multicenter Graduate Program in Chemistry (PPGMQ-MG), and to the Minas Gerais State Research Support Foundation – FAPEMIG (Process RED-00056-23) for all the support.

REFERENCES

BLANCO, A.; DAHLQUIST, E.; KAPPEN, J.; MANINNEN, J.; NEGRO, C.; RITALA, R. Use of modelling and simulation in the pulp and paper industry. **Mathematical and Computer Modelling of Dynamical Systems**, v. 5, n. 15, p. 409–423, 2009. DOI: https://doi.org/10.1080/13873950903375387

CARVALHO, D. M.; SILVA, M. R.; COLODETTE, J. L. Estudo da branqueabilidade da polpa kraft de eucalipto. **Ciência Florestal**, v. 25, n. 1, p. 185-197, jan-mar., 2015, DOI: http://dx.doi.org/10.1590/10.1590/1980-509820152505185

CARVALHO, D. M.; SILVA, M. R.; COLODETTE, J. L. Efeito da qualidade da madeira no desempenho da polpação Kraft. **Ciência Florestal**, v. 24, n. 3, p. 677-684, jul.-set., 2014, DOI: https://doi.org/10.1590/1980-509820142403015

CORREIA, F. M.; D'ANGELO, J. V. H.; ALMEIDA, G. M.; MINGOTI, S. A. Predicting Kappa Number In A Kraft Pulp Continuous Digester: A Comparison Of Forecasting Methods. **Brazilian Journal Of Chemical Engineering.** São Paulo, v. 3, n. 35, jul-sep 2018, DOI: https://doi.org/10.1590/0104-6632.20180353s20160678

COSTA, V. M. **Análise das variáveis de entrada de uma rede neural usando teste de correlação e análise de correlação canônica**. 2011. 110p. (Dissertação de Mestrado em Ciências, Tecnologia Nuclear – Reatores) – USP – IPEN, São Paulo, 2011.

D'ALMEIDA, M. L. O. **Celulose e papel: tecnologia de fabricação da pasta celulósica**. 2.ed., São Paulo: IPT, p. 45-106, 1988.

FENGEL, D.; WEGENER, G. **Wood—Chemistry, Ultrastructure, Reactions**. 2nd Edition, Berlin: Walter de Gruyter, 1989.

FOELKEL, C. E. B. Processo kraft. *In:* **Química dos processos de produção de celulose.** Belo Oriente: CENIBRA/UFV, 73 p., 1977.

GOMIDE, J. L.; COLODETTE, J. L.; OLIVEIRA, R. C.; SILVA, C. M. Caracterização tecnológica, para produção de celulose, da nova geração de clones de *Eucalyptus* do Brasil. **Revista Árvore**, Viçosa, vol 29, nº 1, Jan/Feb 2005. DOI: https://doi.org/10.1590/S0100-67622005000100014

MORAIS, P. H. D. **Effect of age of the wood of eucalyptus in its chemical and pulpability, and bleachability and properties of the pulp**. 2008. 79 f. Dissertação (Mestrado em Agroquímica analítica; Agroquímica inorgânica e Físico-química; Agroquímica orgânica) - Universidade Federal de Viçosa, Viçosa, 2008.

OKAN, O. T.; DENIZ, I.; TIRYAKI, S. Application Of Artificial Neural Networks For Predicting Tensile Index And Brightness In Bleaching Pulp. **Maderas-Ciencia Y Tecnologia**. Concepción, v. 3, n. 17, p. 571-584, jul. 2015. DOI: http://dx.doi.org/10.4067/S0718-221X2015005000051

R DEVELOPMENT CORE TEAM (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: http://www.R-project.org, Accessed in: 27 de fev, de 2026.

RUBINI, B.R.; and YAMAMOTO, C.I. Development of Prediction Oxygen Delignification Models Using Kinetic Expressions and Neural Networks. **Tappi Journal**, v. 5, p. 3-6, 2006.

SANTOS, S. R.; SANSÍGOLO, C. A. Influência da densidade básica da madeira de clones de *Eucalyptus grandis* x *Eucalyptus urophylla* na qualidade da polpa branqueada. **Ciência Florestal**, v. *17, n.* 1, p. 53–63, 2007. DOI: https://doi.org/10.5902/198050981935

Scandinavian Pulp, Paper and Board Testing Committee - **SCAN-CM 43:95** - Basic density - Wood Chips for Pulp Production - Estocolmo: SCAN, 1995.

SHIMOYAMA, V. R. S.; BARRICHELO, L. E. G. – **Densidade básica da madeira, melhoramento e manejo florestal – IPEF – Série Técnica – Vol. 6**; Piracicaba: IPEF, Nº 20; pag. 1-22, 1990.

StatSoft, Inc. (2007) STATISTICA (Data Analysis Software System), Version 7.0. Available from: http://www.statsoft.com. Accessed in: 26 de fev, de 2026.

TRUGILHO, P. F.; LIMA, J. T.; MENDES, L. M. Influência da Idade nas Características Físico-Químicas e Anatômicas da Madeira de *Eucalyptus saligna*. **Cerne,** Lavras, v. 2, p. 94-111, 1996.

VITAL, B. R.; PEREIRA, A. R.; LUCIA, R. M. D.; ANDRADE, D.C. **Efeito da idade da árvore na densidade da madeira de Eucalyptus grandis cultivado na região do cerrado de Minas Gerais**. Brasília: IBDF, (8): 49 - 52, 1984 (Boletim Técnico).

WOOD, S. N. **Generalized Additive Models: An Introduction with R**. London: Chapman and Hall/CRC, 2006.

Authorship Contribution

1 Fabiano da Rocha Stein

Forestry Engineer, MSc in Forestry Engineering, Forestry Planner

https://orcid.org/0009-0002-1509-9368 • fabiano.stein@ufv.br

Contribution: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Visualization; Writing – original draft

2 Gustavo Eduardo Marcatti

Forestry Engineer, PhD. in Forestry Engineer, Professor

https://orcid.org/0000-0003-1455-4859 • gustavomarcatti@ufsj.edu.br

Contribution: Data curation; Formal analysis; Software; Validation

3 Marcelo Moreira da Costa

Forestry Engineer, PhD. in Forestry Engineer, Professor

https://orcid.org/0000-0002-0132-1231 • mmdc@ufv.br

Contribution: Project administration; Resources

4 Cassiano Rodrigues de Oliveira

Chemist, PhD. in Agrochemistry, Professor

https://orcid.org/0000-0003-3306-1451 • cassiano.oliveira@ufv.br

Contribution: Conceptualization; Project administration; Writing – original draft; Writing – review & editing

5 Hélio Garcia Leite

Forestry Engineer, PhD. in Forestry Engineer, Professor

https://orcid.org/0000-0002-6357-9174 • hgleite@ufv.br

Contribution: Conceptualization; Data curation; Supervision; Validation

How to quote this article

STEIN, F. R.; MARCATTI, G. E.; COSTA, M. M.; OLIVEIRA, C. R.; LEITE, H. G. Prediction of kraft pulp industrial production with generalized additive models and artificial neural networks. **Ciência Florestal**, Santa Maria, v. 36, e95533, p. 1-19, 2026. DOI 10.5902/1980509895533. Available from: https://doi.org/10.5902/1980509895533. Accessed in: day month abbr. year.

Data Availability Statement:

Datasets related to this article will be available upon request to the corresponding author.