



ADRIELE APARECIDA PEREIRA

**ESTIMAÇÃO DE PARÂMETROS DE MODELOS
NÃO LINEARES COM RESÍDUOS
AUTOCORRELACIONADOS**

LAVRAS - MG

2016

ADRIELE APARECIDA PEREIRA

**ESTIMAÇÃO DE PARÂMETROS DE MODELOS NÃO LINEARES COM RESÍDUOS
AUTOCORRELACIONADOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr. Augusto Ramalho de Moraes
Orientador
Profa. Dra. Thelma Sáfydi
Coorientadora

**LAVRAS - MG
2016**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Pereira, Adriele Aparecida.

Estimação de parâmetros de modelos não lineares com resíduos autocorrelacionados / Adriele Aparecida Pereira. - Lavras : UFLA, 2016.

66 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2016.

Orientador: Augusto Ramalho de Moraes.

Bibliografia.

1. Coffea Arábica L.. 2. Modelo Brody. 3. Modelo Logístico. 4. Reamostragem ponderada. 5. Regressão não linear. I. Universidade Federal de Lavras. II. Título.

ADRIELE APARECIDA PEREIRA

**ESTIMAÇÃO DE PARÂMETROS DE MODELOS NÃO LINEARES COM RESÍDUOS
AUTOCORRELACIONADOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 14 de dezembro de 2016.

Dra. Carla Regina Guimarães Brighenti UFSJ

Dra. Izabela Regina Cardoso de Oliveira UFLA

Dr. Márcio Balestre UFLA

Dra. Myriane Stella Scalco UFLA

Prof. Dr. Augusto Ramalho de Moraes
Orientador
Profa. Dra. Thelma Sáfyadi
Coorientadora

**LAVRAS - MG
2016**

*Ao meu esposo, Tales,
aos meus pais, José Antônio e Bernadete,
aos meus irmãos, Ramon e Cleófas e
ao meu sobrinho, Kauan Gabriel.*

DEDICO.

AGRADECIMENTOS

Primeiramente, agradeço à Santíssima Trindade e à Nossa Senhora, por iluminarem minha vida pessoal e profissional, me abençoando a cada dia com saúde e perseverança. Tudo é possível para quem Neles creem!

Aos professores Augusto Ramalho de Moraes e Thelma Sáfadi, por me orientarem ao longo do doutorado.

Aos professores membros da banca, de qualificação e defesa, por contribuírem para o enriquecimento desse trabalho.

A todos os professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, por contribuírem para a minha formação profissional.

À Universidade Federal de Lavras, por possibilitar a realização do doutorado, assim como o mestrado e a graduação.

Ao CNPq, pela concessão de bolsa de estudos.

Ao meu esposo Tales, pelo amor, companheirismo, palavras de otimismo e incentivo, e por estar ao meu lado em todos os momentos. Você faz parte dessa realização profissional!

Aos meus pais José Antônio e Bernadete, aos meus irmãos Ramon e Cleófas e ao meu sobrinho Kauan Gabriel, que me deram todo o apoio quando fui morar em outra cidade para estudar. Essa conquista também é de vocês!

*Buscai primeiro o reino de Deus e a sua justiça,
e tudo mais vos será acrescentado.*

Mateus 6:33

RESUMO GERAL

Devido à importância econômica do café para o Brasil, um dos principais focos de pesquisa do setor agrícola refere-se ao desenvolvimento do cafeeiro. Uma variável relevante a ser considerada em tais pesquisas é a altura de plantas, pois além de ser representativa do desenvolvimento vegetativo está correlacionada com a produtividade. Para o cafeicultor é essencial caracterizar o padrão de crescimento da cultura, uma vez que o manejo e a produção estão relacionados com o seu desenvolvimento ao longo do tempo. Dentre os modelos que podem ser utilizados para modelar o crescimento destacam-se os modelos de regressão não lineares. No entanto, para obter estimativas coerentes e modelar corretamente o padrão de crescimento do cafeeiro, deve-se considerar que dados longitudinais podem apresentar autocorrelação residual, sendo que, se tal característica não é considerada, os resultados e inferências podem ser comprometidos. Assim, os objetivos deste trabalho foram: No primeiro capítulo, apresentar um resumo sobre modelos não lineares e suas principais características. No segundo capítulo, identificar, dentre os modelos de regressão não lineares Logístico, Brody, von Bertalanffy e Richards, o que melhor caracteriza o padrão de crescimento em altura de plantas do cafeeiro irrigadas e não irrigadas ao longo do tempo, considerando a autocorrelação dos resíduos, sendo a estimação dos parâmetros realizada pelo método de mínimos quadrados (abordagem clássica). E, no terceiro capítulo, objetivou-se apresentar a modelagem não linear bayesiana do crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, também considerando a autocorrelação residual e os modelos não lineares citados anteriormente; e apresentar um estudo de simulação com o intuito de validar o método de reamostragem ponderada utilizado na estimação de parâmetros, considerando como distribuição candidata uma aproximação da verossimilhança. Nos resultados do capítulo 2, verificou-se, com base nos critérios adotados, desvio padrão residual e critério de informação de Akaike e nos modelos avaliados, que o modelo Brody é o que melhor representa o crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, ao longo do tempo. Dentre os resultados do capítulo 3, verificou-se que, no estudo de simulação, os intervalos de máxima densidade a posteriori obtidos para todos os parâmetros dos modelos avaliados, inclusive os parâmetros que modelam a autocorrelação residual, contiveram os valores paramétricos pré-estabelecidos, validando a eficiência do método da reamostragem ponderada, considerando como distribuição candidata uma aproximação da verossimilhança. Para as plantas não irrigadas, o critério de informação bayesiano e o critério da densidade preditiva ordenada, indicaram que, dentre os modelos avaliados, o modelo Logístico é o que melhor descreve o crescimento em altura do cafeeiro ao longo do tempo. E, para as plantas irrigadas, esses mesmos critérios indicaram o modelo Brody.

Palavras-chave: Coffea Arábica L.. Modelo Brody. Modelo Logístico. Reamostragem ponderada. Regressão não linear.

GENERAL ABSTRACT

Due to the economic importance of coffee for Brazil, one of the main focuses of research in the agricultural sector refers to the coffee development. In such research, the height of plants is a relevant variable to be considered, as well as being representative of the vegetative growth it is correlated with productivity. For the coffee grower, it is essential to characterize the growth pattern of the crop, since the management and production are related to its development over time. The non-linear regression models stands out among the models that can be used to model the growth. However, in order to obtain coherent estimates and thus to correctly model the growth pattern of the coffee tree, it should be considered that longitudinal data may present residual autocorrelation, and if this characteristic is not considered, the results and inferences can be compromised. The objectives of this study were: In the first chapter, present a summary on non-linear models and their main characteristics. In the second chapter to identify among Logistic, Brody, von Bertalanffy and Richards non-linear models that best describes the height growth pattern of irrigated and non-irrigated coffee plants over time, considering the residual autocorrelation, performing the estimation of the parameters by the method of least squares (classical approach). In the third chapter, we present the Bayesian nonlinear modeling of growth in height of irrigated and non-irrigated coffee plants, also considering the residual autocorrelation and Logistic, Brody, von Bertalanffy and Richards non-linear models. We also present a simulation study with the purpose of validating the weighted resampling method in the parameters estimation, considering a likelihood approximation as candidate distribution. In the results of chapter 2, we verified that among the studied models, the Brody model is the one that best represents the growth pattern of irrigated and non-irrigated coffee plants over time based on the criteria of residual standard deviation and Akaike's information criterion. Among the results of chapter 3, we found that the high posterior density intervals obtained for all the parameters of the evaluated models, including the parameters of residual autocorrelation, contained the pre-established parametric values, validating the efficiency of the weighted resampling method for parameter estimation, considering as candidate distribution an approximation of likelihood. For non-irrigated system, the Bayesian information criterion and the criterion of ordered predictive density indicated, among the studied models, the Logistic model as the one that best describes the growth of the coffee plant height over time. In addition, for the irrigated system, the same criteria indicated the Brody model.

Keywords: Coffea Arabica L.. Brody model. Logistic model. Weighted resampling. Nonlinear regression.

SUMÁRIO

	CAPÍTULO 1: INTRODUÇÃO GERAL	10
	REFERÊNCIAS	18
	CAPÍTULO 2: MODELAGEM CLÁSSICA DA CURVA DE CRESCIMENTO DO CAFEIEIRO	19
1	INTRODUÇÃO	21
2	MATERIAL E MÉTODOS	22
3	RESULTADOS E DISCUSSÃO	25
4	CONCLUSÕES	31
	REFERÊNCIAS	32
	CAPÍTULO 3: MODELAGEM BAYESIANA DA CURVA DE CRESCIMENTO DO CAFEIEIRO	34
1	INTRODUÇÃO	36
2	MATERIAL E MÉTODOS	38
3	RESULTADOS E DISCUSSÃO	46
3.1	Análise dos dados simulados	46
3.2	Ajuste bayesiano de modelos não lineares à altura de plantas do cafeeiro .	49
4	CONCLUSÕES	58
	REFERÊNCIAS	59
	ANEXOS	62

CAPÍTULO 1: INTRODUÇÃO GERAL

Em diversas situações, deseja-se estudar a relação entre duas ou mais variáveis quantitativas. O método estatístico utilizado para estudar tal relação é denominado análise de regressão. Assim, os modelos de regressão descrevem o comportamento de uma variável, denominada variável resposta (ou dependente) em função de uma ou mais variáveis, denominadas variáveis independentes.

Segundo Draper e Smith (1998) os modelos de regressão, em relação aos parâmetros, são classificados em:

- i) Linear: um modelo é considerado linear, se todas as suas derivadas parciais de primeira ordem com relação aos parâmetros não dependem de nenhum parâmetro;
- ii) Não linear: se pelo menos uma das derivadas parciais de primeira ordem do modelo com relação aos parâmetros, depende de algum parâmetro, então o modelo é considerado não linear;
- iii) Linearizável: é um modelo que a princípio é não linear, mas que através de alguma transformação se torna linear.

Uma das aplicações de modelos de regressão é no estudo de curvas de crescimento, a qual é definida por uma sequência de medidas de determinada espécie ou indivíduo, geralmente peso ou altura, tomadas ao longo do tempo. Dentre esses modelos, os modelos não lineares têm sido preferidos, pois descrevem de forma satisfatória diferentes curvas de crescimento e apresentam parâmetros com interpretação biológica (FERNANDES et al., 2014).

No entanto, para modelar corretamente uma curva de crescimento, deve-se considerar que dados longitudinais podem apresentar autocorrelação residual, ocasionada pela dependência entre as observações tomadas ao longo do tempo, sendo que, se tal característica não é considerada, os resultados e inferências podem ser comprometidos (MUNIZ; NASCIMENTO; FERNANDES, 2017).

Modelos de regressão não lineares: algumas características e propriedades

Na Tabela 1 são apresentados alguns modelos não lineares mais utilizados no estudo de curvas de crescimento.

Tabela 1 Modelos de regressão não lineares utilizados na modelagem de curvas de crescimento.

Modelo	Parametrização
Brody	$y_i = \alpha[1 - \beta e^{-k \cdot t_i}] + e_i$
Logístico	$y_i = \frac{\alpha}{1 + e^{-k \cdot (t_i - \beta)}} + e_i$
von Bertalanffy	$y_i = \alpha[1 - \beta e^{-k \cdot t_i}]^3 + e_i$
Richards	$y_i = \alpha[1 - (1 - \delta)e^{-k \cdot (t_i - \beta)}]^{1-\delta} + e_i$

Nas parametrizações da Tabela 1, tem-se que: y_i corresponde aos valores observados da variável resposta ao longo do tempo t_i ; α representa a assíntota superior do modelo; nos modelos Logístico e Richards β indica a abscissa do ponto de inflexão, nos demais modelos é um parâmetro de escala sem interpretação biológica direta; δ é o parâmetro que define o formato da curva e determina em que proporção de α ocorre o ponto de inflexão; k representa o índice de crescimento, de forma que, quanto maior for este valor menos tempo será necessário para a variável resposta atingir a assíntota α ; e, e_i é o erro aleatório associado à observação y_i , o qual pressupõe-se que seja independente e identicamente distribuído segundo uma distribuição Normal com média zero e variância homogênea, ou seja, $e_i \sim N(0, \sigma_e^2)$ (MISCHAN; PINHO, 2014).

A interpretabilidade do parâmetro α é comprovada com o cálculo do limite com a variável tempo tendendo ao infinito ($t \rightarrow \infty$). Para o modelo Logístico, por exemplo, tem-se que:

$$\lim_{t \rightarrow \infty} \frac{\alpha}{[1 + e^{-k \cdot (t - \beta)}]} = \lim_{t \rightarrow \infty} \frac{\alpha}{[1 + \frac{1}{e^{k \cdot (t - \beta)}}]} = \frac{\alpha}{[1 + 0]} = \alpha$$

O modelo Logístico é um dos modelos não lineares mais conhecidos e utilizados (SEBER; WILD, 1989). Esse modelo apresenta formato sigmoideal e é simétrico em relação ao ponto de inflexão; ou seja, quando t assume o valor β (abscissa do ponto de inflexão), a variável resposta y atinge metade do seu valor máximo esperado, ou seja, $\frac{\alpha}{2}$.

Outro modelo que também apresenta formato sigmóide é o de von Bertalanffy (BERTALANFFY, 1957). Entretanto, esse modelo não apresenta simetria como o Logístico, pois o ponto de inflexão ocorre em $t = \frac{\ln(3\beta)}{k}$, momento em que a variável resposta atinge o valor $\frac{8}{27}\alpha$, aproximadamente 30% da assíntota.

Já o modelo Brody é um exemplo de modelo não linear que não apresenta comporta-

mento sigmoial (BRODY, 1945). Ele não possui ponto de inflexão e apresenta concavidade voltada para baixo, sendo caracterizado como um modelo de crescimento restrito. Calculando as derivadas de 1ª e 2ª ordem desse modelo com relação a t , tem-se que:

$$\begin{aligned}\frac{dy}{dt} &= \alpha \cdot (-\beta) \cdot e^{-k \cdot t} \cdot (-k) = \alpha \cdot \beta \cdot k \cdot e^{-k \cdot t} \\ \frac{d^2y}{dt^2} &= -\alpha \cdot \beta \cdot k^2 \cdot e^{-k \cdot t}\end{aligned}$$

Assim, como $\alpha > 0$, $\beta > 0$, $k > 0$ e a função exponencial não se anula, certifica-se que o modelo Brody não possui ponto de inflexão.

Apresentando quatro parâmetros e formato variável, o modelo Richards representa uma generalização dos principais modelos não lineares (KOYA; GOSHU, 2013; RICHARDS, 1959). Assim, quando δ assume os valores 0 , $\frac{2}{3}$ e 2 ele se decompõe, respectivamente, nos modelos Brody, von Bertalanffy e Logístico.

Substituindo os valores comentados acima na expressão do modelo Richards, tem-se que:

i) $\delta = 0$

$$\begin{aligned}y_i &= \alpha [1 - (1 - 0)e^{-k \cdot (t - \beta)}]^{1-0} + e_i \\ y_i &= \alpha [1 - e^{-k \cdot (t - \beta)}] + e_i\end{aligned}\tag{1}$$

que é outra parametrização do modelo Brody.

ii) $\delta = \frac{2}{3}$

$$\begin{aligned}y_i &= \alpha [1 - (1 - \frac{2}{3})e^{-k \cdot (t - \beta)}]^{1-\frac{2}{3}} + e_i \\ y_i &= \alpha [1 - \frac{1}{3}e^{-k \cdot (t - \beta)}]^3 + e_i\end{aligned}$$

que é uma reparametrização do modelo von Bertalanffy, em que a constante $\frac{e^{k \cdot \beta}}{3}$ é representada pelo parâmetro β na expressão apresentada na Tabela 1.

iii) $\delta = 2$

$$y_i = \alpha[1 - (1 - 2)e^{-k \cdot (t - \beta)}]^{1/2} + e_i$$

$$y_i = \alpha[1 + e^{-k \cdot (t - \beta)}]^{-1} + e_i$$

que é exatamente a parametrização do modelo Logístico apresentada na Tabela 1.

Calculando a 2ª derivada do modelo Richards com relação a t e igualando-a a zero obtém-se o seu ponto de inflexão. Desse modo,

$$\begin{aligned} \frac{dy}{dt} &= \alpha \cdot \left(\frac{1}{1 - \delta}\right) \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta - 1} \cdot - (1 - \delta) \cdot e^{-k \cdot (t - \beta)} \cdot (-k) \\ &= \alpha \cdot k \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta - 1} \cdot e^{-k \cdot (t - \beta)} \end{aligned}$$

A 2ª derivada é dada por:

$$\begin{aligned} \frac{d^2y}{dt^2} &= \alpha \cdot k \cdot \left\{ \left(\frac{1}{1 - \delta} - 1\right) \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta - 2} \cdot - (1 - \delta) \cdot e^{-k \cdot (t - \beta)} \cdot (-k) \cdot e^{-k \cdot (t - \beta)} + \right. \\ &\quad \left. (-k) \cdot e^{-k \cdot (t - \beta)} \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta - 1} \right\} \end{aligned}$$

Igualando a zero tem-se que:

$$\alpha \cdot k \cdot \left\{ \left(\frac{\delta}{1 - \delta}\right) \cdot (1 - \delta) \cdot k \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{2\delta - 1/1 - \delta} \cdot e^{-2k \cdot (t - \beta)} - \right.$$

$$\left. k \cdot e^{-k \cdot (t - \beta)} \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta} \right\} = 0$$

$$\delta \cdot k \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{2\delta - 1/1 - \delta} \cdot e^{-2k \cdot (t - \beta)} = k \cdot e^{-k \cdot (t - \beta)} \cdot [1 - (1 - \delta)e^{-k \cdot (t - \beta)}]^{1/1 - \delta}$$

$$\delta \cdot e^{-k \cdot (t - \beta)} = 1 - (1 - \delta)e^{-k \cdot (t - \beta)}$$

$$\delta \cdot e^{-k \cdot (t - \beta)} + (1 - \delta)e^{-k \cdot (t - \beta)} = 1$$

$$e^{-k \cdot (t - \beta)} = 1$$

$$\ln[e^{-k \cdot (t - \beta)}] = \ln(1)$$

$$-k \cdot (t - \beta) = 0$$

$$t = \beta$$

Portanto, verifica-se que $t = \beta$ é a abcissa do ponto de inflexão do modelo Richards. Substituindo na expressão do modelo, tem-se que, nesse momento, a variável resposta atinge o valor:

$$y(\beta) = \alpha \cdot \delta^{\frac{1}{1-\delta}}$$

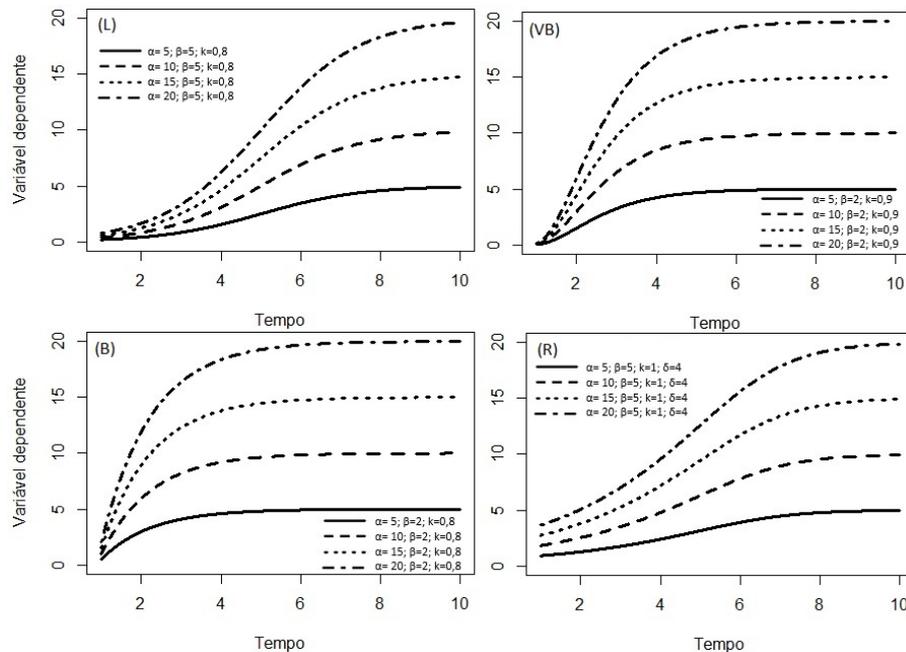
Desse modo, o parâmetro δ , além de definir o formato da curva, determina em que proporção de α ocorre o ponto de inflexão.

Estudo sobre variação de parâmetros em modelos de regressão não lineares

A seguir, é apresentado um estudo sobre a variação dos parâmetros α , β e k nos modelos não lineares Logístico, Brody, von Bertalanffy e Richards, possibilitando melhor compreensão sobre o comportamento desses modelos.

De acordo com a Figura 1, para os quatro modelos considerados, verifica-se que, a variação do parâmetro α não influencia no formato dos modelos, mas sim no valor máximo a ser atingido pela variável dependente ao longo do tempo.

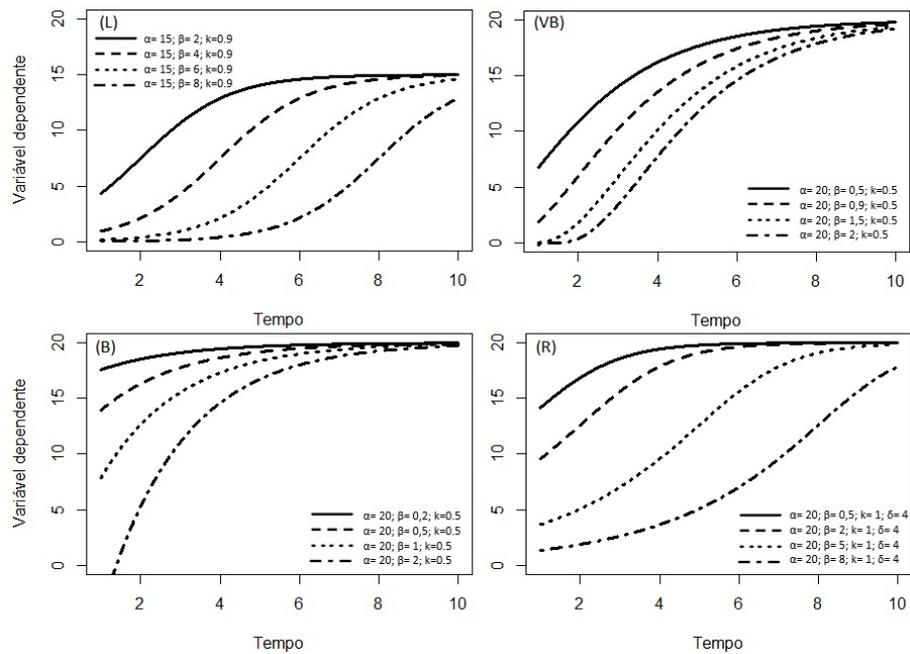
Figura 1 Variação do parâmetro α nos modelos Logístico (L), von Bertalanffy (VB), Brody (B) e Richards (R).



Já a variação do parâmetro β , apresentada na Figura 2, ocasiona um deslocamento na curva estimada (à medida que se aumentou os valores de β , os gráficos se deslocaram

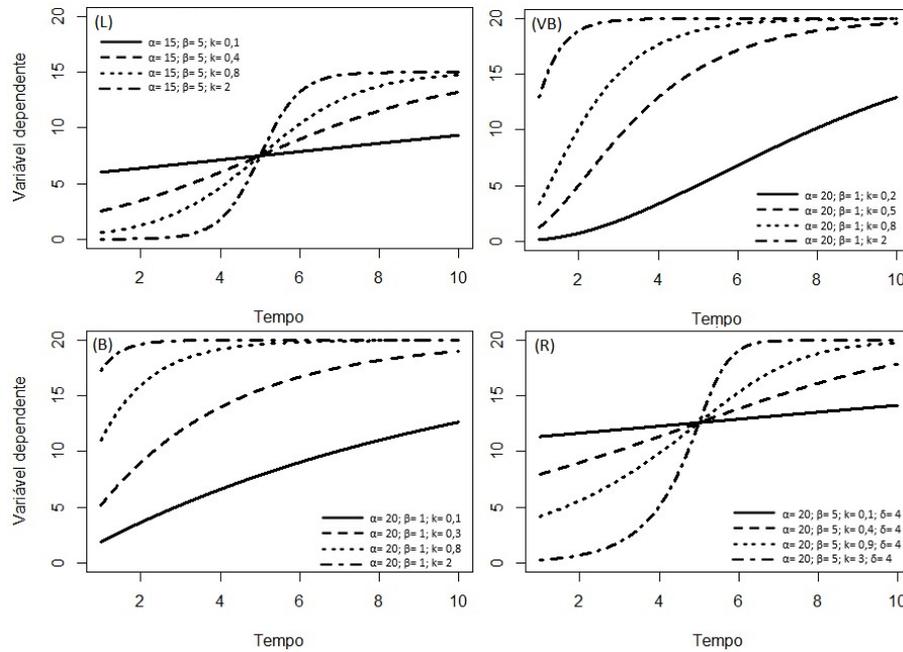
para a direita), podendo influenciar no comportamento dos modelos. No modelo Logístico, esse parâmetro indica exatamente a abscissa do ponto de inflexão; assim, conforme β varia, o momento em que ocorre a mudança de concavidade também varia.

Figura 2 Variação do parâmetro β nos modelos Logístico (L), von Bertalanffy (VB), Brody (B) e Richards (R).



Com base na Figura 3, observa-se que a variação do parâmetro k é determinante no comportamento dos modelos, podendo afetar os formatos.

Figura 3 Variação do parâmetro k nos modelos Logístico (L), von Bertalanffy (VB), Brody (B) e Richards (R).



As curvas estimadas com valores de k próximos a zero (linhas contínuas) tendem a apresentar um comportamento linear simples. À medida que k aumenta, é possível observar maior crescimento no período estudado. Desta forma, pode-se concluir que, quanto menor o valor estimado para k menos intenso é o crescimento ao longo do tempo e, de forma equivalente, quanto maior o valor de k mais intenso (acentuado) é o crescimento.

Outra maneira de compreender como o valor estimado para k influencia no comportamento do modelo é calculando a diferença da variável resposta em um intervalo de tempo. Por exemplo, ao considerar o modelo Richards com $k = 0,1$ e $k = 3$, tem-se que o crescimento ocorrido entre os tempos 2 e 6, é de, aproximadamente, 1 unidade para o modelo Richards considerando $k = 0,1$, e 18 unidades para o modelo considerando $k = 3$. Desta forma, fica evidente que quanto maior o valor de k maior é o crescimento ao longo do tempo.

Estimação de parâmetros de modelos de regressão não lineares

A estimação dos parâmetros de um modelo de regressão pode ser feita por diferentes métodos, sendo que, a abordagem clássica é a mais comum. Tal abordagem considera que o parâmetro a ser estimado é fixo, ou seja, é igual a um valor particular, e toda a inferência se

resume nas informações contidas nos dados amostrais.

Outra abordagem que pode ser utilizada na estimação é a bayesiana, que considera tanto as observações quanto os parâmetros do modelo como variáveis aleatórias. Desse modo, toda inferência é feita com base na distribuição a posteriori dos parâmetros, que é obtida combinando a distribuição a priori dos parâmetros com a distribuição conjunta dos dados amostrais (PAULINO; TURKMAN; MURTEIRA, 2003).

Inserido nesse contexto, esse trabalho foi estruturado em três capítulos, sendo que, neste primeiro, objetivou-se apresentar um resumo sobre modelos não lineares e suas principais características. No segundo e terceiro capítulos, buscou-se apresentar a modelagem não linear da curva de crescimento do cafeeiro considerando resíduos autocorrelacionados, sendo a estimação de parâmetros feita por meio das abordagens clássica e bayesiana, respectivamente.

REFERÊNCIAS

BERTALANFFY, L. Quantitative laws in metabolism and growth. **Quantitative Review of Biology**, Chicago, v. 32, p. 230-270, Sept. 1957.

BRODY, S. **Bioenergetics and growth**. New York: Reinhold, 1945. 1023 p.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: J. Wiley, 1998. 706 p.

FERNANDES, T. J. et al. Seleção de modelos não lineares para descrição das curvas de crescimento do fruto do cafeeiro. **Coffee Science**, Lavras, v. 9, n. 2, p. 207-215, abr./jun. 2014.

KOYA, P. R.; GOSHU, A. T. Generalized mathematical model for biological growths. **Open Journal of Modelling and Simulation**, Dhahran, v. 1, n. 4, p. 42-53, 2013.

MISCHAN, M. M.; PINHO, S. Z. de. **Modelos não lineares: funções assintóticas de crescimento**. São Paulo: Cultura Acadêmica, 2014, 184 p.

MUNIZ, J. A.; NASCIMENTO, M. da S.; FERNANDES, T. J. Nonlinear models for description of cacao fruit growth with assumption violations. **Caatinga**, Mossoró, v. 30, n. 1, p. 250-257, jan./mar. 2017.

PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

RICHARDS, F. J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Elmsford, v. 10, p. 290-300, 1959.

SEBER, G. A. F.; WILD, C. J. **Non linear regression**. New Jersey: J. Wiley, 1989. 752 p.

CAPÍTULO 2: MODELAGEM CLÁSSICA DA CURVA DE CRESCIMENTO DO CA- FEEIRO

RESUMO

A cafeicultura figura como uma das mais relevantes atividades agrícolas do Brasil, garantindo ao país a liderança no ranking internacional de produção e exportação de café. Para o cafeicultor é essencial caracterizar o padrão de crescimento da cultura, uma vez que o manejo e a produção estão relacionados com o seu desenvolvimento ao longo do tempo. Dentre os modelos que podem ser utilizados para modelar o crescimento, os modelos de regressão não lineares se destacam por se adequarem a variados padrões de crescimento e por resumirem as informações presentes nos dados em alguns parâmetros interpretáveis biologicamente. Um dos métodos que podem ser utilizados na estimação dos parâmetros é o de mínimos quadrados, que consiste em estimá-los de forma que a soma de quadrados dos resíduos seja mínima. No entanto, para obter estimativas coerentes e modelar corretamente o padrão de crescimento do cafeeiro, é essencial considerar na modelagem a natureza dos dados e suas características. Uma dessas características, que em grande parte das pesquisas é desprezada, é a autocorrelação residual, ocasionada pela dependência entre as observações tomadas ao longo do tempo. Desse modo, o objetivo deste capítulo foi identificar, dentre os modelos de regressão não lineares Logístico, Brody, von Bertalanffy e Richards, o que melhor caracteriza o padrão de crescimento em altura de plantas do cafeeiro irrigadas e não irrigadas ao longo do tempo, considerando a autocorrelação dos resíduos. Com base nos critérios adotados, desvio padrão residual e critério de informação de Akaike e nos modelos avaliados, o modelo Brody é o que melhor representa o crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, ao longo do tempo.

Palavras-chave: Autocorrelação residual. Crescimento vegetativo. Modelos não lineares.

ABSTRACT

Coffee production is one of the most important agricultural activities in Brazil, ensuring the country's leadership in the international coffee production and export ranking. For the coffee grower it is essential to characterize the growth pattern of the crop, since management and production are related to its development over time. Among the models that can be used to model growth, the nonlinear regression models stand out for adapting to various growth patterns and for summarizing the information present in the data in some biologically interpretable parameters. The least squares is one of the methods that can be used to estimate the parameters. The method consists of estimating the parameters that make the sum of squares of the residuals minimal. However, in order to obtain coherent estimates and thus to model the growth pattern of the coffee tree correctly, it is essential to consider in the modeling the nature of the data and its characteristics. The residual autocorrelation caused by the dependence of the observations taken over time is one of these characteristics. However, many researchers despise it. In this context, the objective of this chapter were identify among Logistic, Brody, von Bertalanffy and Richards nonlinear models, that the best characterizes the growth pattern in height of coffee plants irrigated and non-irrigated over time, considering the residual autocorrelation. Among the studied models, the Brody model is the one that best represents the growth pattern of irrigated and non-irrigated coffee plants over time based on the criteria of residual standard deviation and Akaike's information criterion.

Keywords: Residual autocorrelation. Vegetative growth. Nonlinear models.

1 INTRODUÇÃO

A cafeicultura figura como uma das mais relevantes atividades agrícolas do Brasil (RODRIGUES et al., 2016), garantindo ao país a liderança no ranking internacional de produção e exportação de café. Para 2016, a safra que está estimada entre 49,13 e 51,94 milhões de sacas beneficiadas, pode ser a segunda maior da história (CONAB, 2016).

Diferente de outras culturas, o cafeeiro leva dois anos para completar seu ciclo fenológico (CAMARGO; CAMARGO, 2001). Esse ciclo é caracterizado pela alternância anual entre altas e baixas produtividades (PEREIRA et al., 2011), que afeta diretamente o crescimento vegetativo das plantas. Assim, para o cafeeiro é essencial caracterizar o padrão de crescimento da cultura, uma vez que o manejo e a produção estão relacionados com o seu desenvolvimento ao longo do tempo.

Existe uma gama de modelos que podem ser utilizados para modelar o crescimento ao longo do tempo. Dentre esses, os modelos de regressão não lineares se destacam por se adequarem a variados padrões de crescimento e por resumirem as informações presentes nos dados em alguns parâmetros interpretáveis biologicamente (BATISTA et al., 2013; ESPIGOLAN et al., 2013; FERNANDES et al., 2014; SOUSA et al., 2014).

Assim, culturas que apresentam um padrão de crescimento caracterizado por uma fase inicial acelerada, com taxa de crescimento crescente e, posteriormente uma fase inibitória, com taxa de crescimento decrescente tendendo à zero no fim do período avaliado, são bem representados por modelos com formato sigmoidal, como por exemplo, os modelos Logístico, Gompertz e von Bertalanffy (FERNANDES et al., 2015; PEREIRA et al., 2014).

Já o modelo Brody, por apresentar concavidade voltada para baixo, representa bem o padrão de crescimento de espécies que apresentam taxa de crescimento decrescente desde a fase inicial.

Com quatro parâmetros e formato variável, o modelo Richards pode ser utilizado para caracterizar diversos padrões de crescimento (SILVA et al., 2011). A flexibilidade desse modelo é devida ao quarto parâmetro, diferenciando-o dos modelos citados anteriormente, que possuem apenas três.

Para a estimação dos parâmetros desses modelos podem ser utilizados o método dos momentos, da máxima verossimilhança e o de mínimos quadrados. Esse último, consiste em estimar os parâmetros de forma que a soma de quadrados dos resíduos seja mínima e é um dos

mais utilizados (DRAPER; SMITH, 1998).

No entanto, para obter estimativas coerentes e modelar corretamente o padrão de crescimento do cafeeiro (ou de qualquer outra planta ou animal) é essencial considerar na modelagem a natureza dos dados e suas características. Uma dessas características, que em grande parte das pesquisas é desprezada, é a autocorrelação residual, ocasionada pela dependência entre as observações tomadas ao longo do tempo (MAZZINI et al., 2005; MENDES et al., 2008; MUNIZ; NASCIMENTO; FERNANDES, 2017). Segundo Prado, Savian e Muniz (2013), ao ignorar tal fato, pode-se obter estimativas viesadas e a subestimação das variâncias dos parâmetros.

Desse modo, o objetivo deste capítulo foi identificar, dentre os modelos de regressão não lineares Logístico, Brody, von Bertalanffy e Richards, o que melhor caracteriza o padrão de crescimento em altura de plantas do cafeeiro irrigadas e não irrigadas ao longo do tempo, considerando a autocorrelação dos resíduos.

2 MATERIAL E MÉTODOS

Os dados analisados são provenientes de experimento realizado com a cultivar Rubi MG 1192, na área experimental do Departamento de Agricultura da Universidade Federal de Lavras, em Lavras-MG.

O plantio da lavoura foi realizado em janeiro de 2001, utilizando o espaçamento 4m x 1m, que corresponde ao plantio de 2500 plantas por hectare. Durante a condução do experimento, os tratos culturais e o controle fitossanitário foram executados de acordo com as exigências da cultura. As aplicações de adubos e a calagem foram feitas de acordo com a análise de solo e os micronutrientes fornecidos por meio da adubação foliar.

O delineamento utilizado foi em blocos casualizados, com quatro repetições. Em cada bloco foram sorteados seis regimes de irrigação; no entanto, nesse trabalho foram analisados apenas dois regimes: não irrigado (NI) e irrigado à 60 kPa (ao longo do texto será utilizado apenas o termo "irrigado" para fazer referência a esse regime).

O sistema de irrigação implantado na lavoura foi o por gotejamento. As irrigações à 60 kPa ocorreram quando o tensiômetro, na profundidade de 25 cm, registrou no tensímetro esse valor de tensão, que é medido em kilos Pascal (kPa).

Cada unidade experimental foi constituída por três linhas de plantio com dez plantas em cada linha, sendo consideradas úteis as oito plantas centrais da linha central.

A variável altura de planta foi representada pela média das oito plantas úteis, sendo mensuradas trimestralmente, no período que compreendeu maio de 2001 e agosto de 2006, totalizando 22 medições.

Inicialmente os dados foram submetidos à análise de variância. Posteriormente, realizou-se o desdobramento da interação (regimes de irrigação*tempo), que apresentou efeito significativo, estudando o crescimento em altura em função das épocas de avaliação (tempo), em cada regime de irrigação. Para estudar (modelar) esse crescimento, foram utilizados os modelos de regressão não lineares apresentados na Tabela 2.

Tabela 2 Modelos de regressão não lineares ajustados à altura de plantas do cafeeiro em função do tempo.

Modelo	Parametrização
Brody	$y_i = \alpha[1 - \beta e^{-k \cdot t_i}] + e_i$
Logístico	$y_i = \frac{\alpha}{1 + e^{-k \cdot (t_i - \beta)}} + e_i$
von Bertalanffy	$y_i = \alpha[1 - \beta e^{-k \cdot t_i}]^3 + e_i$
Richards	$y_i = \alpha[1 - (1 - \delta)e^{-k \cdot (t_i - \beta)}]^{1/(1-\delta)} + e_i$

Na modelagem do crescimento em altura de plantas do cafeeiro, os termos apresentados na Tabela 2 são interpretados na seguinte forma: y_i corresponde à altura média de plantas do cafeeiro (em cm) observadas ao longo do tempo t_i ; t_i indica as épocas de avaliação, em dias após o plantio, $t_i = 90, 180, \dots, 1980$ dias; α representa a altura máxima a ser atingida pelas plantas na fase adulta; nos modelos Logístico e Richards β indica a abscissa do ponto de inflexão, ou seja, o momento em que as plantas atingem a taxa de crescimento máxima, nos demais modelos é um parâmetro de escala sem interpretação biológica direta; δ é o parâmetro que define o formato da curva e determina em que proporção de α ocorre o ponto de inflexão; k representa o índice de crescimento, de forma que, quanto maior for este valor menos tempo será necessário para as plantas atingirem α ; e e_i é o erro aleatório associado à observação y_i , o qual pressupõe-se que seja independente e identicamente distribuído segundo uma distribuição Normal com média zero e variância homogênea, ou seja, $e_i \sim N(0, \sigma_e^2)$ (MISCHAN; PINHO, 2014).

Inicialmente os parâmetros foram estimados pelo método de mínimos quadrados ordinários, que tem por objetivo minimizar a soma de quadrados dos resíduos (DRAPER; SMITH, 1998). Assim, para um modelo não linear, cuja representação na forma matricial é $\mathbf{Y} =$

$f(X, \boldsymbol{\theta}) + \varepsilon$, tem-se que a soma de quadrados dos resíduos é dada por:

$$\begin{aligned} S(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= [\mathbf{Y} - f(X, \boldsymbol{\theta})]^T [\mathbf{Y} - f(X, \boldsymbol{\theta})] \\ &= \mathbf{Y}^T \mathbf{Y} - 2f(X, \boldsymbol{\theta})^T \mathbf{Y} + f(X, \boldsymbol{\theta})^T f(X, \boldsymbol{\theta}) \end{aligned}$$

em que \mathbf{Y} denota o vetor formado pelas observações, $\boldsymbol{\theta}$ é o vetor formado pelos parâmetros do modelo e ε representa o vetor de resíduos (erro). Derivando a expressão acima e igualando-a a zero, obtém-se o chamado sistema de equações normais não linear, o qual não apresenta solução explícita. Assim, faz-se necessário a utilização de algum método iterativo, como por exemplo, o de Gauss-Newton, de Marquardt e Gradiente, para obtenção das estimativas dos parâmetros que minimizam $S(\boldsymbol{\theta})$ (SILVA et al., 2007).

Para verificar se os resíduos apresentaram autocorrelação significativa foi analisado o gráfico da função de autocorrelação (ρ_k), $k = 0, 1, 2, \dots$, definida por:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

em que, $\gamma_k = Cov[e_i, e_{i-k}]$ é a função de autocovariância. Os resíduos que apresentam autocorrelação significativa devem ser modelados. Uma forma de modelar essa autocorrelação é incorporando parâmetros autoregressivos de ordem (p) na modelagem. A ordem a ser considerada em cada ajuste é obtida analisando o gráfico da função de autocorrelação parcial (ϕ_{kk}), dada por:

$$\phi_{kk} = Cov[e_i, e_{i-k} | e_{i-1}, e_{i-2}, \dots, e_{i-k+1}]$$

Assim, os ajustes que apresentaram autocorrelação residual de ordem 1 foram refeitos pelo método de mínimos quadrados generalizados, incorporando na modelagem o parâmetro autoregressivo de ordem 1, dado por:

$$e_i = \phi_1 e_{i-1} + u_i$$

Analogamente, aos que apresentaram autocorrelação de ordem 2, incorporou-se o parâmetro autoregressivo de ordem 2:

$$e_i = \phi_2 e_{i-2} + \phi_1 e_{i-1} + u_i$$

em que e_i , e_{i-1} e e_{i-2} correspondem aos resíduos gerados nos tempos t_i , t_{i-1} e t_{i-2} , respectivamente; ϕ_1 e ϕ_2 são os parâmetros autoregressivos de ordem 1 e 2, respectivamente; e, u_i é o ruído branco, ou seja, $E[u_i] = 0$, $E[u_i \cdot u_{i-k}] = 0$ para $k \neq 0$ e $E[u_i^2] = \sigma_u^2$ (MORETTIN; TOLOI, 2006).

Para avaliar a qualidade dos ajustes e identificar o modelo que melhor caracteriza o padrão de crescimento em altura do cafeeiro, foram utilizados o desvio padrão residual (DPR) e o critério de informação de Akaike (AIC).

O desvio padrão residual é proporcional ao quadrado médio do erro (QME), sendo calculado da seguinte forma:

$$DPR = \sqrt{QME}$$

Desse modo, quanto menor o valor do DPR, melhor é o ajuste fornecido pelo modelo (PRADO; SAVIAN; MUNIZ, 2013).

O AIC é baseado na minimização da informação (ou distância) de Kullback-Leibler, que é uma medida de proximidade entre o modelo ideal (perfeito) e o modelo candidato. Uma estimativa dessa distância é calculada pela seguinte expressão:

$$AIC = -2\ln(L(\boldsymbol{\theta})) + 2p$$

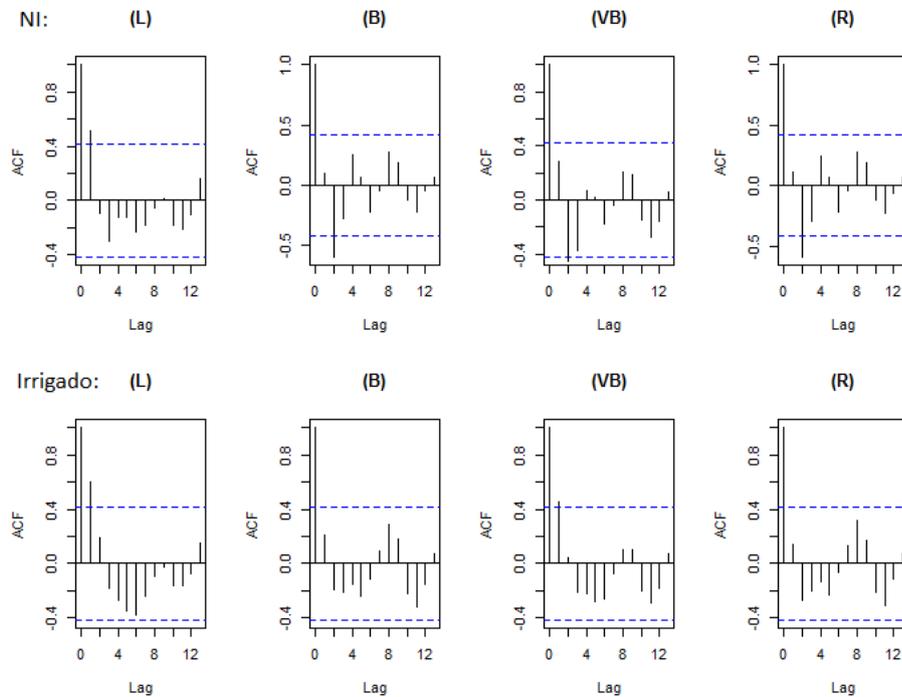
em que $L(\boldsymbol{\theta})$ é a estimativa da função de máxima verossimilhança e p é o número de parâmetros do modelo avaliado. O modelo que apresentar o menor valor de AIC é considerado o mais próximo do modelo ideal e, assim, o que fornece melhor ajuste (AKAIKE, 1974).

Todos os procedimentos necessários para execução desse trabalho, como a estimação dos parâmetros, obtenção de gráficos, entre outros, foram feitos utilizando o software R (R DEVELOPMENT CORE TEAM, 2016). No Anexo A é apresentada uma rotina com as funções utilizadas.

3 RESULTADOS E DISCUSSÃO

Após estimar os parâmetros dos modelos pelo método de mínimos quadrados ordinários (considerando o método iterativo de Gauss-Newton), foram obtidos os gráficos das funções de autocorrelação (FAC) para os resíduos (FIGURA 4).

Figura 4 Função de autocorrelação para os resíduos provenientes dos ajustes dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R) considerando os regimes não irrigado (NI) e irrigado.

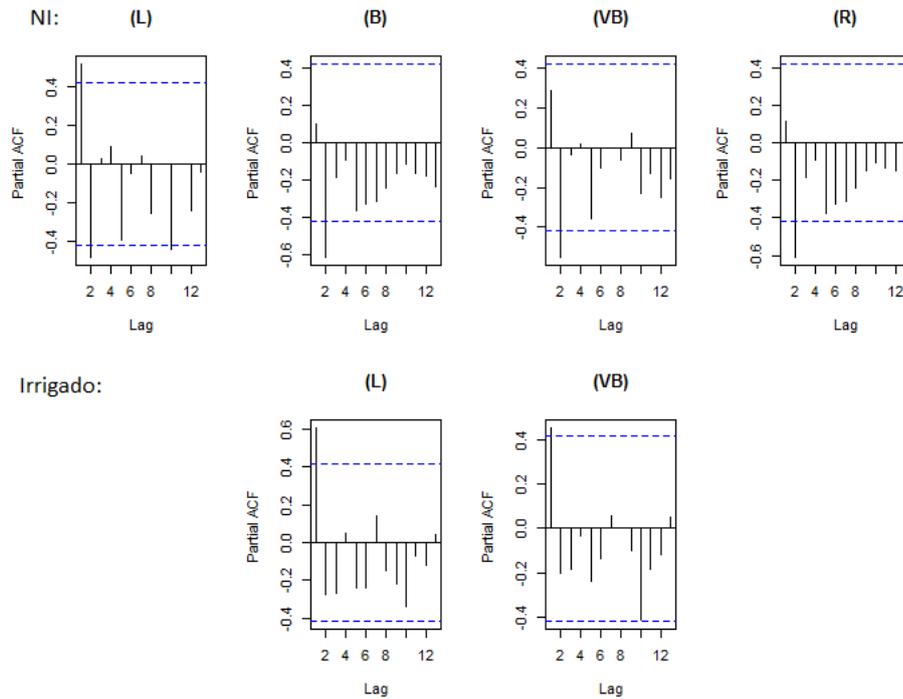


Na Figura 4, a análise da FAC indica que todos os ajustes, referentes ao regime NI, apresentaram autocorrelação residual significativa, uma vez que, outros lags, além do zero, foram significativos (estão fora do intervalo). Em relação ao regime irrigado, verificou-se que apenas os resíduos gerados pelos modelos Logístico e von Bertalanffy são significativamente autocorrelacionados. Logo, como os ajustes dos modelos Brody e Richards referentes ao regime irrigado não apresentaram autocorrelação significativa (pois os lags estão dentro do intervalo) não é necessário incorporar parâmetros autoregressivos no ajuste destes modelos.

De acordo com os gráficos obtidos para a função de autocorrelação parcial (FACP), apresentados na Figura 5, deve-se considerar um processo autoregressivo de ordem 2, denotado por AR(2), para modelar a autocorrelação dos resíduos de todos os modelos ajustados ao regime NI. No entanto, antes de incorporar um processo AR(2), foi considerado o processo AR(1) e verificado se tal ordem foi suficiente para modelar a autocorrelação. Assim, analisando novamente o gráfico da FAC (não exibida), constatou-se que os resíduos gerados pelo ajuste do modelo Logístico com AR(1) mostraram-se independentes, sugerindo que não é necessário considerar um processo AR(2). Nos demais modelos, o processo AR(1) não foi suficiente,

sendo então considerado um processo AR(2). Em relação ao regime irrigado, a FACP indicou a incorporação de processos AR(1) nos ajustes dos modelos Logístico e von Bertalanffy.

Figura 5 Função de autocorrelação parcial para os resíduos provenientes dos ajustes dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R) considerando o regime não irrigado (NI), e para os modelos Logístico (L) e von Bertalanffy (VB) considerando o regime irrigado.



Na tabela a seguir, são apresentadas as estimativas obtidas para os parâmetros dos modelos reajustados, considerando o regime não irrigado (TABELA 3).

Tabela 3 Estimativas dos parâmetros e os respectivos intervalos de confiança (IC) considerando os ajustes dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R) à altura de plantas do cafeeiro não irrigado.

Modelos	Parâmetros	Estimativas	IC _{95%} (.)
L com AR(1)	α	179,2**	[170,16703 ; 188,29687]
	β	725,4**	[662,58329 ; 788,13217]
	k	0,00230**	[0,00198 ; 0,00253]
	ϕ_1	0,5745*	[0,16640 ; 0,98269]
B com AR(2)	α	242,3**	[217,57300 ; 267,03182]
	β	0,9558**	[0,93794 ; 0,97359]
	k	0,00061**	[0,00049 ; 0,00073]
	ϕ_1	0,1636	[-0,21263 ; 0,53988]
	ϕ_2	-0,6146**	[-0,99099 ; -0,23820]
VB com AR(2)	α	198,1**	[186,45990 ; 209,80908]
	β	0,5402**	[0,51575 ; 0,56460]
	k	0,00121**	[0,00106 ; 0,00137]
	ϕ_1	0,4555*	[0,05698 ; 0,85400]
	ϕ_2	-0,5807*	[-0,98510 ; -0,17622]
R com AR(2)	α	233,6**	[176,55949 ; 290,60789]
	β	17,2	[-559,97470 ; 594,43531]
	k	0,00068*	[0,00019 ; 0,00117]
	δ	0,0751	[-0,45049 ; 0,60079]
	ϕ_1	0,1805	[-0,20771 ; 0,56868]
	ϕ_2	-0,6159**	[-1,00463 ; -0,22721]

Teste t: ** significativo a 1%; * significativo a 5%

Ao analisar a Tabela 3, observa-se que, os modelos Brody e Richards apresentam estimativas para o parâmetro α próximas da altura média esperada para o cafeeiro na fase adulta, que segundo o Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (2016), é em torno de 240 cm (2,4 m) para cultivares de baixo porte, como é o caso da Rubi MG 1192. Em contrapartida, os modelos Logístico e von Bertalanffy tendem a subestimar esse parâmetro, apresentando valores inferiores a 200 cm.

Apesar do parâmetro δ possibilitar ao modelo Richards flexibilidade no ajuste, este

parâmetro não foi significativo pelo teste t, o que de certa forma, indica que na modelagem do crescimento em altura do cafeeiro esse quarto parâmetro não seja relevante e, provavelmente, pouco contribuiu na melhoria de qualidade de ajuste.

Em todos os ajustes realizados, as estimativas obtidas para os parâmetros ϕ_1 e ϕ_2 estão coerentes com as condições de estacionariedade estabelecidas na literatura, sendo que, para processos AR(1) deve-se ter $-1 < \phi_1 < 1$ e para processos AR(2) é desejável que $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ e $-1 < \phi_2 < 1$ (MORETTIN; TOLOI, 2006).

Ao estudar o crescimento do fruto do cacau, Muniz, Nascimento e Fernandes (2017) verificaram que, ao incorporar parâmetros que modelam a autocorrelação residual o erro padrão das estimativas e também o desvio padrão residual diminuíram, propiciando estimativas mais precisas e melhor qualidade de ajuste.

As estimativas obtidas para os parâmetros dos modelos (re)ajustados, considerando o regime irrigado, são apresentadas na Tabela 4.

Tabela 4 Estimativas dos parâmetros e os respectivos intervalos de confiança (IC) considerando os ajustes dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R) à altura de plantas do cafeeiro irrigadas.

Modelos	Parâmetros	Estimativas	IC _{95%} (.)
L com AR(1)	α	190,6**	[182,85717 ; 198,42093]
	β	603,7**	[553,87121 ; 653,60685]
	k	0,00240**	[0,00209 ; 0,00265]
	ϕ_1	0,7664**	[0,38959 ; 1,14322]
B	α	222,4**	[207,74948 ; 236,97933]
	β	0,9385**	[0,91180 ; 0,96514]
	k	0,00089**	[0,00075 ; 0,00103]
VB com AR(1)	α	204,0**	[194,00997 ; 214,02016]
	β	0,5024**	[0,47525 ; 0,52963]
	k	0,00140**	[0,00122 ; 0,00158]
	ϕ_1	0,5793*	[0,13106 ; 1,02763]
R	α	234,8**	[186,60480 ; 282,99520]
	β	-273,3	[-447,51400 ; 463,91350]
	k	0,00072**	[0,00020; 0,00124]
	δ	-0,1934	[-0,77283; 0,38603]

Teste t: ** significativo a 1%; * significativo a 5%

Independente do modelo considerado, as maiores estimativas para o parâmetro k foram obtidas no regime irrigado (TABELAS 3 e 4). No entanto, apenas o ajuste do modelo Brody indica que, ao longo do período avaliado, as plantas irrigadas se desenvolveram de forma mais rápida que as não irrigadas, pois os intervalos de confiança obtidos para esse parâmetro não se sobrepõem.

Com base na Tabela 4, verifica-se que, as estimativas obtidas para o parâmetro autoregressivo de ordem 1, incorporado no ajuste dos modelos Logístico e von Bertalanffy, também atendem à condição de estacionariedade: $-1 < \phi_1 < 1$. Para o ajuste do modelo Logístico, por exemplo, a estimativa obtida para ϕ_1 (0,7664), indica que os resíduos apresentaram correlação positiva forte.

Na Tabela 5, são apresentadas as estimativas obtidas para o desvio padrão residual e critério de informação de Akaike, considerando os ajustes dos modelos Logístico, Brody, von Bertalanffy e Richards para ambos os regimes de irrigação estudados.

Tabela 5 Valores obtidos para os avaliadores de qualidade de ajuste: critério de informação de Akaike (AIC) e desvio padrão residual (DPR), para ambos os regimes de irrigação.

Modelos	NI		Modelos	Irigado	
	AIC	DPR		AIC	DPR
L com AR(1)	136,572	4,749	L com AR(1)	138,301	4,940
B com AR(2)	118,270	3,081	B	124,845	3,706
VB com AR(2)	122,258	3,373	VB com AR(1)	128,869	3,987
R com AR(2)	120,173	3,169	R	126,122	3,745

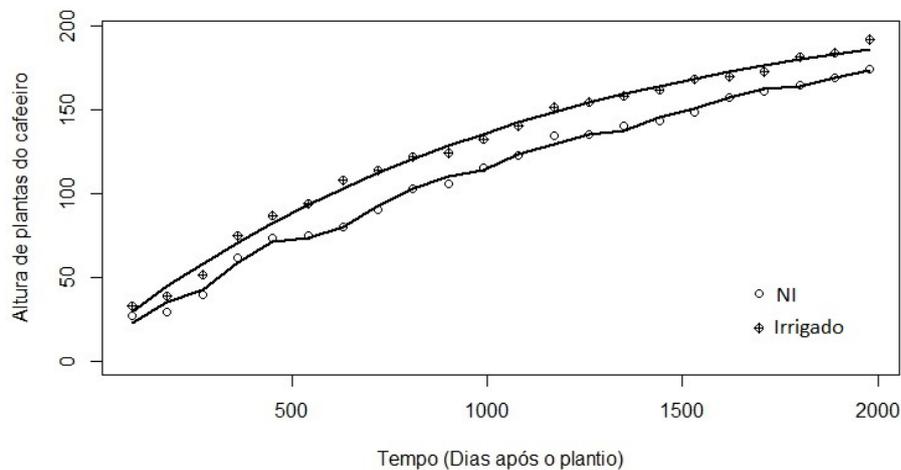
Com base na Tabela 5, verifica-se que, para ambos os regimes estudados, os modelos Brody e Richards apresentaram os menores valores tanto para o AIC quanto para o DPR. No entanto, o modelo Brody se destaca apresentando os melhores resultados, sendo considerado, dentre os modelos avaliados, o que melhor representa o padrão de crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, ao longo do tempo.

Esses resultados divergem dos obtidos por Carvalho et al. (2006) e Pereira et al. (2016), que ao estudarem o crescimento vegetativo do cafeeiro ao longo do tempo, verificaram que os modelos Logístico e Gompertz, respectivamente, descreveram de forma satisfatória o crescimento em altura de plantas. No entanto, vale ressaltar que, tais pesquisas não compararam o ajuste de diferentes modelos com formatos variados em suas análises.

A equivalência entre os modelos Richards e Brody quando $\delta = 0$ (EQUAÇÃO 1) pode possivelmente, explicar o fato de o modelo Richards apresentar o segundo melhor ajuste, uma vez que as estimativas obtidas para esse parâmetro não diferem de zero, conforme verificado pelo teste t (TABELAS 3 e 4).

Os valores observados para a altura média de plantas do cafeeiro, não irrigadas e irrigadas, e a representação gráfica dos ajustes do modelo Brody são apresentados na Figura 6.

Figura 6 Altura média de plantas do cafeeiro e ajustes do modelo Brody considerando os regimes de irrigação NI e irrigado.



O ajuste do modelo Brody considerando o regime NI (FIGURA 6), evidencia como o modelagem da autocorrelação propicia maior qualidade no ajuste, conseguindo captar praticamente toda a variação dos dados.

4 CONCLUSÕES

Considerando os modelos avaliados e os critérios de seleção desvio padrão residual e critério de informação de Akaike, o modelo Brody é o que melhor representa o crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, ao longo do tempo.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v. 19, n. 6, p. 716-723, 1974.
- BATISTA, E. L. da S. et al. Modelagem do crescimento de cultivares de cana-de-açúcar no período de formação da cultura. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 17, n. 10, p. 1080-1087, 2013.
- CAMARGO, A. P.; CAMARGO, M. B. P. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, Campinas, v. 60, n. 1, p. 65-68, 2001.
- CARVALHO, C. H. M. de et al. Evolução do crescimento do cafeeiro (*Coffea arábica* L.) irrigado e não irrigado em duas densidades de plantio. **Ciência e Agrotecnologia**, Lavras, v. 30, n. 2, p. 243-250, mar./abr. 2006.
- CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: café**. Disponível em: <<http://www.conab.gov.br>> Acesso em: 16 Mar. 2016.
- CONSÓRCIO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO DO CAFÉ. **Cultivar Rubi MG 1192**. Disponível em: <<http://www.consorciopesquisacafe.com.br/index.php/tecnologias/separador-8/cultivares/506-rubi-mg-1192>>. Acesso em: 18 Mar. 2016.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd ed. New York: J. Wiley, 1998. 706 p.
- ESPIGOLAN, R. et al. Aplicação de modelos não-lineares para descrever a evolução de características de crescimento e carcaça em bovinos da raça Hereford. **Ciência Rural**, Santa Maria, v. 43, n. 3, mar. 2013.
- FERNANDES, T. J. et al. Parameterization effects in nonlinear models to describe growth curves. **Acta Scientiarum. Technology**, Maringá, v. 37, n. 4, p. 397-402, out./dez. 2015.
- FERNANDES, T. J. et al. Seleção de modelos não lineares para descrição das curvas de crescimento do fruto do cafeeiro. **Coffee Science**, Lavras, v. 9, n. 2, p. 207-215, abr./jun. 2014.
- MAZZINI, A. R. A. et al. Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.
- MENDES, P. N. et al. Modelo logístico difásico no estudo do crescimento de fêmeas da raça Hereford. **Ciência Rural**, Santa Maria, v. 38, n. 7, p. 1984-1990, out. 2008.

MISCHAN, M. M.; PINHO, S. Z. de. **Modelos não lineares: funções assintóticas de crescimento**. São Paulo: Cultura Acadêmica, 2014, 184 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006. 538 p.

MUNIZ, J. A.; NASCIMENTO, M. da S.; FERNANDES, T. J. Nonlinear models for description of cacao fruit growth with assumption violations. **Caatinga**, Mossoró, v. 30, n. 1, p. 250-257, jan./mar. 2017.

PEREIRA, A. A. et al. Descrição do crescimento vegetativo do cafeeiro cultivar Rubi MG 1192 utilizando modelos de regressão. **Coffee Science**, Lavras, v. 9, n. 2, p. 266-274, 2014.

PEREIRA, A. A. et al. Modelagem não linear do crescimento em altura do cafeeiro irrigado e não irrigado em diferentes densidades. **Irriga**, Botucatu, v. 1, n. 1, p. 140-149, 2016.

PEREIRA, S. P. et al. Crescimento, produtividade e bienalidade do cafeeiro em função do espaçamento de cultivo. **Pesquisa Agropecuária Brasileira**, Brasília, v. 46, n. 2, p. 152-160, fev. 2011.

PRADO, T. K. L. do; SAVIAN, T. V.; MUNIZ, J. A. Ajuste dos modelos Gompertz e Logístico aos dados de crescimento de frutos de coqueiro anão verde. **Ciência Rural**, Santa Maria, v. 43, n. 5, p. 803-809, 2013.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 jan. 2016.

RODRIGUES, R. R. et al. Crescimento inicial do cafeeiro conilon sob déficit hídrico no solo. **Coffee Science**, Lavras, v. 11, n. 1, p. 33-38, 2016.

SILVA, N. A. M. et al. Aplicação do Método Bayesiano na estimação de curva de crescimento em animais da raça nelore. **Revista Ceres**, Viçosa, v. 54, n. 312, p. 192-199, mar./abr. 2007.

SILVA, N. A. M. et al. Seleção e classificação multivariada de modelos de crescimento não lineares para bovinos nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 63, n. 2, p. 364-371, 2011.

SOUSA, I. F. et al. Fitting nonlinear autoregressive models to describe coffee seed germination. **Ciência Rural**, Santa Maria, v. 44, p. 2016-2021, 2014.

CAPÍTULO 3: MODELAGEM BAYESIANA DA CURVA DE CRESCIMENTO DO CAFEIEIRO

RESUMO

Na modelagem de curvas de crescimento deve-se considerar que dados longitudinais podem apresentar autocorrelação residual, sendo que, se tal característica não é considerada, os resultados e inferências podem ser comprometidos. A abordagem bayesiana, que considera tanto as observações quanto os parâmetros de um modelo como variáveis aleatórias, tem se mostrado eficiente na estimação de parâmetros. No entanto, como geralmente não é possível obter as distribuições marginais de forma analítica, faz-se necessário a utilização de algum método, como o método de reamostragem ponderada, para gerar amostras dessas distribuições e assim obter uma aproximação para as mesmas. Dentre as vantagens desse método, destaca-se a geração de amostras independentes e o fato de não ser necessário avaliar convergência. Diante desse contexto, os objetivos deste capítulo foram: apresentar a modelagem não linear bayesiana do crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, considerando a autocorrelação residual e os modelos não lineares Logístico, Brody, von Bertalanffy e Richards; e apresentar um estudo de simulação com o intuito de validar o método de reamostragem ponderada utilizado na estimação de parâmetros, considerando como distribuição candidata uma aproximação da verossimilhança. Dentre os resultados, verificou-se que, no estudo de simulação, os intervalos de máxima densidade a posteriori obtidos para todos os parâmetros dos modelos avaliados, inclusive os parâmetros que modelam a autocorrelação residual, contiveram os valores paramétricos pré-estabelecidos, validando a eficiência do método da reamostragem ponderada, considerando como distribuição candidata uma aproximação da verossimilhança. Para as plantas não irrigadas, o critério de informação bayesiano e o critério da densidade preditiva ordenada, indicaram que, dentre os modelos avaliados, o modelo Logístico é o que melhor descreve o crescimento em altura do cafeeiro ao longo do tempo. E, para as plantas irrigadas, esses mesmos critérios indicaram o modelo Brody.

Palavras-chave: Autocorrelação residual. Metodologia bayesiana. Modelos não lineares.

ABSTRACT

In the modeling of growth curves, we must consider that the longitudinal data may present residual autocorrelation, and if this characteristic is not considered, the results and inferences can be compromised. The Bayesian approach, which considers both the observations and the parameters of a model as random variables, has been shown to be efficient in parameter estimation. However, it is not possible to obtain marginal distributions analytically. Therefore, a method of use, such as the weighted resampling method, is necessary to generate samples from these distributions and then obtain an approximation for them. Among the advantages of this method, we highlight the generation of independent samples and the fact that it is not necessary to evaluate convergence. In this context, the purposes of this chapter were: to present a simulation study with the purpose of validating the method of weighted resampling in the parameters estimation of the nonlinear models considering as candidate distribution a likelihood approximation. We also present the Bayesian nonlinear modeling of height growth of irrigated and non-irrigated coffee plants, considering the residual autocorrelation by Logistic, Brody, von Bertalanffy and Richards models. Among the results, we verified that, in the simulation study, the high posterior density intervals obtained for all the parameters of the evaluated models, including the parameters of residual autocorrelation, contained the pre-established parametric values, validating the efficiency of the weighted resampling method, considering as candidate distribution an approximation of likelihood. For non-irrigated system, the Bayesian information criterion and the criterion of ordered predictive density indicated, among the studied models, the Logistic model as the one that best describes the growth of the coffee plant height over time. In addition, for the irrigated system, the same criteria indicated the Brody model.

Keywords: Residual autocorrelation. Bayesian methodology. Nonlinear models.

1 INTRODUÇÃO

Devido à importância econômica do café para o Brasil, um dos principais focos de pesquisa do setor agrícola refere-se ao desenvolvimento do cafeeiro. Uma variável relevante a ser considerada em tais pesquisas é a altura de plantas, pois além de ser representativa do desenvolvimento vegetativo está correlacionada com a produtividade (ASSIS et al., 2014; PEREIRA et al., 2016).

A modelagem da curva de crescimento do cafeeiro por meio de modelos de regressão não lineares permite ao pesquisador a descrição do crescimento ao longo do tempo e a obtenção de informações de interesse prático, como a altura máxima na fase adulta e a velocidade de crescimento, que são importantes para caracterizar o desenvolvimento das plantas (PEREIRA et al., 2014; REIS et al., 2014).

Como já comentado na seção 1 do capítulo 2, para modelar corretamente a curva de crescimento deve-se considerar que os dados de crescimento tomados ao longo do tempo podem apresentar autocorrelação residual (PEREIRA et al., 2016). No entanto, em diversas pesquisas aplicadas, verifica-se que essa característica não é incorporada na modelagem, podendo comprometer os resultados e inferências (MAZZINI et al., 2005; MENDES et al., 2008; MUNIZ; NASCIMENTO; FERNANDES, 2017).

Uma abordagem que tem se mostrado eficiente na estimação dos parâmetros é a bayesiana, que considera tanto as observações quanto os parâmetros do modelo como variáveis aleatórias. Dessa forma, especifica-se uma distribuição a priori para os parâmetros, denotada por $P(\theta)$ - que representa em termos de probabilidade o conhecimento pré-existente sobre os parâmetros - e uma distribuição de probabilidade conjunta para os dados amostrais, chamada função de verossimilhança $L(\theta|Y)$ (BOX; TIAO, 1992; PAULINO; TURKMAN; MURTEIRA, 2003). A combinação dessas informações, pelo teorema de Bayes, resulta na distribuição a posteriori de θ . Essa distribuição, indicada por $P(\theta|Y)$, representa o conhecimento atualizado sobre os parâmetros, associando o conhecimento inicial do pesquisador com a informação proveniente da amostra:

$$P(\theta|Y) \propto L(\theta|Y).P(\theta)$$

Dentre as vantagens da abordagem bayesiana, destaca-se a modelagem satisfatória mesmo com uma amostra relativamente pequena e a obtenção de estimativas intervalares mais precisas (MACEDO et al., 2014; MARTINS FILHO et al., 2008; SILVA et al., 2007).

Para fazer inferências sobre determinado parâmetro θ_i , $i = 1, 2, \dots, p$, deve-se obter a distribuição marginal a posteriori desse parâmetro, que é dada pela integral da distribuição a posteriori com relação aos demais parâmetros. Devido à complexidade dessas integrais, na maioria das vezes, a obtenção das distribuições marginais não é feita de forma analítica, fazendo-se necessário a utilização de algum método de aproximação. Os métodos de Monte Carlo via Cadeias de Markov (MCMC) têm sido os mais utilizados na geração de amostras da distribuição marginal a posteriori, possibilitando, desta forma, a obtenção aproximada da densidade. Entretanto, por serem iterativos, esses métodos podem demorar para convergir e exigem que o pesquisador utilize diagnósticos de avaliação da convergência (COLE et al., 2012; FERNANDES, 2016).

Um método alternativo aos MCMC, que gera amostras independentes, não requer avaliação de convergência e é computacionalmente mais rápido, é o método de reamostragem ponderada, proposto por Rubin (1987) (COLE et al., 2012; LOPES; POLSON; CARVALHO, 2012). A geração das amostras por esse método, também conhecido por amostragem-reamostragem por importância (SIR), é feita em duas etapas: inicialmente gera-se valores de uma distribuição candidata, e, na segunda etapa, é feita uma reamostragem considerando o peso (probabilidade) atribuído a cada valor gerado.

Apesar de existirem na literatura pesquisas aplicadas apresentando o ajuste de modelos de regressão por meio de métodos bayesianos, geralmente tais pesquisas ou utilizam modelos não lineares e se limitam à pressuposição de independência residual (ANDRADE FILHO et al.; 2010; FORNI et al., 2009; GUEDES et al., 2004; MARTINS FILHO et al.; 2008; MENDES, 2011) ou consideram a autocorrelação e utilizam modelos lineares (CHIB, 1993; CHIB; GREENBERG, 1994; MENZEFRICKE, 1999). Sendo que, especificamente no estudo do crescimento do cafeeiro, não se encontram na literatura pesquisas com enfoque bayesiano.

Diante desse contexto, os objetivos deste capítulo foram: apresentar a modelagem não linear bayesiana do crescimento em altura de plantas do cafeeiro, irrigadas e não irrigadas, considerando a autocorrelação residual e os modelos não lineares Logístico, Brody, von Bertalanffy e Richards; e apresentar um estudo de simulação com o intuito de validar o método de reamostragem ponderada utilizado na estimação de parâmetros, considerando como distribuição candidata uma aproximação da verossimilhança.

2 MATERIAL E MÉTODOS

Antes de realizar a modelagem dos dados reais, foi feito um estudo de simulação de modo a validar a metodologia da reamostragem ponderada, proposta para a estimação dos parâmetros no contexto bayesiano.

O método da reamostragem ponderada consiste em obter uma amostra de uma distribuição $f(\theta)$, que não possui forma conhecida, gerando valores de uma distribuição candidata, denotada por $g(\theta)$, que seja possível de ser amostrada e que esteja definida no mesmo domínio de $f(\theta)$ (SMITH; GELFAND, 1992). Fundamentado na teoria de amostragem por importância (RUBIN, 1987), ao multiplicar e dividir $f(\cdot)$ por $g(\cdot)$ obtém-se:

$$\frac{f(\theta)}{g(\theta)} \cdot g(\theta)$$

Assim, ao gerar amostras de $g(\theta)$ e ponderá-las por pesos obtidos com base em $\frac{f(\theta)}{g(\theta)}$ obtém-se uma amostra aproximada de $f(\theta)$. Os passos necessários para execução do método de reamostragem ponderada são apresentados a seguir:

i) Gerar uma amostra de tamanho M de $g(\theta)$:

$$(\theta_1, \theta_2, \dots, \theta_M)$$

ii) Calcular os pesos, denotados por w_j , $j = 1, 2, \dots, M$, para cada amostra gerada:

$$w_j = \frac{f(\theta_j)}{g(\theta_j)}$$

iii) Obter as probabilidades q_j , $j = 1, 2, \dots, M$, normalizando os pesos:

$$q_j = \frac{w_j}{\sum_{t=1}^M w_t}$$

iv) Reamostrar de $(\theta_1, \theta_2, \dots, \theta_M)$ com base nas probabilidades q_j , obtendo uma nova amostra de tamanho m , com $m \leq M$:

$$(\theta_1^*, \theta_2^*, \dots, \theta_m^*)$$

Smith e Gelfand (1992) mostraram que, a amostra obtida pelo método de reamostragem ponderada no passo iv) constitui uma amostra de $f(\theta^*)$ que, a medida que $m \rightarrow \infty$, converge

em distribuição para $f(\boldsymbol{\theta})$. Em análises bayesianas, têm-se que a distribuição $f(\cdot)$ de interesse é a distribuição a posteriori $P(\boldsymbol{\theta}|\mathbf{Y})$.

Esses mesmos autores sugerem utilizar como candidata a distribuição a priori, $P(\boldsymbol{\theta})$. No entanto, nesse trabalho, considerou-se como distribuição candidata uma aproximação da verossimilhança, denotada por $Q(\boldsymbol{\theta})$, criada com base no ponto de máximo da verossimilhança ($\hat{\boldsymbol{\theta}}$) e na decomposição de Cholesky da matriz de covariâncias dos parâmetros, Σ . Assim, as amostras foram obtidas da seguinte forma:

i) Assumindo que $Q(\boldsymbol{\theta}) \sim N(\hat{\boldsymbol{\theta}}, \Sigma)$, utilizou-se a seguinte transformação linear para gerar uma amostra de tamanho $M=11000$ da distribuição candidata:

$$Q(\boldsymbol{\theta}_j) = \hat{\boldsymbol{\theta}} + L \cdot \mathbf{Z}_j$$

em que, L é a matriz triangular inferior resultante da decomposição de Cholesky da matriz Σ e \mathbf{Z}_j é o vetor simulado, $\mathbf{Z}_j \sim N(\mathbf{0}, I)$, sendo I uma matriz identidade. Segundo Fernandes (2016), ao considerar essa distribuição candidata, o cálculo dos pesos e das probabilidades são simplificados, sendo obtidos diretamente da distribuição a priori, pois como assume-se normalidade para o modelo, tem-se que $Q(\boldsymbol{\theta})$ converge em distribuição para $L(\boldsymbol{\theta}|\mathbf{Y})$ a medida que $M \rightarrow \infty$. Assim, os passos ii) e iii) foram calculados da seguinte maneira:

ii) Para $1 \leq j \leq 11000$, calculou-se w_j e q_j :

$$w_j = \frac{P(\boldsymbol{\theta}_j|\mathbf{Y})}{Q(\boldsymbol{\theta}_j)} = \frac{L(\boldsymbol{\theta}_j|\mathbf{Y}) \cdot P(\boldsymbol{\theta}_j)}{Q(\boldsymbol{\theta}_j)} \simeq \frac{L(\boldsymbol{\theta}_j|\mathbf{Y}) \cdot P(\boldsymbol{\theta}_j)}{L(\boldsymbol{\theta}_j|\mathbf{Y})} = P(\boldsymbol{\theta}_j)$$

iii)

$$q_j = \frac{P(\boldsymbol{\theta}_j)}{\sum_{t=1}^M P(\boldsymbol{\theta}_t)} = P(\boldsymbol{\theta}_j)$$

No passo iii), o denominador das probabilidades q_j tende a um quando $M \rightarrow \infty$. Logo, as probabilidades q_j são iguais aos pesos w_j .

iv) Considerando as probabilidades calculadas, foi feita a reamostragem com reposição da amostra gerada no passo i), obtendo uma nova amostra de tamanho $m = 10000$.

De acordo com Ehlers (2007) e Straub e Papaioannou (2015), por ser um método não iterativo, o método de reamostragem ponderada tem por vantagens a geração de amostras independentes e o fato de não ser necessário avaliar convergência.

No estudo de simulação foram considerados os mesmos modelos de regressão não

lineares vistos no capítulo 2, sendo que, para realizar um estudo abrangente envolvendo a autocorrelação, foram analisadas quatro situações: Logístico com resíduos autocorrelacionados de ordem 1 - AR(1), Brody com AR(2), von Bertalanffy com resíduos autocorrelacionados de ordem 1 e 2 e, Richards com resíduos independentes.

Para cada modelo, foram simulados 22 valores de altura de plantas do cafeeiro (y_i), sendo que o i -ésimo valor gerado corresponde ao tempo t_i , $t_i = 90, 180, \dots, 1980$ dias; foram gerados também valores para os resíduos autocorrelacionados e_i . Os valores paramétricos para α , β , k e δ , apresentados na Tabela 6, foram determinados com base em resultados (clássicos) da literatura e os valores de ϕ_1 e ϕ_2 foram escolhidos de modo a caracterizar as quatro situações de autocorrelação analisadas para os resíduos.

Tabela 6 Valores paramétricos atribuídos aos parâmetros dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R) no estudo de simulação.

Parâmetros	L	B	VB	R
α	250	250	250	250
β	900	1	1	900
k	0,004	0,004	0,004	0,004
ϕ_1	0,8	0	0,8	0
ϕ_2	0	-0,6	-0,4	0
δ	-	-	-	2,5

Considerando que os resíduos (ruído branco) seguem uma distribuição Normal, ou seja, $u_i \sim N(0, \sigma_u^2 = 4)$, tem-se a seguinte expressão para a função de verossimilhança:

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{i=1}^{22} \frac{1}{\sqrt{2\pi\sigma_u^2}} \cdot \exp \left[\frac{-1}{2\sigma_u^2} (y_i - \mu_i)^2 \right]$$

em que \mathbf{Y} representa o vetor formado pelos valores simulados e o valor esperado μ_i e $\boldsymbol{\theta}$ - que denota o vetor formado pelos parâmetros de interesse - variam conforme o modelo (TABELA 7).

Tabela 7 Valores esperados considerando os modelos Logístico, Brody, von Bertalanffy e Richards.

Modelos	Valor esperado (μ_i)
Logístico com AR(1)	$\frac{\alpha}{1+e^{-k \cdot (t_i-\beta)}} + \phi_1 e_{i-1}$
Brody com AR(2)	$\alpha[1 - \beta e^{-k \cdot t_i}] + \phi_2 e_{i-2}$
Von Bertalanffy com AR(1) e AR(2)	$\alpha[1 - \beta e^{-k \cdot t_i}]^3 + \phi_1 e_{i-1} + \phi_2 e_{i-2}$
Richards	$\alpha[1 - (1 - \delta)e^{-k \cdot (t_i-\beta)}]^{1-\delta}$

Nas parametrizações apresentadas na Tabela 7, t_i indica as épocas de avaliação, em dias após o plantio: $t_i = 90, 180, \dots, 1980$ dias; α representa a altura máxima esperada; β indica a abscissa do ponto de inflexão nos modelos Logístico e Richards, nos demais modelos é um parâmetro de escala sem interpretação biológica direta; δ é o parâmetro que define o formato da curva e determina em que proporção de α ocorre o ponto de inflexão; k denota o índice de crescimento e determina a eficiência do crescimento: quanto maior for este valor menos tempo será necessário para α ser atingido; ϕ_1 e ϕ_2 são os parâmetros autoregressivos de ordem 1 e 2 respectivamente; e e_i é o resíduo autocorrelacionado.

As distribuições a priori foram elicítadas com base na técnica conhecida como priori de especialista, que visa determinar um intervalo de ocorrência plausível para os parâmetros e alguma medida de posição, como por exemplo a média, de acordo com o conhecimento de um especialista sobre o fenômeno em estudo. Segundo Moala e Penha (2016) a elicitação de distribuições a priori para os parâmetros constitui uma fase importante na realização da análise bayesiana.

Assim, para o parâmetro α , que indica uma estimativa da altura máxima a ser atingida pelas plantas do cafeeiro, logo $\alpha > 0$, foi considerada como distribuição a priori uma Gama com hiperparâmetros 10 e 0,04.

Para o parâmetro β , que nos modelos Logístico e Richards indica o momento em que a taxa de crescimento é máxima (ponto de inflexão) e nos demais modelos não apresenta interpretação prática direta, também foi considerada uma distribuição Gama com hiperparâmetros 1,5 e 0,003.

O parâmetro k , que é interpretado como índice de crescimento, em geral assume valores pequenos, concentrados próximo ao limite inferior do intervalo (0,1). Desta forma, como distribuição a priori foi atribuída uma Beta(2,25; 1500).

Ao utilizar o modelo Richards, a especificação de uma distribuição a priori para o

parâmetro δ também é necessária. Para esse parâmetro, que pode assumir valores positivos ou negativos, foi considerada uma Normal com média 0,11 e desvio padrão 10.

Foi definida como priori para ϕ_1 uma distribuição Beta generalizada, cujo intervalo de ocorrência é $[-1, 1]$, pois segundo Morettin e Toloi (2006), para que o processo seja estacionário a seguinte condição deve ser satisfeita:

$$|\phi_1| < 1$$

A obtenção da distribuição Beta generalizada é feita da seguinte forma: seja $V \sim Beta(h_1, h_2)$ definida no intervalo $(0, 1)$, então a variável aleatória definida por $W = (n - m)V + m$ no intervalo finito $[m, n]$ segue uma distribuição Beta generalizada (h_1, h_2) (MCDONALD; XU, 1995). Deste modo, ao considerar $V \sim Beta(3, 2)$, a variável aleatória ϕ_1 definida como $\phi_1 = (1 + 1).V - 1$ tem distribuição Beta generalizada $(3, 2)$ definida no intervalo $[-1, 1]$. Portanto, a distribuição a priori de ϕ_1 é dada por $[2.Beta(3, 2)-1]$.

A distribuição a priori estabelecida para o parâmetro ϕ_2 foi uma Uniforme $(-1, 1)$, sendo observadas as seguintes condições de estacionariedade:

$$\phi_1 + \phi_2 < 1$$

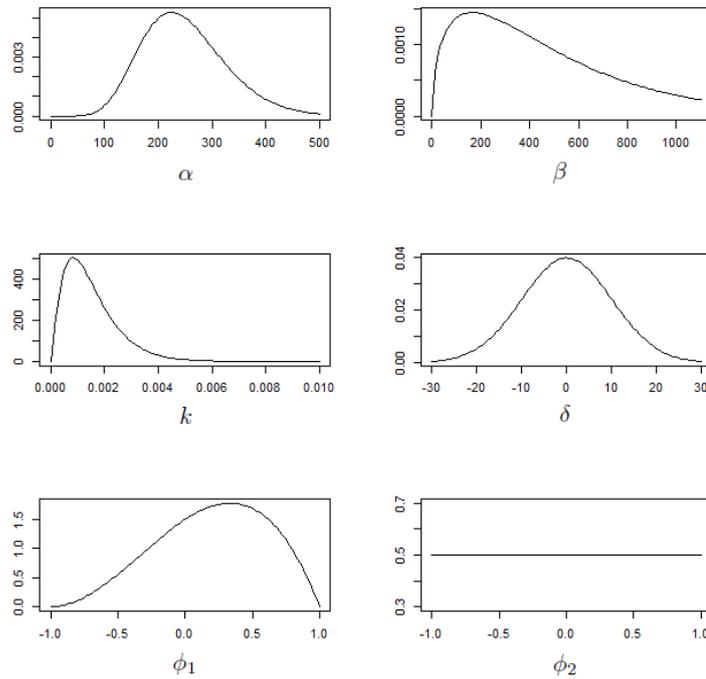
$$\phi_2 - \phi_1 < 1$$

$$|\phi_2| < 1$$

Vale ressaltar que, a determinação dos hiperparâmetros das distribuições a priori (que foi feita de modo a obter variância grande) e a atribuição de uma distribuição pouco informativa para o parâmetro ϕ_2 , se baseia no fato da literatura se mostrar escassa em pesquisas na área vegetal relacionadas a modelagem bayesiana de curvas de crescimento considerando a autocorrelação entre os resíduos.

O formato das distribuições a priori, utilizadas tanto no estudo de simulação quanto na análise dos dados reais, pode ser observado na Figura 7.

Figura 7 Representação gráfica das distribuições a priori atribuídas aos parâmetros α , β , k , δ , ϕ_1 e ϕ_2 .



Para cada modelo analisado, considerou-se independência entre os parâmetros. Assim, a distribuição a priori conjunta para o vetor de parâmetros de interesse θ , $P(\theta)$, é dada pelo produto da priori de cada parâmetro.

Utilizando o teorema de Bayes e substituindo as expressões obtidas para a função de verossimilhança e a distribuição a priori conjunta, tem-se que a distribuição a posteriori $P(\theta|\mathbf{Y})$, para cada caso, é dada por:

$$P(\theta|\mathbf{Y}) \propto \exp \left[\frac{-1}{2\sigma_u^2} \sum_{i=1}^{22} (y_i - \mu_i)^2 \right] \cdot P(\theta)$$

Para obter um resumo informativo sobre a distribuição marginal a posteriori de cada parâmetro, faz-se necessário a resolução de integrais da forma:

$$E[h(\theta|\mathbf{Y})] = \int h(\theta|\mathbf{Y}) \cdot P(\theta|\mathbf{Y}) d\theta$$

em que $h(\theta)$ é uma função convenientemente escolhida. No entanto, como não é possível resolver essas integrais de forma analítica, foi utilizado o método de reamostragem ponderada, descrito anteriormente, para gerar amostras. Especificamente no estudo de simulação, os passos do método foram repetidos 100 vezes, e como amostra final, foi tomada em cada posição, a média dessas repetições. Posteriormente, foram obtidas as estimativas da média, moda e

do intervalo de credibilidade de máxima densidade a posteriori (HPD) para os parâmetros. Se os intervalos HPD obtidos contiverem os valores paramétricos a eficiência do método é comprovada.

A análise dos dados reais foi feita considerando os mesmos conjuntos de dados e os mesmos modelos utilizados no capítulo 2.

Para cada modelo considerado, a presença e a ordem da autocorrelação residual foram verificadas por meio dos gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) (MORETTIN; TOLOI, 2006). (Esses gráficos já foram apresentados na seção 3 do capítulo 2).

Novamente, considerando que os resíduos (ruído branco) seguem uma distribuição Normal, ou seja, $u_i \sim N(0, \sigma_u^2)$, tem-se a seguinte expressão para a função de verossimilhança:

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{i=1}^{22} \frac{1}{\sqrt{2\pi\sigma_u^2}} \cdot \exp\left[\frac{-1}{2\sigma_u^2}(y_i - \mu_i)^2\right]$$

em que \mathbf{Y} representa o vetor formado pelos valores observados y_i e, o valor esperado μ_i e $\boldsymbol{\theta}$ variam conforme o modelo e a estrutura de autocorrelação indicada pelos gráficos da FAC e FACP (TABELA 8).

Tabela 8 Valor esperado e estrutura de autocorrelação para os resíduos dos modelos Logístico, Brody, von Bertalanffy e Richards considerando os dados reais referentes aos regimes NI e irrigado.

NI	Valor esperado (μ_i)
Logístico com AR(1)	$\frac{\alpha}{1+e^{-k \cdot (t_i - \beta)}} + \phi_1 e_{i-1}$
Brody com AR(2)	$\alpha[1 - \beta e^{-k \cdot t_i}] + \phi_1 e_{i-1} + \phi_2 e_{i-2}$
Von Bertalanffy com AR(2)	$\alpha[1 - \beta e^{-k \cdot t_i}]^3 + \phi_1 e_{i-1} + \phi_2 e_{i-2}$
Richards com AR(2)	$\alpha[1 - (1 - \delta)e^{-k \cdot (t_i - \beta)}]^{1-\delta} + \phi_1 e_{i-1} + \phi_2 e_{i-2}$
Irrigado	Valor esperado (μ_i)
Logístico com AR(1)	$\frac{\alpha}{1+e^{-k \cdot (t_i - \beta)}} + \phi_1 e_{i-1}$
Brody	$\alpha[1 - \beta e^{-k \cdot t_i}]$
von Bertalanffy com AR(1)	$\alpha[1 - \beta e^{-k \cdot t_i}]^3 + \phi_1 e_{i-1}$
Richards	$\alpha[1 - (1 - \delta)e^{-k \cdot (t_i - \beta)}]^{1-\delta}$

Para cada modelo que apresentou autocorrelação residual, e_i denota os resíduos correlacionados provenientes do ajuste do próprio modelo realizado pelo método de mínimos quadrados

ordinários.

Analogamente, a distribuição a posteriori $P(\boldsymbol{\theta}|\mathbf{Y})$, para cada caso, é dada por:

$$P(\boldsymbol{\theta}|\mathbf{Y}) \propto \exp \left[\frac{-1}{2\sigma_u^2} \sum_{i=1}^{22} (y_i - \mu_i)^2 \right] \cdot P(\boldsymbol{\theta})$$

Após a geração das amostras pelo método de reamostragem ponderada, foram obtidas as estimativas da média e moda a posteriori e intervalo de credibilidade de máxima densidade a posteriori (HPD) para os parâmetros.

Para indicar o modelo que apresentou melhor ajuste, considerando a abordagem bayesiana, foram utilizados o critério de informação bayesiano (BIC) e o critério da densidade preditiva ordenada (CPO).

O BIC é um critério de seleção que pondera entre o máximo do valor esperado a posteriori da função log-verossimilhança e a quantidade de parâmetros do modelo, sendo calculado da seguinte forma:

$$BIC = -2.E[\ln(L(\boldsymbol{\theta}|\mathbf{Y}))] + (p + 1).ln(n)$$

em que $E[\ln(L(\boldsymbol{\theta}|\mathbf{Y}))]$ é o valor esperado da função log-verossimilhança com relação a distribuição a posteriori, p é o número de parâmetros do modelo e n é o tamanho da amostra. O modelo que apresentar o menor valor para o BIC é considerado o mais adequado (CARLIN; LOUIS, 2000).

O CPO é baseado na densidade preditiva de uma observação y_i dada as demais $\mathbf{Y}_{(-i)}$. O cálculo dessa densidade é dado por:

$$c_i = f(y_i|\mathbf{Y}_{(-i)}) = \int f(y_i|\mathbf{Y}_{(-i)}, \boldsymbol{\theta}) \cdot P(\boldsymbol{\theta}|\mathbf{Y}_{(-i)}) d\boldsymbol{\theta}$$

em que $P(\boldsymbol{\theta}|\mathbf{Y}_{(-i)})$ é a densidade a posteriori de $\boldsymbol{\theta}$ condicionado aos dados $\mathbf{Y}_{(-i)}$. Utilizando as amostras geradas $\boldsymbol{\theta}_j^*$, $j = 1, 2, \dots, m$, a expressão 2 pode ser aproximada por:

$$\hat{c}_i = \hat{f}(y_i|\mathbf{Y}_{(-i)}) = \frac{1}{m} \sum_{j=1}^m f(y_i|\mathbf{Y}_{(-i)}, \boldsymbol{\theta}_j^*)$$

Se c_i é um valor pequeno, isso indica que a observação y_i é pouco provável para o modelo em questão. Dessa forma, o CPO consiste em escolher o modelo que apresentar o maior valor para

$$\ln(C) = \sum_{i=1}^n \ln(c_i) \text{ (ANDRADE FILHO et al., 2010).}$$

Todos os procedimentos necessários para execução desse trabalho, como a geração das amostras, resumo das distribuições, obtenção de gráficos, entre outros, foram feitos utilizando o software R (R DEVELOPMENT CORE TEAM, 2016). No Anexo B é apresentada uma rotina com as funções utilizadas.

3 RESULTADOS E DISCUSSÃO

Os resultados são divididos em duas seções: a primeira é referente ao estudo de simulação e, a segunda refere-se à análise dos dois conjuntos de dados reais.

3.1 Análise dos dados simulados

Nessa seção são apresentados os resultados obtidos para as simulações referentes às quatro situações consideradas: modelo Logístico com AR(1), Brody com AR(2), von Bertalanffy com AR(1) e AR(2) e Richards com resíduos independentes.

As estimativas obtidas para a média e moda a posteriori e para o intervalo HPD são exibidas na Tabela 9.

Tabela 9 Estimativas da média e moda a posteriori e intervalo HPD para os parâmetros dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R), considerando os dados simulados.

Modelos	Valores paramétricos	Média	moda	HPD _{95%}
L com AR(1)	$\alpha=250$	250,4	250,3	[249,06502; 251,78976]
	$\beta=900$	902,7	902,5	[897,34402; 908,16256]
	$k=0,004$	0,00393	0,00393	[0,00387; 0,00410]
	$\phi_1=0,8$	0,6223	0,6373	[0,40055; 0,84253]
B com AR(2)	$\alpha=250$	250,1	250,1	[249,64000; 250,66328]
	$\beta=1$	0,9978	0,9979	[0,98474; 1,01041]
	$k=0,004$	0,00398	0,00398	[0,00391; 0,00404]
	$\phi_2=-0,6$	-0,4069	-0,4153	[-0,72750; -0,11319]
VB com AR(1) e AR(2)	$\alpha=250$	249,7	249,7	[248,83456; 250,49408]
	$\beta=1$	0,9861	0,9863	[0,96008; 1,01192]
	$k=0,004$	0,00398	0,00398	[0,00391; 0,00406]
	$\phi_1=0,8$	0,5709	0,5895	[0,31642; 0,82239]
R	$\phi_2=-0,4$	-0,3823	-0,3895	[-0,68359; -0,07201]
	$\alpha=250$	250,3	250,3	[248,61520; 251,90922]
	$\beta=900$	913,1	913,7	[898,95892; 927,32920]
	$k=0,004$	0,00409	0,00409	[0,00386; 0,00433]
	$\delta=2,5$	2,6290	2,6318	[2,46596; 2,80050]

Com base na Tabela 9, pode-se verificar que, de modo geral, as estimativas da média e da moda a posteriori para os parâmetros α , β , k estão bem próximas, indicando simetria para as distribuições marginais. Essa simetria pode ser percebida também analisando os histogramas construídos para cada parâmetro, apresentados nas Figuras 8, 9, 10 e 11.

Ainda em relação à Tabela 9, nota-se que as estimativas obtidas tanto para a média quanto para a moda dos parâmetros que modelam a autocorrelação, ϕ_1 e ϕ_2 , estão em conformidade com as condições de estacionariedade estabelecidas na literatura, sendo que, para processos AR(1) deve-se ter: $-1 < \phi_1 < 1$ e, para processos AR(2) é desejável que: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ e $-1 < \phi_2 < 1$ (MORETTIN; TOLOI, 2006).

O fato de todos os intervalos de máxima densidade a posteriori conterem os valores paramétricos pré-estabelecidos (TABELA 9) validam a eficiência do método da reamostragem

ponderada, considerando como distribuição candidata uma aproximação da verossimilhança, na estimação de parâmetros de modelos não lineares.

Figura 8 Histogramas das distribuições marginais dos parâmetros do modelo Logístico com resíduos autocorrelacionados de ordem 1, considerando dados simulados.

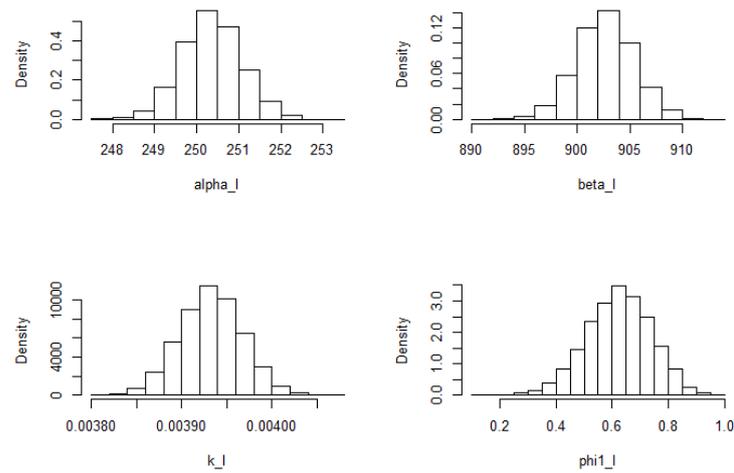


Figura 9 Histogramas das distribuições marginais dos parâmetros do modelo Brody com resíduos autocorrelacionados de ordem 2, considerando dados simulados.

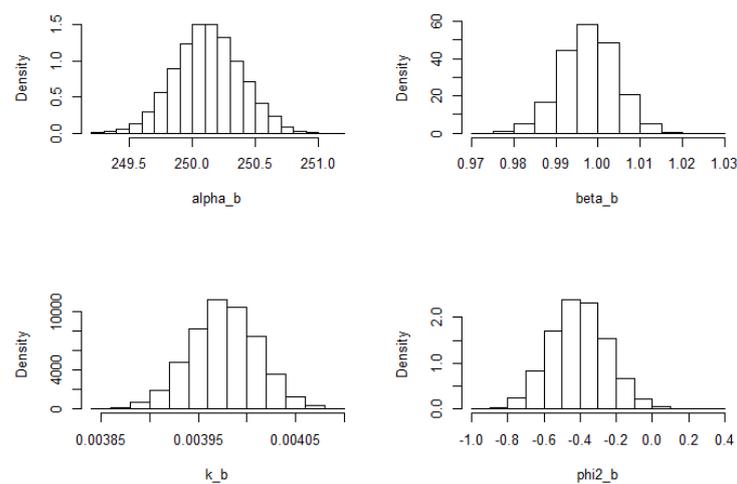


Figura 10 Histogramas das distribuições marginais dos parâmetros do modelo von Bertalanffy com resíduos autocorrelacionados de ordem 1 e 2, considerando dados simulados.

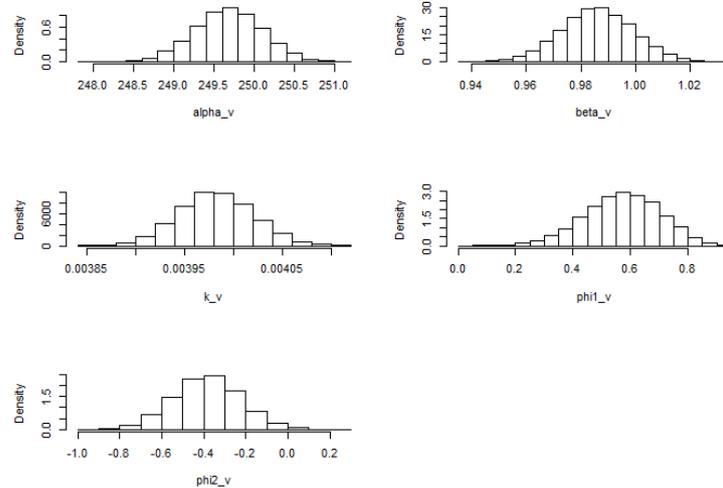
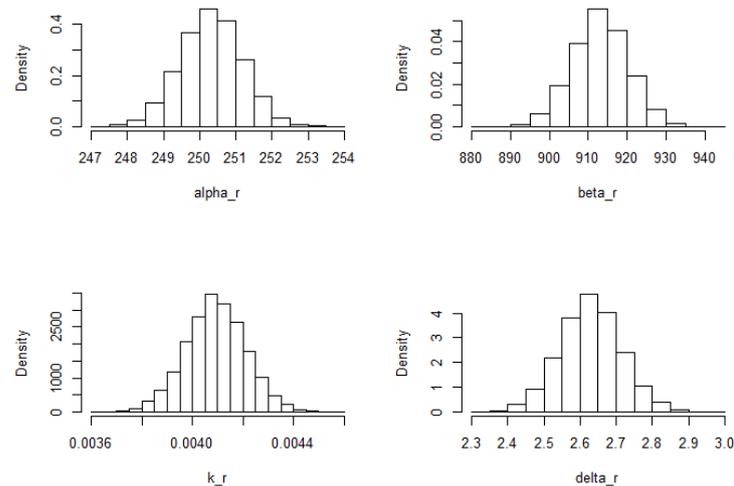


Figura 11 Histogramas das distribuições marginais dos parâmetros do modelo Richards com resíduos independentes, considerando dados simulados.



3.2 Ajuste bayesiano de modelos não lineares à altura de plantas do cafeeiro

Na Tabela 10, são apresentadas as estimativas obtidas para a média e moda a posteriori assim como o intervalo HPD, para todos os parâmetros dos modelos ajustados, considerando os dados referentes ao regime não irrigado (NI).

Tabela 10 Estimativas a posteriori para a média, moda e intervalo HPD dos parâmetros dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R), considerando os dados referentes ao regime não irrigado.

Modelos	Parâmetros	Média	Moda	HPD _{95%}
L com AR(1)	α	179,5	179,3	[171,50677 ; 186,60914]
	β	726,2	726,0	[672,05286 ; 778,59692]
	k	0,00224	0,00223	[0,00202 ; 0,00247]
	ϕ_1	0,5194	0,5190	[0,21992 ; 0,81319]
B com AR(2)	α	241,5	242,4	[221,09566 ; 261,97139]
	β	0,9563	0,9563	[0,94188 ; 0,97142]
	k	0,00061	0,00061	[0,00051 ; 0,00071]
	ϕ_1	0,1794	0,1758	[-0,13248 ; 0,46234]
	ϕ_2	-0,6010	-0,5952	[-0,91725 ; -0,32907]
VB com AR(2)	α	198,5	198,9	[188,70933 ; 208,02190]
	β	0,5399	0,5394	[0,51928 ; 0,55977]
	k	0,00121	0,00120	[0,00108 ; 0,00134]
	ϕ_1	0,4292	0,4381	[0,14370 ; 0,74024]
	ϕ_2	-0,5577	-0,5547	[-0,89605 ; -0,25650]
R com AR(2)	α	215,4	219,5	[189,42799 ; 242,42765]
	β	217,9	157,3	[9,93506 ; 425,34463]
	k	0,00084	0,00081	[0,00064 ; 0,00105]
	δ	0,2575	0,1950	[0,06191 ; 0,44509]
	ϕ_1	0,1879	0,2019	[-0,12893 ; 0,49771]
	ϕ_2	-0,6048	-0,6002	[-0,90362 ; -0,28628]

Analisando a Tabela 10, verifica-se que, de modo geral, as estimativas para a média e a moda a posteriori dos parâmetros β e k para os modelos Logístico, Brody e von Bertalanffy são bem próximas, indicando que a distribuição marginal a posteriori desses parâmetros é do tipo simétrica. Ao analisar o crescimento do fruto do tomateiro, Mendes (2011) também verificou que as marginais dos parâmetros β e k , para o modelo Logístico são do tipo simétricas. Os histogramas apresentados nas Figuras 12, 13 e 14 mostram esse comportamento, indicando uma semelhança com a distribuição Normal, por exemplo.

No entanto, ao comparar o quantil superior do intervalo HPD da distribuição marginal do

parâmetro k referente ao modelo Brody (q_{hpd}), cuja área abaixo dele é de 97,5%, com o quantil da distribuição Normal (q_n), considerando $\bar{k} = 0,00061$ e $s_k = 0,00006$ (não apresentado na tabela) têm-se que: $q_{hpd} = 0,00071$ e $q_n = 0,000727$. Assim, como $q_n > q_{hpd}$ conclui-se que a Normal possui caudas mais pesadas que a marginal do parâmetro k .

Analogamente, para o parâmetro β do modelo Brody, têm-se que $q_{hpd} = 0,97142$ e $q_n = 0,974174$ (considerando $\bar{\beta} = 0,9563$ e $s_\beta = 0,00912$). Logo, como $q_n > q_{hpd}$ segue que a Normal possui caudas mais pesadas que a marginal do parâmetro β .

Já em relação às estimativas obtidas para os parâmetros do modelo Richards, nota-se uma diferença acentuada entre os valores da média e moda a posteriori, indicando que as marginais dos parâmetros são assimétricas, o que pode também ser verificado pela análise gráfica da Figura 15. Para os parâmetros β , k , δ e ϕ_2 nota-se distribuições assimétricas a direita.

Ainda em relação à Tabela 10, pode ser verificado que, de modo geral, os intervalos HPD obtidos para os parâmetros foram significativos, reafirmando a eficiência da metodologia bayesiana na estimação de parâmetros, já relatada em diversos trabalhos, tais como Andrade Filho et al. (2010), Martins Filho et al. (2008) e Savian et al. (2009), entre outros.

Ao analisar as Figuras 12, 13, 14 e 15, nota-se que, em geral, ocorre uma modificação no formato das distribuições marginais dos parâmetros quando comparadas com as distribuições utilizadas a priori (FIGURA 7). Isso ocorre pois a informação presente nos dados ($L(\theta|Y)$) atualiza o conhecimento a priori ($P(\theta)$).

Figura 12 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Logístico considerando autocorrelação residual de ordem 1 e os dados referentes ao regime NI.

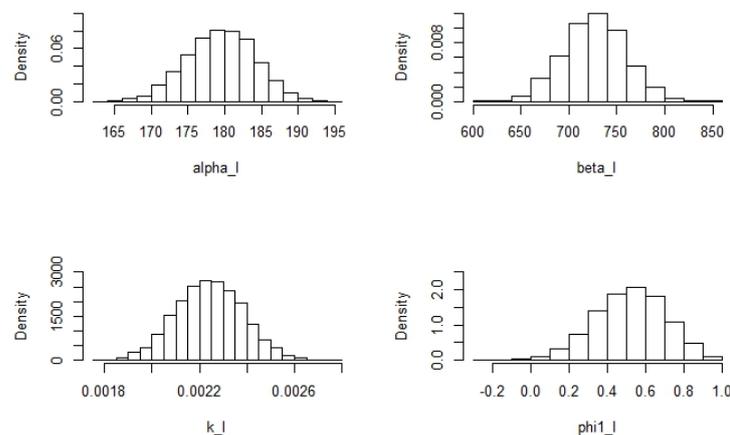


Figura 13 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Brody, considerando autocorrelação residual de ordem 2 e os dados referentes ao regime NI.

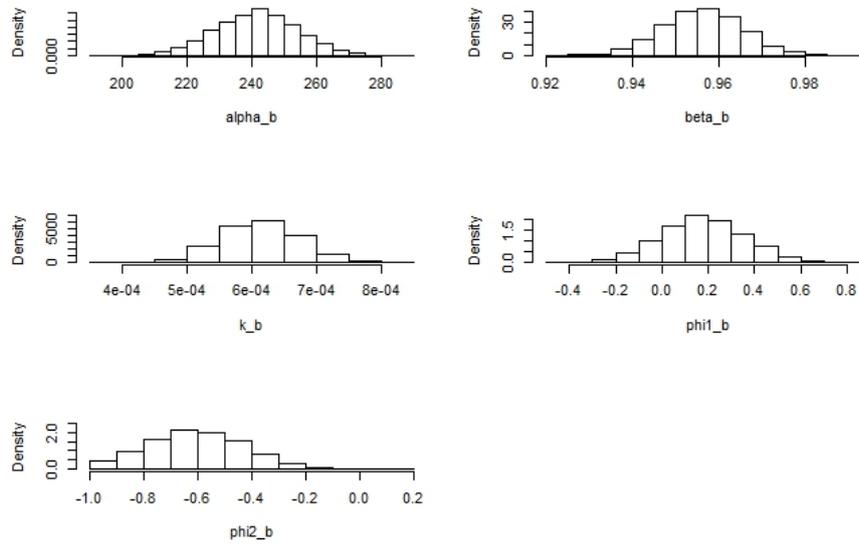


Figura 14 Histogramas amostrais das distribuições marginais dos parâmetros do modelo von Bertalanffy, considerando autocorrelação residual de ordem 2 e os dados referentes ao regime NI.

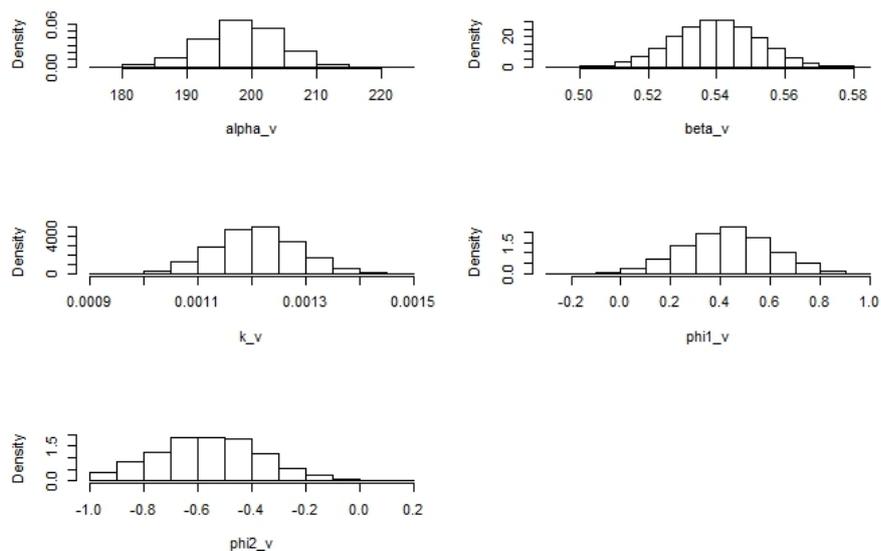
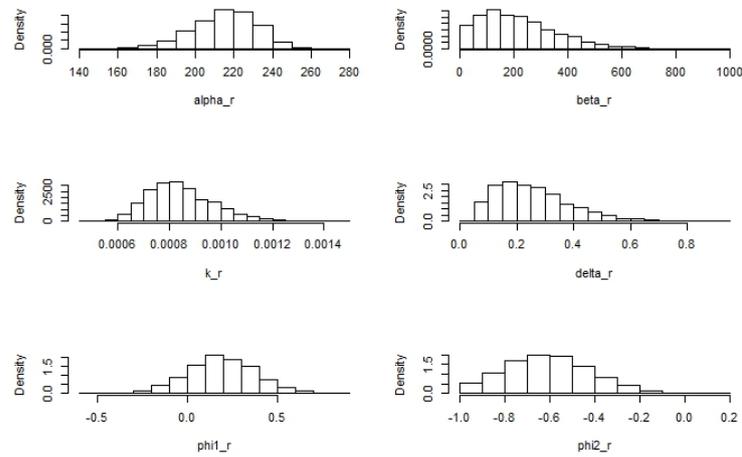


Figura 15 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Richards considerando autocorrelação residual de ordem 2 e os dados referentes ao regime NI.



Na Tabela 11 são apresentadas as estimativas obtidas para a média e moda a posteriori assim como o intervalo HPD, para todos os parâmetros dos modelos ajustados, considerando os dados referentes ao regime irrigado.

Tabela 11 Estimativas a posteriori para a média, moda e intervalo HPD dos parâmetros dos modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R), considerando os dados referentes ao regime irrigado.

Modelos	Parâmetros	Média	Moda	HPD _{95%}
L com AR(1)	α	190,5	190,7	[184,00125 ; 196,90064]
	β	602,4	599,8	[557,92723 ; 642,93790]
	k	0,00237	0,00237	[0,00212 ; 0,00259]
	ϕ_1	0,6597	0,7019	[0,43056 ; 0,91784]
B	α	222,3	222,1	[210,58702 ; 234,77761]
	β	0,9386	0,9396	[0,91767 ; 0,96146]
	k	0,00089	0,00088	[0,00077 ; 0,00100]
VB com AR(1)	α	204,1	204,2	[195,48090 ; 212,08980]
	β	0,5021	0,5013	[0,47916 ; 0,52327]
	k	0,00140	0,00140	[0,00125 ; 0,00154]
	ϕ_1	0,5139	0,5419	[0,19440 ; 0,81738]
R	α	206,5	208,8	[188,23719 ; 223,16551]
	β	189,5	98,9	[0,91410 ; 376,65874]
	k	0,00104	0,00101	[0,00088 ; 0,00119]
	δ	0,1693	0,1211	[0,00858 ; 0,33983]

Com base na Tabela 11, verifica-se que, também no regime irrigado, os intervalos HPD foram significativos para todos os parâmetros. O intervalo obtido para o parâmetro α , considerando o ajuste do modelo Brody, por exemplo, indica que as alturas máximas mais prováveis, para plantas do cafeeiro na fase adulta, estão entre 210,6 cm e 234,8 cm.

Os intervalos HPD obtidos para todos os modelos (TABELA 11) apresentaram menor amplitude quando comparados com os respectivos intervalos de confiança obtidos na modelagem clássica (TABELA 4), indicando que os intervalos obtidos pela metodologia bayesiana são mais precisos (MARTINS FILHO, et al., 2008; SILVA et al., 2005). Essa maior precisão, pode possivelmente, estar associada ao fato de que, nos métodos clássicos, a estimação é feita exclusivamente com base nas informações presentes nos dados; já nos métodos bayesianos, a estimação é feita com base na combinação da informação presente nos dados com o conhecimento inicial sobre os parâmetros (ROSSI, 2011).

Ainda em relação à Tabela 11, verifica-se uma diferença acentuada entre as estimativas

para a média e a moda a posteriori do parâmetro ϕ_1 para os modelos von Bertalanffy e Logístico, sugerindo que as marginais são do tipo assimétricas. Essa assimetria (à esquerda) pode ser confirmada pela análise gráfica dos histogramas apresentados nas Figuras 16 e 17.

As diferenças entre os valores obtidos para a média e moda a posteriori para todos os parâmetros do modelo Richards também se mostraram acentuadas, indicando que as distribuições marginais para os seus parâmetros são do tipo assimétricas, conforme evidencia a Figura 18. Tal comportamento também foi verificado para as distribuições marginais dos parâmetros desse modelo considerando o regime não irrigado.

O formato das distribuições marginais para todos os parâmetros dos quatro modelos não lineares ajustados pode ser visualizado nas Figuras 16, 17, 18 e 19.

Figura 16 Histogramas amostrais das distribuições marginais dos parâmetros do modelo von Bertalanffy considerando autocorrelação residual de ordem 1 e os dados referentes ao regime irrigado.

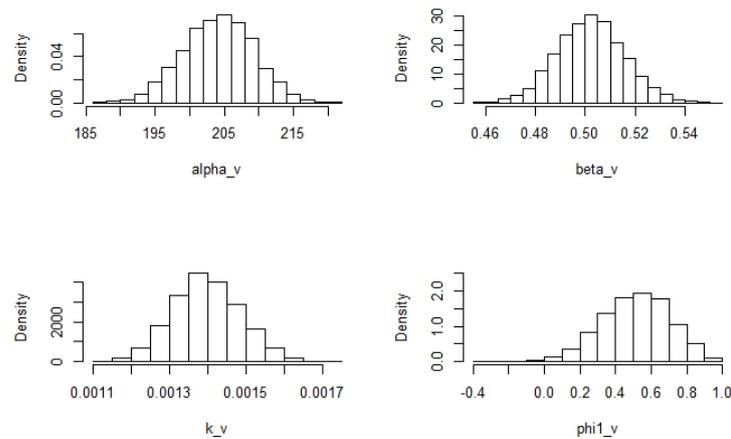


Figura 17 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Logístico, considerando autocorrelação residual de ordem 1 e os dados referentes ao regime irrigado.

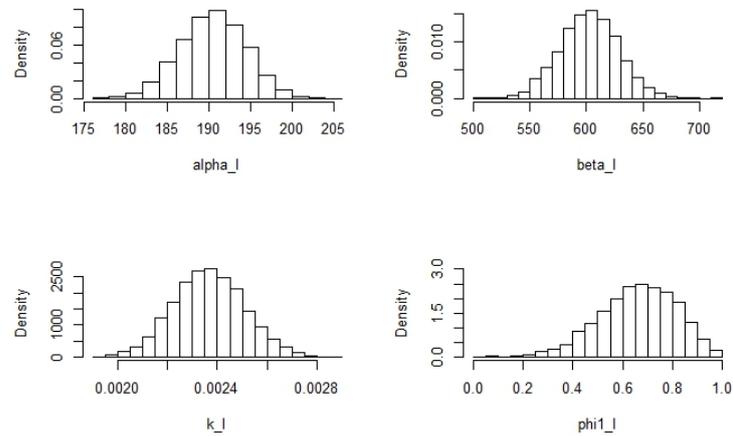


Figura 18 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Richards considerando os dados referentes ao regime irrigado.

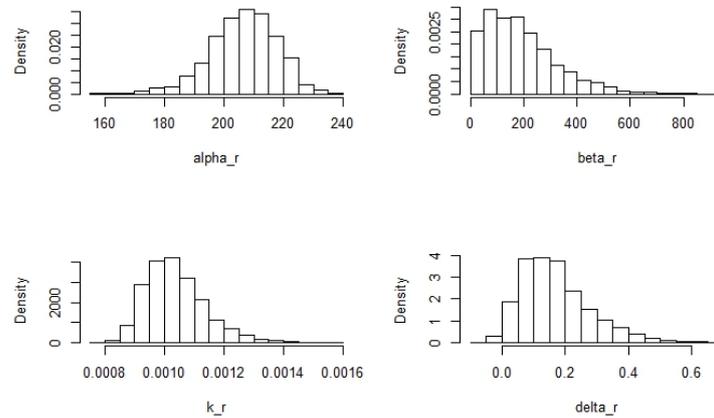
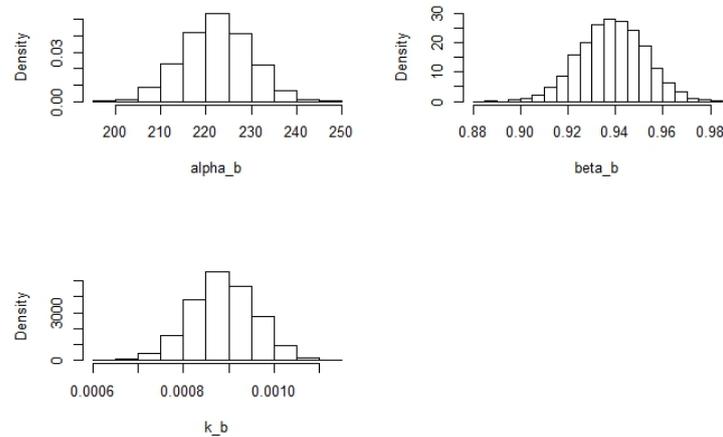


Figura 19 Histogramas amostrais das distribuições marginais dos parâmetros do modelo Brody considerando os dados referentes ao regime irrigado.



Para comparar a qualidade de ajuste fornecida pelos quatro modelos analisados, foram calculados o critério de informação bayesiano (BIC) e o critério da densidade preditiva ordenada (CPO), cujos resultados são apresentados na Tabela 12.

Tabela 12 Estimativas para os critérios de seleção BIC e CPO, considerando os dados referentes aos regimes não irrigado (NI) e irrigado e os modelos Logístico (L), Brody (B), von Bertalanffy (VB) e Richards (R).

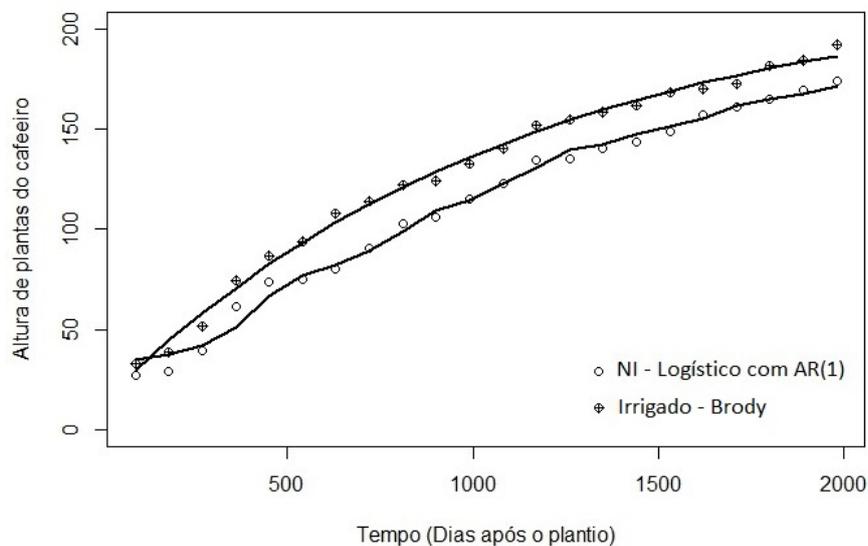
Modelos	NI		Irrigado		
	CPO	BIC	Modelos	CPO	BIC
L com AR(1)	-267,779	36,573	L com AR(1)	-268,981	36,851
B com AR(2)	-309,362	39,569	B	-260,085	33,366
VB com AR(2)	-284,650	39,582	VB com AR(1)	-267,883	36,563
R com AR(2)	-402,479	63,915	R	-316,759	59,042

Considerando o regime NI, as estimativas obtidas para ambos os critérios adotados indicam que, dentre os modelos avaliados, o Logístico é o que melhor descreve o crescimento em altura de plantas do cafeeiro, pois tal modelo apresentou o menor valor para o BIC e o maior para o CPO (TABELA 12). Em relação ao regime irrigado, as estimativas do CPO e BIC indicaram que o modelo Brody, dentre os modelos avaliados, é o que melhor representa a curva de crescimento em altura do cafeeiro irrigado. Esses resultados divergem dos obtidos

na modelagem clássica realizada no capítulo 2, pois com base nos critérios adotados, o modelo Brody foi o que apresentou melhor ajuste para ambos os regimes estudados (TABELA 5).

Os valores observados para a altura média de plantas do cafeeiro não irrigadas e a representação gráfica do ajuste bayesiano do modelo Logístico com AR(1), assim como os valores observados referentes às plantas irrigadas e o ajuste bayesiano do modelo Brody são apresentados na Figura 20.

Figura 20 Altura média de plantas do cafeeiro e ajustes bayesianos dos modelos Logístico com AR(1) e Brody considerando os regimes NI e irrigado, respectivamente.



4 CONCLUSÕES

A eficiência do método de reamostragem ponderada, considerando como distribuição candidata uma aproximação da verossimilhança, foi verificada pelo estudo de simulação.

Para as plantas não irrigadas, o critério de informação bayesiano e o critério da densidade preditiva ordenada, indicaram que, dentre os modelos avaliados, o modelo Logístico é o que melhor descreve o crescimento em altura do cafeeiro ao longo do tempo. E, para as plantas irrigadas, esses mesmos critérios indicaram o modelo Brody.

REFERÊNCIAS

- ANDRADE FILHO, M. G. de et al. Uma abordagem bayesiana para curvas de crescimento com diferentes densidade a priori. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 2, p. 161-181, 2010.
- ASSIS, G. A. de et al. Correlação entre crescimento e produtividade do cafeeiro em função do regime hídrico e densidade de plantio. **Bioscience Journal**, Uberlândia, v. 30, n. 3, p. 666-676, 2014.
- BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: J. Wiley, 1992. 588p.
- CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical Bayes methods for data analysis**. 2nd ed. Boca Raton: Chapman and Hall Press, 2000. 418p.
- CHIB, S. Bayes regression with autoregressive errors: a Gibbs sampling approach. **Journal of Econometrics**, North Holland, v. 58, p. 275-294, 1993.
- CHIB, S.; GREENBERG, E. Bayes inference in regression models with ARMA (p, q) errors. **Journal of Econometrics**, North Holland, v. 64, p. 183-206, 1994.
- COLE, S. R. et al. Bayesian posterior distributions without markov chains. **American Journal of Epidemiology**, Oxford, v. 175, n. 5, p. 368-375, Feb. 2012.
- EHLERS, R. S. **Introdução à inferência bayesiana**. Curitiba: Departamento de Estatística - UFPR, 2007, 102p. Disponível em: <<http://leg.ufpr.br/ehlers/bayes>>. Acesso em: 28 mai. 2014.
- FERNANDES, T. J. **Modelos duplo sigmoidais para o crescimento de frutos do cafeeiro**. 2016. 88 p. Tese (Doutorado em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras. 2016.
- FORNI, S. et al. Comparison of different nonlinear functions to describe Nelore cattle growth. **Journal of Animal Science**, Champaign, v. 87, n. 2, p. 496-506, 2009.
- GUEDES, M. H. P. et al. Estudo das curvas de crescimento de cordeiros das raças Santa Inês e Bergamácia considerando heterogeneidade de variâncias. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 2, p. 381-388, mar./abr. 2004.
- LOPES, H. F.; POLSON, N. G.; CARVALHO, C. M. Bayesian statistics with a smile: a resampling-sampling perspective. **Brazilian Journal of Probability and Statistics**, São Paulo, v. 26, n. 4, p. 358-371, 2012.

- MACEDO, L. R. de et al. Modelagem hierárquica bayesiana na avaliação de curvas de crescimento de suínos genotipados para o gene halotano. **Ciência Rural**, Santa Maria, v. 44, n. 10, p. 1853-1859, out. 2014.
- MARTINS FILHO, S. et al. Abordagem bayesiana das curvas de crescimento de duas cultivares de feijoeiro. **Ciência Rural**, Santa Maria, v. 38, n. 6, p. 1516-1521, 2008.
- MAZZINI, A. R. A. et al. Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.
- MCDONALD, J. B.; XU, Y. J. A generalization of the beta distribution with applications. **Journal of Econometrics**, Amsterdam, v. 66, n. 1, p. 133-152, Mar./Apr. 1995.
- MENDES, P. N. **Ajuste dos modelos Gompertz e Logístico na descrição das curvas de crescimento de tomates: uma abordagem bayesiana**. 2011. 82 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2011.
- MENDES, P. N. et al. Modelo logístico difásico no estudo do crescimento de fêmeas da raça Hereford. **Ciência Rural**, Santa Maria, v. 38, n. 7, p. 1984-1990, out. 2008.
- MENZEFRICKE, U. Bayesian prediction in growth-curve models with correlated errors. **Test**, Pamplona, v. 8, n. 1, p. 75-93, 1999.
- MOALA, F. A.; PENHA, D. L. Métodos de elicitação da distribuição à priori Beta. **Revista Brasileira de Biometria**, Lavras, v. 34, n. 1, p. 49-62, jan. 2016.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006. 538 p.
- MUNIZ, J. A.; NASCIMENTO, M. da S.; FERNANDES, T. J. Nonlinear models for description of cacao fruit growth with assumption violations. **Caatinga**, Mossoró, v. 30, n. 1, p. 250-257, jan./mar. 2017.
- PAULINO, C. D.; TURKMAN, A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.
- PEREIRA, A. A. et al. Descrição do crescimento vegetativo do cafeeiro cultivar Rubi MG 1192 utilizando modelos de regressão. **Coffee Science**, Lavras, v. 9, n. 2, p. 266-274, abr./jun. 2014.
- PEREIRA, A. A. et al. Modelagem não linear do crescimento em altura do cafeeiro irrigado e não irrigado em diferentes densidades. **Irriga**, Botucatu, v. 1, n. 1, p. 140-149, 2016.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <<http://www.r-project.org>>. Acesso em: 20 jan. 2016.

REIS, R. M. et al. Modelos de regressão não linear aplicados a grupos de acessos de alho. **Horticultura Brasileira**, Vitória da Conquista, v. 32, n. 2, p. 178-183, abr./jun., 2014.

ROSSI, R. M. **Introdução aos métodos bayesianos na análise de dados zootécnicos com uso do WinBUGS e R.** Maringá: Eduem, 2011. 191 p.

RUBIN, D. B. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. **Journal of the American Statistical Association**, New York, v. 82, n. 398, p. 543-546, 1987.

SAVIAN, T. V. et al. Análise bayesiana para modelos de degradabilidade ruminal. **Ciência Rural**, Santa Maria, v. 39, n. 7, p. 2169-2177, out. 2009.

SILVA, F. F. e et al. Abordagem bayesiana da curva de lactação de cabras Saanen de primeira e segunda ordem de parto. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 1, p. 27-33, 2005.

SILVA, N. A. M. et al. Aplicação do Método Bayesiano na estimação de curva de crescimento em animais da raça nelore. **Revista Ceres**, Viçosa, v. 54, n. 312, p. 192-199, mar./abr. 2007.

SMITH, A. F. M.; GELFAND, A. E. Bayesian statistics without tears: a sampling-resampling perspective. **The American Statistician**, Washington, v. 46, n. 2 p. 84-88, May 1992.

STRAUB, D.; PAPAIOANNOU, I. Bayesian updating with structural reliability methods. **Journal of Engineering Mechanics**, Houston, v. 141, n. 34, p. 1-13, Mar. 2015.

ANEXOS

ANEXO A - Rotina no R referente à modelagem não linear clássica.

```

# pacote necessário
library(nlme)

# valores observados para a variável
# resposta (altura de plantas)
y<-c(y1,y2,...,y22)

# valores da variável independente (tempo)
t<-c(t1,t2,...,t22)

# ajuste clássico do modelo Brody
# (estabelecer os valores iniciais)
b<-nls(y~A*(1-B*exp(-K*t)), start=c(A=a1,B=b1,k=k1))

# FAC para verificar se a autocorrelação é significativa
acf(residuals(b), main="(B)")

# FACP para verificar a ordem do AR(p)
# a ser incorporada na modelagem
pacf(residuals(b), main="(B)")

# ajuste clássico do modelo Brody incorporando
# parâmetros autoregressivos - AR(2)
bc2<-gnls(y~A*(1-B*exp(-K*t))+P1*e1_b+P2*e2_b,
start=c(A=a1,B=b1,K=k1,P1=p1,P2=p2))

# estimativas e teste t para os parâmetros
# e critérios de seleção
summary(bc2)

# intervalos de confiança (com 5 casas decimais)
round(confint(bc2), digits=5)

# gráficos
plot(t,y,xlab="Variável independente",
ylab="Variável dependente")
lines(t,fitted(b))
lines(t,fitted(bc2), lty=2)

```

ANEXO B - Rotina no R referente à modelagem não linear bayesiana (dados reais).

```

# pacotes necessários
library(nlme)
library(coda)
library(modeest)

# valores observados para a
# variável resposta (altura de plantas)
y<-c(y1,y2,...,y22)

# valores da variável independente (tempo)
t<-c(t1,t2,...,t22)

# ajuste clássico do modelo Brody
# (estabelecer os valores iniciais)
b<-nls(y~A*(1-B*exp(-K*t)), start=c(A=a1,B=b1,k=k1))

# ajuste clássico do modelo Brody
# incorporando parâmetros autoregressivos - AR(2)
bc2<-gnls(y~A*(1-B*exp(-K*t))+P1*e1_b+P2*e2_b,
start=c(A=a1,B=b1,K=k1,P1=p1,P2=p2))

# Definição das prioris logaritimizadas
# (estabelecer hiperparâmetros)
Priori_b<-function(theta){
dgamma(theta[1],shape=c1,rate=c2,log=TRUE)+# alpha: gama
dgamma(theta[2],shape=c3,rate=c4,log=TRUE)+# beta: gama
dbeta(theta[3],c5,c6,log=TRUE)+           # k: beta
dbeta((theta[4]+1)/2,c7,c8,log=TRUE)+     # phi1: beta gen.
dunif(theta[5],c9,c10,log=TRUE)         # phi2: uniforme
}

# Estimativa do ponto de máximo da verossimilhança
theta_b<-bc2$coefficients

# Matriz de covariâncias dos parâmetros
sigma2_b<-bc2$varBeta

# decomposição de Cholesky da matriz
# de covariâncias dos parâmetros
L_b<-t(chol(sigma2_b))

# Método de reamostragem ponderada
M<-11000
m<-10000
w<-0*1:N

```

```

THETA<-matrix(0,M,5)

# geração de amostras da dist. candidata
# (aproximação da verossimilhança)
for(i in 1:M){
  THETA[i,1:5]<-theta_b+L_b%*%rnorm(5)
  # cálculo dos pesos
  w[i]<-lPriori_b(THETA[i,1:5])
}
posicao<-sample(1:M,m,replace=TRUE,prob=exp(w))
# reamostragem com base nas probabilidades
posteriori_b<-THETA[posicao,]

# amostra final das marginais
Afb<-posteriori_b[,1]
Bfb<-posteriori_b[,2]
Kfb<-posteriori_b[,3]
P1fb<-posteriori_b[,4]
P2fb<-posteriori_b[,5]

# verificação das condições de estacionariedade para AR(2)
estac<-numeric(10000)
for (i in 1:10000){
  if (P1fb[i]+P2fb[i]<1)
  if (P2fb[i]-P1fb[i]<1)
  if (abs(P2fb[i])<1)
  {
    estac[i]<-0
  }else estac[i]<-1
}
sum(estac)

# obtenção dos intervalos HPD (com 5 casas decimais)
amfinal_b<-mcmc(posteriori_b)
round(HPDinterval(amfinal_b,prob=0.95),digits=5)

# obtenção das médias a posteriori
summary(amfinal_b)

# obtenção das modas a posteriori
mlv(amfinal_b[,1])
mlv(amfinal_b[,2])
mlv(amfinal_b[,3])
mlv(amfinal_b[,4])
mlv(amfinal_b[,5])

# histogramas
par(mfrow=c(3,2))
hist(Afb,prob=T,xlab="alpha_b",main="")

```

```
hist(Bfb,prob=T,xlab="beta_b",main="")  
hist(Kfb,prob=T,xlab="k_b",main="")  
hist(P1fb,prob=T,xlab="phi1_b",main="")  
hist(P2fb,prob=T,xlab="phi2_b",main="")
```