

**Modelos simétricos transformados bayesianos para descrever o crescimento em altura
de *eucalyptus urophylla***

**Bayesian transformed symmetric models to describe the height growth of *eucalyptus
urophylla*.models**

**Modelos bayesianos transformados simétricos para describir el crecimiento en altura de
eucalipto *urophylla***

Recebido: 07/07/2020 | Revisado: 15/07/2020 | Aceito: 16/07/2020 | Publicado: 22/07/2020

Kleber Napoleão Nunes de Oliveira Barros

ORCID: <https://orcid.org/0000-0003-2515-3292>

Universidade Estadual da Paraíba, Brasil

E-mail: kleberbarros@cct.uepb.edu.br

Mácio Augusto de Albuquerque

ORCID: <https://orcid.org/0000-0002-0113-9130>

Universidade Estadual da Paraíba, Brasil

E-mail: marcioaa@uepb.edu.br

José Antônio Aleixo da Silva

<http://orcid.org/0000-0003-0675-3524>

Universidade Federal Rural de Pernambuco

E-mail: jaaleixo@uol.com.br

Resumo

É abordado neste trabalho o modelo de crescimento em altura não linear de Chapman-Richards com distribuição dos erros seguindo a nova classe de modelos simétricos transformados e inferência Bayesiana para os parâmetros. O objetivo foi aplicar essa estrutura, via algoritmo de Metropolis-Hastings, com o intuito de selecionar a equação que melhor estimasse as alturas de clones de *Eucalyptus urophylla* provenientes de experimento implantado no Instituto Agrônomo de Pernambuco (IPA), na cidade de Araripina. O Polo Gesseiro do Araripe é uma zona industrial, situada no alto sertão pernambucano, que consome grande quantidade de lenha proveniente da vegetação nativa (caatinga) para calcinação da gipsita. Nesse cenário, há grande necessidade de uma solução, econômica e ambientalmente, viável que possibilite uma minimização da pressão sobre a flora nativa. O gênero *Eucalyptus* se apresenta como alternativa, pelo seu rápido desenvolvimento e versatilidade. A altura tem

se revelado fator importante na prognose de produtividade e seleção de clones melhores adaptados. Uma das principais curvas de crescimento em altura é o modelo de Chapman-Richards com distribuição normal para os erros. Os dados foram retirados de uma plantação, com 72 meses. Foram realizadas as inferências e diagnósticos para modelo transformado e não transformado com diversas distribuições simétricas. Após a seleção da melhor equação, foram mostrados alguns gráficos da convergência dos parâmetros e outros que comprovam o ajuste aos dados do modelo simétrico transformado t de Student com 5 graus de liberdade utilizando inferência Bayesiana nos parâmetros.

Palavras-chave: Florestas energéticas; Modelo de Chapman-Richards; Transformação de dados.

Abstract

It is presented in this work the growth model nonlinear Chapman-Richards with distribution of errors following the new class of symmetric models processed and Bayesian inference for the parameters. The objective was to apply this structure, via Metropolis-Hastings algorithm, in order to select the equation that best predicted heights of clones of Eucalyptus urophylla experiment established at the Agronomic Institute of Pernambuco (IPA) in the city of Araripina. The Gypsum Pole of Araripe is an industrial zone, located on the upper interior of Pernambuco, which consumes large amount of wood from native vegetation (caatinga) for calcination of gypsum. In this scenario, there is great need for a solution, economically and environmentally feasible that allows minimizing the pressure on native vegetation. The genus Eucalyptus presents itself as an alternative for rapid development and versatility. The height has proven to be an important factor in prognosis of productivity and selection of clones best adapted. One of the main growth curves, is the Chapman-Richards model with normal distribution for errors. However, some alternatives have been proposed in order to reduce the influence of atypical observations generated by this model. The data were taken from a plantation, with 72 months. After selecting the best equation, was shown some convergence of graphics and other parameters that show the fit to the data model transformed symmetric Student's t with 5 degrees of freedom in the parameters using Bayesian inference.

Keywords: Energetic forests; Chapman-Richards model; Data transformation.

Resumen

El modelo de crecimiento de altura no lineal de Chapman-Richards se aborda en este trabajo con distribución de errores siguiendo la nueva clase de modelos simétricos transformados y la

inferencia bayesiana para los parámetros. El objetivo era aplicar esta estructura, utilizando el algoritmo Metropolis-Hastings, para seleccionar la ecuación que mejor estimaba las alturas de los clones de *Eucalyptus urophylla* a partir de un experimento implementado en el Instituto Agronómico de Pernambuco (IPA), en la ciudad de Araripina. El Polo Gesseiro do Araripe es un área industrial, ubicada en las tierras altas de Pernambuco, que consume una gran cantidad de leña de la vegetación nativa (caatinga) para la calcinación del yeso. En este escenario, existe una gran necesidad de una solución económica y ambientalmente viable que minimice la presión sobre la flora nativa. El género *Eucalyptus* se presenta como una alternativa, debido a su rápido desarrollo y versatilidad. La altura ha demostrado ser un factor importante en el pronóstico de la productividad y la selección de clones mejor adaptados. Una de las principales curvas de crecimiento en altura es el modelo Chapman-Richards con distribución normal de errores. Sin embargo, se han propuesto algunas alternativas para reducir la influencia de las observaciones atípicas generadas por este modelo. Los datos fueron tomados de una plantación de 72 meses. Después de seleccionar la mejor ecuación, se mostraron algunos gráficos de la convergencia de los parámetros y otros que demuestran el ajuste a los datos del modelo transformado simétrico del Estudiante con 5 grados de libertad utilizando la inferencia bayesiana en los parámetros.

Palabras clave: Bosques energéticos; Modelo Chapman-Richards; Transformación de datos.

1. Introdução

A madeira é uma importante fonte renovável de energia. O manejo adequado de florestas e o uso racional da madeira como energia podem promover a oferta de energia renovável e de grande qualidade ecológica, diferente dos derivados de petróleo e carvão mineral, que são fontes esgotáveis de energia e que difundem na atmosfera, por ocasião da queima, o carbono entre outros poluentes, causando problemas de aquecimento global (Gatto, 2003). A mineral gipsita é um hidrato do sulfato de cálcio ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) amplamente utilizado na fabricação de cimento e revestimentos em habitações (Araujo et al., 2016). O gesso é um pó branco, derivado da gipsita, que ao ser misturado à água, forma uma pasta que endurece rapidamente. É empregado na composição de estuques, fabricação de estátuas, na imobilização de membros fraturados etc. Serve também para envolver os moldes de cera dos dentes feitos pelos dentistas e para muitos outros fins. Ferreira (2017) indicam alguns outros usos da gipsita e derivados em alguns dos setores produtivos e científicos da sociedade. A cadeia produtiva do Araripe é, altamente, dependente do gesso, sendo gerados 13,2 mil

empregos diretos e 66 mil indiretos, resultantes da extração em 42 minas, 140 indústrias de calcinação e, aproximadamente, 726 indústrias de pré-moldados, com renda anual de cerca de US 300 milhões (Sindusgesso, 2009). Por outro lado, O setor de produção secundária, no qual se enquadram as calcinadoras do Araripe, consome, predominantemente, a biomassa florestal como fonte energética. A degradação ambiental no Araripe está fortemente associada ao beneficiamento mineral, predominantemente, a gipsita. Em consequência, os estoques lenheiros da região foram reduzidos ao longo do tempo. Desta forma, as indústrias são obrigadas a adquirir lenha de estados vizinhos, encarecendo os custos e alastrando o problema para outras áreas (Gadelha et al., 2010). Diante da pressão que a flora nativa vem sofrendo e como o aumento da oferta de outras fontes energéticas é inviável alguma medida para aumentar a oferta de lenha na região tem prioridade. Os manejos sustentáveis de florestas de rápido crescimento em altura passam a ter uma importância na minimização da pressão sobre a mata nativa, pois podem suprir a demanda de lenha da indústria gesseira e, assim, atenuar a devastação florestal da região pela oferta de madeira em menor espaço de tempo e maior volume. Desta forma em função dos altos índices de produtividade e características energéticas, como densidade da madeira e poder calorífico, árvores do gênero *Eucalyptus* vem sendo utilizados para a formação de florestas com fins energéticos e, desta forma, diminuir a pressão sobre as florestas nativas. As principais espécies de eucaliptos utilizadas são: *Eucalyptus grandis*, *E. urophylla*, *E. urograndis* (híbrido), *E. camaldulensis*, *E. citriodora*, *E. cloeziana*, *E. globulus*, *E. maculata*, *E. paniculata*, *E. pellita*, *E. pilularis*, *E. saligna* e *E. tereticornis* (Lima et al., 2010). O objetivo foi aplicar essa estrutura, via algoritmo de Metropolis-Hastings, com o intuito de selecionar a equação que melhor estimasse as alturas de clones de *Eucalyptus urophylla* provenientes de experimento implantado no Instituto Agrônomo de Pernambuco (IPA), na cidade de Araripina.

2. Material e Métodos

A Região do Araripe, localizada na Mesorregião do Sertão Pernambucano representa 18,8% do território estadual com uma área de 1857,9 km². As plantas utilizadas na pesquisa foram oriundas de experimento realizado na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA), localizada na Chapada do Araripe no município de Araripina. As coordenadas geográficas de posição possui 07°27'37"S e 40° 24'36"W e altitude de 831 metros. A precipitação média anual é de 650 mm, com concentração de 70% entre os meses de dezembro a março, provocando deficiências hídricas ao longo do ano. O clima é do tipo Bshw' (semi-árido quente e seco com chuvas de verão-

outono, segundo classificação climática e Köppen-Geiger), registrando temperatura média anual de 24°. O solo é do tipo latossolo vermelho-amarelo (Araújo, 2004).

O Módulo de Experimentação Florestal para a região do Araripe, que foram implantado em março de 2002, em uma área de 2,352 ha. As parcelas contêm 49 plantas, sendo que as áreas úteis foram compostas de 25 plantas, contendo 15 clones, entre híbridos e espécies do gênero *Eucalyptus* plantadas no espaçamento 3x2 m, perfazendo 21 metros de largura e 14 de comprimento num total de 294 m² de parcela.

Os dados do modulo de experimentação florestal foram coletados aos dois meses e a seguir passaram a ser coletados a cada seis meses até serem completados 78 meses. Dentre as diversas variedades e cruzamentos plantados no instituto, selecionou-se o híbrido de *Eucalyptus urophylla* com cruzamento natural, o qual retornou 882 observações de altura de árvores homogêneas provenientes de experimento inteiramente ao acaso.

Modelagem

Seja $\beta = (\varphi, \tau)$, em que: $\varphi = \{\alpha, \beta, k, \delta, \theta, \lambda\}$ é o vetor de parâmetros do preditor não linear definido pelo modelo de Chapman-Richards

$$n_i = \alpha \left\{ 1 - \exp[-k(x_i - \delta)] \right\}^{-\theta} \quad (1)$$

com $\theta = (1 - y)^{-1}$ e $\tau = \phi^{-1}$ é o parâmetro de precisão dos modelos simétricos transformados. Note que este é o modelo de Chapman-Richards supondo $\beta = 1$ e $\delta = 0$. A função de verossimilhança (14) para os modelos simétricos transformados seguindo a equação (1) pode ser escrita como

$$L(\varphi, \tau | y_i^{(\lambda)}, x_i) = \tau^{n/2} \prod_{i=1}^n h \left[\left(\frac{y_i^{(\lambda)} - n_i}{\tau^{-1}} \right)^2 \right] |y_i|^{\lambda-1} \quad (2)$$

em que: n_i é o preditor não linear (14) e $y_i^{(\lambda)}$ é a transformação usual de Box-Cox (3) com $\lambda_2 = 0$.

Foram utilizadas distribuições a priori para φ e τ nas formas:

$$\varphi \sim N(\hat{\varphi}, \hat{\phi}) \text{ e } \tau \propto \hat{\phi}^{-1} \quad (3)$$

em que: $\hat{\varphi}$ e $\hat{\phi}$ são as estimativas de máxima verossimilhança para φ e ϕ , respectivamente. A partir das observações das alturas medidas entre 2 e 78 meses de vida das árvores (num total de 14

medidas ao longo do tempo), aplicou-se os modelos simétricos transformados via logaritmo da função (13) e se obteve estimativas clássicas para o vetor de parâmetros via método de máxima verossimilhança supondo diversas distribuições simétricas. Os graus de liberdade da distribuição t de Student foram estimados por verossimilhança perfilada.

A verossimilhança (13) e as densidades a priori (14), não apresentam forma analítica para a densidade a posteriori, dessa forma, utilizamos o algoritmo de Metropolis-Hastings para obter estimativas das densidades condicionais a posteriori de cada parâmetro e realizamos as inferências do modelo obtido. Metropolis-Hastings foi escolhido por ser de fácil implementação, enquanto o algoritmo amostrador de Gibbs requer distribuições condicionais a posteriori, que neste caso não são de fácil determinação.

Modelagem Estatística

Existem certos conjuntos de dados em que a suposição de normalidade na distribuição dos erros não é possível, mesmo ao se empregar algum artifício. Isso pode ocorrer mesmo para dados que apresentem simetria. Desta forma, se for aplicado um modelo, normal para esses dados, pode ser que se encontre diversos pontos extremos não explicados pelo modelo, designados “outliers”. Uma solução seria utilizar distribuições que pudessem minimizar a influência de tais observações. É com esse intuito que se define uma nova classe que contém diversas distribuições desse tipo. A variável aleatória Y tem distribuição simétrica, com suporte em \mathcal{R} , com parâmetros de locação $\mu \in \mathbb{R}$ e de escala $\phi > 0$, se sua função densidade é da forma

$$f(y_i; \mu_i, \phi_i) = \frac{1}{\sqrt{\phi_i}} h \left[\left(\frac{y_i - \mu_i}{\sqrt{\phi_i}} \right)^2 \right], \quad y_i \in \mathbb{R} \quad (5)$$

para alguma função $h(\cdot)$ denominada função geradora de densidade, com $h(u) > 0$, para $u > 0$ e $\int_0^\infty u^{1/2} h(u) du = 1$. Essa condição é necessária e suficiente para que $f(y_i; \mu_i, \phi_i)$ seja uma função densidade de probabilidade. Denota-se $Y_i \sim S(\mu_i, \phi_i)$ e denomina-se Y_i de variável aleatória simétrica. Um modelo estatístico de regressão que segue a distribuição (14), com a função de ligação μ_i é denominado Modelo Simétrico. A nova classe de modelos simétricos permite ajustar inúmeros modelos e uma ampla variedade de dados, que não são passíveis de modelagem via teoria clássica dos modelos não lineares. Esta nova classe de modelos inclui distribuições tais como normal, t de Student, de Laplace ou exponencial dupla (ED), logísticas tipo I e II, Cauchy e os modelos normais contaminados (Cysneiros et al., 2005). Observa-se na Tabela 1 estão dispostas as formas algébricas de $h(u)$ para alguns membros da classe de distribuições simétricas.

Tabela 1: Funções geradoras de densidade de algumas distribuições simétricas.

Distribuição	Função Geradora de Densidade
Normal	$h(u) = (2\pi)^{-1} \exp(-u/2)$
Cauchy	$h(u) = [\pi(1 + u)]^{-1}$
Laplace (ED)	$h(u) = 2^{-1} \exp(-\sqrt{ u })$
Logística tipo I	$h(u) = c \frac{\exp(-u)}{[1 + \exp(-u)]^2}, \quad c \approx 1,4843$
Logística tipo II	$h(u) = \frac{\exp(-u^{1/2})}{[1 + \exp(-u^{1/2})]^2}$
t_v de Student	$h(u) = v^{v/2} B(1/2, v/2)^{-1} (v + u)^{-(v+1)/2}, \quad v \in \mathbb{Z}_+$

Fonte: Autores.

O uso de transformações em análise de regressão é muito comum e pode ser útil quando o modelo original não satisfaz as suposições usuais. A família de transformação proposta por Box e Cox (1964) representa uma ferramenta útil para estatísticas aplicadas tratando dos aspectos de homogeneidade de discrepância, aditividade e normalidade. Considere a família paramétrica de transformações \mathcal{G} da variável Y expressa por

$$\mathcal{G} = \{\Lambda = \Lambda(Y, \lambda) = Y \sim S(\mu, \phi), \lambda = (\lambda_1, \dots, \lambda_m)^T\},$$

em que: μ e ϕ são os, respectivos, parâmetros de locação e escala do modelo simétrico e λ é o vetor de m parâmetros de transformação de Y . Seja $Y^{(\lambda)}$ a imagem de Λ pela função

$$Y^{(\lambda)} = \Lambda(Y, \lambda). \quad (5)$$

Para cada conjunto de parâmetros λ , $Y^{(\lambda)}$ é uma função monotônica de Y . A transformação mais utilizada é a de Box e Cox (1964) expressas por:

$$Y^{(\lambda)} = \begin{cases} \frac{(Y - \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0, \lambda_2 \geq 0 \\ \log(\lambda_1), & \lambda_1 = 0, \lambda_2 \geq 0 \end{cases} \quad (6)$$

em que: $\lambda = (\lambda_1, \lambda_2)$, tal que $\forall y \in Y \Rightarrow \min(y + \lambda_2) > 0$. Na prática, quando não há dados negativos se escolhe $\lambda_2 = 0$.

Cordeiro e Andrade (2007) desenvolveram uma nova classe de modelos lineares generalizados transformados (MLGTs) que visam estender os modelos de Box e Cox (1964) para os modelos da família exponencial. Cordeiro e Andrade (2009) propõem outra classe de modelos de regressão denominada modelos simétricos transformados (MSTs), que também generalizam os modelos de Box e Cox. Para introduzir uma estrutura regressora na classe de distribuições simétricas, toma-se

a componente sistemática para o vetor da média $\mu = E(Y^{(\lambda)})$ expressa por (Cordeiro & Andrade, 2009)

$$g(\mu) = \eta_i(\beta) = h(x_i; \beta, \phi) \quad (7)$$

sendo $g(\cdot)$ a função de ligação conhecida, monótona e diferenciável até a segunda ordem, \mathbf{X} é a matriz $n \times p$ de covariáveis de posto completo $p < n$ e $\beta = (\beta_1, \dots, \beta_n)^T$, ϕ e λ são os $p + 2$ parâmetros desconhecidos a serem estimados.

Definem-se os modelos simétricos transformados (MSTs) pelas famílias de transformações (5) e distribuição (14) assumindo componente sistemática (4). Isto é, os MSTs assumem que existe algum valor de λ tal que as variáveis aleatórias $Y_1^{(\lambda)}, \dots, Y_n^{(\lambda)}$ podem ser tratadas como independentes e identicamente distribuídas com componente sistemática (4) e aleatória (1).

Conforme Cordeiro e Andrade (2009) a log-verossimilhança para os parâmetros é expressa por

$$\ell(\beta, \phi, \lambda) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log \left\{ h \left[\phi^{-1} (y_i^{(\lambda)} - \mu_i)^2 \right] \right\} + \sum_{i=1}^n \log [J(\lambda, y_i)]. \quad (8)$$

Considerando os modelos simétricos com transformação segundo Box e Cox, tem-se que $J(\lambda, y_i) = |y_i|^{\lambda-1}$.

A função (8) pode ser maximizada incondicionalmente usando algum pacote estatístico como o SAS ou R. Um intervalo de confiança assintótico, com nível de significância γ para cada parâmetro θ_i , é expresso por $IC_{(\theta_i, 100(1-\gamma)\%)} = \hat{\theta}_i \pm \tilde{\alpha}/2\sqrt{\kappa^{\theta_i}}$, em que: $\kappa^{\theta_i} = \kappa^{\theta_i, \theta_i}$ é o i -ésimo componente da diagonal da inversa da informação esperada de Fisher, $K_{\theta_i}^{-1}$ e $z_{\gamma/2}$ é o quantil $1 - \gamma/2$ da distribuição normal padrão.

Critérios de Seleção de modelos

Muitos procedimentos têm sido propostos com o intuito de comparar modelos. Chatfield & Xing (2019) apresentou seu método de identificação de modelos. Esse critério de informação de Akaike (AIC) é uma estatística bem conhecida e de fácil interpretação para seleção de modelos de regressão. Desta forma, para comparar todos os modelos não transformados e transformados ajustados aos dados, pode-se usar o critério de informação de Akaike definido por:

$$AIC = -2\ell(\hat{\theta}) + 2r, \quad (9)$$

em que: $\ell(\hat{\theta})$ é a log-verossimilhança, dada pela equação (8) no caso dos modelos simétricos transformados, maximizada em $\theta = (\beta, \phi, \lambda)$. Já $r = p + 2$ ou $r = p + 1$, para os modelos simétricos transformados e não-transformados, respectivamente, sendo p o número de parâmetros β do preditor não-linear η_i . A equação com o menor valor do AIC, entre todos os modelos ajustados, pode ser considerada como a que melhor explica os dados.

Outro critério de seleção de modelos popular é o BIC (*Bayesian Information Criterion*), proposto por Schwarz (1978). Este critério pondera entre o número de parâmetros do modelo e a função de log-verossimilhança. O melhor modelo, segundo este critério, será aquele que apresentar o menor BIC expresso por:

$$BIC = -2\ell(\hat{\theta}) + r \cdot \log(n), \quad (10)$$

Cordeiro e Andrade (2009) calculam o erro quadrático médio (EQM) e o erro percentual absoluto médio (EPAM), para avaliar o melhor modelo. Estes critérios são expressos, respectivamente, por

$$EQM = \frac{100\%}{n s_{y(\hat{\lambda})}^2} \sum_{i=1}^n (y_i^{(\hat{\lambda})} - \hat{\eta}_i)^2 \quad (11)$$

e

$$EPAM = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i^{(\hat{\lambda})} - \hat{\eta}_i}{y_i^{(\hat{\lambda})}} \right|, \quad (12)$$

em que: $\bar{s}_{y(\hat{\lambda})}^2$ é a variância amostral de $y^{(\hat{\lambda})}$. O modelo que apresenta menores valores de EQM e/ou EPAM produz a característica desejável de possuir menores erros de ajustes.

Os resíduos de Pearson definidos em Cordeiro e Andrade (2009) para os modelos simétricos transformados são dados por

$$r_i = \frac{y_i^{(\lambda)} - \hat{\eta}_i}{\sqrt{k_2 \phi}} \quad (13)$$

em que: $k_2 = -2\psi'(0)$, sendo $\psi(\cdot)$ a parte real da função característica da distribuição. No caso da distribuição logística tipo II, por exemplo, $k_2 = \frac{\pi^2}{3}$. Com os resíduos podemos

reproduzir o gráfico $\bar{r}_i \times i$, em que i é o índice em que cada observação aparece. Assim, a presença de resíduos fora da faixa ± 2 indica a presença de *outliers*.

Inferência Bayesiana

Na inferência clássica, também denominada frequentista, os parâmetros, uma vez estimados da amostra, tornam-se valores fixos, dos quais se exprimem todo conhecimento estatístico acerca da população. A inferência Bayesiana, no entanto, trata o problema da estimação de parâmetros por uma abordagem diferenciada. A amostra, para a inferência Bayesiana, é um dos componentes da estimação, representada pela função de verossimilhança, denotada $L(\theta|x)$ e obtida por:

$$L(\theta|x) = \prod_{i=1}^n f(x_i|\theta), \quad (14)$$

sendo $f(x_i|\theta)$ é uma função densidade avaliada no ponto x_i .

A outra parte, designada distribuição a priori, $\pi(\theta)$, representa a informação subjetiva do especialista/pesquisador acerca do parâmetro ou problema em questão (BOX; TIAO, 1973). Esta informação sobre o conjunto paramétrico é atualizada pelo conhecido teorema de Bayes:

$$\pi(\theta|x) = \frac{L(\theta|x)\pi(\theta)}{\int L(\theta|x)\pi(\theta) d\theta}, \quad (15)$$

em que: $\pi(\theta|x)$ é a distribuição θ atualizada, denominada densidade a posteriori. Na prática, a integral do denominador, sendo uma constante e podendo ser obtida posteriormente, é suprimida e a regra (12) pode ser obtida, proporcionalmente, por:

$$\pi(\theta|x) \propto L(\theta|x)\pi(\theta). \quad (16)$$

Quando θ é um vetor, a forma analítica da expressão (15) ou (16) se torna difícil, ou comumente impossível, de ser encontrada. Neste caso, recorre-se a algoritmos intensivos, tal como o de Metropolis-Hastings (Paulino et al., 2003). Metropolis e associados, na década de 1950 (Metropolis et al., 1953), desenvolveram este método, empregado em Matemática, Física e Estatística para obter uma sequência de valores que convergem para uma distribuição de equilíbrio $\pi(\cdot)$. Na década de 1970, Hastings (Hastings, 1970) entre outros generalizaram o procedimento (Little & Rubin, 2019). Geman & Geman (1984) desenvolveram o amostrador

de Gibbs. Após 35 anos da sua criação, os pesquisadores Bayesianos redescobriram-no e então passou a ser a principal ferramenta computacional em inferência estatística (Ntzoufras, 2009). Devido à alta flexibilidade e generalidade, o algoritmo de Metropolis-Hastings (MH) se torna, particularmente, interessante quando, a amostragem direta da distribuição de probabilidade é complexa ou quando não possui forma analítica fechada como é típico em inferência Bayesiana (Ahelegbey, Billio & Casarin, 2016).

Existem muitas recomendações conflitantes, quanto à quantidade de iterações do MH e ao número de simulações para cada parâmetro (cadeias). Alguns autores recomendam amostrar muitas cadeias de tamanho pequeno (Gelfand & Smith, 1990), outros trabalham com diversas cadeias longas (Gelman & Rubin, 1992) e outra vertente utiliza uma única cadeia muito longa (Geyer, 1992). Quanto aos valores iniciais, Gelman & Rubin (1992), sugerem que, ao se utilizar a metodologia das múltiplas cadeias, os valores iniciais sejam escolhidos de forma dispersa (Gilks, 1996). É recomendado que se descarte as primeiras iterações (burn-in) para eliminar a influência dos valores iniciais. O método mais utilizado para determinar o tamanho do burn-in é a inspeção visual (Gilks, 1996). Geyer (1992) sugere que o cálculo do tamanho do burn-in é irrelevante e utiliza entre 1% e 2% do tamanho total da cadeia. Entretanto, existem diversas formas de obter o tamanho do burn-in de forma analítica (Cowles & Carlin, 1996). Existem diversas variações do MH. Em sua versão genérica, o algoritmo Metropolis é descrito a seguir.

Suponha que $\pi(\theta|x)$ seja a densidade de interesse. Seja então $q(\theta|\cdot)$, a menos de uma constante, uma densidade, conhecida como distribuição proposta para estimá-la. O algoritmo pode ser implementado da seguinte forma:

- (i) especifique um valor inicial $\theta^{(0)}$;
- (ii) para $i = 1, \dots, T$:
 - (a) faça $\theta = \theta^{(i-1)}$;
 - (b) proponha um novo valor θ' da distribuição proposta $q(\theta'|\theta)$;
 - (c) Calcule
$$A(\theta', \theta) = \min\left(1, \frac{\pi(\theta'|x)q(\theta|\theta')}{\pi(\theta|x)q(\theta'|\theta)}\right)$$
 - (d) Atualize $\theta^{(i)} = \theta'$ com probabilidade $A(\theta', \theta)$;
 - (e) Caso contrário, faça $\theta^{(i)} = \theta$.

Paulino et al. (2003) consideram a análise de modelos lineares e análise de variância sob o enfoque Bayesiano. Dey et al. (2000) reúnem, vários artigos que estendem o problema para os modelos lineares generalizados. Box e Tiao (1973) definem distribuições a priori e

posteriori para λ , supondo resposta normal, para modelos com transformação uniparamétrica e biparamétrica realizando análise dos efeitos das estimativas após transformação. Pericchi (1981) trata o mesmo problema, supondo uma priori não-informativa alternativa.

Diagnósticos Bayesianos

Os diagnósticos Bayesianos constituem uma forma de avaliação quantitativa da convergência das cadeias obtidas a partir da simulação de um algoritmo MCMC. Dois dos principais critérios com esta finalidade são o critério de Raftery e Lewis e o diagnóstico de Heidelberger e Welch (COWLES & CARLIN, 1996).

(1) Critério de Raftery e Lewis

Define quantidades mínimas que deveriam ser utilizadas para se atingir baixa autocorrelação em cada parâmetro. Define também, o fator de dependência I . Se $I \geq 5$, há indícios de forte autocorrelação que pode vir de uma escolha fraca dos valores iniciais.

(2) Diagnóstico de Heidelberger e Welch

Este teste utiliza a estatística de Cramer-von-Mises para testar a hipótese de que a amostra tomada siga uma distribuição estacionária. É aplicado a toda cadeia e depois aos 10%, 20%, ... primeiras iterações até que a hipótese nula seja aceita, ou até atingir 50% das iterações, quando o teste falha, caso em que não se pode dizer que houve estacionariedade na cadeia.

Modelagem

Seja $\beta = (\varphi, \tau)$, em que: $\varphi = (\alpha, \beta, \kappa, \delta, \theta, \lambda)$ é o vetor de parâmetros do preditor não linear definido pelo modelo de Chapman-Richards

$$\eta_i = \alpha \{1 - \beta \exp[-\kappa(x_i - \delta)]\}^{-\theta} \quad (17)$$

com $\theta = (1 - \gamma)^{-1}$ e $\tau = \phi^{-1}$ é o parâmetro de precisão dos modelos simétricos transformados. Note que este é o modelo de Chapman-Richards supondo $\beta = 1$ e $\delta = 0$. A função de verossimilhança (14) para os modelos simétricos transformados seguindo a equação (1) pode ser escrita como

$$L(\boldsymbol{\varphi}, \tau | y_i^{(\lambda)}, x_i) = \tau^{n/2} \prod_{i=1}^n h \left[\left(\frac{y_i^{(\lambda)} - \eta_i}{\tau^{-1}} \right)^2 \right] |y_i|^{\lambda-1} \quad (18)$$

em que: η_i é o preditor não linear (14) e $y_i^{(\lambda)}$ é a transformação usual de Box-Cox (6) com $\lambda_2 = 0$. Foram utilizadas distribuições a priori para $\boldsymbol{\varphi}$ e τ nas formas:

$$\boldsymbol{\varphi} \sim N(\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\phi}}) \quad e \quad \tau \propto \hat{\boldsymbol{\phi}}^{-1} \quad (19)$$

em que: $\hat{\boldsymbol{\varphi}}$ e $\hat{\boldsymbol{\phi}}$ são as estimativas de máxima verossimilhança para $\boldsymbol{\varphi}$ e $\boldsymbol{\phi}$, respectivamente. A partir das observações das alturas medidas entre 2 e 78 meses de vida das árvores (num total de 14 medidas ao longo do tempo), aplicou-se os modelos simétricos transformada via logaritmo da função (18) e se obteve estimativas clássicas para o vetor de parâmetros via método de máxima verossimilhança supondo diversas distribuições simétricas. Os graus de liberdade da distribuição t de Student foram estimados por verossimilhança perfilada.

A verossimilhança (18) e as densidades a priori (19), não apresentam forma analítica para a densidade a posteriori, dessa forma, utilizamos o algoritmo de Metropolis-Hastings para obter estimativas das densidades condicionais a posteriori de cada parâmetro e realizamos as inferências do modelo obtido. Metropolis-Hastings foi escolhido por ser de fácil implementação, enquanto o algoritmo amostrador de Gibbs requer distribuições condicionais a posteriori, que neste caso não são de fácil determinação. Os ajustes foram realizados com o software R 2.10.1 e para a realização dos diagnósticos Bayesianos foi utilizado o pacote coda versão 0.13-3. Esse pacote provê uma série de diagnósticos, dentre os quais os de Raftery-Lewis, Heidelberger-Welch e Geweke.

3. Resultados e Discussão

Após aplicação do algoritmo de máxima verossimilhança para os modelos não transformados e transformados com as distribuições da Tabela 1, variando os graus de liberdade ν , no caso da distribuição t de Student. Ainda em relação a distribuição t de Student, perfilando-se a verossimilhança se observou que a verossimilhança era máxima quando $\nu = 3$ ou 5 para o modelo não transformado e transformado, respectivamente. Os resultados das log-verossimilhanças, AIC, BIC, EPAM e EQM estão dispostos para todas as distribuições testadas na Tabela 2.

Tabela 2: Comparação das estatísticas de seleção das equações e erros dos ajustes.

Distribuição	Modelo	$\ell(\theta)$	AIC	BIC	EPAM	EQM
normal	CNT	-1696,1	3402,196	3426,107	14,337	10,313
	BNT	-1764,35	3536,708	3555,837	15,252	11,637
	CT	-1412,48	2834,968	2858,879	7,282	5,883
	BT	-1359,47	2728,948	2752,858	8,102	4,477
Cauchy	CNT	-2518,94	5047,871	5071,782	14,95	10,224
	BNT	-2738,64	5485,274	5504,403	15,535	11,007
	CT	-1834,26	3678,527	3702,438	6,792	6,546
	BT	-1374,78	2759,552	2783,463	7,007	3,496
Laplace	CNT	-1946,39	3902,779	3926,69	14,902	10,186
	BNT	-2029,16	4066,317	4085,445	15,269	10,502
	CT	-3079,05	6168,108	6192,019	105,876	114,7
	BT	-2650,44	5310,886	5334,797	113,941	141,257
logística tipo I	CNT	-1743,35	3494,708	3513,837	14,019	11,131
	BNT	-2096,15	4200,292	4219,421	15,44	13,601
	CT	-1431,94	2873,877	2897,788	15,42	5,856
	BT	-1532,3	3074,595	3098,506	9,381	6,448
logística tipo II	CNT	-1694,85	3397,691	3416,819	14,338	10,313
	BNT	-1715,87	3439,746	3458,875	14,565	10,693
	CT	-1860,54	3731,083	3754,994	13,784	6,04
	BT	-1502,85	3039,61	3015,699	8,322	4,82
t de Student	CNT ($\nu = 3$)	-447,252	902,505	921,634	14,481	9,837
	BNT ($\nu = 3$)	-1793,05	3596,108	3620,019	14,806	10,224
	CT ($\nu = 5$)	1479,66	-2949,32	-2946,93	7,143	13,957
	BT ($\nu = 5$)	1491,963	-2973,93	-2950,02	9,921	3,252

Fonte: Autores.

As siglas CNT, BNT, CT e BT significam, respectivamente, modelo clássico não transformado, Bayesiano não transformado, modelo clássico transformado e

Bayesiano transformado. Pode-se ver que as estimativas Bayesianas para a distribuição t de Student com $\nu = 5$ e uso de transformação produz valor máximo para a log-verossimilhança e os menores valores de AIC, BIC e EQM, ajustando-se melhor, portanto, por estes quatro critérios. Note que as três melhores equações, segundo os critérios de AIC e BIC, são aquelas que possuem distribuição t de Student BT, CT e CNT. Seguindo, exclusivamente, o EPAM, a melhor equação seria a que possui transformação usando inferência Bayesiana e distribuição de Cauchy para a variável dependente.

Observa-se na Tabela 3 o teste de Raftery e Lewis sugere valores para o burn-in e o total de iterações para o modelo Bayesiano transformado com distribuição t de Student com $\nu = 5$ graus de liberdade.

Tabela 3: Diagnósticos Bayesianos.

Parâmetro	Raftery-Lewis			Heidelberger-Welch
	Burn-in (B)	Total (N)	Fator (I)	p-valor
α	240	257040	6,86	0,122
κ	200	217780	5,81	0,117
θ	240	253040	6,75	0,747
λ	200	215000	5,74	0,293
τ	480	518020	13,8	0,835

Fonte: Autores.

Para τ , por exemplo, ele sugere um total de 518020 iterações com burn-in de 480 iterações. Na prática, os valores utilizados foram, até, quatrocentas vezes maiores na intenção de diminuir autocorrelações e aplicar os outros testes que necessitam de um número maior de iterações, já que, na aplicação dos mesmos, particiona-se os valores simulados. O fator de dependência de Raftery e Lewis aponta que as escolhas podem não ter sido as melhores para as estimativas iniciais. O teste de Heidelberger e Welch obteve p-valores superiores a 10%, indicando a aceitação da hipótese de estacionaridade para este teste. Apesar do diagnóstico de Raftery e Lewis apontar para uma possível má escolha nas estimativas iniciais, os outros diagnósticos foram atendidos para o modelo selecionado. Portanto, a Tabela 4 apresenta as estimativas para a média, desvio padrão, assimetria, curtose e intervalo de credibilidade de 95%, seguindo um modelo simétrico com transformação de Box-Cox, preditor não linear de Chapman-Richards e inferência Bayesiana nos parâmetros, utilizadas para representar o comportamento das alturas dos híbridos *Eucalyptus urophylla*, com cruzamento natural, ao longo dos 72 meses.

Tabela 4: Estimativas Bayesianas dos parâmetros do modelo de crescimento em altura Chapman-Richards com distribuição t de Student com uso de transformação.

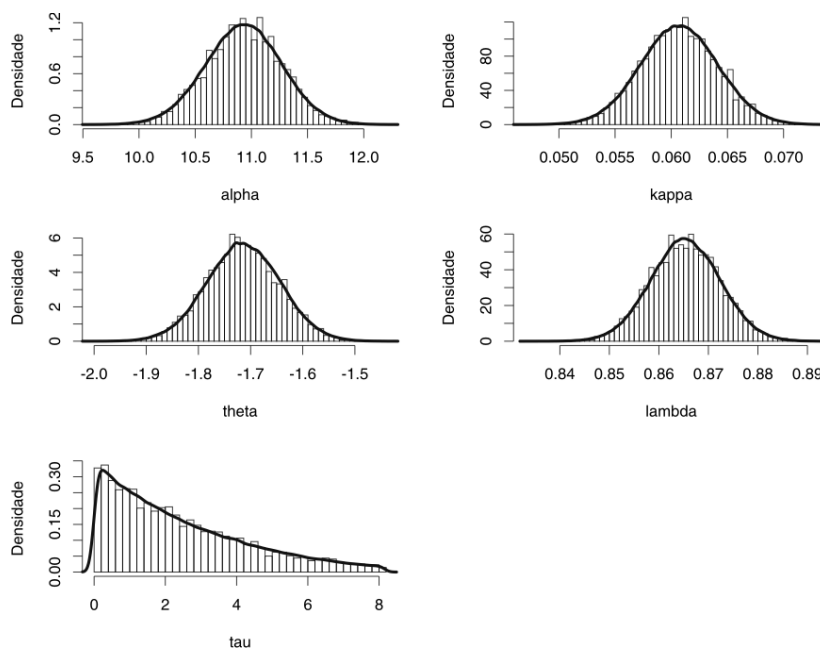
Parâmetro	Média	D. Padrão	Assimetria	Curtose	I. Cred.
α	10,9384	0,3659	-0,2039	0,0893	(10,1708, 11,6569)
κ	0,061	0,0035	-0,0734	-0,5105	(0,0540, 0,0673)
θ	-1,7106	0,0684	-0,0695	0,1954	(-1,8511, -1,5742)
λ	0,8658	0,0072	-0,1726	0,5813	(0,8507, 0,8809)
τ	2,1733	1,7841	0,9099	0,1525	(0,0684, 6,2864)

Fonte: Autores.

A estimativa de λ sob $H_0: \lambda = 1$ contra $H_1: \lambda \neq 1$ são significativas ao nível de 95% de credibilidade.

Na Figura 1 são mostradas as densidades condicionais a posteriori do vetor de parâmetros do modelo escolhido.

Figura 1: Histogramas das densidades a posteriori do modelo BT com distribuição t de Student.



Fonte: Autores.

Que as densidades a posteriori continuam normalmente distribuídas entre os valores estimados, exceto para τ cuja distribuição é assimétrica positiva, diferindo da distribuição a priori, que foi ajustada por uma distribuição uniforme.

Na Figura 2 (esquerda) se encontram os ajustes das estimativas Bayesianas aos dados transformados pelo modelo de Chapman-Richards e distribuição t de Student cujo preditor não linear é

$$\eta_i = 10,938[1 - \exp(-0,061x_i)]^{1,711}. \quad (17)$$

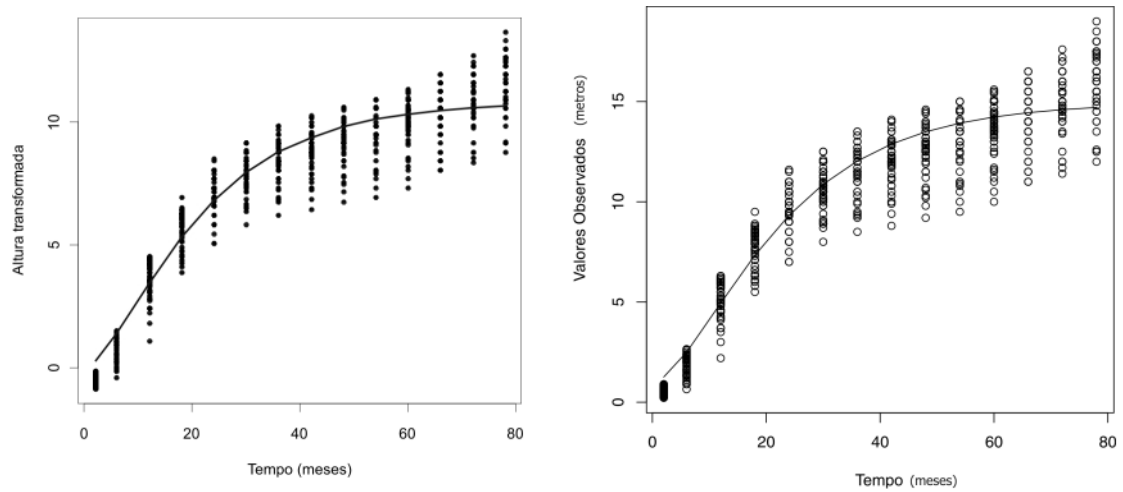
Nota-se um bom ajuste, exceto nos meses iniciais. Note que ao se executar a transformação de Box-Cox, as medidas para o segundo mês se tornaram negativas, pois, antes da transformação, eram quantidades inferiores a 1 metro. O mesmo vale para a estimação clássica. Temos ainda, que a estimativa do parâmetro assintótico é 10,94, fazendo com que o modelo perca a interpretação física. Isso acontece devido à variação da escala que a transformação proporciona (Box & Cox, 1964).

Para efeito de avaliação na escala original, realiza-se a combinação das equações (6) e (17), fazendo $y_i^{(\lambda)} = \eta_i$, tal que:

$$\hat{y}_i^* = \{9,471[1 - \exp(-0,061x_i)]^{1,711} + 1\}^{1,155}, \quad (18)$$

o que simplifica a interpretação do modelo. A equação (18) é conhecida como retransformação (Cordeiro, 2009). Na Figura 2 (direita) se encontra o ajuste aos dados originais pela equação (18).

Figura 2: Ajuste das estimativas Bayesianas aos dados transformados (esquerda) e das estimativas retransformadas aos dados originais (direita).



Fonte: Autores.

Note que essa equação, embora não pareça, a estimativa do parâmetro assintótico condiz com sua interpretação física. Para ver isso, subtraia a unidade do termo entre chaves, obtendo uma aproximação grosseira $\hat{y}_i^\#$ de \hat{y}_i^* :

$$\hat{y}_i^\# = 13,722[1 - \exp(-0,061x_i)]^{1,976}, \quad (18)$$

sendo a nova estimativa do parâmetro assintótico igual 13,722.

4. Considerações Finais

De um modo geral, observou-se que os modelos simétricos transformados com distribuição t de Student com 5 graus de liberdade, função de ligação não linear seguindo o modelo de Chapman-Richards e inferência Bayesiana via algoritmo de Metropolis-Hastings se ajustam bem aos dados de crescimento em altura de Eucalyptus no Polo Gesseiro do Araripe.

Referências

- Ahelegbey, D. F., Billio, M., & Casarin, R. (2016). Sparse graphical vector autoregression: a Bayesian approach. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, (123/124), 333-361.
- Araujo, S. M. S. D. (2004). O Pólo Gesseiro do Araripe: unidades geo-ambientais e impactos da mineração. TESE (Doutorado em Geociências) - Universidade Estadual de Campinas, Campinas, 2004, 305.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-43.
- Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.
- Carvalho Araujo, S. K., Oliveira, D. R., da Silva, T. J., Barbosa, J. C., & Gava, M. (2016). Desempenho acústico de painéis de gesso incorporados com fibras de celulose. *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, 20(1), 456-463.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R*. CRC press.
- Cordeiro, G. M., & Andrade, M. G. (2011). *Transformed symmetric models*. *Statistical Modelling*, 11(4), 371-388.
- Cordeiro, G. M., & de Andrade, M. G. (2009). Transformed generalized linear models. *Journal of Statistical Planning and Inference*, 139(9), 2970-2987.
- Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 883-904.
- Cysneiros, F. J. A., Paula, G. A., & Galea, M. (2005). Modelos simétricos aplicados. *Escola de Modelos de Regressao*, 9.

Dey, D. K., Ghosh, S. K., & Mallick, B. K. (Eds.). (2000). *Generalized linear models: A Bayesian perspective*. CRC Press.

Ferreira, F. C. (2017). *Estudo de caracterização do gesso para revestimento produzido no Polo Gesseiro do Araripe* (Master's thesis, Universidade Federal de Pernambuco).

Gadella, F. H. D. L., Silva, J. A. A. D., Ferreira, R. L. C., Melo, I. V. D., Jorge, D. L., Tavares, J. A., & Silva, S. P. R. D. (2012). Rendimento volumétrico e energético de clones de híbridos de *Eucalyptus* sp. no polo gesseiro do Araripe, PE. *Ciência Florestal*, 22(2), 331-341.

Gatto, D. A., Santini, E. J., Haselein, C. R., & Durlo, M. A. (2003). Características da lenha produzida na região da quarta colônia de imigração italiana do Rio Grande Do Sul. *Ciência Florestal*, 13(2), 7-16.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410), 398-409.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4), 457-472.

Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical science*, 473-483.

Gilks, R., & Richardson, S. (1996). Spiegelhalter. *Markov Chain Monte Carlo in practice*.

Hastings, W. K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Lima, C. G. D. R., Carvalho, M. D. P., Narimatsu, K. C. P., Silva, M. G. D., & Queiroz, H. A. D. (2010). Atributos físico-químicos de um Latossolo do cerrado brasileiro e sua relação com características dendrométricas do eucalipto. *Revista Brasileira de Ciência do Solo*, 34(1), 163-173.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* (Wiley: New York.).

Paulino, C. D., Turkman, M. A. A., & Murteira, B. (2003). *Estatística bayesiana* Fundação Clouste Gulbenkian Lisboa.

Pericchi, L. R. A. (1981). Bayesian Approach to Transformations to Normality. *Biometrika Series A*, London, 68, 1, 35-43.

Schneider, P. R., Schneider, P. S. P., & Souza, C. A. M. (2009). *Análise de regressão aplicada à engenharia florestal*. Santa Maria, Universidade Federal de Santa Maria, 294.

Schwarz, G. (1978). Estimating the dimension of the model. *Annals of Statistics*, Hayward, 6 461 – 464.

Porcentagem de contribuição de cada autor no manuscrito

Kleber Napoleão Nunes de Oliveira Barros– 50%

Mácio Augusto de Albuquerque – 30%

José Antônio Aleixo da Silva – 20%