

## Importância relativa das variáveis preditoras no processo de modelagem da produtividade florestal

*Daniel Henrique Breda Binoti<sup>1</sup> Helio Garcia Leite<sup>2</sup> Valdir Carlos Lima de Andrade<sup>3</sup> Marcio da Conceição<sup>4</sup> Nairam Felix de Barros Filho<sup>4</sup> Leonardo Machado Pires<sup>4</sup> Luiz Otávio Rodrigues Pinto<sup>1</sup> Thuliany Fernandes Araujo Paes<sup>1\*</sup>*

<sup>1</sup> Eldorado Brasil, Rod. BR-158 – Km 231, Três Lagoas - MS, Brasil.

<sup>2</sup> Universidade Federal de Viçosa, Departamento de Engenharia Florestal, Av. Ph Rolfs, SN, Viçosa - MG, Brasil.

<sup>3</sup> Universidade Federal de Tocantins, Departamento de Engenharia Florestal, Av. NS-15, Quadra 109, Norte, SN, Palmas - TO, Brasil.

<sup>4</sup> The Timber Group, Rua Santa Lucia, 250, Montes Claros - MG, Brasil.

### Original Article

#### \*Corresponding author:

thuliany.paes@eldoradobrasil.com.br

#### Keywords:

Eucalyptus

Artificial Neural Networks

Forest productivity models

#### Palavras-chave:

Eucalipto

Redes Neurais Artificiais

Modelos de crescimento e produção

#### Received in

2022/03/22

#### Accepted on

2022/09/28

#### Published in

2022/12/31



DOI:

<http://dx.doi.org/10.34062/afs.v9i4.13519>

**RESUMO:** A modelagem do crescimento e produção florestal é um grande desafio para os gestores florestais em função da grande quantidade de variáveis envolvidas e da importância das estimativas geradas para a tomada de decisão no empreendimento florestal. Diversos métodos estatísticos e de inteligência artificial podem ser utilizados visando a verificação da importância das variáveis e seleção das mesmas para o processo de modelagem florestal. Neste estudo é demonstrado o uso do método de perturbação em modelos de Redes Neurais Artificiais na definição da importância relativa de variáveis preditoras (silviculturais, climáticas e de manejo) da produtividade de povoamentos de eucalipto ao final da rotação (produção florestal). Foram utilizados dados de 320 talhões de plantios de eucalipto localizados no norte do Estado de Minas Gerais, com idade superior a sete anos. A precipitação distribuída em diversas idades e o teor de argila do solo foram as variáveis de maior importância para a predição do volume na idade de corte.

### Relative importance of predictor variables in the forest productivity modeling

**ABSTRACT:** Modeling forest growth and production is a major challenge for forest managers due to the large number of variables involved and the importance of the generated estimates for decision making in the forestry enterprise. Several statistical and artificial intelligence methods can be used to verify the importance of variables and their selection for the forest modeling process. This study demonstrates the use of the perturbation method in defining the relative importance of predictor variables (silvicultural, climatic and management) in predicting the productivity of eucalyptus stands at the end of the rotation. Data from 320 eucalyptus plantations located in the north of the State of Minas Gerais, aged over seven years, were used. Precipitation distributed at different ages and soil clay content were the most important variables for the prediction of volume at cutting age.

## Introdução

O manejo de florestas equiâneas envolve variáveis biológicas, como o potencial de crescimento de determinado genótipo em um tipo de solo específico. Mas também envolve variáveis de ordem econômica, como o valor de determinada atividade ou preço de determinado fertilizante. Há ainda variáveis que estão relacionadas a questões sociais, como a regulação da mão de obra de determinada empresa ou região na qual o empreendimento exerce influência. Frequentemente essas variáveis são correlacionadas, e a alteração de uma dessas pode ocasionar uma série de alterações nas demais variáveis (Buongiorno e Guilles, 2003).

Um dos elementos essenciais para o manejo de florestas equiâneas é a predição ou projeção de estoques de colheita de madeira (Campos e Leite, 2017). A previsão do estoque (prognose), para cada unidade de manejo, é feita usualmente com emprego de modelos de crescimento e produção (Clutter et al., 1983). A construção desses modelos de forma assertiva para estimação volumétrica em plantios florestais é fundamental para controle de produção, planejamento do abastecimento fabril, atendimento as demandas internas e externas ao empreendimento, bem como análises de índices produtivos. Os resultados da predição auxiliam em questões financeiras como negociações de compra e venda, avaliações econômicas e análises de investimentos (Bettinger et al., 2009).

Diversas ferramentas podem ser utilizadas para a modelagem do crescimento e produção de plantios florestais. Dentre essas ferramentas destacam-se modelos oriundos de ferramentas estatísticas e de aprendizado de máquinas. Modelos estatísticos são amplamente utilizados no processo de modelagem gerando resultados satisfatórios na grande maioria das vezes. Situações como a necessidade de estimação da produtividade na idade de corte tendo como base informações de inventário florestal a partir de uma idade muito precoce (Cruz et al., 2008), ou simulações de curvas de crescimento e produção considerando variáveis climáticas e suas anomalias (Alcântara, 2015; Dávila Veja, 2018), ou ainda a previsão dos impactos da alteração de práticas silviculturais na produtividade (Vanclay, 1994; Magalhães, 2014), têm exigido a utilização de ferramentas de aprendizagem de máquinas para a modelagem do crescimento e produção florestal.

Dentre os métodos de aprendizado de máquinas mais utilizados para prognose em povoamentos equiâneos destacam-se os modelos conexionistas (Campos e Leite, 2017). Dentre esses modelos destacam-se as redes neurais artificiais (RNA) e as máquinas vetores de suporte (SVM). Ambos os modelos podem representar um ganho no processo de modelagem do crescimento de produção quando comparados com modelos estatísticos, pois esses extraem o conhecimento a partir dos dados não necessitando ao usuário da ferramenta o processo de

seleção de variáveis ou conhecimento avançado no processo estatístico (Kordon, 2010).

A modelagem do crescimento e produção utilizando métodos de aprendizado de máquinas, apesar de apresentar bons resultados, traz alguns riscos no caso de uso indiscriminado (Marwala, 2015). É necessário atentar para as relações reais de causa e efeito entre as variáveis dependentes e variáveis independentes; para a necessidade de inclusão de variáveis seguindo o princípio da parcimônia, para o realismo biológico dos modelos matemáticos gerados e para a sensibilidade quando da simulação de intervenções na floresta. Isto evidencia a importância da seleção preliminar de variáveis de entrada para RNA e SVM. Essa seleção visa diminuir a dimensionalidade dos dados (Weston et al., 2000), reduzir o custo computacional e aumentar a exatidão das saídas (Olden e Jackson, 2002; Olden et al., 2004). O papel da seleção é definir quais variáveis dependentes são mais adequadas para representar a informação contida nos dados, possivelmente em contraste com a adoção de todas elas (Fávero e Belfiore, 2017).

Diversos métodos podem ser utilizados para a seleção de variáveis em modelos de aprendizado de máquinas, contudo, o método da perturbação é o que apresenta uma melhor aplicação prática (Olden et al., 2004). No caso de uma RNA, o método de perturbação avalia a mudança no erro quadrático médio ao adicionar uma pequena quantidade de ruído branco (dados sabidamente alterados) a cada neurônio de entrada, enquanto os demais neurônios de entrada permanecem inalterados. A mudança resultante no erro quadrático médio, para cada perturbação de entrada, ilustra a importância relativa das variáveis de entrada (Olden e Jackson, 2002; Olden et al., 2004).

Neste estudo é demonstrado o uso do método de perturbação na definição da importância relativa de variáveis preditoras (silviculturais, climáticas e de manejo) na produtividade de povoamentos de eucalipto ao final da rotação.

## Material e Métodos

Os dados utilizados foram provenientes de 320 talhões de povoamentos de clones de eucalipto, localizados no norte do Estado de Minas Gerais, com idades superiores a sete anos. Os dados possuem em média aproximadamente 25 ha sendo utilizado cerca de 3 parcelas por talhão. Para este estudo foram utilizadas apenas as produtividades da idade de sete anos, que foi a rotação definida para a floresta onde se encontram os povoamentos. As referências de idade para realização de atividades silviculturais nos talhões são apresentadas na Tabela 1.

A precipitação anual e distribuição da precipitação durante o ano de obtenção dos dados foram utilizadas como preditoras para a produtividade na idade de corte. Os dados climáticos

dos talhões foram obtidos na plataforma NASA POWER, coletados automaticamente por script escrito, utilizando a linguagem de programação

Python, com uso de API disponibilizado pela NASA (Stackhouse et al., 2017).

Tabela 1. Padrões operacionais para as principais práticas silviculturais aplicadas aos talhões

<b>Atividade</b>	<b>Nº de dias para realização da atividade, em relação ao momento do plantio</b>
Primeira adubação de manutenção	90
Primeira capina manual em linha total	90
Segunda adubação de manutenção	366
Segunda capina manual em linha total	180
Subsolagem com fosfatagem	-30
Terceira capina manual em linha total	360
Terceira adubação de manutenção	720
Capina mecânica em linha com pré-emergente	60

Como variáveis edáficas e de nutrição mineral, utilizou-se os níveis de adubação de nitrogênio, fósforo e potássio, teor de argila e matéria orgânica do solo. As variáveis espaçamento, altitude e genótipo também tiveram sua importância avaliada. Após a determinação da importância das variáveis estudadas, avaliou-se o aumento do

coeficiente de correlação entre produtividades observadas e previstas, com a inclusão de cada variável e seguindo sua ordem de importância. Na Tabela 2 são apresentadas as 25 variáveis utilizadas como representativas da produtividade média aos sete anos.

Tabela 2. Variáveis consideradas para estimação da produtividade florestal aos sete anos

<b>Variável</b>	<b>Tipo</b>	<b>Descrição</b>
Capina mecânica em linha total com pré-emergente	Contínua	Dias até data de referência
Primeira adubação de manutenção	Contínua	Dias até data de referência
Primeira capina manual em linha total	Contínua	Dias até data de referência
Segunda adubação de manutenção	Contínua	Dias até data de referência
Segunda capina manual em linha total	Contínua	Dias até data de referência
Subsolagem com fosfatagem	Contínua	Dias até data de referência
Terceira adubação de manutenção	Contínua	Dias até data de referência
Terceira capina manual em linha total	Contínua	Dias até data de referência
Genótipo	Nominal	Identificação do clone
Teor de argila	Contínua	Porcentagem
Teor de matéria orgânica	Contínua	Porcentagem
Altitude	Contínua	Metros acima do nível do mar
Área útil por planta	Contínua	m <sup>2</sup> por planta
Quantidade de adubação de Fósforo	Contínua	kg/ha
Quantidade de adubação de Nitrogênio	Contínua	kg/ha
Quantidade de adubação de Potássio	Contínua	kg/ha
Precipitação até um ano de idade (Idade <=1 ano)	Contínua	mm
Precipitação com um ano de idade	Contínua	mm
Precipitação com dois anos de idade	Contínua	mm
Precipitação com três anos de idade	Contínua	mm
Precipitação com quatro anos de idade	Contínua	mm
Precipitação com cinco anos de idade	Contínua	mm
Precipitação com seis anos de idade	Contínua	mm
Precipitação com sete anos de idade	Contínua	mm
Porcentagem de chuva no período seco	Contínua	mm

**Redes neurais artificiais e validação**

Foi treinada uma rede neural artificial com as seguintes especificações (Kordon, 2010):

- Uma camada oculta com 12 neurônios;
- 300 épocas no treinamento;
- Algoritmo de treinamento RPROD+;
- Validação cruzada k-folder com quatro grupos;
- 30 execuções;
- Uso da raiz quadrada do erro quadrático médio (RMSE) como medida de desempenho.

O processo de perturbação foi utilizado para a determinação da importância relativa de cada variável, sendo avaliada a exatidão das RNA's utilizando-se a análise gráfica de resíduos e, também, o coeficiente de correlação linear entre valores observados e estimados de produtividade média ao final da rotação regulatória.

**Resultados e Discussão**

Com todas as variáveis independentes obteve-se a RNA para a estimação da produtividade aos 7 anos. O desempenho médio das redes obtidas apresentou um coeficiente de correlação de 0,9466 e uma RMSE de 39,30%.

As dispersões das produtividades observadas e estimadas, dos erros relativos percentuais e a frequência de casos por classe de erro relativo, são apresentadas na Figura 1. Consta nesta figura o resultado da comparação entre os vetores de produtividades observadas e previstas pela RNA, com p-valor igual a 0,9922. A importância relativa e cada variável de entrada está indicada na Figura 2.

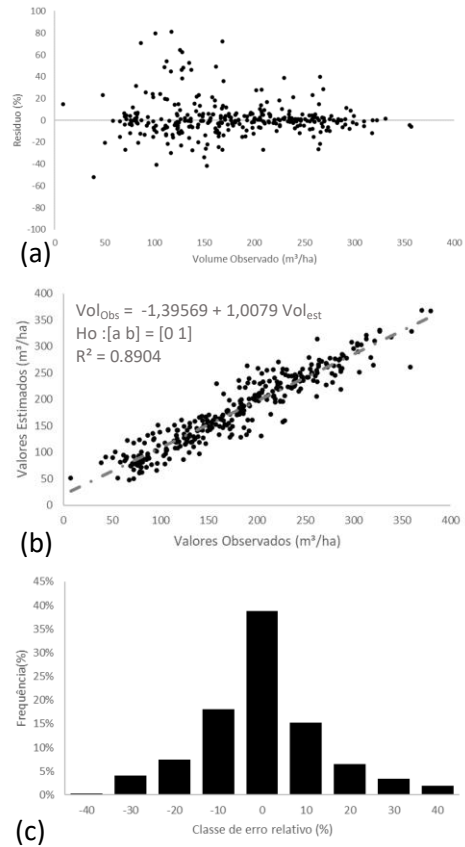


Figura 1. Análise gráfica da estimação volume na idade de corte em função das variáveis observadas. (a) Erro relativo em porcentagem entre volumes observados e estimados; (b) Relação entre volumes observados e estimados e (c) Histograma da distribuição do erro relativo em porcentagem.



Figura 2. Importância relativa das variáveis analisadas para a produtividade aos sete anos.

Na Figura 3 é apresentado o incremento do coeficiente de correlação linear de Pearson entre valores observados e preditos, em função do acréscimo de variáveis predictoras. A inclusão das variáveis seguiu o critério de importância estabelecida pelo processo de perturbação.

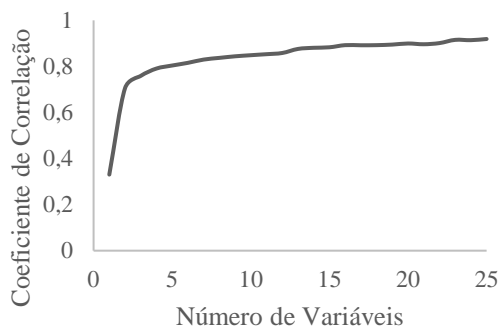


Figura 3. Variação do coeficiente de correlação linear de Pearson entre valores estimados e observados em função da adição de variáveis predictoras.

A predição da produtividade de povoamentos equiâneos exige exatidão, estimativas não viesadas e emprego de modelos que permitam simular cenários futuros e fazer análises econômicas. Em função da grande quantidade de variáveis e das interações entre

elas, que determinam o crescimento das plantas, a estimação de produtividade demanda a identificação e inclusão daquelas mais importantes, começando pela própria idade das árvores (Billings, 1952).

A construção de um modelo para predição da produtividade requer a inclusão dos fatores e variáveis determinantes do crescimento. Por outro lado, dimensionar os efeitos é importante para implementar ações antrópicas visando aumentar a produtividade dos povoamentos. Assim, isolar uma variável ou modelar a interação entre variáveis predictoras nem sempre é trivial, principalmente se forem incluídas variáveis relacionadas às práticas silviculturais e operações florestais.

Algumas variáveis e recursos de crescimento não podem ser modificados, em escala comercial, para níveis preestabelecidos. Por exemplo, déficit hídrico, amplitude de temperatura e altitude. Mas outras podem ser modificadas por ações antrópicas e, dentre elas, é importante identificar aquela mais impactantes. As variáveis, que podem ser modificadas e que influenciam, são aquelas relacionadas ao plantio, regeneração e condução dos povoamentos, incluindo o genótipo, o espaçamento de plantio, o regime de corte e às práticas silviculturais.

Dentre as variáveis contempladas neste estudo, as de maior contribuição relativa para a produtividade ao final da rotação de 7 anos, foi a precipitação, não podendo ser controladas ou

definidas por ações antrópicas. Entretanto, cenários de predição da precipitação para anos futuros, podem ser simulados possibilitando os estudos e análises sobre os efeitos técnicos e econômicos dessa. As ações antrópicas alteram os níveis dessas variáveis a longo prazo, mas não há como administrar para níveis desejados. Apenas algumas variáveis, dentre aquelas contempladas na modelagem, podem ser modificadas e controladas; a maioria relacionada à fertilidade e física do solo.

O conjunto de variáveis e suas interações utilizado neste estudo, na maioria das vezes desconhecidas, definem a capacidade produtiva do lugar, expressa pelo índice de sítio. Por isso a modelagem do crescimento e da produção é usualmente feita em função das variáveis idade, grau de ocupação da área (área basal ou frequência) e índice de sítio.

A inclusão do índice de sítio nos modelos quase sempre é suficiente para garantir a exatidão desejada nas estimativas de produtividade e o polimorfismo natural das curvas de produção de povoamentos de eucalipto. Com a inclusão do sítio, é possível avaliar o efeito da capacidade produtiva e, também, inferir sobre quanto aumentar a capacidade produtiva para se atingir determinada produtividade ao final da rotação, para uma espécie ou clone específico. No entanto, permanece a pergunta: como modificar os níveis das variáveis que em conjunto definem a capacidade produtiva? Não é uma tarefa simples, em função das interações muitas vezes desconhecidas, mas o primeiro passo é determinar a contribuição relativa de cada variável na produtividade ao final da rotação, ainda que sem considerar as interações.

À medida em que mais variáveis vão sendo adicionadas como entradas no modelo preditivo é esperado um aumento na correlação entre produtividades preditas e observadas. Isto é um fato e não é, portanto, relevante. Porém, ao modelar a relação entre coeficiente de correlação e acréscimo de variáveis é possível verificar que a partir de certa quantidade de variáveis não há um retorno significativo em produtividade (Figura 3). Em certo momento, haverá uma estagnação na correlação, mostrando que a adição de novas variáveis não contribuirá para aumentar a correlação e, portanto, um aumento da produtividade seria conseguido caso fossem feitas alterações nos níveis daquelas já incluídas na análise (Olden e Jackson, 2002; Olden et al., 2004).

O resultado obtido neste estudo, do incremento do coeficiente de correlação em função da adição de variáveis, mostrou que a partir da quinta variável adicionada o ganho no coeficiente de correlação foi pouco significativo (Figura 3). Até a quinta variável adicionada a correlação já era superior a 80%. Os benefícios da utilização de um número reduzido de variáveis incluem, uma melhor visualização e entendimento dos dados, redução do

tempo e esforço computacional para a geração dos modelos, além dos benefícios práticos da redução das variáveis mensuráveis em campo (Fávero e Belfiore, 2017).

Os resultados desse estudo podem ainda ser utilizados para a determinação das variáveis controláveis que possam ser manejadas pelos gestores florestais visando uma melhor produtividade futura. Variáveis de maior importância devem receber maior prioridade nos processos de manejo e gestão de recursos, o que poderá ser utilizado como critério inicial na definição de variáveis a serem amostradas e monitoradas para a previsão da produtividade futura. Contudo, a análise pontual e particular de cada plantio florestal deve ser considerada para que se possa ter uma adequação do processo de predição.

Considerando que algumas das variáveis importantes não podem ser alteradas os gestores devem se adaptar a essas. As variáveis ambientais merecem atenção pela conhecida importância e principalmente com relação aos cenários de mudanças climáticas. Keenan (2015) descreve alternativas para adaptação e minimização dos efeitos das mudanças climáticas no manejo florestal (i) prever respostas de espécies e ecossistemas ao clima futuro, (ii) ações de adaptação no manejo florestal, (iii) novas abordagens e ferramentas para tomada de decisão sob incerteza e parcerias mais fortes entre pesquisadores e profissionais e (iv) arranjos de políticas para adaptação na gestão florestal.

Esse trabalho foi desenvolvido com informações coletadas por um sistema de informações operacionais, permitindo a partir do mesmo inferir sobre questões importantes para a gestão e manejo dos plantios florestal. Esse fato demonstra a necessidade das empresas florestais da utilização desse tipo de sistema para a definição de estratégias e para a tomada de decisões de maneira eficiente e eficaz (Nobre et al., 2003; Angelo et al., 2007).

### Conclusões

O processo de perturbação é eficiente para a definição do grau de importância relativa de variáveis utilizadas para a predição da produtividade de plantios florestais na idade de corte.

A precipitação distribuída em diversas idades e o teor de argila do solo foram as variáveis de maior importância para a predição do volume na idade de corte.

### Referências

Alcântara AEM (2015) *Redes neurais artificiais para prognose do crescimento e da produção de povoamentos de eucalipto em Minas Gerais*. 2015. 43f. Tese (Doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa.

Angelo H, Almeida AN, Fontes PJP, Souza NA (2007) Determinantes para um Sistema de Informação Florestal. *Floresta e Ambiente* 2017; 24: e00000033

Bettinger P, Boston K, Siry LP, Grebner RL (2009) *Forest Management and Planning*. Elsevier. 331p.

Billings WD (1952) The environmental complex in relation to plant growth and distribution. *The Quarterly Review of Biology*, v.27, n.3, p.251-264.

Buongiorno J, Gilles JK (2003) *Decision methods for forest resource management*. San Diego, CA: Academic Press. 439p.

Campos JC, Leite HG (2017) *Mensuração Florestal: Perguntas e respostas*. Viçosa: UFV, 5. Ed, 636 p.

Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL (1983) *Timber management: A quantitative approach*. New York: John Wiley & Sons. 333 p

Cruz JP, Leite HG, Soares CPB, Campos JCC, Smit L, Nogueira GS, Oliveira MLR (2008) Modelos de crescimento e produção para plantios comerciais jovens de *Tectona grandis* em Tangará da Serra, Mato Grosso. *Revista Árvore*, Viçosa, v.32, n.5, p.821-828.

Dávila Vega AE (2018) *Modelagem de povoamentos de eucalipto no nordeste da Bahia utilizando regressão e redes neurais artificiais*. 2018. 52f. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa.

Fávero LP, Belfiore P (2017) *Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®*, 1216p.

Keenan RJ (2015) Climate change impacts and adaptation in forest management: a review. *Annals of Forest Science* 72, 145–167.

Kordon AK (2010) *Applying Computational Intelligence: How to Create Value*, Springer,-Verlag, Berlin/Heidelberg. 470p.

Magalhães JGS (2014) *Prognose e regulação de florestas de eucalipto utilizando modelo de crescimento e inteligência computacional*. Dissertação (Mestrado em Ciência Florestal)- Universidade Federal de Viçosa, Minas Gerais.

Marwala T (2015) *Causality, Correlation and Artificial Intelligence for Rational Decision Making*. Singapore: World Scientific. 208p.

Nobre SR, Rodriguez LCE, Assis MLR, Amaral TM. (2003) Avaliação do uso da tecnologia da informação nos sistemas de administração da produção florestal. *In: Anais do 8º Congresso Florestal Brasileiro: Informações gerais e Trabalhos Voluntários*; São Paulo. São Paulo: SBEF, SBS 12 p

Olden JD, Jackson DA (2002) Illuminating the 'black-box': a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*. 154:135-150

Olden JD, Joy MK, Death RG (2004) An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*. 178:389-397.

Stackhouse JR, Barnett T, Bristow W, Chandler JH, Westberg, Zhang TM, Tisdale BT (2017) A revitalized GIS-enabled NASA POWER web site with featuring updated data parameters, expanded data accessibility, and analysis functionality for the renewable energy and other applications. *Am Sol Energy Soc Meet* 1:9-13

Vanclay JK (1994) *Modelling forest growth and yield*. Wallingford, UK: CAB International, 312 p.

Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V (2000) *Feature selection for SVMs.*, v.13, p.668-674.