



## AJUSTE DE UM MODELO DE CRESCIMENTO E PRODUÇÃO DE *Tectona grandis* L. f. CONSIDERANDO A MULTICOLINEARIDADE

Lucas Araujo Moura<sup>1</sup>, Allan Libanio Pelissari<sup>1</sup>, Luan Demarco Fiorentin<sup>1</sup>, Gabriel Agostini Orso<sup>1</sup>

1 Universidade Federal do Paraná, Curitiba, PR, Brasil. E-mail: mouraforestal@gmail.com; allanpelissari@gmail.com; luandfiorentin@gmail.com; gabrielorso16@gmail.com  
Autor correspondente: Lucas Araujo Moura. E-mail: mouraforestal@gmail.com.

### RESUMO

A predição do crescimento e produção volumétrica de plantios comerciais é essencial ao manejo de florestas de *Tectona grandis*. Todavia, constantemente os pressupostos da regressão linear não são verificados. Diante disso, este trabalho objetivou construir um modelo linear considerando a presença de multicolinearidade para conceber estimativas precisas e inferências corretas da produção volumétrica em um plantio de *T. grandis*. Inicialmente, foram selecionadas as covariáveis e as variáveis transformadas que apresentaram relação com o volume e ajustado um modelo completo com todas as covariáveis. Posteriormente, foi utilizado o fator de inflação de variância corrigido para verificar o nível de multicolinearidade das variáveis. Utilizou-se o método *stepwise* para seleção das variáveis do modelo de produção, selecionando apenas as covariáveis com coeficiente de regressão significativo ao nível de 5%. Constatou-se que o modelo completo apresentou alta multicolinearidade por duas ou mais covariáveis. Entretanto, o método *stepwise* conseguiu reduzir a multicolinearidade, selecionando as covariáveis área basal, logaritmo do diâmetro médio quadrático e idade.

**Palavras-chave:** Covariáveis, Fator de inflação de variância corrigido, Método Stepwise, Produção volumétrica.

## ADJUSTMENT OF A GROWTH AND PRODUCTION MODEL OF *Tectona grandis* L. f. CONSIDERING MULTICOLLINEARITY

### ABSTRACT

The prediction of growth and volumetric production of commercial plantations is essential to the management of *Tectona grandis* forests. However, constantly the assumptions of linear regression are not verified. Therefore, this work aimed to build a linear model considering multicollinearity to conceive accurate estimates and correct inferences of volumetric production in a *T. grandis* plantation. Initially, the covariates and transformed variables that showed a relationship with volume were selected and a complete model with all covariates was fitted. Subsequently, the corrected variance inflation factor was used to check the level of multicollinearity of the variables. The stepwise method was used to select the variables for the production model, selecting only the covariates with significant regression coefficient at the 5% level. It was found that the complete model presented high multicollinearity by two or more covariates. However, the stepwise method was able to reduce multicollinearity by selecting the covariates basal area, logarithm of quadratic mean diameter and age.

**Key words:** Covariates, Corrected variance inflation factor, Stepwise method, Volume production.

### ASSOCIAÇÃO BRASILEIRA DE MENSURAÇÃO FLORESTAL



### INTRODUÇÃO

A *Tectona grandis* L. f., conhecida popularmente como Teak, Teca, May sak e Otaji, é originária do Sudeste Asiático e Subcontinente Índico (Índia, Mianmar, Tailândia e Laos), sendo considerada uma das mais importantes espécies arbóreas comerciais, sendo cultivada na Ásia, Oceania, África e América Latina (Gaitan-Alvarez *et al.*, 2019). O Brasil é um dos principais países onde a *T. grandis* é cultivada comercialmente, totalizando cerca de 93.957 hectares de área plantada (IBÁ, 2019).

Considerando a importância econômica da *Tectona grandis*, modelos de crescimento e produção em nível de povoamento total (MPT) são essenciais para fornecer estimativas e projeções futuras, auxiliando o planejamento de forma eficiente (Weiskittel *et al.* 2011; Mendonça *et al.*, 2017). Entretanto, para obter corretas inferências acerca de um modelo de regressão linear, é necessário atender alguns pressupostos básicos, como independência residual, normalidade dos resíduos, homogeneidade das variâncias e linearidade dos parâmetros (Schmidt & Finan, 2018).

Para o ajuste de modelos de crescimento e produção em nível de povoamento total, variáveis dendrométricas são coletadas, de tal modo que essas variáveis podem apresentar algum grau de colinearidade (Gujarati & Porter, 2011). De acordo com Willy *et al.* (2022), a multicolinearidade é uma forte correlação de duas ou mais variáveis independentes em um modelo de regressão. Essa interrelação influencia na significância dos coeficientes de regressão modelo, gerando valores imprecisos dos parâmetros estimados.

Dessa forma, o objetivo deste trabalho foi construir um modelo linear para crescimento e produção, considerando a multicolinearidade, para estimativas precisas e inferências corretas da produção volumétrica em um plantio comercial de *Tectona grandis*.

### MATERIAL E MÉTODOS

A área de estudo está localizada no município de Nossa Senhora do Livramento, estado do Mato Grosso. O clima local é do tipo tropical - Aw, com estação seca no inverno (Köppen 1948). A precipitação média é de 1.300 mm ao ano, ao passo que a temperatura média anual é de 25°C, com solos predominantemente do tipo PLANOSSOLO HÁPLICO Eutrófico (Embrapa, 2006).

O estudo foi desenvolvido em um povoamento seminal de *Tectona grandis* implantado no ano de 2001, totalizando 223,16 ha com espaçamento de 3 m x 3 m. Durante a implantação da amostragem, 46 parcelas permanentes de 30 m x 30 m foram alocadas na área, onde foram coletados os valores médios anuais do diâmetro a 1,3 m do solo (cm) e da altura total das árvores (m).

As covariáveis utilizadas para a construção do modelo de produção volumétrica em nível de povoamento ( $V$ ) foram: área basal ( $G$ ), média aritmética dos diâmetros a 1,3 m do solo ( $\bar{D}$ ), diâmetro dominante ( $D_{dom}$ ), diâmetro médio quadrático ( $D_g$ ), média aritmética das alturas totais

( $\bar{H}$ ) e altura dominante ( $H_d$ ).

O Fator de Inflação de Variância Corrigido (CVIF – *Corrected Variance inflation factor*) foi aplicado para identificar o nível de colinearidade entre as covariáveis. Assim,  $CVIF > 10$  pode representar um elevado grau de multicolinearidade em um modelo de regressão linear. Segundo Curto & Pinto (2011), o CVIF é definido conforme a Equação (1).

$$CVIF = \frac{1}{1-p^2} \cdot \frac{1-R^2}{1-R^2_j} \quad j = 1, 2, \dots, p \quad (1)$$

Em que:  $p$  é o número das variáveis preditoras;  $R^2_j$  é o coeficiente de correlação múltipla, resultante da regressão de  $X_j$  nos outros  $p - 1$  regressores.

Em seguida, um modelo final foi ajustado pelo método dos mínimos quadrados ordinários considerando o grau de multicolinearidade entre as covariáveis. O processo *stepwise* foi aplicado para auxiliar na seleção das variáveis preditoras, elegendo aquelas com coeficientes de regressão estatisticamente significativos ( $p$ -valor  $< 0,05$ ).

## RESULTADOS E DISCUSSÃO

O modelo inicial de produção, com as variáveis que apresentaram correlação linear significativa com volume, ao nível de 5% de significância, foi composto pelas variáveis área basal ( $G$ ), média aritmética dos diâmetros a 1,3 m do solo ( $\bar{D}$ ), diâmetro dominante ( $D_{dom}$ ), diâmetro médio quadrático ( $D_g$ ), média aritmética das alturas totais ( $\bar{H}$ ) e altura dominante ( $H_d$ ) (Tabela 1). Os coeficientes de regressão estimados apresentaram significância ao nível de 5% pelo teste  $t$ , exceto para  $H_d$ .

**Tabela 1.** Parâmetros estimados para o modelo inicial de produção volumétrica em povoamento de *Tectona grandis*

Coeficiente de regressão	Valor estimado	Erro padrão	p-valor	CVIF
$\hat{\beta}_0$	-76,449*	2,521	< 0,05	
$\hat{\beta}_1 G$	8,768*	0,146	< 0,05	21,45
$\hat{\beta}_2 \bar{D}$	-25,206*	10,030	< 0,05	36.033,86
$\hat{\beta}_3 D_{dom}$	-3,710*	0,734	< 0,05	162,13
$\hat{\beta}_4 D_g$	30,613*	10,147	< 0,05	36.761,50
$\hat{\beta}_5 \bar{H}$	3,505*	1,533	< 0,05	443,84
$\hat{\beta}_6 H_d$	-1,721 <sup>ns</sup>	1,185	0,143	226,50

Em que:  $\hat{\beta}_i$  é o coeficiente de regressão do modelo, \* é significativo ao nível de 5%, <sup>ns</sup> é não significativo e CVIF é o fator de inflação de variância corrigido.

Ao avaliar o grau de multicolinearidade por meio do Fator de Inflação de Variância Corrigido (CVIF), foi possível observar que, com exceção da área basal, as demais covariáveis apresentaram colinearidade muito alta ( $CVIF > 10$ ) com as variáveis preditoras do modelo (Tabela 1). Cruz *et al.* (2019) relatam que modelos com muitas covariáveis podem implicar em complicação nos ajustes devido à alta correlação entre as variáveis preditoras, tornando os coeficientes de regressão imprecisos.

Desse modo, após a aplicação do processo *stepwise* e considerando o limite de CVIF próximo

a 10, foram selecionadas a área basal ( $G$ ) e o logaritmo do diâmetro médio quadrático -  $\log(D_g)$ . Entretanto, por se tratar de um modelo de produção volumétrica em nível de povoamento, a idade ( $t$ ) também foi incorporada (Tabela 2). Essas variáveis foram estatisticamente significativas, ao nível de 5%.

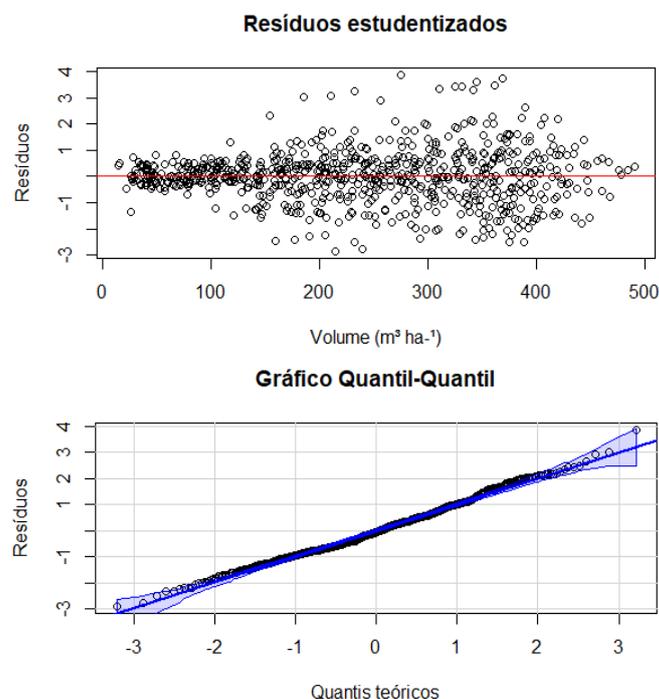
**Tabela 2.** Parâmetros estimados para o modelo final de produção volumétrica em povoamento de *Tectona grandis*

Coefficiente de regressão	Valor estimado	Erro padrão	p-valor	CVIF
$\hat{\beta}_0$	103,173*	8,656		
$\hat{\beta}_1 G$	10,546*	0,097	< 0,05	11,59
$\hat{\beta}_2 \log(D_g)$	-77,725*	3,954	< 0,05	10,45
$\hat{\beta}_3 t$	6,557*	0,183	< 0,05	6,06

Em que:  $\hat{\beta}_i$  é o coeficiente de regressão do modelo, \* é significativo ao nível de 5%, <sup>ns</sup> é não significativo e CVIF é o fator de inflação de variância corrigido.

O uso de um modelo final de produção volumétrica que seleciona número menor de variáveis é vantajoso por ser mais parcimonioso e atender aos objetivos de estudo. Segundo Campos & Leite (2009), modelos mais simples e que necessitam menos esforço computacional e a campo são mais gerais e não sofrem menor influência da multicolinearidade, sendo recomendados aos ajustes de modelos de regressão.

Por fim, no diagnóstico do modelo, foi possível constatar coeficiente de determinação ajustado ( $R^2_{aj.}$ ) de 0,995 e erro padrão da estimativa (SEE%) de 3,62%. Os resíduos estudentizados apresentaram ausência de viés, bem como corroboraram a presença de variâncias homogêneas (Figura 1). Por outro lado, foi possível observar que os resíduos tenderam à distribuição normal por meio do gráfico Quantil-Quantil, apesar uma pequena fuga da normalidade pelos extremos.



**Figura 1.** Gráficos de resíduos estudentizados e quantil-quantil para o modelo final de produção volumétrica em povoamento de *Tectona grandis*

## CONCLUSÃO

As covariáveis área basal, logaritmo do diâmetro médio quadrático e idade foram selecionadas para compor o modelo de crescimento e de produção volumétrica em nível de povoamento para *Tectona grandis*.

O método *stepwise* possibilitou gerar um modelo mais parcimonioso e com o Fator de Inflação de Variância Corrigido (CVIF) próximo a 10. O modelo selecionado proporcionou atender os critérios de normalidade dos resíduos e homogeneidade das variâncias para corretas inferências estatísticas.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Campos, J. C. C.; Leite, H. G. **Mensuração florestal: perguntas e respostas**. 3.ed. Viçosa: Editora UFV, 2009. 542 p.
- Curto, J. D.; Pinto, J. C. The corrected VIF (CVIF). *Journal of Applied Statistics*, v. 38, n. 7, p. 1499-1507, 2011. <https://doi.org/10.1080/02664763.2010.505956>.
- Cruz, G. S.; Ribeiro, R. B. S.; Gama, J. R. V.; Almeida, B. R. S.; Ximenes, L. C.; Gomes, K. M. A.; Bezzera, T. G. Ajuste e avaliação na estimativa volumétrica para *Lecythis lurida* (Miers) S.A. Mori em uma área de manejo florestal. **Advances in Forestry Science**, v. 6, n. 1, p.549-554, 2019. <https://doi.org/10.34062/afs.v6i1.7247>.
- Empresa Brasileira de Pesquisa Agropecuária - Embrapa. **Sistema brasileiro de classificação de solos**. 2.ed. Rio de Janeiro: Embrapa, 2006. 306p.
- Gaitan-Alvarez, J.; Moya, R.; Berrocal, A. The use of X-ray densitometry to evaluate the wood density profile of *Tectona grandis* trees growing in fast-growth plantations. **Dendrochronologia**, v. 55, p. 71-79, 2019. <https://doi.org/10.1016/j.dendro.2019.04.004>.
- Gujarati, D. N.; Porter, D. C. **Basic econometrics**. 5.ed. New York: McGraw-Hill, 2012. 923p.
- Indústria Brasileira de Árvores - IBÁ. **Relatório 2019**. São Paulo: IBÁ, 2019. 79p. Disponível <https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioanual2019.pdf>. Acesso em: 18 Mai. 2023.
- Mendonça, A. R.; Calegario, N.; Silva, G. F.; Carvalho, S. P. C. Growth and yield models for Eucalyptus stands obtained by differential equations. **Forestry Science**, v. 74, n. 5, p.364-370, 2017. <https://doi.org/10.1590/1678-992x-2016-0035>.
- Schmidt, A. F.; Finan, C. Linear regression and the normality assumption. **Journal of Clinical Epidemiology**, v. 98, p. 146-151, 2018. <https://doi.org/10.1016/j.jclinepi.2017.12.006>.
- Weiskittel, A. R.; Hann, D. W. H.; Kershaw, J. A.; Vanclay, J. K. **Forest growth and yield modeling**. Oxford: Wiley-Blackwell, 2011. 430p. <https://doi.org/10.1002/9781119998518>.
- Willy, P. H. A.; Alvim, C. A. N.; Yoshi, C. V. H.; Bechara, M. T. Problema da multicolinearidade na avaliação econométricas de bens por inferência estatística. **Revista Terra & Cultura**, v.38, p. 139-151, 2022.