

MODELAGEM MARGINAL CONJUNTA DA ALTURA E VOLUME PARA *Araucaria angustifolia*

JOINT MARGINAL MODELING OF HEIGHT AND VOLUME FOR *Araucaria angustifolia*

Luan Demarco Fiorentin¹, Wagner Hugo Bonat², Allan Libanio Pelissari¹,
Sebastião do Amaral Machado¹, Saulo Jorge Téó¹

¹Universidade Federal do Paraná, Curitiba, Paraná, Brasil – luanfiorentin@hotmail.com,
wbonat@ufpr.br, allanpelissari@gmail.com & samachado@ufpr.br

²Universidade do Oeste de Santa Catarina, Xanxerê, Santa Catarina, Brasil – sauloteo@yahoo.com.br

RESUMO

Variáveis mensuradas em florestas normalmente apresentam algum grau de correlação. Logo, ajustar modelos para estimar variáveis biométricas de forma independente não é a abordagem mais adequada. Assim, modelos multivariados ganham relevância devido à capacidade de quantificar associações entre variáveis respostas. Nesse contexto, o objetivo da presente pesquisa foi ajustar modelos lineares generalizados de covariância multivariada (MCGLMs) univariados e multivariados para estimar altura e volume de árvores. As variáveis altura (H), volume (V) e diâmetro (D) foram coletadas da *Araucaria angustifolia* em floresta nativa, localizada no estado de Santa Catarina, Brasil. Os MCGLMs foram ajustados para estimar H e V, em abordagem univariada e multivariada. O preditor linear dos modelos foi fixado previamente em função da covariável D, para ambas as variáveis. Devido a um aparente padrão de variância não constante das duas respostas, diferentes estruturas do preditor linear matricial foram testadas, com efeito da covariável D variando até polinômio de grau três. Ainda, um parâmetro de potência foi estimado nas duas abordagens, com a finalidade de obter uma função de variância para cada variável. Os parâmetros estimados nas abordagens univariadas e multivariadas foram similares. Em geral, o erro padrão dos parâmetros foi menor para os modelos multivariados, sendo consequência da correlação entre as variáveis respostas. Os resultados também sugeriram que uma função de variância Poisson-Gama composta é adequada para variável V, bem como uma função constante para variável H. O modelo mais adequado foi obtido com preditor linear matricial somente em função de um parâmetro de dispersão associado a uma matriz identidade.

PALAVRAS-CHAVE: Distribuição Tweedie, Floresta Ombrófila Mista, regressão multivariada.

ABSTRACT

Variables measured in a forest usually presented some degree of correlation. So, fitting models for estimating biometric variables in an independent way is not the most suitable approach. Thus, multivariate models become more interesting due to the ability of quantifying associations between response variables. In this context, the main objective of this research was to fit univariate and multivariate regression models based on multivariate covariance generalized linear models (MCGLM), for estimating the trees height and volume. Height (H), volume (V) and diameter (D) variables were obtained from *Araucaria angustifolia* specie in native forest, located in Santa Catarina state, Brazil. The MCGLMs were fitted for estimating H and V in univariate and multivariate approach. The linear predictor of the models was previously fixed as a function of covariate D for both responses. Due to the apparently non-constant covariance pattern for both variables, we tested different structures for the matrix linear predictor, where the effect of covariate D changing until a third-degree polynomial model. Still, a power parameter was estimated in both approaches where the aim was to obtain a variance function for each covariate. The estimated parameters of the univariate and multivariate approach were similar for some models. In general, the standard error of the parameters was lower for multivariate models, what is a consequence of the correlation between responses variables. The results also suggested that a composed Poisson-Gama variance function is suitable for V, but a constant function for H. The most suitable model was obtained with matrix linear predictor as a function of a dispersion parameter associated to identity matrix.

KEYWORDS: Tweedie distribution, Mixed Ombrophilous Forest, multivariate regression.

INTRODUÇÃO

A *Araucaria angustifolia* (Bert.) O. Ktze. é uma espécie nativa do Brasil e pertencente à família Araucariaceae, sendo a única representante desta família na flora brasileira (MARCHIORI, 2005). A *Araucaria angustifolia* ocorre em associações diversificadas, as quais compreendem grupamentos de espécies com características próprias, formando estágios sucessionais distintos (WEBER et al., 2017). Apesar da grande importância que essa espécie representa para as florestas nativas, especialmente as da região Sul do Brasil, ela está ameaçada de extinção atualmente, em decorrência da exploração excessiva, sem a reposição adequada (SCHEEREN et al., 1999).

Devido ao elevado valor econômico da madeira de *Araucaria angustifolia*, é fundamental quantificar os estoques volumétricos desta espécie nos ecossistemas florestais. O volume constitui uma das informações mais importantes para o conhecimento do potencial disponível em uma floresta, visto que, o volume individual fornece subsídios para a avaliação do estoque de madeira e para análise do potencial produtivo das florestas (THOMAS et al., 2006). Ainda, as variáveis diâmetro à altura do peito e altura total também são atributos fundamentais ao nível de árvore para muitos aspectos do manejo florestal.

No entanto, quantificar o volume e a altura das árvores em florestas nativas é uma atividade onerosa e que demanda tempo para coleta, enquanto a mensuração do diâmetro é um procedimento relativamente mais simples e de menor custo. Nesse contexto, é comum o ajuste de modelos estatísticos que expressam o comportamento de variáveis de difícil obtenção em função de variáveis como o diâmetro das árvores.

A abordagem mais comum para caracterizar a altura e o volume das árvores em florestas nativas ou plantadas consiste na aplicação de modelos lineares generalizados (FU et al., 2017), modelos não lineares (LAM et al., 2017) e modelos de efeitos-mistos (MEHTÄTALO et al., 2015). Assim, a modelagem das variáveis é normalmente realizada de forma individual, em que as variáveis são tratadas de forma independente. No entanto, como as árvores são organismos biológicos, é esperado que as variáveis mensuradas em um indivíduo apresentem algum tipo de associação, além de um padrão não constante de variância em função do aumento nas dimensões dos indivíduos. Portanto, tratar a altura e o volume da árvore de forma independente não é a abordagem mais adequada.

Os modelos de regressão multivariados são

generalizações dos modelos univariados e permitem estudar simultaneamente duas ou mais variáveis de interesse. Porém, a aplicação desses modelos é bastante restrita em pesquisas florestais (LAPPI, 2006). Recentemente, BONAT & JØRGENSEN (2016) desenvolveram os chamados modelos lineares de covariância multivariada (*Multivariate Covariance Generalized Linear Models* - MCGLM). Os MCGLM são uma classe de modelos estatísticos multivariados que permitem modelar variáveis respostas de diferentes naturezas simultaneamente. Assim, a partir dos MCGLM é possível descrever o comportamento de dados contínuos, como o diâmetro à altura do peito, altura total e volume individual, além de dados discretos, normalmente provenientes de algum tipo de contagem, como o número de árvores atacadas por algum tipo de praga e a abundância das espécies florestais.

A principal vantagem dos MCGLM está justamente na sua flexibilidade em modelar simultaneamente diversas variáveis respostas de diferentes naturezas, além de quantificar a associação entre elas por meio de parâmetros de correlação. A outra vantagem está relacionada com a matriz de covariância, especificada no chamado preditor linear matricial. Essa estrutura é bastante flexível e permite modelar dados correlacionados temporal e espacialmente, por meio de uma combinação linear de matrizes conhecidas, as quais podem descrever diferentes comportamentos e graus de associação. Dessa forma, é possível incluir diferentes estruturas de correlação entre as observações, além de funções de variância para diferentes tipos de variáveis respostas, e modelar de forma adequada a variabilidade natural dos dados (BONAT & JØRGENSEN, 2016; BONAT et al., 2017; BONAT, 2018).

No presente estudo, modelos univariados e multivariados foram ajustados para as variáveis altura total e volume individual das árvores em função do diâmetro à altura do peito. Ainda, diferentes estruturas da matriz de variância foram testadas, bem como a inclusão de funções de variância. A principal hipótese formulada na presente pesquisa foi que a inclusão da função de variância melhora o ajuste dos modelos de regressão univariados e multivariados. Portanto, os objetivos foram: I) avaliar o ajuste de modelos de regressão linear univariado e multivariado para caracterizar o comportamento das variáveis altura e volume da espécie *Araucaria angustifolia* em florestas nativas; e II) analisar a influência da função de variância na qualidade do ajuste dos modelos de regressão linear univariados e multivariados.

Na seção de material e métodos são descritos o conjunto de dados utilizados como motivação e uma breve

introdução aos MCGLM. Na sequência, os principais resultados e discussões do ajuste dos modelos univariados e multivariados são apresentados, seguidos pelas conclusões.

MATERIAL E MÉTODOS

Banco de dados

Os dados utilizados na presente pesquisa foram coletados no município de Xanxerê, Santa Catarina, Brasil. A região faz parte do Bioma Mata Atlântica, sobre o domínio da Floresta Ombrófila Mista (FOM). A floresta em estudo apresenta uma área total de, aproximadamente, 400 hectares. Atualmente, o fragmento florestal está sendo enriquecido com a espécie *Ilex paraguariensis* A. St.-Hill, com a finalidade de produzir Erva-mate.

A seleção das árvores para compor a amostra foi realizada de forma aleatória no fragmento florestal. O banco de dados foi constituído por 169 observações independentes, que corresponderam a indivíduos com grande amplitude de variação em suas dimensões. Em cada árvore, foram quantificadas as variáveis aleatórias diâmetro à altura do peito (D, em centímetros - cm), altura total (H, em metros - m) e volume individual (V, em metros cúbicos - m³) com casca.

O volume de cada seção da árvore foi calculado por meio do método de cubagem de Huber. O volume total com casca foi obtido pelo somatório do volume de cada seção com o volume da ponta da árvore (MACHADO & FIGUEIREDO FILHO, 2009). Na porção basal da árvore, a circunferência foi mensurada com fita métrica, e posteriormente transformada para diâmetro, onde as seções apresentaram comprimento menor, com amplitude de 0,1 – 0,3 m; 0,3 – 0,5 m; 0,5 – 0,7 m; 0,7 – 0,9 m; 0,9 – 1,3 m. Posteriormente, as seções foram definidas com comprimento de 2 m de comprimento, em que o Relascópio de Espelho de Bitterlich, modelo banda estreita, foi utilizado para medir indiretamente os diâmetros nas posições superiores do tronco das árvores. Assim, as medidas foram tomadas nas alturas de 0,2 m; 0,4 m; 0,6 m; 0,8 m; 1,1 m; 2,3 m; e a cada 2 m até à altura total da árvore.

Para realizar a medição do diâmetro com o Relascópio de Espelho de Bitterlich, a borda esquerda da árvore deve coincidir com a borda esquerda de uma das bandas ou estreitas do Relascópio, de modo que a borda direita da árvore está dentro das bandas ou estreitas. Assim, se faz a contagem do número de Unidades Relascópicas (UR). O valor obtido na contagem das UR foi aplicado na expressão

a seguir, com a finalidade de se estimar o diâmetro em diferentes alturas da árvore:

$$d_j = \frac{L}{50} UR_j,$$

em que: d_j é o diâmetro da árvore na altura j (cm); L é a distância da árvore ao observador (cm); 50 é um valor constante para o Relascópio de banda estreita; UR_j são as unidades relascópicas na altura j da árvore.

Modelos lineares generalizados de covariância multivariada (MCGLMs)

A modelagem de variáveis florestais por meio de modelos de regressão multivariados não é uma abordagem comum em pesquisas florestais. Portanto, essa subseção tem como finalidade principal apresentar a estrutura dos modelos lineares de covariância multivariada (MCGLM). Essa classe de modelos requer a especificação do valor esperado e a variância da variável resposta (BONAT & JØRGENSEN, 2016). Assim, uma formulação genérica do modelo para o caso multivariado é dada por:

$$E(\mathbf{Y}) = \mathcal{M} = \{g_1^{-1}(\mathbf{X}_1\boldsymbol{\beta}_1), \dots, g_R^{-1}(\mathbf{X}_R\boldsymbol{\beta}_R)\}$$

$$\text{Var}(\mathbf{Y}) = \mathbb{C} = \boldsymbol{\Sigma}_R \otimes \boldsymbol{\Sigma}_b,$$

em que: $\mathbf{Y}_{N \times R} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_R\}$ é uma matriz de respostas; $\mathcal{M}_{N \times R} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_R\}$ é uma matriz de valores esperados; \mathbf{X}_R é uma matriz de delineamento $N \times K_R$; $\boldsymbol{\beta}_R$ é uma matriz de vetores de regressão $K_R \times 1$; g_r é uma função de ligação diferenciável duas vezes e monótona; $\boldsymbol{\Sigma}_R \otimes \boldsymbol{\Sigma}_b = \text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1, \dots, \tilde{\boldsymbol{\Sigma}}_R) (\boldsymbol{\Sigma}_b \otimes \mathbf{I}) \text{Bdiag}(\tilde{\boldsymbol{\Sigma}}_1^T, \dots, \tilde{\boldsymbol{\Sigma}}_R^T)$ é o produto Kronecker generalizado; $\boldsymbol{\Sigma}_R$ é a matriz de covariância $N \times N$ dentro da resposta $r = 1, \dots, R$; $\boldsymbol{\Sigma}_b$ é a matriz de correlação entre variáveis respostas; $\tilde{\boldsymbol{\Sigma}}_R$ é a matriz triangular inferior da decomposição de Cholesky de $\boldsymbol{\Sigma}_R$; o operador Bdiag representa uma matriz bloco diagonal; e \mathbf{I} é a matriz identidade $R \times R$.

A especificação da matriz de covariância $\boldsymbol{\Sigma}_r$ para cada resposta é dada por:

$$\boldsymbol{\Sigma}_R = V(\boldsymbol{\mu}_r; \mathbf{p}_r)^{1/2} h\{\boldsymbol{\Omega}(\boldsymbol{\tau}_r)\} V(\boldsymbol{\mu}_r; \mathbf{p}_r)^{1/2},$$

em que: $V(\boldsymbol{\mu}_r; \mathbf{p}_r) = \text{diag}\{\vartheta(\boldsymbol{\mu}_r; \mathbf{p}_r)\}$ é uma matriz diagonal, cujas entradas principais denotam a função de variância $\vartheta(\boldsymbol{\mu}_r; \mathbf{p}_r)$ aplicada a cada elemento do vetor de respostas $\boldsymbol{\mu}_r$; \mathbf{p}_r é um vetor de parâmetros de potência; $\boldsymbol{\tau}_r$ é um vetor de parâmetros de dispersão; $h\{\boldsymbol{\Omega}(\boldsymbol{\tau}_r)\} =$

$\tau_0 Z_0 + \dots + \tau_D Z_D$; h é uma função de ligação de covariância; Z_d são matrizes conhecidas que descrevem a estrutura de covariância com $d = 0, \dots, D$; e τ_D é um vetor de parâmetros $(D + 1) \times 1$. A estrutura que especifica a média, $E(\mathbf{Y})$, é chamada de preditor linear ou não-linear, enquanto a estrutura que especifica a covariância do modelo, $\text{Var}(\mathbf{Y})$, é denominada de preditor linear matricial.

A função de ligação g é responsável por relacionar o preditor linear com os valores esperados da variável resposta. Escolhas apropriadas da função de ligação permitem garantir valores adequados para esperança da variável resposta. De forma semelhante, a função de ligação de covariância h associa o preditor linear matricial com a covariância da variável resposta.

A função de variância é um componente importante dos MCGLMs, uma vez que diferentes suposições são realizadas quanto a distribuição de probabilidade da variável resposta para diferentes valores da função de variância (BONAT & JØRGENSEN, 2016). O parâmetro de potência da função de variância $\vartheta(\cdot; p_r) = \mu_r^{p_r}$ caracteriza a família de distribuição Tweedie e os casos mais importantes são as distribuições Normal ($p = 0$), Poisson ($p = 1$), Poisson-Gamma composta ($1 < p < 2$), Gamma ($p = 2$) e Normal Inversa ($p = 3$).

Análise estatística dos dados

Os MCGLM foram utilizados para modelar as variáveis dendrométricas mensuradas no fragmento florestal. As duas variáveis respostas foram altura (H) e volume (V) das árvores, as quais são de natureza contínua. O diâmetro (D) foi utilizado como única variável preditora dos modelos, também de natureza contínua, com o seu efeito podendo variar até o terceiro grau.

Inicialmente, o ajuste dos modelos de regressão foi realizado de forma univariada, isto é, as variáveis H e V foram consideradas de forma independente e os modelos foram tratados separadamente. Em seguida, os modelos foram ajustados conjuntamente em uma abordagem multivariada. Nesse contexto, a finalidade foi analisar a influência da correlação entre as variáveis respostas nas estimativas pontuais e nos erros padrões dos parâmetros dos modelos.

As duas variáveis respostas têm como característica natural uma variância não constante ao longo do tempo, a qual pode ser atribuída ao natural processo de crescimento das árvores. Assim, é esperado que a variância das observações aumente em função das dimensões dos indivíduos. Para acomodar o padrão de variância não

constante, duas abordagens foram testadas: a primeira consistiu em modelar a variância diretamente no preditor linear matricial como uma função da covariável D_i , sendo o efeito representado por um polinômio de terceiro grau; e a segunda consistiu em incluir uma função de variância $\vartheta(\mu_r; p_r)$ no preditor linear matricial, a qual permitiu modelar diretamente a relação entre média e variância das variáveis respostas.

A qualidade do ajuste dos modelos foi comparada por meio da pseudo verossimilhança gaussiana (PV) e o pseudo critério de informação Bayesiano (PBIC). A PV corresponde a uma medida similar ao valor de log-verossimilhança do contexto de estimação por máxima verossimilhança. Portanto, maiores valores indicam melhores ajustes dos modelos. O PBIC é uma medida que apresenta a vantagem de punir modelos com maior quantidade de parâmetros, e quanto menor os seus valores melhores são os ajustes (BONAT, 2018).

O ajuste dos modelos de regressão foi realizado no software estatístico R (R CORE TEAM, 2009), por meio do pacote *mcglm*, versão 0.5.0 (BONAT, 2018). O pacote foi construído com uma interface bastante intuitiva e possui diversas funções que auxiliam no ajuste dos modelos e na construção do preditor linear e dos componentes do preditor linear matricial. Os gráficos foram construídos usando o pacote *ggplot2*, versão 3.1.0.

RESULTADOS E DISCUSSÃO

Análise exploratória de dados

A análise exploratória dos dados utilizados como motivação na pesquisa foi realizada com a finalidade de entender o comportamento das variáveis respostas e a sua relação com a covariável. Os histogramas apresentados na Figura 1 sugerem que as variáveis não possuem a mesma distribuição de probabilidade. A variável V apresentou uma distribuição fortemente assimétrica, o que sugere a inclusão de uma função de ligação logarítmica para o preditor linear do modelo de regressão. Ainda, é possível notar que a relação entre o volume médio das árvores e a sua variância não é constante para diferentes valores de diâmetros, com tendência de aumentar a variabilidade para árvores de maiores dimensões.

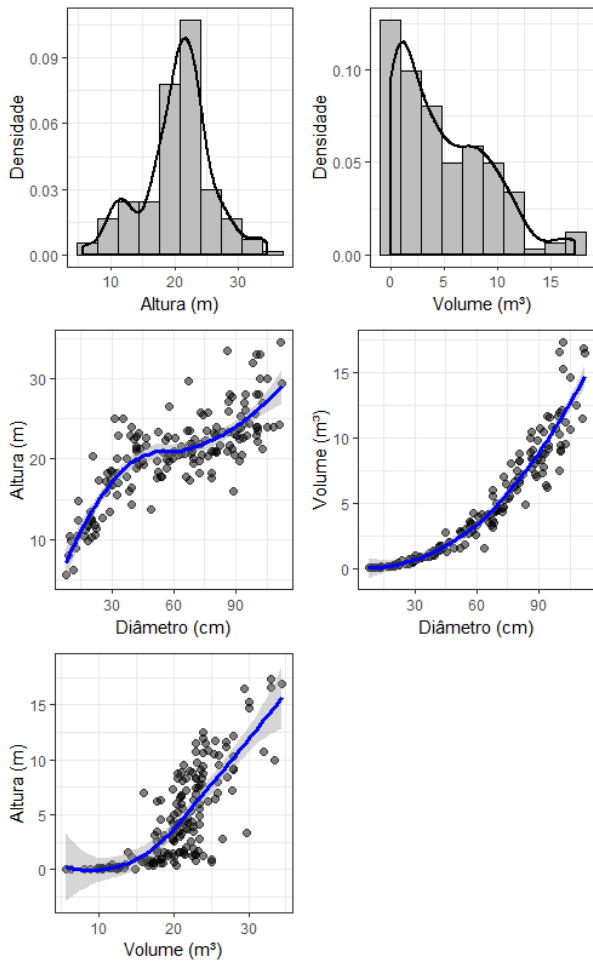


Figura 1. Histograma das variáveis resposta altura (H) e volume (V) e gráficos de dispersão entre as variáveis respostas e a covariável diâmetro (D). Linha sólida em azul é um polinômio local suavizado, com intervalos de 95% de confiança.

Esse resultado sugere a inclusão de uma função de variância Tweedie, uma vez que ela caracteriza diversas distribuições assimétricas contínuas, e permite modelar de forma adequada a relação entre média e variância das respostas. Por outro lado, a variável H apresentou uma distribuição aparentemente simétrica, o que sugere a inclusão de função de ligação identidade para o preditor linear, além de uma função de variância constante, que é uma suposição assumida quando a variável segue uma distribuição normal.

O maior interesse na presente pesquisa foi modelar os componentes do preditor linear matricial. Portanto, o preditor linear foi especificado em análise preliminar e a mesma estrutura de média foi utilizada no ajuste univariado e multivariado. O valor esperado da observação i da variável H_i foi especificado a partir do efeito cúbico da variável D_i , enquanto o valor esperado da variável V_i foi especificado com base em um efeito de segundo grau.

Assim, o preditor linear com a função de ligação para a variável volume V foi definido por:

$$E(H_i) = \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \beta_3 D_i^3,$$

$$E(V_i) = \exp(\beta_0 + \beta_1 D_i + \beta_2 D_i^2).$$

Além da função de variância utilizada para modelar a relação entre a média e a variância das respostas, estimada por meio de um parâmetro de potência, diferentes estruturas para a matriz de variância foram testadas no preditor linear matricial. Assim, as combinações lineares de matrizes conhecidas foram dadas por:

$$M1 = \Omega(\tau_r) = \tau_0 I,$$

$$M2 = \Omega(\tau_r) = \tau_0 I + \tau_1 Z,$$

$$M3 = \Omega(\tau_r) = \tau_0 I + \tau_1 Z + \tau_2 Z^2,$$

$$M4 = \Omega(\tau_r) = \tau_0 I + \tau_1 Z + \tau_2 Z^2 + \tau_3 Z^3,$$

em que: I é uma matriz de identidade de dimensão $N \times N$, sendo N o número de observações; Z é uma matriz com a diagonal principal constituída das observações dos diâmetros (D) das árvores, de dimensão $N \times N$, com seu efeito variando até o terceiro grau.

O resultado do ajuste dos modelos univariados e multivariados para as variáveis altura e volume da espécie *Araucaria angustifolia* estão apresentados nos tópicos a seguir.

Modelos univariados

Os parâmetros estimados e o erro padrão dos modelos univariados ajustados para as variáveis H e V estão apresentados nas Tabelas 1 e 2, respectivamente. Em geral, o intercepto dos preditores lineares para H não apresentaram significância, ao nível de 95% de confiança (fixado em todas as análises), enquanto os demais parâmetros foram significativos. Os parâmetros de dispersão foram significativos, mas com efeito negativo para $\hat{\tau}_1$ do modelo M3 e $\hat{\tau}_2$ do M4 para variável H .

Para variável V , os parâmetros do preditor linear foram todos significativos, enquanto os parâmetros de dispersão $\hat{\tau}_0$ e $\hat{\tau}_1$ do preditor linear matricial não foram para o modelo M2. O parâmetro estimado de potência não diferiu de 1,5 para o modelo M1, quando considerado a variável V . Esse valor caracteriza uma distribuição da família

Tweedie do tipo Poisson-Gama composta, indicando que a variância dos dados aumenta mais rapidamente que os valores de média.

Tabela 1. Estimativas pontuais e erros padrões dos parâmetros estimados para os modelos univariados de altura (H), para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	Parâmetro	Estimativa	Erro padrão
M1	β_0	0,3776	1,7087
	β_1	0,8937	0,1114
	β_2	-0,0131	0,0021
	β_3	0,00007	0,00001
	τ_0	8,7814	1,1021
M2	β_0	0,1853	1,4851
	β_1	0,9082	0,1019
	β_2	-0,0133	0,0020
	β_3	0,00007	0,00001
	τ_0	5,0865	1,8687
	τ_1	0,0600	0,0340
M3	β_0	0,0205	1,6592
	β_1	0,9235	0,1095
	β_2	-1,3679	0,002077
	β_3	0,00007	0,00001
	τ_0	9,3133	2,6558
	τ_1	-0,1236	0,1174
M4	β_0	0,3125	1,3065
	β_1	0,9053	0,0962
	β_2	-1,1340	0,0019
	β_3	0,00007	0,00001
	τ_0	-2,4738	2,3795
	τ_1	0,7560	0,2815
	τ_2	-0,0164	0,0065
	τ_3	0,0001	0,00004

Por outro lado, a não significância do parâmetro de potência do modelo M2 indicou uma distribuição Normal ($p = 0$), o que sugere que a variância é constante. Por fim, o intervalo de confiança do parâmetro de potência do modelo M3 sugeriu uma distribuição de Poisson ($p = 1$), indicando que a variância cresce de forma proporcional à média dos dados. Ainda, o modelo M4 não apresentou convergência, mesmo quando o algoritmo de estimação foi flexibilizado com a finalidade de facilitar o ajuste.

Tabela 2. Estimativas pontuais e erros padrões dos parâmetros estimados para os modelos univariados de volume (V), para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	Parâmetro	Estimativa	Erro padrão
M1	β_0	-3,1707	0,1108
	β_1	0,0988	0,0036
	β_2	-0,0004	0,00003
	τ_0	0,0830	0,0102
	p	1,6350	0,0829
M2	β_0	-3,1614	0,1110
	β_1	0,0986	0,0036
	β_2	-0,0004	0,00003
	τ_0	0,0465	0,3032
	τ_1	0,0009	0,0083
	p	1,4389	1,5765
M3	β_0	-2,9687	0,1131
	β_1	0,0939	0,0037
	β_2	-0,0004	0,00003
	τ_0	0,0742	0,0264
	τ_1	-0,0036	0,0017
	τ_2	0,00008	0,00005
	p	0,9376	0,3377

Modelos multivariado

Os parâmetros estimados e o erro padrão dos modelos multivariados ajustados para as variáveis H e V estão apresentados nas Tabelas 3 e 4, respectivamente. Apenas o modelo M1 para variável H apresentou todos os parâmetros significativos ao nível de 95% de probabilidade. No entanto, o modelo M1 e M3 para V apresentaram todos os parâmetros significativos, sendo que os parâmetros de potência resultaram em valores similares aqueles observados para o ajuste univariado. Ainda, também foi observado efeito negativo do parâmetro de dispersão $\hat{\tau}_1$ do modelo M3 para variável V. O modelo M4 não foi apresentado porque o algoritmo de estimação não convergiu.

Como apresentado previamente, os MCGLM são baseados em uma especificação marginal, isto é, são modelos que requerem a especificação de um componente para média e outro para a variância. A especificação marginal dos modelos apresenta a vantagem de não testar a hipótese nula dos parâmetros de dispersão $\tau = 0$ na borda do espaço paramétrico, como os testes de razão de verossimilhança, Wald e escore (BONAT, 2017). Assim, é possível observar efeitos negativos dos parâmetros

dispersão, como aqueles observados na presente pesquisa.

Tabela 3. Estimativas pontuais e erros padrões dos parâmetros estimados para os modelos multivariados de altura (H), para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	Parâmetro	Estimativa	Erro padrão
M1	β_0	5,7281	1,5568
	β_1	0,5201	0,0994
	β_2	-0,0061	0,0018
	β_3	0,00003	0,00001
	τ_0	9,3747	1,1124
M2	β_0	5,1783	1,4466
	β_1	0,5543	0,09526
	β_2	-0,0067	0,0018
	β_3	0,00003	0,00001
	τ_0	7,0741	2,0709
M3	τ_1	0,0353	0,0366
	β_0	4,4628	1,4632
	β_1	0,6141	0,0968
	β_2	-0,0079	0,0018
	β_3	0,00004	0,00001
	τ_0	7,6747	1,6314
	τ_1	-0,0036	0,0867
τ_2	0,0003	0,0010	

Tabela 4. Estimativas pontuais e erros padrões dos parâmetros estimados para os modelos multivariados de volume (V), para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	Parâmetro	Estimativa	Erro padrão
M1	β_0	-3,1528	0,1122
	β_1	0,0982	0,0036
	β_2	-0,0004	0,00003
	τ_0	0,0843	0,0095
	p	1,6152	0,0781
M2	β_0	-3,1541	0,1116
	β_1	0,0984	0,0036
	β_2	-0,0004	0,00003
	τ_0	0,0426	0,2130
	τ_1	0,0010	0,0059
	p	1,4080	1,0991
M3	β_0	-3,0332	0,1113
	β_1	0,0953	0,0036
	β_2	-0,0004	0,00003
	τ_0	0,0527	0,0063
	τ_1	-0,0034	0,0014
	τ_2	0,0001	0,00004
	p	0,7738	0,2631

Os modelos univariados e multivariados apresentam a mesma interpretação quanto aos efeitos estimados dos parâmetros do preditor linear e o preditor linear matricial. A diferença observada entre as estimativas pontuais e os erros padrões dos parâmetros ajustados em ambas as abordagens pode ser explicada pela correlação entre as duas variáveis respostas, como pode ser observado na Tabela 5. A correlação entre as respostas H e V foi significativa e próxima de 0,5 em todos os ajustes, indicando que essas variáveis compartilham informações. Como consequência, os erros padrões dos modelos multivariados foram menores em relação aos univariados para quase todos os parâmetros estimados.

Tabela 5. Correlação (ρ) estimada entre as variáveis respostas altura (H) e volume (V) no ajuste multivariado, para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	Estimativa	Erro padrão
M1	0,5075	0,0595
M2	0,4945	0,0604
M3	0,4715	0,0619

Qualidade do ajuste

A qualidade do ajuste dos modelos univariados e multivariados para as variáveis respostas H e V estão apresentadas na Tabela 6. Para tornar as medidas de qualidade comparáveis, os valores obtidos nos ajustes univariados foram somados. Os valores da pseudo verossimilhança (PV) tenderam a aumentar em função da inclusão de novos parâmetros no preditor linear matricial, indicando melhor desempenho para os modelos mais parametrizados. No entanto, o pseudo critério de informação Bayesiano (PBIC) penalizou a PV dos modelos em função do número de parâmetros.

Assim, os melhores resultados foram obtidos nos modelos mais simples, em que o preditor linear de matriz foi composto apenas de uma matriz identidade, com uma dispersão estimada comum a todas as observações. Porém, é importante ressaltar novamente que a função de variância estimada conjuntamente para a variável V e o parâmetro de potência (p) também sugeriu uma distribuição Poisson-Gama composta para a variável.

Esses resultados indicaram que é fundamental a inclusão de componentes para modelar a variância da variável resposta V, uma vez que ela não é constante e tende a crescer mais rapidamente que os valores de média. Ainda, é possível notar que a função de variância foi capaz

de absorver grande parte da variabilidade dos dados, evitando a necessidade de se modelar a variância em função da covariável D, conforme realizado nos modelos 2 a 4.

Tabela 6. Pseudo verossimilhança e pseudo critério de informação Bayesiano para os modelos de altura (H) e volume (V) univariados e multivariados, para espécie *Araucaria angustifolia* em floresta nativa.

Modelo	PV	PBIC
Modelo para H e V : univariado		
M1	-570,57	1199,31
M2	-567,80	1205,41
M3	-564,70	1210,84
Modelo para H e V : multivariado		
M1	-550,98	1165,95
M2	-549,95	1175,52
M3	-548,09	1183,44

O M1 mostrou-se mais adequado para modelar o comportamento das variáveis respostas no ajuste univariado e multivariado, principalmente devido a sua formulação simplificada em relação aos demais, aliado ao menor valor de PBIC. A partir desse modelo, intervalos com 95% de confiança para esperança das variáveis respostas foram construídos na Figura 2. É possível notar que os intervalos foram bastante estreitos, indicando alta precisão dos modelos ajustados.

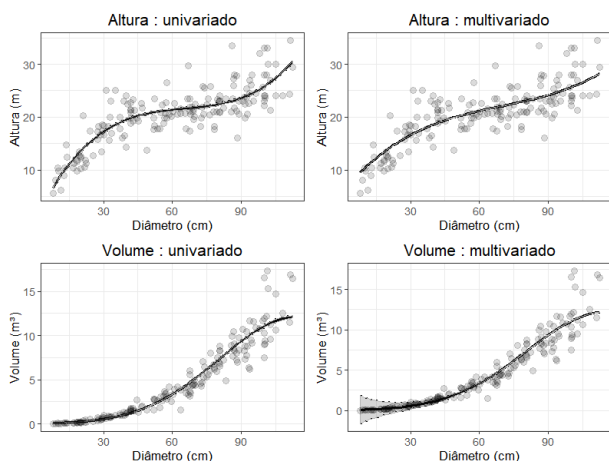


Figura 2. Intervalos de 95% de confiança para esperança das variáveis respostas altura (H) e volume (V) dos modelos univariados e multivariados, para espécie *Araucaria angustifolia* em floresta nativa.

CONCLUSÕES

Os modelos de regressão linear univariado e multivariados ajustados para descrever o comportamento das variáveis altura e volume da espécie *Araucaria angustifolia* mostram-se adequados e com elevado potencial de aplicação também a diversas espécies florestais. A função de variância aumenta a qualidade do ajuste dos modelos em ambas as abordagens e permite uma modelagem adequada da variância das variáveis respostas. Portanto, os modelos lineares generalizados de covariância multivariada é uma classe de modelos com excelente potencial de aplicação aos dados de biometria florestal.

AGRADECIMENTOS

O primeiro autor é apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com bolsa de pesquisa de Doutorado.

REFERÊNCIAS

BONAT, W.H. et al. Modelling the covariance structure in marginal multivariate count models: hunting in Bioko Island. **Journal of Agricultural, Biological, and Environmental Statistics**, v.22, n.4, p.446-464, 2017.

BONAT, W.H. Multiple response variables regression models in R: The mcglm package. **Journal of Statistical Software**, v.84, n.4, p.1-30, 2018.

BONAT, W.H.; JØRGENSEN, B. Multivariate covariance generalized linear models. **Journal of the Royal Statistical Society. Series C: Applied Statistics**, v.65, n.5, p.649-675, 2016.

BONAT, W.H. Modelling Mixed types of outcomes in additive genetic models. **The international journal of biostatistics**, v.13, n.2, p.1-16, 2017.

FU, L. et al. A generalized nonlinear mixed-effects height to crown base model for Mongolian oak in northeast China. **Forest Ecology and Management**, v.384, p.34-43, 2017.

LAM, T.Y. et al. Evaluating and modelling genus and species variation in height-to-diameter relationships for Tropical Hill Forests in Peninsular Malaysia. **Forestry**, v.90, n.2, p.268-278, 2017.

LAPPI, J.A. multivariate, nonparametric stem-curve prediction method. **Canadian Journal of Forest Research**, v.36, n.4, p.1017-1027, 2006.

MACHADO, S.A.; FIGUEIREDO FILHO, A. **Dendrometria**. Guarapuava: UNICENTRO, 2009.

MARCHIORI, J.N.C. **Dendrologia das gimnospermas**. 2005.

MEHTÄTALO, L. et al. Modeling height-diameter curves for prediction. **Canadian Journal of Forest Research**, v.45, n.7, p.826-837, 2015.

SCHEEREN, L.W. et al. Crescimento em altura de *Araucaria angustifolia* (Bert.) O. Ktze. em três sítios naturais, na região de Canela (RS). **Ciência Florestal**, v.9, n.2, p.23-40, 1999.

WEBER, K.S. et al. Variação volumétrica e distribuição espacial do estoque de carbono Em Floresta Ombrófila Mista. **Revista Acadêmica: Ciência Animal**, v.3, n.2, p.77-85, 2017.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, 2009.

THOMAS C. et al. Comparação de equações volumétricas ajustadas com dados de cubagem e análise de tronco. **Ciência Florestal**, v.16, n.3, p.319-327, 2006.

Recebido em 17-10-2019 Aceito em 25-10-2019